Complementarity and Similarity in Complex Networks

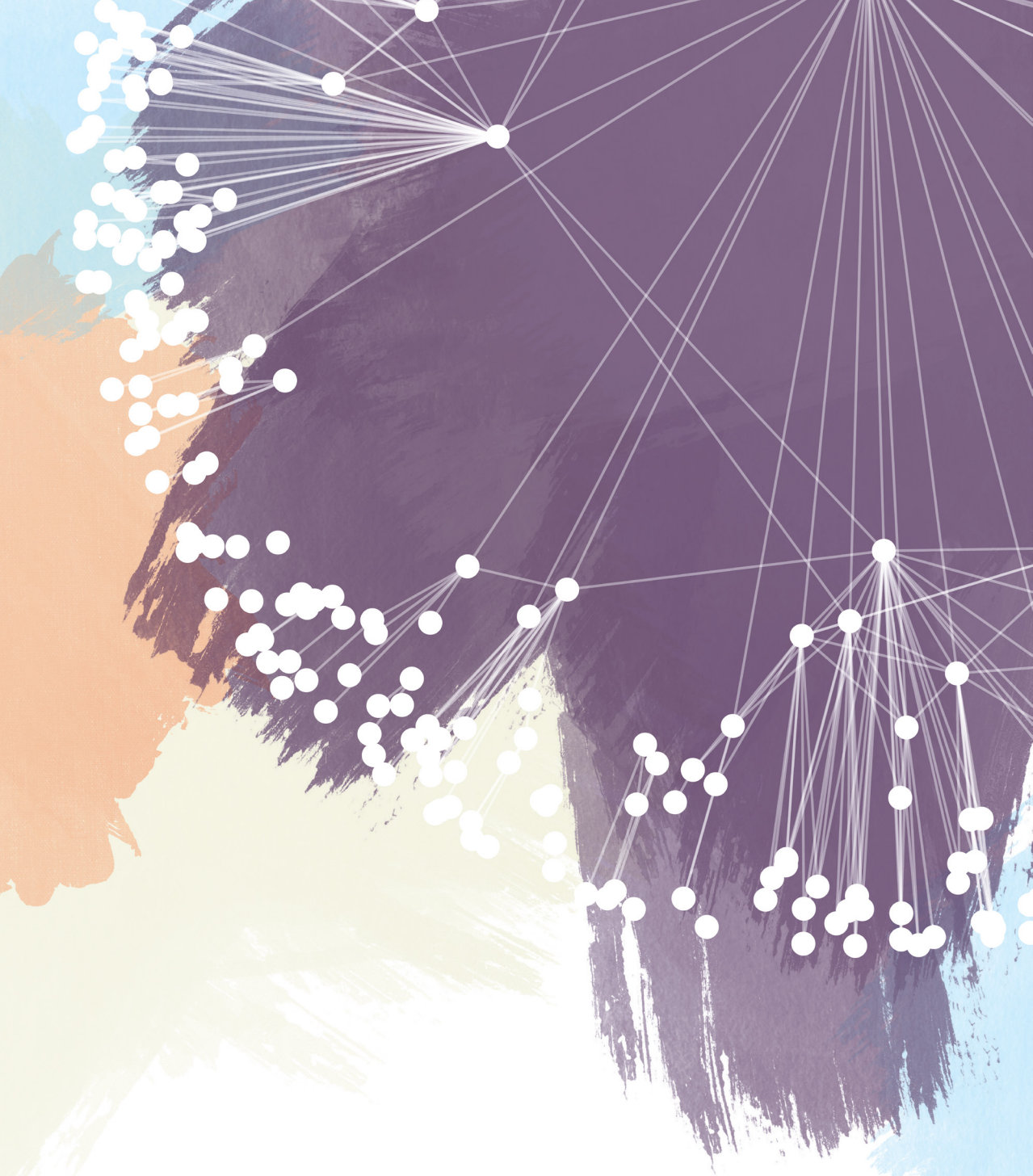
Budel, G.J.A.

