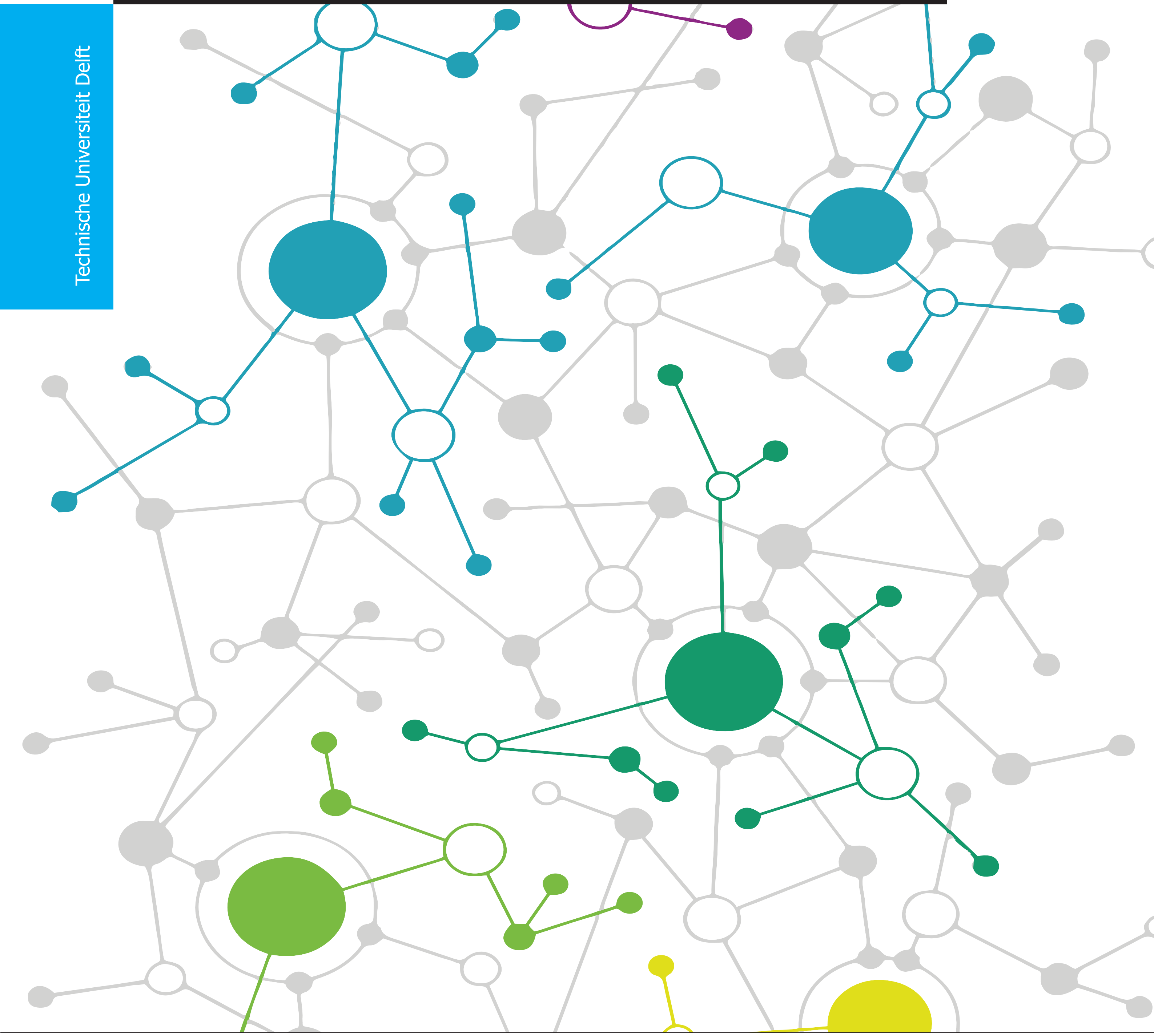
Resilience in Information Centric Networks and the Analogy with Human Collaborative Networks

Eline Neeltje Rietberg

Technische Universiteit Delft



TU Delft Delft University of Technology

Challenge the future

Resilience in Information Centric Networks and the Analogy with Human Collaborative Networks

by

Eline Neeltje Rietberg

in partial fulfillment of the requirements for the degree of

Master of Science
in Applied Mathematics

and in partial fulfillment of the requirements for the degree of

Master of Science
in Science Communication

at the Delft University of Technology,
to be defended publicly on Wednesday June 21st, 2017 at 11.00h.

Student number:	4114922
Project duration:	July 11th, 2016 – June 21st, 2017

Thesis committee Applied Mathematics

Chairman:	Prof. dr. ir. K. I. Aardal,	TU Delft, Applied Mathematics, Optimization
1st supervisor:	Prof. dr. ir. R. E. Kooij	TNO & TU Delft, Intelligent Systems, Netwerk Architectures and Services
2nd supervisor:	Prof. dr. ir. H. van der Berg	TU Delft, Science Education and Communication

Thesis committee Science Communication

Chairman:	Prof. dr. M. de Vries,	TU Delft, Applied Mathematics, Optimization
1st supervisor:	Dr. M. C. A. van der Sanden,	TU Delft, Science Education and Communication
2nd supervisor:	Drs. C. Wehrmann	TU Delft, Science Education and Communication
3rd supervisor:	Prof. dr. ir. R. E. Kooij	TNO & TU Delft, Intelligent Systems, Netwerk Architectures and Services

Preface

Over the past year, I have worked on my two master theses, of which the final result is in front of you. It is my final exam after seven years of studying at Delft University of Technology. These seven years were full of inspiring, challenging and insightful moments, and studying at this university has shaped me into the person I am today.

I have performed two researches that are linked through a shared theme; Information Centric Networking. This overarching theme of the two researches has come to life in an interesting and interactive meeting with Rob Kooij and Maarten van der Sanden. Together, we have formulated the two research directions.

I would like to thank TNO for the graduation position they have offered me. Furthermore, there are many people who have contributed to the successful completion of my study program, and my thesis in particular. First of all, I want to thank my two thesis committees, and Maarten and Rob in particular. The enthusiasm and curiosity of Rob has greatly influenced my curiosity for the mathematical part of the thesis. I have found working with you highly inspiring. Maarten has taught me to trust in the creative process of doing research. Without your supervision, I would have never been able to trust that the outcome of the research would be satisfactory, even though the path towards the outcomes was unclear from time to time.

Secondly, I would like to thank my parents. In the past seven years, they have been of great support. They have always provided me with the freedom and support to change my mind, to choose my own path, and to stay close to myself. I believe that one of the most important things to be a successful human being is to have the ability to think independently and to have the tools and skills to be autonomous. These are skills I have definitely learned from you, and no university could ever do so.

Furthermore, I want to thank Bram and Marlien for their critical reading, and their useful feedback on the thesis in progress, and the people at the 'afstudeerhok' for their daily support, both practically and mentally. Bart, your creative and enthusiastic help with the creation of the first experiment was not only very helpful, but it has also motivated me to a great extend. Bart, David, Frouke, Martine, Matthijs, Richard, Sanne and Tim, thank you for participating in the experiment. Daantje Derks, Raquel Fernandez and Tom Postmes have been so kind to provide me with input on my methodology. Particularly, I would like to thank Daantje Derks, who has spoken with me multiple times, and really helped me with her open and constructive thoughts.

Lastly, Luuk has helped me al lot with structuring my thoughts, fixing technical issues, and channelling my moods and insecurities into something productive again. Together, we are ready for new adventures, wherever the wind might take us.

Equipped with the knowledge gained, the life lasting friendships that have their roots in Delft, and the critical but open mind that I have developed over the past seven years, I feel more than ready to broaden my view and find a way to apply all this in a way that suits me.

I hope you will enjoy reading this thesis, and that you will celebrate my graduation with me!

Eline Neeltje Rietberg
Delft, June 2017

Summary

In this thesis, we look into two different researches with the same basis; information centric networking. Enhanced resilience is one of the often mentioned advantages of this paradigm shift in internet networks. However, this resilience is not quantified in literature so far. In this thesis the information resilience of hierarchical ICN topologies is analytically approached, allowing us to quantify the enhanced information resilience. Furthermore, understanding information sharing in human collaborative networks is shown to be very complex. The caching mechanisms in ICN are very suitable for building up an analogy with human collaborative networks, to increase the understanding of information sharing in human collaborative networks. We build up the analogy, and investigate the notion of resilience in a human collaborative network through 3 consecutive experiments. The aim of this part of the research is to enhance insights in human collaborative networks from a cybernatic point of view.

By using a simulation engine, called Icarus, we have been able to simulate ICN behaviour. By simulating link failures, and investigating the extent to which information requests can still be delivered to the end users of the ICN, the information resiliency can also be simulated. Furthermore, we have constructed a two step approach to analytically determine this information resiliency. In the first step, the so called cache hit ratio has been approximated. In the second step, the probability of an existing path is being computed. Combining these two results in a remarkably good approximation of the information resilience of hierarchical ICNs.

For the analogy, a systematic literature research was performed, to determine 7 factors that influence information sharing in human collaborative networks. These seven factors serve as a basis to interpret the results upon. Furthermore, we measured the seven factors at the participants of the experiments. We have disturbed our human collaborative network in the experiments by taking one of these seven factors in literature as guiding. The data obtained in the experiments have served as a basis to make a connection to the ICN simulator. The connection established show that the analogy is valid.

The results of the experiments (and its connection to the ICN simulator) allow us to describe the analogy in great detail, and find out that measuring resilience in human collaborative networks is different in nature than it is in ICNs. Hence, we find another boundary of the ICN. With the gained knowledge and insights from both researches, we can use the analogy to develop a different view on certain aspects of both research fields, leading to a new proposal of a caching strategy for ICN.

The results of this research are twofold. Firstly, we are able to analytically quantify the (enhanced) information resiliency in hierarchical ICN networks. The analytical model developed performs well for the hierarchical networks considered, and shows (among other things) that for our hierarchical realistic network (Figure 6.2), with a cache of 5% of the content population, a population build from Wikipedia traces (size 1000), and a LCD strategy, the percentage of successful requests increases from 43% to 55% when 20% of the links is broken (Figure 6.8).

Secondly, the research has resulted in an extensively investigated analogy between human collaborative networks and ICNs. With the help of 7 selected factors from literature as well as the experiments, we have substantiated the analogy and found some limitations in the analogy. With finding the essence of the analogy, we find out what items between human collaborative networks and ICNs are comparable, and how they can be compared. This essence was the starting point of the experiments. In the experiments, we have seen that we can trace the spread of human information sharing, and compare it to the spread in an ICN. Furthermore, we can proceed doing so when we disturb the network.

This analogy allows for a way of thinking, that could ultimately lead to a decision support system on the composition of collaborative teams. By using the analogy as a way of thinking, we came up with a suggestion for a new caching strategy in ICN, based on information sharing of humans in a collaborative network.

We have submitted a paper to the ACM ICN conference 2017 in Berlin for the information resilience research in ICN. At the moment of writing, the paper is still under review. The paper elaborates on this two step analytical approach and the quantifications of the enhanced information resiliency for different configurations of ICNs.

Contents

List of Abbreviations

API	Application Programming Interface
CDN	Content Delivery Network
DSLAM	Digital Subscriber Line Access Multiplexer
EMA	Exploratory Modelling and Analysis
ICN	Information Centric network/ networking
IoT	Internet of Things
IP/MPLS	Internet Protocol/Multiprotocol Label Switching
IPTV	Internet Protocol Television
ISP	Internet Service Provider
KPN	Koninklijke PTT Nederland
LCD	Leave Copy Down
LCE	Leave Copy Everywhere
LFU	Least Frequently Used
LRU	Least Recently Used
RAP	Robust Adaptive Planning
SEC	Science Education and Communication
SNA	Social Network Analysis
TNO	Nederlandse Organisatie for Toegepast Natuurwetenschappelijk Onderzoek
TRA	Theory of Reasoned Action
VoD	Video on Demand

I

General Introduction

1
Introduction

Research is to see what everybody else has seen, and to think what nobody else has thought

Albert Szent-Gyorgyi

This thesis was written as part of the double degree program Applied Mathematics and Science Communication at Delft University of Technology. The research takes a common starting point, namely caching strategies in Information Centric Networks (ICNs), but develops in two different directions. For the mathematical part, the resilience of ICNs is being researched. For the science communication part, an analogy is set up between ICNs and human collaborative networks.

Since the two research directions belong to the same graduation project, various attempts for a sensible crossover between the two research directions have been made during the research. In this thesis, the two research directions are distinguished between where necessary, and are considered together where possible.

1.1. Information Centric Networking

As just explained, the starting point of both research directions is Information Centric Networking (ICN). Note here that ICN refers to both information centric networking and information centric network, depending on the context. In Chapter 2, we will take a closer look at ICN, its developments and its research challenges. In this section, the main ideas will be explained to allow the reader to get acquainted with the subject.

Over the past 25 years, Internet has evolved rapidly, from the first non-military email on the European continent (CWI, 2008) in 1988, till today, where even everyday objects interact with each other through the Internet (Internet of Things (Cavalcante et al., 2016)). Not only the amount of connections has increased drastically, also the behaviour of users has changed. According to Cisco (V.N.I, 2016), the share of Internet IP video traffic in the current global traffic is around 70%, but is expected to grow up to 82% by 2020 (V.N.I, 2016). This contributes greatly to the overall increase of Internet traffic. As Pan, Paul, and Jain (2011) state, "[…] it has become extremely difficult to support the ever increasing demands for security, performance reliability, social content distribution, mobility, and so on[…]".

As the use of the Internet evolves, a need for new Internet architectures to meet those new requirements arises. One of these new architectures is Information Centric Networks (ICN). In this architecture, content is routed through the network based on a content attribute (for example a name) rather than an address, as is the case in traditional IP and telephony-based architectures (Kurose, 2014). A great advantage of ICN is that content can easily be cached (stored) in the network, leading to a quicker and more efficient delivery of the content to the users (Xylomenos et al., 2014). How efficient ICNs are, depends largely on the choice where to cache which content; the so-called caching strategy. Many caching strategies have been developed, such as Leave Copy Everywhere (LCE), Leave Copy Down (LCD) (Laoutaris, Che, & Stavrakakis, 2006) and ProbCache (Psaras, Chai, & Pavlou, 2012). These caching strategies will be discussed in Section 2.4.

Another expected advantage of ICN is the enhanced information resilience. However, relatively little is known about how to quantify this enhances resiliency. Researchers have proposed some adjustments to the ICN to become more resilient (Tyson, Bodanese, Bigham, & Mauthe, 2014) (Rak, Niedermayer, Papadimitriou, & Romero, 2016), but the potentially increased information resilience is hardly quantified. Al-Naday, Reed, Trossen, and Yang (2014) state that "quantifying the benefits of this [ICN] approach to resilience is as yet unclear", followed by quantification for one very specific ICN architecture.

The aim of this research is to decrease that gap and quantify the possible enhancement of the information resilience in ICNs. Having a physical connection between the content and the end user is a first requirement for content to be available. Other technical aspects of the availability, such as the knowledge of how to route the cached information to the requester (as discussed in Rak et al. (2016)), are left aside in this research in order to focus on the mathematical aspects of resiliency, rather than on the network architecture side. Once the information resilience is being quantified, different caching strategies will be compared to assess their performance on information resilience.

1.1.1. Caching

As said before, caching is the phenomenon of storing copies of information in the network. The added value of a cache is found in the reduction of traffic, as the locality (being the average distance of the content to the end users) of the content increases when the content is cached. The first articles about (web-)caching date back to the end of the last century (for example (Breslau, Cao, Fan, Phillips, & Shenker, 1999)). Since then, a lot of theoretical research has been done, and some small ICNs are being implemented. However, these implementations remain small in size and/or application. How effective a caching system is, depends on many things. We will briefly discuss the most important ones.

First of all (and as mentioned above), the effectiveness of a cache depends largely on the caching strategy. Deciding where to leave a copy of which content has a large impact. We will return to different caching strategies in Section 2.4. Secondly, the distribution of the content in the networks influences the effectiveness of the cache. Often, a Zipf distribution is being considered, as many authors (for example (Breslau et al., 1999)) have shown that the content of the internet follows this distribution. The (mathematical) details on the Zipf distribution can be found in Section 3.4. Thirdly, the structure (topology) of the network influences the effectiveness of the caches. Lastly, the cache size has a large effect. The larger the amount of items that can be stored, the more effective the cache is.

We will mainly restrict ourselves to these four important factors, even though other factors might influence the effectiveness. To the best of our knowledge, these factors impact the caching effectiveness the most. In the remainder of the thesis, the four factors above will be varied, and the impact on the cache effectiveness will be measured. The impact is often measured by the cache hit ratio, which expresses the fraction of content requests that has been dealt with by a cache. An example of a measurement of the cache hit ratio can be found in Figure 5.1.

1.2. Human Collaborative Networks

The other branch of this research will focus on human communication networks within organisations. Within the communication field, network theory (Littlejohn & Foss, 2010) is used to describe groups of people that relate to each other in a certain way. Another way of looking at those communication networks is done by for example Onnela et al. (2007). They have analysed a large set of phone calls, to set up a network of human interaction, where each node would represent an individual.

Not only the shape of these networks is of importance; also the amount of communication between individuals is important (in other words, the strength of the links). Furthermore, the kind of information that is shared is of interest.

It is therefore interesting to investigate the information sharing in a human professional network. However, restricting ourselves to a psychological and/or sociological view would leave opportunities like social network analysis (Carrington, Scott, & Wasserman, 2005) aside. Social network analysis uses mathematical concepts like graph theory to understand and investigate social structures (Otte & Rousseau, 2002). Setting up an analogy between a human professional network and an ICN is another way to include mathematics in investigating information sharing in a human professional network. However, as we compare the two in an analogy, we are forced to also take the human factors into account.

It is possible to determine the topology (with its different functions (Stephenson, 2005)) of a network in the professional environment. The theory of ICNs and its in-network caching property will be used to compare the network topology with the spread of information through this network. As a result, we are able to compare the information spread through a human collaborative network with the information spread through an ICN.

In this research, a clear distinction will be made between knowledge and information. The reason for this is that, unlike IT-systems, human communication is always vulnerable for interpretation. Many different definitions of these two concepts exist. Alavi and Leidner (2001) state that "[...] knowledge is information possessed in the mind of individuals: it is personalized information (which may or may not be new, unique, useful, or accurate), related to facts, procedures, concepts, interpretations, ideas, observations and judgements". As the analogy between the human network and the ICN should be as clean as possible, the focus will be on information, as the personalisation part is not something that happens in an ICN either. It is worth noticing however, it's difficult - if not impossible - to completely separate knowledge and information, as humans tend to always personalize and interpret where possible. This will complicate the analogy. In Section 11.1, we will come back to this and elaborate on the distinction between knowledge and information.

1.3. Research Questions

1.3.1. Problem Statement

We have identified two problems that relate to each other. Firstly, little is known about the availability of information in an ICN when disturbances in the network occur. If we would be able to quantify the enhanced information resilience in ICNs for different situations, this can be taken into account in ICN simulations and analyses. Secondly, understanding human information sharing in human collaborative networks is shown to be very complex. If we are able to build an analogy between human collaborative networks and ICN, we can use the latter in thinking about information sharing in human collaborative networks. This should lead to better understanding of information sharing in human collaborative networks. Furthermore, aspects of human collaborative networks are applicable to ICN with the use of the analogy. This also enhances our way of thinking about ICN. It is therefore that the aim of the research is formulated as follows:

1.3.2. Aims of Research Project

1. Gain insights into the information resilience in ICNs for different caching strategies and different network topologies.

2. Gain insights into the information spread through, and information resilience in a human collaborative network. This will be done by setting up an analogy between an ICN and a human collaborative network, and by examining the possibilities and boundaries of the analogy.

1.3.3. Research Questions

The aims described above lead to to the following two research questions:

1. To what extent is it possible to analytically compute the availability of the content in an information centric network when breakdowns occur?

2. To what extent can the information spread in a human collaborative network and an ICN be applied to each other, to enhance insights into information spread and information resilience?

The two questions relate to each other as shown in Figure 1.1. The first part of the research aims to prove and quantify the enhanced information resilience in an ICN compared to a 'regular' internet network. The second part of the research tries to identify the boundaries and possibilities of an analogy between the human network on the one hand, and an ICN with its caching capacity on the other hand.

Within the communication research, the focus will lie on setting up the analogy. This is done in two phases. First of all, the information spread is measured in a human professional network. Secondly, the experiment is being repeated while the same network is being disturbed. We will come back to that in Part III.

1.4. Relevance

Little is known about the enhanced resilience of ICNs over networks without caching. Investigating this enhanced resiliency is of added value to the scientific and theoretical knowledge on ICNs.

It remains very hard to predict or even understand human behaviour in cooperation settings, especially when distrubances in the composition of the group occur. One way of explaining the human behaviour is by looking at each individual, and find out what moves them to act the way they do. This is a cybernetic approach. In the cybernetic approach, the system of parts that influence each other is considered. The ultimate goal is to achieve a balance or a change (Littlejohn & Foss, 2010). This cybernetic approach is for example followed by the theory of reasoned action (Fishbein & Ajzen, 1977). Unfortunately, these approaches have shown to be too complex when trying to understand a whole system of human behaviour. It is therefore that we should consider the whole system, and try to find out how the system as a whole will behave. Or, as M. van der Sanden and Flipse (2016) put it: "In science communication many professionals and researchers recognize the same gap between theory and practice, and thereby implicitly the need for system thinking and its whole/part thinking. Not to say that science communication is in a theoretical crisis, but in a strong need to make system thinking explicit."

Grasping these systems is difficult, and help from other fields is needed to build models or tools that help us evaluate these systems. In this research, we aim to do that, by comparing the information spread and the resilience in human collaborative networks with the information spread and resilience in ICNs. If the two are comparable up to a high level, behaviour in the ICN could help explaining behaviour in the human collaborative network, and vice versa.

1.5. Relation Between Two Projects

The relation between the two projects is illustrated in Figure 1.1. The impact of breaking the links in the ICN is subject of investigation in the mathematical part of this research. The two green arrows depict the comparison made between the caching behaviour in human collaborative networks and ICNs. This is what we research in the science communication part. In the comparison between the two networks, literature from the sociology, psychology and communication field will be used. Attention will be given to this in Part III of this research.

1.6. Research Approach

In this section, we will elaborate on the methodology of the research. Furthermore, we will make some observations regarding reliability, validity and the ethics of this research.

1.6.1. Study Design, Data Collection and Limitations

For the information availability part, the simulation package called 'Icarus' (Saino, Psaras, & Pavlou, 2014) will be used. This simulator allows comparing different caching strategies and it has been designed in order to be easily extensible. With Icarus, caching strategies are compared measuring cache-hit ratios as well as server-hit ratios. However, the version of the simulator used does not natively support metrics and evaluation scenarios specifically targeted at studying information availability in ICNs. Therefore, these will be added to the simulator throughout the course of this research[1].

[1]Supposedly, the recently released version of the simulator (released after the start of this research project) does support some metrics.

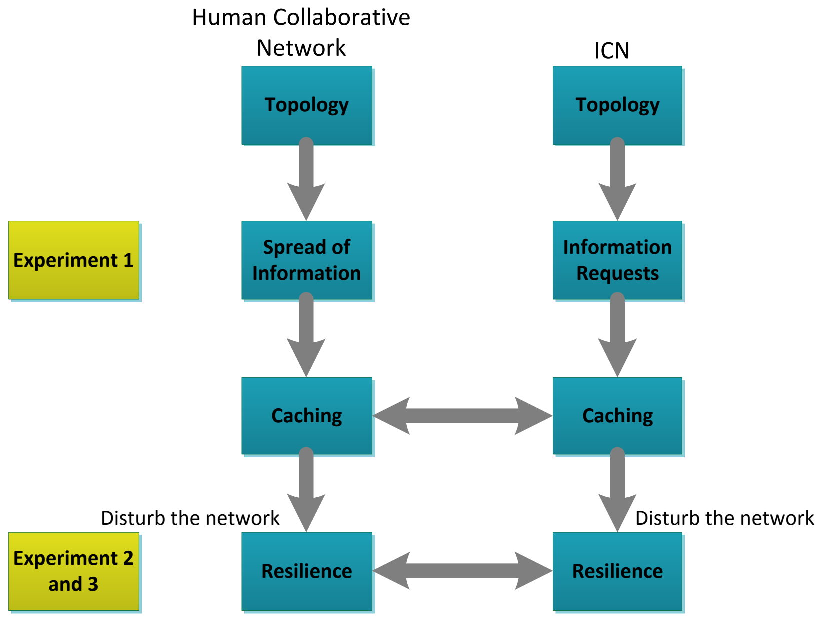


Figure 1.1: The crossover between the two projects

For example, the ability to break down some links during a simulation run will need to be introduced. This way, it is possible to analyse which nodes are connected to each other. As a result, it can be inferred what content is available in which parts of the network. This experiment can be carried out using different values for the fraction of the links that have been removed. When comparing the graphs obtained this way, one can assess which caching strategies are most effective in increasing the information availability under breakdowns.

During the course of this research, both the time component and the caching component will be stopped before starting the analysis. This means that no further cache replacements will take place once the analysis starts. In addition, no download times, recovery times etc. will be taken into account.

It has been noted that hostile attacks on the nodes or random failures have a different impact on the network (Albert, Jeong, & Barabási, 2000). From this study, we know there is a large difference in the network behaviour depending on the choice of the nodes to break down. However, the exact results are hard to predict for the resilience measure of information availability. We will limit ourselves to breaking down links at random. This means that targeted attacks will not be part of the research.

For studying the information spread in a human collaborative network, systematic literature research has been performed, and an experiment will be set up. The systematic literature review provides us with the factors from literature that influence human information sharing behaviour in a collaborative network. Furthermore, we use the theory of ICN to set up the experiment. Within an ICN, content is shared in a specific way, and within a simulation, it is also traceable. What is specific in an ICN, is that content is sent from A to B, but is cached in one or more different nodes it traverses on the way from A to B. The behaviour, location and number of caches can be tracked in an ICN simulation. As a result, content is not only available in the source, but also in other nodes within the network.

Within a human collaborative network however, it is not so easy to track this information. Besides, human components like interpretation make this tracking even harder. On top of that, according to Ipe (2003), the decision to share information also depends on motivation. What could be similar though, is the way caches are used in the network. These influencing factors will be researched in Chapter 11.

We set up an experiment where different roles (hubs, gatekeepers, pulsetakers, and possibly more/different ones) in the network are analysed. This has been done before (for example by (Cassi, Corrocher, Malerba, & Vonortas, 2008)), and will be applied to a group of professionals working together in this case. This has largely to do with different kinds of centrality[2] that can be measured in the network. Furthermore, some factors such as willingness to share and trust are indicated before the experiment. Next, we trace some parts of information with different sources. It becomes possible to analyse if the behaviour in a human network is somewhat similar to the behaviour within an ICN. As a next phase, a correlation between these specific roles in the human network and the roles of the nodes in an ICN (router, source, receiver etc.) is sought.

In the second phase, a disturbance in the network is realised by removing some people from the group. These people are selected on the basis of the factors that we measured before. These factors will be selected from literature, and will be discussed in Section 11.6. One measure will be picked, and the influence of the others will be discussed in the discussion section (Chapter 16).

1.6.2. The Experiment

For a group of master students all team members will be asked to state the team members they socialize with the most. They can list any number of colleagues up to 25% of the total number of participants participating in the experiment. The limit of 25% is chosen to make sure that the network remains manageable. It could be argued that other limits could be used too, but in this particular case 25% seems to be more than enough.

This information can then be plotted in a directed graph. Differences between the centrality of the participants in the network (roles) can be found using mathematical analysis (social network theory). Next, all participants are asked to work together on an assignment. Besides, they get a personal assignment that they would want to share with others. By asking them afterwards who has told what to whom, the information spread can be measured. This allows tracking which information has reached whom, and whether this information has reached the end user from a source or a cache.

1.6.3. Reliability, Validity and Ethics

There are no foreseen problems in the ethical part of the research. The privacy of the participants in the experiment should be taken into account though. It might be that participants do not want the network to be shared with other colleagues, as it shows which work and or social contacts are considered important. Furthermore, some questions are asked to the participants before the experiment, to map the factors that influence information sharing according to literature. These questions are slightly uncomfortable sometimes, and it might make the participants rethink some of the opinions they have about each other. This is the reason why a notification was made to the ethical committee of Delft University of Technology.

The Mathematical part is fully replicable (when we take sample sizes and number of simulations large enough). Hence, the research is reliable. For the research to be valid, we need the parameters to be as realistic as possible. For now, real Internet data (usage, number of nodes, number of requests per minute, size of the content population etc. etc.) are not available to the researchers. This is the reason why some time will be spent on acquiring the best (variety of) parameters. For this, expertise of fellow researchers as well as real data (small bits that are available) and literature will be used. Furthermore, for validation of the mathematical analytical model, the 1000 most visited Wikipedia pages on a certain date were retrieved from the Wikipedia API. These figures were used as input data to the model.

For the communication part, the outcomes of the research depend highly on the group of people researched. It is therefore not replicable. If the results seem to be promising, more groups need to be researched to get statistically sound results. The methods of mapping the network for different settings have been used more often. This would suggest the method to be quite reliable and valid. No problems in validity of the research have been foreseen. However, triangulation is difficult in this research, meaning that we cannot easily validate our results.

[2]Centrality is a measurement in a network that says something about how central a node is in the network. Different ways of measuring centrality reflect different roles in the network. We will elaborate on that in Section 11.6.1.

1.7. Outline of the Report

This report is divided into four parts. The first part consists of a general introduction to the two research directions. As you have just read, Chapter 1 introduces the research topics and the research questions. Furthermore, a more specific introduction to ICNs will be given in Chapter 2.

In Part II, the mathematical core of the research can be found. This part starts with a more specific introduction to the mathematical part, some background and literature considerations in Chapter 3. In the Chapters 4, 5, 6 and 7, the results of the research can be found, before moving on to the conclusion (Chapter 8) and discussion (Chapter 9).

Part III deals with the communication part of this research. After an introduction (Chapter 10), an extended literature review can be found in Chapter 11. A chapter on analogies (Chapter 12) leads us towards the methodology and design of the experiments in Chapter 13. This then leads to the results of the experiments in Chapter 14. Finally, the conclusion can be found in Chapter 15 and the discussion in Chapter 16.

In Part IV, we elaborate on the the synergy between the two researches. In Chapter 17, we use the analogy established in Part III as a way of thinking. We show the added value of the analogy there. In Chapter 18, a reflection can be found on doing two studies in parallel. Also, the way the two researches enhance each other is being elaborated on. What could the mathematical research community learn from the science communication research, and vice versa?

2

Introduction to Information Centric Networks

Information is not knowledge
A. Einstein

Information centric networks (ICNs) have been briefly discussed in Chapter 1. As ICNs are central in both researches, this section is dedicated to familiarize the reader with ICNs. In order to do so, Section 2.1 explains the concept of ICN, and its development. Section 2.2 compares the ICN with content delivery networks and peer to peer networks, to explain what makes ICN unique. After we treat the added value of caching in an ICN in Section 2.3, we elaborate on caching strategies in Section 2.4, and cache replacement strategies in Section 2.5.

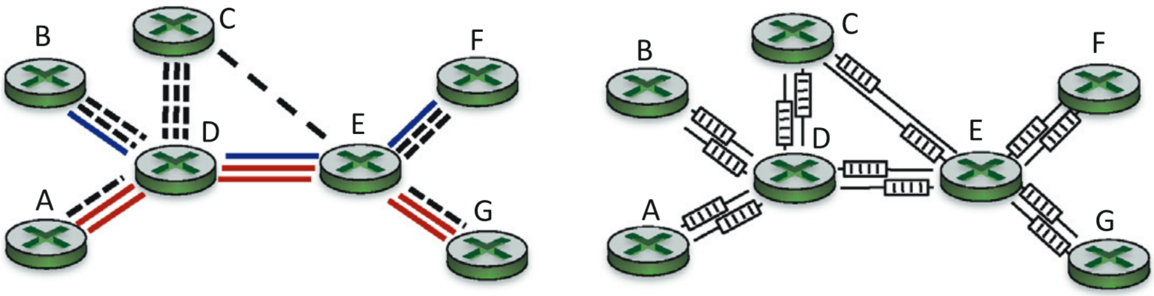
2.1. What is Information Centric Networking?

Since the development of communication networks, a major paradigm shift has been established. Where telephone networks one hundred years ago were so-called circuit switched, we nowadays communicate through packet switched networks. In a circuit switched network, a dedicated connection is established between the two users that want to communicate. This connection is then unavailable for all other communication while in use[1]. In Figure 2.1a (Kurose, 2014), such a network example can be found. In this example, three calls are being made: one from B to F, and two from A to G. An extra call from A to F would now be blocked (if no dynamic alternating routing schedules are used), as there is no connection available along their shortest path (all connections between D and E are occupied). A serious drawback of this way to communicate is the number of links needed if the network grows both in size and density. The Erlang distribution (Erlang, 1909) allows us to compute the probability of a call being blocked.

In the beginning of the 1960s Baran came up with the idea of segmenting information into packets before sending them. Shortly after, Kleinrock (1961) starts to publish about the more effective information flows through a packet switched network instead of a circuit switched network. In a packet switched network, the information is being wrapped into packages, that will all individually have to find their way through the network. Each router (node D and E) that receives a package forwards it to the next router in the right direction. However, it is still necessary that two hosts make a connection. This is shown in Figure 2.1b. The current internet network still largely is packet switched. However, as video traffic becomes more and more popular, receiving the content becomes more important than connecting to a specific terminal. Or, as Kurose (2014) puts it, “[...] what a person wants, rather than where it is located, is what matters most”.

This is where ICNs start to play a role. In ICN, it is no longer required to have a connection between two hosts. Where the content is located becomes less relevant, but the speed and distance to

[1] Of course, multiple connections between users might exist.



(a) An example of circuit switching (Kurose, 2014) (b) An example of packet switching (Kurose, 2014)

Figure 2.1: Two paradigms in communication networks

the user become more important. In an ICN, "[...] information can be named, addressed and matched independently of its network physical location [...]". (Rak et al., 2016) According to Carofiglio, Morabito, Muscariello, Solis, and Varvello (2013), the current research into ICN aims to "architect a network that automatically interprets, processes, and delivers content (information) independently of its location."

In an ICN, all requested content may be stored in multiple places at the same time, as each node in the network is equipped with a so called cache that can store information. Having the 'copies' in the network allows end users (requesters of information) to retrieve their information from multiple places (both from caches and sources). In the type of ICNs that we will consider, each item of information still has one source, which will never lose its information.

Once an end user requests information, each node along the path to the source checks for the requested information in its own cache. If the information is available in the cache, the information is sent from the cache to the end user. In the process of sending information to the requester through the network, this information can be cached along its paths. The so-called caching strategy determines the specifics of how this is achieved. We will come back to that in Section 2.4.

An example of an ICN can be found in Figure 2.2, which is retrieved from (Kurose, 2014). Here, the same system as in Figure 2.1 with 7 nodes is being considered. The content population consists of three items, named 1,2 and 3, of which respectively F, G and C are the sources. Each node in the network can store two pieces of information (note that this is the cache size).

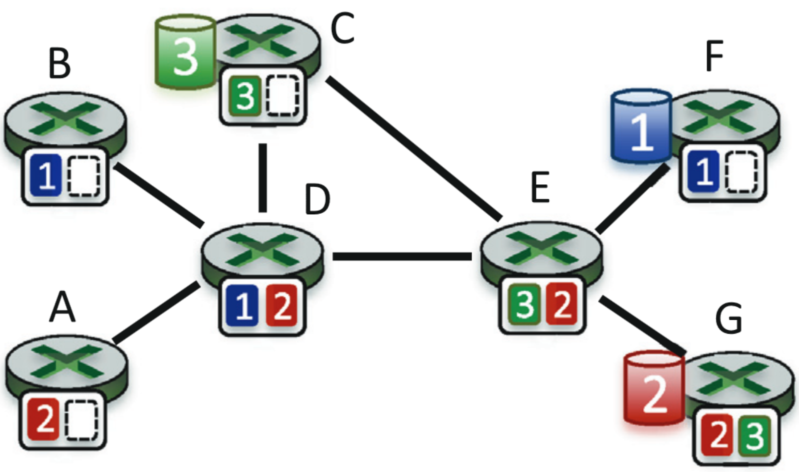


Figure 2.2: An example of an ICN. (Kurose, 2014)

Now, node B requests information item 1, node A requests information piece 2, and node G requests information piece 3. This happens sequentially. The caching strategy here is Leave-Copy-Everywhere

(LCE), which means that the information is cached in every traversed node. Once a cache is full (which is the case with node E after the first two information requests), a cache replacement strategy is needed to decide which information piece is being removed from the cache once a new piece of information appears. The most common of these is the Least-Recently-Used (LRU) strategy, which is also applied in this case. This means that, once information piece three requests a place in the cache, information piece 1 is replaced by item 3, as item 1 has been in use least recently.

2.2. Content Delivery Networks and Peer to Peer Networks

Although ICNs are new, and up to this point mainly theoretical in nature, certain elements of ICNs are not entirely new. Especially web caching is already widely implemented in the current internet network. In this section, we first discuss content delivery networks (CDNs) and proxies (which are like CDNs) and afterwards peer to peer networks.

CDNs and proxies are widely used in the current internet, and both work as a caching mechanism. They serve a different purpose though (Vakali & Pallis, 2003). Proxies are used from the side of Internet Service Providers (ISPs), and cache the most frequently (or most recently) used content. CDNs on the other hand, are used by content providers to cache specific information with the purpose of serving a large amount of users and serving the information faster. We focus on the latter here, as they come much closer to ICNs than Proxies do.

CDNs have two key components, namely request routers and surrogates. A surrogate acquires the content from the cache server, while a request router assigns an end user request to a surrogate. The surrogate can find the closest cache of the requested information (Lee, Jeon, Yoon, & Song, 2012). As a result, when an end user requests information that is in a CDN, the CDN itself will redirect the request to the closest CDN server that contains a copy.

What makes a CDN fundamentally different from an ICN? CDNs are a way to make information central in a network architecture that is built to provide host-to-host communication (Carofiglio et al., 2013). ICN on the other hand, is a new architecture that is designed to deal with information delivery rather than host-to-host communication. Where the caching capacity of CDNs is limited to designated (large) CDN servers, all routers in an ICN would have a (much smaller) caching capacity. Hence, the caching capacity changes from a layer of servers over the current network, to a property of the network itself.

Another difference between CDNs and ICNs is that CDN are application specific, so they are only able to store content for specific application (e.g. as specified by the content provider). In the case of ICN, every piece of information may be stored. This difference does not apply to proxies.

Where CDNs are always working with an overlay structure of caches (dedicated caching servers), the caching capacity can also be allocated at the end users itself in a peer to peer caching network (P2P web cache) (Iyer, Rowstron, & Druschel, 2002). Its concept lies close to the concept of using proxies by ISPs (Lua, Crowcroft, Pias, Sharma, & Lim, 2005). It is more similar to ICN than CDN or proxies in the sense that the amount of caches is larger, and the caching capacity per node is smaller. However, this is still a way of caching in current internet models instead of building a new architecture as proposed with ICN. Also, there is no in-network caching, but rather a caching capacity at the end users.

2.3. The Added Value of Caching

Caching has a number of advantages. The first couple of advantages relate to the efficiency of the network. As on average the content is closer to the end user, bandwidth consumption and traffic is reduced (Vakali & Pallis, 2003). For the same reason, latency (the time it takes from request to delivery) and the user-perceived delays reduce, as the information has to travel over a shorter distance. Davison (2001) mentions that also the load on the source servers reduces when part of the requests do not reach the source server but one of the caches.

The second set of advantages deals with the performance of the network in case of disturbances. The resilience increases, as multiple copies of the same information are available in the network. Resilience in this case is broad, but is mainly related to the question: how well the network still performs in case of a disturbance? It turns out to be hard to quantify this increased resilience. In this thesis, we make an attempt to quantify this resilience.

2.4. Caching Strategies

The strategy that decides which content should be cached where, is called the caching strategy. Many strategies are possible, but we will discuss four different strategies here, namely Leave Copy Everywhere (LCE), Leave Copy Down (LCD), ProbCache (caching on the basis of probabilities) and BidCache. The latter has been developed by TNO, and uses different parameters to let each node 'bid' on getting the privilege to cache the information traversing that node.

Leave Copy Everywhere

The simplest and also most aggressive way to decide where to cache information is by simply storing it everywhere. In the LCE strategy, a copy of the information is cached in every node along the path from source to end user, until it is replaced. Although very simple in understanding, the strategy is not very effective. Laoutaris et al. (2006) show that there is no need to keep a copy in every node along the path, but that a subset of nodes suffices and is in fact even more effective. It is therefore that they propose a strategy that only caches in a subset of nodes: Leave Copy Down (LCD).

Leave Copy Down

In the LCD strategy, not all nodes downstream get a copy of the information, but only the direct neighbour of the node where the information was retrieved. This node of retrieval can either be a source, or a node that already contained the information in its cache. Note that a copy is cached in the neighbouring node, and the content remains available in the node where it was retrieved from.

In this strategy the information slowly lowers down from the source. More importantly, the information that is requested more often is situated lower in the hierarchy than the information that is requested less often. According to Laoutaris et al., the reason why LCD outperforms LCE is twofold. First of all, in one request, information is only stored in one cache, therefore replacing information in only one cache. Secondly, the number of information replications along the path are smaller than with an LCE strategy. If one uses an LCE strategy, and obtains information from a source, then it caches this information along the entire path to the user, which leads to a decrease in the diversity of information cached along the path.

ProbCache

Much more complex is the strategy ProbCache (Psaras et al., 2012), where the decision of whether to cache is based on a probability. This probability is static, but different per node. Per node, an estimate of the caching capability of paths that include this node is calculated, as well as a cache weight. By multiplying these two, the probability that an item traversing the node will be cached is computed.

We will briefly elaborate on the two factors that determine this probability. To be able to calculate the estimate of the caching capability of paths, two values are needed. First of all, the time since interception (TSI) value is given by the hop distance from the end user requesting the information to the source. Secondly, the time since birth (TSB) factor is equal to the hop distance from the server that contains the information to the considered router. The caching capability of the paths is now equal to

$$CacheCap(x) = \frac{\sum_{i=1}^{TSI-(TSB-1)} N_i}{N_x}, \tag{2.1}$$

where N_i is equal to the caching capacity in node i. This factor which we call CacheCap is named 'times in' by Psaras et al. because it estimates how often the path can afford to cache the item.

Next, Psaras et al. argue that by deciding where to cache, a weight should be added, that increases the probability of caching close to the end user, and decreased the probability of caching far away from the end user. This weight is equal to the fraction of the TSB and TSI value. Hence

$$CacheWeight(x) = \frac{TSB}{TSI}. \tag{2.2}$$

Psaras et al. furthermore show that this strategy outperforms LCD and LCE for specific parameters.

BidCache

The last caching strategy we will discuss here is BidCache. It was developed at TNO in 2016 (Gill, D'Acunto, Trichias, & van Brandenburg, 2016). ProbCache used a strategy to compare one node with others (by the TSI and TSB factors). BidCache also does that, but on a basis of an auction. Each node has to bid on caching a certain content, and the highest bidding node wins the auction, and caches the content.

How a bidding to the auction is computed can be adjusted to the objectives of the network. Examples of possible bidding parameters (all accompanied with their own weight of importance) are the cache size, the hop count and distance to the source and/or the end user.

2.5. Cache Replacement Strategies

The cache replacement strategy prescribes what content is deleted from the cache when space needs to be freed to cache a new item of information. There are many possible replacement strategies, of which a survey can be found in (Podlipnig & Böszörmenyi, 2003). We discuss the two most popular replacement strategies. The most common replacement strategy is Least Recently Used (LRU). In this strategy, the item that is been used least recently is removed from the cache. This strategy will be used in the majority of the examples in this thesis.

The other replacement strategy that is very common is the Least Frequently Used strategy. In this strategy, the item that has been requested from the cache least frequently, is removed. A serious drawback of this strategy is that the items that have been added to the cache recently (and therefore have not been requested very frequently) are removed first too. There are multiple options to adjust to this problem, such as ageing principles (Arlitt, Cherkasova, Dilley, Friedrich, & Jin, 2000).

In the next part, attention will be given to the information resilience in ICNs. For those readers mainly interested in the communication research, we suggest to move to part 3.

II

Resilience in Information Centric Networks

3
Background and Contribution

In Chapter 1, we have stated the research question. Let us repeat it here:

> To what extent is it possible to analytically compute the availability of the content in an information centric network when breakdowns occur?

To answer this question we will first link the question to the concept of information resilience (Section 3.1). We relate it to robustness and compare network resilience with information resilience. Furthermore, this Chapter will give the necessary background information (Section 3.1 -3.3), and elaborate on our contribution to the scientific community.

The main network topologies used in this thesis will be introduced in Section 3.2. In Section 3.3 we pay some attention to hierarchical structures, which are the kind of network structures that we will take into consideration. The simulator used will be introduced in Section 3.4 and the distribution behind the popularity of the content population in Section 3.5. With knowledge about these concepts, the reader will be introduced to the mathematical contribution to the scientific world by the end of the chapter. The remainder of Part II aims to elaborate on this mathematical contribution, in order to answer the research question.

3.1. Information Resilience

In this section we will pay attention to the definition of information resilience, which we want to quantify in this thesis. In network theory, the notion of robustness is often used. Network robustness is the extent to which a network can deal with perturbations imposed onto the network (Albert et al., 2000). Barabási (2013) shows that this robustness depends on the size and topology of the network among other things.

Percolation theory is closely related to this (Barabási, 2013). This theory says that if a fraction of the nodes is removed from a network, this network stays largely connected for a long time. After the threshold is reached, the network suddenly falls apart in many components. This phase transition can be quantified with percolation theory. The effect of node failure in a network depends largely on which nodes are broken. It is therefore that in network robustness calculations we distinguish between the effects of random failures and attacks.

Gao, Barzel, and Barabási (2016) define the resilience of a network as "a system's ability to adjust its activity to retain its basic functionality when errors, failures and environmental changes occur." Cohen, Erez, Ben-Avraham, and Havlin (2000) have investigated this resilience for the internet network.

Unlike the items above, we are not interested in computing the resilience of the network itself, but we are interested in the extent to which the network can still deliver the requested information to the end users in case of a disturbance. Hence, we are not interested in the network resilience, but in the information resilience of the network.

3.1.1. How to Measure Information Resilience?

As discussed before, information resilience is the measure of interest. There are several possible measurements of this resilience. One of these measurement methods is discussed in Sourlas, Tassiulas, Psaras, and Pavlou (2015). When disturbances occur in the network, information can get lost with a certain probability. This mechanism is called absorption. In their paper, Sourlas et al. (2015) show that there is a lower bound for this probability, and a corresponding time to it. As a measure of resilience, they use the percentage of absorbed items as well as the mean absorption time.

While this seems a good resilience measure at first, there are some drawbacks. First of all, the fact that information gets absorbed by the network, and thus the information gets lost, is not necessarily a bad thing. When information become redundant, there is no need to keep it 'alive' in the network. Imagine for example a sensor that measures the temperature every x minutes, and sends it to the network. What is interesting for most end users (requesters) is the temperature of the last measurement. All earlier measurements are redundant, and therefore do not necessarily have to be preserved in the network.

Let us turn back to the analogy with the human collaborative network. Also in a human network, information might become redundant, either because it is no longer accurate, or because there is no more need for or interest in the information. Therefore, measuring the absorption time and probability does not seem to fulfil our requirements.

So we search for another measure of resilience. One could investigate per end user what percentage of all information is still available to it. When taking the average over all end users, this indeed is another measure of information resilience. However, the objection made for the resilience measure of Sourlas et al. (2015) still holds here. If we consider all information, the redundant information is taken into account too.

It is therefore that we focus on a third way of measuring information resilience. It is similar to the measure suggested above, but instead of taking all information into account, a workload is taken into account. This workload could be either a generated workload, or a real workload from historic data. Hence, we define the information resilience as:

> The percentage of information requests that still can be delivered to the end users in the network, given a disturbance in the network

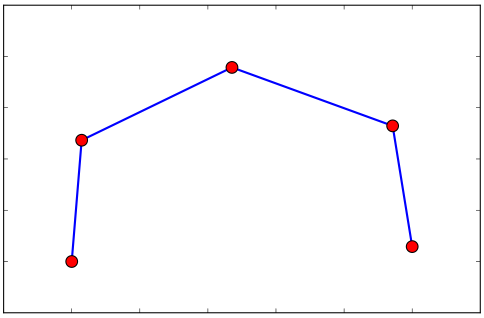
This disturbance can be an attack or a random failure. Furthermore, the failure could be in the links, the nodes, or even just the caches. In this thesis, we consider random link failure. Note that this measure is inspired by Gao et al. (2016), where the resilience is also expressed as certain behaviour given a failure of a fraction of the links (or nodes) in a network.

3.2. Network Topologies

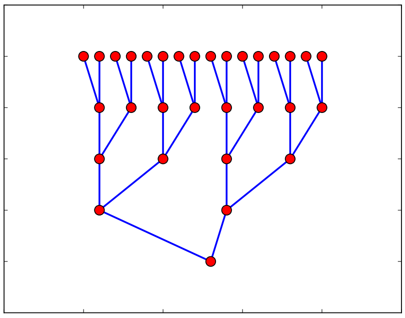
When analysing the information resilience in ICNs, a certain topology needs to be considered. In this thesis, a couple of topologies will be considered. First of all, a path topology is considered in some cases (Figure 3.1a). This topology consists of n nodes and $n-1$ edges, connecting the nodes in a single path. The path topology is far from realistic for an ICN, but as it is simple, the analytical results are easier to understand as well.

Secondly, the tree topology is used (Figure 3.1b). A tree topology consists of n nodes, with $n-1$ edges, forming no cycles (hence connecting all nodes to at least one other node). This is slightly more realistic than the path topology, but it still does not approach a real internet network. Also here, the analytical results are relatively simple, hence the consideration.

The third topology considered is the Barabási-Albert topology (Figure 3.2). The Barabási-Albert model is and example of how to generate a scale-free topology. A scale-free topology, as explained in (Barabási, 2013), is a topology of which the degree distribution of the nodes follows the power law,

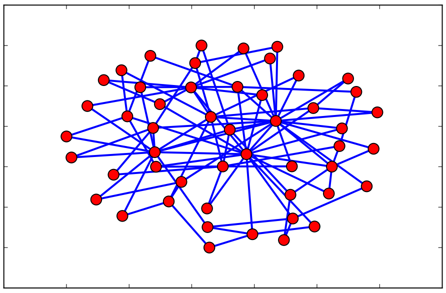


(a) A path topology

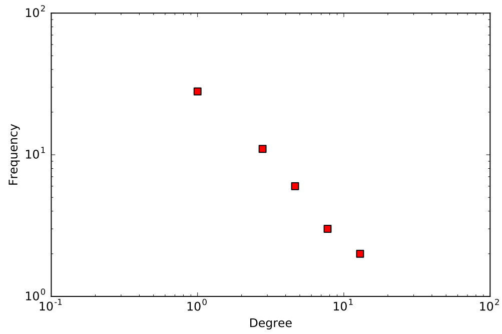


(b) A tree topology

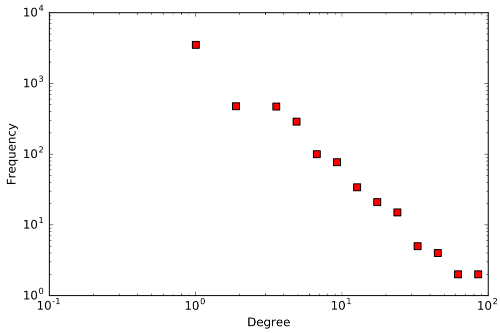
Figure 3.1: Topologies used



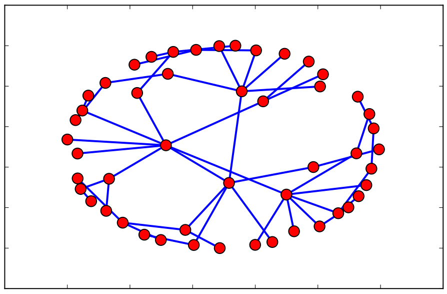
(a) A Barabasi-Albert Topology with 50 nodes



(b) The degree distribution of the topology in 3.2a



(c) The degree distibution with 5000 nodes



(d) A Barabási-Albert topology with m=1.

Figure 3.2: Barabasi-Albert topology

meaning that many nodes have a very small degree, and a few nodes have a very high degree. Hence, the degree distribution can be approximated by Equation 3.1.

$$p_k \sim k^{-\gamma} \tag{3.1}$$

Next, Barabási describes how to build a network that follows the power law. The network starts with an initial network, which is often fully connected. Then, the nodes are added one by one, and are connected to exactly m other nodes. This is however not done at random, but with so called preferential attachment (Barabási, 2013). This means that a new node connects to node i in the existing network heavily depends on the degree of node i, k_i. To be more precise, the probability of a new node to connect to node i is:

$$p(i) = \frac{k_i}{\sum_j k_j} \tag{3.2}$$

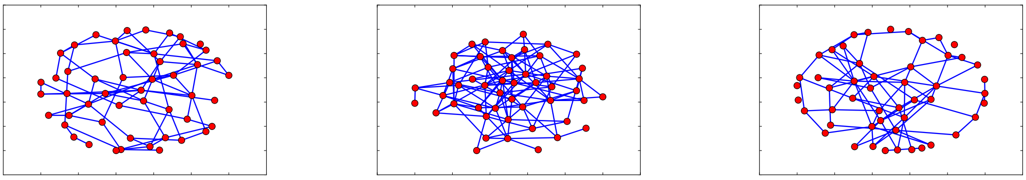
In Figure 3.2a we see a Barabási-Albert topology of 50 nodes. The initial network consisted of 4 nodes, each new node is connected to the network with two edges (m). In Figure 3.2b, the degree distribution is plotted too. As we see, the power law is indeed followed. To further clarify that latter statement, we have repeated the experiment with 5000 nodes in Figure 3.2c. Obviously, this does not provide any information any more when plotted into a similar figure as Figure 3.2a, but the degree distribution is still according to the power law.

What is important to notice, is that when taking $m = 1$, we obtain a network that looks a bit like a tree, but has an interior that is more fully connected. We see that in Figure 3.2d.

The last topology we will consider is the Waxman topology (Waxman, 1988). The Waxman topology considers n nodes that have a random distance $\delta \in [0, L]$ between each other. Furthermore, two other parameters $\alpha \in (0, 1]$ and $\beta \in (0, 1]$ need to be defined. Now, the probability of an edge between any two nodes is equal to

$$p = \alpha \cdot e^{-\frac{\delta}{\beta \cdot L}} \tag{3.3}$$

In Figure 3.3, a Waxman topology of 50 nodes is drawn for different values of α and β.



(a) A Waxman topology with $\alpha = 0.2$, $\beta = 0.3$
(b) A Waxman topology with $\alpha = 0.9$, $\beta = 0.1$
(c) A Waxman topology with $\alpha = 0.1$, $\beta = 0.9$

Figure 3.3: Examples of a Waxman topology

3.3. Hierarchical Structures

The analytical results will be limited to hierarchical networks. In such an hierarchical network, all nodes belong to a specific layer in the network. All tree topologies are hierarchical in nature, but cycles are allowed in hierarchical networks. An example of an hierarchical network can be found in Figure 3.4, which depicts a communication network.

In a hierarchical network, each layer can be reached from the end users. Furthermore, in the shortest path from an end user to a source, exactly one node of each layer is visited. This means that each node in a layer can be connected to nodes in the same layer, the layer above and the layer beneath the node itself, but not to other layers. Further more, it does not prohibit cycles. If links in the same layer would be prohibited, the graph would be bipartite.

Using hierarchical networks allows us in a later stage to analyse all nodes in a layer at once. This simplifies the analysis. We define a hierarchical network as:

> A network where layers can be defined in such a way that each shortest path from top to bottom of the network passes exactly one node from each layer.

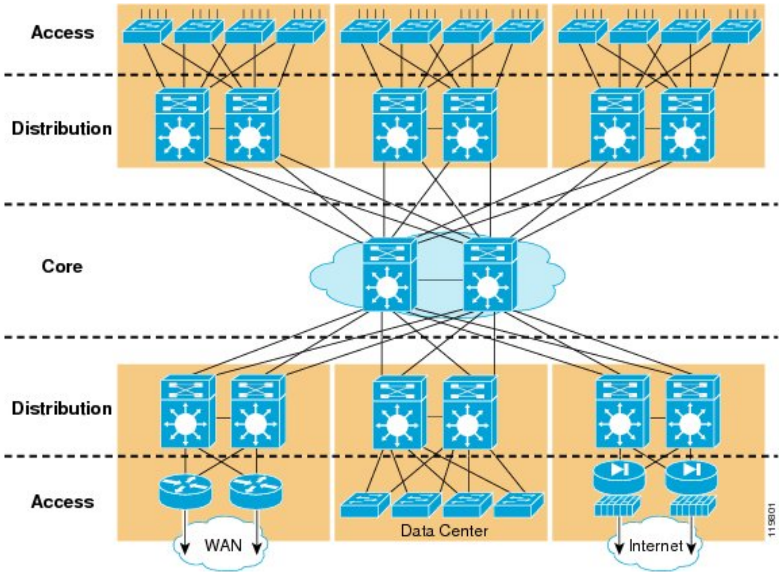


Figure 3.4: An example of a hierarchical communication network, retrieved from (cisco, 2008).

3.4. Icarus

Saino et al. (2014) have developed a caching simulator for ICNs, based on Python. The code is open source, and well documented. As a result, it is possible to add caching strategies or other elements to the simulator.

In the simulator, the first phase is to configure a file with all variables as discussed before. Examples of settings that can be varied with are the cache size, the topology, the caching strategy, the in-cache replacement strategy, the content population, the popularity parameter and the number of information requests that should be generated. In the second phase, the simulation runs, resulting in a number of figures, as well as some logging data.

For more information on the simulator, the reader is referred to GitHub[1].

3.5. Zipf Distribution

The Zipf distribution (Kingsley, 1949) is a special case of the power law. In the ICN simulation software Icarus, the Zipf distribution is used to generate a distribution of the probability that a specific content item is being requested by one of the end users. In other words, the Zipf distribution is used to simulate that certain content is more popular than other content. The extent to which some content is more popular than others is called the skewness of the content popularity. If the popularity is very skewed, certain content items are much more popular than others.

To compute this probability, there is a variable, called α, which is key. We consider a content population $i \in \{1, 2, 3, \cdots, n\}$, and define $k(i) = \frac{1}{i^{\alpha}}$. Now to obtain the probability that item i is being requested it is necessary to normalise such that the probabilities sum up to one. Hence, this probability is equal to:

$$p(i) = \frac{k(i)}{\sum k(j)} \tag{3.4}$$

Now, let us illustrate the importance and impact of α quickly with a numeric example. Suppose we have a content population of 4. In Table 3.1, three values of α are compared, being $\alpha = 0.1, \alpha = 1$ and $\alpha = 10$. The first and the last are extreme values.

[1]The GitHub can be found at `https://icarus-sim.github.io/`.

Table 3.1: Table of comparison popularity of data for different α

alpha	0,1		1		10	
content item	k(i)	p(i)	k(i)	p(i)	k(i)	p(i)
1	1	0,270303728	1	0,48	1	0,999006537
2	0,933032992	0,252202296	0,5	0,24	0,000976563	0,000975592
3	0,89595846	0,242180912	0,333333333	0,16	1,69351E-05	1,69183E-05
4	0,870550563	0,235313063	0,25	0,12	9,53674E-07	9,52727E-07

Let us elaborate on a realistic value of α. In 1999, Breslau et al. (1999) have shown that the number of requests is indeed distributed according to a Zipf distribution, and that realistic values of α lie between 0.6 and 0.8. It is therefore that $\alpha = 0.8$ will often be used in the remainder of this thesis.

3.6. Contribution

The scientific contribution of the mathematical part of this thesis lies in the resilience part. In this thesis, we use both simulation and analytical (approximation) approaches to compute the resilience of different networks. To be able to compute this resilience, we need to know what content is retrieved from what cache.

This can be done in a simulation in which the method is rather straight forward. As has been explained in more detail in Chapter 3.4, existing simulation software called Icarus can simulate information requests from the end users. Next, Icarus can track which of these requests will be retrieved from which cache. The percentage of content requests that is retrieved from the cache is called the cache hit ratio.

However, there is another way to find this cache hit ratio. By solving a (large) set of equations, the cache hit ratio can be approximated (very well as we will see later). This method is based on the so called Che's approximation.

Once the cache hit ratio is being computed, the resilience can be computed. Again, we can use a simulation or an analytical approach for this. The simulation is a new contribution. The approach is rather brute force, which makes it accurate but slow once the system grows. The analytical approach is also new. By comparing it to the simulation approach, we find that in many cases the analytical approach is accurate too. One of the main advantages is that the analytical approach is much quicker. A schematic overview of this can be found in Figure 3.5.

The research has resulted in a scientific paper submitted to the 4th ACM Conference on Information-Centric Networking. At the time of writing, the paper is still under review.

3.7. Influence of Different Parameters

In this section, let us elaborate on the different parameters that play a role in the ICNs. Furthermore, let us address which of the blocks in Figure 3.5 they influence. These parameters should not be new to the reader (for more elaboration the reader is referred back to Chapter 2). This Section aims to gain an overview of the sphere of influence of these parameters. Note here that the cache hit ratio influences the resilience. Parameters might influence the resilience directly, or through the cache hit ratio.

Cache Size

If the cache size increases, more items can be stored in the cache. As a result, the probability that a certain content item can be found in the cache increases. Hence, the cache hit ratio will increase. In an extreme setting, the cache size is as large as the content population. If this is the case, all content can be stored in the cache, and the cache hit ratio will be equal to one. Hence, the cache size influences the cache hit ratio directly.

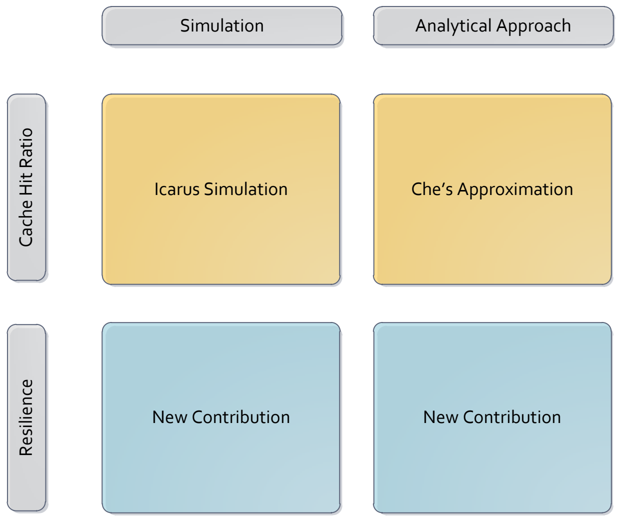


Figure 3.5: Schematic overview of the research

Content Population

If the population size increases, a smaller fraction of the content can be stored in the caches. Hence, the cache hit ratio decreases once the content population increases. However, often the cache size is chosen as a fraction of the content population. If that is the case, the content population hardly has any influence. If that is not the cache, the content population influences the cache hit ratio.

Content Popularity

The content popularity has influence on what content can be found in the caches. If the popularity is very skewed (hence some content is very popular), this skewed content has a large likelihood to be found in the caches as well as a high probability to be the requested content. Hence, the content popularity influences the cache hit ratio.

Caching Strategy

Not only the factors above influence the cache hit ratio. One of the most influential factors is the caching strategy. This set of rules on where the content is being cached has already been elaborated on in Section 2.4. This note is only for the purpose of being complete on the factors that influence the cache hit ratio and/or the resilience. The influence of the caching strategy on the cache hit ratio is large. The set of equations that need to be solved in Che's equation to approximate the cache hit ratio change significantly when the caching strategy changes. We will come back to that in Section 5.1.

Topology

All factors above only influence the resilience through the cache hit ratio. The topology however, influences both the resilience and the cache hit ratio directly.

First of all, the topology influences the cache hit ratio, because the configuration of the nodes determines the path along which content is sent from the source to the end user. This is the same path along which is searched for content in the caches. It is therefore that the equations to be solved when approximating the cache hit ratio (Che's approximation) change when the topology changes. See also Section 5.1.

The resilience is also directly influenced by the topology. We will pay a lot of attention to that in the remainder of this thesis, but let us explain the matter with an intuitive example here. In this example, we limit ourselves to obtaining content from the source. Hence, the imaginary caching capacity is equal to zero.

Imagine a path graph. If one link along the path breaks down, no content can be delivered to the end user. Secondly, imagine that all links are being doubled along this path graph. If one link breaks down now, we are certain that the content can still be delivered. Hence, by changing the topology (duplicate each link) we have influenced the impact of a breaking link.

4

Analytical Results: Simple Network Structures

In the next two chapters, the analytical results of the resilience of the ICN are central. Firstly, in Section 4.1 some special cases will be considered. For these special cases, the behaviour can be predicted accurately. An example of a situation for which it is possible to predict its behaviour when links break down, is a tree topology with very high skewness. This is because it is then possible to predict which content is in the caches, allowing one to analyse what its behaviour will be under disturbances.

As already stated above, before being able to analyse the resilience of a topology, it is needed to know the behaviour of the caches. This so-called cache performance can be analysed with full accuracy by using Markov chains (Section 4.2). This is done in the second section of this chapter. However, as we will see, there is one big drawback of this method, namely its exponentially increasing complexity as the system complexity increases (for example more content, more nodes in the network, more caches etcetera). It is for that reason that the method can only be used for extremely small illustrative examples.

It is therefore that we will have to reach out to approximations to analyse the cache performance. This is done in the Chapter 5. The so called Che's approximation (Che, Tung, & Wang, 2002) appears to be very suitable for this. However, we need to extend the approximation in order to be suitable for systems with more than one cache.

4.1. Special Cases

Before moving to a more general approach, let us consider some special cases, in which we can fully understand the behaviour.

4.1.1. High Skewness - Tree Topology

The popularity of the content population has great influence on the way information spreads through the caches. If some content is much more popular (hence more often requested) than the rest of the content, the probability that this popular content is in the caches is higher than the probability that the other content items are in the cache. If the skewness increases (and the belonging parameter α also increases), the difference between content items on how often they are requested by the end users also increases.

Assuming the warm up time[1] to be large enough, and the skewness to be high (large α), the behaviour of the network with edges breaking down can be predicted. When α is high, some content pieces are much more popular than others. The least recently used (LRU) strategy is being applied as replacement strategy, meaning that the least recently used content piece is being replaced in the

[1]the number of simulations performed before the measuring begins. This is done to get content in all the caches.

cache. As some content pieces are very popular, these probably never leave the cache. Hence, after some time, all caches are filled with the few (possibly even one) very popular pieces of information. Furthermore, almost all requests have identifier 1, as the requests are so skewed.

Now, let us assume the following example, as shown in Figure 4.1. The tree consists of 15 nodes, of which node 0 is the source, and node 7 till 14 are end users. The remaining nodes are routers. As explained above, the caches all contain the content with identifier 1. Furthermore, all end users are evenly active, meaning that they all request the same amount of content. Now what happens with a breakdown of an edge?

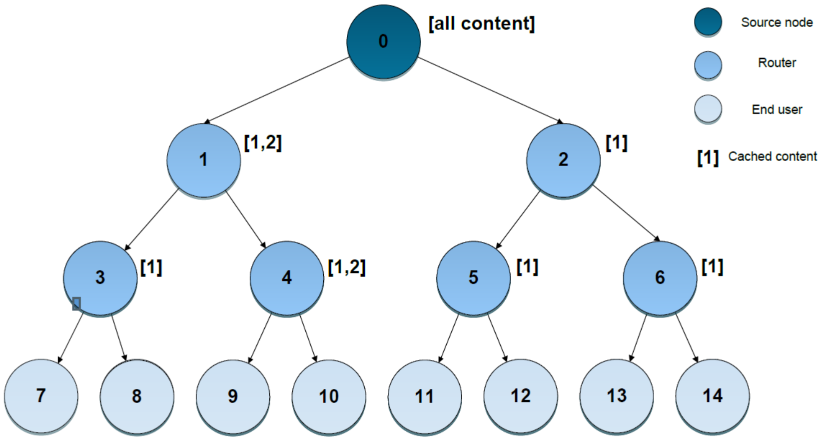


Figure 4.1: tree topology with caches filled with high skewness

When breaking the edges between the routers or between the source and a router, this has no effect on the information resilience, as the requests are still available in the caches of node $3, 4, 5$ or 6. When breaking an edge between a router and an end user, $1/8$ of the requests can no longer be dealt with. Hence, there is a probability of $6/14$ that the broken edge has no effect on the information resilience, and a probability of $8/14$ that one of the end users has no access to the information anymore. This results in the following:

$$\mathbb{E}[\text{percentual decay in information handling }] = \frac{6}{14} \cdot 0 + \frac{8}{14} \cdot \frac{1}{8} = \frac{1}{14} \tag{4.1}$$

Hence, when simulating over a number of runs, we expect a linear decay from 100% towards 0% when breaking all 14 links.

This can indeed be concluded from Figure 4.2, where the results of 100 runs are averaged. Here a value of $\alpha = 100$ is taken. In Section 3.5, we have compared the α values 0.1, 1 and 10. So one should note that this value is extremely large. As a result, it is practically certain that the first content item will be the only item requested. This is exactly what we want to happen.

We generalise the results of Equation 4.1 to an arbitrarily large binary tree with $n + 1$ nodes and n edges. If we have n edges, then there are $\frac{n}{2} + 1$ that are directly connected to the end users, and which therefore have a direct impact once they break. For the other $\frac{n}{2} - 1$ edges, there is no impact on the information resilience once they break. As in Equation 4.1, the impact of a breaking link connected to an end user, is equal to disconnecting one end user from the rest of the network. Hence, the generalisation of Equation 4.1 is:

$$\mathbb{E}[\text{percentual decay in information handling }] = \frac{\frac{n}{2} - 1}{n} \cdot 0 + \frac{\frac{n}{2} + 1}{n} \cdot \frac{1}{\frac{n}{2} + 1} = \frac{1}{n} \tag{4.2}$$

Hence, the result obtained for a graph with 15 nodes generalizes to graphs of arbitrary size.

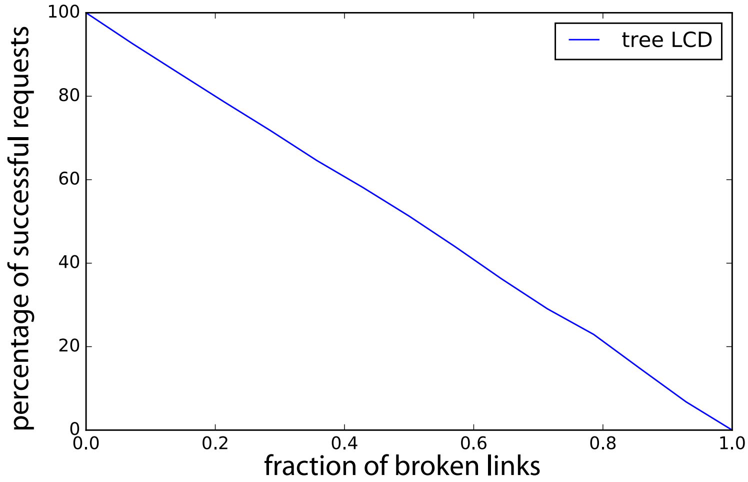


Figure 4.2: Information resilience of a tree topology with high skewness

4.1.2. No Skewness - Tree Topology

Now that we have analysed what happens when the content is very skewed, the question arises what happens when there is no preference in the content at all. First of all, when all content is requested with the same probability, but the caches are infinitely large and enough time has passed, we end up in a similar situation as in Section 4.1.1. The difference lies in the fact that the caches are filled with all content.

The interesting part is what happens when the cache size is not infinite, but rather limited. Let us consider again the tree from Figure 4.1. If a fraction f of all links is broken, then the probability that a link is still working is equal to $1-f$. Furthermore, let C be the absolute size of the caches, and let $cs = \frac{C}{P}$, where P is the size of the content population. Hence, the probability that node 7 can find the content requested in node 3 when a fraction f of the links is broken is equal to $(1-f)\cdot cs$, being the probability that the link works, times the probability that the content requested is situated in node 3.

If the requested content is not found in node 3, it might be that the information can be found in node 1. The probability that the information is retrieved from node 1 is equal to $(1-f)\cdot(1-cs)\cdot(1-f)\cdot cs = (1-f)^2\cdot cs(1-cs)$, being the probability that the link between 7 and 3 is still intact, times the probability that the content is not in node 3, times the probability that the link between node 3 and 1 still works, times the probability that the content can be found there. Note, that here the caching strategy has not been taken into account. Due to the caching strategy, the probability of a piece of content being found in node 3 is probably not independent from the probability that the same piece of content is found in node 1. However, for the time being, we assume it to be independent.

Generalizing the result above, the probability that a request can still be retrieved, given that a fraction f of the links is broken, can be given by:

$$p(\text{succesfull request}) = \sum_{i=1}^{n-1}\left((1-f)^i\cdot cs\cdot(1-cs)^{i-1}\right) + (1-f)^n(1-cs)^{n-1}, \tag{4.3}$$

where n is the depth (number of layers below the source) of the tree (hence $n = 3$ in Figure 4.1). When we apply this to the tree of Figure 4.1, we obtain the results in Figure 4.3. As we can see, the calculated results are very similar to the simulated results.

However, two rather important assumptions have been made. First of all, as stated above, the caches are not filled independently from each other, as the caching strategy LCD that has been used here fills the caches rather specifically, only filling a cache when its parent node has the information piece cached already. Secondly, returning to the example of Figure 4.1, when the information piece requested by node 7 is not found in node 3 or 1, there is a possibility for it to be found in node 0, but also in node 4. The latter has not been taken into account here either. However, when returning to the obtained results in Figure 4.3, the two assumptions either compensate for each other, or they can both be neglected. As we are not able to fully understand what is happening here, we will return to this matter in the discussion (Chapter 9).

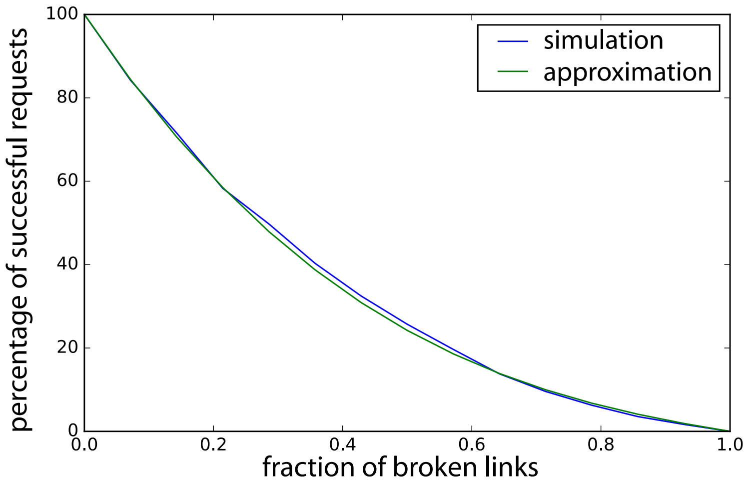


Figure 4.3: Information resilience of a tree topology with high skewness

4.1.3. Some Skewness - Tree Topology

So, we can predict the behaviour of a tree topology when there is no skewness or when there is a very high skewness. But what happens when there is some skewness? There are some different cases here for which we can give an intuitive analytical approach. However, the general case is too complex to catch in an analytical model here.

Table 4.1: Content and probability of requesting

Content	Probability
a	0.1
b	0.2
c	0.3
d	0.4

We will start with the spread of the content over the caches. We already know this behaviour for no or very large skewness, but what can be said about the spread of content over the caches when there is some skewness in the data? Again, we take the tree from Figure 4.1. Also, we take only 4 pieces of content into consideration, namely a, b, c and d. The probability that an end user requests this content can be found in Table 4.1. Furthermore, let us take LCD as the caching strategy. Now, if the cache size $C = 1$, the only way an item can be cached in for example cache 3 is if it is requested by node 7 and/or 8 twice, before nodes $7, 8, 9$ or 10 request any other content. This probability can be calculated. In the case of the taken example, the probability that content b is requested twice in a row is equal to

$$P(\text{content b is requested twice in a row}) = \left(\frac{2}{10}\right)^2 = \frac{4}{100} \tag{4.4}$$

However, the probability that two following requests are the same is equal to

$$P(\text{same content request two times in a row}) = \left(\frac{1}{10}\right)^2 + \left(\frac{2}{10}\right)^2 + \left(\frac{3}{10}\right)^2 + \left(\frac{4}{10}\right)^2 = \frac{30}{100} \tag{4.5}$$

Now, the probability that an item is already in cache 3 can be computed, although more complex than the computations above. This already indicates that the system quickly becomes very difficult to solve. Let alone a situation where the cache size C increases to $C = 2$. It should be clear now, that we run into a large set of problems when continuing this way of reasoning. It is therefore that we will move our attention to Markov chains and Che's approximation.

4.2. Markov Chains

In this paragraph, we will consider a path graph with 3 nodes: one source, one router and one end user (see also Figure 4.4). The cache will have size $C = 2$. Furthermore, we consider the same content population as in the previous case, with the same popularity distribution (Table 4.1).

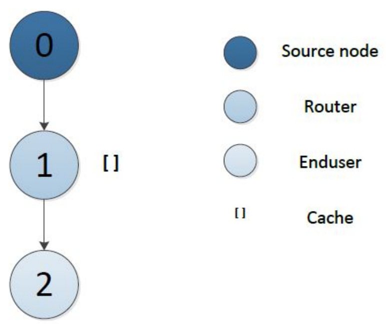


Figure 4.4: A path graph with one router

It is possible to recognize a Markov chain in the different states of the system. The LRU strategy, however, is not entirely memoryless, which is a key condition for letting the Markov chains be of use. To use the LRU, we need to track which content was the least recently used one. It is possible to work around that by letting the position in the cache denote the least recently used content item. That means that $[a, b]$ is a different state of the cache then $[b, a]$, where the latter means that content item a is the least recently used content item.

We now obtain the following set of states of the system.

$$S = \{[a, b], [a, c], [a, d], [b, a], [b, c], [b, d], [c, a], [c, b], [c, d], [d, a], [d, b], [d, c]\} \tag{4.6}$$

It is now possible to construct the transition matrix P. We will explain the row $p_{[a,b]}$ of this matrix, as the rest is filled with the same logic. Here, we assume a time homogeneous system. Content d is being requested with probability 0.4. If the system is in the state where the cache is filled with $[a, b]$, and content d is being requested, the cache will contain $[d, a]$ afterwards. Hence, $P_{[a,b][d,a]} = 0.4$. Likewise, $P_{[a,b][c,a]} = 0.3$, and $P_{[a,b][b,a]} = 0.2$. Now, if content a is being requested, noting changes, as a is already in the first place of the cache. Hence $P_{[a,b][a,b]} = 0.1$.

The transition matrix can be filled continuing this reasoning. The matrix can be found next.

$$P = \begin{array}{c} \\ [a,b] \\ [a,c] \\ [a,d] \\ [b,a] \\ [b,c] \\ [b,d] \\ [c,a] \\ [c,b] \\ [c,d] \\ [d,a] \\ [d,b] \\ [d,c] \end{array} \begin{array}{c} \begin{array}{cccccccccccc} [a,b] & [a,c] & [a,d] & [b,a] & [b,c] & [b,d] & [c,a] & [c,b] & [c,d] & [d,a] & [d,b] & [d,c] \end{array} \\ \begin{pmatrix} 0,1 & 0 & 0 & 0,2 & 0 & 0 & 0,3 & 0 & 0 & 0,4 & 0 & 0 \\ 0 & 0,1 & 0 & 0,2 & 0 & 0 & 0,3 & 0 & 0 & 0,4 & 0 & 0 \\ 0 & 0 & 0,1 & 0,2 & 0 & 0 & 0,3 & 0 & 0 & 0,4 & 0 & 0 \\ 0,1 & 0 & 0 & 0,2 & 0 & 0 & 0 & 0,3 & 0 & 0 & 0,4 & 0 \\ 0,1 & 0 & 0 & 0 & 0,2 & 0 & 0 & 0,3 & 0 & 0 & 0,4 & 0 \\ 0,1 & 0 & 0 & 0 & 0 & 0,2 & 0 & 0,3 & 0 & 0 & 0,4 & 0 \\ 0 & 0,1 & 0 & 0 & 0,2 & 0 & 0,3 & 0 & 0 & 0 & 0 & 0,4 \\ 0 & 0,1 & 0 & 0 & 0,2 & 0 & 0 & 0,3 & 0 & 0 & 0 & 0,4 \\ 0 & 0,1 & 0 & 0 & 0,2 & 0 & 0 & 0 & 0,3 & 0 & 0 & 0,4 \\ 0 & 0 & 0,1 & 0 & 0 & 0,2 & 0 & 0 & 0,3 & 0,4 & 0 & 0 \\ 0 & 0 & 0,1 & 0 & 0 & 0,2 & 0 & 0 & 0,3 & 0 & 0,4 & 0 \\ 0 & 0 & 0,1 & 0 & 0 & 0,2 & 0 & 0 & 0,3 & 0 & 0 & 0,4 \end{pmatrix} \end{array}$$

As we have assumed a steady state matrix, it is now possible to find the stationary distribution.

π is the stationary distribution belonging to S if the following conditions are satisfied:

1. $0 \leq \pi_j \leq 1$
2. $\sum_{j \in S} \pi_j = 1$
3. $\pi_j = \sum_{i \in S} \pi_i P_{ij}$

It is not hard to verify that the following vector (Equation 4.7) satisfies all three conditions, and is therefore our steady state vector.

$$\pi = \begin{pmatrix} 0.022 \\ 0.033 \\ 0.044 \\ 0.025 \\ 0.075 \\ 0.100 \\ 0.043 \\ 0.086 \\ 0.171 \\ 0.067 \\ 0.133 \\ 0.200 \end{pmatrix} \tag{4.7}$$

This means that if we look at a random moment in time, there is a probability of 0.2 that the state of the cache is $[d, c]$. This leads to the following probabilities of content being in the cache.

$$\begin{aligned} P(a \text{ in cache}) &= 0.235 \\ P(b \text{ in cache}) &= 0.441 \\ P(c \text{ in cache}) &= 0.608 \\ P(d \text{ in cache}) &= 0.716 \end{aligned} \tag{4.8}$$

Now, let us check this with a simulation. We copy the parameters as taken before, and run the Icarus software 500 times. In each run of Icarus, we check the cache after 20 iterations. This leads to the results in Equation 4.9. As we see, the results are satisfying, indicating that the analytical approach of Markov Chains align with the simulation results.

$$\begin{aligned}
\#(a \text{ in cache}) &= 110 = 22.0\% \\
\#(b \text{ in cache}) &= 224 = 44.8\% \\
\#(c \text{ in cache}) &= 309 = 61.8\% \\
\#(d \text{ in cache}) &= 357 = 71.4\%
\end{aligned} \tag{4.9}$$

Once the distribution in the cache is known, it is also possible to calculate the resilience. Notice that two links can break down. If the link between the end user and the rest of the network breaks, the information resilience reduces to 0. We name this link B for now. If the other link breaks down (A), the request can still be handled iff the requested information is in the cache.

This leads to the following calculation.

$$\begin{aligned}
P(\text{succesfull request}) = \\
P(\text{B breaks})P(\text{succes} \mid \text{B breaks}) + P(\text{A breaks})P(\text{succes} \mid \text{A breaks}) = \\
0 + \frac{1}{2} \sum_{i \in \{a,b,c,d\}} P(\text{i requested})P(\text{i in cache}) = \\
\frac{1}{2}(0.1 \cdot 0.24 + 0.2 \cdot 0.44 + 0.3 \cdot 0.61 + 0.4 \cdot 0.70) = 0.29
\end{aligned} \tag{4.10}$$

Let us verify that with a simulation too. We run 300 runs of the ICN simulation. After the caches were filled, the edges were randomly broken one by one. Taking the average success rate of the information requests for each fraction of broken nodes, leads to the result of Figure 4.5. If one node is broken ($f = 0.5$), on average 28.967% of the content requests could be dealt with. This aligns perfectly with the previously calculated result.

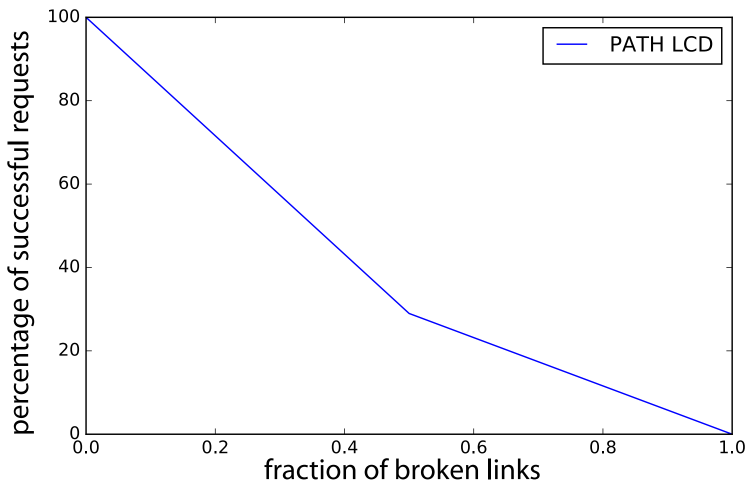


Figure 4.5: Resilience simulation of path graph with 3 nodes

Now that we have successfully computed this small example, it would be interesting to compute more difficult examples too. However, the possible different states are increasing rapidly as the size of the network, content population or caches increases. If we would add one more node to our network, the transition matrix would already consist of 144 rows instead of 12. One can imagine that if we consider tree networks with 20 nodes, a content population of 100 and a cache size of $C = 5$, it becomes impossible to compute the steady state, as the transition matrix is too large. It is therefore hard to use this method for all networks in general.

5

Analytical Results: Hierarchical Network Structures

In this Chapter, we consider more general hierarchical topologies and network configurations than we did in Chapter 4. As was described in Section 3, the analytical approximation of the information resilience goes through two steps. First of all, the cache hit ratio needs to be approximated. Secondly, the information resilience can be computed with the use of the cache hit ratio. In Section 5.1, a mathematical introduction to Che's approximation will be given. In Section 5.2, we use Che's approximation to approximate the cache hit ratio's. Section 5.3 uses this to compute the information resilience. Finally, in Section 5.4, the obtained results are applied to slightly more complex network topologies.

5.1. Che's Approximation

In 2002, Che et al. proposed an approximation to evaluate both the P_{in} (probability a piece of content is in the cache) and P_{hit} (the probability of having a cache hit). Back then, it was shown the approximation works very well, but it was not fully understood why it worked so well. This is further elaborated on in (Rosensweig, Kurose, & Towsley, 2010) and (Fricker, Robert, & Roberts, 2012).

Originally, Che's approximation is only suitable for investigating the behaviour of networks with only one cache. If an item is requested by the end user, it is first checked whether this item can be retrieved from the cache. If this is not the case, it is retrieved from the source, and stored in the cache. Let $1 \leq i \leq N$ be the item of interest from the content population of size N. The corresponding popularity proportion (in this thesis taken according to the Zipf distribution, but this is no requirement for Che's approximation) is called $q(i)$ (Fricker et al., 2012). Then

$$P_{in}(i) \approx 1 - e^{-q(i)t_C}, \tag{5.1}$$

where t_C is the root of

$$\sum_{i=1}^{N} P_{in}(i) = C \tag{5.2}$$

Che's approximation makes use of the pact that the cache replacement strategy is LRU. For other replacement strategies, it does not work. It is important here to notice that the P_{in} and P_{hit} are equal in the case of one cache. This follows from the PASTA property, as discussed in the next section.

It is possible to understand t_C as a simplification of $t_C(i)$ which is the time needed to let exactly C different items that do not include i be requested by the end users from that cache. In other words, $t_C(i)$ is the cache eviction time, the time needed to let item i disappear from the cache again. As one might argue, this eviction time is not a constant, as it depends on the specific sequence of requests done by the end user. However, Fricker et al. (2012) prove that this is an acceptable simplification of the model. The strength of Che's approximations lies in the fact that all the dynamics of different content is clustered in this one number t_C, which represents the respons of the cache to a content request (Garetto, Leonardi, & Martina, 2016).

5.1.1. Poisson Process

As said before, the article of (Garetto et al., 2016) distinguishes between the P_{in} and P_{hit}. The arrival times of requests have a Poisson distribution. Now, the P_{in} is the time-average probability that a certain piece of content is in the cache, while the P_{hit} denotes the probability of a hit. One could observe that the two are the same, if the Poisson Arrival See Time Average property (often referred to as PASTA property) holds (Wolff, 1982). In (Wolff, 1982), the definition of the PASTA property is given as: "the expected fraction of arrivals who find (observe or see) the process in a state is equal to the corresponding fraction of time". The property holds, however, only when the arrivals are independent of the state of the system. This is the case in the situation with one cache, but it is not generally true in the case with more caches. The arrival of requests in the caches closer to the source become dependent of the state of the caches closer to the end users (if often requested information is in the caches close to the end user, there will be fewer requests that go upstream to the caches closer to the sources). This implies that the P_{in} and P_{hit} are different in these situations. Furthermore, we need to estimate the arrival times in the caches closer to the source to be able to use Che's approximation again.

5.1.2. Nature of Che's Approximation

Fricker et al. (2012) explain some more on the background of Che's approximation.Here, we will repeat some of their explanation of the nature of Che's approximation. In particular, we will see that some assumptions are made in the approximation, which are justified.

First of all, Fricker et al. introduce two random variables. First of all, they introduce

$$X_n(t) = \sum_{i=1, i\neq n}^{N} \mathbb{1}_{\tau_i < t}, \tag{5.3}$$

where τ_n denotes the exponentially distributed inter-request interval of object n. The equation represents the number of different objects that have been requested up to time t, excluding item n. Secondly,

$$T_C(n) = \inf\{t > 0 : X_n(t) = C\} \tag{5.4}$$

means the time it takes until exactly C different items other than n have been requested. This plays a role in the cache hit ratio, as item n will not be in the cache anymore at time $T_C(n)$. Now, suppose that object n is being requested at $t = 0$ (which can be done without loss of generality). If the next request for object n is done before time $T_C(n)$, it will result in a cache hit. Hence, we want τ_n to be smaller than the $T_C(n)$ in order to have a cache hit. But if the next request is indeed in the interval $(0, \tau_n)$, then clearly $X_n(\tau_n) < C$. Combining all this, we obtain a cache hit if

$$X_n(\tau_n) < C \tag{5.5}$$

Which is true iff

$$T_C(n) > \tau_n. \tag{5.6}$$

Hence,

$$P_{in}(n) = P(T_C(n) > \tau_n) = \mathbb{E}\left(1 - e^{-q(n)T_C}\right). \tag{5.7}$$

Which is already very similar to equation 5.1. However, we still need to find an approximation for C. Since time $T_C(n)$, there are exactly C objects in the cache. Filling in Equation 5.3 in Equation 5.4 for time $T_C(n)$, we find that

$$C = \sum_{i=1, i\neq n}^{N} \mathbb{1}_{\tau_i < T_C(n)}. \tag{5.8}$$

Furthermore, if we take expectation also here, we obtain

$$C = \sum_{i=1, i \neq n}^{N} \mathbb{E}\left(1 - e^{-q(i)t}\right). \tag{5.9}$$

So far, we have not been approximating anything, but have been precise. However, to be able to compute these equations, we wish to make some simplifications. First of all, we assume the $T_C(n)$ to be deterministic. Therefore, it is possible to replace the random variable $T_C(n)$ by a constant, that we call $t_C(n)$. This constant then satisfies

$$C = \sum_{i=1, i \neq n}^{N} \left(1 - e^{-q(i)t}\right). \tag{5.10}$$

Furthermore, the constant $t_C(n)$ approximates

$$P_{in}(n) = 1 - e^{-q(n)t_C(n)}. \tag{5.11}$$

Secondly, we assume that our constant above ($t_C(n)$) is not dependent on the content item we consider. But if that is the case, we also have no need to exclude the particular content item from the summation. We get $t_C(n) = t_C$, which gives

$$C = \sum_{i=1}^{N} \left(1 - e^{-q(i)t}\right). \tag{5.12}$$

Note that this is only reasonable if the contribution of the popularity of one content item is relatively small to the entire popularity. Now that we have build up the Equations 5.1 and 5.2, we move on to an extension for more than one cache.

5.1.3. Che's Approximation for Two Caches

Here, we discuss the extension of Che's approximation in number of caches and in caching strategies as proposed by Garetto et al. (2016). We do this for both LCE and LCD.

Leave Copy Everywhere

The Leave Copy Everywhere caching strategy can be applied using Che's approximation. In the article of Garetto et al. (2016), a standard approach is mentioned which we will shortly repeat here. In the remainder of this mathematical derivation, cache number 2 is the cache close to the source, while cache number 1 is the one close to the end user (following Figure 4.4). As the Leave Copy Everywhere policy has the effect that the first cache is not influenced by the second one, the behaviour of the first cache is equal to the situation of only one cache, and so P_{in} can be found in Equation 5.1. Let the Poisson arrival times in the second cache be approximated by

$$\overline{q}_2(i) = q_1(i)(1 - p_{hit}(1, i)) \tag{5.13}$$

Like $P_{in}(1, i)$, $P_{in}(2, i)$ is also given by the Poisson arrival times and $t_C^{(2)}$:

$$P_{in}(2, i) \approx 1 - e^{-\overline{q}_2(i)t_C^{(2)}} \tag{5.14}$$

where $t_C^{(2)}$ is the root of

$$\sum_{i \in \{a,b,c,d\}} \left(1 - e^{-\overline{q}_2(i)t}\right) = C \tag{5.15}$$

Hence, the $P_{in}(2, i)$ is calculated very similar to the $P_{in}(1, i)$. As said before, $P_{hit}(1, i) = P_{in}(1, i)$, but this is not the case for the second cache, as the PASTA property does not hold there. We obtain a cache hit in cache two if an item is requested, available in cache 2, but not available in cache 1 (as that would cause a cache hit in cache 1). Note that, especially when the LCE strategy is used, in practice the probability of a cache hit in cache two is much smaller than the probability of an item being present in this cache. Theoretically, when the $t_C^{(2)}$ is much larger than $t_C^{(1)}$, this is not necessarily the case.

Garetto et al. (2016) explain that, if a request for item i arrives at cache two at time t, this can only be the case if the item i is not present in cache one. Hence, no request for item i was made in the interval $[t-t_C^{(1)}, t]$, as that would cause the item to be available in cache one. Furthermore, to have the item available in cache two (which is a requirement for a cache hit), a request for item i should have arrived at cache two in the interval $[t-t_C^{(2)}, t-t_C^{(1)}]$, which only makes sense if $t_C^{(2)} > t_C^{(1)}$. The arrival times at cache two in that interval are not Poisson, as they depend on the state of cache 1. However, we approximate the arrival times to be still Poisson, by using the $\overline{q}_2(i)$ as stated in Equation 5.13.

If we incorporate the condition that $t_C^{(2)}$ should be greater than $t_C^{(1)}$, it follows that $P_{hit}(2,i)$ can be approximated by

$$P_{hit}(2,i) \approx 1 - e^{-\overline{q}_2(i)\max\left((t_C^{(2)}-t_C^{(1)},0\right)}. \tag{5.16}$$

Leave Copy Down

The difficulty with the Leave Copy Down strategy is that the behaviour of the first cache depends on the content of the second cache, but the content in the second cache is influenced by the first cache (hence, they are interdependent). As a result of that, we will end up with interdependent equations as well. This can be solved by using a fixed point iterative method to solve the equations.

Before moving to the equations specific for LCD, let us write a more general formula for $t_C^{(j)}$. $t_C^{(j)}$ is the root of

$$\sum_{i=1}^{N} P_{in}(j,i) = C. \tag{5.17}$$

Equation 5.17 is just a generalisation of Equation 5.2, where a system with only one cache was stated. There are two ways to have an item present in cache 1 at time t. One way is that the item was last requested in the interval $[t-t_C^{(1)}, t]$, and was hit in cache 2 (and then copied in cache 1 too). The probability for this is equal to $(1-P_{in}(1,i))P_{hit}(2,m)$, being the probability that the item is not in cache 1, but is hit from cache 2. The other option is that the item was last requested in the interval $[t-t_C^{(1)}, t]$, and was already in cache 1. In that case, due to a cache hit, it also remains in cache 1. The probability for this situation to occur is simply $P_{in}(1,i)$ For the first cache, the PASTA property holds again. We compute the $p_{in}(2,i)$ again as in equation 5.14. Hence,

$$\begin{aligned} P_{in}(1,i) \approx P_{hit}(1,i) &= [(1-P_{in}(1,i))P_{hit}(2,m) + P_{in}(1,i)]\left(1-e^{-\overline{q}_1(i)t_C^{(1)}}\right) \\ P_{in}(2,i) &= 1-e^{-\overline{q}_2(i)t_C^{(2)}}. \end{aligned} \tag{5.18}$$

Now, how do we obtain a cache hit in cache 2? We elaborate on the line of reasoning for $t_C^{(2)} > t_C^{(1)}$, but the same way of reasoning leads to the case where $t_C^{(2)} \leq t_C^{(1)}$. There are again two ways in which we can obtain a cache hit at time t in the second cache. First of all, it could be that the item was stored in both caches, but was already removed again from the first cache. In that case, the previous request arrived in the interval $[t-t_C^{(2)}, t-t_C^{(1)}]$. Secondly, it could be that the second cache was filled, but there was no other request that caused the item to be stored in cache one. In that case the request arrived at cache two in interval $[t-t_C^{(1)}, t]$. Combining this with the approximation of the Poisson process as done already at Leave Copy Everywhere, this leads to the following equations:

$$P_{hit}(2,i) \approx \begin{cases} \left(1-e^{-\overline{q}_2(i)\left(t_C^{(2)}-t_C^{(1)}\right)}\right)e^{-\overline{q}_2(i)t_C^{(1)}} + \left(1-e^{-\overline{q}_2(i)(1-P_{hit}(2,i))t_C^{(1)}}\right), & \text{if } t_C^{(2)} > t_C^{(1)} \\ 1-e^{-\overline{q}_2(i)(1-P_{hit}(2,i))t_C^{(2)}}, & \text{otherwise} \end{cases} \tag{5.19}$$

In the next sections, we will elaborate on these equations with some examples. Furthermore, we will use this knowledge to compute the information resilience in some more general topologies than a path graph of length 4.

5.2. Application of Cache Hit Ratio's

In this section, we elaborate on computing the cache hit ratio. We do this for situations with only one cache as well as situations with more caches. In the latter case, we will also distinguish between the LCE and LCD caching strategy.

5.2.1. Single Cache

Let us first consider an example of the Che's approximation by approximating the cache behaviour of the example in Section 4.2 (Figure 4.4). Here, we have indeed a system with only one cache (node 1). The Che's approximation assumes a LRU in-cache policy, which matches the above example.

We will make use of the explanation of the approximation by Fricker et al. (2012). Like above, let the content population be $\{a, b, c, d\}$, the cache size $C = 2$, and the popularity distribution be as in Table 4.1. This distribution is denoted with $q(i)$, where $i \in \{a, b, c, d\}$.

According to Fricker et al., this hit rate is by construction equal to the P_{in}, which we have already computed with the help of Markov chains. The results of that are shown in Equation 5.1. Hence, in order to repeat this result, we first need to find the root of Equation 5.2. In other words, we need to find t such that

$$\begin{aligned} (1-e^{-0.4t}) + (1-e^{-0.3t}) + (1-e^{-0.2t}) + (1-e^{-0.1t}) = 2 \iff \\ e^{-0.4t} + e^{-0.3t} + e^{-0.2t} + e^{-0.1t} = 2 \end{aligned} \tag{5.20}$$

Since we know that the root is unique, the equation can be solved by for example Newtons method. We find that $t \approx 2.994$. Now, it follows that

$$\begin{aligned} P_{in}(a) \approx P_{hit}(a) = 1 - e^{-0.1*2.994} = 0.259 \\ P_{in}(b) \approx P_{hit}(b) = 1 - e^{-0.2*2.994} = 0.451 \\ P_{in}(c) \approx P_{hit}(c) = 1 - e^{-0.3*2.994} = 0.593 \\ P_{in}(d) \approx P_{hit}(d) = 1 - e^{-0.4*2.994} = 0.698 \end{aligned} \tag{5.21}$$

Comparing this to the results in 5.1, we see some differences, but $|P_{in}(i)_{Markov} - P_{in}(i)_{Che}| < 0.03$ for all content items.

5.2.2. Multiple Caches

The relatively easy extension that we will consider first, is adding one node (a router) to the network of Figure 4.4, such that the network remains a path graph. As there is a substantial difference in the LCD en LCE strategy and its cache behaviour, we will consider them both, starting with the LCE strategy. Note here that we are still only considering the probability of a cache being hit, and not the resilience yet. The resilience will be considered in Section 5.3.

Leave Copy Everywhere

Let us verify again if these results align with the simulation results. For this particular example, $P_{in}(1, i)$ is equal to the same expression in the situation with one cache. Hence, it can be found in Equation 5.21. Now, $\overline{q}_2(i)$ is given by:

$$\begin{aligned} \overline{q}_2(a) = 0.1(1 - 0.259) = 0.0741 \\ \overline{q}_2(b) = 0.2(1 - 0.451) = 0.1098 \\ \overline{q}_2(c) = 0.3(1 - 0.593) = 0.1221 \\ \overline{q}_2(d) = 0.4(1 - 0.698) = 0.1208 \end{aligned} \tag{5.22}$$

Next, $t_C^{(2)}$ can be approximated, again using Newtons method. We obtain $t_C^{(2)} \approx 6.5757$. And so:

$$\begin{aligned} P_{in}(2, a) \approx 1 - e^{-0.0741*6.5757} \approx 0.386 \\ P_{in}(2, b) \approx 1 - e^{-0.1098*6.5757} \approx 0.514 \\ P_{in}(2, c) \approx 1 - e^{-0.1221*6.5757} \approx 0.552 \\ P_{in}(2, d) \approx 1 - e^{-0.1208*6.5757} \approx 0.548 \end{aligned} \tag{5.23}$$

Table 5.1: Comparing simulation results with approximation path graph of length 4

	Approx-imation	Simulation	Absolute difference	Relative difference
$P_{in}(1,a)$	0,259	0,238	0,021	0,088235294
$P_{in}(1,b)$	0,451	0,418	0,033	0,078947368
$P_{in}(1,c)$	0,593	0,61	0,017	0,027868852
$P_{in}(1,d)$	0,698	0,734	0,036	0,049046322
$P_{in}(2,a)$	0,386	0,342	0,044	0,128654971
$P_{in}(2,b)$	0,514	0,57	0,056	0,098245614
$P_{in}(2,c)$	0,552	0,556	0,004	0,007194245
$P_{in}(2,d)$	0,548	0,532	0,016	0,030075188

Interesting to notice is that the distribution of the second cache is more evenly distributed than the distribution of the first cache. Let us again check this result with a simulation. Doing the same as before (again 500 simulations), we obtain the results in Table 5.1. Here, both the approximation and the simulation results are shown. These results are promising enough to continue working on the approximation.

However, we are interested in the probability of a cache hit, rather than the probability of an item being stored in the cache. For that, we need equation 5.16, repeated here: $P_{hit}(2,i) \approx 1 - e^{-\overline{q}_2(i)\max\left((t_C^{(2)} - t_C^{(1)}, 0\right)}$

As $t_C^{(2)} \approx 6.5757 > 2.994 \approx t_C^{(1)}$, and $t_C^{(2)} - t_C^{(1)} \approx 6.5757 - 2.994 = 3.5817$, we obtain:

$$\begin{aligned} P_{hit}(2,a) &\approx 1 - e^{-0.0741*3.5817} \approx 0.233 \\ P_{hit}(2,b) &\approx 1 - e^{-0.1098*3.5817} \approx 0.325 \\ P_{hit}(2,c) &\approx 1 - e^{-0.1221*3.5817} \approx 0.354 \\ P_{hit}(2,d) &\approx 1 - e^{-0.1208*3.5817} \approx 0.351 \end{aligned} \tag{5.24}$$

Note here, that in a situation where $t_C^{(2)}$ and $t_C^{(1)}$ are almost equal, the probability for a cache hit in cache two reduces to almost zero. This is logical, because a cache hit in cache two can only occur in a situation where the item is stored in cache 2 but not in cache 1, which becomes less probable if the time needed to request C distinct objects from the cache are almost equal.

So far, the equations are only applicable to path graphs, but Garetto et al. (2016) already commented on the differences when considering more complex topologies. First of all, the considered Poisson arrival rates ($\overline{q}_j(i)$) at a point in the system should be build up as before (see Equation 5.13) with the difference of summing up for all other caches that forward the particular content piece to the considered cache in case of a miss. Also, the probability of a hit is computed slightly more advanced. We will come back to it in Section 5.4.

Leave Copy Down

Here, we repeat the example of the previous section, but use another strategy, namely LCD. As the caches are interdependent of each other, we cannot compute this case by hand as we did before. We need to use a fixed point iteration for that. The pseudo code for the fixed point iteration can be found in Algorithm 1. Note that this algorithm is specific for the path example considered in this section. In the algorithm, we write down the set of equations that we want to solve, but deduct the two sides of the equation from each other, and put the result vector on the other side of the equation. For example, res= $(1 - e^{-q2*(1-p_{hit}(2))*t_C^{(2)}}) - p_{hit}(2)$. Note here that the result vector should be equal to zero, in order to solve the equation. It is therefore that the function F (line 3-17) only contains the equations that should be solved. By finding the root of this set of equations (line 18), the algorithm is complete.

The different components in the set of equations are interdependent. As such, the results printed below are interdependent as well. One can only check if the results found indeed lead to correct

Algorithm 1 Algorithm of finding the cache hit ratio's of a path graph of length 4, with the LCD strategy

$q1 = [q11, q12 \cdots q1n]$
n=len(q1)
Def(F):
$\triangleright$ Initiate all variables $(P_{in}(1), q2, P_{hit}(2), P_{in}(2))$, with the length of the size of the content population, as well as a result vector
$\triangleright$ Next, the different set of equations that need to be solved are stated
res[0:n] = $q1 * (1 - P_{in}(1)) - q2$
$t_C^{(1)}$ = optimize.root($\sum((1 - e^{(-q1*T)}) * ((1 - P_{in}(1)) * P_{hit}(2) + P_{in}(1)))$-cachesize)
$t_C^{(2)}$ = optimize.root($(1 - e^{-q2*T})$-cachesize)
res[2*n:3*n] = $((1 - P_{in}(1)) * p_{hit}(2) + P_{in}(1)) * (1 - e^{-q1*t_C^{(1)}}) - P_{in}(1)$
res[3*n:4*n] = $1 - e^{-q2*t_C^{(2)}} - P_{in}(2)$
if $t_C^{(2)} > t_C^{(1)}$: **then**
    res[5*n:6*n] = $\left(1 - e^{-q2*(t_C^{(2)} - t_C^{(1)})}\right) * e^{-q2*t_C^{(1)}} + \left(1 - e^{-q2*(1-p_{hit}(2))*t_C^{(1)}}\right) - p_{hit}(2)$
end if
if $t_C^{(2)} \leq t_C^{(1)}$: **then**
    res[5*n:6*n]= $(1 - e^{-q2*(1-p_{hit}(2))*t_C^{(2)}}) - p_{hit}(2)$
end if
Return res
optimize.root(F)

equations. We will do that below. Recall that $\overline{q}_2(i) = q_1(i)(1 - p_{hit}(1, i))$. We find, after a fixed point iteration, that:

$$\begin{aligned} \overline{q}_2(a) &= 0.1 * (1 - 0.1924) = 0.0808 \\ \overline{q}_2(b) &= 0.2 * (1 - 0.4383) = 0.1123 \\ \overline{q}_2(c) &= 0.3 * (1 - 0.6220) = 0.1134 \\ \overline{q}_2(d) &= 0.4 * (1 - 0.7473) = 0.1011 \end{aligned} \tag{5.25}$$

Next, let us verify the results for $P_{in}(1, i) \approx P_{hit}(1, i) = [(1 - P_{in}(1, i))P_{hit}(2, m) + P_{in}(1, i)] \left(1 - e^{-\overline{q}_1(i)t_C^{(1)}}\right)$.

$$\begin{aligned} P_{in}(1, a) &\approx [(1 - 0.1924) * 0.3250 + 0.1924] * \left(1 - e^{-0.1*t_C^{(1)}}\right) \approx 0.1924 \\ P_{in}(1, b) &\approx [(1 - 0.4383) * 0.3897 + 0.4383] * \left(1 - e^{-0.2*t_C^{(1)}}\right) \approx 0.4383 \\ P_{in}(1, c) &\approx [(1 - 0.6220) * 0.3915 + 0.6220] * \left(1 - e^{-0.3*t_C^{(1)}}\right) \approx 0.6220 \\ P_{in}(1, d) &\approx [(1 - 0.7473) * 0.3688 + 0.7473] * \left(1 - e^{-0.4*t_C^{(1)}}\right) \approx 0.7473 \end{aligned} \tag{5.26}$$

These equations are true if $t_C^{(1)} = 5.5022$. Next, $P_{in}(2, i) = 1 - e^{-\overline{q}_2(i)t_C^{(2)}}$ is being computed.

$$\begin{aligned} P_{in}(2, a) &\approx 1 - e^{-0.0808*t_C^{(2)}} \approx 0.4246 \\ P_{in}(2, b) &\approx 1 - e^{-0.1123*t_C^{(2)}} \approx 0.5364 \\ P_{in}(2, c) &\approx 1 - e^{-0.1134*t_C^{(2)}} \approx 0.5397 \\ P_{in}(2, d) &\approx 1 - e^{-0.1011*t_C^{(2)}} \approx 0.4993 \end{aligned} \tag{5.27}$$

These equations are true if $t_C^{(2)} = 6.8423$. Lastly, we verify $P_{hit}(2, i)$ from equation 5.19. As $t_C^{(2)} = 6.8423 > 5.5022 = t_C^{(1)}$, the formula to solve is

$$P_{hit}(2, i) \approx \left(1 - e^{-\overline{q}_2(i)\left(t_C^{(2)} - t_C^{(1)}\right)}\right) e^{-\overline{q}_2(i)t_C^{(1)}} + \left(1 - e^{-\overline{q}_2(i)(1 - P_{hit}(2,i))t_C^{(1)}}\right)$$

$$
\begin{aligned}
P_{hit}(2,a) &\approx \left(1 - e^{-0.0808*(6.8423-5.5022)}\right) e^{-0.0808*5.5022} + \left(1 - e^{-0.0808*(1-0.3250)*5.5022}\right) \approx 0.3250 \\
P_{hit}(2,b) &\approx \left(1 - e^{-0.1123*(6.8423-5.5022)}\right) e^{-0.1123*5.5022} + \left(1 - e^{-0.1123*(1-0.3897)*5.5022}\right) \approx 0.3897 \\
P_{hit}(2,c) &\approx \left(1 - e^{-0.1134*(6.8423-5.5022)}\right) e^{-0.1134*5.5022} + \left(1 - e^{-0.1134*(1-0.3915)*5.5022}\right) \approx 0.3915 \\
P_{hit}(2,d) &\approx \left(1 - e^{-0.1011*(6.8423-5.5022)}\right) e^{-0.1011*5.5022} + \left(1 - e^{-0.1011*(1-0.3688)*5.5022}\right) \approx 0.3688
\end{aligned}
\tag{5.28}
$$

We see here, that the two strategies indeed result in different caching behaviour. In the case of LCD, there is a cache hit more often in the second cache, resulting in a higher cache hit in general. In the next section, we will take the tree of Figure 4.1, and approximate the cache performance for this topology, where the cache size will be varied, and the content population will be increased. Furthermore, we will compute the information resilience of this topology.

5.3. Information Resilience Using Che's Approximation: Path and Tree

In this section, the information resilience of some networks is being approximated with making use of the cache hit ratio. These results are being compared to the results of a simulation. This simulation uses the Icarus simulation as input, and computes the information resilience from there. The simulations have been performed on a Oracle Virtual Machine, with Python version 2.7.12. Per simulation, we intend to specify the number of simulations and give an indication for the simulation time. Also, the results are being compared to the situation where there is no caching capacity in the network.

5.3.1. Approximation of Information Resilience

First of all, we investigate the caching behaviour of the tree of Figure 4.1. As said before, the cache size will be varied, and the content population will be increased towards 400. The content popularity will follow the Zipf distribution, as discussed in Section 3.4. For the caching strategy, we take LCE in this case.

By using the fixed point iteration to solve the system of equations, we are now able to run a simulation to compare the hit rates of the simulation with the approximated cache hits. Notice here, that we have computed the total hit rate as in Equation 5.29. All individual values have been calculated with the fixed point iteration.

$$
P_{HIT} = P_{hit}(total) = \sum_{m \in \text{content}} \left(q_1(m) * P_{hit}(1,m) + \overline{q}_2(m) * P_{hit}(2,m)\right) \tag{5.29}
$$

In Figure 5.1, the cache hit ratio of the binary tree with depth 3 is drawn. We define l_0 to be the layer of end users, and l_3 to be the source. l_1 and l_2 are the layers with routers. The total cache hit ratio of the approximation and simulation are drawn in blue and red (upper two lines). The combined cache hit ratio of the 4 caches at l_1 (node 3,4,5,6 in Figure 4.1) is represented by the simulation/approximation first layer. Likewise, the combined cache hit ratio of l_2 (node 1,2, in Figure 4.1) is also shown. As we can see, the approximation for both sets of caches as well as the total cache hit ratio are very close to the cache hit ratio in the simulation.

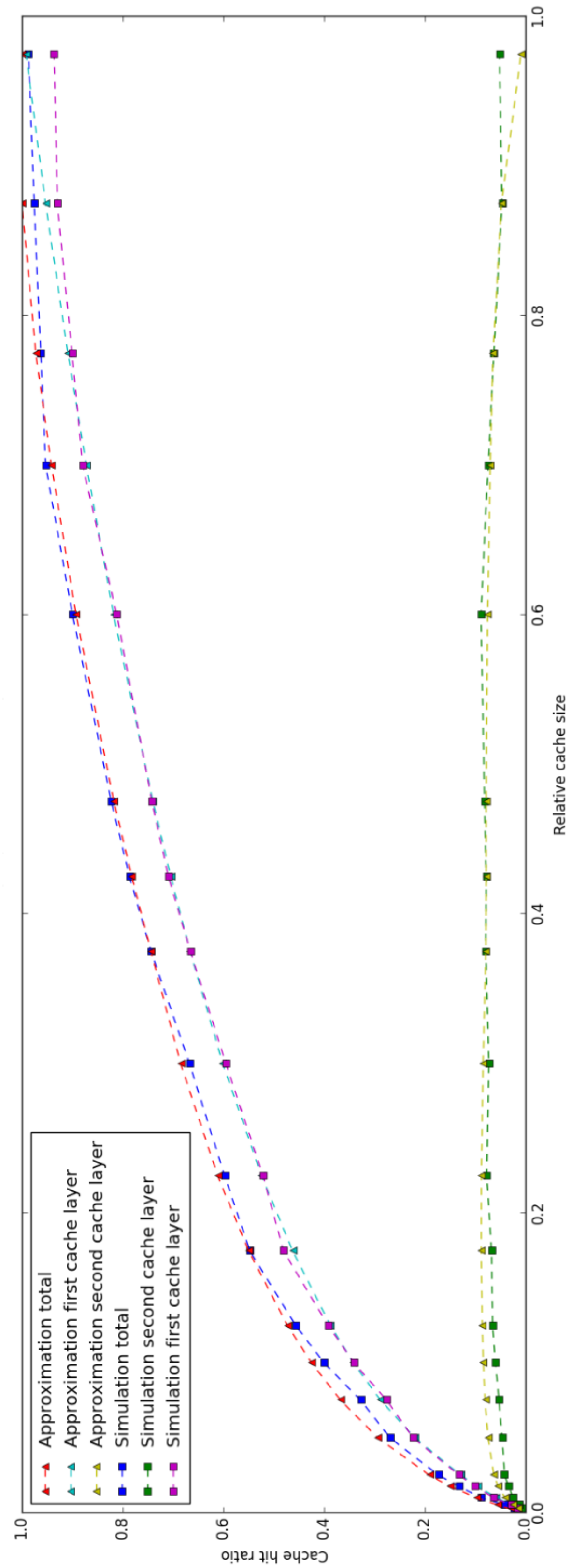


Figure 5.1: Cache hit ration comparison LCE strategy

Having knowledge about the behaviour of the caches is sufficient to approximate the resilience as well. Calculating the probability for a successful request can be done in a similar way as was done in Section 4.1.2. We state the probability of a successful request in Equation 5.30, where P_{HIT} is known per cache layer from the approximation, f is the fraction of the broken links, N is the total number of edges, and n the number corresponding to the layer number. hence, $n = 0$ corresponds to l_0, the end users. The probability of a successful request in the first layers of caches (l_1) is equal to the probability of a link being existent $(1 - f)$ times the probability of a hit in that cache (P_{HIT} in layer 1). Extending this for all layers of caches, we obtain Equation 5.30.

$$P(\text{succesful request}) = \sum_{i=1}^{n} \left(\prod_{j=0}^{i-1} \frac{N * (1-f) - j}{N - j} \right) * P_{HIT}(i) \tag{5.30}$$

Nicely enough, since the resilience calculation uses the hit rate, the dependency of the caches is already encountered for. We are interested in how well the approximation of the resilience is, compared to the simulation. To get a good and general view of that, in Section 5.3.4, an overview of the performance can be found. For a specific setting (topology, size, strategy etcetera), the cache hit ratio for different caches is being shown, and afterwards the resilience for 3 choices of cache size too. As the ultimate goal is to compare the enhancements with having no caching capacity at all, the red lines show the resilience for no cache.

5.3.2. Simulation of Information Resilience

In the previous subsection, attention was given to the approximation of the information resilience. In this subsection, some attention will be given to the simulation of the information resilience. Simulations are run.

In the simulation, we use the Icarus simulator to fill the caches. Furthermore, we use it to generate a workload. Next, one link is broken at random. The algorithm detects the connected components that remain. Per connected component, it collects all information items. Note here, that if the component includes the source node, all information items are included in the component per definition.

Now, we can check per workload item if the end user and the information item requested are available in the same component. If this is the case, we have a successful information request. This way, it is possible to compute the percentage of successful requests over all workload items. Next, an extra link is broken, and the impact is computed the same way. This continues until no links remain.

The process above shows us one run of the simulation. There are three ways we have the possibility to perform an extra run in the simulation. The first option of these three is to generate a unique workload every time we break a link and check its impact. Secondly, it is possible to break all links once again and see per broken link what the impact is. Note that in these cases the caches are not refilled. The other option is to refill the caches (by running Icarus) and then break the links one by one again. The first option incorporates the randomness in the workload. The second option of making another run does incorporate the randomness in breaking the links. However, it does not incorporate the randomness in filling the caches. Although the cache filling largely depends on the strategy chosen, there is some randomness involved, which is also closely related to the workload generation (by the Zifp distribution).

In every simulation it is possible to decide whether a new workload generation should be done every time we break a link. Furthermore, the number of times we keep the same cache filling but break the links one by one can be decided upon. The same holds for the number of times that we can restart the whole simulation and refill the caches by running Icarus. Finally, we take the mean over all results. Unless stated differently, the default setting is to refill the caches 10 times, per refill break all links 10 times, and not generate a new workload every broken link. Note that this way we have 100 different simulations. This is the default setting because running Icarus (refilling the caches) and generating a new workload are rather time costly operations.