DOI
10.4233/uuid:404db306-a85f-46ac-80d0-ba2a4f24a26e

Publication date
2024

Document Version
Final published version

Citation (APA)
Budel, G. J. A. (2024). *Complementarity and Similarity in Complex Networks*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:404db306-a85f-46ac-80d0-ba2a4f24a26e



Complementarity and Similarity in Complex Networks

Gabriel Budel

COMPLEMENTARITY AND SIMILARITY IN COMPLEX NETWORKS

COMPLEMENTARITY AND SIMILARITY IN COMPLEX NETWORKS

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof.dr.ir. T.H.J.J. van der Hagen,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op dinsdag 16 april 2024 om 10:00 uur

door

Gabriël Johannes Augustinus BUDEL

Master of Science in Econometrics & Management Science,
Erasmus Universiteit Rotterdam, Nederland
en
Master of Science in Finance & Investments,
Rotterdam School of Management, Erasmus University, Nederland
geboren te Rheden, Nederland.

Dit proefschrift is goedgekeurd door de

promotor: prof.dr.ir. P.F.A. Van Mieghem
copromotors: dr.ir. E. Smeitink, dr. M.A. Kitsak

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Prof.dr.ir. P.F.A. Van Mieghem,	Technische Universiteit Delft, *promotor*
Dr. M.A. Kitsak,	Technische Universiteit Delft, *copromotor*
Dr.ir. E. Smeitink,	Technische Universiteit Delft, KPN, *copromotor*
Onafhankelijke leden:	
Prof.dr.ir. R.E. Kooij,	Technische Universiteit Delft
Prof.dr. M. Boguñá Espinal	Universitat de Barcelona, Spain
Prof.dr. D. Krioukov	Northeastern University, Boston, USA
Dr. C. De Bacco,	Max Planck Institute, Germany
Prof.dr.ir. G. Gaydadjiev,	Technische Universiteit Delft, *reservelid*

Keywords:	Complex networks, complementarity, similarity, semantic networks, hyperbolic geometry, network clustering
Printed by:	Ipskamp Printing, Enschede
Cover and Illustrations:	Designed by Terri van der Zwan

ISBN 978-94-6366-845-3

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

CONTENTS

SUMMARY

Networks are becoming ever more important in today's highly interconnected society, from telecommunication networks and social networks to power grids and the Internet. The field of Network Science seeks to uncover structure within the complex topologies of networks and the processes that govern their link formation. Many methods and models in the field are founded on link-formation principles that are driven by similarity, drawing inspiration from social network theory. In this dissertation, we discuss various network representations based on similarity, and we introduce and illustrate an alternative link formation principle that is based on complementarity.

The first part of this dissertation focuses on clustering the nodes of a network or community detection. Here, the nodes of a network are partitioned into several clusters and the objective is to precisely determine the cluster memberships based on only the network topology. Many clustering methods assume that the *number* of clusters is known *a priori*. In Chapter 2, we investigate how exactly to find this number of clusters for a given graph. We discuss several modularity maximization and spectral clustering methods, and we outline how they can be used to find the number of clusters. We compare the performance of several different algorithms by evaluating these methods on benchmark graph models where the ground truth clusters are known.

In the second part, we explore network representations in the hyperbolic space. In Chapter 3, we extend the 2-dimensional random hyperbolic graph model to a hyperbolic space of arbitrary dimensionality. Our rescaling of the model parameters and variables casts the random hyperbolic graph model of any dimension to a unified mathematical framework, such that the degree distribution is invariant to the dimensionality of the space. We analyze the different connectivity regimes of the model and their limiting cases. In Chapter 4, we describe how hyperbolic graphs are built on a connection principle based on similarity, and we identify a class of real-world networks in which the links are driven by principles of complementarity rather than similarity. We propose a framework for embedding complementarity-driven networks into hyperbolic space and we describe the ensuing complementarity random hyperbolic graph model. In Chapter 5, we further investigate the topological properties of the complementarity random hyperbolic graph.

The third and final part of the dissertation centers on semantic networks, which describe semantic relations between words or concepts. In Chapter 6, we systematically analyze the topological properties of a large, multilingual dataset of semantic networks. Our investigation covers both universal and language-specific structural properties of these networks. We examine the roles that the connection principles of similarity and complementarity play in their link formation, and we discuss how a deeper understanding of these organizing principles benefits applications in natural language processing.

SAMENVATTING

In een steeds nauwer verbonden samenleving worden netwerken, variërend van telecommunicatienetwerken en sociale netwerken tot energienetten en het internet, almaar belangrijker. De netwerkwetenschap streeft ernaar structuur te ontdekken binnen de complexe topologieën van netwerken en de processen waarin hun verbindingen tot stand komen. Veel methoden en modellen in dit veld zijn gebaseerd op verbindingsprincipes die gedreven worden door patronen van gelijkenis, ingegeven door de theorie van sociale netwerken. In dit proefschrift bespreken we verschillende netwerkrepresentaties op basis van gelijkenis en introduceren we een alternatief verbindingsprincipe gebaseerd op complementariteit.

Het eerste deel van dit proefschrift richt zich op het clusteren van de nodes van een netwerk. Hier zijn de nodes van een netwerk opgedeeld in meerdere clusters en het doel is om de clusterlidmaatschappen nauwkeurig te bepalen op basis van enkel de netwerktopologie. Veel clusteringmethoden gaan ervan uit dat het *aantal* clusters *a priori* bekend is. In Hoofdstuk 2 onderzoeken we hoe we het aantal clusters in een graaf precies kunnen vinden. We bespreken verschillende methoden voor modulariteitsmaximalisatie en spectrale clustering, en we schetsen hoe deze kunnen worden gebruikt om het aantal clusters te vinden. We vergelijken de nauwkeurigheid van verschillende algoritmen door ze te evalueren op grafen waar de werkelijke clusters bekend zijn.

In het tweede deel verkennen we netwerkrepresentaties in de hyperbolische ruimte. In Hoofdstuk 3 breiden we de 2-dimensionale hyperbolische toevalsgraaf uit naar een hyperbolische ruimte van arbitraire dimensionaliteit. Onze transformatie van de variabelen en modelparameters brengt de hyperbolische toevalsgraaf van arbitrarire dimensionaliteit in een uniform wiskundig raamwerk, zodat de graadverdeling onafhankelijk is van de dimensionaliteit. We analyseren de verschillende connectiviteitsregimes van het model en hun limietgevallen. In Hoofdstuk 4 beschrijven we hoe hyperbolische grafen zijn gebaseerd op een verbindingsprincipe gebaseerd op gelijkenis, en we identificeren een klasse van netwerken waarin de verbindingen zijn gebaseerd op complementariteit in plaats van gelijkenis. We introduceren een raamwerk om deze op complementariteit gebaseerde netwerken in de hyperbolische ruimte in te bedden, en we beschrijven de resulterende complementaire hyperbolische toevalsgraaf. In Hoofdstuk 5 onderzoeken we de belangrijkste topologische eigenschappen van dit graafmodel.

Het derde en laatste deel van het proefschrift richt zich op semantische netwerken, welke semantische relaties tussen woorden of concepten beschrijven. In Hoofdstuk 6 voeren we een systematische analyse uit van de topologische eigenschappen van een grote, meertalige dataset van semantische netwerken. Ons onderzoek bestrijkt zowel universele als taalspecifieke eigenschappen van deze netwerken. We onderzoeken de rollen die de verbindingsprincipes van gelijkenis en complementariteit spelen in de totstandkoming van hun verbindingen en bespreken hoe een beter begrip van deze principes de natuurlijke taalverwerking (NLP) kan verbeteren.

1

INTRODUCTION

I would have written a shorter letter, but I did not have the time.

Blaise Pascal

COMPLEX networks are ubiquitous in today's increasingly interconnected world. With the rise of modern technologies such as the Internet, distances in our social networks have become shorter, and understanding the structures of the networked systems that connect us has become consistently more important [1]. Other real-world examples of complex networks include biological networks, telecommunication networks, transportation networks, brain networks, and power grids [2]. Typically, networks are separated into the topology of the underlying graph and the process that runs on top of the network [3].