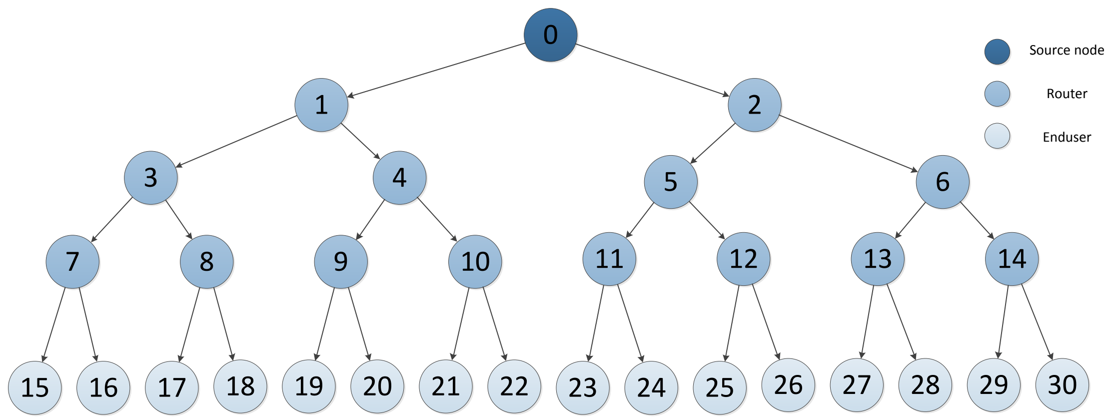


Figure 5.2: A binary tree with 3 layers of caches

5.3.3. No Cache

A quick note should be made on approximating the results in case there is no cache. If there is no cache, there should be a connection between the end user and the source. If we compare with the Equation 5.30, we obtain the following equation for no cache:

$$P(\text{succesful request}) = \left(\prod_{j=1}^{n-1} \frac{N * (1-f) - j}{N - j} \right), \tag{5.31}$$

where again f is the fraction of broken links, N is the total number of edges and n the total number of layers, where 1 is the layer of end users.

5.3.4. Results

In the following section, some results are shown for different networks (paths and trees) and different cache sizes ($10\%, 20\%$ and 50% of the content population).

In Figure 5.3, we find the cache hit ratio of a path graph with 3 caches, content population 100, and Leave Copy Everywhere as the strategy. As the path contains only 4 links, the resilience figures (Figures 5.4, 5.5 and 5.6) have only 5 data points. In Figure 5.7, the same is done, but for an increase content population from 100 to 400. As we will see in the rest of the results, the shape of the graphs do not differ a lot when increasing the population size. Naturally, the computation time does increase.

In Figure 5.11, the results for the binary tree topology with 3 layers of caches can be found (see Figure 5.2 for the topology). The LCE strategy is used in this case. We see similar behaviour as before, where the resilience is approximated very well, mainly because the cache hit ratio's can be approximated. Again, we see a large increase in resilience when the cache size increases. In Figure 5.15, the content population is again increased to 400. The patterns is comparable to a smaller population.

Next, we consider the path topology with a content population of 100, but with the Leave Copy Down strategy. We do this only for a path topology, as Leave Copy Down will return later in this thesis. The results are found in Figures 5.19, 5.20, 5.21 and 5.22. Comparing these results to LCE, the resilience of LCD is higher.

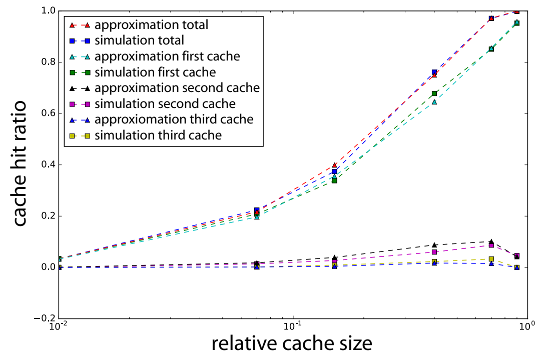


Figure 5.3: Cache hit ratio of path graph with 3 caches, a content population of 100 and caching strategy LCE

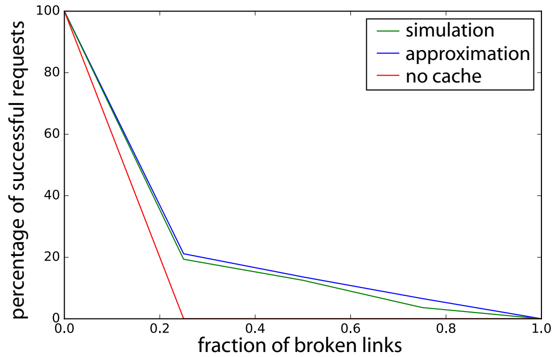


Figure 5.5: The resilience for cache size of 10 (10% of content population)

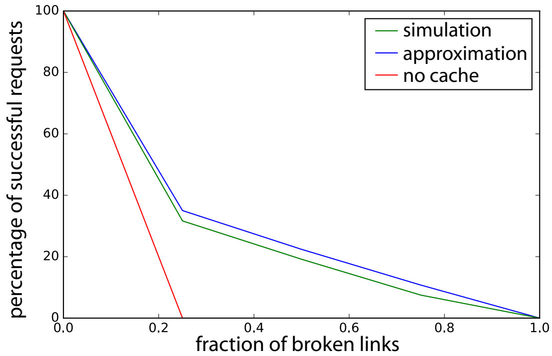


Figure 5.4: The resilience for cache size of 20 (20% of content population)

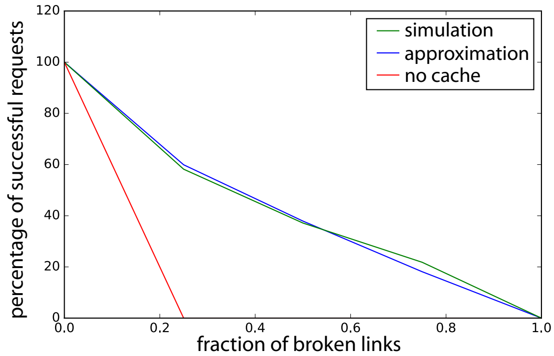


Figure 5.6: The resilience for cache size of 50 (50% of content population)

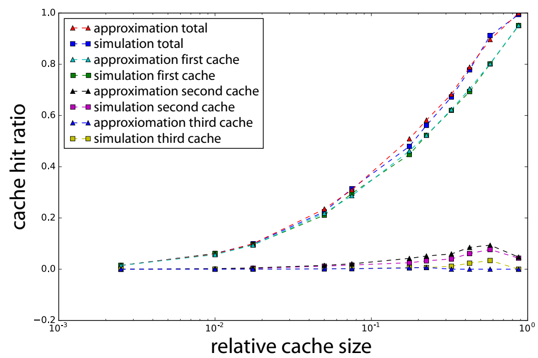


Figure 5.7: Cache hit ratio of path graph with 3 caches, a content population of 400 and caching strategy LCE

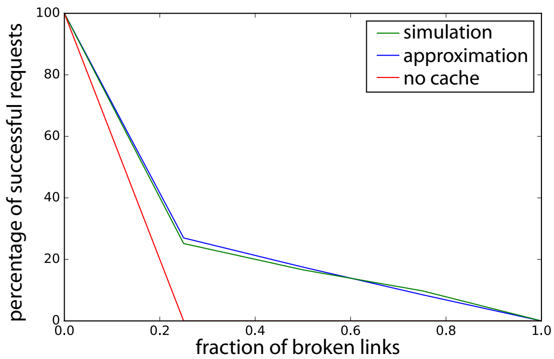


Figure 5.9: The resilience for cache size of 40 (10% of content population)

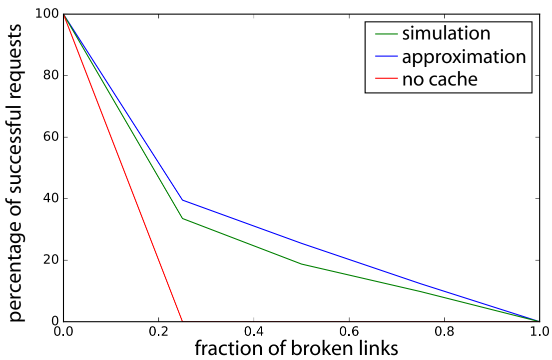


Figure 5.8: The resilience for cache size of 80 (20% of content population)

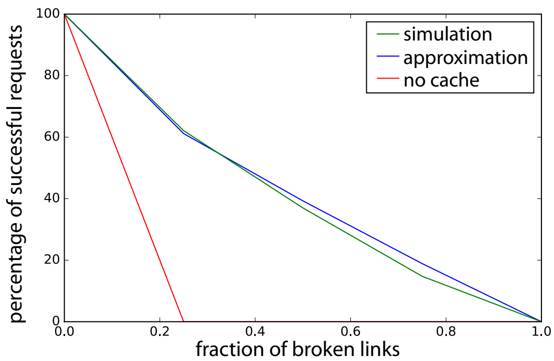


Figure 5.10: The resilience for cache size of 200 (50% of content population)

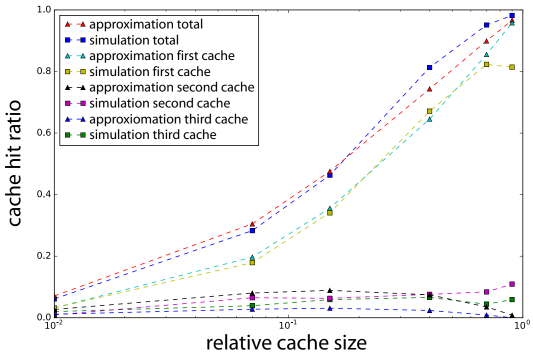


Figure 5.11: Cache hit ratio of tree graph with 3 layers of caches (30 links), a content population of 100 and caching strategy LCE

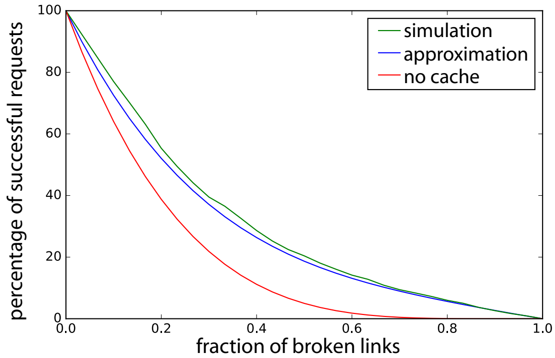


Figure 5.13: The resilience for cache size of 10 (10% of content population)

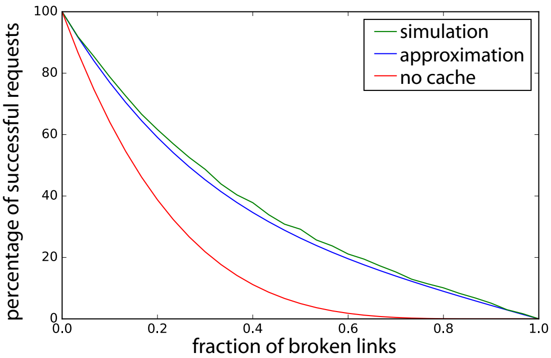


Figure 5.12: The resilience for cache size of 20 (20% of content population)

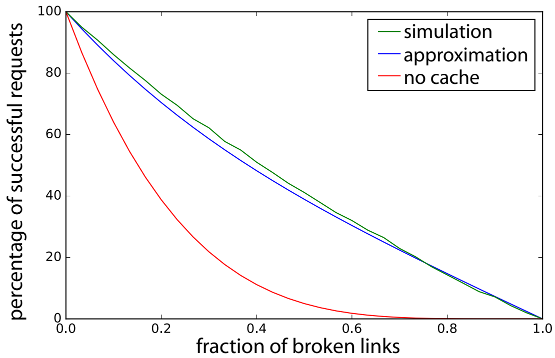


Figure 5.14: The resilience for cache size of 50 (50% of content population)

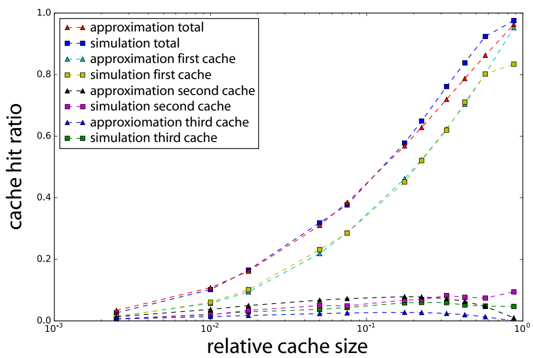


Figure 5.15: Cache hit ratio of tree graph with 3 layers of caches (30 links), a content population of 400 and caching strategy LCE

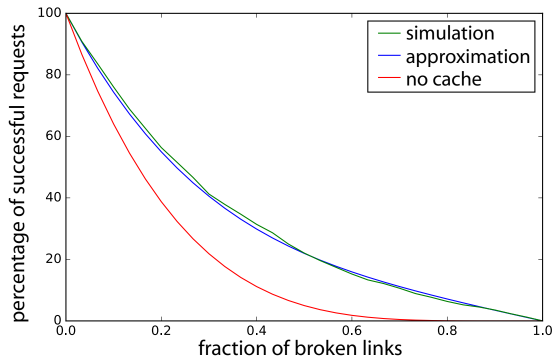


Figure 5.17: The resilience for cache size of 40 (10% of content population)

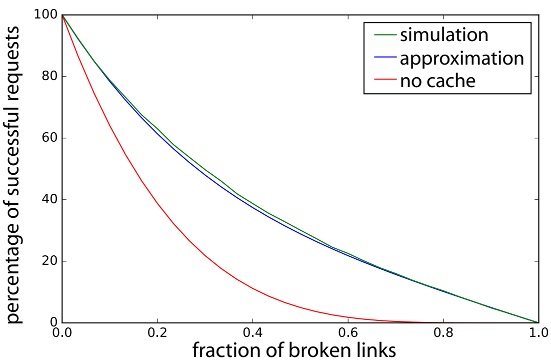


Figure 5.16: The resilience for cache size of 80 (20% of content population)

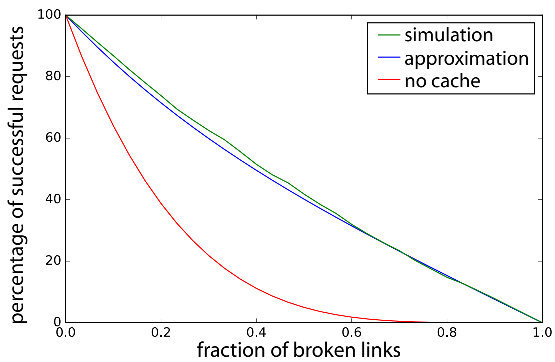


Figure 5.18: The resilience for cache size of 200 (50% of content population)

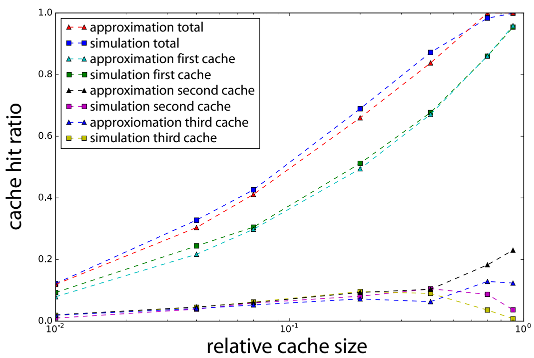


Figure 5.19: Cache hit ratio of path graph with 3 caches, a content population of 100 and caching strategy LCD

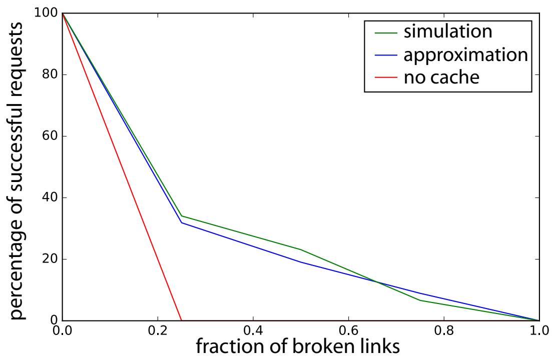


Figure 5.21: The resilience for cache size of 10 (10% of content population)

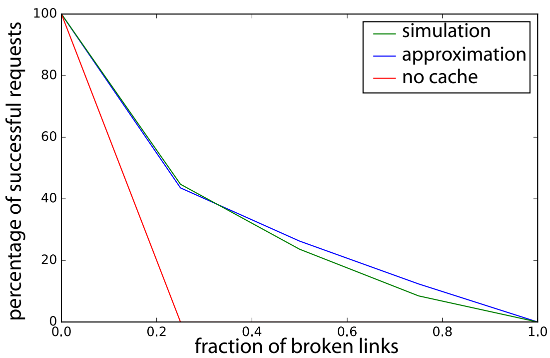


Figure 5.20: The resilience for cache size of 20 (20% of content population)

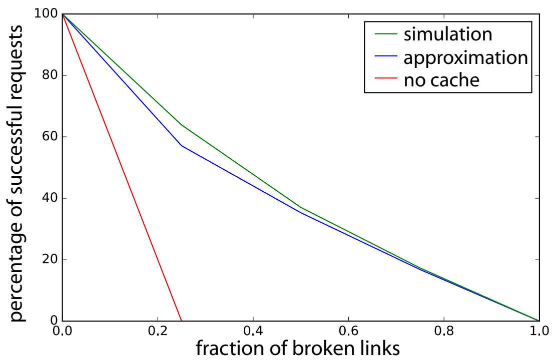


Figure 5.22: The resilience for cache size of 40 (40% of content population)

5.4. Information Resilience Using Che's Approximation: Hierarchical Structure

Now that we can approximate the information resilience of paths and trees, we turn our attention to slightly more complex networks.

5.4.1. Topology

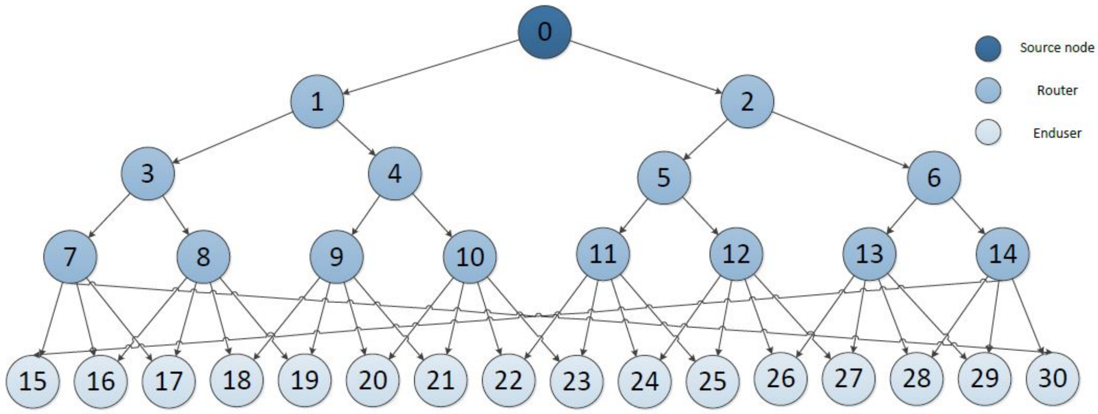


Figure 5.23: cache hit ration comparison LCE strategy

In this section, we consider the topology of Figure 5.23. In this topology, the basis is a binary tree with 5 layers of nodes (hence, the basis is Figure 5.2). Furthermore, each end user has a second connection to the lowest layer of routers. This is done in a systematic way such that all routers in the lowest level have 4 children.

This is done because of several reasons. First of all, by adding extra links, the number of possible shortest paths to the source increases. This adds complexity to the network considered. In the simulation, we see that the simulation software systematically chooses the same shortest path to follow the requests.

Secondly, the network remains symmetric this way, which makes the analyses easier in the sense of having less dimensions to solve. This is of course a simplification of realistic networks, but when the networks become more complex, the computation itself does not become more complicated. The complexity of the amount of different computations that needs to be done does increase however.

Before diving into both the simulation and approximation in more detail, we need to take one important thing into consideration. When considering the tree networks as before, we considered the information to be available when the information item was available in the whole connected component that included the end user. This means that in essence a larger component was used to simulate the resilience than was used to approximate the resilience. The approximation namely made only use of the path in the direction of the source (so no paths down the tree were considered).

This was no significant problem up to this moment. This is mainly because in the tree structure, the number of links going down is limited, and the content in the caches is partly overlapping with their parents. Therefore, the added value of considering the nodes down the tree is very limited. However, now that we consider this new topology, this becomes a problem. Because of the many extra added links, the connected component is much larger than the paths up to the source.

It is therefore that we need to make a choice of the conditions that allow a network to retrieve information. One could say that on the short term, looking along the set of shortest paths to the source are the only allowed nodes to look for information items. On the longer term however, the existence of a path between the end user and the node that contains the information item should be enough to retrieve the data. For now, we focus on short term effects, and will thus only consider the nodes along the shortest paths to the source when computing the information resilience. In Section 6.4, the pseudo code of the information resilience will be elaborated on. For now, the reader should keep the assumptions made above in mind.

5.4.2. Approximation

In this section, we will elaborate on the the approximation of the cache hit ratio and the information resilience of the network.

Cache Hit Ratio

For this topology we pick the LCE caching strategy. The cache hit ratio is again being computed with the approximations from Garetto et al. (2016). This time however, each end user is connected to two routers. One could assume that they send half of their requests to each of these two routers. The routers on the other hand, have twice as many links to their children as the other routers in the network. The result of this cancels out in the cache hit ratio. Hence, the computation of the cache hit ratio of the topology in Figure 5.23 is equal to the one in Figure 5.2. This is explicitly not always the case when extra links are added, but in this specific case it is.

No Cache

When there is no caching capability in the network, an information request can be dealt with when there is a connection between the end user and the source node. There are exactly two possible paths from each end user to the source. Also, to reach the source, there needs to be a path of length 4. For reasons of simplification, we will consider the paths to be entirely unique, even though they might (largely) overlap.

If we define f to be the fraction of links that has been broken, we obtain the following equation for the approximation of the probability of a path from the end user to the source.

$$\begin{aligned} P(\text{a path from the end user to the source}) &\approx 1 - P(\text{no path}) \\ &= 1 - (1 - P(\text{one path}))^2 = 1 - (1 - (1 - f)^4)^2 \end{aligned} \tag{5.32}$$

Note here, that the probability of a path, $(1-f)^4$, is a simplification. Furthermore, the two paths are not independent from each other, which is assumed here. But even though we are using a simplification, we shall see in Section 5.4.3 that the results give no reason for reconsidering this simplification.

Resilience Approximation With Cache

Having a cache increases the resilience, as we have seen before. Now that we know the probability of a path to a layer (see Section 5.4.2), we can extend it easily to a formula for approximating the information resilience in an ICN. This is done equally as in equation 5.30. Again, let f be the fraction of broken nodes, and let n be the number of layers (including the source). We obtain:

$$P(\text{succesfull request}) \approx \sum_{i=1}^{n} \left((1 - (1 - (1 - f)^i)^2) * P_{HIT}(i) \right) \tag{5.33}$$

5.4.3. Results

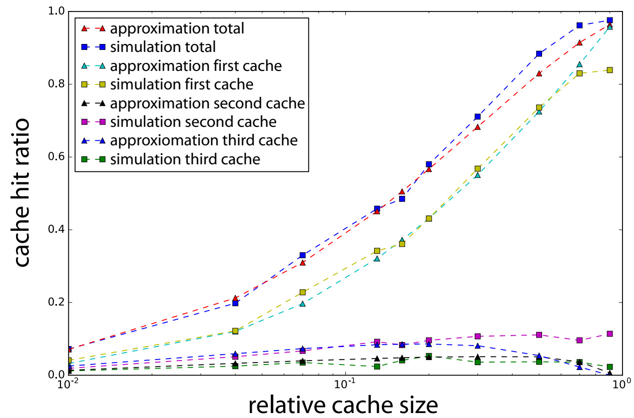
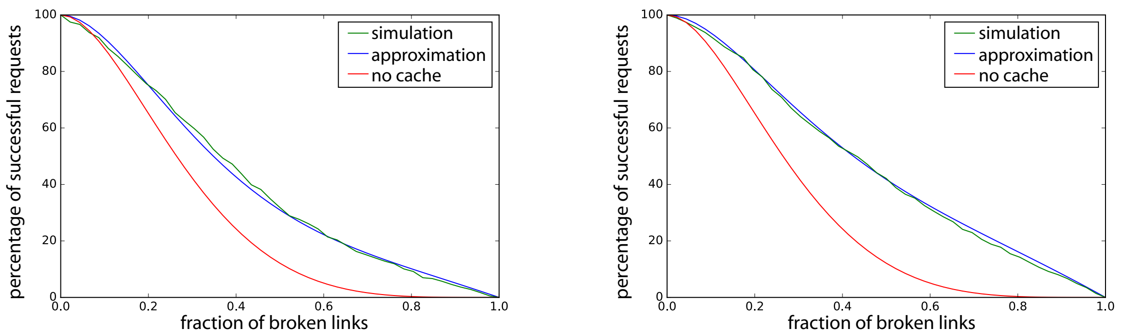


Figure 5.24: Cache hit ratio of hierarchical structure, a content population of 100 and caching strategy LCE



(a) The resilience for cache size of 10 (10% of content population)

(b) The resilience for cache size of 20 (20% of content population)

Figure 5.25: The resilience of Figure 5.23 for different cache sizes

The cache hit ratio as well as the information resilience for two cache sizes (10 and 20) are depicted in the Figures 5.24-5.25b. Note that the information resilience has greatly increased with respect to the tree network without the extra links. Let us compare this topology to the topology without extra links (Figure 5.2). If we compare Figure 5.12 with Figure 5.25a. We see that when 1/5th of all links break, about 55% of all information requests in the original network (no extra links) is successfully dealt with, whereas this is 75% in the topology with extra links. In case that the cache size is 20 instead of ten (hence we compare Figure 5.13 with Figure 5.25b), the increase of the information availability with a fraction of broken links of 1/5 increases from 60% to over 80%. We clearly see a large advantage in adding the extra links. This increase is mostly due to the impact of the existence of paths rather than the impact from the cache hit ratio.

However, as the cache hit ratio has not changed, this increase is solely due to the different topology. We see that also when we compare the two networks in case of no cache (the red lines in i.e. Figure 5.12 and Figure 5.25a). If we again would break 1/5th of all links, we would have a little less than 40% successful requests in the tree topology, and about 65% in the topology of Figure 5.23.

6

Applying the Analytical Results to a Realistic ICN

In this chapter, a realistic network will be considered. To what extent can we apply our theoretical results to find the information resilience of this network? In this chapter, we first select a network, then select a dataset. We compute some simulation results as well as some analytical results. Next, we compare the results.

6.1. Realistic ICN

As ICN is not deployed on a real scale yet, our "realistic" network will have to remain artificial. Often, when ICN's are used in calculations, networks concerning internet of things (IoT) or IPTV are being used as an inspiration for such a network. Due to accessibility reasons, we consider an IPTV network here. The following structures appear in most IPTV networks:

1. One or multiple sources
2. A so-called core network
3. A so-called access network, of which at least the DSLAM (digital subscriber line access multiplexer) is part
4. Home networks

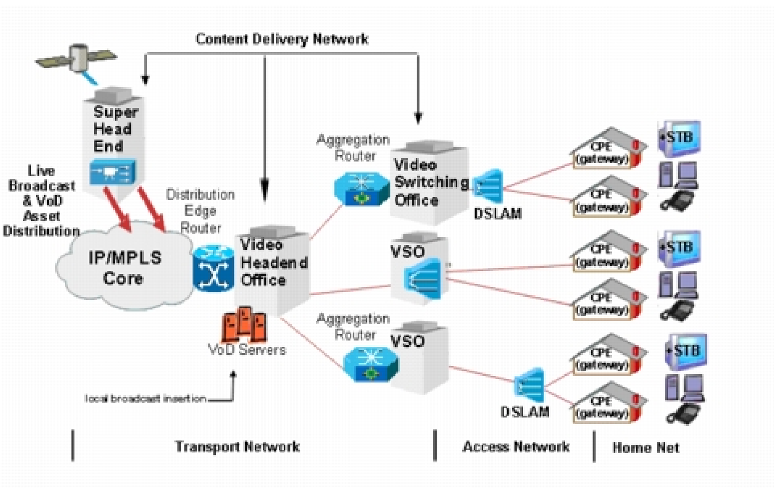


Figure 6.1: An IPTV Network, retrieved from (Cisco, 2012)

An example of a network that includes al these factors is being depicted by Figure 6.1. It is depicting a typical IPTV network, according to Cisco (2012). When generalizing all this information to a fictive though realistic network, we obtain the network of Figure 6.2.

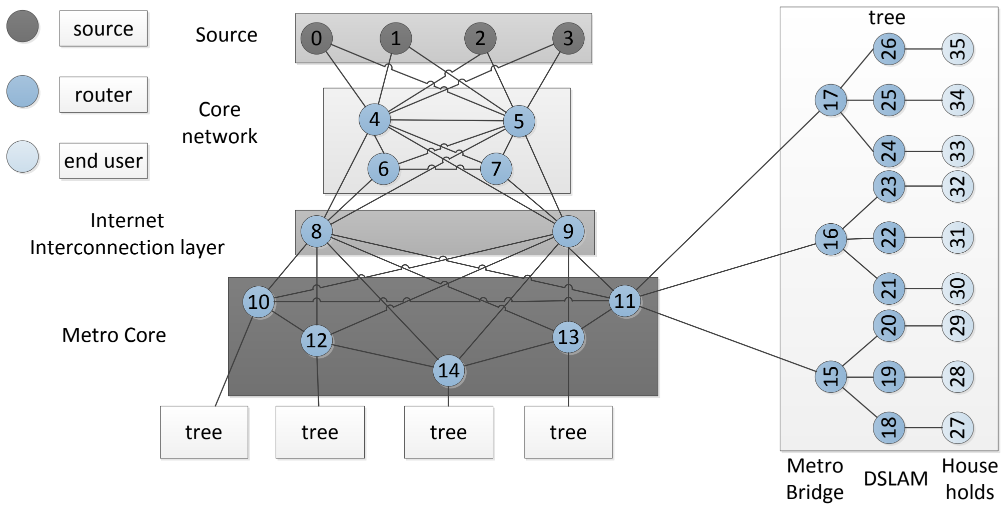


Figure 6.2: Realistic network for analyses, inspired by Figure 6.1

6.2. Realistic Data

In the previous section, we have selected a realistic network. As there is also a need for selecting data, this will be the matter of concern for this section. As for example shown by (Adamic & Huberman, 2002), the number of pages visited is following Zipf's law. In this section we will select a trace of information requests. As we are considering an IPTV network, it would be ideal to take a trace from for example Youtube to use for our analysis. Unfortunately, this information is not available. It is therefore that we will use traces of Wikipedia page views. For that, we have used the Wikipedia API[1] to retrieve the number of page views for the most popular Dutch Wikipedia pages on January 16th 2017. We sort this data on popularity, and remove the page views, such as the main page of Wikipedia, that are clearly not the page that visitors wanted to visit, but merely a means to reach the page of interest. Combining all this, we obtain the results in Figure 6.3. Here, the 1000 pages are sorted based on their number of views. They are depicted on a logarithmic scale on the horizontal axes. The number of corresponding page views is depicted on the vertical axes.

If we use the same amount of total page views, and we also take 1000 pages (items), the Zipf-law data generation with $\alpha = 0.8$ as built in in Icarus obtains the results that can be seen in Figure 6.3. We see here that the two figures seem to be similar. However, if we do not consider the first hit in the Wikipedia traces, but we do consider the second to the 1000th item, it can be seen that the Wikipedia data seems to be less skewed than the Zipf data. By that, we mean that the second most visited item from the Zipf generated list is visited more often than the Wikipedia data, and the 1000th most visited item from the Zipf generated list is visited less often than the same item from the Wikipedia data. The real data is compared to the generated data later on, to see if the differences in the data have a large impact on the resilience.

6.3. Cache Hit Ratio

As discussed in Chapter 5, computing the information resilience is done in a two step approach. First of all, the cache hit ratio is computed, which we will do here. Although computing the cache hit ratio

[1]See also https://wikitech.wikimedia.org/wiki/Analytics/PageviewAPI for more information.

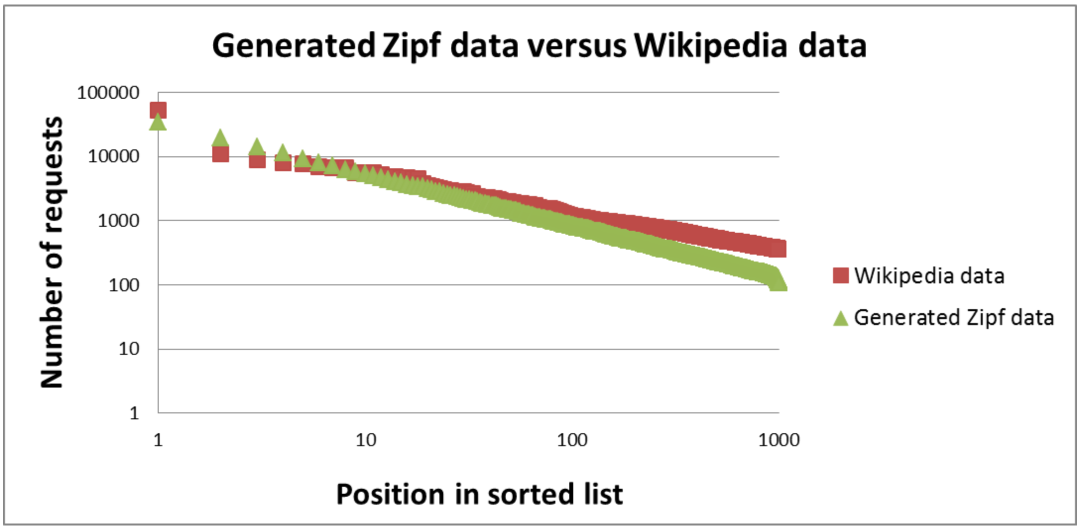


Figure 6.3: Representation of the 1000 most popular Dutch Wikipedia pages on January 16th 2017 as well as generated data with Zipf law using the same amount of items and total views.

here is not very different from what it was before, we will give some attention to it. Especially, the approximation is in some sense unique as the set of equations is different for every topology.

The simulation however, does still work the same way. After one run of Icarus, the cache hit ratio is being determined by Icarus itself.

6.3.1. Approximation

Let us consider the pseudo code of the approximation. The strategy LCE is used here. Algorithm 2 is optimizing the set of equations, and is returning the cache hit ratio. The equations that need to be solved are depicted in the function in Algorithm 3. For each iteration in finding the solution, the hit ratio and t_C need to be computed with the help of Algorithm 4 and 5. Note that the numbers in the subscripts of the variables in the algorithms correspond to the nodes in one layer in the topology (see Figure 6.2). Note furthermore that the variables are written as vectors, where each entry in the vector represents a different content item. It is therefore that we do not use specific content items in this section. Rather, all variables are vectors, and when a multiplication of vectors is written, we mean in fact an entry-wise multiplication rather than a vector multiplication.

In Algorithm 3, we have initiated the result function, of which the root is found in Algorithm 2. Take for example row 7 of the algorithm. If the result vector is equal to zero, this means that $3 \cdot (\lambda_{1826} \cdot (1 - Pin_{1826})) = \lambda_{1517}$, which is in line with solving the set of equations for solving the cache hit ratios from (Garetto et al., 2016). Note that in every iteration, all λ except from λ_1 change, and by that all t_C change. That in turn causes the P_{hit} to change. This is exactly why we need an iterative solver for this.

Now, let us pay some specific attention to the factors used in the equations in line 6,7,8,9 and 10 of Algorithm 3. For general cache networks, Garetto et al. (2016) point out that

$$\overline{\lambda}_i = \sum_j \overline{\lambda}_i (1 - P_{hit}(j)) r_{j,i} \tag{6.1}$$

where $r_{j,i}$ is the fraction of requests for object m that are forwarded from cache j to cache i. Note that this fraction is zero if the two nodes are not in consecutive layers. With this in mind, let us consider λ_{1517}, and in particular node 15. Note that $r_{j,i} = 0$ for all nodes j, except from node 18,19 and 20. For these three nodes, $r_{j,i} = 1$. Hence,as the λ is equal for all three nodes,

$$\overline{\lambda}_{1517} = 3 * \overline{\lambda}_{1826}(1 - P_{hit}(1826)) * 1. \tag{6.2}$$

The same line of reasoning applies to the other $\bar{\lambda}$'s in the equations.

For the P_{hit} that is computed in Algorithm 5, we also refer to (Garetto et al., 2016). They state that, the probability that item i hits cache m given that it had been forwarded from cache n is equal to

$$P_{hit}(m, i|j) \approx 1 - \ e^{A_{m,n}}, \tag{6.3}$$

where

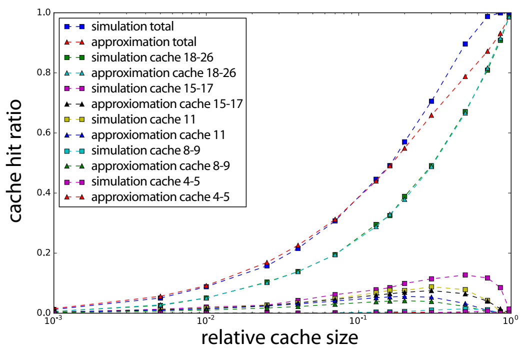
$$A_{m,n} = r_{n,m}\bar{(\lambda)}(n)(1 - p_{in}(n, i)) \max(0, t_C^{(m)} - t_C^{(n)}) + \sum_{p \neq n} \bar{(\lambda)}(i)(1 - p_{in}(p, i)) t_C^{(m)}. \tag{6.4}$$

Algorithm 2 Algorithm of computing the cache hit ratio

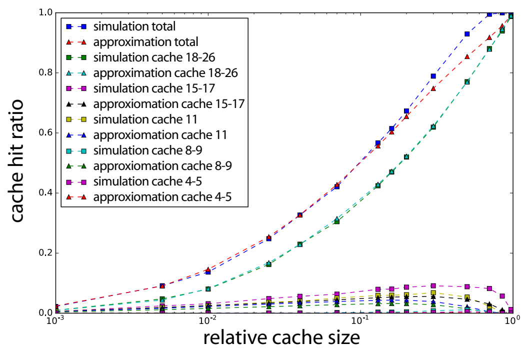
▷ Optimize function f (as specified in Algorithm 3)
solve = optimize.root(f, args = (lambda1,x0,), method='excitingmixing')
result =solve.x
▷ compute the cache hit ratio by using the result vector $(\lambda \cdot P_{hit})$ and the number of nodes per layer
$s = 9 \cdot (\lambda_{1826} \cdot P_{hit}(1)) + 3 \cdot (\lambda_{1517} \cdot P_{hit}(2)) + (\lambda_{11} \cdot P_{hit}(3)) + 2 \cdot (\lambda_{89} \cdot P_{hit}(4)) + 2 \cdot (\lambda_{45} \cdot P_{hit}(5))$
▷ Do the same per layer instead of the total
$P_{hit} = \sum(s)$
return P_{hit}

6.3.2. Results of Cache Hit Ratio

In this section, we simulate and approximate the cache hit ratio's for both the generated Zipf data and the Wikipedia data. Again, the simulation and approximation fit together very well. In Figure 6.4a, the results with Wikipedia data can be found. In Figure 6.4b, the same can be found for the generated Zipf data (as we did before).



(a) Cache hit ratio of a realistic network, using Wikipedia data, a content population of 1000, strategy is LCE



(b) Cache hit ratio of a realistic network, using generated data, a content population of 1000, strategy is LCE

Figure 6.4: Cache hit ratios for two types of popularity distribution

What we see here, is a good fit between the simulation and the approximation. However, between the generated data and the Wikipedia data, there is quite a large difference. We have seen before that the popularity between the data seems to be alike (with the right parameters in the generated data). However, the cache hit ratio is less alike. The cache hit ratio is substantially higher in the generated data than it is in the Wikipedia data. This can be explained as follows: although the graphs of Figure 6.3 seem to be rather similar, they do differ a bit. If we sum the 20 most popular messages, we find that there is a probability of 0.2045 for the Wikipedia data that a request for information is one of these first 20 items. However, for the generated data, we find that this probability is equal to 0.3102. As we can see in Figure 6.4a and 6.4b, the impact of this is rather large.

Algorithm 3 Algorithm of solving the set of equations for cache hit ratio of the realistic network

Load the proper λ
n=len(λ)
def $f(x, \lambda_1, x_0)$:
initiate $\lambda_{1826}, \lambda_{1517}, \lambda_{11}, \lambda_{89}, \lambda_{45}, Pin_{1826}, Pin_{1517}, Pin_{11}, Pin_{89}, Pin_{45},$
$P_{hit}(1517), P_{hit}(11), P_{hit}(89), P_{hit}(45),$ a result vector
▷ Now we set up the computation
res[0:n] = $\lambda_1 - \lambda_{1826}$
res[$n : 2 \cdot n$]=$3 \cdot (\lambda_{1826} \cdot (1 - Pin_{1826})) - \lambda_{1517}$
res[$2 \cdot n : 3 \cdot n$]=$3 \cdot (\lambda_{1517} \cdot (1 - P_{hit}(1517))) - \lambda_{11}$
res[$3 \cdot n : 4 \cdot n$]=$\frac{1}{2} \cdot (\lambda_{11} \cdot (1 - P_{hit}(11))) - \lambda_{89}$
res[$4 \cdot n : 5 \cdot n$]=$(\lambda_{89} \cdot (1 - P_{hit}(89))) - \lambda_{45}$
▷ compute the TC with Algorithm 4
Tc_{1826} = optimize.root(compute$_{tc}$, $x0, args = (lambda1826,)$)
Tc_{1517} = optimize.root(compute$_{tc}$, $x0, args = (lambda1517,)$)
Tc_{11} = optimize.root(compute$_{tc}$, $x0, args = (lambda11,)$)
Tc_{89} = optimize.root(compute$_{tc}$, $x0, args = (lambda89,)$)
Tc_{45} = optimize.root(compute$_{tc}$, $x0, args = (lambda45,)$)
▷ Compute the P_{in}
res[$5 \cdot n : 6 \cdot n$]= $(1 - e^{(-\lambda_{1826} \cdot * Tc_{1826})}) - Pin_{1826}$
res[$6 \cdot n : 7 \cdot n$]= $(1 - e^{(-\lambda_{1517} \cdot Tc_{1517})}) - Pin_{1517}$
res[$7 \cdot n : 8 \cdot n$]= $(1 - e^{(-\lambda_{11} \cdot Tc_{11})}) - Pin_{11}$
res[$8 \cdot n : 9 \cdot n$]= $(1 - e^{(-\lambda_{89} \cdot Tc_{89})}) - Pin_{89}$
res[$9 \cdot n : 10 \cdot n$]= $(1 - e^{(-\lambda_{45} \cdot Tc_{45})}) - Pin_{45}$
▷ Compute the P_{hit} with Algorithm 5
res[$10 \cdot n : 11 \cdot n$]= $compute_{P_{hit}}(P_{hit}(1517), Pin_{1826}, \lambda_{1826}, Pin_{1826}, Tc_{1517}, Tc_{1826},$
$\lambda_{1826}, Pin_{1826}, Tc_{1517}, \lambda_{1826}, Pin_{1826}, Tc_{1517})$
res[$11 \cdot n : 12 \cdot n$] =$compute_{P_{hit}}(P_{hit}(11), Pin_{1517}, \lambda_{1517}, Pin_{1517}, Tc_{11}, Tc_{1517}, \lambda_{1517},$
$Pin_{1517}, Tc_{11}, \lambda_{1517}, Pin_{1517}, Tc_{11})$
res[$12 \cdot n : 13 \cdot n$] =$compute_{P_{hit}}(p_{hit}(89), Pin_{11}, \lambda_{11}, Pin_{11}, Tc_{89}, Tc_{11}, 0, 0, 0, 0, 0, 0, 0.5)$
res[$13 \cdot n : 14 \cdot n$] =$compute_{P_{hit}}(p_{hit}(45), Pin_{89}, \lambda_{89}, Pin_{89}, Tc_{45}, Tc_{89}, 0, 0, 0, 0, 0, 0, 0.5)$
return res

Algorithm 4 Definition of computing Tc

def $compute_{tc}(T, \lambda)$:
$G = 1 - e^{(-\lambda * T)}$
return $\sum(G)$ − cache size

Algorithm 5 Definition of computing P_{hit}

def $compute_{P_{hit}}(P_{hit}(5), P_{hit}(1), \lambda_1, Pin_1, Tc_9, Tc_1, \lambda_2, Pin_2, Tc_{10}, \lambda_3, pin_3, Tc_2,$ratio=1):
S = ratio$\cdot(\lambda_1 \cdot (1 - Pin_1) \cdot max(0, Tc_9 - Tc_1))$+ratio$\cdot(\lambda_2 \cdot (1 - Pin_2) \cdot Tc_{10})$+ratio$\cdot(\lambda_3 \cdot (1 - pin_3) \cdot Tc_2)$
return $((1 - P_{hit}(1)) * (1 - e^{(-S)})) - p_{hit}(5)$

6.4. Resilience

After the computation of the cache hit ratio, the information resilience can again be computed. Before moving to the approximation, let us first give some attention to the simulation.

6.4.1. Simulation

The simulation is brute force, meaning that we break all links one by one and evaluate the effects. For the number of simulations, two different parameters are important. First of all, the number of runs from the Icarus simulator needs to be determined. In a new simulation, the caches are refilled according to the algorithm in Icarus. This part of the simulation is not shown in the Algorithm 6. Next, per Icarus run, the links are broken at random one by one. This is incorporated in Algorithm 6. After each newly broken link, the impact is analysed. When all links are broken, one run of the simulation is finished. For the next iteration, the entire network is restored, and the procedure is repeated. The number of times this procedure is repeated is the second parameter that is important.

Now, how do we analyse the impact of breaking a link? First of all, the set of descendants from all nodes is being determined, and all nodes in such a set are being determined. Next, find out which items are stored in the caches in each of these sets. Now, we take the receivers of these sets into account, and check for each request of information if the end user and the requested piece of information can be found in the same set. The number of succesful requests is counted against the total number of requests, and this is averaged over all runs.

6.4.2. Approximation

Now, like we did before, let us build the resilience formula. Like before, the formula consists of the probability that a path to a certain layer exists times the probability of a cache hit. Hence,

$$
\begin{aligned}
P(\text{succesful request}) \approx \sum_{i=1}^{5} (P(\text{path to layer i exists}) \cdot P_{HIT}(i)) + \\
P(\text{path to layer 6 exists}) \cdot (1 - \sum_{i=1}^{5} P_{HIT}(i)).
\end{aligned}
\tag{6.5}
$$

For the first three layers, the probability that there exists a path to a specific node in the layer i, given that a fraction f of the nodes is being broken, is equal to: $P(\text{path to layer i exists}) = (1-f)^i$. For layers 4 and 5 as well as for the source, the probability of a path, given f, is equal to:

$$
P(\text{path to layer i exists}) = ((1-f)^3 \cdot \left(2 \cdot (1-f) - (1-f)^2)\right)^{i-3}) = (1-f)^3 \cdot (1-f^2)^{i-3}. \tag{6.6}
$$

Now, if we put all this together, we obtain the formulas in Equation 6.7 and 6.8

$$
\begin{aligned}
P(\text{succesful request}) \approx \left(\sum_{j=1}^{3} (1-f)^j \cdot P_{HIT}(j) + \sum_{k=4}^{5} (1-f)^3 \cdot (1-f^2)^{k-3}) \cdot P_{HIT}(k) \right) \\
+ \left((1-f)^3 \cdot (1-f^2)^3) \cdot \left(1 - \sum_{i=1}^{5} P_{HIT}(i) \right) \right)
\end{aligned}
\tag{6.7}
$$

Hence, the formula for resilience without caching capacity is

$$
P(\text{succesful request}) \approx ((1-f)^3 \cdot (1-f)^3 \cdot (1-f^2)^3). \tag{6.8}
$$

Algorithm 6 Algorithm of computing resilience

```
while l ≤ Number of Simulations: do
    build DiGraph from topology that points from endusers to sources
    while j ≤ Number of Edges: do
        remove one random edge
        for Node ∈ end users: do
            ▷ Find all decendants
            Sort[node]=nx.descendants(G,node)
        end for
        for  k,v in Sort do
            make list of all nodes in an element
            for  t in list: do
                ▷ make dict “a”: per node all content items that can be reached in de caches
                a[itt][int(t)]=cachesFinal[int(t)]
            end for
        end for
        supercounter=0
        while i ≤ len(eventlist): do
            for  k ,v in a: do
                ▷ only take those into account where the receiver is in the connected component.
                if  int(eventlist[i]['receiver']) in v: then
                    for  r,ss in v.items(): do
                        if  int(eventlist[i]['content']) in ss: then
                            ▷ Hence we have a succesful information request
                            supercounter+=1
                            ▷ the request is succesful, move to next item in eventlist
                            break
                        end if
                    end for
                end if
            end for
        end while
        ▷ Number of succesful requests divided by total number of requests
        yax = (supercounter / (len(eventlist)) * 100)
    end while
    save all yax items in a list called “measure”
end while
final = mean(measure)
```

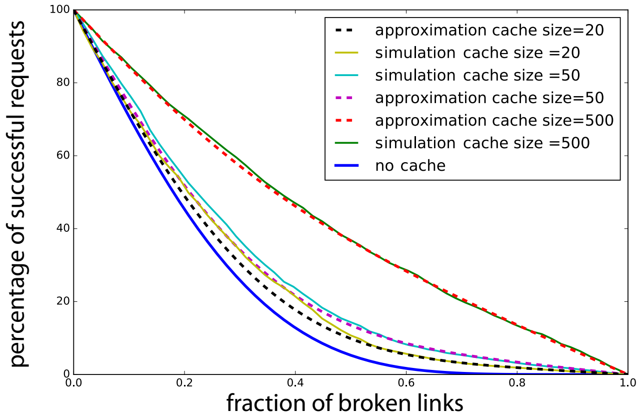


Figure 6.5: Resilience of realistic network with Wikipedia data with different cache sizes. Caching strategy is LCE

6.4.3. Results

In Figure 6.5 both the simulation and the approximation of the resilience can be found for different cache sizes. As expected, the added value of the caches increases when the cache size increases. Furthermore, the simulation and approximation align very well. The computation time of the simulation is much longer than the computation time of the approximation. So if we consider a larger content population, it is desirable to restrict ourselves to the approximation. As we will see in the next section, the caches are much more effective when changing the cache policy. When considering Leave Copy Down instead of Leave Copy Everywhere, the impact of the caches increases.

6.5. Influence of different parameters

In this section, we discuss the influence of three parameter changes. First of all, we change towards Leave Copy Down. Secondly, we increase the content population. Lastly, we increase the network size.

6.5.1. Leave Copy Down

In the Leave Copy Down strategy, nothing happens to the way we compute or simulate the resilience. The way the cache hit ratio is being computed differs. This is because the set of equations, as explained by Garetto et al. (2016), changes. The consequence is that the Algorithms 3, 4 and 5 change.

We see here that also for Leave Copy Down the approximation gives very reasonable results. The results show that the resilience increases here with regard to the LCE strategy. To make that visile, Figure 6.10 depict that the LCD and LCE strategy for the Wikipedia data population of size 1000, and the cache size of 20. We see that, when 40% of the links is broken, the successful requests increase from 18% to 21%.

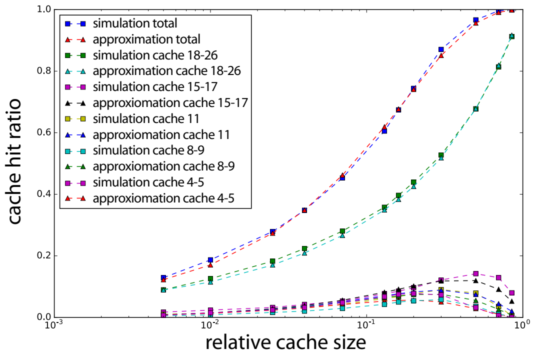


Figure 6.6: Cache hit ratio of realistic network with Wikipedia data with Leave Copy Down strategy

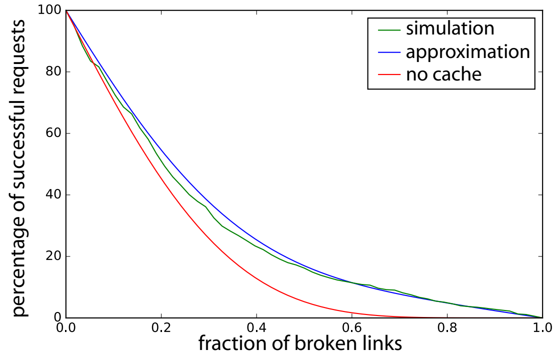


Figure 6.8: The resilience for cache size of 50 (5% of content population)

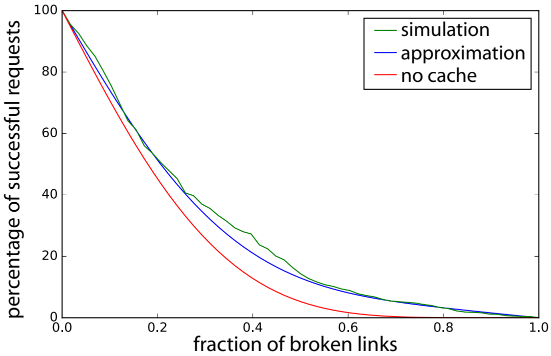


Figure 6.7: The resilience for cache size of 20 (2% of content population)

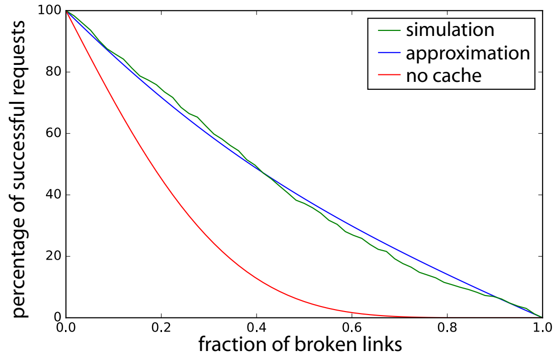


Figure 6.9: The resilience for cache size of 500 (50% of content population)

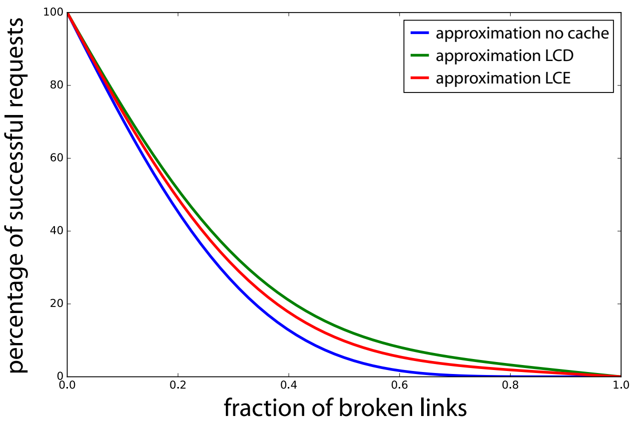


Figure 6.10: Resilience of the realistic network, Wikipedia data, 1000 content items, and different caching strategies.

Table 6.1: Some time indications of running the simulations

cache size	50	500	5	10	25	40	50	70
content size	1000 (wiki)	1000 (wiki)	1000 (wiki)	1000 (wiki)	1000 (wiki)	1000 (wiki)	1000 (wiki)	1000 (wiki)
network	KPN	KPN	KPN	KPN	KPN	KPN	KPN	KPN
strategy	LCE	LCE	LCD	LCD	LCD	LCD	LCD	LCD
simulation (sec)	31 sec	49 sec	23	20	21	23	22	26
cache hit ratio analytical (sec)	1,5 sec	14 sec	-	8	18	6	16	13
simulate resilience*	6.35 min	8 min	-	-	-	-	5.20 min	-

Table 6.2: More time indications of running the simulations

cache size	130	160	200	300	500	700	850
content size	1000 (wiki)	1000 (wiki)	1000 (wiki)	1000 (wiki)	1000 (wiki)	1000 (wiki)	1000 (wiki)
network	KPN	KPN	KPN	KPN	KPN	KPN	KPN
strategy	LCD	LCD	LCD	LCD	LCD	LCD	LCD
simulation (sec)	29	31	31	34	37	42	41
cache hit ratio analytical (sec)	6	5	7	21	72	20	25
simulate resilience*	-	-	-	-	5.20 min	-	-

The Tables 6.1 and 6.2 provide us with an indication of run times for some situations. It is just indicative.

6.5.2. Larger Content Population

For increasing the content population to 10000, the same still works. Now, we increase to 100.000, but only for the approximation. This is because the simulation simply takes too long now. The approximation also takes long, but still doable. Per cache size, the approximation takes between 40 and 6500 seconds.

See also Table 6.3 for the approximation time of one specific run. Note here that the computation times vary greatly. Furthermore, we do not see an absolute increase in the computation times when the cache size increases. First of all, note that this is a result due to one reason only. Secondly, we cannot fully explain this odd way of computation time distribution for different cache sizes. It is therefore that we will come back to this in the discussion section (Section 9) of this thesis.

Table 6.3: Time it takes to approximate the cache hit ratio, LCE, content population: 100.000

Cache size	time (sec)
1	42
50	1061
1500	599
7000	130
13000	504
20000	6449
50000	490
70000	352
85000	1139

Furthermore, for this large content population, the cache hit ratio is plotted for the first three layers of caches of the network in Figure 6.11. Note here that the simulation only is depicted for small cache sizes. This is because the number of measured/warm up requests needs to be much larger than the content population for larger cache sizes. This is especially a problem for larger cache sizes, as it might be that these caches are not entirely filled yet is the number of warm up requests is too small. For smaller cache sizes, this problem does not occur. It is therefore that we have added the simulation results for smaller cache sizes.

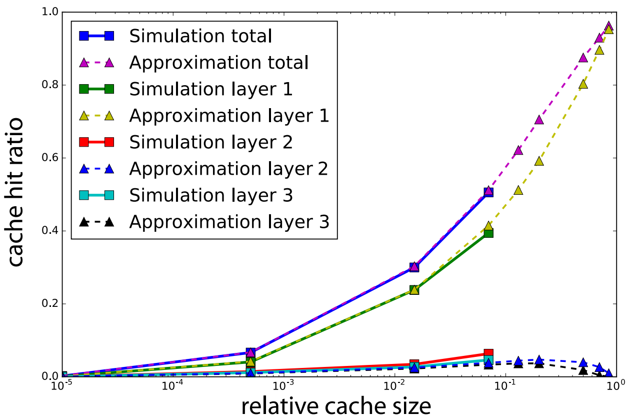


Figure 6.11: The cache hit ratio of the larger realistic network

6.5.3. Larger Network Topology

We have extended the topology in Figure 6.2. This network is similar to the network of IPTV-network considered there (same amount of layers), but has more nodes on the lower layers. Specifically, the source layer, core layer and internet interconnection layer remain the same. However, the metro core layer is extended to 8 nodes, each having 5 disjoint children nodes on the metro bridge layer, leading to 40 metro bridge nodes in total. Each metro bridge node has 5 disjoint children on the DSLAM layer resulting in a total of 200 DSLAM nodes. Each DSLAM is in turn connected to 6 distinct house holds, yielding to a total of 1200 house holds. To summarize, this topology has 1458 nodes: 4 sources, 254 routers, and 1200 end users. We use a content set of 1000 items (requested using a Zipf distribution with $\alpha = 0,8$) and the LCD caching strategy (cache size C = 20). Figure 6.12 reports the analytically computed information availability, as a function of broken links, for this network in this setup.

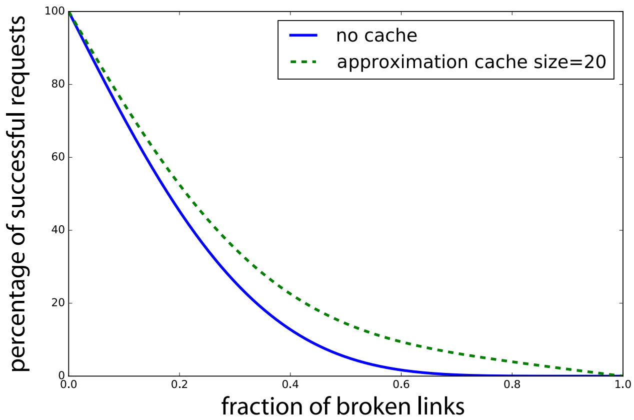


Figure 6.12: The resilience for cache size of 20 (2% of content population)

Here, having a caching capacity of two percent of the content population per node, increases the successful requests significantly only when a large fraction of the links is broken. If for example 20% of the links is broken, the information resilience increases from 45% to a little over 50%. Hence, we have an increase of about 11% of the information resilience.

7
Extension of Results to Non-Hierarchical Networks

In the previous sections, the analytical approximation of the information resilience in hierarchical networks has been key. In this chapter, we will extend some of the analytical results with simulation results. This is done for two reasons. First of all, the simulation results enhance the insights of the effects of certain parameters. Secondly, for non hierarchical network topologies, such as Barabási-Albert topologies (See Chapter 3) the information resiliency cannot yet be approximated with the analysis developed in this thesis.

As explained in Section 3.4, the caching part of the simulations are already developed for the ICN simulator Icarus. By leaving the caching part to Icarus, the analysis and impact of a breakdown should be implemented separately.

In this phase, the model remains static. No time or continuous caching is playing a role in this model. As soon as Icarus has run, the caches are filled with the workload. This can be done either by the generated workload according to the Zipf distribution, or according to the Wikipedia traces. In the analysis, the resilience is been measured during a breakdown of edges, as explained in Section 3.1.1. The results depend on several variables, as has been already discussed previously.

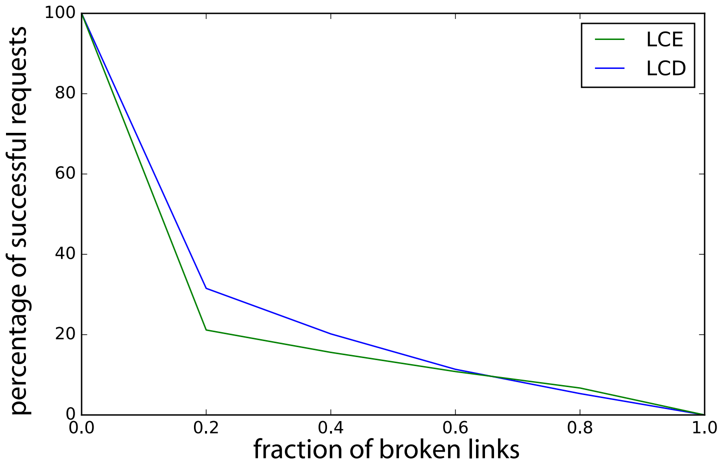
For the results, we have varied between a Waxman and a Barabási-Albert topology, of which more can be found in Chapter 3. Furthermore, there has been a variation in the size of the networks.

In the next section, a comparison between LCE and LCD is being simulated for different topologies. We see that LCD outperforms LCE in all cases. This is in line with the cache hit ratio of the different strategies. However, in individual break downs, this is not necessarily the case, as for certain topologies a specific set of broken links might have a very large impact on the information resilience. This will be elaborated on in Section 7.2.1. In Section 7.2, the simulation results of non hierarchical networks will be elaborated on.

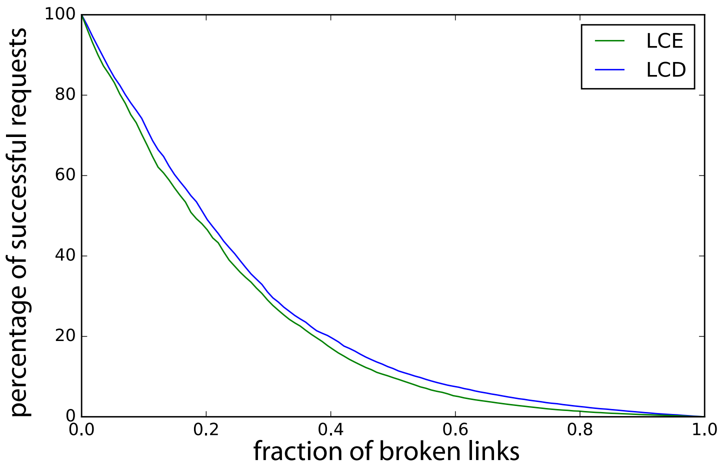
7.1. Comparing LCE and LCD

In line with the analytical results of Chapter 5, we compare the LCE and LCD strategy for several topologies. First of all, a path graph of length 6 ($n = 6$) is being considered, even though this is a hierarchical network. 10 iterations are performed. Per iteration, the caches are getting filled. Then, another 10 times, the edges are broken at random, and a new workload is being constructed. The average amount of workloads that can be finished is saved. The content population is the are the Wikipedid traces of 1000 items. The cache size is equal to 10% (100 items). In Figure 7.1a, the results can be found. This is in line with what we know about the performance of LCE and LCD.

However, when we change towards a complex network, we expect the LCE strategy to be more effective. This is because in that case, the popular content is closer to the end users, hence the



(a) LCE versus LCD in a path graph of length 6, cache size is 10%



(b) LCE versus LCD in a Barabási-Albert network with 108 nodes, $m = 1$, and cache size 1%

Figure 7.1: The resilience for different topologies, different caching strategies and different cache sizes

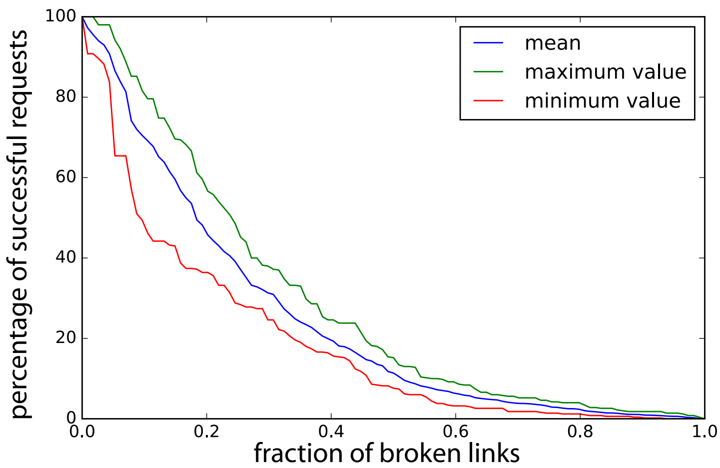
popular content is less sensitive for disturbances than in case of the LCD strategy. In Figure 7.1b, the same is depicted for a Barabási-Albert network with 107 links and 100 initial nodes (the number of links initially attached to the network $m = 1$), with extensions as end users (total of 108 nodes). This is done because end users can only be linked to one other node (which is not a source or an end user) according to the simulator Icarus. The content population is, like in the example above, 1000 Wikipedia traces, and the cache size is equal to 1% (10).

Furthermore, we have made a major adjustment with regards to the hierarchical networks. In the beginning of this thesis, we have considered all the components one end user is connected to as possible nodes to retrieve the information from. When considering more complex networks, we have changed this into the set of shortest paths. The way this has been simulated made use of the fact that the networks were hierarchical, and the numbers were systematically given to the nodes. As this is no longer possible in Barabási-Albert networks or Waxman networks, we will return to the connected component measurements as used in the beginning of the thesis. Note that the simulation results can therefore not be compared to the results of the hierarchical networks. We can, however, compare different parameters with each other. We see indeed that the LCE and LCD strategy function much more alike here than in the path strategy. It would most probably be possibly to build a simulation that takes the shortest paths into account, also for Barabási-Albert and Waxman networks. However, due to time limitations, we have chosen to restrict ourselves to the connected components way of simulating.

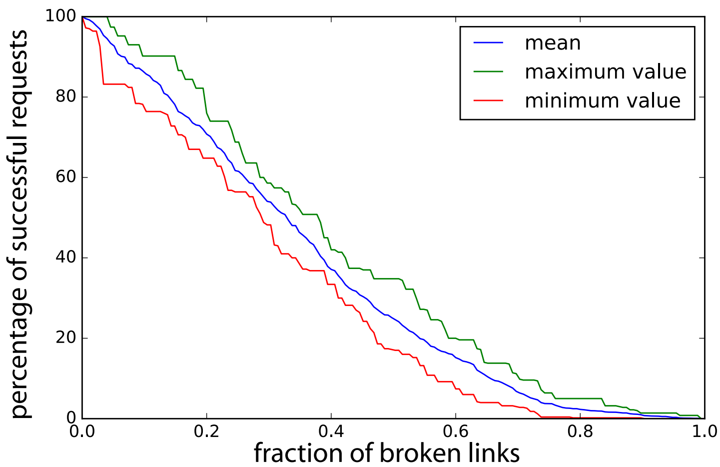
7.2. Barabási-Albert and Waxman Topology

In this section, we will consider Waxman and Barabási-Albert topologies. First, the error bars are considered. Secondly, different cache sizes are considered.

7.2.1. Error Bars of Broken Links



(a) Barabási-Albert network with $m = 1$ with errorbar, LCE



(b) Barabási-Albert network with $m = 2$ with errorbar, LCE

Figure 7.2: The resilience of Barabási-Albert networks for cache size 20, population 1000 (Wikipedia traces)

In this section, we will consider the Waxman and Barabási-Albert topology and take a look at the error bars of these results. We depict the maximum and minimum results as well as the mean. We compare two types of a Barabási-Albert topology with as well as an instance of the Waxman topology.

In Figure 7.2a, the mean, maximum and minimum values of the Barabási-Albert network with 107 links as described in the previous section is depicted. The LCE strategy is used here. Also here, 100 iterations have been performed. We see that the minimum and maximum value are less smooth than the mean. This is in line with our expectations, as the mean depicts the average over multiple simulations, and the extremes depict one particular extreme value. Secondly, in Figure 7.2b, we consider a different Barabási-Albert network. Also here, we have 108 final nodes. However, the number of links initially attached to the network, m is increased to $m = 2$. To achieve a situation with 108 nodes, we then need to reduce the initial network to 72 nodes instead of 100. This is because there are no nodes in the initial network with degree 1 left, which is, for the ICN simulator a requirement for the end users.

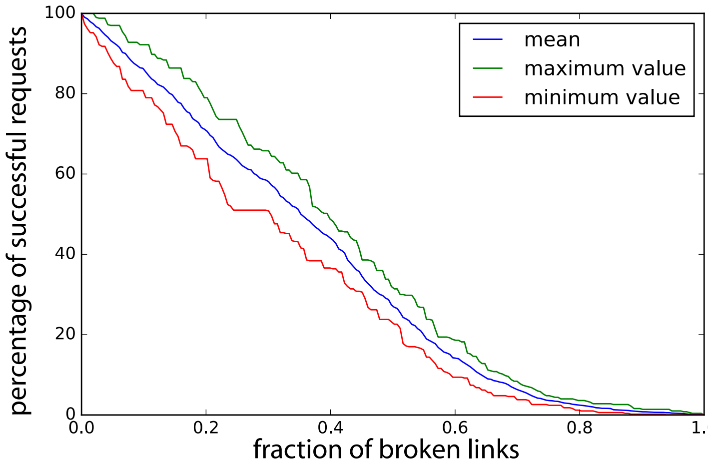


Figure 7.3: Waxman network with $\alpha = 0.05$, $\beta = 0.9$ with error bars, LCE, population of 1000 (Wikipedia traces) cache size 20

Because of the different parameter m, the number of links increases to 171. The other parameters remain the same.

We find an increase of the information resilience if we increase the parameter m. This is in line with our expectations, as the extra links reduce the effect of links being broken. However, if we would not consider the fraction of broken links, but the absolute number of broken links, the Barabás-Albert network with $m = 1$ is more resilient. If, for example, 20% of the links in Figure 7.2a is broken, there are 86 links left, and the information resilience is about 45%. If we want 86 links left in Figure 7.2b, we need to break 50% of the links. Then, the information resilience is about 25%. Although it is good to keep in mind that this is the case, we do not wish to change our measurement. This is because we are interested in the impact of broken links on the information resilience for different parameters. If we have a network with more links, then that is just another parameter. We are, given these parameters, interested in the impact of breaking links, which is exactly what is depicted in the figures.

In Figure 7.3, we have considered a Waxman network with 100 original nodes, and parameters $\alpha = 0.05$ and $\beta = 0.9$. Recall from Section 3.2 that the probability of an edge between any two nodes is equal to

$$p = \alpha \cdot e^{-\frac{\delta}{\beta \cdot L}}. \tag{7.1}$$

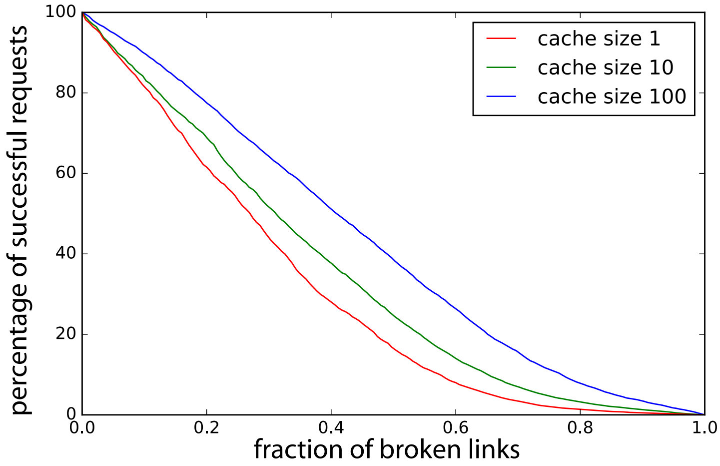
Also here, end users are added to the topology, to preserve the simulator assumption that end users have node-degree 1. As a result, we have 156 nodes and 213 links. The cache size depicted here is 20, which is 2% of the content population of 1000 Wikipedia article traces.

7.2.2. Different Cache Sizes

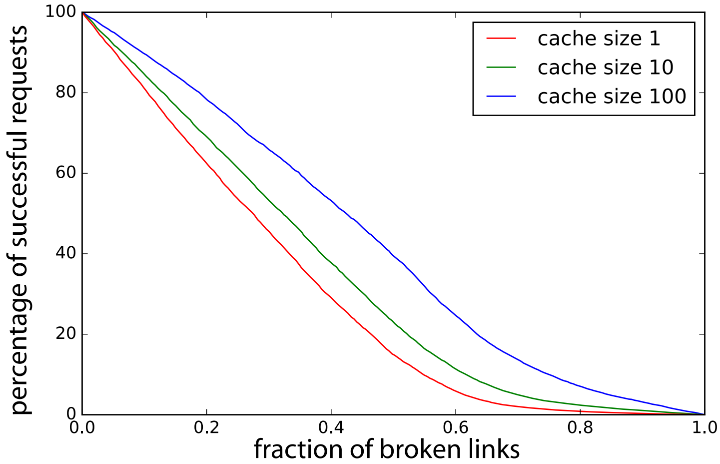
In this section, we briefly compare the information resilience for different cache sizes of both the Barabási-Albert and the Waxman topology. We consider the same Waxman topology as we did in the previous section. Also the Barabási-Albert topology with $m = 2$ is the same as in the previous section.

In Figure 7.4a, the Barabási-Albert network is considered for three different cache sizes, and a content population of 1000 (Wikipedia traces). The cache sizes are 0.1%, 1% and 10%. Increasing the cache size has a positive influence on the information resilience.

In Figure 7.4b, the Waxman network is considered for three different cache sizes, and also a content population of 1000 (Wikipedia traces). The cache sizes are 0.1%, 1% and 10%. We hardly see any difference between the two networks. We see for large cache sizes that the information resilience of the Waxman topology remains a bit higher, but the differences remain very small.



(a) Information resilience of Barabási-Albert network for different cache sizes



(b) Information resilience of Waxman network for different cache sizes

Figure 7.4: Information resilience for different cache sizes and different topologies

In this Chapter, we have seen that we can simulate the information resilience for other topologies, such as the Barabási-Albert topology and the Waxman topology. For that, we have, due to practical reasons, returned to the largest component way of simulating rather than the shortest paths. It would be very interesting to make an attempt to analytically compute the information resilience for these kind of networks as well.

8

Conclusions

In this section we answer the research question of the mathematical part of this research to conclude this research. In Chapter 4, 5 and 6, we have focussed on analytically computing the information resilience of the networks. In Chapter 7, we focussed on more complex networks, of which we need to simulate results as analytical approaches fall short. The research question to be answered here is:

To what extent is it possible to analytically compute the availability of the content in a network when breakdowns occur?

The first step towards answering this question has been to define the term information resiliency, and use that as a measure of the availability of content. Information resiliency has been defined as

The percentage of information requests that still can be delivered to the end users in the network, given a disturbance in the network.

Armed with a measure of the availability of information, some first intuitive situations have been investigated. In case that information has a very specific popularity, or in case that the network has a very specific shape or in case that the caches have a specific size, the information resiliency can be computed in a straightforward manner. However, to move beyond these very specific situations, more complex computation is needed. For hierarchical structures (see Section 3.3 for a definition), we have validated our approach using simulations and various topologies and caching strategies. We have shown that these computations are rather accurate. As we have seen, the information resilience in hierarchical networks depends on the probability of a layer to be reachable from the requester, as well as the probability that the content is hit in that cache.

The probability that a path to a certain layer exists can be computed rather easily in hierarchical structures. Let f be the fraction of links that is broken. Given one unique path of length m between 2 nodes, the probability that a path exists between these two nodes is approximated by

$$P(\text{path}) \approx (1-f)^m. \tag{8.1}$$

If there are two (independent) paths between two nodes, both with length m, then the probability that a path exists between two nodes, given that f is again the fraction of broken links, is approximated by

$$P(\text{path}) \approx 1-(1-(1-f)^m)^2\,. \tag{8.2}$$

In case that parts of the paths are unique, Equation 8.1 and 8.2 can be combined.

The probability that the content is hit in a cache can be approximated by Che's approximation. In Chapter 5 we have given a lot of attention to this set of equations. The set of equations can be solved with a fixed point iteration.

The cache hit ratio per layer can be combined with the probability that a path exists to obtain the information resilience of a hierarchical structure:

$$P(\text{succesful request}) = \sum_{k=1}^{n} (P(\text{path to layer}k) * P_{HIT}(k)) \tag{8.3}$$

The difficulty with non-hierarchical structures is that it is much harder to compute the probability of an existing path, as the path length differs for all end users. This indicates that non-hierarchical structures would be much harder to compute. However, there is no reason yet to assume that the computation would not be possible for non-hierarchical structures. However, applying the analytical results to non-hierarchical structures would be more complex, and the number of equations that should be identified increase rapidly. Furthermore, the cache hit ratio is much harder to compute for non-hierarchical (or highly non-symmetric) structures, as the equations to be solved are also dependent on the number of paths reaching to the node.

For non-hierarchical structures, we have investigated the information resilience by simulation. The two structures under investigation were the Barabási-Albert topology and the Waxman topology. With simulations, the information resilience there can also be approximated. In the discussion section (Section 16), we will elaborate on analytical approximations with these topologies.

In this thesis, we have found analytical results for two caching strategies, namely 'Leave Copy Down' (LCD) and 'Leave Copy Everywhere' (LCE). It should also be possible to incorporate a strategy where probabilities play a role. When the caching strategy becomes very complex, however, the equations that need to be solved to obtain the cache hit ratio become also more (and possibly too) complex. Note here, that if we are not able to approximate the cache hit ratio, it is also not possible to compute the information resilience for that topology.

We have also analysed the trend of information availability as a function of broken links. Our analysis shows that the relative increase in the information availability due to caching grows with the faction of broken links. We have quantitatively shown the increase in resilience for different parameters. To give an example of that, for our hierarchical realistic network (Figure 6.2), we have shown that with a cache of 5% of the content population, a population build from Wikipedia traces (size 1000), and a LCD strategy (Figure 6.8), the percentage of successful requests increases from 43% to 55% when 20% of the links is broken.

To summarise, we are able to analytically compute the availability of the content in a network when breakdowns occur for hierarchical network structures that have either LCE or LCD as caching strategy. For all other network configurations, the availability can be approximated by simulations. As we have not concluded that it is impossible to compute the availability for other network configurations, further research might show ways to expand the extent to which the availability of content can be computed analytically.

9
Discussion

In this Chapter, we will discuss some cases that stood out during the research. First of all, we will reflect on some issues in the research itself. Which issues remain unsolved, and do we have ideas of how to solve them? Secondly, we will turn our attention to items that are still open for investigation. One of these things is that the analytical results could be applied to non-hierarchical structures as well, but we do not fully oversee its consequences yet. What would happen if we analytically compute the information resilience for a Barabási-Albert topology?

9.1. Open Ends in the Research

The analytical results rise some questions with regard to the scalability. When computing the cache hit ratio's for the LCD strategy, one to three equations need to be solved per layer of the network, per content item in the population. Hence, when we consider a content population of 1000 items in the realistic topology, already 15.000 equations need to be solved. As the equations need to be solved with an iterative solver (the equations all depend on each other) the computation time becomes an issue for large content populations.

In Table 6.3 we have seen computation times for varying cache sizes. The computation times vary a lot, and there does not seem to be any relation between the cache size and the computation time. We think this is due to the fixed-point iteration computation, that we have considered to be a black box in this thesis. For future research, we would strongly advise to do some more research into a suitable fixed-point iteration method.

There are multiple ways to help the fixed-point iterative solver towards a solution. The first way is to come up with a precise start vector, as we would expect it to converge quicker when the start vector is more accurate. This is not always the case, but it often does help. Secondly, the solving method might increase the speed of solving. The solving matrix is non-linear, but the matrix of equations to be solved is sparse. Hence, some non-linear method should be used. In this research, not much attention was put into the fixed-point iterative solver. The Krylov method chosen worked well in practise, but has not been tested on a proof of convergence. In fact we are not even certain that the iterative method does converge. In practice, however, this has not shown to be a problem.

If the size of the content population increases, the method becomes slower, and for population sizes larger than 1000, the method becomes too slow to be practical. This could possibly be solved by using a more appropriate fixed-point iterative method as well as using a more precise start vector. It is known for fixed point iterations that some set of equations might have a convergence region. Outside of this region, the fixed point iterative method no longer converges.

9.1.1. Assumptions

In Section 4.1.2, we have considered the network topology of Figure 9.1. We have made two assumptions on the network that do not seem to have any influence on the outcome. One reason for that could be that they are non-restrictive. The other reason could be that they cancel each other out. First of all, we had assumed that the caches were filled independently, which is not the case. Secondly, it was assumed that the search for an item could also be performed off the shortest path.

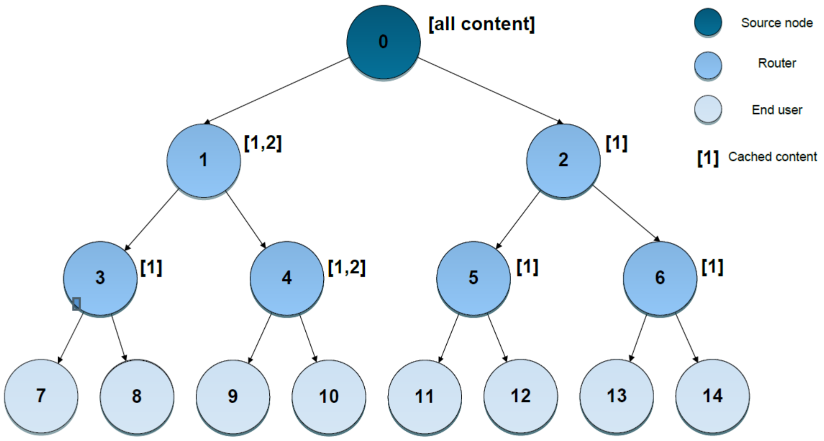


Figure 9.1: Tree topology with caches filled with high skewness

A hint that the assumptions indeed cancel each other out is that when looking at one simulation run, we saw in the nodes $1, 3, 4$ together, 21 unique items with a population size of 40, and a cache size $C = 10$. We run a small simulation to compute this probability. Let us pick 30 (total cache size in the 3 nodes) random integers from 40 numbers, with replacement. In the simulation, we see that on average, we pick indeed 21 unique numbers.

However, when taking only node 1 and 3 into consideration, we see 17 unique items. When running the simulation again, we would expect 15 unique numbers instead. This suggests that indeed the caches are not filled independently from each other, but the two assumptions made above cancel each other out.

9.2. Future Research

9.2.1. Time and Place Dependent Popularity Function

Some simplifications have been made during the course of this research. These simplifications are common in ICN simulations, but it would be interesting to investigate the impact of these simplifications. First of all, we have considered the content to be equally popular over the entire network, over the entire duration of the simulation. In reality, this is not necessarily the case. If these simplifications are removed, what would be the impact? Besides incorporating these items in the simulation, it might also be possible to incorporate these items in the analytical approximations. It should be possible to incorporate these item in a function for the content popularity that is both place and time dependent. If that is established, this function can also be used for solving the equations that determine the cache hit ratio.

9.2.2. Analytically Determining the Information Resilience for Non-Hierarchical Structures

What complicates random structures over hierarchical structures is the number of different equations that need to be solved. First of all, recall that the Cache hit ratio in a node is equal to the sum over de cache hit ratios per population item in that node. Hence,

$$P_{HIT}(k) = \sum_{d \in D} (\lambda_k(d) * P_{hit}(k, d)) . \tag{9.1}$$

In a hierarchical network that happens to also be symmetric, the cache hit ratio of all nodes in a layer can be computed at once. With symmetric in this instance, we mean that, per layer, each node is connected to the same number of parents and children. In that case, we can consider a whole layer at once, because each cache in each layer is visited equally often, resulting in the same approximated cache hit ratio per node.

Now, if the structure is not symmetric, some adjustments need to be made within a layer to be able to compute the cache hit ratio. In practice however, we see that hierarchical structures have a tree structure in the parts of the network closer to the end user (as we have seen in Chapter 6), which simplifies the adjustments that need to be made again. In Figure 6.2, we see that the topology there is not entirely symmetric either, as we have only considered one tree structure instead of all five. Hence, the nodes 10, 11, 12, 13 and 14 do not all have the same amount of children. The equations of Garetto et al. (2016) provide us with a suitable tool to deal with that.

However, let us consider a Barabási-Albert topology with $n = 30$ nodes, and $m = 2$, the minimal degree of each node (see Section3.2 for a specific explanation of the parameter m). We consider node 0 to be the only source, and nodes 20-29 the end users. See Figure 9.2 for the exact network considered.

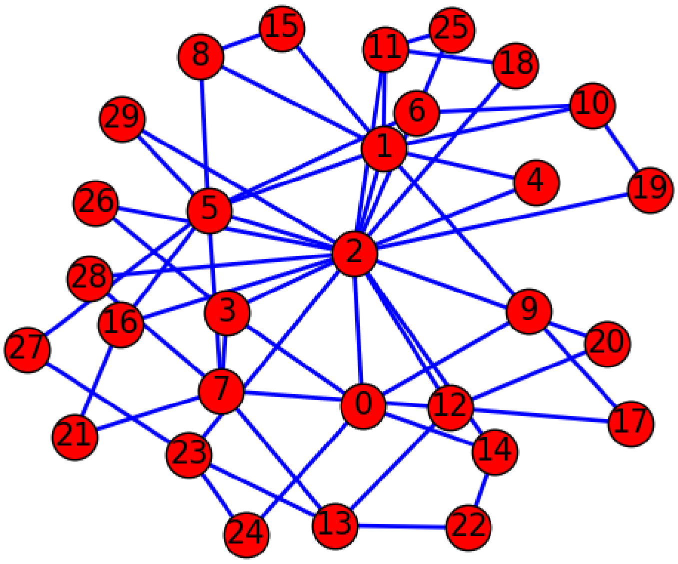


Figure 9.2: The considered Barabási network

Now, the only way to compute the cache hit ratios of these nodes is to consider the shortest paths from the end user to the source individually. When multiple shortest paths occur, we need to consider them both. For this example, with ten end users, it is doable. But for larger networks this is highly complex.

What complicates the matter further is that a node can be on different positions in the shortest path for different end users. To give an example of that, the shortest path for node 25 is 25-6-2-0. However, for node 28, this is 28-2-0. Hence, node two has a different position in these sequences, resulting in a difference in both the caching behaviour as well as the resilience computation.

For the latter, we have satisfied in computing the probability that a path exists to that node for now, but this would have to be separated into two statements for the network of Figure 9.2, one for each different path length.

As we see, the computation is complicated when the structure becomes highly asymmetric and non-hierarchical. However, there is no reason to expect the computation to be impossible. It would be interesting to see if we can extend the analytical results to be suitable for more general topologies too.

9.2.3. Rerouting

It would be interesting to consider the impact of link failures on routing, and incorporate rerouting schemes in the analysis. With this, we mean that for most of the analysis in this thesis, we have only considered the set of shortest paths from the end users to the sources as the set of available nodes to retrieve content from. However, in case of a disturbance, rerouting schemes could also be considered. Ultimately, all nodes that are still connected to an end user can be used to retrieve information from. But also intermediate concepts such as only allowing rerouting over the same layer could be considered.

9.2.4. Forecasting 'Good' Networks

Ultimately, it would be nice to forecast which ICN network topologies and strategies would be most effective for certain purposes of an ICN. Not only the simulation tools such as Icarus are very effective for that; also the analytical results can enhance the insights in what parameter changes in a topology work well.

What we have seen is that the impact of the first cache is very large. Also the second layer of caches still adds to the information resiliency in the network. However, caches higher up in the hierarchy are hardly ever hit, and therefore not adding much value to the ICN. If caching is of importance, there is little added value in having a lot of caches that are positioned consecutively in the hierarchy (Figure 9.3). Having these caches in a parallel structure is in fact more effective for the information resilience in the network, as they provide the end user with alternative paths, resulting in a higher probability to still reach the source in case there is no cache hit.

For an example of that, see Figure 9.3. So when investing in caches (suppose a choice needs to be made), we would advice for as many parallel caches as possible.

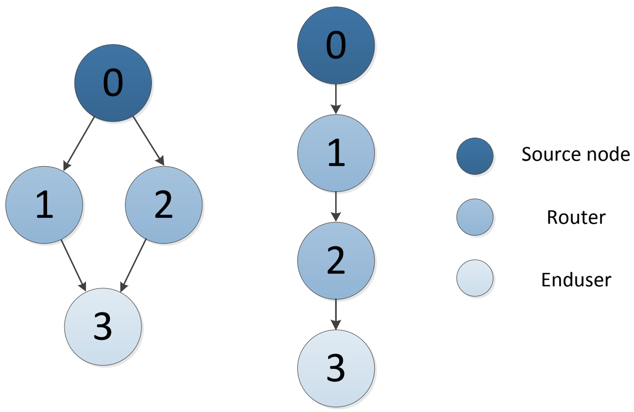


Figure 9.3: The parallel routers (left) versus the path graph (right).

Furthermore, we have seen the cache hit ratio to be a very important factor in the performance of the ICN. This cache hit ratio depends also largely on the cache replacement policy. Even though least recently used is intuitively spoken a good replacement strategy, we have not investigated other replacement strategies. The equations of the cache hit ratio rely largely on the reasoning of when an item will enter a cache. This depends to a large extent on the presence of the content in other caches. Hence, when to throw out an item from the cache again is also important.

At this moment, the analytical results are mathematically interesting, but relate to rather abstract graph structures, which makes it a bit hard to predict a network that behaves best under certain conditions. However, it would be interesting to research the possibilities to extend it in such a way that it is possible to do that.

III

The Analogy Between ICNs and Human Collaborative Networks

10

Introduction

Power, today, comes from sharing information, not withholding it

Keith Ferrazzi

The third part of this research uses the mathematical concepts and experience we have gathered in Part II. By now, the reader should be acquainted with Information Centric Networking, its key aspects and key functions. Most of all, the reader should be familiar with the in-network caching capacity of ICNs. For a recap of this, the reader is referred to Chapter 2 of this thesis. As was introduced in Chapter 1, these concepts will be used to build an analogy between ICNs and human collaborative networks. Let us use this place to make this analogy intuitive.

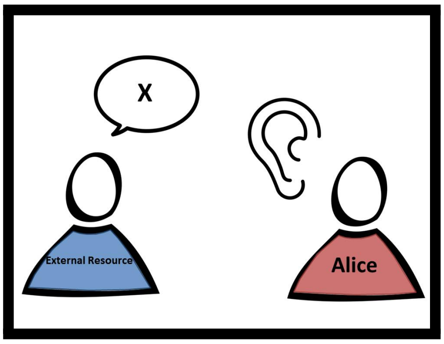


Figure 10.1: (A)

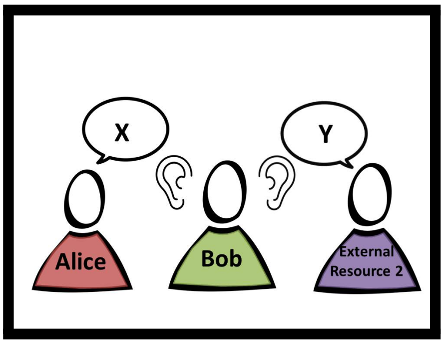


Figure 10.2: (B)

10.1. Research Context

10.1.1. Intuitive Analogy

In this section, we will pay some attention to the intuition behind the analogy, as well as the notion of resilience in a human collaborative network.

In the following figures, an intuitive explanation of the analogy is given. Here, the in-network caching capacity is key.

We consider the information spread of Alice and Bob. They work together in a group with Charlie and Dylan. Alice and Bob are good friends, and they talk together a lot, also outside official meetings. In particular they are in the same sports class.

For their project, the group needs novel information. It is therefore that Alice decides to talk to an external resource. She obtains new information, which we call information X (Figure 10.1). Here, the information is treated as one piece, that is not interpreted by Alice at all. She just absorbs the information, and uses it in the exact same form later. This is a simplification of reality, but for the sake of comparison it works well. Relating this to an ICN, the external resource acts as a source of information.

After meeting with the external resource, Alice meets Bob at their sports class, and tells him about information X. In ICN terminology, Alice becomes a router of information at this point (Figure 10.2).

Later that day, Bob calls with another external resource, and obtains information Y (Figure 10.2).

The next day, Bob meets with Charlie and Dylan, and tells them about the new information he has obtained since the last time they met. By telling them, he is also behaving like a router at that particular moment. Charlie and Dylan on the other hand, take the role of end users (Figure 10.3).

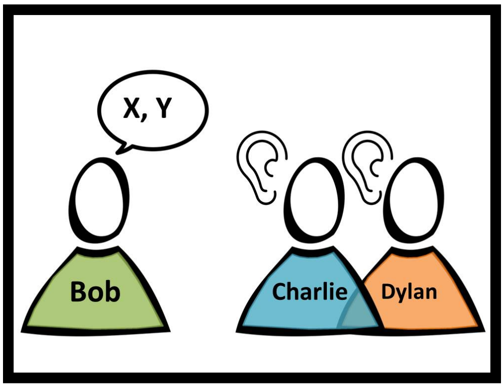


Figure 10.3: (C)

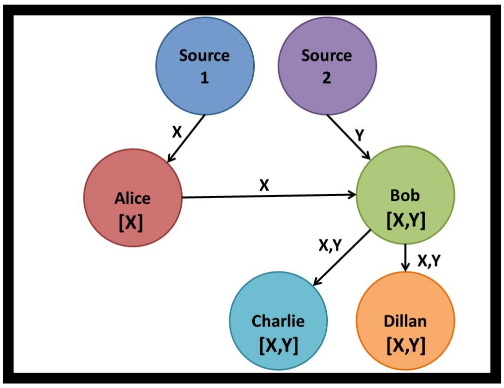


Figure 10.4: (D)

After this information spread, Alice is only familiar with information X. The other three, however, have knowledge of information X and Y. On the right side of this page, a graphic representation of this information spread can be seen (Figure 10.4). The first three pictures depict the sequence of events. Figure 10.4 depicts the schematic, more ICN-like approach of the information sharing.

This is the very basic approach of the analogy between human collaborative networks and ICNs. However, there are many simplifications made here. As already mentioned, humans will always interpret information, add their knowledge to it, and retell the information with some personal touch.

10.1.2. Resilience

In the mathematical part of the thesis, we were mostly interested in the impact of a disturbance in the network on the spread of information. We considered random link failure in the mathematical part, but we will consider specific node failure here. In an ICN, the random link failure is equivalent to a connection between two nodes failing at random. This can be because of many reasons. However, as we are considering random failures, we specifically do not consider an attack. In this part of the thesis, we will consider node failure, as it is easier established when dealing with human beings. The impact will be larger, and therefore better measurable. Furhtermore, node failure is equivalent to some people not being part of the network any longer. This happens often, due to many reasons such as illness or career changes. Furthermore, the failures are no longer random, but specific. This is because we want to connect the theoretical factors in human information sharing to the actual behaviour when a network gets disturbed. To be able to do that, it is preferable to select the node failures according to one of these factors that theoretically influence information sharing in human collaborative networks.

In dealing with these disturbances, the human network topology is key. This is one of the most influential factors on the resilience of information spread. The human network topology is another (besides interpretation of information) factor that limits the analogy. Where the topology in an ICN does not change itself once the network is disturbed, humans might look for other contacts when their usual ways of communication are not available. This will complicate the analogy in the case of a disturbance.

10.2. Research Questions

The main research question was already formulated in Section 1. We will repeat it here.

To what extent can the information spread in a human collaborative network and an ICN be applied to each other, to enhance insights into information spread and information resilience?

The research question is extended with some sub-questions that will help in building this analogy. These are listed next.

1. What is the essence of the analogy between ICNs and human collaborative networks?
2. What factors influence the way humans share information in a collaborative network?
3. What is the influence of disturbing the network on the information spread in human collaborative networks?
4. How does the information spread in a disturbed human collaborative network relate to information resilience in ICNs?

Let us review these questions. First of all, there will be some natural boundaries of the analogy. Also, we need to make clear how the different aspects of human collaborative networks can be translated to the ICN. In other words, what are the subjects of comparison in the analogy? By answering these kind of questions, we build ourselves a platform from where we start thinking and investigating about the analogy. This platform is what we call the essence of the analogy.

In the second research question, we set the analogy aside for a while, and wonder what factors influence information sharing in human collaborative networks. The answers will be found in literature. We will pick some of these factors for our theoretical framework. The items will be picked with the analogy in mind.

In the third question, the effects of a disturbance are being questioned. What happens when we disturb the human collaborative network, and how does this affect the information sharing? This closely relates to the fourth question, where the information spread in a disturbed network is compared to information resilience in ICNs.

The first two sub questions aim to find a rudimentary level of the analogy, whereas the third and fourth question are meant to investigate the possibilities of adding another layer of complexity to the analogy. If positive results about the extra layer are obtained, it could imply that other layers could also be added in new research.

10.3. Research Scope

Information sharing in human networks can be compared to information sharing in an ICN. However, there are natural differences. More attention to these differences will be given in Chapter 12 where we discuss the nature of the analogy. The theoretical lens will be developed in Chapter 11.

In this research, we consider the analogy at a high level. By that we mean that many details are not considered. This already starts with the literature research, where we find many factors that influence the information sharing in collaborative networks, but where we only take a couple of them into consideration. The same holds for the collaborative networks considered. The networks are considered in a closed setting, with a specific task. The network is relatively small in size (7-8 people). This all is done to enable ourselves to focus on the high level of the analogy instead of getting tangled in the many details.

As explained in Section 1.4, the problem in approached from a cybernetic point of view, where the system as a whole is the most important thing considered. Hence, individual considerations play less of a role. Due to theoretical limitations, the individual considerations to share information is reviewed theoretically in Chapter 11. But from there we take a systems approach towards the experiment and the analogy.

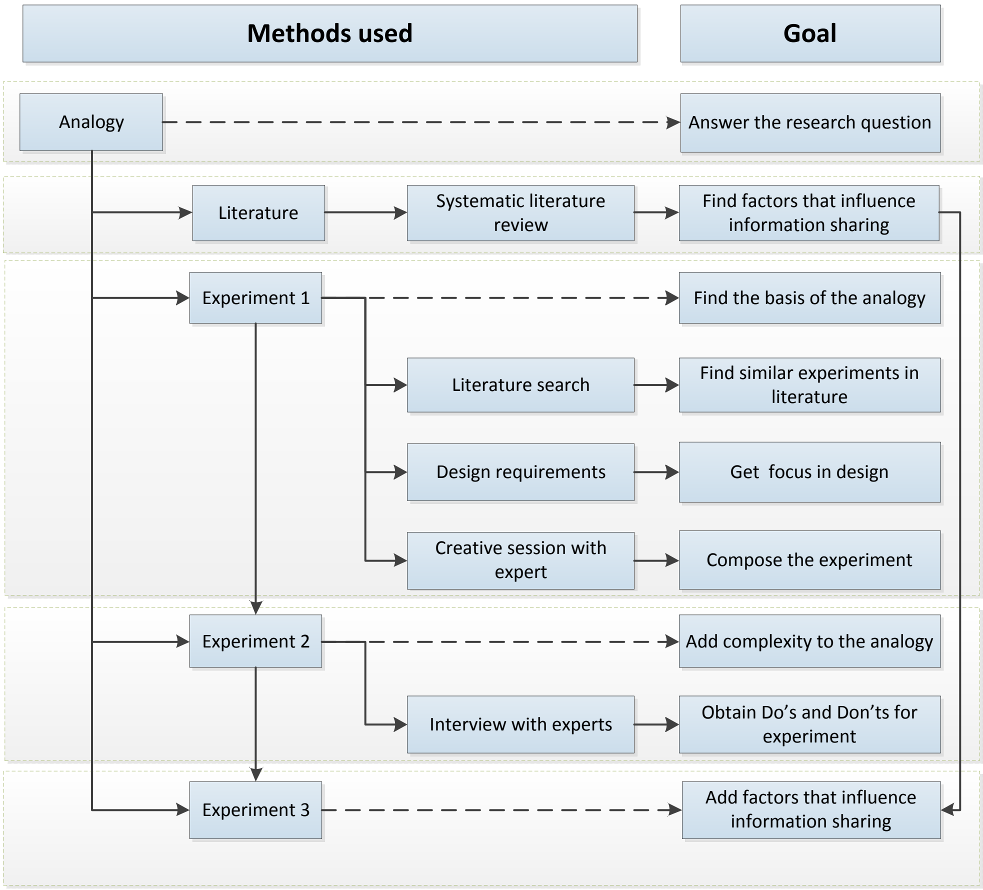


Figure 10.5: An overview of the methods and their goal

10.4. Research Methodology

In this research, several methods have been applied. In this section, we will not describe the methods in great detail, but this section will provide the reader with overview of the methods. In Figure 10.5, an overview of the methodology in the different stages of the research can be found.

To answer our main research question (To what extent can the information spread in a human collaborative network and an ICN be applied to each other, to enhance insights into information spread and information resilience?), we build up an analogy in which we compare human collaborative networks with information centric networks. We build up the analogy through 4 different methods.

First of all, in Chapter 11, a systematic literature review is performed with the goal to determine the most important factors that influence the information sharing within human collaborative networks. Performing a systematic literature review allows us to systematically limit our search.

Furthermore, to add detail to the analogy, three experiments will be performed. The experiments are consecutive in the sense that the results and feedback from participants of the first experiment are being used for the second experiment, and the second experiment will be used for the third experiment. We perform experiments to find the borders and the possibilities of the analogy, and to add detail to the analogy.

To design the experiments, several methods have been used. First of all, we have studied literature to find out that experiments as such are not yet available in literature. As a result, we have to design the experiment ourselves. Secondly, design requirements have been composed. Thirdly, with these design requirements, a creative session with an expert specialised in gamification for education purposes, the first experiment was designed to measure the spread of information through a human collaborative network.

In the second experiment, we wanted to add some complexity, in order to let human behaviour play a larger role in the experiments. To design the second experiment, we have interviewed 3 experts from the university of Amsterdam, Groningen and the Erasmus University.

An overview for which method is used for what purpose and in what stage of the research van be found in Figure 10.5.

10.5. Outline of the Experiments

Once it is set in stone what the objects of comparison are, and with what purpose the comparison is to be found, one can start building the analogy. This needs to be done step by step, to keep the analogy as clean as possible. Therefore, 3 experiments have been performed, in which different aspects and complexity levels of the analogy were emphasized and researched. The first experiment is very clean, and is meant to build up the analogy for the first time. However, the way human communication is modelled there is rather artificial. In the experiments 2 and 3 the human aspect is brought back into the experiment, which is of added value, but has some drawbacks as well. These advantages and drawbacks will be discussed later on.

In the experiments, we want to measure the spread of information in a human collaborative network. In an ICN simulator, the information spread can be traced. If we are able to do the same in human collaborative networks, the two can be compared. Furthermore, if the more human aspects of communication we can add to the analogy, the more valueable the analogy becomes.

That brings us to a brief description of the 3 experiments before diving into it with more detail. First of all, an experiment was set up in the minor class to investigate if human collaborative networks and ICNs are comparable at all. The students had spread information in a traceable way. This spread of information was then collected and compared to the spread of information in an ICN.

In the second experiment, more human aspects of human communication were added. In adding these elements, the theoretical framework as discussed before was used. This experiment has been performed with a group op 8 students of Delft University of Technology. It has been used as a trial run for the final experiment. Furthermore, the information resilience was researched in this experiment as well. To do that, some people were excluded from the experiment the second time. Which people were excluded was based on one of the aspects of the theoretical network, namely the extent to which the person was liked by the rest of the group. Based on an evaluation form and personal observations, adjustments were made towards the final experiment.

In the third experiment, a group of students that work together in two teams were participating. In this experiment, the step towards a real collaborative human network has been made. The three experiments together give great insights in the possibilities of the analogy, but some boundaries were found as well. By using the theoretical framework, some insights regarding the spread of information in a human collaborative network were obtained. These insights are part of the conclusion, but give great input to the discussion section as well.

The first experiment builds the essence of the analogy. However, this experiment lacks a lot of human detail and behaviour. This detail and behaviour is being added in the second experiment. The second experiment is merely a trial run for the third experiment, where the participants are selected from an actual collaborative network.

11 Literature

A point of view can be a dangerous luxury when substituted for insight and understanding.

Marshall McLuhan, Canadian Communications Professor

In this chapter, we construct a theoretical framework. But before that, we pay some attention to two of the concepts from the main research question: Information sharing and information resilience. In Section 11.1, we dive deeper into the philosophical concept of information sharing. How can we distinguish information sharing and knowledge sharing? These rather difficult and questionable boundaries will be discussed.

In Section 11.2, we define the concept of information resilience in human collaborative networks. In Sections 11.3 to 11.6, a systematic literature review is being performed with the aim of developing a theoretical framework in Section 11.6. This theoretical framework will include the most important factors that influence the way information is spread through a human collaborative network. Factors one should think about are individual motivation to spread, or the way the network connections are spread.

We specifically aim for finding factors from literature that help us in building up the analogy. As a results, these factors should be measurable, and preferably not too complex. This will be part of the motivation for some factors in the following sections.

11.1. Information Sharing

In this part, the notion of information sharing will be discussed. In doing so, the article by Pilerot (2012) will be leading, as he has published an article that discussed issues regarding information sharing in great detail. Furthermore, the choices made for our research will be elaborated on.

We aim to keep our used definition of information sharing in human collaborative networks as close to the definition of information spread in ICN, as we will use it for our analogy. To this, there are a number of choices that we need to make with respect to this definition. First of all, sharing can be interpreted as something that is done between groups of people, but it can also be done between individuals. As Pilerot (2012) shows, examples of both practices can be found in literature. As we have interest in mapping the streams of information through the network, we consider information sharing as a sharing process between two individuals rather than groups. What complicates the situation, is that in real life information is sometimes shared with multiple people at once (in a meeting for example). As a result of our choice, these options are not automatically being taken into account. If we wish to incorporate them in a specific setting, we need to consider it as one person sharing with multiple individuals at once.

The second issue regarding sharing that we take into account here, is the difference between sharing and exchanging. As shown by Pilerot, different meanings are assigned in literature to the word sharing. On the one hand, sharing can be seen as a two-way interaction, where the two persons involved both bring in a part of the shared information (exchanging). On the other hand, sharing can be seen as a one-directional transfer of information, where one entity (sender) transfers its information to another entity (receiver). Even more detailed distinct meanings of sharing can be found. Examples of this are information spread, flow, transfer and diffusion. In essence, these terms are more specific and detailed, and are all overarched by the term information sharing.
In our research, the focus lies on information sharing seen as a one way transfer of information. The spread of information is seen by us as the collective result of all information shared. So what information is where after a certain amount of time, and how did it get there? This lies very close to the notion of information diffusion, although this notion is more often used in combination with computer sciences. Note that this is our personal interpretation, as literature is not conclusive about this.

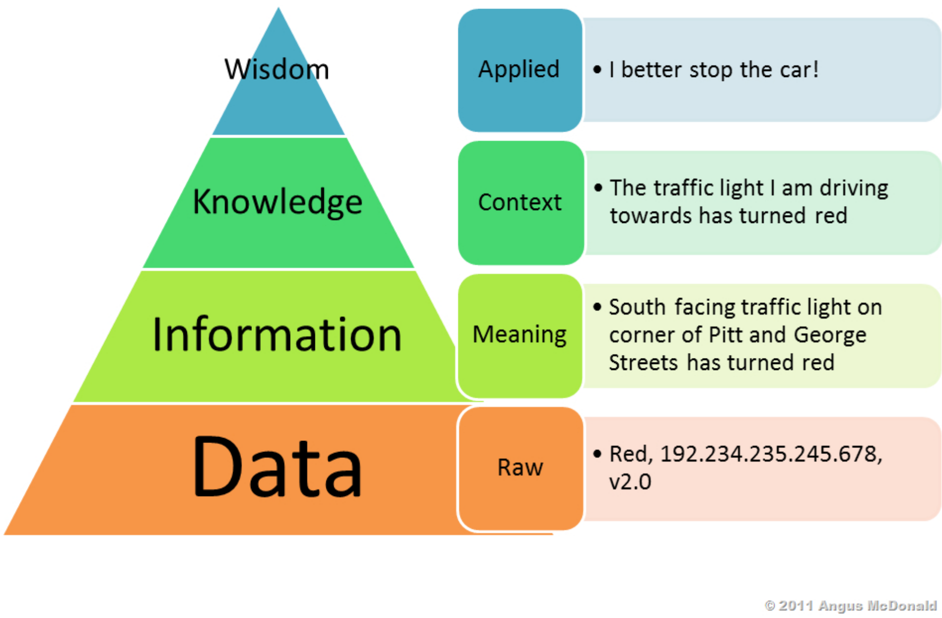


Figure 11.1: Knowledge versus information

Now that the word sharing is discussed, let us move on to the word information. Also here, there are two main considerations. First of all, there is the difference between knowledge and information. Secondly, information can, arguably, be seen as a 'materialistic stance' (Pilerot, 2012), which is what we prefer to do. We will come back to that later.

In literature, the difference between knowledge sharing and information sharing is not always clear. However, here we will make a clear distinction between the two. We follow the work of Davenport and Prusak (1998), of which a visualisation can be found in Figure 11.1. Information, according to Davenport and Prusak, is a message. It is materialistic in nature (a document, or an audio message), and it has some sort of shape in itself. The message has meaning to the receiver. Knowledge however, is broader in essence. Davenport and Prusak give the following working definition: "Knowledge is a fluid mix of framed experience, values, contextual information, and expert insight that provides a framework for evaluating and incorporating new experiences and information. It originates and is applied in the minds of knowers. In organizations, it often becomes embedded not only in documents or repositories but also in organizational routines, processes, practices, and norms."

This is not the place to discuss this definition in great detail, but emphasis should be put on the fact that knowledge embraces much more than information. Interpretation, context and values are just examples of things that are part of knowledge. As it is not the aim of the research to capture all these features, we restrict ourselves to information.

What remains is an answer to what information is. A book full of physics constants is information, but one of these specific constants is information too. Hence, information can be very large in nature, but also very small. Furthermore, information can be intentional (a book), or unintentional (reading somebody's body language provides possibly unintended information). What we pick here, is Buckland's (1991) 'information as a thing' meaning of information. By that, we mean that we see information as things that are informative. This could be a note, an email, or (and mostly in this research) an oral information transfer. By choosing this interpretation of information, we enable ourselves to talk about a piece of information, no matter how large this piece is. The reason we choose this paradigm is that by materialising the matter, we can compare pieces of human information with pieces of information sent over the internet. Also there, we assume the size of the piece (or package as it is called in ICNs) to be irrelevant, even though that is questionable in real life.

11.2. Definition of Information Resilience

Resilience is a concept that is used in many fields. In all these fields, the meaning is similar, but differs slightly from field to field. Oxford dictionary states that resilience is "The capacity to recover quickly from difficulties; toughness." These difficulties differ from field to field. Resilience has its roots in physics, where resilience is a property of a material to bounce back in its original shape when deformed. Furthermore, in psycology, Windle (2002) define resilience as the "successful adaptation to life tasks in the face of social disadvantage or highly adverse conditions." Hence, the difficulties considered here apply on the individual rather than the group or the situation. In sociology, on the other hand, Home and Orr (1997) state that resilience is a fundamental quality of not only individuals, but also groups and organisations.

In general, resilience in all fields has something to do with the ability to bounce back from a disruption. However, for our research, we have specific interest in the information resilience of human collaborative networks. This information resilience is not used in literature in a similar meaning as it is used in ICNs. However, since we do want to make a comparison between the two by investigating the analogy, we have a need of a clear definition of information resilience in a human collaborative network.

In the remainder of this section, we will investigate some other sorts of resilience in collaborative networks (team resilience and organisational resilience), and use aspects of these definitions to come up with our own definition of information resilience in human collaborative networks.

Definition of Team Resilience

The term information resiliency in human collaborative networks is not represented in current literature. It is therefore that we need to make use of concepts related to collaborative networks and resilience to construct our own defintion. One of these concepts is *team resilience*.

As explained by King, Newman, and Luthans (2015), individual resilience is defined by different authors in different ways. We embrace the idea that individual resilience is the ability of an individual to bounce back from adversity (Smith, Tooley, Christopher, & Kay, 2010). Team resilience however, is different from the sum of individual resilience in a team. According to Shankar Sankaran, Rodríguez-Sánchez, and Vera Perea (2015), team resiliency is "[...] their ability to adapt and emerge strengthened from difficult situations".

These aspects of team resilience encompass part of what we mean with information resiliency in human collaborative networks. However, two main differences need to be mentioned here. First of all, the difficult situations as mentioned in the aforementioned definition are external influences in practice. These could be as extreme as nuclear emergence situations (Gomes, Borges, Huber, & Carvalho, 2014), but also smaller like market shifts (Amaral, Fernandes, & Varajão, 2015). The type of disturbances we are looking for, have their nature inside the team itself. Examples of this are people leaving the company unexpectedly, or internal struggle. These internal disturbances complicate the collaboration within a team.

The second difference between information resilience and team resilience is that the definition on team resilience mentioned above has no emphasis on the information spread in a team. What we research is the way the information spread is influenced by the disturbance within the team. This is possibly a part of team resiliency, but it is not being emphasized in literature as far as we know.

Definition Organisational Resilience

Organisational resilience encompasses a much broader field than team resilience does, even though its definition is rather similar to the definition on team resilience. According to Seville et al. (2011), "[o]rganisational resilience is a function of the overall vulnerability, situation awareness and adaptive capacity of an organisation in a complex, dynamic and interdependent system". This is then elaborated on by saying "[a] resilient organisation is one that is still able to achieve its core objectives in the face of adversity". Also Shankar Sankaran et al. (2015) state that resiliency in organisations is "a concept that might aid organisations survive and thrive in difficult or volatile economic times".

In practice, this encompasses a lot, including the relationship of an organisation with other organisations, its information security, its leadership abilities and many more things. In our belief, this term is so broad, that it is of little added value to our definition of information resilience in teams.

Definition Information Resilience in Human Collaborative Networks

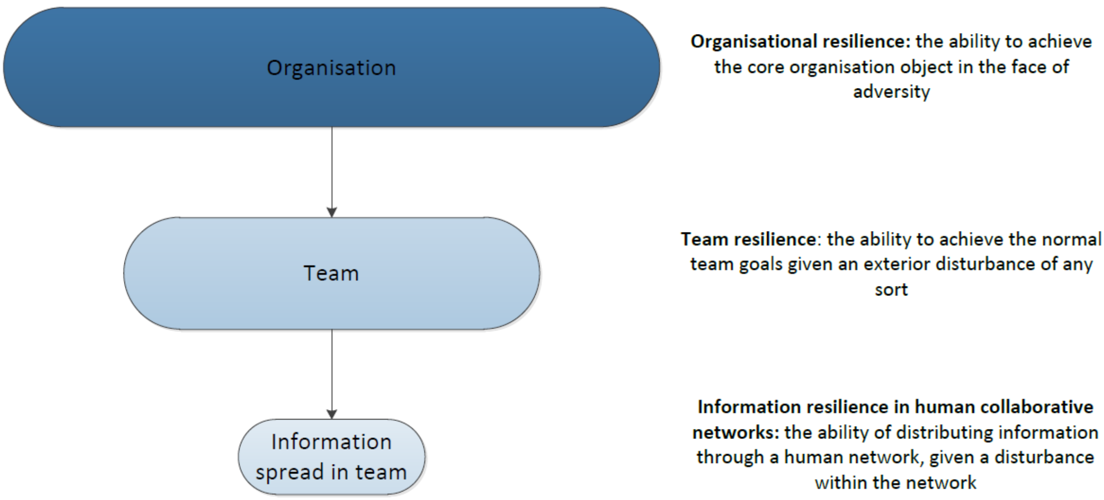


Figure 11.2: Information resilience definition

In line with our objective (setting up the analogy) as well as the definitions above, we shape our own definition of information resiliency in human collaborative networks. As said before, information resiliency differs from team resiliency in two main directions. First of all, the disturbance is exterior in team resilience, while it is a disturbance in the network itself in information resilience. Secondly, the focus on information distribution is missing in team resilience.

What is similar in all definitions concerning resilience is that there should be some kind of disturbance or adversity which needs to be dealt with. The ability to deal with this adversity is considered to be resilience. By whom this is to be dealt with is not explicitly specified by these definitions. If we use the information above, we can reformulate a definition for team resilience, which does not contradict the current definition, but puts a different emphasis on the matter. Our own definition of team resilience is:

"Team resilience is the ability to achieve the team goals given an exterior disturbance or adversity of any sort."

Now, we use this to formulate a working definition of information resilience in human collaborative networks. See also Figure 11.2.

"Information resilience in human collaborative networks is the ability of distributing information through a human network, given a disturbance within the network, such that achieving the team goals remains possible."

11.3. Systematic Literature Review

For the communication part of this research, a systematic literature review was performed. The reason for this is twofold. On the one hand, it is a way to limit the research field. By systematically looking into a set of keywords and concepts, all articles within these keywords are considered, and articles outside it will not be considered.

The second reason for a systematic literature review is to support the experiment with a theoretical framework, based on the literature review. This framework can assist in setting up the experiment, in selecting the participants for the experiments on information resilience, and in drawing the conclusions.

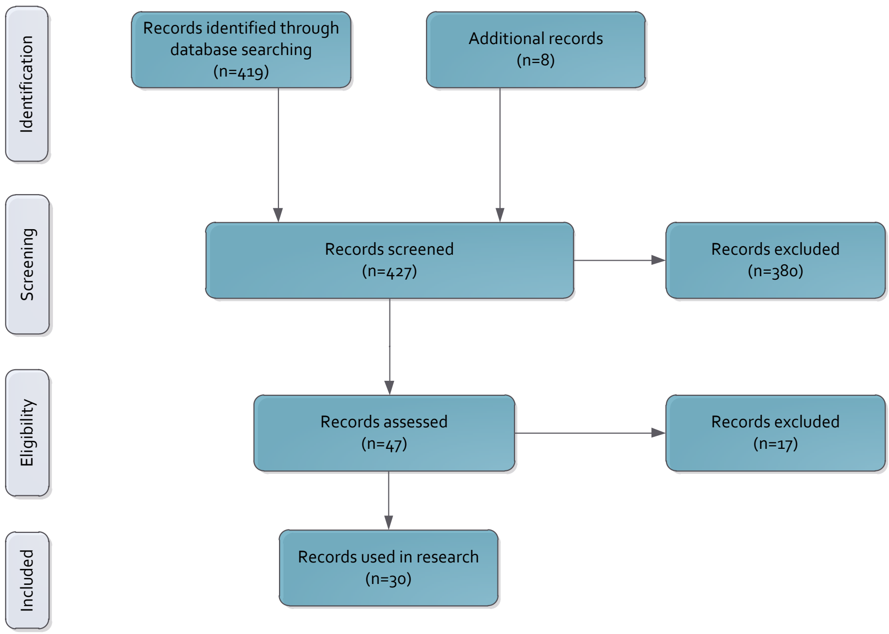


Figure 11.3: PRISMA flow diagram with the actual numbers of this research

Before moving to the key concepts of our initial framework, let us elaborate more on the method behind this literature review. The database Scopus is used to find articles with a set of keywords. This is the identification phase according to the method of Moher, Liberati, Tetzlaff, and Altman (2009). This method is called PRISMA, and a flow diagram of it can be found in Figure 11.3. In this phase, also some articles that have been found in the orientation phase will be added to the body of literature considered. If the title sounds promising, the records are screened in the second phase, by looking into the abstract. Records that are not considered after reading the abstract are excluded. In the eligibility phase, articles are read/ extensively looked into. Articles that are excluded from that moment onwards need a reason to be excluded.