Networks are said to be complex because they are not straightforward to model, due to the numerous interdependencies and non-linearities they contain. The field of Network Science aims to find structure in the seemingly chaotic nature of network topologies and processes. One notable approach is the study of graph spectra [4], which play a key role in, for example, the analysis of epidemic spreading [5]. In this dissertation, we discuss several network representations that are based on principles of similarity, and we identify and describe an alternative link formation principle that is based on complementarity.

1.1. NETWORK CLUSTERING

Social networks and several other types of networks often consist of communities: groups of densely connected similar nodes that together constitute the full network [6]. In community detection or network clustering, the goal is to recover the community memberships given the topology of the graph. Network clustering is an active area of research, with new approaches still appearing regularly [7, 8].

There exists a wide range of community detection algorithms: modularity maximization [9, 10], spectral clustering [4, 11], hierarchical clustering [12], statistical clustering [13–15], attraction-repulsion processes [8], and more [6]. The outcome of a community detection algorithm relies heavily on the definition of 'community', thus making the clustering problem scientifically ill-posed. How to define a community is driven by the motivation behind employing community detection: different motivations lead to different definitions, none of them is the best, but each of them is potentially useful for a specific goal [16]. Recently, it was reiterated by Fortunato and Newman [7] that community detection remains a challenging problem because of the lack of a general agreement on what is a community. Commonly used definitions of graphs with community structures are the stochastic block model (SBM) [17, 18] with a Poisson degree distribution, the degree-corrected stochastic block model (DCSBM) [19] with a given degree sequence, and the Lancichinetti–Fortunato–Radicchi benchmark [20] with a power-law degree distribution.

1.2. HYPERBOLIC GEOMETRY

Besides community structures, many large real-world networks exhibit characteristics such as sparse average degrees, power-law degree distributions, the small-world property, and large graph clustering coefficients [1–3]. These topological properties imply a latent hyperbolic geometry with constant negative curvature underlying the network [21]. Conversely, the random hyperbolic graph (RHG), a graph model where the nodes have a latent representation in the hyperbolic disk $\mathbb{H}^2$, possesses these exact properties [21–27]. Real-world networks can be mapped to hyperbolic spaces using the RHG as a null model [28–30]. Such hyperbolic network mappings, or hyperbolic embeddings, find applications in routing, shortest path finding, and navigation [28, 31–37], link prediction [29, 38–44], and network scaling [22, 25, 35], among others [45]. Because RHGs have large clustering coefficients that do not approach zero in the thermodynamic limit, connected nodes exhibit similar neighborhoods in the graph, which we refer to as 'similarity'.

1.3. SEMANTIC NETWORKS

Technological advances in recent years saw a vast increase in the availability of digital content, thereby amplifying the demand for computers to interpret and process unstructured texts written by humans. The field of Natural Language Processing (NLP) aims to confer to computers the capability to answer questions, analyze sentiment, retrieve information, and translate texts [46]. Semantic networks, or knowledge graphs, play a fundamental role in NLP by representing human knowledge in a structured way as a graph, for example by denoting 'is-a' or 'related-to' relationships between real-world concepts [47, 48]. The applications of semantic networks are numerous, as they enable enhanced Internet search [49], applications of artificial intelligence [50, 51], and digital assistants [52, 53]. Recent developments in Transformer-based large language models have brought about a significant leap forward in the effectiveness of computational text interpretation tasks [54, 55]. As Transformer models do not leverage semantic networks, they are not yet fully interpretable in their operations. Whereas many successful NLP

algorithms based on semantic networks have been proposed, there have been relatively few works focusing on the structure of these networks [56, 57].

1.4. NOTATION

The following notation is used throughout this dissertation. Vectors, multidimensional points, and matrices are printed in bold, while scalars are printed in Roman. The $N \times 1$ all-one vector is denoted by $\mathbf{u}$, and the $N \times N$ all-one matrix by $\mathbf{J}$. A graph G consists of N nodes and L links and is defined by the $N \times N$ adjacency matrix $\mathbf{A}$, where the element $a_{ij} = 1$ if nodes i and j are connected in G, and $a_{ij} = 0$ otherwise. The degree k_i of a node i is the i-th element of the degree vector $\mathbf{k} = \mathbf{Au}$, where the diagonal matrix $\mathbf{\Delta} = \text{diag}(\mathbf{k})$ contains the vector $\mathbf{k}$ on its main diagonal. We denote by $1_{\{x\}}$ the indicator function that equals 1 if statement x is true, and 0 otherwise. A random variable X is capitalized, and its expected value is denoted by $E[X]$, while its variance is denoted by $\text{Var}[X]$. In graphs with a community structure, the parameter c denotes the number of communities or clusters.