Inclusion and Exclusion Criteria

The Appendix A contains a table with articles that are not included in the eligibility phase accompanied with a reason for exclusion. These exclusion criteria could be related to reliability reasons, but mostly are related to relevance and validity.

Goal of Theoretical Framework

Now, what do we need a theoretical framework for? Before we can move the analogy up to the point where the resilience is being taken into account, the basic analogy needs to be set up. In setting up this analogy, we expect there to be some differences between the human collaborative network and the ICN simulator. In an ICN, the way information spreads depends on the topology and the information requests only. In a human collaborative network however, the spread depends on much more than that. To give just one example, the personal motivation of one individual to share or not share information can have a large impact on the information spread through the whole network. As there are many more factors that influence the spread of information in a human collaborative network, we need a theoretical framework that incorporates these factors. Deliberate choices of what to take into account and what not will be made, and those factors taken into account will be added to the framework.

However, this is not the only thing we take into account. We want to study the analogy up to the point where a disturbance in the network occurs. Hence, the problem goes beyond the spread of information, and includes the impact of a disturbance. This impact, measured as information resilience, is influenced by many factors. In an ICN, resilience already depends on many things, but in a human collaborative network, the number of factors influencing the resilience is even higher. This is because human communication incorporates many details, and the slightest detail can greatly influence the communication within a human collaborative network. Hence, the resilience is key in our theoretical framework. However, it is closely intertwined with the information spread. This is because the two concepts are intertwined too. One could say that the resilience is about how well the information spreads, given the disturbance. It is therefore that the final theoretical framework will focus on resilience, but we cannot do that without considering the information spread in the network.

The theoretical framework will guide us which questions should be asked to the participants of the experiments. As we investigate which factors play an important role in information sharing and in resilience, we can ask the participants which factors apply to them. Some of these aspects could be asked straight away, and some should be asked/investigated more implicitly. These answers will guide us in choosing which people to remove from the experiment when testing the resilience. However, for that, not all influencing factors can be taken into account, but we will have to choose one. The other aspects can be used in observing the experiment though, and could lead to some insights in the discussion section of the research.

To summarize, the theoretical framework is needed to guide us in a choice of which people to take away from the experiment in the resilience phase, and it helps us to find possible explanations of behaviour that stands out. Furthermore, the theoretical framework will help us to build up the analogy in detail.

The literature research consists of several steps: first of all, an initial framework is being set up, based on the orientation phase of the research. From this framework, we derive keywords that are researched in a literature review as described above. Also, from the reference list of the articles found in the first round, some additional articles might be included. Finally, with all initial articles combined with the articles from the literature search, a definite framework will be built.

11.4. Initial Framework

In this section we start with an initial framework on the information spread in a human collaborative network. The framework is almost as broad as can be. Secondly, we add the aspects of resilience, and narrow ourselves down to those aspects that are of uppermost importance to the resilience. This will be the starting point of the systematic literature research.

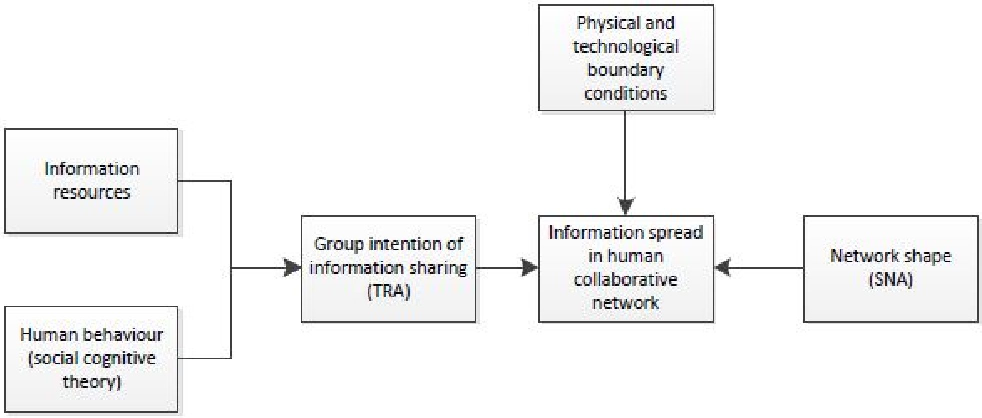


Figure 11.4: The theoretical framework, depicting what influences the information sharing in an organisation

In Figure 11.4, the initial theoretical framework on the information spread is depicted. The framework is inspired on the orientation phase of the research, where different articles about information sharing were read, different people have been consulted about their experience, and common sense is used as a source of inspiration too. Next, let us give a brief explanation of the aspects of this framework, as well as the arguments to include them.

Three aspects are taken as key to the spread of information through human collaboration. First of all, the network shape is very important. With this, we mean that the topology as well as the role and position of the different actors is important in the way information spreads through this human network. We are not going to dive into this matter with large detail here.

As we want to compare the ICN with the human collaborative network, we need some kind of network approach in the framework. This is because the information requests, together with the network topology, define the behaviour of the ICN. Hence, to compare the two, we need to take at least the network topology into account, but probably need to expand more on the network aspects of the cooperation. Network analysis is still a very broad concept. It includes the analysis of what roles are being performed by whom as well as the network topology. But also the notion of resilience from a network perspective could be seen as part of social network analysis (SNA).

This approach towards information sharing is used more in research. Pilerot (2012) mentions it as one of six theoretical approaches to information sharing. In (1996), Haythornthwaite has written an article in which the possibilities of (social) network analyses for the research in information spread (there called information exchange) are being stated. As this is just the initial framework, the exact details of what we include and exclude from Social Network Analysis are yet unclear.

The second aspect we take into account are the physical and technological boundary conditions. By this, we mean that the structure of the interaction between group members matters. If they run into each other on a daily basis on the work floor, the dimensions of their interaction will be different than when they can only meet through email and an occasional call. These boundary conditions of information sharing play a large role, but they have not been thoroughly researched. Fisher and Naumer (2006) have published a set of articles over time, in which they refer to these physical aspects as an information ground.

We take these physical aspects into account for now, as they seem to have a large impact, but can be adjusted rather easily. The less tangible aspects of information sharing such as motivation can less easily be influenced than the physical and technical aspects of the information sharing.

The third aspect taken into account is the group intention of information sharing. This is the intangible part of information sharing. For now, we consider the group intention to exist of two main parts: the information resources (what is the information that is being spread? is it interesting to share?) and the human behaviour. With the latter, we mean that people have all kind of motivations and incentives to share (or not to share) information.

This intangible part of information sharing is also part of the theory of reasoned action. This theory explains how attitude and behaviour coherent (Fishbein & Ajzen, 1977).

It is clear that a lot of concepts are bundled in these chosen concepts. However, we do this intentionally, as it is just the starting point. With the iterations made from this point onwards, more detail will be added to the framework. Also, choices will have to be made, as we cannot incorporate everything.

Although only found out in retrospect, the three key concepts we take into account here, largely overlap with the three main foci as distinguished by (Pilerot, 2012). In the article *'LIS research on information sharing activities – people, places, or information'*, the author concludes his 6 different theoretical views with stating that there are actually just 3 ways of looking at the matter.

1. People: what is the relation and the commonalities between the people who are sharing? What is their common ground to start sharing from?
2. Place: what does the location and atmosphere look like? How are the material conditions of influence to that?
3. Information: what is shared, and how does it 'flow'?

Information Resilience

Given this framework, where does resilience play a role? Before diving into literature, lets keep it intuitive. If a network gets disturbed (because people unexpectedly leave the team for example), what makes that the information still spreads through the network? First of all, the shape of the network plays an important role. If there are many ties between all members of the network, the information can still spread. However, if the number of ties is low, the missing people could have had key positions in the network that now prohibit the information to spread.

Secondly, the intention of the people in the group to share, still plays a big role. When the usual structures are disturbed due to the disturbance of the network, the intention to share might also be disturbed and changed.

The boundary conditions play less of a role in the resilience. Assuming they do not change due to the disturbance in the network, they do not influence the resilience in itself, but only through the way the information is spread.

Adding all this together, we obtain Figure (11.5), in which the information resilience is influenced by the group intention of information sharing, the network shape, and of course by the information spread. The arrow between information spread and information resilience is in two directions, as one could argue that the way information spreads through the network is influenced by the information resilience once a disturbance occurs. If the information resilience is very high, the information spread is probably occurring without problems, even though there is a disturbance.

Keywords

The framework in combination with the literature read in the orientation phase leads to some keywords. They are divided in this section into two type of keywords: primary keywords that are related immediately to one of the aspects in the framework. On the other hand, secondary keywords that are more generally related to the topic are chosen. For details, see Appendix A.

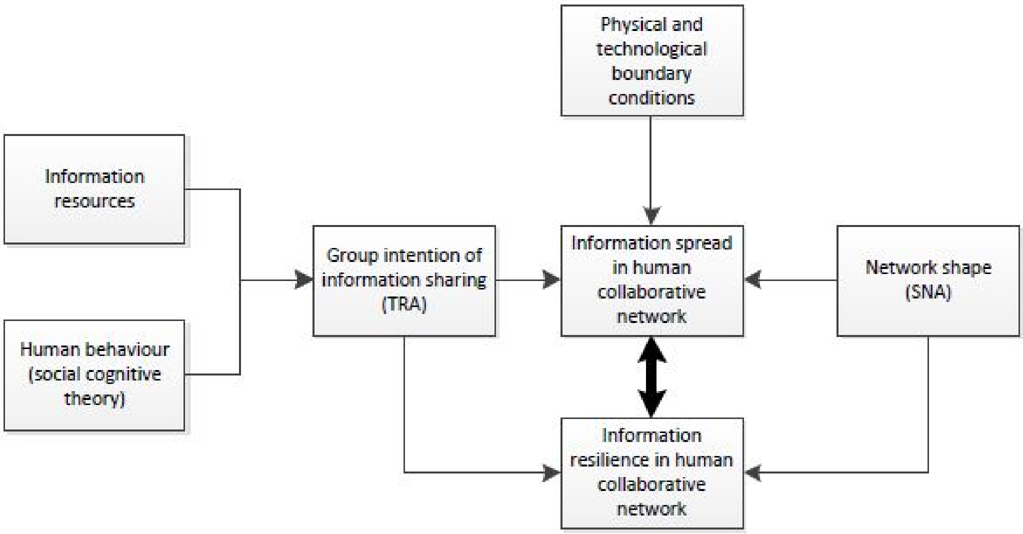


Figure 11.5: The theoretical framework, depicting what influences the information sharing and the resilience in a human collaborative network

11.5. Team Versus Group

For the purpose of our research we aim to keep the analogy as sober as possible. However, the use of the word network indicates that there is some kind of connection between the different members of the group of humans that is considered. In building the analogy, we aim for having a group of people with a connection together. That is the reason a distinction between a group and a team is being made. A team is a special kind of group. According to Cooke and Gorman (2006), "[i]n a team, individual team members have specific and varied roles and interdependence among members is required to perform a specific task". This interdependence is what we need to be able to set up the analogy.

The use of the word team is a specification of the kind of groups that we will consider. The definition of the word team as used here, means that there needs to be some task that can only be executed with the use of different team members, where, once the team is up and running, two team members cannot be interchanged arbitrarily. In the experiment however, this kind of complexity will not be added. The reason a team is still the kind of network we want to consider is that we have interest in the roles they take in their regular teams. Adding this complexity of different roles in the experiment to the analogy should be kept out of scope here, but should definitely be added in possible new iterations of the analogy. However, as we will see in the final theoretical framework, we do take centrality measures into account, that do, to some extent, represent the different positions of the participants in the network.

11.6. Theoretical Framework

In our search for the analogy between information resiliency in ICNs and information resiliency in human collaborative network, we build a theoretical framework. This framework helps us twofold. First of all, it helps us to shape and fine tune the experiment. Secondly, it can be of help in finding the borders of the analogy. Hence, it is used as a theoretical framework, to shape the approach towards the analogy.

In the Section 11.2, we have formulated a definition of information resilience in teams. Furthermore, we have discussed the key aspects of this information resilience. To make our analogy, we make use of these aspects in our theoretical framework.

11.6.1. Overview of Factors Influencing Information Sharing

The information sharing in networks is well described in literature. The next step in building the theoretical framework is to deepen the knowledge about what influences information sharing in a network. Using the reviewed literature (Appendices A and B), we get a network of factors influencing the information sharing in networks. See Figure 11.6. In the figure, the factors are coloured in such a way that they agree with the factors of the initial framework in figure 11.5 (physical and technological boundary conditions, Social Network Analysis and Theory of Reasoned action).

It is necessary to make a choice between all these factors that influence the information spread in human collaborative networks. We will shortly discuss the elements in Figure 11.6, and explain the choices to include or exclude the factor in our theoretical framework. In doing this, the reader needs to take into account that the ultimate goal of the theoretical scoping is to find factors of human behaviour that influence the information sharing and the information resilience. We focus on human behaviour rather than physical aspects, because these factors are variables in our experiment which can be both measured and can help us understanding unexpected differences with ICNs. Furthermore, human behaviour factors can be used to make a choice in which nodes (people) will be removed from the team in the resilience experiment. More about that can be found in Chapter 13, where we discuss the methodology.

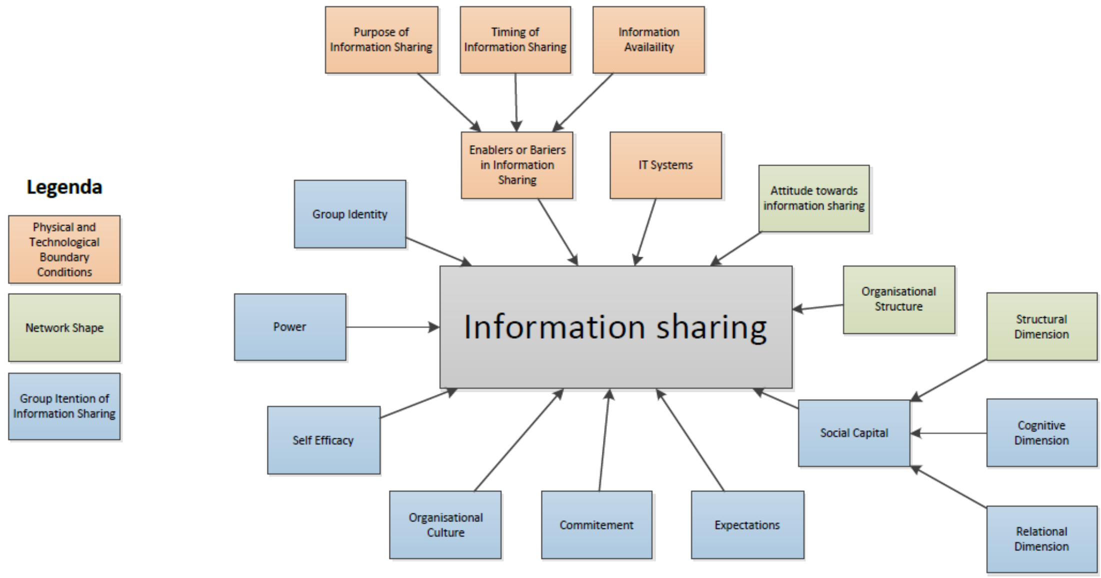


Figure 11.6: The large theoretical network of information spread in human collaborative network

Before moving to the discussion of the factors in Figure 11.6, one note needs to be made. A substantial part of the literature found deals with knowledge sharing instead of information sharing. Even though we have discussed the difference in Section 11.1 in detail, we have chosen to accept knowledge sharing in literature. This has been done because the reasons and factors that influence knowledge sharing and information sharing are largely the same. For example, it does not matter much if we deal with knowledge or with information when we say that trust does play an important role in the decision to share or not to share. It does matter, however, when we deal with it on an individual level (one bit of knowledge or information) but in general you would be more adverse in sharing both knowledge and information to a person you do not trust.

Physical and Technological Boundary Conditions

Let us focus on the red squares of Figure 11.6 first. In the systematic review, we have come across some factors that provide us with physical and technological boundary conditions of information sharing.

Widén-Wulff (2007) mention a lot of bariers and enablers to information sharing. Furthermore, they mention a lot of aspects that play a large role in information sharing. These aspects will come back in the rest of this chapter.

Widén-Wulff mentions the availability of information, timing and purpose as three enablers or barriers. But also identity (both personal and in a group) and culture of the group are mentioned. They could be considered boundary conditions too, but we have chosen to incorporate them in the regions of network shape and intention of sharing.

Kim and Lee (2006) focuses on the IT systems and boundary conditions that need to be met in order to have information sharing within organisations as optimal as possible.

Even though not found in the literature research (which focuses on information sharing in organisations), there are more technical and physical boundary conditions one could think of. An example of this is physical distance. Communication becomes harder (not nessesarily impossible due to technological possibilities) when the distance becomes larger. But as said before, we want to focus on the human behaviour component of information sharing. It is also therefore that the items mentioned in the red squares of Figure 11.6 will not be part of the final theoretical scope. These aspects are boundary condition with a physical or technical ground.

Due to convenience and the line of the story, the blue and green lines will not be strictly separately discussed, but will be mixed.

Organisational Structure and Culture

In the same figure, there are two squares called organisational structure and organisational culture. As already mentioned above, Widén-Wulff mentiones the organisational culture as one of the aspects that influence information sharing. Also in (Kim & Lee, 2006), organisational culture and structure play a large role in the theoretical scope on information sharing in organisations. Both concepts are rather complex in its nature. Many things contribute to these concepts. Kim and Lee considers vision, goals, trust and the social network to be components that belong to the organisational culture. Furthermore, they consider formalization, centralization and performance based reward systems to be part of the organisational structure. For our purpose, organisational culture and structure are therefore too broad and will not be part of the final theoretical scoping. However, as we will see, aspects of it (such as trust) will come back in the final framework.

One could argue that instead of not taking it into consideration because it is too broad, one could also make attempt in narrowing the concepts down to aspects that are concise enough to be taken into account. However, as mapping the organisational structure and culture is very costly time wise, and it will hardly play any role in the way teams interact in the experiment, we have chosen not to do so. This is because the experiment will take place in a closed setting, and so the organisational structure does not play a role at all. The organisational culture will play a role, but it will be embedded in other factors too.

Power

Another square that plays a role in the way and the extent to which information is shared is the power within the network. With this, we mean that if there is a power difference between two people in the network, they will react differently with each other. In their article, Chuang, Chen, and Tsai (2015) focus on Social Exchange Theory in information sharing in supply chain management. In this theory, the decision of sharing depends largely on the expected personal benefit. As such, they define 4 factors that influence the information sharing, namely trust, commitment, reciprocity, and power. As we will see, trust and reciprocity are also mentioned in relation to social capital. We will come back to that later. Also the commitment factor will be discussed later on.

Power can be understood in two different ways. One way is with respect to resource scarcity. If team members possess information that others do not (but want to have), this team member gets a certain power. Another way to understand power is in a more absolute way. A boss has a certain power over its employees as he decides upon their promotion, salary etcetera. Chuang et al. prove that power in its first meaning indeed influences the information sharing. According to Derks (2017b), power relations in the second meaning (boss over employee) also have a very large influence on the way people communicate together. Furthermore, she says that in these kind of experiments such power relations should be prevented. It is therefore that we choose the teams in such a way that they have a flat hierarchy. This way, there is no power difference on paper. Although there might be other ways a power difference occurs, this will be hard to measure as these power relations are probably hidden. It is therefore that we do not take the power relations into consideration in the theoretical scoping.

Group Identity

The group identity factors are mentioned by Widén-Wulff (2007) as one of the factors that influence the information sharing in network. According to Tyler (2003), multiple factors built to the group identity. Examples of these factors are identification, roles, cooperation and values. These factors are complex in nature, and it is hard to measure these factors, let alone measure the combination of the factors. It is therefore that we do not take group identity into account into our theoretical framework.

Self Efficacy

Self efficacy is the extent to which a person beliefs he or she is able to reach the intended results. This influences the way a person acts, motivates himself and think (Chuang et al., 2015).

According to Chuang et al., self efficacy influences the attitude towards knowledge sharing, and the intention of knowledge sharing. As this influence is rather implicit, and therefore we leave self efficacy out of our theoretical framework.

Commitment

Wu, Chuang, and Hsu (2014) mention commitment as one of the factors that influence the information sharing in supply chain performance. According to Yang, Wang, Wong, and Lai (2008) this commitment is commitment towards the relation. In other words, they point out that the information sharing is positively influenced by the commitment of having a strong relationship.

As this influence is indirect, we decide to not take the commitment towards the relationship into account. The commitment towards information sharing itself, is implicitly caught in the willingness and eagerness to share information. Hence, we do not take commitment towards the relationship into account in the theoretical framework.

Expectations

The expectation of what will be the outcomes of knowledge sharing play a role in the intend to share or not. Compeau and Higgins (1995) point out two types of outcome expectations, namely community related and personal expectations. Chiu, Hsu, and Wang (2006) show that mainly the community related outcome expectations have an influence on the knowledge sharing in virtual communities.

As these outcome expectations are very hard to measure, and as the expectations will not be part of the experiments, we also leave them out of the theoretical framework.

Attitude

De Vries, Van den Hooff, and de Ridder (2006) have written an article where they take attitude into consideration in knowledge sharing. They argue that knowledge sharing has both a receiving and a sending component in it. They say that there is a difference between collecting and donating knowledge. Next, this difference is used to also distinguish between two kinds of positive attitudes towards knowledge sharing, namely willingness and eagerness.

They define willingness as "the extent to which an individual is prepared to grant other group members access to his or her individual intellectual capital." Eagerness is defined as "the extent to which an individual has a strong internal drive to communicate his or her individual intellectual capital to other group members." Please note here that there is no reason to suppose that this intellectual capital must be as strong as knowledge, and cannot be limited to information. According to De Vries et al., willingness is rather passive in nature, while eagerness has an active connotation. Those eager for information will more actively search search for it than those willing to share. As we will see in Chapter 12.1, the concepts of willingness and eagerness to share can be linked to the pushing and pulling concepts in the networks.

van den Hooff and Hendrix (2005) uses this primarily theoretical distinction to set up an experiment where they measure the individual eagerness and willingness for knowledge (information) sharing. They show that this distinction is present in practice, and that it influences the way knowledge is being shared. The factors are applicable to the level of our experiments. Furthermore, van den Hooff and Hendrix explain how to measure this eagerness and willingness of knowledges sharing. They come up with a questionnaire that will be discussed in Section 13.3.2.

We decide to take these two aspects of the attitude towards sharing, willingness and eagerness, into account in our theoretical framework. This is done because it is measurable on the one hand, and it is somewhat more sober than for example self efficacy. They are simple in the sense that they are easy to understand and to measure. However, they are complex in the sense that a lot of things can cause a change in the attitude.

It is important to note here that this attitude is towards sharing in general, and not towards sharing to a specific other individual. However, the questions do relate to the group of people they share with in the experiment.

Social Capital

The last squares in Figure 11.6 that have not been discussed yet are social capital, and the three squares connected to that. As we will see, Social Capital theory will play a large role in the remainder of this chapter. Social capital theory is a very broad and well known theory, which encompasses all that is enclosed by a social network. The people in your network might be very useful resources. According to Nahapiet and Ghoshal (1998) Social capital is, "the sum of the actual and potential resources embedded within, available through, and derived from the network of relationships possessed by an individual or social unit. Social capital thus comprises both the network and the assets that may be mobilized through that network." Although the different aspect of social capital can be measured (as is done by Stone (2001) for example), there is no cumulative measurement that allows us to appreciate one teams social capital over another. In our case, we are interested in a few very specific aspects of social capital, so this is no problem.

Social capital is often used to explain the motives of people to share information. Among others, Chiu et al. (2006) have used this for knowledge sharing in virtual communities, whereas Díez-Vial and Montoro-Sánchez (2014) show that the social capital factors play a role in knowledge sharing in organisations. Social capital is usually split into three dimensions. These dimensions are the relational, cognitive and structural dimension (Nahapiet & Ghoshal, 1998), and can also be found in Figure 11.6. These three dimensions will be elaborated on next.

Relational Dimension of Social Capital

The first dimension of social capital considered is the relational dimension. Where the structural dimension we will discuss later (Section 11.6.1) deals with the network and its ties (so what is the structure of the social network looking like), the relational dimension deals with the kind of relation people in the network have.

These relations come with different concepts, such as friendship, liking each other, respect, trust and many more. These relations influence the behaviour of the members of the network. To give one example of that, if you do not like a person, you less likely would want to share a personal experience with this person. Nahapiet and Ghoshal mention 4 key facets of the relational dimension. These are trust and trustworthiness, norms and sanctions, obligations and expectations, and lastly identity and identification. As said before, we focus on those aspects that are both measurable and that are relevant in a closed setting. Therefore, the identity and identification facet will not be included. Furthermore, the norms and sanctions facet is excluded as it is hard to catch these in a closed setting. The other two aspects of the relational dimension will be taken into consideration.

Trust and Trustworthiness

This first facet deals with trusting the people in the network, and being trusted by those in the network. If you do not trust a person, you expect your information to be unsafe with this person. Trust is used in our every day language all the time, but it is hard to define trust. Many definitions exist in current literature. As we do not speak about trust in organisations, institutions or for example a prediction, we focus ourselves on interpersonal trust (being the trust that exists between 2 people). Rotter (1967) defines trust as "an expectancy held by an individual or a group that the word, promise, verbal or written statement of another individual or group can be relied upon." On the other hand, Larzelere and Huston (1980) defines interpersonal trust as "a belief by a person in the integrity of another individual." In this definition, integrity plays a role, while relying on another person's statement is the important factor in the definition by Rotter.

According to Rempel, Holmes, and Zanna (1985), interpersonal trust encompasses three main elements. These elements are predictability (will the other person react as expected?), dependability (to what extent is it possible to depend on the other person?) and faith. The first two elements have some kind of learning aspect in it. With this, we mean that feelings of predictability and dependability are often based on prior experiences. However, as we deal with new situations all the time, to trust someone also encompasses the belief that the other person will be able to deal with these new situations too. This belief is what is called faith by Rempel et al..

We will use the three elements (dependability, predictability and faith) of Rempel et al., as they also provide us with a way of measuring these elements of trust. Sanders, Schyns, Dietz, and Den Hartog (2006) provide us with a very broad overview of interpersonal trust in organisations. They compare fourteen different measures. Al these different approaches measure a different kind of trust. This can be trust in a manager, trust between departments, but also trust between individual employees. in Table 11.1, an overview of this can be found.

As we are interested in the trust between different people in a team, we will pick the questionnaire of (McAllister, 1995) as the one to measure. We come back to that in Section 13.3.

Table 11.1: An overview of the trust measurements, and the entities between which is measured.

Author	Trust between who?		
(McAllister, 1995)	A work colleague (peer)		
(Robinson, 1996)	Employer		
(Cummings & Bromiley, 1996)	Other departments		
(Clark & Payne, 1997)	Employer	management	Most Managers
(Brockner, Siegel, Daly, Tyler, & Martin, 1997)	Employer		
(Mayer & Davis, 1999)	Top Management		
(Spreitzer & Mishra, 1999)	Employees in general		
(Shockley-Zalabak, Ellis, & Winograd, 2000)	Supervisor & Management	Co-workers /Peers	Whole organisation
(Tyler, 2003)	Top Management	Immediate Supervisor	
(Huff & Kelley, 2003)	Everyone in the participating Organisation		
(Gillespie, 2003)	Immediate Supervisor		
(Tzafrir & Dolan, 2004)	Specific core employees (as a group)	Specific core managers (as a group)	

Obligations and Expectations

Nahapiet and Ghoshal (1998) define obligations and expectations as one of the factors of the relational dimension. These factors however, can be decomposed in a lot of smaller factors. Akhavan and Hosseini (2016) for example mention friendship as one of these. On the other hand, reciprocity is also mentioned as one factor by for example Nahapiet and Ghoshal. This reciprocity is defined by Falk and Fischbacher (2006) as follows: "People are reciprocal if they reward kind actions and punish unkind ones."

(Collins & Miller, 1994) has shown that self disclosure and liking influence each other positively. Self disclosure is a way of information sharing, but it focusses on personal information. (Derks, 2017b), however, suggests that liking is also of importance when dealing with other kinds of information sharing. Also in professional teams, the extent to which you like someone influences the way and amount of information sharing.

Relational Factors in Information Sharing

Describing social capital factors as factors that influence the information sharing in a collaborative network is very common. According to Chuang et al. (2015), individuals and groups try to interact when they expect something in return. This theory of social exchange is then used to identify 4 factors that could possibly influence the information sharing within teams. These four factors are trust, commitment, reciprocity and power. However, when considering social capital theory, trust and reciprocity play a role too. Furthermore, commitment is a specific part of shared norms and values, as it is meant to establish a common ground to communicate from.

This relational aspect of information sharing (trust and reciprocity) is seen in many more investigations. Wu et al. (2014) show that anticipated reciprocal relationships influence our attitude towards knowledge (and information) sharing. Hence, it is not only the kind of relationship the members have together that influence the information sharing behaviour. The expectation of the value of the relationship itself is already enough to have positive effect on the information sharing behaviour.

From the relational dimension of social capital we decide to take trust and reciprocity as the two factors of interest that we incorporate in the theoretical framework.

Structural Dimension of Social Capital

As mentioned before, the 'shape' of the network is at least as important as the relational dimension. This structural dimension is less personal than the relational dimension, and deals with the patterns of connections. Also for this dimension, Nahapiet and Ghoshal (1998) have defined the most important facets, being the configuration of network ties, the patterns of linkages in terms of mathematical measures, and the organisation of the network. With the latter, the purpose of the network, and usage for other purposes is being meant. For example, people that are both friends and colleagues might use their friendship network to discuss work related issues.

The structural dimension plays a very large role in the research, as it lies on the boundary of mathematics and social sciences. It is possible to grab the structural dimension of social capital in computer simulations. Using ICN for this is just one option, but is has done much more often, and for various applications. Brynielsson, Högberg, Kaati, Mårtenson, and Svenson (2010), for example, uses simulation to detect the social positions in a network (also in a later article (Brynielsson, Kaati, & Svenson, 2012)). Michalski, Palus, and Kazienko (2011) tries to match the social network with the organisational structure by using large data sets of email communication. A whole field of research is named after this mathematical analyses of networks: Social Network Analyses. This comes back largely to the structural dimension of social capital theory.

Because we deal with a closed setting in the experiment, we will focus on the first two facets of the structural dimension, being the network ties and the mathematical measurements.

Network Ties

The paragraphs about network ties and patterns in the network relate to each other very closely. Within organisations, there are often social networks that do not necessarily follow the hierarchical structure, but that have proven to be very important in solving issues and problems (Cross, Parker, Christensen, Anthony, & Roth, 2004). Mitchell (1969) defines these social networks as " "a specific set of linkages among a defined set of persons, with the additional property that the characteristics of these linkages as a whole may be used to interpret the social behaviour of the persons involved." Hence, the linkages between the persons in the social network should be of such nature that we can interpret the behaviour of these people.

It is possible to map these relations between the people in an organisation. The way this is done is often through direct questioning (Scott, 2012). This direct questioning can be then done in different ways that we will not elaborate on now.

As an analogy with ICNs will always require a network to work with, these network ties do not only play a role in the information sharing part of the theoretical lens, but also in the resilience part. Also the information resilience in a human collaborative network depends largely on the (social) network considered. As we have seen in the mathematical analyses of networks, adding more ties to the network increases the resilience (see Chapter 5). Although not related one on one to a human network, the same principle does hold here too. Dodds, Watts, and Sabel (2003) show in their paper that certain network topologies of an organisational structure are vulnerable to disintegration when a failure occurs. The term robustness is used here, which is, as discussed in Section 11.2 quite similar to resilience.

Patterns in the Network

Where the network ties are concerned with the question "who has contact with who?", we deal with the question "who has what role in the network?" In this paragraph we consider some of these roles, as well as their corresponding way to measure these roles. However, as there are many different roles that can be distinguished between in a network, we have chosen to only consider the most important ones.

First of all, there is a set of roles that are associated with the kind of connection that certain individuals in the network have to other parts of the organisation. a liaison for example, is connected to other departments of the organisation. Gatekeepers have contacts outside the organisation. Being uncoupled with the rest of the network makes an individual an isolator (Tichy, Tushman, & Fombrun, 1979). These roles are less interesting to us, as we will focus on the collaborative network only, and not on it outbound ties.

Secondly, there is a group of roles that are more active within the considered network, namely the centrality of the persons in the network. The general idea is the following: the more central a person is in the network, the more influence he/she has. There are many different ways to compute this centrality (Stephenson & Zelen, 1989). We will discuss the 3 main centrality measures, namely closeness, eigenvector and betweenness centrality, here.

Closeness centrality investigated how close one node is to all the other nodes, using the shortest paths. This can then be ranked (Freeman, 1978). This is then normalised to the sum of minimal possible distances. The interpretation of it is that the person with the highest closeness centrality has on average the shortest paths to the others in the network. Hence, he can reach everyone from the network the quickest/easiest. The built-in function of python package networkx uses the following definition to compute the closeness centrality:

$$C(u) = \frac{n-1}{\sum_{v=1}^{n-1} d(v,u)}, \tag{11.1}$$

where $d(v,u)$ is the shortest-path distance between v and u, and n is the number of nodes in the graph. Note furthermore that a higher number indicates a higher centrality.

Betweenness centrality was defined by Freeman (1977), and is more interested in how often a certain node appears on a shortest path between two other nodes. Bavelas (1948) was the first to describe this centrality measure. According to him, when a person is located strategically on the shortest communication paths between people, this person has a lot of influence to the communication in the network.

The built-in function of python package networkx describes the betweenness centrality function as: the betweenness centrality of a node v is the sum of the fraction of all-pairs shortest paths that pass through v:

$$c_B(v) = \sum_{s,t \in V} \frac{\sigma(s,t|v)}{\sigma(s,t)}, \tag{11.2}$$

where V is the set of nodes, $\sigma(s,t)$ is the number of shortest (s,t)-paths, and $\sigma(s,t|v)$ is the number of those paths passing through some node v other than s, t. If $s = t, \sigma(s,t) = 1$,and if $v \in s,t, \sigma(s,t|v) = 0$.

Eigenvector centrality is mathematically spoken a little more complex. The interpretation however is, that being liked by a person who is very popular himself makes you more likeable (Bonacich & Lloyd, 2001). This way, the people with most influence in the network can be determined. The mathematical calculation goes through the eigenvectors of the adjacency matrix.

We will use all three measures of centrality, as they all have a different interpretation. In our opinion, all three measures add something to our understanding of the social network. Closeness centrality is concerned with how 'far' all others in the network are situated from you. Betweenness centrality is more interested in how well a person in the network can influence the communication in a network. Finally, the eigenvector centrality says more about the power of the individuals in the network. All have their added value, so for now we group the three measures. They will appear in the final framework as centrality factors.

Cognitive Dimension of Social Capital

In the cognitive dimension, the set of values, experiences and other resources that provide the network with a common language, interpretations and a context. These factors cannot be neglected when considering human communication as a system. However, as we focus on the possibilities of building an analogy, we will not take these factors into account, as they are hard to grasp or measure.

One important note should be made however. In this research we try to stick to information sharing in its cleanest shape. However, humans do no just transfer information without interpretation. The factors mentioned above matter a lot to the way the information is interpreted. But since this interpretation (mistakes) are unique every time they occur, they will not be easily measured or even noticed.

11.6.2. Theoretical Framework for Information Sharing

Previously, we have discussed all the factors found in literature that influence the way and extent to which information in a network is being shared. We have made choices, based on our objective, which is setting up the analogy. To reach this objective, we look for aspects that are measurable in a closed setting. These aspects will help us in making a choice of which people to exclude from the network once the resilience part of the experiment starts.

The factors we have included are: trust, liking, reciprocity, centrality, social network ties and attitude (both eagerness and willingness). These factors are largely inspired on social capital theory, which is often used as the main theory to explain the way information sharing happens. The factors can be classified in three sets. First of all, there are the relational factors, namely trust, liking and reciprocity. They are part of the relational dimension of social capital theory. Secondly, there are two factors from the structural dimension of social capital theory, namely centrality and social network ties. Thirdly,

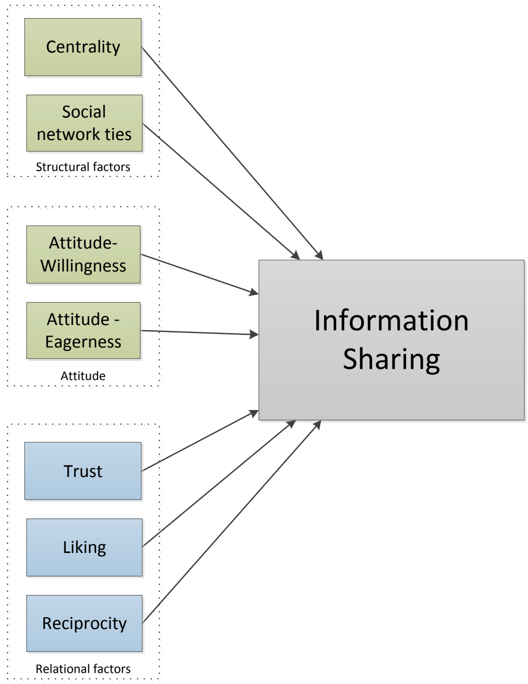


Figure 11.7: The large theoretical network of information spread in human collaborative network

two factors are added that are not part of social capital, namely eagerness and willingness of sharing. These two are both related to the attitude towards sharing. These factors and their classification can be found in Figure 11.7.

Akhavan and Hosseini (2015) have performed a literature research in which they sum up and count all the determinants from social capital mentioned in literature on knowledge sharing. We believe that the outcomes of this research give us an indication that the factors we consider are indeed factors that are common to consider when it comes to information (or knowledge) sharing. In the research performed by Akhavan and Hosseini, 50 articles on knowledge sharing and social capital (in different settings and applications) are taken into account. In Table 11.2, the determinants together with their dimension are being depicted. Furthermore the number of articles that mention this determinant as important as well as a note whether the determinants are being incorporated are shown.

What we see is that the five factors from social capital that we include encompass the top three items as mentioned in literature (as reviewed by Akhavan and Hosseini). Furthermore, the three aspects that we add (namely liking, willingness and eagerness to share) to this are based on an interview (Derks, 2017b), and literature outside the scope of social capital (De Vries et al., 2006).

Now that we have focused our theoretical lens towards information sharing, we will move along towards resilience. In the next section, we will take the information sharing factors as a starting point, and see which of these factors directly influence the information resilience (instead of only through information sharing).

11.6.3. Complete Theoretical Framework

In Figure 11.8, the complete framework can be found. As can be seen, the information resilience has been added to the framework. As said before, the information resilience in a team that we mean here, is more restrained then the definitions on resilience are in literature. This is due to the fact that the

Table 11.2: determinants for knowledge sharing, as in (Akhavan & Hosseini, 2016)

Deteminant	# mentioned	Structural Dimension	Relational Dimension	Cognitive Dimension	picked?
Trust	31		x		yes
Social network ties	21	x			yes
Reciprocity	13		x		yes
Shared language	11			x	no
Identification	11		x		no
Shared vision	11			x	no
Social interaction	10	x			no
Norms	9		x		no
Tie Strengths	8	x			no
Commitment	6		x		no
Obligations	4		x		no
Centrality	4	x			yes
Proximity of employees	3	x			no
Network configuration	3	x			no
Shared narratives	3			x	no
Network density	3	x			no
Training	2			x	no
Appropriable organization	2	x			no
Tenure in the field	2			x	no
Shared values	2		x		no

analogy is central in our research. It is therefore that the factors that usually apply to the resilience do not necessarily apply to the information resilience here.

Because there is little to nothing written in literature about information resilience, we will stick in our theoretical scope to the factors that were also associated with information sharing. The structural factors relate to the information resilience directly. As we have seen in Part II, the structure of the network is very important for its resilience. This encompasses both the ties (also called the edges or links in Part II) in the network and the importance certain people have in the network (hence, the centrality of certain people).

Secondly, the information sharing influences the information resilience directly. If all information sharing in a network goes through a certain person, and that person somehow withdraws from the network, there might be no information sharing at all any more. Hence, the way information spreads through the network is very important for the resilience in the network.

The attitude and relational factors have an influence on the information resilience, but mostly through information sharing. By that, we mean that the attitude and relational factors influence mainly the way information is shared and spread.

Now that the theoretical scope is being set, we will take a look at analogies in general. What are they, and what do we need to make a proper analogy? Furthermore, the natural boundary conditions and subjects of comparison will be determined there.

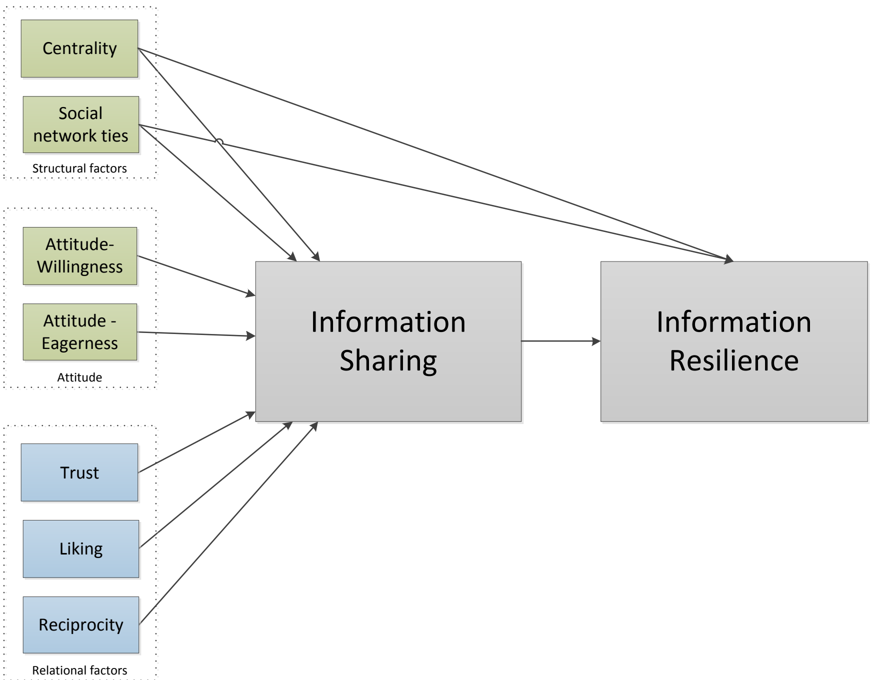


Figure 11.8: The theoretical scope on information resilience

12
Analogies

"[Over het 'motorblok' van VVD, D66 en CDA] Ik raad de verkenner aan om onder de verdeel-en-heers-kap te kijken, en goed de zuigers te beoordelen, en rekening te houden met de partijen die vinden dat de motor op schone stroom moet draaien, en met de lijsttrekkers die natuurlijk geen vijfde wiel aan de wagen willen zijn [...] Vaststaat voorlopig wel dat het motorblok nog te weinig pk heeft [...]. Een metafoor in dit verband is levensgevaarlijk, [..] voordat je het weet maak je teveel toeren en zit er zand in de machine."

Frits Spits

In this chapter we will consider analogies in general in Section 12.1. In Section 12.2, we will consider the analogy specific for us, and find out something about its boundaries and nature. In this section, we will build up the essence of the analogy, which will function as a starting point for the experiments.

12.1. What is an Analogy?

Analogies are used in many fields of expertise such as the biology, psychology or in logic. In all these expertise fields, the definition of the analogy differs slightly. In biology for example, an analogy considers equal functions, coming from different origins (Boyden, 1947) (as opposite to a homology, where different functions and a common origin are considered). An example of this are the wings of bats and insects. Although having the same function (to fly), they have different evolutionary origins. See also Figure 12.1.

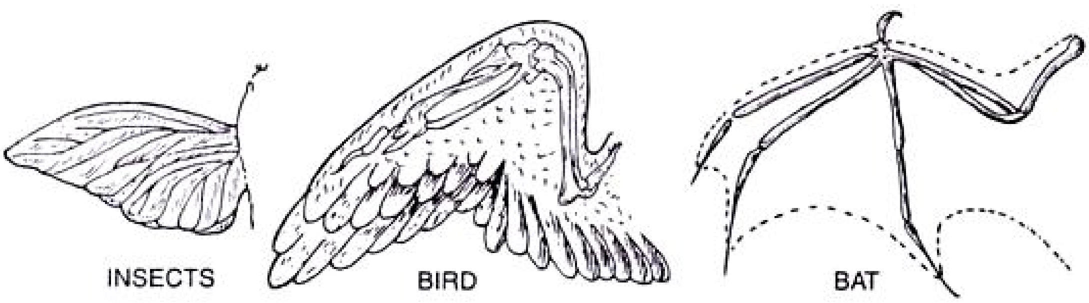


Figure 12.1: Analogy between wings of different animals. Retrieved from http://cdn.yourarticlelibrary.com/wp-content/uploads/2014/06/image_thumb32.png

Now, if we allow us to look further than biology only, but keep this definition, many more analogies can be found. To just give an example, gecko feet and Velcro straps are an analogy too. They have different origins, but their function is similar (sticking).

Even though one intuitively immediately sees the comparison between gecko feet and Velcro straps, there is a large problem here. That is that sticking to walls is not the only function of gecko feet, as they are also meant to walk on. There is no way that Velcro straps are useful to walk on. Hence, the analogy only makes sense when the subject of comparison is (made) clear.

Now, this problem becomes even bigger when we do not stick to the definition of an analogy as used in biology, but the much broader definition as it is given by the Oxford dictionary: "A comparison between one thing and another, typically for the purpose of explanation or clarification." This comparison needs a subject to have meaning. Gentner (1983) has construced a theoretical framework about how humans derive meaning from an analogy. Often, with simple analogies, this meaning becomes clear immediately once we have the two objects that are being compared. For example, in the analogy "tires are like shoes" (Gentner, Bowdle, Wolff, & Boronat, 2001) no explanation is needed to see the subject of the analogy.

Already with slightly more complex analogies however, the subject of comparison must be made explicit. To give an example, the analogy "Facebook is like Starbucks" (by Chris Dixon) does not mean anything. However, when adding the subject of the analogy (internet commerce) and a little explanation ("Facebook is like a Starbucks where everyone hangs out for hours but almost never buys anything"), it does have meaning.

One could also argue that, once the explanation or essence of the analogy is being understood, there are often multiple examples of analogies possible. An analogy helps us to clarify the essence of a certain problem or phenomenon. Once this is done, we can then apply this essence to other examples too. For example, the analogy between "Finding a good man is like finding a needle in a haystack" contains the essence that it is very hard to find these two things. However, we can then also use this to make the analogy for someone who lost her necklace: "finding back your necklace is like finding a needle in a haystack".

So, when we are talking about an analogy between interaction in a human collaborative network and an ICN, we need to make the subject of the analogy clear for each part of the analogy. Furthermore, some explanation might be needed, to explain why the two different aspects are comparable. Or to go back to the definition as used in biology: we need to explain why the two things with a different origin (namely an ICN and a human collaborative network) have the same function. Also, the function should be as explicit as possible.

12.2. Analogy Between ICN and Human Collaborative Network

In this section, we will consider the analogy, its nature, purpose, and its boundaries.

12.2.1. Nature of Analogy

In a human collaborative network, people with different backgrounds work together to obtain a common goal. In this process, they need to share information. In Chapter 11, the different aspects that influence this information sharing process have been discussed already. Some of these aspects, such as trust, are purely human in nature, while others, such as the existence of a relationship between the participators (network ties), can be partly explained from the point of view of an ICN as well. In this section, we compare the functionalities and functions of a human collaborative network with the functions and functionalities of and ICN. This enables us to think about the analogy from a more abstract level. Those aspects that we cannot immediately link in the analogy, or those aspects that we left out intentionally are discussed in the next section (Section 12.2.2). In Table **??**, the parallels between the two sides of the analogy can be found.

Purpose of Both Networks

In order to be able to set up the analogy in a proper way, we need to have both the purposes of the networks clear as well as having very clearly in mind why we want to make the analogy. The reason for this is that some things intrinsically sound good to do in an experiment, but they might influence the ability to make a proper analogy or they miss the goal of the analogy (by having too much or too little detail). It is therefore that we shortly repeat the three elements here.

The purpose of an ICN
An ICN has the purpose of distributing information from sources or caches to the end users. Its aim is to do this as quickly and reliable as possible.

The purpose of a human collaborative network
A human collaborative network uses information sharing as a means to achieve some kind of goal.

Reason for comparison and analogy set up
We want to compare the two in order to gain more insights in the information sharing in human collaborative networks. If we can catch (parts of) the way information spreads through a human collaborative network in an ICN simulation, we could also use the simulation software to research possible effects of disturbances in the network. Hence, if the analogy can be set up, decision makers might be able to get more insights in the information sharing of their team, and information resilience in their team in particular. What needs to be noted here, is that it is only a very first attempt to set up such a tool, much detail will still be missing, and should be added in further research. The guiding principle of this research is to keep the analogy and the variables in the analogy simple, in order to investigate the possibilities and boundaries of the analogy.

Roles in Human Collaborative Networks

Imagine a project group of people working together, developing a communication strategy for a company. One of the members (called Alice) of the group is graphically very strong. One other person (called Bob) has very much experience, and knows a lot about best practices. A third person (called Charlie) manages the contacts with the clients. In the human collaborative network, each person has its own role. Because they have different expertise, they are the source of different kinds of information. For example, when the client has some requests, Charlie is most probably the source of the information within the network. When Alice receives this information, he might decide to go talk to Bob about the information, as it might have an impact on the graphic design of the communication strategy. When the nature of information is different, another person most probably will be the source of information in the network.

We consider someone to be a router in the network (for a specific piece of information) when they receive the information, and (at some point) pass the information along to someone else. What we leave out in the analogy is interpretation of the information as well as the ability of humans to compose new information from different pieces of information. With the latter, we mean that humans draw (logical) conclusions based on information all day. Sometimes it is a typical case of jumping to the conclusion, sometimes it is just a logical conclusion from the premises. This kind of behaviour belongs to the branch of knowledge instead of information, as discussed in Section 11.1. As said before, we come back to the exclusions of the analogy in the next section.

A receiver of information in this analogy is someone who receives a piece of information, and who does not transfer this information to anybody else anymore. Note that a person can take a different role in the network for different pieces of information. While one could argue if it holds for all kinds of networks (for example including mobile networks), the ICNs we consider have one function per node. This is either a source, a router or a receiver. The receiver requests information, which is sent to it from the source through the routers. As said before, these routers possibly have a caching capacity from where the information can also be send.

Workload and Content in Human Collaborative Networks

The workload in an ICN is the set of information requests that are being requested during the measurement period. This workload include, per request an end user, an identifier for the information and a timestamp. As can be found in Table **??**, the analogy between the workload in human collaborative networks and ICNs is hard to make. This is because (at least) some of the information spread in a human collaborative network is pushed whereas the information spread in the ICN is a consequence of a pull principle. In an ICN, an end user requests for information, which is then looked for in the rest of the network and is being transferred (and possibly cached on the way) to the end user. Hence, what information is cached where in the network depends on what information the end users pulled from the network.

In a human collaborative network however, the source of information (A) often decides that some information is of added value to someone else (B). The information is being shared. If B decides that the information is valuable to others as well, then B can transfer the information again to C. As pull mechanisms in a human collaborative network exist parallel to push mechanisms, the workload does not have an obvious analogue in the human collaborative network. It is interesting to think about the concepts of eagerness and willingness to share as discussed in Chapter 11. Willingness to share is associated with the willingness to open up your personal information when requested. In other words, it says something about the individuals' attitude towards sharing information in a pull situation. If the other person requests for information, will you provide it? Eagerness to share, however, is associated with the extent to which you decide to share information with others from an intrinsic drive. Hence, it is related to the push mechanisms described above. The individual decides that the information present to him/her is interesting to others as well, and that it is therefore interesting to share. We will reflect on these two factors in combination with the experiments in Section 16.

The content on the other hand, is largely similar in both networks. In both cases, some choices need to be made with respect to the definition of a 'piece' of content. In ICNs, information is sent in packages, hence one could really talk about a piece of information. However, it is often assumed that all pieces have exactly the same size. In reality, this is not necessarily the case. In human communication, this border is also not always clear. We have discussed this issue already in Section 11.1. In the next section about the boundaries of the analogy, we will add some more detail to our definition of the word 'piece'.

Caching in Human Collaborative Networks

If we assume that people remember everything, then all pieces of information that they ever come in contact with will be added to their infinitely large cache. If we do not assume that, we will have to adjust the caching size and policy for the memory of the humans in the network. There are multiple ways to look at the memory of humans, which we will do in Section 12.2.2.

Disturbances in Human Collaborative Networks

In a human collaborative network, breaking links would mean that two people no longer exchange information for any kind of reason (a fight, physical distance etcetera). Breaking nodes would mean that a person is removed from the network (for example due to long-term illness or because a person quits his job). The third comparison is rather far fetched, but loosing its caching capacity would be similar to losing the ability to save information (memory). Closest to this analogy would be for example the loss of a notebook or a personal computer crash. Suddenly, not all information the person had access to before is available anymore. In this thesis, we will restrict ourselves to people dissappearing from the network, as its expected effects are the largest, and the case is easier to grasp than links or caching capacity getting broken.

12.2.2. Boundaries

In the previous section we have set the nature of the analogy. In this section, we will set and explain the boundaries, what will not be taken into account, and why this is the case. They will return in the discussion section though.

Information

As we need to keep the experiment as manageable and simple as possible, we try to decompose the information into pieces that are as large as a single fact. This is not because humans or ICNs cannot deal with more, but it is to have the information as traceable as possible. Hence, the boundary we

Table 12.1: Overview of the links between ICNs and human cooperative network

Function or parameter in ICN	Parallel in human collaborative network
Topology	Topology of human interaction
Router	Roles in network - a router of information
end user	Roles in network - a receiver of information
Source	Roles in network - a source of information
Content popularity	Content distribution in human communication
Cache	Humans saving information to spread it later again
Disturbance in link, cache or node	Disturbance in contact, person or 'memory'
Caching strategy	What information is memorised by whom[1]
Cache size	How much information is remembered[2]
Workload	The set of information items that is requested from the network[3]
Content population	How much content is being shared?
Cache replacement policy	Irrelevant as the cache size is infinite
Requests per min	Speed of sharing information

choose for the information size is due to practical reasons rather than due to limitations of one of either systems.

Furthermore, humans have the ability to transform and combine information together. Ultimately, this leads to new knowledge. But also before we can talk about knowledge, it is a human feature to combine and interpret information. It is hard to exclude this entirely from the research, even though it is clear that we should not include it when trying to keep the analogy as pure as possible. In designing the experiment, part of the design requirements will be that the possibility of information composition will be ruled out as much as possible.

Push Versus Pull

It has been mentioned already before, but an essential difference between ICNs and human collaborative networks is the notion of pull versus push. In an ICN, all information is being pulled from the caches and sources, whereas humans can decide to pass information through, thereby pushing the information through the network. In the experiment, we will stay as close to the ways humans interact, and allow both pushing and pulling of information. However, in the simulation the information flows will be treated as pull information.

Human Memory

As the experiment needs to take place in a relative short period of time, people are unlikely to forget information. Also on a longer time scale, a discussion could be held on whether we really forget a lot of information (Brady, Konkle, Alvarez, & Oliva, 2008). What surely is true is that some memories can be (temporarily) inaccessible, having the same effect as forgetting. However, for the short time the experiment lasts, we assume the people to be able to remember all information. Furthermore, the questionnaire will allow for mentioning some information is forgotten.

Heterogeneity in Groups of Humans

In an ICN, all routers act in the same way, and therefore are more or less interchangeable. In a human network however, the knowledge (and therefore also part of the information that is spread), personality, experience and many other things differ from user to user. This can be partly dealt with by having different routers and end users for different types of information in the analogy. More about this can be found in Section 14.1.3. However, it remains that humans are much more complex than routers. Their motives to share information are also a lot more complex than it is in ICNs as we have seen in Chapter 11.

[1]On the short term, this is Leave Copy Everywhere. as we assume that people will remember the information
[2]We often assume this to be infinity, as it is very hard to quantify.
[3]Hard, as humans do communicate less through pull and also through push mechanisms.

13 Experiments

Research is creating new knowledge

Neil Armstrong

In this chapter, we discuss the set up and the methodology of the experiments. The literature to design such an experiment is not sufficient to base the experiments on. It is therefore that we need to construct an experiment by ourselves. As a result, we build the experiment in three consecutive steps to use the insights of the earlier experiments to the latter experiments. We elaborate on the methodology as discussed in Section 10.4. Furthermore, we explain the design of the experiments.

In Figure 13.1, a quick overview of the three experiments is given. The next three paragraphs all elaborate on one experiment. Per paragraph, we first explain the goal of the experiment. Next, the methodology to design the experiment is elaborated on. What actions have been taken to come to a design of the experiment, and why did we take these actions? Lastly, the design of the experiment is described.

Per experiment, the goal, methods, sample group and experiment design are stated in just a view words. The next three sections of this chapter elaborate on the items in this figure. The final paragraph of this chapter elaborates on the data analysis of the data obtained in these experiments.

13.1. The First Experiment

In an early stage, some data on the information spread in a human network was collected in the minor class of Science Communication at Delft University of Technology. The reason for this is twofold. First of all, no useful experiment of tracking the analogue information spread was found in literature. Therefore, it is needed to come up with an experiment by ourselves, which needs some phases of testing and revising. Secondly, the data collected was used to set up a first iteration of the analogy. This allowed us to check if a future experiment had any chance of leading to the results we are looking for. If some problems occur, the experiment could be tweaked before the definitive experiments are executed. What we call the first experiment here, is in fact two runs of the same type of experiment, where some adjustments had been made after the first experiment.

13.1.1. Goal of the First Experiment

The goal of the first experiment is to explore the possibilities of catching the information spread in a human collaborative network. The practices found here have been applied in the next experiments.

13.1.2. Methodology of the First Experiment

The first set of experiments can be viewed as exploratory, in order to find out some best practices regarding the experiment. In order to design such an experiment, 4 steps have been taken, which were already discussed in Section 10.4. First of all, a search for literature regarding measuring the

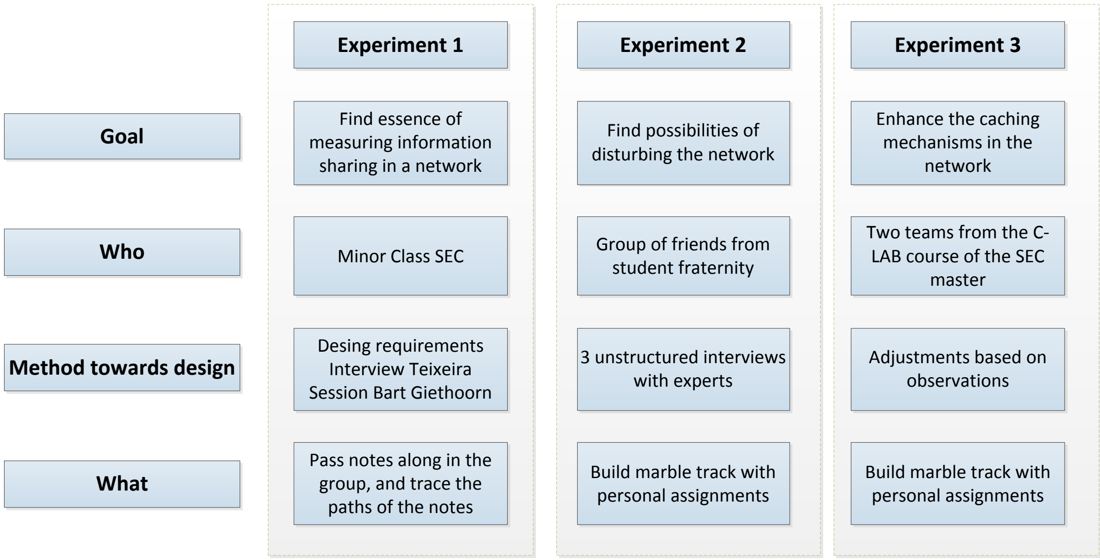


Figure 13.1: Overview of the three experiments

information spread in human collaborative networks was done. Secondly, mr. Teixeira from the Technology, Policy and Management department of Delft University of Technology was interviewed, to get the objectives of the experiment straight. Thirdly, the design objectives were formulated (what should be measured with the experiment, and what are the conditions to that). The last step towards designing a first experiment was in an interactive session with Bart Giethoorn. His company specializes in gamification for education purposes, and together we have composed the set up of the first experiment.

Literature

The search for literature was part of the systematic literature review. We will discuss some literature here that connects to the way information spreads in human collaborative networks. Note here, that not the factors that influence information sharing are of interest (these were already discussed in Chapter 11), but the way the information spread in human collaborative networks can be measured is of interest. This is because we need a way to measure the spread of information in our experiment.

Pentland (2004) has developed the so-called 'sociometer'. This device can be worn by different people in a human network. If two people from the network talk together, the sociometer is able to detect both the person and the duration of the conversation. This data tells us something about who communicates a lot with whom, but the sociometer does not track what the people in the human network communicate about. Hence, caching behaviour cannot be detected by the sociometer.

Digitally, a lot of research has been done on the information spread in a social network. Through Facebook, Twitter, blogs and websites like Reddit, certain messages can be followed when being liked/shared/retweeted/commented on etcetera. For example, Lerman, Ghosh, and Surachawala (2012) have investigated the way news stories cascade through a follower graph (who follows who) on Twitter. Furthermore, Wang et al. (2011) use forwarded emails to investigate the information spread in social networks. These studies have in common that they measure the type of information we want to measure. However, we are merely interested in face to face communication. The reason for this is that face to face communication encompasses a lot more than online communication. Items such as trust and interpretation play a different role in face to face communication. Where humans are able to steer the conversation slightly when they are misinterpreted, this sometimes fails in online communication. We have all had online conversations where the other side of the communication line interpreted the information sent with a different emotion than the sender did. This often results in misunderstandings or even fights that need to be resolved. Furthermore, online communication has different dynamics with regard to timing. In a conversation, one cannot keep silent for 2 minutes. This is perfectly acceptable in a WhatsApp conversation, and with email 2 hours or sometimes days are acceptable as well.

We have not been able to find any literature on measuring information spread (both what and to whom) in face to face communication. It is therefore that we will have to come up with our own experiment.

Unstructured Interview with André Teixeira

We have interviewed André Teixeira (Associate Professor at TPM), expert in communication technologies, to obtain a different perspective on the analogy. At the faculty of TPM, system approach is often used, and the human perspective on technology is part of the research as well, talking to a communication technologies expert is a good way of gaining a new perspective on the analogy. As the result of the interview will be largely intangible, and it will merely contribute to our knowledge and understanding of building up the analogy, the interview will be unstructured, and not coded afterwards.

Where Maarten van der Sanden could approach the ICN from a human network perspective very well, Teixeira was able to view the human network from a ICN perspective. Talking with him enlarged our insights in the analogy. He explained that, in order to be able to design an experiment, it is needed to identify the essence of the analogy, and to define the items of comparison. This has been done in Section 12.1. Furthermore, he pointed out that the aim of the experiment as well as the design requirements should be made explicit, in order to design an experiment. This is what we have done next.

Design Requirements

To know what we try to design, the essence of the analogy needs to be taken into account, as well as the purpose of both networks and the purpose of comparison. To keep the goal of the experiment, as well as the important factors in mind, we have created a list of design requirements. They can be found in Figure 13.2, and will be discussed next.

The main objective of the experiment is to measure what information is spreading through a human collaborative network in what way. With the latter we mean that it should be traceable who has given what information to whom. Furthermore, the participants should behave as naturally as possible, stay engaged, leave existing relationships intact and the experiment should be not too time consuming. The last requirement, as mentioned in Chapter 12.1, says that the experiment should not encourage participants to compose knowledge (e.g. interpretations, conclusion drawing etc.) from the experiment. Note that this latter requirement is in slight contrast with the natural behaviour requirement as interpreting is something humans do all day. It is therefore part of natural behaviour to create knowledge. We have chosen to focus on the composing knowledge requirement in the first experiment, but to focus on the natural behaviour in the second and third experiment. That means that we keep the information simple and tangible in the first experiment, but allow for more human interpretation and more complex information sharing in the second and third experiment. As we will see, in experiment two, the interpretation and composing of new knowledge was a problem in the experiment. The participants, however, were communicating face to face, showing their usual communication patterns to a large extent.

In Figure 13.2, the design requirements are listed. With these requirements in mind, brainstorming about different ideas could get started.

Interactive Session with Bart Giethoorn

Bart Giethoorn (Giethoorn, 2016) has his own company in gamification in education[1]. He uses elements from games (both board an computer games) in his teaching to high-school children(physics). With his experience in implementing these gaming elements he was of great help in brainstorming with me about how to meet all the design requirements.

[1](http://www.playbookgamification.nl)

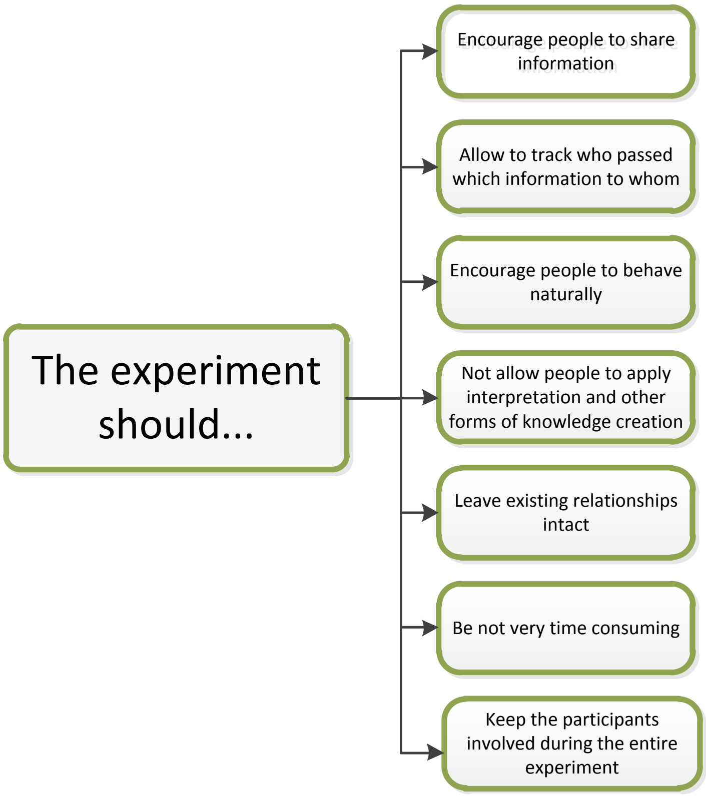
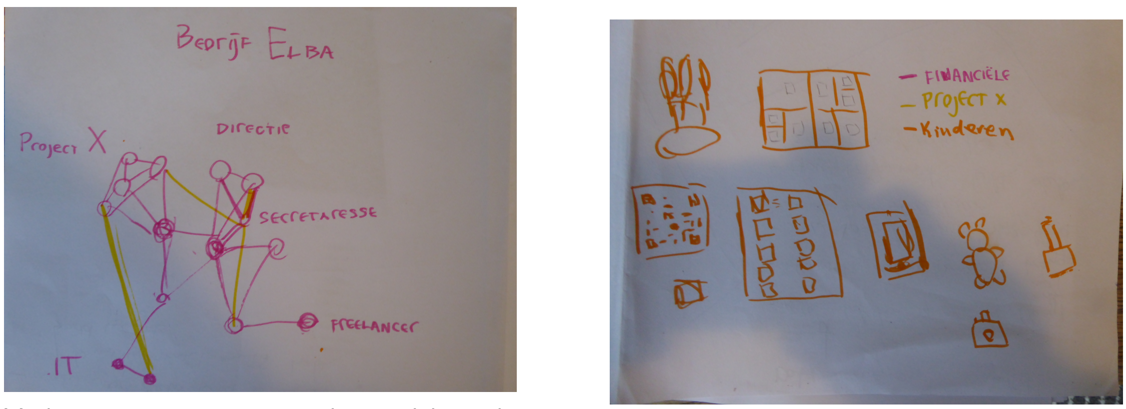


Figure 13.2: The design requirements of the first experiment

During two hours, several drawings and ideas have passed the table, resulting in the set up for the first set of experiments. The basis for the final first idea was created here. The final first idea matches all design criteria. We will reflect on that in Section 13.1.3. In Figure 13.3, some photo's of the creative process can be found. The added value Giethoorn lies in his ability to approach the issue from a systematic design way of thinking. After the session, we had collected enough ideas to finalize the first experiment.



(a) The imaginary company network created during the creative session to help us think about the experiment

(b) Ideas of ways to make the information spread tangible

Figure 13.3: Photo's of creative session

Sample Group of First Experiment

In this stage of the research, not much is known about sample sizes or sample diversity yet. Due to availability reasons, and because so little was known about the group requirements, the students from the science communication minor at the Delft University of Technology were selected as sample group. The group size is about 35 students. The students in this group had just started with the minor, and had known each other for about two weeks at the time of the experiment. Hence, their social network was still fluid.

13.1.3. Design of the First Experiment

In the first experiment, the social network of the minor class described above was mapped, in order to be able to compare the social network ties with the information spread. This was done by asking them to write down 1 to 10 answers to the following question: "Met deze mensen uit de minor heb ik het meeste contact" (With these people from the minor I have the most contact).

Next, they were asked to pass along 2 messages among each other. This message was written down on a piece of paper multiple times. For an example of that, see Figure 14.1. Every time they told the message to somebody else, they could pass half of their messages to the person they told it to. For that, they had half a day (about 4 hours). The goal of the game was to make sure that everyone had at least one note of the messages by the end of the time.

In Appendix C, a schematic overview of the way they had to share information is depicted.

The experiment encourages people to share information because of the gaming element. The second design criterion is that the experiment should allow to track who passed which information to whom. This is possible because the students write their name on the pieces of information before spreading it. This way, it is possible to reconstruct afterwards who has passed what piece of information to whom when all the pieces are collected after the experiment. A drawback is that information might get lost when one or more of the pieces of paper get lost. Hence, theoretically, tracing back who gave what paper to whom is possible, but the data quality might be poor as some of the notes might get lost.

Thirdly, as the experiment is performed in a regular environment for the students, they are encouraged to behave a natural as possible. But as we have discussed before, the fact that they share information through notes rather than by talking to each other, implies that, although the behaviour is natural, the behaviour is not very close to the natural behaviour of information sharing in a network. The fourth requirement, not allowing the participants to apply interpretation to the experimental results worked very well, as the sharing part existed solely on a physical exchange of papers. Hence, no interpretation at all was encouraged.

The existing relationships were not subject of this experiment other than mapping their social network. Furthermore, as the experiment took place during classes, the experiment was not more time consuming than desirable.

The last requirement, keeping the participants involved, is a little bit harder in this experiment, as the participants also have to behave naturally in the meantime (e.g. attend a class). However, as they can be involved in the game by others as well, the participants were generally all well-involved in the experiment.

All in all, the natural behaviour and human interpretation requirements are subject of discussion as they are somewhat contradictory. Other than that, the requirements were met in the first experiment.

13.1.4. Adjustments After the First Experiment

After the first round of the first experiment, the minor students were asked about their experiences during the experiment. They pointed out several things that were adjusted in the next round of the experiment. For example, as the rules were unclear to some students at first, the second round used the graphical explanation of Appendix C. We will point out the other adjustments here. The motivation of these adjustments can be found in Section 14.1.2.

- A graphical explanation of the rules was given to the class.
- Students could only share outside the classroom, and no longer inside the classroom.
- The little fact that was represented by the piece of paper was no longer written on the piece of paper.
- Time of sharing needed to be written down.
- There were less shareable papers, which created a shortage in the papers. Hence, not everybody could have a paper of every color any more.
- Some people received a 'pull'-assignment, where they had to search for information themselves.

13.2. The Second Experiment

In this round, we want to add complexity to the experiment. We do that by adding a second round to the experiment; an original one, and a round where the network participating in the network was disturbed. This is done to be able to research the information resiliency of the network. Furthermore, we measured the extent to which the participants like each other, which is one of the seven factors from the theoretical framework. With this measurement, we selected the participants that were removed from the second part of the experiment.

13.2.1. Goal of the Second Experiment

In the second experiment, the aim is to add complexity to the experiment without jeopardising the design requirements. This added complexity should enable us to observe the natural behaviour of sharing information rather than investigating the possibilities of measuring the information sharing as we did in the first experiment. Furthermore, to extend the analogy, an element of disturbance was added to a second run of the same experiment. By observing the differences between the two runs, more insights could be obtained regarding the information resilience in the human collaborative network.

We aim for testing the experiment set up for the third experiment. Hence, the goal is not to make a lot of changes between the second and third experiment, but merely to adjust some shortcomings.

13.2.2. Methodology of the Second Experiment

In this section, we describe the steps taken to evolve from the first set of experiments to the second experiment. To do that, we have decided to move from a natural to a controlled setting. Furthermore, we have spoken with several experts at different parts of the field.

Natural Versus Controlled Experiment

An experiment in a natural setting is preferable in general, especially since natural behaviour is on our design criteria list. However, as the goal of the research is to find the possibilities and boundaries of the analogy, it is of uppermost importance to have traceable information flows that can be used in the analogy setup. Therefore, it was decided to have the experiment in a controlled setting, where we can make sure that the information is traceable. What happened in the first set of experiments is that people lost their pieces of paper, or left class halfway the experiment due to personal reasons. We expect that he quality of the data suffers more from that than it does from moving towards a closed environment. Furthermore, the closed setup allows a repetition of the same experiment with a small disturbance, to investigate the analogy with respect to the information resilience.

Sample Group of Second Experiment

We consider a group size of 7-8 people. This is because this group is small enough to measure what happens, and to make sure everyone is having a relationship with each other. On the other hand, the group is large enough so that one person behaving odd (because of motivation/tiredness/personal mood or other reasons) will not affect the experiment too much. Furthermore, teams are usually not larger than 8-10 people. As the goal is to remain as close to reality as possible, the maximum size of teams is what we use here. Wittenbaum, Hollingshead, and Botero (2004) show in their research that the information sharing in larger groups (7-8 people) is probably more effective, but they also point out that research is not conclusive on that. Over all, 7 to 8 people seem to be a reasonable as well as realistic and effective group size.

Selection of Participants

We select a sample according to our standards set in Section 10.4. 8 students, all members of student association DSV Sint Jansbrug, were selected to take part in the trial run. Because we want to investigate the impact of them liking each other, we wanted to construct a group in which the people more or less know all participants, but with whom they are not necessary close friends. This was done, by selecting people from 3 different groups. 5 of the participants are member of the same group of friends, which has dinner together once a week. The group is much larger, namely about 30 people. These 5 people were selected from the group by their speed of signing in to the experiment. They are of different age, ranging from 21 to 25.

Two of the participants are in another group, going back for 7 years. They have know each other since the start of their studies. They know some of the people from the other group as they might have been joining in committees or became member in the same year. The last participant was not in any of these groups. However, as he had been in the board of the association 2 years ago, most of the participants knew him rather good.

Unstructured Interviews with Experts in the Field

We have spoken to Daantje Derks, Tom Postmes and Raquel Fernandez to obtain more insights on the psychological and sociological side of this experiment. These three experts all have their expertise in a different corner of of this field, and they were asked to judge the experiment from this field of expertise. Furthermore, they were asked about their view on the experiment regarding validity, feasibility and best practices for the experiment.

The research group in Delft misses experience and knowledge on experimental psychology which we need for the experiments, not only for designing on a higher level, but also for collecting best practices on things to do and not to do. These experts were searched through Dutch universities that offer psychology programmes. The following Universities have been approached: Leiden University, Erasmus University, Free university of Amsterdam, University of Amsterdam and Groningen University. We aimed at approaching universities in proximity of Delft University of Technology as it is easier to visit these universities. However, some of the professors approached recommended to contact somebody else. As a result, Groningen University was also approached. As expected, not all universities were willing to help. However, we are very pleased with the help of two experts on experimental psychology (Daantje Derks, assistant professor at the Erasmus graduate school of social sciences and the humanities, and Tom Postmes, professor in social psychology at the University of Groningen), and one on human communication patterns (and linguistics) (Raquel Fernandez, associate professor at the Institute for Logic, Language & Computation of the University of Amsterdam). As the interviews were exploratory in nature, and the focus and field of expertise of the experts was not entirely clear before the conversations, the interviews were unstructured. For that reason, they also have not been coded, as the aim was to get some advices, and not to find patterns or to compare two experts.

The first expert spoken to was Daantje Derks, assistant professor at the Erasmus graduate school of social sciences and the humanities. She pointed out that information that has either a high intensity, or a high value leads to emotional responses (Duprez, Christophe, Rimé, Congard, & Antoine, 2015). That should be kept in mind when designing an experiment. She made a distinction between task driven assignments and emotional assignments. They trigger different behaviour, and the emotional assignments trigger the participants to use their social network more, as they only want to tell emotionally heavy things (so intense and/or high value) to the people they trust. By asking them to fulfil emotional tasks, the emotional valence plays a role. An example of doing that is to create a conflict of interest or conflicting assignments. However, we will see that these emotional assignments are not part of the third experiment. This is because adding this layer of complexity made the experiment in its current shape too complex. The conflicting assignments tested in assignment two did create a lot of chaos and interpretation, which made it impossible for those two participants to participate in the experiment properly. For that reason, the third experiment focusses on task driven assignments.

Furthermore, she advised us to take people on one hierarchical level (so no managers over the others) as this influences the behaviour, and unnecessarily complicates the experiment. This was also pointed out in Chapter 11, when discussing power relations. We proposed giving the participants a task that they would have to fulfil without it being the main goal of the experiment. This to encourage the students to behave naturally. She agreed with this, and mentioned that usually something as simple as building a Lego shape or building something from blocks is enough to let people forget about the artificial setting.

As we have seen in Chapter 11, Derks also proposed liking as a control variable. Because participants feel uncomfortable quickly, it is important to know how much they like each other. Lastly, she gave some best practices. She advised to keep the experiment short (about 10 minutes), as participants easily get distracted after about 10 minutes. Also, show them how much time they have left. Furthermore, we should never award the participants as this stimulates competitive behaviour.

The second conversation was with Tom Postmes, professor in social psychology at the University of Groningen. He focused very much on making sure to measure what you want to measure. His main concern was that there is no way to make sure that the behaviour in the controlled experiment still has anything to do with natural behaviour. We think he has a valid point here, but the behaviour of the people in the experiment is most certainly more natural than the behaviour of the ICN simulator. Furthermore, the behaviour of the participants is natural, but by leaving details of a natural setting out (such as the amount of attention paid, or other people from the network interfering), their behaviour might not reflect their behaviour in a natural setting. This is not necessarily a problem for building up an analogy between an ICN and a human collaborative network. Secondly, Postmes pointed out that the data sets will be way too small to make any statistical claims. He called the research a search for anecdotal evidence.

The third conversation was with Raquel Fernandez, associate professor at the institute for logic, language & computation of the University of Amsterdam. She knows a lot on the modelling part of the research. Furthermore, she could provide me with data sets to analyse, if we would decide to move in the direction of analysing online social communication instead of face to face communication. However, we have decided to use face to face communication.

13.2.3. Design of the Second Experiment

In literature, we have distinguished seven factors of interest that influence the information sharing in human collaborative networks. One of them, liking, will be measured in this experiment. We pick this factor because it was suggested by our interviewed expert (Derks, 2017b). Furthermore, we attempt to measure the information resilience in the human collaborative network by measuring the effects when we remove some people from the experiment in a second round.

Before the Experiment

Before the experiment, a questionnaire was filled in by the 8 participants. In this questionnaire, they had to answer 4 questions on a 9 point Likert scale about each participant. These questions were:

1. I have great confidence in [person]'s good judgement.
2. I would vote for [person] in a class or group election.
3. [Person] is one of the most likeable people I know.

4. [Person] is the sort of person whom I myself would like to be.

These 4 questions are picked from Rubin's scale of liking (Rubin, 1970). In his article, he states 13 questions on a 9 point Likert scale that together measure how much the person of interest is liked by the person filling in the questionnaire. As filling in 13 questions per person is not doable in a group of 8 people, we have decided to pick the 4 most appropriate questions according to our judgement. Note here that it is meant to find an indication of liking, as we want to study the impact of liking on the resilience of the system.

The answers to these questions were summed up over all participants and all questions to give every participant an individual score. Hence, the personal liking score for person k, P_k is equal to

$$P_k = \sum_{i \in \text{ participants}} \sum_{j \in \text{ questions about k}} \text{Score}_{i,j}.$$

The 3 most liked persons (hence, the 3 people with the highest overall score) were picked as the people that would be excluded from the second part of the experiment. One of the other seven factors from the theoretical scope could have been chosen as well, but as this experiment was merely a test run for the third experiment, we have only measured one factor.

Design of the First Part of the Experiment

The experiment consisted of 4 rounds. In the first round, the participants had to build a marble track. The following rules were given to them. As it was originally explained in Dutch, the Dutch translation can be found directly after the English rule.

1. Build a marble track with the building materials that you can find on this table. (Bouw samen een knikkerbaan met de materialen op deze tafel)

2. 10 minutes time, after that, the marble track is tested. (10 minuten de tijd, daarna testen we de knikkerbaan.)

3. Hidden, personal assignment: (Verborgen, persoonlijke opdracht:)

   You can talk about your asignment with others, but not more than one person at the same time. (Je mag het er wel over hebben met anderen, maar niet met meer dan 1 persoon tegelijk.)

   Don't let anyone else read your assignment. Talking about it is allowed (if you wish in the hallway)(Laat je kaartje aan niemand anders lezen. Erover praten mag wel (evt. op de gang).)

4. Try to fulfill both the group assignment (build a marble track) as the personal assignment. (Probeer zowel de groepsopdracht (bouw een knikkerbaan) als je persoonlijke opdracht tot een goed einde te brengen.)

In the first round, 8 hidden assignments were spread in envelopes with the names of the participants written down on it. The name was only written down to prevent them from having (almost) the same assignment twice. The 8 hidden assignments can be found in Appendix D. An example of an assignment is: "Zorg ervoor dat de knikkerbaan breder wordt dan hoog. Zorg ervoor dat maximaal 3 andere mensen te weten komen dat dit jouw opdracht is." (Make sure that the marble track is wider than high. Make sure a maximum of three people get to know that this is your assignment) These assignments are task driven, as this keeps the participants from feeling uncomfortable. This way, the experiment stays manageable. It would be desirable to extend the experiments with intense assignments (emotionally driven or with a high value), but for the course of this experiment, that is too complex. The assignments could be interchanged with other task driven assignments.

However, it is important to keep in mind that the assignments should be formulated in such a way that it is in the participants' interest to share his assignment with part of the group, but not with all of them. After the first experiment, the participants all have filled in a data collection form. We come back to this in the next section (13.2.3).

Design of the Second Part of the Experiment

In the second experiment, there were 2 more rules added to the 4 rules above. First of all, 3 people were excluded from participation. This was done to be able to measure the impact of a disturbance in the information sharing in the network (the information resilience). The participants that were excluded were invited to watch the second experiment, and to use their observational skills to find out about the other personal assignments. They were, however, explicitly told to not interfere (talk, ask questions, help building) with the marble track.

The second rule added was that no two building materials that were following up each other in the previous marble track could follow up again. This was done to make sure they would have to build another marble track than they did in the first experiment.

Since building the marble track was only meant to distract the participants and to let them behave as normal as possible, we explicitly did not add too many rules to the building procedure. As we will see later, this was not received positively by all members. Again, the assignments can be found in the Appendix D.

After the second experiment, there was another data collection form, followed up by an evaluation form.

Data Collection Forms

The data collection forms after both parts of the second experiment were the same. First of all, there were questions about their personal assignment. They had to state whether they fulfilled the assignment, and what went wrong if they had not. Also, they had to state who they had told about their personal assignment, and who else had figured out their assignment (and how they did that).

Next, they had to answer questions about the assignments of the other participants. For each participant they had to answer the following 3 questions:

1. Do you know what [person]'s personal assignment is?
2. If yes, what is it?
3. If yes, how did you find out?

With this information, we are able to construct the flows of information. This is needed to be able to compare the flows of information in the human collaborative network with those in an ICN.

Evaluation

After both data collection forms and experiments, the participants were asked to fill in an evaluation form. This form will help us in reshaping the experiment into its final form. In the evaluation, the participants were asked to agree to some statements on a 7 point Likert scale. After that, 2 open questions were asked to give the participants the opportunity to share some feelings or suggestions for improvement.

13.2.4. Design Requirements

Before continuing to the third experiment, we reflect upon the design requirements, to make sure that the experiment is still in line with our objectives.

The first objective is that the experiment should encourage people to share information, which is clearly met with this design. Secondly the experiment should allow to track who passed which information to whom. The data collection forms detect the way of sharing. This is not flawless, as the participants might not recognize the information sharing as such. Thirdly, the experiment should encourage people to behave naturally. We try to achieve this by distracting the participants from their personal assignment by letting them build a marble track.

The fourth requirement, not allowing the participants to interpret the information, is much harder to establish. The results will show that this did happen a lot. By keeping the questions in the data collection form afterwards as clean as possible, and by pointing out that the participants should answer what they know instead of what they think, we attempt to keep the interpretation factor as small as possible in the third experiment.

The other factors, namely keeping the existing relationships in tact, be not too time consuming, and keeping the participants involved, are expected to be no problem at all in this experiment. The participants are not asked to disagree or anything like that, and they are being held busy during the course of one hour.

13.3. The Third Experiment

In the third and last experiment, the second experiment is adjusted based on the feedback from Derks (2017a), and the observations made in the second experiment. Furthermore the questionnaire filled in before the experiment was expanded, to measure all the factors that were taken into account in the theoretical framework.

13.3.1. Goal of the Third Experiment

The third and final experiment has the goal to obtain data on the seven factors of influence in combination with the information sharing data. Furthermore, we want to obtain data on the influence of a disturbance in the social network. With this data, we want to add some more detail to the analogy. This data can help us in thinking about the information sharing in a human collaborative network with the analogy as a starting point. An example of that is thinking about the notion of trust from an ICN perspective in a human collaborative network. Furthermore, it enhances our insights in the relation between information resilience in human collaborative networks and ICNs. The last adjustment with respect to the second experiment is that conflicting assignments caused a lot of agitation and confusion, and are therefore excluded from the experiment this time.

13.3.2. Methodology of the Third Experiment

Two main changes have been made in the third experiment. We have selected a new sample and the other 6 factors have been added to the questionnaire beforehand. Furthermore, we have made some minor adjustments with respect to the second experiment.

Like we did in the second experiment, we select questions from existing questionnaires, for reasons of validity. To keep the questionnaire manageable in size, we select some questions from all these questionnaires, and combine them into one questionnaire. Liking remains part of the questionnaire as it was in the second experiment. The other factors will be discussed below.

Social Network

To map the social network of the participants, they are asked with whom they have the most contact in their private life. It is compulsory to indicate two people, but more people can be added manually.

Reciprocity

For measuring the reciprocity, we ask for a reciprocity feeling towards the whole group, rather than to individuals. This is because, as stated in the theoretical framework, the group establishes a norm of reciprocity, which makes it hard to measure differences towards individuals in the group, especially since the questions are very limited. Perugini, Gallucci, Presaghi, and Ercolani (2003) have established a questionnaire that measures this personal norm of reciprocity. We make slight differences to this questionnaire in two ways. First of all, the number of questions is too large (27) to fit in this experiment. Secondly, the authors focus on a very general norm of reciprocity, whereas we will focus on this norm within this group. The questions will be adjusted such that this becomes clear.

The original 27 questions can be categorised into two groups, namely questions about behaviour and questions about belief. The latter is less of interest for us, so we leave these questions out. That leaves us with 18 questions, which is still slightly too many. Within these questions, a distinction is being made between positive and negative reciprocity. This difference is made because some people tend to act heavily on positive behaviour, while others are more sensitive towards negative reactions and behaviour.

Table 13.1: Questions asked on reciprocity

Question nr.	**Positive reciprocity**
1	If someone in this group does a favour for me, I am ready to return it [2]
2	I'm ready to do a boring job to return previous help of someone in this group
3	When someone in this group does me a favour, I feel committed to repay him/her
4	If someone in this group lends me money as a favour, I feel I should give him/her back something more than what is strictly due
5	I go out of my way to help somebody in this group who has been kind to me before
	Negative reciprocity
6	I am kind and nice if others of this group behave well with me, otherwise it's tit-for-tat
7	If somebody in this group puts me in a difficult position, I will do the same to him/her
8	If somebody in this group offends me, I will offend him/her back
9	If somebody in this group is impolite to me, I become impolite
10	The way I treat others in this group depends much on how they treat me

We take 5 statements from both aspects. This results in the questions of Table 13.1.

Trust

Rempel et al. (1985) has created a questionnaire to measure the three elements of trust. This questionnaire focuses on close relationships, which is not necessarily the case in our experiment setting. Unfortunately, the relationships described here are so close, that they can not be used in our experiment. On the other hand, McAllister (1995) have created a questionnaire to measure trust on colleagues. We have categorized them to the three aspects of trust (Rempel et al., 1985). The results of this can be found in Table E.1 of Appendix E. We pick one question of each aspect, on the basis of relevance to our experiment. The questions selected are:

1. We would both feel a sense of loss if one of us was transferred and we could no longer work together. (dependability)
2. If I shared my problems with this person, I know that (s)he would respond constructively and caringly. (predictability)
3. Given this person's track record, I see no reason to doubt her/his competence and preparation for the job. (faith)

Note that we will have to adjust the questions to the setting of the experiment.

Willingness and Eagerness of Sharing

van den Hooff and Hendrix (2005) have composed a questionnaire that measure the willingness and eagerness of knowledge sharing. The questions are asked on a 5 point Likert scale. A table with the questions can be found in Table 13.2. These questions are the literate questions as proposed by van den Hooff and Hendrix. For the experiment itself, the questions will be tweaked in formulation in such a way that they related directly to the people they perform the experiment with.

Sample Group of Third Experiment

As there is no reason to rethink the group size, again 7-8 participants were selected for the experiment. They work together in two project groups. They know each other from the master, and work together in this project group since 10 weeks. The two project groups do not work together on a regular basis. However, one project group was too small for our aim, so we have decided to combine two groups. This is not ideal, but for the aim of building the analogy it suffices.

[2] We are aware this is not proper English. However, the question has been copied directly from (Perugini et al., 2003), and has therefore also been asked this way to the participants.

Table 13.2: Questions for eagerness and willingness of knowledge sharing

Eagerness	
1	I tell my colleagues about things that I consider important - even if they don't ask me for it.
2	My colleagues can learn a lot from me.
3	I tell my colleagues what I know more often than they tell me what they know.
4	I keep my colleagues informed of what I know, even if they don't ask me about it.
Willingness	
1	I like helping my colleagues.
2	I like being appreciated for what I know or what I can do.
3	I like sharing my knowledge and capabilities with my colleagues.
4	I can learn a lot from my colleagues.
5	I try to share my knowledge and ideas with colleagues if I can help them.

13.3.3. Design of Third Experiment

The rules of the experiment were the same as the rules in the second experiment, with one small exception. As caching (passing information that was not originally yours) was not very present in the second experiment, it was pointed out to the students that this was a possibility in the third experiment.

The assignments were like the assignments in the second example, and can be found in Appendix H. As not much changed with respect to the second experiment, we do not reflect on the design requirements here again. The only note made is that in the third experiment, some more attention was given in the explanation of the experiment to the caching and interpretation of the information, to prevent the participants from drawing there own conclusions.

The factors measured as above can all be used to determine who will be removed from the experiment in the second part. We pick one of these factors here, but any other could have been picked as well. We pick centrality measures[3] because these measures lie closest to the analysis that is common in ICN. This is quite arbitrarily, and it is also therefore that we will come back to it in the discussion section (Section 16).

13.4. Data Analysis

In the experiments, we collect data in which we track who has shared what information with whom. This data can be used to imitate in the ICN simulator Icarus. In doing this, the wish is to set as little restrictions to Icarus as possible. The more restrictions are set, the more artificial the analogy becomes.

The data obtained in the experiment can be incorporated in the simulator by reconstructing the social network, the sources and other things in the ICN simulator. Next, the data paths are constructed. These paths depict one stream of information (the messages or information about personal assignments spread) from the source to the last user in that stream. By entering the end users of each stream of information to Icarus, the simulator will itself construct a path along which it sends the information to these end users.

How this is done, and how the two problems that occurred can be solved, is discussed in Chapter 14.

[3]The centrality measures considered are Eigenvector centrality, betweenness centrality and closeness centrality. For more information, see Section 11.6.1.

14

Results

> Not everything that can be counted counts, and not everything that counts can be counted.
>
> A. Einstein

In this chapter, the results of the three experiments are depicted. Per experiment, we first show the results, then the interpretation of the results, and finally the link to the simulator. For the third experiment, we also discuss the results of the questionnaire beforehand and the interpretation of these results.

14.1. First Experiment

The aim of the first experiment is to find out if we can develop a method of tracing information spread in human collaborative networks. Furthermore, this first data set can be used to enter the information spread into the simulator, to see what kind of adjustments need to be made in the simulator in order to be able to compare the information spread in an ICN with a human collaborative network.

Firstly, the results of the experiment itself are shown, and interpreted. Secondly, the links to the simulator are discussed. The first experiment consisted of two runs, in between of which small adjustments were made. It is therefore that we call it the first set of experiments, consisting of a first and a second part.

14.1.1. Experiment Results

For the first set of experiments, the social network of the science communication minor class was mapped. Furthermore, the participants were asked twice to pass around some notes, that allows us to map the information spread afterwards. Some examples of that are shown in Figure 14.1. The social network found is shown in figure 14.2a

Two different experiments were performed. The ties over which the information has spread are being investigated both times. In the first experiment, two different notes were passed. In the third experiment, however, 3 coloured notes were passed along. The results of the latter experiment can be found in Figure 14.2b.

We see that the traces of the notes are indeed measurable this way.

14.1.2. Interpretation of Results

The first experiments are exploratory, and the results are exploratory as well. We mention a number of interpretations, as they directly or indirectly contribute to the analogy and/or insights to the experiments.

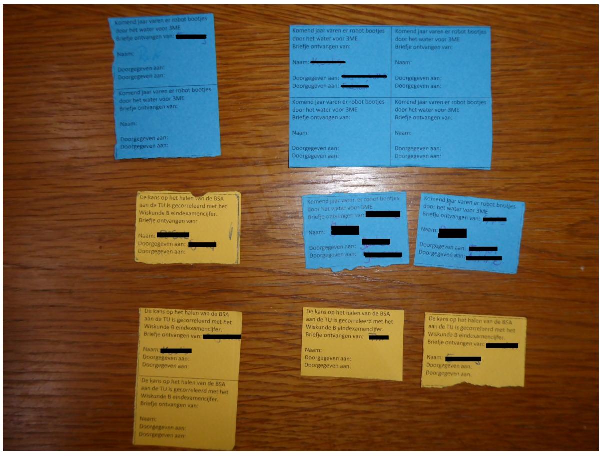
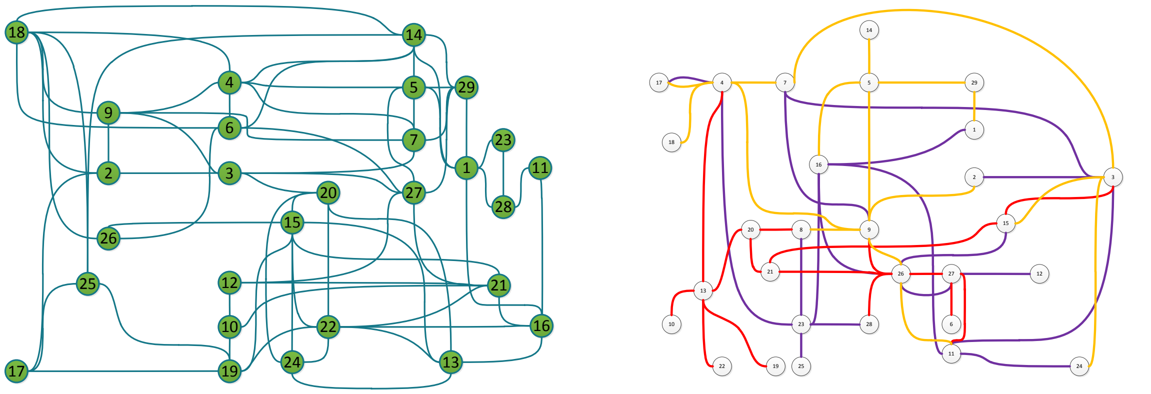


Figure 14.1: The notes the participants shared amoung each other



(a) The social network of the minor class

(b) The network of information spread in the minor class

Figure 14.2: The social network and its corresponding information spread

What we see is a social network with many ties, whereas the expectation was that the students would answer on average 2 to 3 names to be important in their social network. In reality however, most of the students would answer 5-9 people in their social network. This way, the network becomes unusable as there are too many ties that are redundant in reality. Also, as the students hardly knew each other by that time, the ties were still rather fluid. By that, we mean that the students would easily use other ties than the ones they had written down in the experiment since their network was not really established yet.

Furthermore, we see that there are cliques[1] present in the social network. Usually, these cliques indicate a very strong relationship between the members. Typically, these are groups of good friends. Here however, we see that the cliques are formed from the project groups they work in together, as the network is only existing for 1,5 weeks at this point. This does not necessarily mean strong friendship at this point yet.

If we turn our attention to the experiments, we note a couple of remarkable things. First of all, in both networks, the different notes are spread through a different part of the network. See for example the yellow and the red information items of Figure 14.2b. This is probably quite realistic, as different information items spread differently through a network, and are not necessarily of added value to all people in the network.

[1]a clique is a part of the network where all nodes are directly connected to all other nodes. An example of that is participants 1,5,14 and 29 that form a clique.

More problematic, however, is that in the first experiment, the information spread of one type spread is disjunct[2]. the information spread among 6 participants are not connected to the rest of the network. This is physically impossible, as it would mean that the two groups independently received the information. However, the information was only given to 1 person at the beginning. Hence, there is a problem with the data quality of the experiment here. This problem has two natures. First of all, not all students returned their notes by the end of the experiment. They either lost it, forgot to return it, or they might have lost their interest to cooperate. Secondly, the students needed to fill in their name on the papers in a certain order. This might have gone wrong also, resulting in pollution of the data.

The latter is most probably an important cause of the low data quality. This is because the participants expressed that the rules of the sharing were unclear to them in the first experiment. Some of them found out during the evaluation that they had not followed the rules of the experiment. The students expressed that it would be better to explain the rules of the game either visually or with a role play. This advise was (with good results) followed in the second run of the experiment, where the visuals of Figure C were shown to the students.

In the first experiment, the students were excited to share the information with each other, and according to the teacher of that class, they were done sharing within 15 minutes. It was said to the participants that disturbing the lecture was not allowed, but this turned out not to limit the participants from spreading the information during the lecture. In the second run it was therefore prohibited to share information in the classroom.

In the second experiment of this first run of experiments, some adjustments were made with regard to the first experiment. The rules were explained differently and a rule was added that sharing inside the classroom was not allowed. Furthermore, as the experiment should be about information sharing, and not about passing notes around, a little fact that belonged to the note were added, and when handing over a piece of the sharing paper, the note itself should vocally be transferred. By the end of the experiment, the participants had to write at the back of their paper what information was represented by each piece of paper. Also, by reducing the amount of shareable papers by 50%, there no longer was enough paper for everyone to participate. Lastly, the participants had to write down the time of sharing, and some participants received a 'pull' assignment, in which they had to search for certain information. These two aspects have not played any further role in the remainder of the experiments as the time part was chosen not to be part of the analogy. Also the pulling part of the assignment was difficult to implement, and felt less natural than the pushing part of information sharing. Here, we have found an obvious difference between ICN and human collaborative networks that cannot fully be accounted for in the analogy.

In the second experiment, the data quality has increased largely. The rules of the experiment were clear to everyone, and the importance of returning the papers was made clear to the participants. There are no outstanding events or anomalies here, so the data obtained in the second experiment was used to make the link to the simulator.

In the central evaluation of the experiment, the participants were asked if they thought the information sharing in this experiment was close to their natural behaviour. One student made the just remark that each individual makes judgement calls on whether or not to share all day. When this person decides that sharing certain information is either nice to tell, or useful, he or she is likely to share the information. If the information does not meet these requirements, he or she would not share. In the experiment, the facts representing the notes were not useful to know. Furthermore, whether one thinks they are nice to share depends to a large extent to the personal preferences of both the receiver and sender of the note.

Hence, in retrospect, we need to conclude that the experiment did not entirely meet all the design criteria. The participants were not entirely encouraged to behave naturally. This problem is acknowledged, but it is hard to reach full natural behaviour while still being able to measure something.

[2]A disjunct network is a network where two or more sets of people can be distinguished, and where no links exist between these two sets of people.

14.1.3. Link to Simulator

The stream of information in the ICN is not equal to the stream of information in the human network. This is due to two problems, which we will have to adjust. We will discuss the two problems in more detail, and also discuss their connection, and a solution to the problems.

Router Versus End User

As said before, the ICN simulator does not allow one node to be a router and an end user at the same time. This is mainly a shortage of the simulator, as it can happen in ICNs that a mobile device is a router at one moment, but is an end user in the next moment.

In the human network, there is a clear possibility of a person being only a receiver of one kind of information (hence an end user), but being a router for another stream of information (hence, the information is passed through to others). By introducing a dummy person, this issue can be easily solved. In essence the person that acts in two ways (router and end user) is split into two pieces in the ICN simulator: one router, and one end user. The end user is only added to the corresponding router. By combining the two again in the final analyses, one person can be represented as an end user and a router at the same time in the ICN.

Shortest Path Issues

Another problem is that the ICN simulator will always choose the shortest path between the source and the end user to pass information. If we take a look at Figure 14.2b, participant 26 has send the 'red' information to participant 11 through participant 27. However, they have shared 'yellow' information already directly. The ICN simulator will now choose the shortest path from participant 26 to 11, which is the direct path.

Now, as we want the results to be equal, we need to tweak the ICN simulator. Originally, the simulator will assume all lengths of the paths to be equal (hence, the hop count is the way of counting the shortest path). This is where the adjustment is being made to the simulator.

From the theory of social network analysis, we know that information streams will follow different paths for different types of information. The gossips will spread through other people and ties in a company than the latest financial developments. If we consider the three streams of information to be different kinds of information, we can add a weight to the ties, depending on the type of information that is being spread. This weight could be perceived as some kind of path length. So by decreasing the path lengths of the red lines when the red information is being considered, there is a larger chance that the red information will indeed follow the red ties.

In the first experiment, we have found the possibilities to measure the spread of information, and the possibility of linking this data to the ICN simulator Icarus. In the second experiment, we will enhance the complexity of the experiment, and measure actual information sharing rather that note passing.

14.2. Second Experiment

In this section, we first state the results of the questionnaire beforehand as well as the information sharing experiment and the resilience experiment. Secondly, this is interpreted and the results of the evaluation form are discussed. Note here that the link to the simulator is not explicitly made since the results do not give any reason to doubt the possibilities to link it to the simulator. The third experiment will again be linked to the simulator.

14.2.1. Experiment Results

In the second experiment, the participants filled in questions about liking the other participants beforehand. The participants that are being liked best by the others were excluded from the experiment the second time. Four questions about each participant are answered on a 9 point Likert scale. The sum of the questions (hence, maximum 36 points) is taken as the liking score. In Table 14.1, the results are depicted.

Table 14.1: Aggregated scores on the liking questions

Points given \ Points received	A	B	C	D	E	F	G	H
A	0	26	28	30	23	17	21	30
B	17	0	18	21	20	27	23	26
C	21	0	0	23	22	20	0	27
D	22	24	26	0	23	24	23	25
E	0	25	0	30	0	31	25	26
F	24	24	20	26	28	0	23	22
G	29	30	0	32	32	32	0	34
H	17	22	21	32	31	11	27	0
Points received total	130	151	113	194	179	162	142	190

If two participants do not know each other, they award 0 points on liking. Participant D,E and H received most points, and were therefore excluded from the experiment the second time.

In Table 14.2, the results of the first experiment can be found. In this table, it is stated who has shared his assignment with whom. The table should be read horizontally. Furthermore, the assignment number is written on the diagonal. This aligns with the numbers of the assignments as given before. To give an example, participant A (with assignment 6) has told about his assignment to participant D and G. Furthermore, participant C has told about her assignment to participant A, and participant F thinks that participant A has overheard him.

In Table 14.3, it is stated who knows whose assignment. If there would be no difference in perception, memory and quality of the filled out data collection forms, the second table should be a transpose[3] of the first table. However, as is indicated in red in the two tables, they do not always overlap. Observations, explanations and ideas about how that can be the case are discussed in section 14.2.2.

In Appendix G.1, and G.2, you will find the same information about the second assignment. Note here that, as three people were removed from the group, the first table is much smaller. However, in the second table, the three people aside were allowed to observe, so they are part of the table again.

As the caching mechanism in this experiment was hardly present, the link to ICN simulator has not been made here. This is because the link would not be interesting at all.

[3]A matrix (or table) transpose is a reflection over the diagonal axes of the original matrix (or table)

Table 14.2: Table with answers about their personal assignment

Who \ To whom told	A	B	C	D	E	F	G	H	Succes?
A	6			yes			yes		yes
B		2						yes	no, too much focus on marble track
C	yes		8	yes			yes	heard	more or less
D			yes	5	yes	heard	heard	heard	no, too busy, too much counteraction
E					3				no, not enough questioning
F	heard	heard	heard	heard	heard	4	heard	yes	yes, obvious oposite goal than participant D
G							7		yes
H		yes				yes		1	yes

Table 14.3: Table with answers about the personal assignments of the others

Who \ About whom known	A	B	C	D	E	F	G	H
A	x	no	no	no	no	yes	no	no
B	no	x	no	yes, told	wrong	yes	no	no
C	no	no	x	yes, told	no	yes, by counteraction participant D	no	no
D	yes, told	no	yes, told	x	no	yes, squabble	no	wrong
E	no	no	no	yes, told	x	yes, squabble	no	wrong
F	no	no	no	yes, opposite to myself	no	x	no	yes, told
G	yes, told	no	no	yes, asked	no	yes, obvious	x	wrong
H	no	yes	no	yes, asked	no	yes	no	x

14.2.2. Interpretation of Results

Here we will discuss some remarks. These were either our observations during the experiment, or things that did not go as expected.

Human Interpretation

Very often it happened that people thought they knew what somebody else's assignment was, but were, in fact, wrong. This is largely due to human interpretation. Let us give an example of that.

In the second experiment, one person accidentally finished his assignment after 1 minute, and had 9 minutes left to just build the marble track. Due to his own preferences and ideas of a nice marble track, he was very much in favour of using a certain building material everywhere it was possible. Many other participants therefore (incorrectly) thought that it was his assignment to use this building material as often as possible in the marble track. In explaining how they found out about this person's assignment, one participant stated: "hij was er alles behalve ondoorzichtig in".

It is interesting to see how this works. In the next experiment we have to choose to either block it out as we have now found this to be a boundary of the analogy, or we have to include it, and find a way to deal with it.

Engagement

In this experiment, the participants were very engaged and involved. As can be seen also in the evaluation forms, they liked doing it very much. The marble tracks did not work entirely in all cases, but the results were impressive for only 10 minutes of work.

In this group, there were no two people disliking each other. Not everyone knew all the others, but there were no signs of possible negative feelings noticed. Also, the evaluation form afterwards does not indicate negative feelings as we will see in question 16 of this questionnaire.

Missing Questions

After the experiment, I found out that one important question was missing in the data collection forms. What is important to also track, is if people have told others about an assignment that is not theirs (caching). This should be added next time.

Level of Assignments

Some more attention should be given to investigation the level of assignments. In the evaluation questions asked there, people did not all agree on whether it was too hard or too easy. This should be checked again with the actual assignment people had next to it.

14.2.3. Link to Simulator

We want to link the data to the simulator, but we run into a problem. People in this experiment tend not to cache. By that, we mean that people in the experiment did not continue to talk to others about 3rd people's assignment. One exception to that was participant A helping out somebody else with the experiment. As a result, people thought that it was his assignment. The lack of caching is a problem as it is a key feature of ICN to cache. As a result, the strength of the analogy increases when caching is a part of the experiment. This is also the reason that we have not linked these results to the simulator, as the results would be obvious and not very interesting.

This could be solved by mentioning in the beginning that participants can help others out with their assignment if they wish to. As we will see, that is enough stimulation to let the participants talk to each other.

14.2.4. Evaluation

In Table F.1 of Appendix F you will find the questions of evaluation (in Dutch), together with the score given by each individual. All questions were answered on a 7 point Likert scale. In general, we see that people enjoyed the experiment (question 2), thought it was very clear what was expected from them (question 4), and had enough time for all elements in the experiment (question 1). Some people indicate that they had a hard time concentrating on the personal assignment (question 3), but others had no problem with this at all. The variety in assignments also lead to a variety in perceived difficulty in the assignments (question 5 and 6).

We see that most of the participants had enough motivation to try to fulfil their assignment (question 7). All but one participant thought the explanation beforehand was clear (question 9). This one participant had trouble with the marble track building without further requirements. Where all participants in both the second and third experiment were ready to start building immediately, this participant wanted more explanation about what he exactly had to build. In the open questions he elaborated on this, saying that engineers in general need more guidelines for building a marble track. In total, however, 15 engineering participants have participated in experiment 2 or 3, an only one person commented on this.

The questionnaire that the participants filled in before the experiment was experienced as uncomfortable (question 12). Furthermore, according to most of the participants, the questionnaire beforehand did not influence their behaviour during the experiment (question 15). One participant however, mentioned that he used the questionnaire beforehand to decide who to talk to in order to fulfil his personal assignment.

14.2.5. Interpretation of Evaluation

The perceived difficulty differences in the assignments (question 5 and 6) could be due to 3 things. First of all, the assignments were designed in such a way that an element of luck was built in. If a participant asks the right person in the first attempt, the assignment can be successful immediately. Secondly, there was a potential difficulty difference in the assignments. Thirdly, how difficult an assignment is perceived, might also be depended on the individual. An assignment can be perceived more difficult if the relation with the rest of the participants is tenser, or if the participant feels shy to participate actively. Many more personal reasons could be given that influence the perceived difficulty of the assignment. all in all, the results of the questionnaire do not give ground to change the experiment heavily.

In the second experiment, we have found enough ground to perform a third experiment with small adjustments to the second experiment. In this second experiment, we have added complexity to the experiment with respect to the tasks the participants have to complete as well as the complexity of the information that should be shared. In the third experiment, we fine tune the caching behaviour and some other details discussed above.

14.3. Third Experiment

In the third experiment, some details were changed with regard to the second experiment. In the explanation of the experiment, some more attention was given to the possibility to cache information. Also, the seven factors of our theoretical framework are indicated in a questionnaire before the start of the experiment. Furthermore, the group of participants are working together in two project teams. In this section, we first pay attention to the social network. Secondly, the results from the questionnaire beforehand are given. Thirdly, the results from the experiment itself are given. Lastly, the results are again linked to the simulator in Section 14.3.5.

14.3.1. Results from Questionnaire Beforehand

We will go over the factors one by one, stating the results of indicating them before pointing out some observations. After that, some more general observations, that link some of the factors, are given.

Social Network

In Table 14.4, we find the results of the social network as indicated by the participants. If we map these results onto a network, we obtain the result in Figure 14.3. Note that participant H has mentioned himself as the second person he has most contact with. We remove this link from the network. The reason for this is that the two questions were compulsory, and not answering anything was not possible. Most likely, the participant did not want to answer a second person.

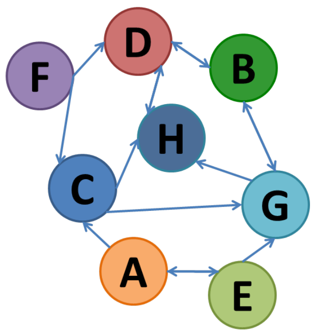


Figure 14.3: The social network of the participants

Table 14.4: Social network of participants

Person	Most contact in private life	Second most contact in private life
A	E	C
B	G	D
C	H	G
D	H	B
E	G	A
F	C	D
G	B	H
H	D	H

Centrality

In Table 14.5 the three main centrality measures are depicted, as well as the degree[4] of each node in the network. It is clear that participant C and G are the most central in the network. Only when considering Eigenvector centrality[5], participant H also scores among the highest. As participant H was not present in the final experiment, we have decided to nominate participant C and G to be removed from the experiment the second run.

Table 14.5: The centrality of the participants

Person	betweenness centrality	closeness centrality	eigenvector centrality	degree
A	0.048	0.500	0.161	2
B	0.048	0.538	0.276	2
C	**0.310**	**0.700**	**0.478**	4
D	0.095	0.538	0.338	3
E	0.048	0.500	0.236	2
F	0.048	0.538	0.291	2
G	**0.310**	**0.700**	**0.443**	4
H	0.095	0.636	**0.467**	3

Attitude Towards Sharing

In the second and third column of Table 14.6, the eagerness and willingness to share per individual are given. Recall that, as proposed by van den Hooff and Hendrix, four questions were asked for the eagerness and five to measure the willingness to share. The questions were answered on a 7 point Likert scale. Since the number of questions are different, we need to compensate for that in order to

[4] the degree of a node is the number of links that are connected to it. Node E, for example, has degree two as it has a link to both A and G but to nobody else.

[5] The measure that can be interpreted as that being liked by a person who is very popular himself makes you more likeable

be able to compare the two. This is done by averaging the scores, as shown in the fourth and fifth column of Table 14.6.

Table 14.6: The eagerness and willingness to share per individual

Respondent	Eagerness to share	Willingness to share	Average eagerness score per question	Average willingness score per question
A	14	29	3.5	5.8
B	17	28	4.25	5.6
C	19	24	4.75	4.8
D	19	31	4.75	6.2
E	18	24	4.5	4.8
F	18	28	4.5	5.6
G	22	32	5.5	6.4
H	21	29	5.25	5.8

In the results, we see that all participants are more willing to share information than they are eager to do so. Furthermore, participant G scores highest in both eagerness and willingness to share. Participant A stands out in being by far the least eager to share information. When considering willingness to share, however, participant A scores among the highest.

Trust

In Table 14.7 the individual trust scores are given. On the vertical axes, the awarded points are given, and on the horizontal axes, the received points are depicted. To give an example, participant C awarded 15 points to participant F. The rewarded points are the aggregate of the 3 questions that were asked on a 9 point Likert scale (hence, 27 points is the maximum score).

Table 14.7: Trust table

Points received / Points given	A	B	C	D	E	F	G	H
A		10	11	-	10	7	11	11
B	-		-	18	-	-	19	-
C	16	17		11	17	15	17	17
D	13	22	14		18	16	17	26
E	17	10	16	-		15	13	18
F	-	-	-	-	-		-	-
G	14	22	15	9	3	15		17
H	20	15	18	21	21	15	21	
Points Received	80	96	74	59	69	83	98	89

It stands out that multiple "-" can be found in the table. This happens when the participant did not answer the three questions about another participant. Participant F decided to not answer any of these questions, and Participant B answered only for two other participants. The other two zeros in the table are probably due to a mistake, or because the participants A and E did not feel comfortable to answer this question about participant D.

Liking

In Table 14.8, the aggregate results of the individual liking scores are depicted. The table works similar to Table 11.6.1. However, this time, 4 questions were asked on the 9 point Likert scale (maximum score 36).

Table 14.8: Liking table

Points given \ Points received	A	B	C	D	E	F	G	H
A		14	14	-	11	12	17	16
B	-		-	20	-	-	22	-
C	21	21		22	20	22	20	22
D	21	26	25		24	14	14	29
E	23	14	21	-		20	18	25
F	-	-	-	-	-		-	-
G	16	33	23	17	4	20		25
H	26	20	28	29	28	21	27	
Points Received	107	128	111	88	87	109	118	117

Table 14.9: The positive and negative reciprocity per individual

Respondent	Positive reciprocity	Negative reciprocity
A	14	27
B	21	27
C	21	7
D	28	13
E	25	17
F	21	17
G	21	15
H	26	19

Reciprocity

In Table 14.9, the scores for both positive and negative reciprocity can be found. We see that, in contrast to the other participants, participant B and A value negative reciprocity higher than positive reciprocity. Also, participant C says that he/she hardly responds to negative reciprocity.

14.3.2. Interpretation of Questionnaire Results

Let us reflect on some of these individual results, as well as the aggregate results in this Section. The results have been discussed with (Derks, 2017, May 30th), and together we have made some interpretations of things that stand out, and that will serve as input for the discussion (Chapter 16 and 17).

Interpretation of Social Network

In the social network, we see that some participants have pointed out each other as ties, but some of the ties are one directional. Derks points out that person B has only ties in two directions, which suggest a lot of reciprocity in their relation. Person F on the other hand, has not been pointed out by any of the participants as the person they have most contact with.

Interpretation of Centrality Measurements

It is interesting to see that the nodes with the highest degree also have the highest centrality on all three measures. This is coincidental, as there is no guarantee that the different measures will lead to the same results. However, as they all measure how central a person is in the network, these small networks can lead to roughly the same results in centrality for all three measures.

Interpretation of Trust Measurements

This table allows us to compare the extent to which participants trust each other. To give an example, participant H awards 18 points on the trust scale to participant C. Participant C, in return, awards participant H with 17 points. We can conclude that the participants C and H trust each other roughly to the same extent. Participant D and H award each other as most trustworthy.

Following this line of reasoning, the trust figures of participant B are interesting. B has only answered the trust question for two participants, namely D and G. Now, if we see what points were received by participant B, we note that participant D and G have awarded their trust in participant B with the highest points in the column of participant B.

Participant A awards significantly lower points to all participants than the other participants. Why this is, is hard to say. Furthermore, there are a couple of places where the mutual trust does not align well. participant G awards participant D with only 9 points, whereas participant D awards participant G with 17 points. Apparently, they so not trust each other to the same extent. The same holds for participants E and A, participants E and G, and participants H and A.

Now, given this data, participant G and B are considered to be most trustworthy, as they have the highest overall aggregate score. If trust would be chosen as the factor to decide who to exclude from a second run of the experiment, these participants would be picked. However, the data quality is questionable, as some zeros appear. There are multiple ways we could try to compensate for that, but there is no evidence that any of these adaptation methods is valid. One example of possible adaptation methods would be to compute not an overall trustworthiness score, but an averaged score, where the number of answering participants is taken into account. In that case, we would obtain the results in Table 14.10. That way, participant H and G, shortly followed by both B and A are considered most trustworthy.

Table 14.10: Trust table average

Participant	Average score
A	16.0
B	16.0
C	14.8
D	14.8
E	13.8
F	13.8
G	16.3
H	17.8

Interpretation of Liking Measurements

We see many similar things as we saw with trust. We have zeros at exactly the same positions in this table as we had in Table 11.6.1. Also, we see a very high trust between participant H and D, and the same observation for participant B can be made.

Also here, the mutual liking does not always align. For example, participant E likes participant A a lot, but this is not mutual (liking scores 23 versus 11). The same holds for Participant E and G (scores 18 versus 4), and to a lower extent also for participants H and A (scores 26 versus 16).

Participants G and B are liked most, according to this data. In an experiment where liking would be the factor of exclusion, these would be excluded. If we compensate for zeros in the same way as we did for trust, we obtain the results in Table 14.11. Note here that the scores are higher than in the trust table, as there is one extra question asked. This time, we see that participants C and H score highest.

Interpretation of Reciprocity Measurements

What is interesting to notice, is that the lowest score in positive reciprocity belongs to the same person as the highest score in negative reciprocity, namely participant A. The other way around, the highest positive reciprocity score belongs to the second lowest negative reciprocity score through participant D. Hence, the participants have strong feelings about their reciprocity.

Table 14.11: Liking table average

Participant	Average score
A	21.4
B	21.3
C	22.2
D	22
E	17.4
F	18.2
G	19.7
H	23.4

Cross-factor Interpretations

Here, we make some observations on the connection between the seven factors. First of all, notice the large extent to which the factors of liking and trust are alike. The correlation between the two tables is equal to 0.85. This indicates that the two depend on each other, but the sample size is too small to conclude anything on it. For future research it might be interesting to research the nature of these two factors, a correlation between them, and a causality (e.g. does liking the other person cause you to trust the person too?).

The second observation we make is that the highest scores among the different factors vary a lot. Although some participants pop up more at the extremes than others, there is no clear pattern that can be depicted.

Reciprocity and being liked and trusted also seem to be connected. Those who have a large difference between their scores on negative an positive reciprocity are liked and trusted less than those who score more equally to it. Indeed the correlation between the liking (and trust) scores and the difference between positive and negative reciprocity has a correlation of -0.48 (-0.71)(please note the small data set, and thus the impossibility to draw conclusions upon this number). Derks suggests that people having a strong sense of negative reciprocity often are less proactive and more introvert; they only come to action when they get hurt. The four people with the highest difference in their reciprocity scores are person A,C,D and E, they all score among the 5 least liked people. According to Derks, this can be explained through the notion of stability. People who react non-proportionally heavy to positive or negative incentives by others are considered to be unstable. This makes them both less liked and less trusted. One participant really stands out in this line of reasoning. That is participant E, who is least liked, but has the smallest difference between positive and negative reciprocity. This could be due to the small data set, or due to some kind of hypocrisy of participant E.

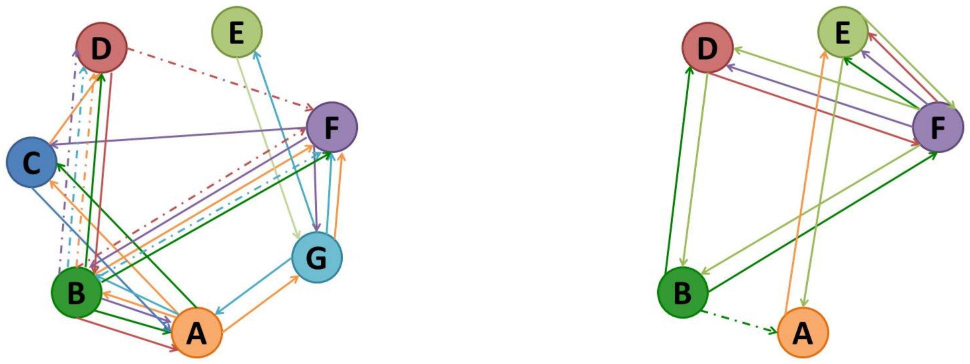
Another thing pointed out by Derks is that people who do not trust each other probably also do not like each other, and vice versa. If you do trust somebody, however, this does not mean that you also like that person. We do not directly see that in the data though.

14.3.3. Experiment Results

The results of the experiment can be found in Figure 14.4a and 14.4b. The colors belong to the participants, and show which participant have passed what information to whom. If the lines are dashed, this means that the sending participants notes that he/she has sent the information, but the receiving participant did not notice.

14.3.4. Interpretation of Results

The experiment was successful in the sense that the caching pattern clearly returns. Furthermore, the participants clearly understood what was meant from them. Let us attempt to link the three categories of factors to the data we observe.



(a) The information sharing in the original experiment (b) The information sharing in the disturbed experiment

Figure 14.4

Structural Factors

Participant H is missing here, since this participant was not present for the final experiment. As a result, the network consists of 7 nodes. A complete network[6] of 7 nodes has 21 links. Of these 21 links, 10 are used for communication in the first experiment (Figure 14.4a).

In the social network as depicted in Figure 14.3, 8 relational ties between the final nodes of the network are present. Of these 8 ties, 5 are in use in the experiment, and three (between C and G, A and E and between G and E) are not in use. The probability that 5 or more ties of these eight are used when at random 10 communication ties are picked from 21 possible ties is 26.8%[7]. This is not enough to conclude that the social network has any influence, but even if the network has any influence, it is limited, and the participants do not hesitate to use other ties than the ties existing in their social network.

For the second experiment, only four links from the original social network are left since two nodes were removed from the network. All these four links are being used. The complete network here has maximum 10 links, of which 6 are in use. The probability that if at random 6 out of 10 communication channels are selected, 4 of them belong to the specific paths are equal to only 7.1%. This would suggest that the links have a real added value here, but the dataset is too small to be conclusive about it.

Before continuing to the attitude relations, if we take the dashed links as full links, the degree of all nodes in both experiments are depicted in Table 14.12. We use this as a measure of how much the participants shared.

The least sharing participants in the network are C and E in the first experiment, and A, B and D in the second experiment. The most sharing participants are A, B and F in the first experiment, and participants E and F in the second.

Attitude and Relational Factors

In Table 14.13, the highest and lowest scoring participants per factor were measured. There is no pattern recognized here. In the Discussion section (Chapter 16), we suggest reasons why that is not the case, and how we could find patterns between these. Derks points out that one way of looking at these different extremes is to consider the team as very complementary to each other. They have different strengths and not one person stands out clearly.

[6] a complete network is a network where all nodes are connected through a direct link with all other networks

[7] This is a simulated result

Table 14.12: The degrees of the nodes in the experiment

Participant	Original experiment	Disturbed experiment
A	10	3
B	15	5
C	5	
D	7	5
E	2	6
F	10	9
G	7	

Table 14.13: The lowest and highest score per factor

	highest	lowest
original experiment	A,B,F	C, E
disturbed experiment	E,F	A,B,D
eagerness	G, D,C	A, B
willingness	G,D	C,E
trust	B,G	D,E
liking	B,G	D,E
positive reciprocity	D,E	A, (B,C,F,G)
negative reciprocity	B,A	C,D

If we compare the willingness to share with the actual sharing in this experiment, we see that the two participants that were the least willing to share (namely participant C and E) indeed share the least information in this experiment. For the eagerness to share, this does not show that clearly.

14.3.5. Link to Simulator

The data of the experiment as depicted in Figure 14.4b was inserted in Icarus. As before, the end users were added as dummy variables. This time however, as there were multiple sources, the same type of dummy variables were used for the sources, that way allowing the participants in the simulator to be both a source and a receiver/router of information. If no weight is added to the links yet, we obtain the following information has been cached in the nodes after completing the simulation (Table 14.14). Note in the third column the cached content in the experiment.

Table 14.14: What information is with whom after the experiment and the simulation

respondent	assignment	experiment	simulation	with weight	with added ties
A	1	1,2,3,4,6,7	1,3,4,6	1,3,4,6	1,2,3,4,6,7
B	2	1,2,4,6,7	2	2,7	1,2,4,6,7
C	3	1,2,3,6	1,2,3,4,6,7	1,2,3,4,6,7	2,3,6
D	4	1,2,4,6,7	1,2,4,6,7	1,2,4,6,7	1,2,4,6,7
E	5	5,7	5,7	5,7	5,7
F	6	1,2,4,6,7	1,2,6,7	1,2,4,6,7	1,2,4,6,7
G	7	1,5,6,7	2,5,6,7	5,6,7	1,2,4,5,6,7

As can be seen, the results are not entirely the same. This can be explained by the restriction on the social network. We have seen that many participants were feeling comfortable enough to share information outside their social network ties. This is partly because the assignments were task driven. As a result, the information can never travel along the exact same route in the simulation (with only the social network ties available) as they do in the experiment.

Now, if we add the used ties to the social network in the simulation, we obtain the results in the last column. These results are mostly the same as the results of the experiment.If we would add different weights for different messages, we would be able to rebuild the entire experiment in the ICN simulation.

In linking the data of the disturbed experiment to the simulator, we run into a problem. The choice of the participants that were removed was based on their centrality. However, by removing these two from the experiment, their social network became disjoint. Now, participant A and E are separated from the rest of the network. This was no problem in the experiment, but it is a problem if we want to link the data to the simulator, as it does not work with disjoint graphs.

Picking these two participants in retrospect was a mistake. If foreseen, two participants should have been selected that keep the network from forming two disjoint graphs. Furthermore, the measure we used to select the participants increases the probability that this happens. The betweenness centrality is a measure for how many of the shortest paths between nodes pass this particular node. If this is high, that implies there are not a lot of other options. Removing this node from the network therefore decreases the options to get from one end of the network to the other. Hence, using the betweenness centrality measure for the participant selection increases the probability that we accidentally create a disjoint network.

We solved the problem artificially, but effectively. In linking the data to the ICN simulator, we added one extra link to the social network, namely from E to F. This way, the network does no longer exist of two disjoint parts.

The results of the information cached in both the experiment and the simulator of the disturbed experiment can be found in Table 14.15. What we see here, is that even without adding weights to the nodes, the caching behaviour is the same in both cases.

Table 14.15: What information is with whom after the disturbed experiment and the simulation

respondent	assignment	experiment	simulation
A	1	1,2,5	1,2,5
B	2	2,5	2,5
D	4	2,4,5,6	2,4,5,6
E	5	1,2,4,5,6	1,2,4,5,6
F	6	2,4,5,6	2,4,5,6

15

Conclusions

> Research is formalized curiosity. It is poking and praying with a purpose.
>
> Zora Neale Hurston

In this chapter we draw the conclusions of this research. This is done by first answering the four sub-questions. Secondly, the main research question is being answered.

15.1. Answers to the Research Sub-Questions

First of all, we answer the 4 research sub-questions.

15.1.1. What is the Essence of the Analogy Between ICNs and Human Collaborative Networks?

In Section 12.1, we have considered the essence and nature of the analogy. In both human collaborative networks and ICN, information is shared among different people/nodes in the network. In an ICN simulator, the way information spreads can be detected. In a human network, this is more complicated, but possible in the experiments we have set up. These two kinds of information spread can be compared in an analogy, in light of different aspects of both networks. In Table 15.1, the items of comparison can be found.

Table 15.1: Overview of the links between ICNs and human cooperative network

Function or parameter in ICN	Parallel in human collaborative network
Topology	Topology of human interaction
Router	Roles in network - a router of information
end user	Roles in network - a receiver of information
Source	Roles in network - a source of information
Content popularity	Content distribution in human communication
Cache	Humans saving information to spread it later again
Disturbance in link, cache or node	Disturbance in contact, person or 'memory'
Caching strategy	What information is memorised by whom[1]
Cache size	How much information is remembered[2]
Workload	The set of information items that is requested from the network[3]
Content population	How much content is being shared?
Cache replacement policy	Irrelevant as the cache size is infinite
Requests per min	Speed of sharing information

Besides the items of similarities, there are some essential differences between human collaborative networks and ICN. In Section 12.1, we have pointed out some of these differences. First of all, we treat information in this thesis as a 'piece'. This is done to make a clear analogy possible, but it is a simplification on both sides of the analogy. In ICNs, different pieces of information might have a different size. However, we treat them all as equally large, occupying the same amount of cache memory. In a human network, information is often embedded in much more than a simple fact. Among others, emotions and context play a role in understanding the information that is passed along. Furthermore, there is a difference between information, data and knowledge.

The second essential difference between human collaborative networks and ICNs is that in an ICN, the information is pulled out of the network, whereas humans also push information to the next person. Thirdly, humans do not have an analogue concept such as a caching strategy or cache replacement strategy. Humans have no control over what information they remember, and what they forget again. We assumed humans to remember everything. However, in the experiments we saw that sometimes person A would claim to have told something to person B, but person B did not recall receiving this information. This will be elaborated on in the next chapter.

In essence, both networks spread information through their networks in an analogue way. The way this information is processed and spread has been subject of this research. Furthermore, certain patterns and behaviour occur when we disturb both networks. This information resilience has been part of the research, but it is not considered to be part of the essence of the analogy.

15.1.2. What Factors Influence the Way Humans Share Information in a Collaborative Network?

In the systematic literature research, seven main factors that influence information sharing behaviour in human collaborative networks are determined. The factors are subdivided in relational factors, attitude factors and structural factors. The relational factors are: reciprocity, linking and trust. The attitude factors are willingness to share and eagerness to share. Finally, the structural factors are centrality measures and social network ties.

These seven factors were chosen from a larger set of factors found in literature (see Figure 11.6). The factors were used to determine which people were excluded from the experiment in the second parts of experiments two and three. Furthermore, the factors will be used in the next chapter to link these factors to the ICN in a thought experiment.

15.1.3. To What Extent Can the Information Spread in Human Collaborative Networks be Incorporated by an ICN Simulation?

In the different experiments, we have captured the information spread in human collaborative networks. The results from these experiments have been used as input for an ICN simulator. Some detail was lost in transforming from a natural setting to an experiment. Furthermore, some adjustments needed to be made to be able to incorporate the data obtained in the experiments in an ICN. However, when done so, the two can be mapped onto each other.

To answer the question, we can capture the information spread in human collaborative networks in an ICN simulation if we use the social network as well as the results of the experiment as input data to the simulation. Furthermore, we need to make the two adjustments mentioned above. That is, we need to accept the 'translation' from a natural information spread to the experimental setting, as well as making some adjustment from data obtained in the experiment to data that can be dealt with in an ICN simulator. We will elaborate on both adjustments next.

Differences Between Natural and Experimental Setting

Natural behaviour differs from behaviour in an experimental setting. By distracting the participants with building a marble track in the experiments, their behaviour becomes more natural. But issues such

[1]On the short term, this is Leave Copy Everywhere. as we assume that people will remember the information

[2]We often assume this to be infinity, as it is very hard to quantify.

[3]Hard, as humans do communicate less through pull and also through push mechanisms.

as adaptive behaviour and learning effects remain present. Furthermore, people share information through more than just face to face contact. Nowadays, a lot of information sharing happens through email or document sharing as well. This is not represented in the experiments. For the level of detail in the analogy at this point, the multiple channels through which humans share information is no problem for the analogy.

The third way in which the experiment differs from the natural setting is that the goals for sharing in these experiments are somewhat artificial. This was also pointed out by the participants in the first set of experiments.

Differences Between Obtained Data and ICN-Suitable Data

We need to add dummy variables for source and receiver as humans can be sources, receivers and routers simultaneously, but the ICN simulator can not deal with that. Furthermore, in some cases, we need to add weights to the paths, as not all information is treated as equally important to all people in the network. Lastly, we need to add ties if the (disturbed) social network contains disjoint connected components[4]. However, it would be much better to prevent the network from becoming disjoint in the first place. This can be done by selecting the removed persons differently.

If accepting the elements above, the information spread in a human collaborative network can be incorporated in an ICN simulation.

15.1.4. How Does the Information Spread in a Disturbed Human Collaborative Network Relate to Information Resilience in ICN?

Where the first three questions were mainly used to build the analogy, the fourth question is meant to investigate how well information resilience in an ICN and information resilience in a human collaborative network can be compared. In other words, is it possible to extend the analogy with the concept of information resilience?

In ICN the information resilience is defined as the percentage of information requests that can be delivered successfully, given a disturbance. In ICN simulations, a list of these information requests can be generated according to some parameters. In the experiment on the other hand, we have only been able to trace specific flows of information. We cannot compare the flows with and without disturbance one-to-one, as we have only measured very specific flows. Furthermore, there is no such thing as a 'failed' information request in the human collaborative network, as human communication in the experiments are both push and pull in nature, and ICNs are solely pull in nature.

As we cannot measure failed information requests in human collaborative networks with these experiments, there is no possibility to measure the information resilience in human collaborative networks in the same way as we can in ICNs. Hence, a boundary in its current state is found. However, in a later state, one could come up with another measure of the information resilience, with that allowing a numerical comparison between the two.

Even though a numerical comparison is not possible, some remarks can be made about the information resiliency in human collaborative networks. Unlike ICN simulators, the humans in the experiments were not limited to the social network ties they had among each other. In the light of the experiment, the people were happy to talk to others if that would be beneficial for the experiment. We come back to this in de the discussion section.

Even though a disturbed human collaborative network is functioning different from a disturbed ICN, the results can still be incorporatedly in an ICN. Even without adding weights to the links, the ICN caching data fully aligned with the data of the human collaborative network in the third experiment. This suggests that even though the disturbed experiments might not be conclusive on the information resilience, they do support the analogy and the simulation of information spread through an ICN simulator.

[4]disjoint connected components means that there are at least two nodes between whom there exists no path

15.2. Answer to the Main Research Question

"To what extent can the information spread in a human collaborative network and an ICN be applied to each other, to enhance insights into information spread and information resilience?"

In this thesis, three experiments have been executed to set up an analogy between information sharing in both ICN and human collaborative networks. The research aimed at finding the possibilities, and researching one possibility in particular; the information resilience in both networks. In Figure 15.1, we find an overview of different possibilities and boundaries. Boundaries occur either because the experiment is too complex to perform in its current shape, or because the analogy is too complex. The axis represent the complexity of the experiment as well as the complexity of the analogy. On the left/under part of the axis, the complexity is small enough to be incorporated in the current analogy. On the right/lower side of the axis, the experiment or analogy becomes too complex.

After this research, we are located at the centre of this figure. However, the axes are fluid, and after performing more research, the axis might go up or right or both.

As we have seen already in the first experiment, the data obtained there can function perfectly as input data to an ICN simulator, if two adjustments are made. First of all, dummy variables need to be made to simulate that a node in a human collaborative network (a person) can act as a source, an end user and a router at the same time. Secondly, different weights need to be assigned to the different streams of information to simulate that not all information is equally interesting to everyone.

The sub-question on the nature of the analogy show that the different variables and items of interest in an ICN have a natural variable or item in the human collaborative network. However, the other way around, this is not always the case. We have seen that human interpretation often plays a role in the streams of information in human collaborative networks, but interpretation is not happening in the ICN simulator that we have considered. Furthermore, humans tend to make new connections in their network if there are clear benefits to do so (for example because the preferred people are not available at the moment). In ICN this is possible if we consider mobile networks. However, in this thesis, we have limited ourselves to a more traditional network (and a corresponding ICN simulator). We will come back to this in the discussion section (Chapter 16).

The ICN simulation features that we have considered fit nicely on a human collaborative network. However, there are features of ICN that do not fit in a human collaborative network easily. One of these features is the caching strategy. In an ICN, routers can choose to store a piece of information. This choice is based on predetermined algorithms. A human being however, cannot actively decide to not remember something. Hence, caching in a human collaborative network is imminent once the information is shared.

In the third experiment, we have incorporated the 7 items of our focus (selected in Chapter 11) that influence information sharing in human collaborative networks according to literature. We have collected an indication on all seven items before the experiment. One of the items, namely the centrality in the network, was used to pick the participants that were excluded from the experiment the second time. However, we have learned in doing so that the social network of the participants should remain one connected component instead of falling apart in several disjoint components. If the latter is the case, the ICN simulator does not accept the network.

To conclude the research, we can say that ICN theory and simulators can largely be applied to human collaborative networks. Some adaptations needed to be made, and some limitations found here might be solvable in a next iteration of the research. This could be because some human features have an equivalent in IT networks, which might be applicable in a next iteration of the research. However, applying the information resilience in its current state to the analogy is not possible due to two reasons. First of all, as the assignments were task driven, the participants felt comfortable enough to step outside their social network, and talk to others as well. Secondly, as the experiment was designed with information sharing in a pushing way, and ICNs work with pulling information out of the network, the measure used for computing the information resilience in an ICN does not apply on the data gathered in the experiment.

The ultimate (future) goal of this research was to allow decision makers to gain insights in their network (e.g. their team) with respect to information sharing and information resilience. To that extent, mapping the social network of the team as well as their professional network certainly enhances insights. Having knowledge about influencing factors such as trust and centrality also enhances the insights. However, at this point in the research it is too early to conclude that certain factors are more importantly than others, or that it becomes possible to predict how the team will deal with a disturbance within the team. All items of the comparison that we have considered or researched are mapped in Figure 15.1. They are scored in complexity, both with respect to the experiment needed to measure, and the the complexity of the analogy. The items on the 'simple' side of the lines (in both directions) were simple enough for this research. The items on the other side of the lines are too complex for this research. Hence, the items in the left lower corner of the diagram have been researched in this research, and further research is needed for the other items, as they have shown to be too complex.

At this point, the analogy does allow for a way of thinking that can be very valuable for decision makers. On this way of thinking will be elaborated in Chapter 17.

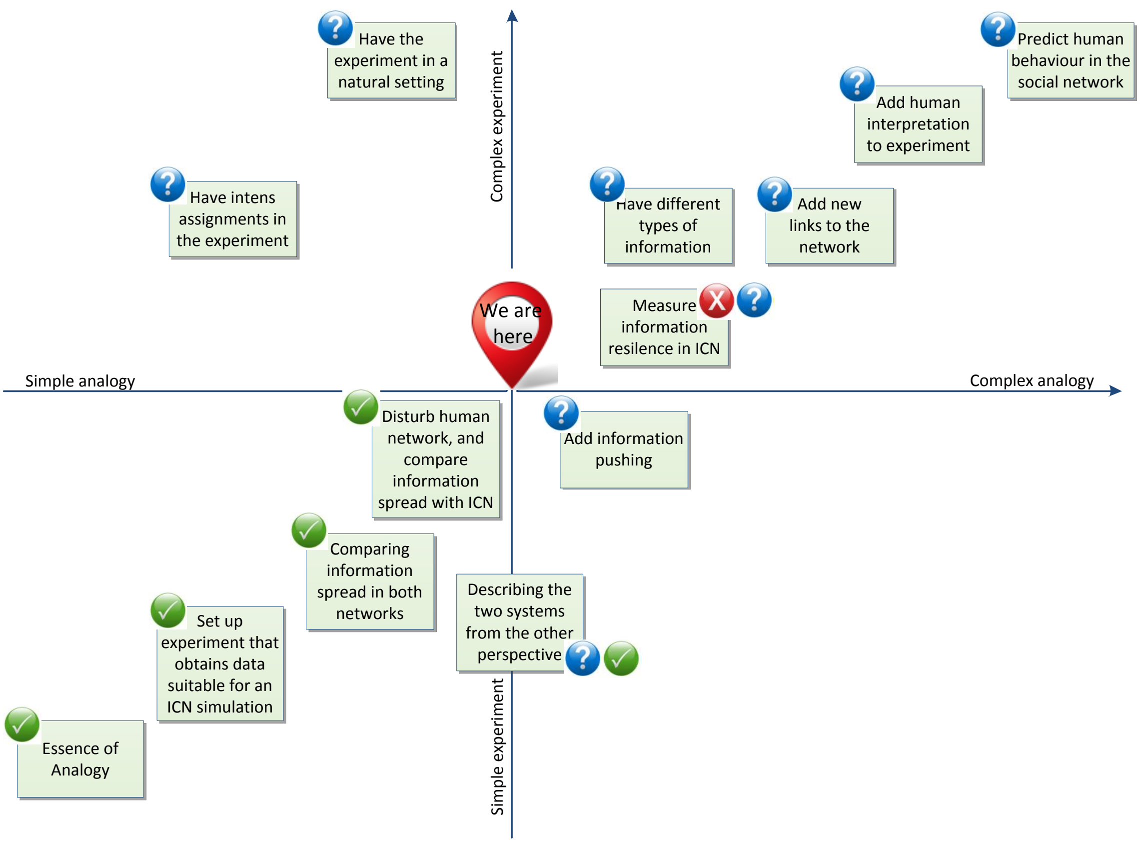


Figure 15.1: An overview of the conclusion

16
Discussion

The important thing is to never stop questioning
A. Einstein

In this chapter, we will discuss particular elements from the research that stand out. In doing so, we will go through the research backwards. We will start with critically reflecting on the results. Furthermore, the methodology, and with that also the experiments will be elaborated on. Next, the literature is discussed. By that time, we have reached the beginning of our research.

From there, we will look into the future. What could be possible extensions of the analogy that could be researched? What other future research is of interest, not only extend the depth of the analogy, but maybe also its use, both for decision makers, and as a way of thinking as well.

In chapter 17, we will elaborate on the use of the analogy and the added value of it for both research fields. These items are also part of the discussion, but because they are of added value to both research fields, they will be discussed in Part IV of the research.

16.1. Results

The results of the experiments give ground to further investigate the possibilities to build up the analogy to a higher level. Furthermore, further research into the connection between the seven factors and the measure of information spread could enhance insights upon the way they influence the information spread.

Information Resiliency

Even though a numerical comparison is not possible, some remarks can be made about the information resiliency in human collaborative networks. Unlike ICN simulators, the humans in the experiments were not limited to the social network ties they had among each other. In the light of the experiment, the people were happy to talk to others if that would be beneficial for the experiment. This is explained by (Derks, 2017b), who pointed out that intense or emotional assignments have a higher impact on the intentions of information sharing than task driven information sharing assignments as we have used in the experiments. If the information is intenser or emotional, people will stick to their social network ties more strictly. Due to ethical and practical reasons, we have not been able to perform the experiment with more emotionally driven assignments, but it would be interesting to see what happens in that situation. If the participants stick to their social network more strictly (which we expect to be the case when considering more intense assignments), one of the two problems that occurred in computing the information resilience would be solved. If the assignments remain task driven, the information resilience in the experiments is very high, as the participants have no problem in talking to each other about the tasks.

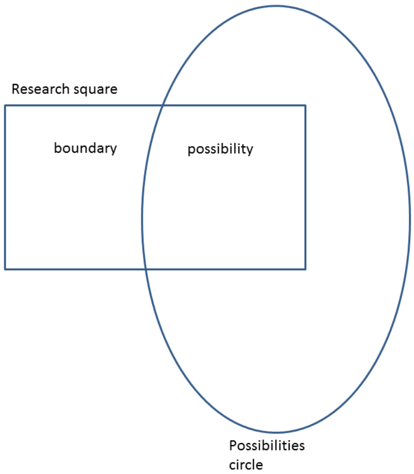


Figure 16.1: What has been researched versus what has been asked

To What Extent...

The main research question starts with the words "to what extent". Answering this question completely is difficult, as it implies that something can be measured. In Figure 16.1 this is depicted. Let the oval symbolise the whole extent to which ICN theory can be applied on human collaborative networks. Let the square depict everything that has been researched. Now, how can we ever be sure that we have researched the whole oval? Probably we cannot. It is therefore that the words to what extent should be interpreted in this context as 'to what extent can we...'. This extend has been part of the conclusion (recall Figure 15.1). In Section 16.5, we elaborate further on the possibilities to extend the analogy. We will use Figure 15.1 as a starting point, and elaborate on the impact of, and the reason for increasing the complexity over each of the axes.

16.2. Methodology

In this section we reflect on the methodology of the thesis. Before we pay attention to some details of the experiments, we will reflect on the methodology as a whole.

16.2.1. Effectiveness of Methodology

The methodology should allow us to achieve the goal of the research. Let us reflect the goal of this part of the research before reflecting on it. The aim of the research is to "gain insights into the information spread through, and information resilience in a human collaborative network."

The methodology used for it was to build up an analogy between human collaborative networks and ICNs. This analogy was build up with the help of three consecutive experiments. The first two experiments were used to construct the third experiment. The third experiment allowed us to complete the analogy. The analogy did provide us with insights on the information spread through and information resilience in human collaborative networks. However, as elaborated on in Chapter 17, the way of thinking connected to this analogy also provides us with a lot of insights.

Now, how would the results be different without the use of the experiments? For that, let us (again) turn our attention to Figure 15.1. The essence of the analogy was researched without the use of the experiments. However, all other items are not possible without the course of an experiment. Already comparing the information spread in the two networks needed the data we obtained in the experiment. By going back and forth between the two networks, we have been able to build up the analogy up to the point we have reached now. Without the experiments, it would have been very hard to make this continuous comparison.

To conclude, the experiments were effective for building the analogy, and the analogy indeed enhances the insights in both information resilience and information spread in the human collaborative network.

16.2.2. Experiment

Natural Behaviour

One could wonder to what extent the experiment influenced the behaviour of the participants. As discussed before, due to practical reasons as well as our abilities to make the information sharing tangible, the experiment was held in a closed setting rather than a natural setting. This always influences the behaviour. The marble track supposedly encouraged the participants to behave as natural as possible, but this remains somewhat artificial.

Questionnaires of Seven Factors From Literature

In the systematic literature research, we have found seven factors that influence the information sharing in human collaborative networks. In the second experiment, we used an existing questionnaire to question the participants on the extend they liked each other. In the third experiment, we added the other six factors. As these six factors were also measured from existing questionnaires, we did not expect any problems in this way of questioning the participants, we did only add it in the third experiment. As expected, no issues occurred in questioning the seven factors.

We will pay attention to the impact of the seven factors and what would be different when one of the other seven factors would be used in Section 16.5.

Resilience in Both Networks

In the experiments 2 and 3, we have attempted to also measure the information resilience in the human collaborative network. We have done that by disturbing the human collaborative network by taking away people from the experiment in a second run of the experiment. In doing so, the network unintentionally became disjoint. Because the assignments given were task driven, this turned out to not be a large problem for the participants, as they has no problem in sharing their information with others. Hence, this potentially large problem, was solved by the participants themselves (and we have not observed any problems in doing so).

In our attempt to measure the information resilience, we believed at first that we should think of disturbing the network as explicitly as possible. The larger the disturbance, the larger the effects, the larger the probability to catch the effects. This made us decide to disturb the network differently than we did in the mathematical part of this research. In the mathematical part, we had removed random links. In this part, we disturbed the networks by removing particular nodes. Which nodes, was decided upon the centrality of the persons in the network.

In retrospect, one could wonder whether this really was a good idea. One of the issues was, as mentioned already, that the network became disjoint. Furthermore, the effects have shown to be difficult to measure even with these disturbances. If we would perform the experiment again, we would reconsider how to disturb the network. However, we believe that replacing the task driven tasks by more emotional tasks is probably more effective in making the step towards measuring the information resilience in the human collaborative network.

Influence of the Questionnaires on the Experiment

One could wonder how, and if the questionnaires before the experiment have influenced the way humans behaved during the experiment. One of the respondents in the second experiment has pointed this out in the evaluation. That means it made him actively rethink his actions. The questions, however, could also implicitly influence the behaviour of the other participants.

16.3. Literature

Level of the Variables

In the selection of the variable of our theoretical framework, we have already excluded some of the factors because they are too complex for the experiment. However, it is worth wondering whether all seven factors in the theoretical framework are on the same level.

The items trust and reciprocity are very complex in nature. Even though we have divided the item of trust into three sub items (faith, dependability, predictability), the notion of trust is in essence much more complex than these three items. The social network ties, on the other hand, can have a very complex cause, but the result is rather simple. Hence, the variables do not all function on exactly the same level. However, all seven factors are simple enough to be caught in a questionnaire.

Eagerness and Willingness to Share

Eagerness and willingness to share are rather uncommon factors in literature when considering information sharing. They are pointed out by van den Hooff and Hendrix (2005). They are both a subset of the attitude towards knowledge sharing, which is a rather common factor considered in knowledge sharing in organisations. We have chosen to narrow the attitude towards knowledge sharing down towards the eagerness and willingness to share, because these factors are more tangible than the attitude towards sharing. Furthermore, they are measured more easily.

Even though the two factors do not seem to be one of the most important factors in information sharing in human collaborative networks, they are in fact part of the attitude towards information sharing, which is an important, but which is difficult to grasp (measure). It is therefore that we have chosen to use these more specific factors. Ideally, we would do the same for trust, as this is also a very complex factor. However, it is much harder to break down the trust factor into smaller items. This leads to a set of 7 factors that are not all on the same level. The factors centrality, social network ties, eagerness to share, willingness to share and liking are on a lower dimensional scale than trust and reciprocity are. Let us elaborate on that in the next subsection.

Eclectic Research

The theoretical framework is constructed by selecting elements from a wide range of possible elements for the theoretical framework (recall Figure 11.6). This selection is partly based on the relevance for our research, but it is also based on the need for factors that are measurable and simple enough to deal with. This could lead to the impression that the theoretical framework is eclectic in nature. With that we mean that the framework is build up by picking those factor that suit us best for not only research, but also practical reasons.

Even though this is the case, and the research can therefore be called eclectic this way, this is not necessarily a problem. The basis of the research is still solid, and the choices have been made on clear arguments. With more time or a different experiment set up, one could decide upon different factors listed in Figure 11.6, but feasibility always play a role in doing research.

The two factors that have a higher complexity than the other five, namely trust and reciprocity, are studied a lot in literature. Even though their complexity, questionnaires to measure them have been developed. For future research, however, we suggest to look into the possibility to narrow down these factors as well.

What we did specifically not take into account here, is the interaction between the factors. One could argue that if one trusts someone else more, he/she most probably would also be more willing to share information. Furthermore, if you like somebody, the chance is probably large that this person also forms part of your social network. If we would take these factors into consideration, the theoretical framework would have become too complex to deal with. It could be interesting, however, to investigate these cross-links, possibly from literature, but also from experiments.

Hidden Profile Paradigm

During the literature research, we have been particularly eager to find an experiment in literature that could be used for our research. In this search, we have found one particular set of experiments that have suitable elements. These experiments are all part of the hidden profile paradigm. In these kind of experiments, group decision making is subject of research. Typically, the group is asked to make a decision (upon what differs greatly), but part of the information that could influence the decision is not shared with all participants (Stasser & Titus, 2003). Certain elements of this paradigm are in line with the design objectives of our experiment. For example, in the experiments, information sharing is present, the existing relationships remain intact and the participants are involved all time. However, there is also a problem with these experiments. As far as we have found, these experiments are always sharing all the information with all people at the same time, with that preventing caching behaviour to happen. Furthermore, the aim of the research is to show that decision making is often poor due to a shared information bias. Examples of this bias are over valuing the information of people with a higher power relation, or discussing the information known, rather than the information unknown. Hence, the aim is not to measure the information spread, but to investigate the decision making.

As a result of this, the hidden profile paradigm is not very suitable to investigate the information sharing in a human collaborative network.

Why ICN?

In this research, we have compared the information spread in a human collaborative network with the spread in an ICN (simulator). The reason for taking an ICN is twofold. The first reason is availability. A mathematical research towards information spread in ICNs was set up already, making the comparison to and ICN better accessible. Secondly, in an ICN, caching is imminent. In many other type of networks we could have considered, this caching is not an in-network property. For example, in Content Delivery Networks (CDNs), caching is present, but is an overlaying feature of the network. This is less analogue to human collaborative networks, as there are no people who are dedicated to only cache information and share it again. Always, humans are part of the set of end users as well.

The arguments mentioned above do not mean that the only possibility between an ICT network and a human collaborative network is an ICN. Just one example of a possibility could be a mobile network, where the devices connect to different places to the rest of the network, depending on their physical location. But due to feasibility and practical reasons, we have chosen to compare to ICNs. To the best of our knowledge, there are no imminent advantages of picking any other network. However, as said before, some networks are clearly unsuitable for comparison.

16.4. The Power of Analogies

During the course of this research, the analogy has been a point of reflection over and over again. The strength of the analogy has mostly been a theoretical question, as using the analogy has not been done until after the third experiment. However, at one point we started to use the analogy as a way of thinking. We elaborate on this in Chapter 17. Already here, however, we would like to use this chapter to show the power of the analogy.

When starting to write Section 17.5, the idea was to make the comparison between the pushing behaviour in a human collaborative network and ICN. The aim of the piece at first was to point out a vague idea of the comparison, basically showing that it is a hard concept. However, when writing the piece, and actually forcing ourselves to use the analogy to make the comparison, the idea of a new caching strategy that uses the two concepts of pushing and pulling information arose. We are confident that it was only due to the analogy that we came up with the caching strategy. The scepticism we had after the research about the true added value of the analogy was resolved entirely when actually using the analogy.

Analogies are used is science more often. Especially getting inspired by nature as an analogy is popular. This has become a whole field of research on its own, called biomimicry or biomimical design (Hacco & Shu, 2002). To give an example, A termite hill stays within 1 degree Celsius from the inside, no matter what the whether conditions outside are. For humans, to obtain the same with our homes, we need to heat and cool down our homes manually. What can we learn from termite hills with respect to the isolation of our homes that can help us lowering the needed amount of energy to keep our home at the desired temperature? (example from `https://biomimicry.org/biomimicry-examples/`) Making the analogy between termite hills and human housing/buildings, allows us to compare the techniques used in the first to be applied on the second.

This is exactly what we do in our analogy between ICNs and human collaborative networks. This analogy, however, is more abstract in nature than the one described above. On the one hand, this makes it more complex to compare items. On the other hand, the possibilities of comparison and learning from the analogy are much larger than in the analogy described above. On multiple levels as well as with multiple items of comparison, items can be compared in their analogue network, and new insights arise from that.

16.5. Extensions of the Analogy and Future research

The research could be extended in many possible ways. We could look at researches that apply the way of thinking established with the analogy. However, this will be done in Chapter 17, and not in this chapter. Furthermore, we could think of research that extends the analogy (e.g. adding a time component). But we could also focus on ways to establish a measurement for information resilience in both human collaborative networks and ICN. Furthermore, some more research on the influence and correlation of the seven factors selected from literature that influence information sharing in human collaborative networks should be performed.

16.5.1. The Analogy for Developing a Communication Platform

Technological, multi disciplinary innovation projects often have a need for a communication platform that encourages the different parties involved to communicate across the borders of each discipline. Here, we would like to propose a heuristic that could be used to design such a communication platform. The key idea is that the communication platform and technological innovation should be intertwined as much as possible, to connect the communication and the technological innovation.

The proposed heuristic consists of 3 steps. In the first step, the essence of the technological innovation and the essence of the particular human collaborative network should be mapped. In the second step, the analogy between the research group (human collaborative networks) and the technological innovation should be build up. This can be used in the third step as a starting point to design a communication platform that uses the integration of the technology and the communication as a starting point.

16.5.2. Extend the Analogy

Adding Time

One of the ways the analogy can be extended is by adding a notion of time. In the current ICN simulator, time is already incorporated. This is expressed in two ways. First of all, the interval in which information requests arrive can be simulated. Furthermore, the latency[1] can be simulated. These two notions have an equivalent in human collaborative networks. The intervals between the requests can directly be measured in an experiment. Furthermore, the notion of latency could be measured if pull plays a role in the experiment too. One could measure the interval between the start of a search for information, and the moment this has been found.

Seven Factors Selected from Literature

In Chapter 17, the seven factors will be used in thinking about the two networks through the analogy. In this Section, however, we will reflect on the role of the seven factors we have selected from literature. In particular, we perform a thought experiment about the 6 factors we did not pick for disturbing the network. Furthermore, we point out some future research items.

measuring the resilience in the human collaborative network has shown to be difficult in its current shape. As long as we are not able to use emotionally driven tasks in the experiment, we do not think much will change if we change the factors upon which we decide to remove people from the network in order to measure the information resilience. However, if we assume that the personal assignments in the experiment become emotionally driven, we could think about the influence of the factors on the experiment. In doing so, we expect the largest influence of the factors liking, social network ties and trust. The other 3 factors (reciprocity, willingness and eagerness to share) are expected to have less of an impact. The negative reciprocity starts to play only a role when people in the experiment start to treat each other negatively (intended or unintended) which we do not expect to happen a lot. The positive reciprocity becomes important if the participants really need to help each other, which is

[1] Latency is the time between the request and the arrival of the requested information. This time is dependent on many parameters, such as the distance between the source and the requestor.

not the case. The willingness and eagerness to share seem to influence the amount of communication (sharing), which is important. However, this does not mean that, in such an experiment (where sharing is the assignment), people with a low eagerness or willingness to share will stop to share at all.

Let us focus on those items that we believe to have a large impact. First of all, the liking seems to be important. Derks (2017, May 30th) points out that sharing emotionally important information is sometimes difficult to those that we like a bit (or not at all). When we do not know each other, this is sometimes easier. Furthermore, to our closest friends, we will not share this kind of information in a setting like this. Hence, by disturbing the network based on the factor liking, the way the information spread through the network will be greatly influenced as well.

Disturbing the network solely based on the social network ties is rather difficult. If we would want to disturb the network based on the social network, centrality measures quickly start to play a role. These centrality measures have already been taken into account. However, if we would be able to disturb the network based on the social network ties, we expect that people who usually do not have a lot of contact would also share a lot of information. By influencing that, we expect to be able to influence the amount of information sharing.

We would expect to share more emotional information to those people we trust than to those we do not trust. Hence, if we would disturb the network based on the trust measurements, we expect to be able to influence how much information will be shared in the experiment.

We have seen that the seven factors do not serve as forecasters for the sharing behaviour that we have seen per se. More research could show strong correlations, but these sample sizes are too small to conclude anything on that. However, the combination of proven correlations and the researched analogy could possibly lead to forecasts on whether or not a team will function after a given disturbance. We would therefore advise for future research into the correlation between the factors in the theoretical framework. In particular, the correlation between liking and trust as well as the (negative) correlation between negative reciprocity and liking and trust are interesting to investigate further.

Resilience in the Analogy

As we have seen in the experiments, in its current form, information resilience cannot be measured the same way in a human collaborative network and in ICN. It is therefore that the tile "Measure information resilience in ICN" is outside the region of what we have established in the research (Figure 16.2). However, by either adjusting the experiment or the measure, this might be solvable. In future research, it would be interesting to walk one of these two paths to try to extend the analogy.

This could be done by changing elements from the experiments. For example, if the items that need to be shared can be shared over a longer time, and are more personal (emotional) in nature, the resilience might actually show in the experiment. Furthermore, the social network could be extended with adding not only the social ties, but also the work related ties. If roles are then added, it is likely that the participants will behave over these ties. When that is the case, the resilience could be measured in the human collaborative network as well.

Another possibility is changing the measure of the resilience. This, however, needs some more thought, as we want to be certain to measure the extend to which the network can deal with people leaving this network.

Add Different Types of Information

In a human collaborative network, different kinds of information spread differently. This has partly been taken into account with adding weights to the links in the experiment. However, what would happen if a clearer distinction would be made between different types of information in an experiment? For example, what would happen if the experiment would simulate a project, where certain types of information are about financing the projects, others about technical details and a third type of information is about the aesthetic aspects of the project. Furthermore, if we establish roles such as design director and treasurer in the experiment, this would mean things for which information is

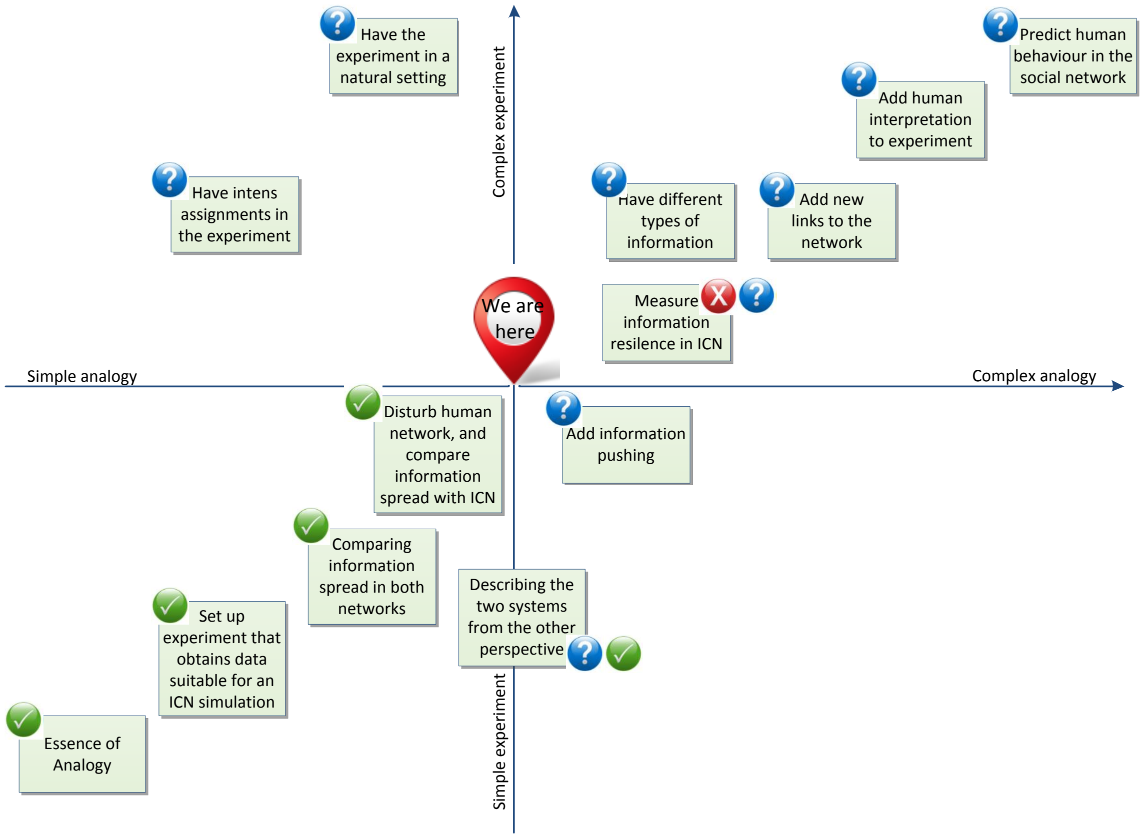


Figure 16.2: An overview of the possibilities to extend the analogy or the experiment

important to whom. As a result, we expect different streams of information in the experiment. It would be interesting to investigate if the patterns really occur the way we predict, and if more sensitive aspects of the information will indeed follow the ties of trust and liking. This is in line with the tile 'different types of information' in Figure 16.2.

16.5.3. Extentions Along the Axes of the Research Overview

In extending the research, it would be interesting to extend in two similar directions, namely adding complexity to both the analogy and the experiment. We have done this already for 'Measure information resilence in ICN' and 'Different Types of Information'. Here, we will explore along these two axes a little bit more.

As has been said in the conclusions (Chapter 15), the position the research is now, is depicted in the centre of Figure 16.2. All items right or above the centre are for now not possible. However, these axes are fluid, and with further research, extensions in these two directions might be interesting. We could use Figure 16.2 as a compass to guide us through the research. Ideally, the research moves in the north eastern direction of this compass, adding complexity both in the experiment and the analogy. However, sometimes, steps in other directions need to be made to be able to do so. For example, measuring the resilence has shown to be difficult. If adding complexity in both the experiment and the analogy is too difficult, it might help to first add complexity to the experiment (by having emotionally intense assignments) before adding the complexity to the analogy too. This way, Figure 16.2 can really be used as a compass.

Ultimately, we would like to end up in the right upper corner of Figure 16.2. A way of performing future research which applies the current analogy would be to create a tool that takes decision makers by the hand to enhance their insights in the information sharing of their team. Who likes who, what relation of trust is there in the network, and how can the information spread be encouraged by paying attention to the current network and its features? However, it is probably more desirable to build such a tool after some more research on the correlation and impact of the tool has been done.

IV

Synergy Between the Two Research Topics

17

Using the Analogy as a Way of Thinking

Science is a way of thinking
C. Sagan

In Part III of the research, we have build up an analogy between human collaborative networks and Information Centric Networks. This analogy allows us to compare information sharing in human collaborative networks and ICNs. But, as we will see in this chapter, the added value of building up the analogy transcends the items researched. In this chapter, we will use the analogy to apply concepts from one field to the other. As we will see, doing this enhances our insights greatly. We will apply the seven factors selected from literature in Chapter 11 to ICN through the analogy. One should note that the real added value of this research lies in the ability to move along the analogy, and approach concepts, problems or challenges from an original perspective. Already in this chapter, we find a new caching strategy, based on information sharing in human collaborative networks, which proofs the added value of the analogy.

This chapter is meant to give an example of the way of thinking this analogy allows. It is not our aim to be complete in showing all its possibilities, but rather to show the nature and direction of these possibilities.

17.1. Trust in an ICN

One of the factors that influence the information sharing in a human collaborative network is trust. People are less likely to share information with people they do not trust. Especially when this information is of high emotional value or has a high intensity (Derks, 2017b). In internet networks, trust is also of high importance. In fact, whole research groups are put in place to research how to ensure that information or other users online can be trusted. New developments such as blockchain (Swan, 2015) (Beck, Stenum Czepluch, Lollike, & Malone, 2016) and eventually in key distribution of Quantum Computing (Gottesman, Lo, Lutkenhaus, & Preskill, 2004) seem to deal with the issue of trust more naturally, but for most internet networks, it is still an issue.

By using the analogy, let us think about the concept of trust in online information in internet networks. The first thing that comes to mind when talking about trust in human networks is the need of a certain kind of relationship between two people before they can trust each other. Furthermore, the three factors of trust as discussed in Section 11.6.1, predictability, dependability and faith should be taken into account.

If we want to think about trust in an internet network, we can link one specific aspect of trust in an internet network through the analogy to trust between humans. Namely, we can think about a way how two nodes in an internet network can trust each other, and with that the information they share, we can use these concepts. Note with this that the trustworthiness of the source of information is no subject of investigation, unless one of the two nodes considered is the source.

In an internet network, a set of protocols, often called TCP/IP (Fall & Stevens, 2011), checks on the procedures in an internet network. It monitors from the physical layer until the application, in between also monitoring things such as the network and the transport of information. The TCP also corrects for data loss due to, for example, time outs. We limit ourselves here to the network and transport layer of the TCP/IP, as this is most analogue to the trust in a human network. Security measures such as private and public key encription (Dierks, 2008) are usually to check if the right message came across, and play a role on the transport level.

In an internet network, one could say that two nodes that are connected directly have a certain relationship together. Communication goes from one to the other, and possibly also back. However, whereas in face to face communication the identity of the other person is usually not under investigation, in internet networks it is sometimes uncertain that the other node really is the node it is claiming to be. In internet networks, several tools to do that are available, but most of them assume that intrusion could appear during transport.

Next, how can we connect the notions of predictability, dependability and faith to communication between two nodes in an internet network? One could think of a system that makes the two nodes depended on each other for receiving the right information. For example, a system with rewards could the thought about. However, a system where certain information can only be passed through one node, also is increasing the dependability of that node. However, the latter would not be a very resilient design.

Furthermore, systems such as anomaly detection have as key fuctionality to detect behaviour that is 'not normal', to find intruders in a network (Patcha & Park, 2007). Hence, these anomaly detection systems see if the behaviour is predictable. The drawback of anomaly detection is the notion of so called false positives. With that, we mean that the system might detect an anomaly that is in fact the actual behaviour, and not caused by an intruder.

Lastly, faith (the belief that the other node will be able to deal with new situation) could be included by valuing the effectiveness of securing information and communication channels of the other node. If the security measures are extensive in the other node, the faith that the node will successfully deal with intruders and false information increases. Hence, if we would design a system that values the security measures of the other node, values the proximity (dependability) to the other node, and that would detect anomalies at the other node, a measure for trust in the other node could be established. If this trust falls under a certain threshold, the information channel is no longer accepted. The advantage of this system is that only neighbouring nodes are taken into account, but by incorporating it in every node, a notion of trust occurs in the whole network.

Now, let us continue with combining the two into networks between humans and computers. Gligor and Wing (2011) have made a serious attempt to build a theory of trust in networks of humans and computers. They point out the same two difficulties as we did above, namely: 1. are the sender and receiver really who they say they are, and 2. is the line of communication really trustworthy? To focus on the first, several practical solutions are in place already. For example, when booking a hotel through a website such as booking.com, it is common to read the reviews of previous customers, to see if the hotel really delivers the quality it is promising to deliver. Furthermore, part of the reason we use a website such as booking.com is that we have faith in this website, and believe that the hotels on the website really exist, will process your booking, and deliver at least a minimum quality. However, when not talking about mass products, but individual communication, this is much harder to establish.

17.1.1. ICN-like Trust Compared to Trust Between Humans

In this section, we turn the way of thinking around. This time, we do not start with the notion of trust in human collaborative networks, which we then try to map on an ICN. Instead, we start with the notion of trust in an internet network and try to find analogue patterns in communication between humans.

Let us turn our attention to the anomaly detection again. The principle in internet and human networks is similar; when something happens we would not expect to happen we pay extra attention to the matter to see if we can trust it. As banks are responsible for the damage of skimming these days, the anomaly detection on your credit card is rather well developed. when a transaction is outside the scope of regular spendings, banks immediately call the owner of the credit card to check upon the transaction. Hence, due to an anomaly, mistrust arises in either the user or the network (is your credit card skimmed, or is the network hacked?). As a results extra steps are being taken to check if the anomaly was intended or not.

Humans do the same kind of things. If you hear something that sounds unrealistic (hence an anomaly) we first check with the node that provides us with the information (are you sure this is the right information?). Secondly, humans often check who the source of information is (who told you that?). Sometimes we get convinced that, even though the information sounds unrealistic, it is probably true, as we trust either the source or the messenger of the information well enough. However, we might also decide to take further steps to check upon the new information.

However, humans are (just like machines) often fooled by our weak anomaly detection. When we click on a phishing mail, because we do not pay attention to the sender, our anomaly detection fails, due to inattention. We let down our anomaly detection. However, we need to distinguish here between trust and anomaly detection, because clicking on a link like this, usually does not happen because we trust the sender, but because we do not pay attention. This is where computers could assist humans in their anomaly detection. At TNO, research is being done about how computers can assist humans in their personal anomaly detection.

17.2. Network Ties in ICN and Human Collaborative Networks

Network ties in human collaborative networks are analogue to links in an ICN. The links in the ICN we have considered are static in nature. With that, we mean that the links remain the same all the time, and except from when one would manually add links to the network, the set of links remains the same. In a human collaborative network, however, the links are more fluid in nature. Not only do the links change over time (due to i.e. friendships becoming less intense), but people tend to also search for new ties if the situation asks for it. As a result, the set of network ties in a human collaborative network is more dynamic than the set of links in ICNs we have considered.

However, some research has been done about mobile devices in ICN. Mobile devices do not have a steady connection to the network. Think about your own mobile telephone. Sometimes, you are not connected to the network with your phone at all, because you are abroad[1] or because you ran out of streaming data for the month. Sometimes, you use your phone in another city than your home, with that connecting to the network at different places. Hence, in an ICN that would allow for mobile devices to connect to the network, the links are more dynamic.

Xylomenos et al. (2014) mention some difficulties with these so called 'non-fixed hosts'. First of all, if the devices are not always connected to the network, they can not serve well as a source of information. One of the assumptions we had made is that the information always remains available in the sources. Secondly, some problems with the different locations of the device might occur.

Xylomenos et al. also elaborate on a possible solution for that. In ICN the so-called publisher/subscribe communication model is used for managing requests. Someone interested in information subscribes to the information, whereas someone having it available publishes that it is available. Note that the latter is just an announcement of availability, no data is transferred yet. In the network, so-called brokers match the subscriptions and the publishers. This way, there is a decoupling between wanting to have information (subscribe) and having information to share (publish). We will come back to that when discussing eagerness and willingness to share information.

[1]Although European legislation will change in June 2017, and streaming prices across the entire union will be equal, this would still hold for being outside of Europe

The dynamic state of the network ties of a human collaborative network have their analogue in allowing mobile devices in the ICN. The same type of issues occur in human collaborative networks as do so in ICN. It is very hard to know what information is where in the network, and who is in need for information. It would be interesting to think of a collaboration support system that has the publisher/subscribe communication model at its basis. A difficulty here is, however, that the strength of the model lies in the fact that all communication needs and information possibilities are known to this system. In human communication, it is hard to think of an efficient system that would trace all information. However, it could be feasible (even though it would loose some of its value) to think about a system that does this on a slightly higher level, with some kind of database that would suggest the right person for you, given some keywords.

17.3. Centrality in ICN and Human Collaborative Networks

In an ICN, the most central nodes are ideally the sources of information. This is because the information is closest to the end users that way. However, in the hierarchical networks considered in this thesis, the most central nodes are usually somewhere in the set of routers. In a human collaborative network, the most central nodes are often also taking the role of routers, connecting one person to another for information.

The centrality is closely connected to the network topology, of which we have seen in the mathematical part of this thesis that is has a large influence on the information spread and resilience in an ICN. In social networks, we often make use of the hubs of the network (people with a high centrality), by asking them who to ask for certain information. In that way, centrality plays a large role in our search for information. In an ICN however, the notion of centrality never plays an active role in linking a publisher to a subscriber.

In the previous section (Section 17.2), we mentioned a system where the publisher and the subscriber of information in human collaborative networks would get linked. Without having this system, this role is often taken by the hub in the network. This is a self-sustaining mechanism; the more the hub is used for these purposes, the more the hub will be able to link the right people to each other. However, this is not a very resilient solution, as the effects would be large when this one person (the hub) withdraws from the network. Hence, it would be interesting in human collaborative networks, to minimize the impact of the hubs on the network. In an ICN, this is done by establishing the routing schemes on the basis of shortest paths rather than central nodes. Note here, however, that the shortest path might (and depending on the topology and centrality measure chosen, also will) pass the most central nodes.

17.4. Willingness to Share

It has been shortly touched upon, but Xylomenos et al. describes the publisher/subscribe communication model to explain how ICN connects the requester of information (subscriber) to the sources of information (publisher). One of the factors that we have considered in this thesis is the willingness to share. Recall that willingness to share is 'the extent to which an individual is prepared to grant other group members access to his or her individual intellectual capital'. This is closely related to the function of the publisher in an ICN. The intellectual capital is the information one source or router possesses, and publishing that it is available would be a sign of willingness to share it.

There is also a large difference between the willingness to share and the publisher. In an ICN, the source or router will always be willing to share/ forward the information that is requested. In a human collaborative network, all kinds of office politics, interpersonal problems or other things might influence the willingness to share. Furthermore, the general willingness to share differs from person to person. We have measured the latter in the experiments.

Now, what would happen if we would replace the publishing feature of ICN with willingness to share in the human meaning of the word? The effect would be that, although some information is available, the node could 'decide' not to share it. However, this interferes greatly with the information transparency that is key in the ICN. Deciding to not announce the availability of information seems to be of no added value to the current information centric networks. Even though this use of the analogy does not lead to new insights at this moment, it is a good example of how the analogy can change the perspective on both ICNs and human collaborative networks.

17.5. Eagerness to Share: Pushing Information in an ICN

A lot of attention has been paid to the difference between pushed and pulled information. It was linked to eagerness and willingness to share, and it was found to be one of the two main issues that prevent us from measuring information resilience in a human collaborative network. Now that the analogy is in place, we can use it to think about the effects of adding the eagerness to share to the ICN, and therefore push the information through an ICN.

If information would be pushed through an ICN, the caches would be filled by an algorithm that is steered by the sources, instead of the information requests (as done by the end users). In an ICN simulator, the workload and popularity distribution of the content is known. If this would be the case in a 'real' ICN (hence, no simulator), pushing the content to the caches might actually work very well. However, there is no such thing as an overlying all-knowing layer in ICN that knows the exact popularity of the content, in combination with which information will be requested by which end user when.

So pushing the information from an all-knowing situation would not make any sense. However, this is also not the way humans work. We do not tell information we have based on all information there exists. We make a judgement whether or not to tell something. This judgement is sometimes wrong too, which can lead to telling information the other party already knew, the other party is not interested in, or, worse, in a situation where it would have been better to tell something, but because of the judgement call made, it was not told. All in all, however, since human communication can be complemented by actually asking for information, a balance of information sharing is being established. Now, if we think of an analogue way of dividing pulling and pushing information through the ICN, we automatically come up with a new caching strategy. In this caching strategy, not only the end user requests determine the caching, but also the sources, and possibly also the caches itself. Imagine that a certain information request is arriving in a source often. Then this source could decide to send this piece of information down to all of its neighbouring nodes, letting it be cached there. Note that this is substantially different from other caching strategies considered in this thesis, as it allows to cache off the path of an information request. If another end user now requests the same information (which is apparently popular as it was requested from the source more often than the threshold), but through another path, it can already find its information in a cache.

To the best of our knowledge, a caching strategy where the caching is determined by both the end users requests and the sources is new. It has not yet been tested on effectiveness, but this could be done by extending a simulator. We see here a perfect example of how the analogy has been used to come up with an application in ICN, based on human behaviour.

17.6. Reciprocity and Liking in ICN

Reciprocity is, just as liking, much harder to express in an ICN than the concepts above. This is because reciprocity and liking are both states of mind, mostly connected with emotion.

Now, let us imagine that reciprocity would indeed play a role in ICN. That would mean that the action (categorized in positive and negative) of one node would influence the action of the other (positively or negatively). But that would suggest some kind of emotional state in the ICN. It is hard to imagine this in ICN, but it is possible in human-machine interaction. An example of that is a virtual research environment. To keep the interaction going, some kind of reward is needed[2], to prevent the participants from only 'taking' information, and not 'giving' any. Here, the digital part (whether in the shape of software, a digital platform or a network) of the interaction facilitates the reciprocity on the human side of the network. By giving credits for sharing the information, and taking credits from 'taking' it, the system appeals to our feeling of reciprocity.

[2]as we have seen in the C-LAB course of the science communication master in 2016 on virtual research environments

We could take this one step further, when individuals get represented by computer systems. In order for everyone to remain satisfied, a feeling of reciprocity is needed in these computer systems. This, however, does not imply two nodes of a network behaving reciprocal, but rather the human individuals represented by the computer system.

Note that reciprocity is a very broad concept. It is about how strongly you react on positive or negative reactions, but also about how much you would feel like returning a favour if somebody does something nice to you. Behind a computer screen, we do not always feel as if we should return a favour that we would always return in real life. To give an example of that, if you would need a favour from your friend, it is often much more effective to call him than sending an email or message is. Now, if the other people are not friends, but strangers, and we add to that the possibility of being anonymous online, we find this feeling of reciprocity is even less present. It would be interesting to research these effects with respect to reciprocity, and to design communication tools that would enhance the feeling of reciprocity among the users.

As liking is also an emotional state, it is equally hard to think of liking in an ICN as it is to think about reciprocity. We could think about only sharing information with those we like, and not with those we do not like. In that case, the publisher of information could add some kind of tag, that allows certain information to only be given to a particular set of other nodes. Routers could do the same. This could even be a different set of nodes per type of information. Even though this is analogue with the notion of liking in human collaborative networks, there is not yet any reason to design an ICN like that. As we see, for now, really using the factors reciprocity and liking in an analogue way (of thinking) in an ICN, is difficult.

17.7. Resilience and Collaboration

In this Section, we will not try to use the analogy to consider certain elements, but we will consider the analogy to consider the application of human collaborative networks as well as ICN on the notion of collaboration. We will do that by considering robust adaptive planning.

Human-Machine Interaction

Not in its current state, but eventually, the analogy and the insight on information sharing in human collaborative networks could be used to build networks in which both humans and computers play a role, in which the human-machine interaction is fine tuned, and it is not imminent whether the 'other side of the line' is in fact a human or a machine. The machines should be able to take the seven factors from literature into account, as a system that does not reward trust or has no notion of reciprocity will imminently fail to be human-like.

Using the analogy might be of added value here, as the analogy allows to think about human concepts in an ICN and vice versa. The analogy answers questions such as 'what is needed for humans to share information, and how can this be taken into account in a digital network'? In designing a system that would allow for a human machine interaction where the machines are hardly distinguished from humans, answering these questions is very important.

Robust Adaptive Planning

In robust adaptive planning, the aim is to make a planning (or to decide upon a planning strategy) that is as good as possible, given deep uncertainties. Or, as M. C. van der Sanden and de Vries (2016) put it: " these strategies might not have the best possible option available as any one outcome, but their satisfactory outcomes occur in the largest range of future contingencies". Furthermore, the word adaptive indicates that, after some time, the situation is evaluated, the current situation as well as possibly other newly available information is taken into account, and if necessary, the planning is adapted based on this new information. The process is very iterative in nature.

Multiple tools for Robust adaptive planing have been developed. Bankes, Lempert, and Popper (2001), for example, have developed a tool called computer assisted reasoning, of which the basic idea is "to use simulation models to create a large database of plausible future scenarios where each entry in the database represents one guess about how the world works and one choice among many alternative strategies we might adopt to influence the world". In other words, with the use of such a database, the decision maker is assisted in having an overview of all possible scenarios. With help of the software, knowledge can than be extracted from the network. The more data can be added to the database, the more scenarios possible. One of the types of data that can be added to this database, is the social values, of which our better understanding in the shape of the analogy might be applicable and of added value.

Also at Delft University of Technology, a workbench called Exploratory Modelling and Analysis (EMA) is developed to offers "computational decision support for decision making under deep uncertainty" (Kwakkel, 2015). The analogy we have build in this thesis could be of added value when adding social factors to this software.

We are not familiar with this workbench, but we are able to reason about a possible workbench that would include ICN and human collaborative networks. For that, we specify ourselves towards decision making with respect to a human collaborative network, given the deep uncertainty that is essential in RAP. Part of these uncertainties can be captured and simulated with in an ICN. One could build a system around this ICN simulator, that does use the analogy as a way of thinking. Systems such as EMA could be used to find input parameters for the ICN simulation. From there, a simulation could be used to find a set of possible outcomes from within the team, that can be used in the decision making (planning) process. Here, taking the 7 human factors from our theoretical framework that influence information sharing in a human collaborative network into account seems is key here.

The research has not developed far enough to really incorporate the analogy in workbenches such as EMA. However, we would strongly encourage this to be researched.

18

Synergy Between the Two Research Directions

> Things have changed — now people focus on big problems, and if you go for a big problem you need to be interdisciplinary.
>
> Theodore Brown, vice-chancellor for research at the University of Illinois

In this chapter we will pay attention to the synergy between the two researches. Why did we perform two researches in the first place, and how do they coherent? The two research fields have a different culture, with different approaches and different foci. In our opinion, the two fields can learn from each other.

18.1. Coherence of the Research

The research has a common starting point, namely ICN. This (future) paradigm in internet networks has recently been researched extensively. In the communication part of this research we have considered human collaborative networks, and compared them to ICNs.

In ICN, the enhancement of resilience is hardly ever quantified, but it is mentioned often as one of the advantages of ICN. In the mathematical part of this research, one particular part of resilience, information resilience, has been quantified. Furthermore, with the help of the experiments, we have tried to apply this notion of information resilience to the human collaborative network as well, and to measure the information resilience of the network. As we have seen, this is not possible in its current shape.

As we can see, the two researches are applied on different fields, but are concerned about the same two concepts. Where scientific knowledge has been added to the mathematical world in the mathematical part of this research, the science communication part has potential added value for both the science communication field and the mathematical field. On the one hand, we have contributed to the cybernatic tradition in science communication, with considering human information sharing as a system. By using the analogy, we can think about this system from a different perspective, with that adding awareness on different patterns in the human collaborative network. An example of that has been found in Chapter 17, where we have applied the notion of anomaly detection on a human collaborative network.

On the other hand, the added value of the communication part of the research to the mathematical field lies in the application of specific human patterns/behaviour to the mathematical concepts behind ICN. As an example, in Chapter 17, we have used the notion of pulling and pushing information in human collaborative networks to create a first idea of a new caching strategy, where not only the behaviour of the end users determines the location of caching, but also the sources can push information in the network to be cached.

18.2. Added Value of Interdisciplinary Research

In this section, we reflect on the added value of doing interdisciplinary research. We answer questions such as: What is interdisciplinary research, why and when do we want to perform interdisciplinary research, what are challenges that we did experience, and how did we overcome these challenges? This results in what we experienced as the added value of the interdisciplinary research.

According to the National Academy of Sciences (2005), "Interdisciplinary research is a mode of research by teams or individuals that integrates information, data, techniques, tools, perspectives, concepts, and/or theories from two or more disciplines or bodies of specialized knowledge to advance fundamental understanding or to solve problems whose solutions are beyond the scope of a single discipline or area of research practice." The disciplines however, might be rather related as well as very different in nature. At Delft University of Technology, interdisciplinary research between closely related disciplines is rather common. For example, graduation projects on the boundary of quantum computing and mathematics or on the boundary of aerodynamics and mathematical analyses have recently been performed. However, the less related the two research fields are, the harder it becomes to perform true interdisciplinary research. At the Science Education and Communication department of the same university, this interdisciplinary research is much more common.

But why do we want to perform interdisciplinary research? Although a lot of scientific and quasi-scientific research has been written about interdisciplinary research, surprisingly little has been written about the reasons to perform this type of research. Several quasi-scientific articles note the necessity of solving challenges that transcend the notion of one disciplinary, such as global warming, or the energy transition. Gardy and Brinkman (2003) put the necessity in comparison to the metaphor about the elephant in pieces. "Ask someone to tell you the story of the blind men and the elephant, and they'll tell you a tale of six men, each of whom touched a different part of an elephant, unable to see what their hands were resting on. [...] The six men argued among themselves–was it a snake, a cow, a piece of rope? Only when they worked together, sharing their different ideas and experiences, were they able to discover the truth. Well, it seems as if scientists across the world are increasingly getting the point of that story."

In our personal opinion, true interdisciplinary research allows for an original point of view on certain research problems. In this thesis, and in particulary in Chapter 17, we have attempted to view Information Centric Networks from a social science point op view, and, vice versa, we considered human collaborative networks from and ICN point of view. This has lead to many new insights and points of view. However, we have experienced some difficulties in doing so.

The largest difficulty can be explained by a personal anecdote. Before the start of the project, we were looking for a thesis topic that could integrate both research fields: mathematics (optimization) and Science Communication. In talking to different teacher, proffesionals and professors, the following two type of things were said to me.

> "This mathematical problem is very interesting, you should consider it as a topic for your thesis. Also, in solving the problem, you will have to talk with several people, and explain the research to them. So it is also a communication topic, right?"
>
> "This communication problem is very interesting, you should consider it as a topic for your thesis. Also, in solving the problem, you will have to calculate things, so it is also a mathematics topic, right?"

The examples above show the incomprehension between the two research fields, even though they are performed at the same university. After a talk with some of my supervisors, some mutual understanding was brought into place, but this did certainly not mean that they understood each other fully. As a researcher, I experienced the continuous trouble of a lack of knowledge and experience with each of my supervisors.

The issues mentioned above, are not the only type of barriers and challenges in performing interdisciplinary research. Pellmar and Eisenberg (2000) point out three types of barriers of interdisciplinary research, namely attitudinal barriers, communication barriers and academic and professional barriers. Especially when social sciences meets natural sciences, attitude is sometimes a barrier. Or, as Viseu (2015) puts it "Too many in the physical and life sciences dismiss social sciences as having a 'service' role, being allowed to observe what they do but not disturb it". In our opinion, Viseu mentions a very valid point here.

Also the communication barrier as mentioned by Pellmar and Eisenberg is a barrier that we have experienced. Not only in the use of jargon, but also in the different styles of communicating the research results through a paper or thesis. Let us give two particular examples of that. Within the communications field, the authors are required to make a clear distinction between results and interpretation of results, to prevent misunderstandings. In mathematics, this misunderstanding does not naturally occur, and as such, the distinction is not made that clearly. The second example lies in the difference in length and number of words used to make a point in both research fields, which was hard to adjust to for some of the committee members. These differences are interesting, need some minor adjustments sometimes, but did not cause any real problems.

In our belief, some interdisciplinary research makes any researcher more all round and open to other research disciplines. Especially when the two research disciplines are very far apart in nature. Where the mathematical part of this thesis was substantively very challenging, the communications part was very challenging at another level. To find yourself a good story with a clear backup from literature that aligns at all parts of the research is much harder in social science than it is in mathematics.

18.3. Two Different Research Fields

Interdisciplinary research does not only add to the competences and insights of the researcher. As the two different fields have different approaches and different cultures, they can also learn from each other. In this Section, we will use our own experience to describe differences in the research fields. By pointing out the differences, we aim to also show the directions in which the research fields can learn from each other.

Mathematics is rather black and white in its nature. Either one is able to proof something, in which case it is true, or it is not true because it is not proven. Of course this blurs a little bit when we talk about simulations or approximations. The beauty of mathematics, namely the possibility of proving things wrong or right, makes mathematical research very valid. Science communication on the other hand, sometimes struggles with validity. It is sometimes hard to conclude that one is really measuring what one wants to measure. Furthermore, the experiments are often hard to repeat. However, when the aim of the research is to enhance insights (like in this research), or to qualitatively investigate certain relations, this is much less of a problem.

Duncan Watts, a true interdisciplinary researcher (physics in background) known for his research on the small world networks[1], stated in De Volkskrant in 2016 (Berg, 10 september 2016):

> Ik dacht ook ooit: ik lees de krant, ik kan wel socioloog worden. Maar die natuurkundige arrogantie heb ik lang geleden afgezworen. Mensen denken vaak: het is toch geen raketwetenschap? Nee, in die zin: het is veel ingewikkelder dan raketwetenschap. De sociale wereld is veel complexer dan de fysieke wereld. Het is een rommeltje, alles loopt in elkaar over.

This view on the difference between social science and natural science from Watts is a view we strongly relate to. On the one hand, mathematical research is often more valid than social science, and a clearer distinction between good and bad science/results can be made. On the other hand, this

[1] These type of mathematical graphs have such a structure that there are some large hubs.

is the case because social behaviour is so much harder to understand than physical behaviour (at least at this point in time), which makes the essence of social science much more difficult.

In the research performed here, the method of researching for the mathematical part became clear after some weeks of reading literature. For the science communication part, however, about 6 or 7 months were used to design the research. Questions such as 'what is the proper method to use for this problem', or 'where does my research idea connect to the scientific knowledge available at this point' are much harder to answer in social sciences than they are in mathematics.

We felt that doing science communication research has been more difficult than performing mathematical research. We do not think that many people would believe that up front, as many of us think that 'they can be sociologists because they read newspapers'. Unravelling this mess that human behaviour in all its facets (sociology, anthropology, psychology, science communication etcetera) is, is something that scientists have not even come close to yet. This makes it so difficult to position a social research, and to speak about wrong or right.

With that in mind, the sometimes condescending behaviour of mathematicians (or any natural scientist) towards social scientists is completely inappropriate. However, we want to make a remark here. Because it is so hard to proof right or wrong in social sciences, sloppy science is probably present here more than it is in natural sciences. Sloppy science here could relate to bad statistics, but also bad validity or sloppy experiments accompanied with shady conclusions. It is hard to proof someone wrong in social sciences (also because experiments are so difficult to repeat), and sloppy science gets mixed with thorough (and valid) science more easily than it happens in natural sciences. This feeds the feeling that every hypotheses one would have in social science could also be found proven one way or another. This obviously does not add to the respect there is about social scientists.

In our believe, it would be of added value if both research fields would understand the complexity of doing research in the other field. It seems that interdisciplinary research becomes more and more important, due to increasing complexity on the issues that society faces. These issues have a social component, but also need technological solutions. If we want to establish truly interdisciplinary teams, there is a larger need of mutual understanding than there is a need for researchers that are able to perform both types of research.

18.3.1. Robustness from Two Perspectives

We want to use one term from both research fields, and show how the different meaning is exemplary for the different positions the two fields are in. In robust optimization, a solution of a set of constraints is called a robust solution if it satisfies all possible realisations of the constraints, given the uncertainty set (Ben-Tal, El Ghaoui, & Nemirovski, 2009). In other words, a solution is called robust if it is a good solution, even if there are uncertainties in the constraints. In Section 17.7, we have discussed the notion of robust adaptive planning, and its connection towards the analogy set up in this research. In robust adaptive planning (RAP), the idea is similar to robust optimization, but the uncertainties are much complexer. Lempert, Popper, and Bankes (2002) call a solution robust if it "performs reasonably well, compared to the alternatives, over a wide range of plausible scenarios." The notion of robustness here is obviously similar. However, in robust optimization, we are able (or allow ourselves) to quantify the uncertainty in the constraints. This quantification can become rather complex (for example no convex constraints), but once the uncertainty in the constraints is set, it is 'only' left to solve the set of equations and find a robust solution. Note that the latter can be extremely complex to solve, but the problem is clearly defined and the boundaries of the problem are clear.

If we turn our attention to the definition of a robust solution in RAP, we deal with a lot more uncertainties and undefined (and unquantified) items. First of all, what is a plausible scenario? But also, RAP deals with 'deep uncertainty' (Chandrasekaran, 2008), where the uncertainty goes beyond the known unknowns towards the unknown unknowns. This is very hard to quantify into a constraint that would be solvable by robust optimization.

We see that the notions of robustness in both research fields are similar, but what it encompasses differs greatly. Science communication attempts to understand the system, and get to a solution starting from a larger set of uncertainties. Mathematics, on the other hand, tries to wrap the uncertainties into constraints, and start the research from there.

References

Adamic, L. A., & Huberman, B. A. (2002). Zipf's law and the internet [Journal Article]. *Glottometrics*, *3*(1), 143-150.

Akhavan, P., & Hosseini, S. M. (2015). Determinants of knowledge sharing in knowledge networks: a social capital perspective [Journal Article]. *IUP Journal of Knowledge Management*, *13*(1), 7.

Akhavan, P., & Hosseini, S. M. (2016). Social capital, knowledge sharing, and innovation capability: an empirical study of r & d teams in iran [Journal Article]. *Technology Analysis & Strategic Management*, *28*(1), 96-113. (Cx2io Times Cited:0 Cited References Count:68)

Alavi, M., & Leidner, D. E. (2001). Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues [Journal Article]. *MIS quarterly*, 107-136.

Albert, R., Jeong, H., & Barabási, A.-L. (2000). Error and attack tolerance of complex networks [Journal Article]. *nature*, *406*(6794), 378-382.

Al-Naday, M. F., Reed, M. J., Trossen, D., & Yang, K. (2014). Information resilience: source recovery in an information-centric network [Journal Article]. *IEEE Network*, *28*(3), 36-42.

Amaral, A., Fernandes, G., & Varajão, J. (2015). Identifying useful actions to improve team resilience in information systems projects [Journal Article]. *Procedia Computer Science*, *64*, 1182-1189.

Arlitt, M., Cherkasova, L., Dilley, J., Friedrich, R., & Jin, T. (2000). Evaluating content management techniques for web proxy caches [Journal Article]. *ACM SIGMETRICS Performance Evaluation Review*, *27*(4), 3-11.

Bankes, S. C., Lempert, R. J., & Popper, S. W. (2001). Computer-assisted reasoning [Journal Article]. *Computing in Science & Engineering*, *3*(2), 71-77.

Barabási, A. L. (2013). Network science [Journal Article]. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *371*(1987). (Cited By :34 Export Date: 26 August 2016)

Baran, P. (1964). On distributed communications networks [Journal Article]. *IEEE transactions on Communications Systems*, *12*(1), 1-9.

Bavelas, A. (1948). A mathematical model for group structures [Journal Article]. *Human organization*, *7*(3), 16-30.

Beck, R., Stenum Czepluch, J., Lollike, N., & Malone, S. (2016). Blockchain–the gateway to trust-free cryptographic transactions [Journal Article]. *ECIS 2016 Proceedings*, 1-14.

Ben-Tal, A., El Ghaoui, L., & Nemirovski, A. (2009). *Robust optimization* [Book]. Princeton University Press.

Berg, J. V. d. (10 september 2016). *De sociale wereld is een rommeltje* [Newspaper Article].

Bonacich, P., & Lloyd, P. (2001). Eigenvector-like measures of centrality for asymmetric relations [Journal Article]. *Social networks*, *23*(3), 191-201.

Boyden, A. (1947). Homology and analogy. a critical review of the meanings and implications of these concepts in biology [Journal Article]. *American Midland Naturalist*, 648-669.

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details [Journal Article]. *Proceedings of the National Academy of Sciences*, *105*(38), 14325-14329.

Breslau, L., Cao, P., Fan, L., Phillips, G., & Shenker, S. (1999). Web caching and zipf-like distributions: Evidence and implications [Conference Proceedings]. In *Infocom'99. eighteenth annual joint conference of the ieee computer and communications societies. proceedings. ieee* (Vol. 1, p. 126-134). IEEE.

Brockner, J., Siegel, P. A., Daly, J. P., Tyler, T., & Martin, C. (1997). When trust matters: The moderating effect of outcome favorability [Journal Article]. *Administrative Science Quarterly*, 558-583.

Brynielsson, J., Högberg, J., Kaati, L., Mårtenson, C., & Svenson, P. (2010). Detecting social positions using simulation [Conference Proceedings]. In *2010 international conference on advances in social network analysis and mining, asonam 2010* (p. 48-55).

Brynielsson, J., Kaati, L., & Svenson, P. (2012). Social positions and simulation relations [Journal Article]. *Social Network Analysis and Mining*, *2*(1), 39-52. (Cited By :8 Export Date: 10 August 2016 Correspondence Address: Kaati, L.; Swedish Defence Research AgencySweden)

Buckland, M. K. (1991). Information as thing [Journal Article]. *Journal of the American Society for Information Science (1986-1998)*, *42*(5), 351.

Carofiglio, G., Morabito, G., Muscariello, L., Solis, I., & Varvello, M. (2013). From content delivery today to information centric networking [Journal Article]. *Computer Networks*, *57*(16), 3116-3127.

Carrington, P. J., Scott, J., & Wasserman, S. (2005). *Models and methods in social network analysis* (Vol. 28) [Book]. Cambridge university press.

Cassi, L., Corrocher, N., Malerba, F., & Vonortas, N. (2008). Research networks as infrastructure for knowledge diffusion in european regions [Journal Article]. *Econ. Innov. New Techn.*, *17*(7-8), 663-676.

Cavalcante, E., Pereira, J., Alves, M. P., Maia, P., Moura, R., Batista, T., … Pires, P. F. (2016). On the interplay of internet of things and cloud computing: A systematic mapping study [Journal Article]. *Computer Communications*.

Chandrasekaran, B. (2008). *Designing decision support systems to help avoid biases & make robust decisions, with examples from army planning* (Report). DTIC Document.

Che, H., Tung, Y., & Wang, Z. (2002). Hierarchical web caching systems: Modeling, design and experimental results [Journal Article]. *IEEE Journal on Selected Areas in Communications*, *20*(7), 1305-1314.

Chiu, C.-M., Hsu, M.-H., & Wang, E. T. (2006). Understanding knowledge sharing in virtual communities: An integration of social capital and social cognitive theories [Journal Article]. *Decision support systems*, *42*(3), 1872-1888.

Chuang, S.-S., Chen, K.-S., & Tsai, M.-T. (2015). Exploring the antecedents that influence middle management employees' knowledge-sharing intentions in the context of total quality management implementations [Journal Article]. *Total Quality Management & Business Excellence*, *26*(1-2), 108-122.

cisco. (2008). *Example of an hierarchical network* [Web Page]. Retrieved from http://www.cisco.com/c/en/us/td/docs/solutions/Enterprise/Campus/HA_campus_DG/hacampusdg.html

Cisco. (2012). Network management solution: Optimize infrastructure for iptv services [Journal Article]. *[online]*. Retrieved from http://www.cisco.com/en/US/technologies/tk869/tk769/technologies_white_paper0900aecd80730d28.html

Clark, M. C., & Payne, R. L. (1997). The nature and structure of workers' trust in management [Journal Article]. *Journal of Organizational Behavior*, 205-224.

Cohen, R., Erez, K., Ben-Avraham, D., & Havlin, S. (2000). Resilience of the internet to random breakdowns [Journal Article]. *Physical review letters*, *85*(21), 4626.

Collins, N. L., & Miller, L. C. (1994). Self-disclosure and liking: a meta-analytic review [Journal Article]. *Psychological bulletin*, *116*(3), 457.

Compeau, D. R., & Higgins, C. A. (1995). Computer self-efficacy: Development of a measure and initial test [Journal Article]. *MIS quarterly*, 189-211.

Cooke, N. J., & Gorman, J. C. (2006). Assessment of team cognition [Journal Article]. *International encyclopedia of ergonomics and human factors*, 270-275.

Cross, R., Parker, A., Christensen, C. M., Anthony, S. D., & Roth, E. A. (2004). *The hidden power of social networks* [Book]. Audio-Tech Business Book Summaries, Incorporated.

Cummings, L. L., & Bromiley, P. (1996). The organizational trust inventory (oti) [Journal Article]. *Trust in organizations: Frontiers of theory and research*, *302*(330), 39-52.

CWI. (2008). *Twintig jaar internet in europa.* (Vol. 2016) (Web Page No. 27/7).

Davenport, T. H., & Prusak, L. (1998). *Working knowledge: How organizations manage what they know* [Book]. Harvard Business Press.

Davison, B. D. (2001). A web caching primer [Journal Article]. *IEEE Internet Computing*, *5*(4), 38-45.

Derks, D. (2017a). *Feedback on experiment* [Interview].

Derks, D. (2017b, January 26th). *On the setup of the experiment* [Interview].

Derks, D. (2017, May 30th). *Interpretation of results* [Interview].

De Vries, R. E., Van den Hooff, B., & de Ridder, J. A. (2006). Explaining knowledge sharing the role of team communication styles, job satisfaction, and performance beliefs [Journal Article]. *Communication research*, *33*(2), 115-135.

Díez-Vial, I., & Montoro-Sánchez, . (2014). Social capital as a driver of local knowledge exchange: A social network analysis [Journal Article]. *Knowledge Management Research and Practice*, *12*(3), 276-288. (Academy of Management Perspectives, 24 (1), pp. 62-77)

Dierks, T. (2008). The transport layer security (tls) protocol version 1.2 [Journal Article].

Dodds, P. S., Watts, D. J., & Sabel, C. F. (2003). Information exchange and the robustness of organizational networks [Journal Article]. *Proceedings of the National Academy of Sciences*, *100*(21), 12516-12521.

Duprez, C., Christophe, V., Rimé, B., Congard, A., & Antoine, P. (2015). Motives for the social sharing of an emotional experience [Journal Article]. *Journal of Social and Personal Relationships*, *32*(6), 757-787.

Erlang, A. K. (1909). The theory of probabilities and telephone conversations [Journal Article]. *Nyt Tidsskrift for Matematik B*, *20*(33-39), 16.

Falk, A., & Fischbacher, U. (2006). A theory of reciprocity [Journal Article]. *Games and economic behavior*, *54*(2), 293-315.

Fall, K. R., & Stevens, W. R. (2011). *Tcp/ip illustrated, volume 1: The protocols* [Book]. addison-Wesley.

Fishbein, M., & Ajzen, I. (1977). Belief, attitude, intention, and behavior: An introduction to theory and research [Journal Article].

Fisher, K. E., & Naumer, C. M. (2006). Information grounds: Theoretical basis and empirical findings on information flow in social settings [Book Section]. In *New directions in human information behavior* (p. 93-111). Springer.

Freeman, L. C. (1977). A set of measures of centrality based on betweenness [Journal Article]. *Sociometry*, 35-41.

Freeman, L. C. (1978). Centrality in social networks conceptual clarification [Journal Article]. *Social networks*, *1*(3), 215-239.

Fricker, C., Robert, P., & Roberts, J. (2012). A versatile and accurate approximation for lru cache performance [Conference Proceedings]. In *Proceedings of the 24th international teletraffic congress* (p. 8). International Teletraffic Congress.

Gao, J., Barzel, B., & Barabási, A.-L. (2016). Universal resilience patterns in complex networks [Journal Article]. *Nature*, *530*(7590), 307-312.

Gardy, J., & Brinkman, F. (2003). The benefits of inter-disciplinary research: Our experience with pathogen bioinformatics [Journal Article]. *Science Next Wave*.

Garetto, M., Leonardi, E., & Martina, V. (2016). A unified approach to the performance analysis of caching systems [Journal Article]. *ACM Transactions on Modeling and Performance Evaluation of Computing Systems*, *1*(3), 12.

Gentner, D. (1983). Structure □ mapping: A theoretical framework for analogy [Journal Article]. *Cognitive science*, *7*(2), 155-170.

Gentner, D., Bowdle, B., Wolff, P., & Boronat, C. (2001). Metaphor is like analogy [Journal Article]. *The analogical mind: Perspectives from cognitive science*, 199-253.

Giethoorn, B. (2016). *Working session with b. giethoorn* [Interview].

Gill, A. S., D'Acunto, L., Trichias, K., & van Brandenburg, R. (2016). Bidcache: Auction-based in-network caching in icn [Conference Proceedings]. In *Globecom workshops (gc wkshps), 2016 ieee* (p. 1-6). IEEE.

Gillespie, N. (2003). Measuring trust in working relationships: The behavioral trust inventory [Conference Proceedings]. In *Academy of management conference, seattle, wa.*

Gligor, V., & Wing, J. M. (2011). Towards a theory of trust in networks of humans and computers [Conference Proceedings]. In *International workshop on security protocols* (p. 223-242). Springer.

Gomes, J. O., Borges, M. R., Huber, G. J., & Carvalho, P. V. R. (2014). Analysis of the resilience of team performance during a nuclear emergency response exercise [Journal Article]. *Applied ergonomics*, *45*(3), 780-788.

Gottesman, D., Lo, H.-K., Lutkenhaus, N., & Preskill, J. (2004). Security of quantum key distribution with imperfect devices [Conference Proceedings]. In *Information theory, 2004. isit 2004. proceedings. international symposium on* (p. 136). IEEE.

Hacco, E., & Shu, L. H. (2002). Biomimetic concept generation applied to design for remanufacture [Conference Proceedings]. In *Asme 2002 international design engineering technical conferences and computers and information in engineering conference* (p. 239-246). American Society of Mechanical Engineers.

Haythornthwaite, C. (1996). Social network analysis: An approach and technique for the study of information exchange [Journal Article]. *Library & information science research*, *18*(4), 323-342.

Home, J. F., & Orr, J. E. (1997). Assessing behaviors that create resilient organizations [Journal Article]. *Employment Relations Today*, *24*(4), 29-39.

Huff, L., & Kelley, L. (2003). Levels of organizational trust in individualist versus collectivist societies: A seven-nation study [Journal Article]. *Organization Science*, *14*(1), 81-90.

Ipe, M. (2003). Knowledge sharing in organizations: A conceptual framework [Journal Article]. *Human Resource Development Review*, *2*(4), 337-359.

Iyer, S., Rowstron, A., & Druschel, P. (2002). Squirrel: A decentralized peer-to-peer web cache [Conference Proceedings]. In *Proceedings of the twenty-first annual symposium on principles of distributed computing* (p. 213-222). ACM.

Kim, S., & Lee, H. (2006). The impact of organizational context and information technology on employee knowledge□sharing capabilities [Journal Article]. *Public Administration Review*, *66*(3), 370-385.

King, D. D., Newman, A., & Luthans, F. (2015). Not if, but when we need resilience in the workplace [Journal Article]. *Journal of Organizational Behavior*.

Kingsley, Z. G. (1949). *Human behavior and the principle of least effort: an introduction to human ecology* [Book]. Addison-Wesley Press.

Kleinrock, L. (1961). Information flow in large communication nets [Journal Article]. *RLE Quarterly Progress Report*, *1*.

Kurose, J. (2014). Information-centric networking: The evolution from circuits to packets to content [Journal Article]. *Computer Networks*, *66*, 112-120.

Kwakkel, J. (2015). *Exploratory modelling and analysis (ema) workbench* [Generic].

Laoutaris, N., Che, H., & Stavrakakis, I. (2006). The lcd interconnection of lru caches and its analysis [Journal Article]. *Performance Evaluation*, *63*(7), 609-634.

Larzelere, R. E., & Huston, T. L. (1980). The dyadic trust scale: Toward understanding interpersonal trust in close relationships [Journal Article]. *Journal of Marriage and the Family*, 595-604.

Lee, B., Jeon, H., Yoon, S., & Song, H. (2012). Towards a cdn over icn [Conference Proceedings]. In *Dcnet/ice-b/optics* (p. 46-51).

Lempert, R., Popper, S., & Bankes, S. (2002). Confronting surprise [Journal Article]. *Social Science Computer Review*, *20*(4), 420-440.

Lerman, K., Ghosh, R., & Surachawala, T. (2012). Social contagion: An empirical study of information spread on digg and twitter follower graphs [Journal Article]. *arXiv preprint arXiv:1202.3162*.

Littlejohn, S. W., & Foss, K. A. (2010). *Theories of human communication* [Book]. Waveland Press.

Lua, E. K., Crowcroft, J., Pias, M., Sharma, R., & Lim, S. (2005). A survey and comparison of peer-to-peer overlay network schemes [Journal Article]. *IEEE Communications Surveys & Tutorials*, *7*(2), 72-93.

Mayer, R. C., & Davis, J. H. (1999). The effect of the performance appraisal system on trust for management: A field quasi-experiment [Journal Article]. *Journal of applied psychology*, *84*(1), 123.

McAllister, D. J. (1995). Affect-and cognition-based trust as foundations for interpersonal cooperation in organizations [Journal Article]. *Academy of management journal*, *38*(1), 24-59.

Michalski, R., Palus, S., & Kazienko, P. (2011). *Matching organizational structure and social network extracted from email communication* (Vol. 87 LNBIP) [Generic]. Springer Verlag.

Mitchell, J. C. (1969). *The concept and use of social networks* [Book]. Bobbs-Merrill.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the prisma statement [Journal Article]. *Annals of internal medicine*, *151*(4), 264-269.

Nahapiet, J., & Ghoshal, S. (1998). Social capital, intellectual capital, and the organizational advantage [Journal Article]. *Academy of management review*, *23*(2), 242-266.

National Academy of Sciences. (2005). *Facilitating interdisciplinary research* [Book]. Washington, DC: The National Academies Press. Retrieved from https://www.nap.edu/catalog/11153/facilitating-interdisciplinary-research doi: doi:doi:10.17226/11153

Onnela, J.-P., Saramäki, J., Hyvönen, J., Szabó, G., De Menezes, M. A., Kaski, K., … Kertész, J. (2007). Analysis of a large-scale weighted network of one-to-one human communication [Journal Article]. *New Journal of Physics*, *9*(6), 179.

Otte, E., & Rousseau, R. (2002). Social network analysis: a powerful strategy, also for the information sciences [Journal Article]. *Journal of information Science*, *28*(6), 441-453.

Pan, J., Paul, S., & Jain, R. (2011). A survey of the research on future internet architectures [Journal Article]. *IEEE Communications Magazine*, *49*(7), 26-36.

Patcha, A., & Park, J.-M. (2007). An overview of anomaly detection techniques: Existing solutions and latest technological trends [Journal Article]. *Computer networks*, *51*(12), 3448-3470.

Pellmar, T. C., & Eisenberg, L. (2000). Barriers to interdisciplinary research and training [Journal Article].

Pentland, A. (2004). Learning communities—understanding information flow in human networks [Journal Article]. *BT Technology Journal*, *22*(4), 62-70.

Perugini, M., Gallucci, M., Presaghi, F., & Ercolani, A. P. (2003). The personal norm of reciprocity [Journal Article]. *European Journal of Personality*, *17*(4), 251-283.

Pilerot, O. (2012). Lis research on information sharing activities - people, places, or information [Journal Article]. *Journal of Documentation*, *68*(4), 559-581. (Cited By :8 Export Date: 30 August 2016)

Podlipnig, S., & Böszörmenyi, L. (2003). A survey of web cache replacement strategies [Journal Article]. *ACM Computing Surveys (CSUR)*, *35*(4), 374-398.

Psaras, I., Chai, W. K., & Pavlou, G. (2012). Probabilistic in-network caching for information-centric networks [Conference Proceedings]. In *Proceedings of the second edition of the icn workshop on information-centric networking* (p. 55-60). ACM.

Rak, J., Niedermayer, H., Papadimitriou, D., & Romero, P. (2016). Information-driven network resilience: Research challenges and perspectives [Journal Article]. *Optical Switching and Networking*.

Rempel, J. K., Holmes, J. G., & Zanna, M. P. (1985). Trust in close relationships [Journal Article]. *Journal of personality and social psychology*, *49*(1), 95.

Robinson, S. L. (1996). Trust and breach of the psychological contract [Journal Article]. *Administrative science quarterly*, 574-599.

Rosensweig, E. J., Kurose, J., & Towsley, D. (2010). Approximate models for general cache networks [Conference Proceedings]. In *Infocom, 2010 proceedings ieee* (p. 1-9). IEEE.

Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust [Journal Article]. *Journal of personality*, *35*(4), 651-665.

Rubin, Z. (1970). Measurement of romantic love [Journal Article]. *Journal of personality and social psychology*, *16*(2), 265.

Saino, L., Psaras, I., & Pavlou, G. (2014). Icarus: a caching simulator for information centric networking (icn) [Conference Proceedings]. In *Proceedings of the 7th international icst conference on simulation tools and techniques* (p. 66-75). ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

Sanders, K., Schyns, B., Dietz, G., & Den Hartog, D. N. (2006). Measuring trust inside organisations [Journal Article]. *Personnel review*, *35*(5), 557-588.

Scott, J. (2012). *Social network analysis* [Book]. Sage.

Seville, E., Brunsdon, D., Dantas, A., Le Masurier, J., Wilkinson, S., & Vargo, J. (2011). *Building organisational resilience: a new zealand approach* [Generic].

Shankar Sankaran, P., Rodríguez-Sánchez, A. M., & Vera Perea, M. (2015). The secret of organisation success: a revision on organisational and team resilience [Journal Article]. *International Journal of Emergency Services*, *4*(1), 27-36.

Shockley-Zalabak, P., Ellis, K., & Winograd, G. (2000). Organizational trust: What it means, why it matters [Journal Article]. *Organization Development Journal*, *18*(4), 35.

Smith, B., Tooley, E., Christopher, P., & Kay, V. (2010). Resilience as the ability to bounce back from stress: A neglected personal resource? [Journal Article]. *The Journal of Positive Psychology*, *5*(3), 166-176.

Sourlas, V., Tassiulas, L., Psaras, I., & Pavlou, G. (2015). Information resilience through user-assisted caching in disruptive content-centric networks [Conference Proceedings]. In *Ifip networking conference (ifip networking), 2015* (p. 1-9). IEEE.

Spreitzer, G. M., & Mishra, A. K. (1999). Giving up control without losing control: Trust and its substitutes' effects on managers' involving employees in decision making [Journal Article]. *Group & Organization Management*, *24*(2), 155-187.

Stasser, G., & Titus, W. (2003). Hidden profiles: A brief history [Journal Article]. *Psychological Inquiry*, *14*(3-4), 304-313.

Stephenson, K. (2005). *Trafficking in trust* [Generic].

Stephenson, K., & Zelen, M. (1989). Rethinking centrality: Methods and examples [Journal Article]. *Social networks*, *11*(1), 1-37.

Stone, W. (2001). Measuring social capital [Journal Article]. *Australian Institute of Family Studies, Research Paper*, *24*.

Swan, M. (2015). *Blockchain: Blueprint for a new economy* [Book]. " O'Reilly Media, Inc.".

Tichy, N. M., Tushman, M. L., & Fombrun, C. (1979). Social network analysis for organizations [Journal Article]. *Academy of management review*, *4*(4), 507-519.

Tyler, T. R. (2003). Trust within organisations [Journal Article]. *Personnel review*, *32*(5), 556-568.

Tyson, G., Bodanese, E., Bigham, J., & Mauthe, A. (2014). Beyond content delivery: Can icns help emergency scenarios? [Journal Article]. *IEEE Network*, *28*(3), 44-49.

Tzafrir, S. S., & Dolan, S. L. (2004). Trust me: a scale for measuring manager-employee trust [Journal Article]. *Management Research: Journal of the Iberoamerican Academy of Management*, *2*(2), 115-132.

Vakali, A., & Pallis, G. (2003). Content delivery networks: Status and trends [Journal Article]. *IEEE Internet Computing*, *7*(6), 68-74.

van den Hooff, B., & Hendrix, L. (2005). Eagerness and willingness to share: The relevance of different attitudes towards knowledge sharing [Journal Article]. *OKLC*.

van der Sanden, M., & Flipse, S. M. (2016). A cybernetic dream: how a crisis in social sciences leads us to a communication for innovation-laboratory [Journal Article]. *Democracy*, *28*, 07.

van der Sanden, M. C., & de Vries, M. J. (2016). Innovation in science and technology education and communication through design thinking [Book Section]. In *Science and technology education and communication* (p. 129-145). Springer.

Viseu, A. (2015). Integration of social science into research is crucial: social scientists must be allowed a full, collaborative role if researchers are to understand and engage with issues that concern the public [Journal Article]. *Nature*, *525*(7569), 291-292.

V.N.I, C. (2016). *Cisco visual networking index:forecast and methodology, 2015–2020* (Report).

Wang, D., Wen, Z., Tong, H., Lin, C.-Y., Song, C., & Barabási, A.-L. (2011). Information spreading in context [Conference Proceedings]. In *Proceedings of the 20th international conference on world wide web* (p. 735-744). ACM.

Waxman, B. M. (1988). Routing of multipoint connections [Journal Article]. *IEEE journal on selected areas in communications*, *6*(9), 1617-1622.

Widén-Wulff, G. (2007). Motives for sharing: Social networks as information sources." [Journal Article]. *Advances in Library Administration and Organization*, *25*, 1-31.

Windle, M. (2002). Critical conceptual and measurement issues in the study of resilience [Book Section]. In *Resilience and development* (p. 161-176). Springer.

Wittenbaum, G. M., Hollingshead, A. B., & Botero, I. C. (2004). From cooperative to motivated information sharing in groups: Moving beyond the hidden profile paradigm [Journal Article]. *Communication Monographs*, *71*(3), 286-310.

Wolff, R. W. (1982). Poisson arrivals see time averages [Journal Article]. *Operations Research*, *30*(2), 223-231.

Wu, L., Chuang, C.-H., & Hsu, C.-H. (2014). Information sharing and collaborative behaviors in enabling supply chain performance: A social exchange perspective [Journal Article]. *International Journal of Production Economics*, *148*, 122-132.

Xylomenos, G., Ververidis, C. N., Siris, V. A., Fotiou, N., Tsilopoulos, C., Vasilakos, X., … Polyzos, G. C. (2014). A survey of information-centric networking research [Journal Article]. *IEEE Communications Surveys & Tutorials*, *16*(2), 1024-1049.

Yang, J., Wang, J., Wong, C. W., & Lai, K.-H. (2008). Relational stability and alliance performance in supply chain [Journal Article]. *Omega*, *36*(4), 600-608.

V

Appendices

A

Systematic Literature Review

The following keywords are used in the systematic literature review.

Primary keywords

- Information spread
- Information sharing
- TRA (Theory of reasoned action)
- Social cognitive theory
- Team resilience
- Social network analyses

Secondary keywords

- Behaviour
- Human networks
- Social networks
- Group decision making
- Collaborative information activities
- Knowledge sharing
- Social capital
- Collaborative sencemaking
- Communication
- Sociology
- Psychology
- Information deficiency
- Knowledge conversion
- Team performance
- Information resilience
- Adaption, autonomy
- Social adjustment
- Collaborative behaviour

keywords	keywords	note	identification records	records screened	records excluded
Information spread	Theory of reasoned action	exclude computer science	9	4	3
Information spread	Social cognitive theory	exclude engineering, computer science and medicine	31	6	5
Information spread	Team resilience		7	1	1
Theory of reasoned action	Social cognitive theory	exclude medicine and computer science	42	0	0
Theory of reasoned action	Team resilience		0		
Social cognitive theory	Team resilience		2	0	
articles found by searching for article that cite wittenbaum			3	3	3
Information spread	human network*	limit to social sciences, exclude medicin and computer science and arts and humanities and environmental sciences	26	2	0
Information sharing	human network*	search within documents of communication, limit to social sciences, exclude medicin and computer science and arts and humanities and environmental sciences	61	6	5
information deficienc* AND resilien*		limit to social science	7	0	most results about HIV
deficienc* AND "team performance" AND information			8	2	0
"knowledge conversion" AND resilien*			1	1	0
"information resilien*" AND team performance			0		

Figure A.1: What items have been considered

information resilience AND team performance			23		
information resilience AND team performance		limit to social science	5	0	
resilien* AND team performance		limit to social science	28	3	
social adjustment AND information sharing		limit to social science	15	0	
collaborative behaviour AND "information sharing" And social		limit to social science	23	0	
"information sharing" AND "collaborative behaviour"			9	2	
"collaborative behaviour" AND resilien*			6	0	
"information sharing"AND resilien* and organi*			37	3	
resilien* AND organi* AND team		limit to social science	75	7	
ïnformation resilienc*"		limit to social science	4	2	

Figure A.2: What items have been considered

B

Literature List

Here the articles selected from the systematic literature review are being depicted.

- Akhavan, P., & Hosseini, S. M. (2016). Social capital, knowledge sharing, and innovation capability: an empirical study of R & D teams in Iran. Technology Analysis & Strategic Management, 28(1), 96-113.
- Amaral, A., Fernandes, G., & Varajão, J. (2015). Identifying Useful Actions to Improve Team Resilience in Information Systems Projects. Procedia Computer Science, 64, 1182-1189.
- Barth, S., Schraagen, J. M., & Schmettow, M. (2015). Network measures for characterising team adaptation processes. Ergonomics, 58(8), 1287-1302.
- Baumann, M. R., & Bonner, B. L. (2013). Member awareness of expertise, information sharing, information weighting, and group decision making. Small Group Research, 44(5), 532-562.
- Chiu, C.-M., Hsu, M.-H., & Wang, E. T. (2006). Understanding knowledge sharing in virtual communities: An integration of social capital and social cognitive theories. Decision support systems, 42(3), 1872-1888.
- Chow, W. S., & Chan, L. S. (2008). Social network, social trust and shared goals in organizational knowledge sharing. Information & Management, 45(7), 458-465.
- Chuang, S.-S., Chen, K.-S., & Tsai, M.-T. (2015). Exploring the antecedents that influence middle management employees' knowledge-sharing intentions in the context of total quality management implementations. Total Quality Management & Business Excellence, 26(1-2), 108-122.
- Cooke, N. J., & Gorman, J. C. (2006). Assessment of team cognition. International encyclopedia of ergonomics and human factors, 270-275.
- De Dreu, C. K., & Beersma, B. (2010). Team confidence, motivated information processing, and dynamic group decision making. European Journal of Social Psychology, 40(7), 1110-1119.
- De Vries, R. E., Van den Hooff, B., & de Ridder, J. A. (2006). Explaining knowledge sharing the role of team communication styles, job satisfaction, and performance beliefs. Communication research, 33(2), 115-135.
- Edson, M. C. (2012). A complex adaptive systems view of resilience in a project team. Systems Research and Behavioral Science, 29(5), 499-516.
- Gomes, J. O., Borges, M. R., Huber, G. J., & Carvalho, P. V. R. (2014). Analysis of the resilience of team performance during a nuclear emergency response exercise. Applied ergonomics, 45(3), 780-788.
- Hassan, E. (2015). information sharing framework for national resilience.

- Janhonen, M., & Johanson, J.-E. (2011). Role of knowledge conversion and social networks in team performance. International Journal of Information Management, 31(3), 217-225.
- Karlova, N. A., & Fisher, K. E. (2013). Plz RT": A social diffusion model of misinformation and disinformation for understanding human information behaviour. Inform Res, 18(1), 1-17.
- Kim, S., & Lee, H. (2006). The impact of organizational context and information technology on employee knowledge□sharing capabilities. Public Administration Review, 66(3), 370-385.
- King, D. D., Newman, A., & Luthans, F. (2015). Not if, but when we need resilience in the workplace. Journal of Organizational Behavior.
- Lloyd, A. (2013). Building information resilient workers: The critical ground of workplace information literacy. What have we learnt? Paper presented at the European Conference on Information Literacy.
- Meneghel, I., Salanova, M., & Martínez, I. M. (2016). Feeling good makes us stronger: How team resilience mediates the effect of positive emotions on team performance. Journal of Happiness Studies, 17(1), 239-255.
- Pederson, J. R., & McLaren, R. M. (2015). Managing information following hurtful experiences How personal network members negotiate private information. Journal of Social and Personal Relationships, 0265407515612242.
- Pentland, A. (2004). Learning communities—understanding information flow in human networks. BT Technology Journal, 22(4), 62-70.
- Shankar Sankaran, P., Rodríguez-Sánchez, A. M., & Vera Perea, M. (2015). The secret of organisation success: a revision on organisational and team resilience. International Journal of Emergency Services, 4(1), 27-36.
- Sozen, H. C. (2012). Social networks and power in organizations: A research on the roles and positions of the junior level secretaries in an organizational network. Personnel review, 41(4), 487-512. doi:10.1108/00483481211229393
- van der Kleij, R., Molenaar, D., & Schraagen, J. M. (2011). Making teams more resilient: Effects of shared transformational leadership training on resilience. Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.
- Widén-Wulff, G. (2007). Motives for sharing: Social networks as information sources.". Advances in Library Administration and Organization, 25, 1-31.
- Wittenbaum, G. M., Hollingshead, A. B., & Botero, I. C. (2004). From cooperative to motivated information sharing in groups: Moving beyond the hidden profile paradigm. Communication Monographs, 71(3), 286-310.
- Wu, L., Chuang, C.-H., & Hsu, C.-H. (2014). Information sharing and collaborative behaviors in enabling supply chain performance: A social exchange perspective. International Journal of Production Economics, 148, 122-132.

C

Schematic overview of information sharing

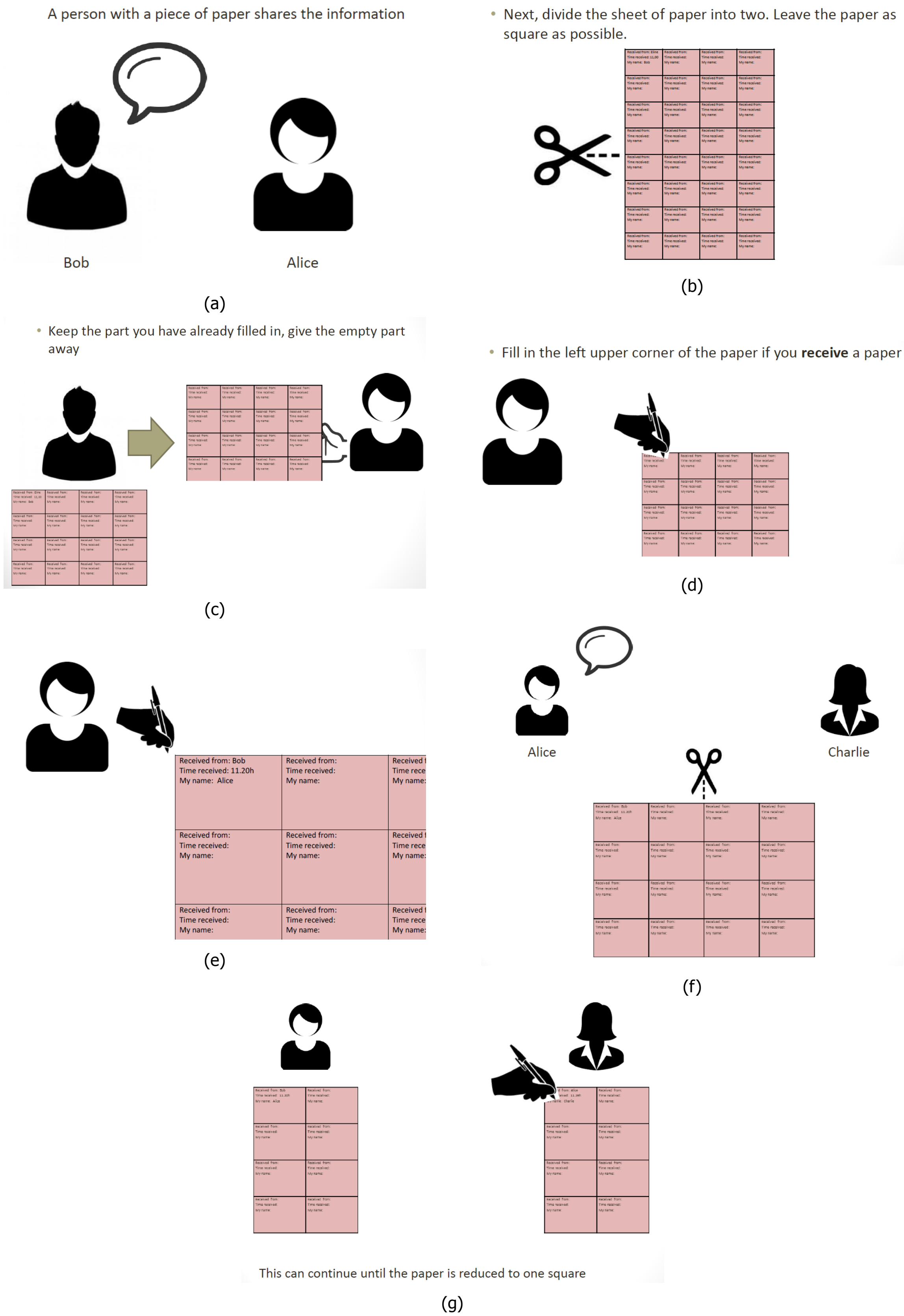


Figure C.1: Schematic view of sharing information in the first experiment

D

Assignments in Second Experiment

Assignments in the first part of the second assignment

1. Probeer uit te vinden welke persoon er net als jij een ster op zijn kaartje heeft. Zorg ervoor dat maximaal 3 andere mensen te weten komen dat jij dat zoekt.

2. Zorg ervoor dat de knikkerbaan een vluchtelement (de knikker komt even los (vliegen) van de baan) heeft. Zorg ervoor dat maximaal 3 andere mensen te weten komen dat dit jouw opdracht is.

3. Zoek uit wat de persoonlijke opdracht van alle anderen is. Zorg ervoor dat maximaal 3 andere mensen te weten komen dat dit jouw opdracht is.

4. Zorg ervoor dat de plastic knikkerbaan die onderdeel is van het bouwmateriaal **niet** gebruikt wordt. Zorg ervoor dat maximaal 3 andere mensen te weten komen dat dit jouw opdracht is.

5. Zorg ervoor dat de plastic knikkerbaan die onderdeel is van het bouwmateriaal gebruikt wordt. Zorg ervoor dat maximaal 3 andere mensen te weten komen dat dit jouw opdracht is.

6. Probeer uit te vinden welke persoon er net als jij een vierkant op zijn kaartje heeft. Zorg ervoor dat maximaal 3 andere mensen te weten komen dat jij dat zoekt.

7. Zorg ervoor dat het hoogste punt van de knikkerbaan minstens 160 cm van de grond is. Zorg ervoor dat maximaal 3 andere mensen te weten komen dat dit jouw opdracht is.

8. Zorg ervoor dat de knikkerbaan breder wordt dan hoog. Zorg ervoor dat maximaal 3 andere mensen te weten komen dat dit jouw opdracht is.

Assignments in the second part of the second experiment

1. Probeer uit te vinden welke persoon er net als jij een driehoek op zijn kaartje heeft. Zorg ervoor dat maximaal 3 andere mensen te weten komen dat jij dat zoekt.

2. Zorg ervoor dat er geen tafel gebruikt wordt als ondeel van (de fundering van) de knikkerbaan. Zorg ervoor dat maximaal 3 andere mensen te weten komen dat dit jouw opdracht is.

3. Zoek uit wat de persoonlijke opdracht van alle anderen is. Zorg ervoor dat maximaal 3 andere mensen te weten komen dat dit jouw opdracht is.

4. Zorg ervoor dat het knikkerpotje dat onderdeel is van het bouwmateriaal **niet** gebruikt wordt. Zorg ervoor dat maximaal 3 andere mensen te weten komen dat dit jouw opdracht is.

5. Zorg ervoor dat het knikkerpotje dat onderdeel is van het bouwmateriaal gebruikt wordt. Zorg ervoor dat maximaal 3 andere mensen te weten komen dat dit jouw opdracht is.

E

Questions About Trust

Table E.1: My caption

Nr	Question	Category
1.	We have a sharing relationship. We can both freely share our ideas, feelings and hopes.	?
2.	I can talk freely to this individual about difficulties I am having at work and know that (s)he will want to listen.	predictability
3.	**We would both feel a sense of loss if one of us was transferred and we could no longer work together.**	dependability
4.	**If I shared my problems with this person, I know that (s)he would respond constructively and caringly.**	predictability
5.	I would have to say that we have both made considerable emotional investments in our working relationship.	dependability
6.	This person approaches her/ his job with professionalism and dedication.	?
7.	**Given this person's track record, I see no reason to doubt her/his competence and preparation for the job.**	faith
8.	I can rely on this person not to make my job more difficult by careless work.	predictability
9.	Most people, even those who aren't close friends of this individual, trust and respect her/him as a co worker	?
10.	Other work associates of mine who must interact with this individual consider her/him to be trustworthy.	?
11.	If people knew more about this individual and her/his background, they would be more concerned and monitor her/his performance more closely.	predictability

F

Questions of the Evaluation After the Second Experiment

Table F.1: Table with results per evaluation question

nr.	Vraag	A	B	C	D	E	F	G	H
1	Ik had genoeg tijd voor alle onderdelen en vragenlijsten van het experiment	6	7	7	7	6	6	7	7
2	Ik vond het leuk om de knikkerbaan te bouwen	5	6	7	7	7	7	7	5
3	Ik vond het moeilijk om me te concentreren op mijn persoonlijke opdracht	2	5	2	6	5	3	1	5
4	Ik vond het duidelijk wat er van mij verwacht werd	6	7	7	7	7	5	7	7
5	Mijn persoonlijke opdracht was te moeilijk	1	2	3	4	5	3	2	3
6	Mijn persoonlijke opdracht was te makkelijk	5	3	4	4	2	4	5	1
7	Ik had geen motivatie om mijn persoonlijke opdracht uit te voeren	2	1	1	3	3	1	1	5
8	Door mijn persoonlijke opdracht vergat ik om actief bij te dragen aan de bouw van de knikkerbaan	2	1	2	1	2	2	5	1
9	Ik vond de uitleg vooraf duidelijk	3	7	6	7	7	6	7	7
10	Ik ben tevreden met het resultaat van de knikkerbaan	7	5	7	6	6	7	7	7
11	Ik vond de vragenlijst die ik vooraf in moest vullen (via google form) erg duidelijk	6	7	5	5	7	7	7	5
12	Ik vond de vragen uit de vragenlijst die ik vooraf in moest vullen (via google form) erg ongemakkelijk	5	6	6	6	7	5	5	4
13	Ik vond het vervelend dat er de tweede keer mensen niet mee mochten doen met het experiment	3	1	3	2	3	2	5	5
14	Ik had liever gehad dat andere mensen dan nu het geval was niet mee mochten doen bij het tweede experiment	2	1	1	3	6	1	1	1
15	Door de vragenlijst vooraf heb ik me anders gedragen in het experiment	1	3	1	2	2	1	1	6
16	Ik vond de aanwezigheid van sommige mensen bij het experiment vervelend	1	1	1	3	1	1	1	1

G

Results of Experiment 2

Table G.1: Table of who send what information at the disturbed experiment

Who \ To whom told	A	B	C	F	G	succes?
A	5		ja			ja
B		1			ja	ja
C	ja		4	ja	ja	nee, verkeerd gelezen
F	gehoord		gehoord	3	ja	nee, david's opdracht niet
G	gehoord	ja	gehoord	ja	2	ja

Table G.2: Table of who received what information at the disturbed experiment

Who \ To whom told	A	B	C	D	E	F	G	H
A		ja, maar fout	ja, verteld			ja, maar fout	ja, geraden	
B	nee	nee	nee			nee	nee	
C	ja, verteld	nee				ja, maar fout	nee	
D	nee	ja, maar fout	nee			gokje, fout	ja, maar fout	
E	ja, maar fout	ja, geobserveerd	nee			ja, maar fout	nee	
F	ja, gevraagd	gokje, fout!	ja, maar fout door Frouke dr verkeerd gelezen opdracht				ja, gevraagd	
G	ja, maar fout	ja, maar fout	ja, verteld			ja (deels), door vertellen		
H	ja, maar fout!	ja, maar fout	ja, maar fout			ja maar fout	nee	

H

Assignments of the Third Experiment

Assignments in the first part of the third experiment

1. Try to find out who has, just like you, a star on his/her card. Make sure that you talk to at maximum three other people about your assignment. Other people can investigate for you. However, no more than 3 people should know that this is your assignment by the end of the building time.
2. Make sure that the marble track has a flying element (the marble 'flies' through the air for a bit). Make sure that you talk to at maximum three other people about your assignment. Other people can investigate for you. However, no more than 3 people should know that this is your assignment by the end of the building time.
3. Try to find out what the personal assignment of all the others is. Make sure that you talk to at maximum three other people about your assignment. Other people can investigate for you. However, no more than 3 people should know that this is your assignment by the end of the building time.
4. Make sure that the plastic marble track that is part of the building material is being used. Make sure that you talk to at maximum three other people about your assignment. Other people can investigate for you. However, no more than 3 people should know that this is your assignment by the end of the building time.
5. Try to find out who has, just like you, a square on his/her card. Make sure that you talk to at maximum three other people about your assignment. Other people can investigate for you. However, no more than 3 people should know that this is your assignment by the end of the building time.
6. Make sure that the highest point of the marble track is at least 200 cm above the floor. Make sure that you talk to at maximum three other people about your assignment. Other people can investigate for you. However, no more than 3 people should know that this is your assignment by the end of the building time.
7. Make sure that the marble track is wider than high. Make sure that you talk to at maximum three other people about your assignment. Other people can investigate for you. However, no more than 3 people should know that this is your assignment by the end of the building time.

Assignments in the second part of the third experiment

1. Try to find out who has, just like you, a triangle on his/her card. Make sure that you talk to at maximum 2 other people about your assignment. Other people can investigate for you. However, no more than 3 people should know that this is your assignment by the end of the building time.

2. Make sure that no table is being used in the (foundation of) the marble track. Make sure that you talk to at maximum 2 other people about your assignment. Other people can investigate for you. However, no more than 3 people should know that this is your assignment by the end of the building time.
3. Try to find out what the personal assignment of all the others is. Make sure that you talk to at maximum 2 other people about your assignment. Other people can investigate for you. However, no more than 3 people should know that this is your assignment by the end of the building time.
4. Make sure that the marble jar (knikkerpotje) is part of the building material used in the marble track. Make sure that you talk to at maximum 2 other people about your assignment. Other people can investigate for you. However, no more than 3 people should know that this is your assignment by the end of the building time.
5. Try to tell as many people about the assignments of others you have found out. Make sure that you talk to at maximum 2 other people about your assignment. Other people can investigate for you. However, no more than 3 people should know that this is your assignment by the end of the building time.