1.5. DOCUMENT STRUCTURE

This dissertation consists of three parts, which are further divided into chapters.

I. Network Clustering The first part covers methods and implications of clustering in complex networks or community detection. Chapter 2 discusses various approaches and models used for partitioning the nodes of a network. In specific, we discuss several modularity maximization and spectral clustering methods, while we focus on their ability to find the number of clusters in a graph. We investigate and compare the performance of several algorithms that estimate the number of clusters within a network by evaluating these methods on benchmark graph models such as the SBM and, in addition, several real-world networks.

II. Hyperbolic Network Representations The second part of this thesis explores hyperbolic representations of complex networks and related network models. In Chapter 3, we extend the 2-dimensional random hyperbolic graph model to a hyperbolic space of arbitrary dimensionality. We provide a rescaling of the model parameters that casts the random hyperbolic graph model of any dimension to a unified mathematical framework, such that the degree distribution is invariant to the dimensionality of the space. We provide a systematic analysis of the different connectivity regimes and their limiting cases for different dimensionalities of the hyperbolic space. In Chapter 4, we describe how the standard hyperbolic graph is built on a connection principle based on similarity, and we remark that there is a class of real-world networks in which the links are driven by principles of complementarity rather than similarity. We propose a framework for embedding complementarity-driven networks into hyperbolic space and derive the complementarity random hyperbolic graph model from our minimal complementarity embedding framework. We show how our embedding method improves upon similarity-based methods in link prediction tasks for complementarity-driven networks. Through the embedding of a semantic network of antonyms, we illustrate

how in our geometric embedding framework similarity is a special case of complementarity. We further investigate the topological properties of the complementarity random hyperbolic graph in Chapter 5. A good understanding of both the similarity and complementarity connection principles is fundamental for understanding the geometric and topological properties of complex networks.

III. Semantic Networks The final part of the dissertation focuses on semantic networks, examining their topological properties systematically and the significance of these networks in various applications. In Chapter 6, we study not only the general but also the language-specific structural properties of semantic networks. We show that the connection principles of similarity and complementarity both play a role in their link formation processes, and we discuss how a better understanding of these organizing principles can aid applications of natural language processing.

I

Network Clustering

2

DETECTING THE NUMBER OF CLUSTERS IN A NETWORK

I don't want to belong to any club that will accept me as a member.

Groucho Marx

Many clustering algorithms for complex networks depend on the choice for the number of clusters and it is often unclear how to make this choice. The number of eigenvalues located outside a circle in the spectrum of the non-backtracking matrix has been conjectured to be an estimator of the number of clusters in a graph. In this Chapter, we compare the estimate of the number of clusters obtained from the spectrum of the non-backtracking matrix with three estimators based on the concept of modularity and evaluate the methods on several benchmark graphs. We find that the non-backtracking method detects the number of clusters better than the modularity-based methods for the graphs in our simulation study, especially when the clusters have slightly different sizes. The estimates of the non-backtracking method are narrowly distributed around the true number of clusters for all benchmark graphs considered. Additionally, for graphs without a clustering structure, the non-backtracking method detects exactly one cluster, which is a convenient property for an estimator of the number of clusters. However, the lack of a well-defined concept of a cluster prevents sharp conclusions.

This Chapter is based on G. Budel and P. Van Mieghem (2021), Detecting the number of clusters in a network, Journal of Complex Networks, 8(6), cnaa047 [58].

2.1. INTRODUCTION

The detection of community structures in complex networks has been a popular topic in Network Science for many years [6]. Finding the number of communities or the number of clusters, however, receives comparably little attention. Many community detection algorithms require the number of communities as an input and their results depend on the chosen number of communities.

The number of communities found by a given algorithm depends on the definition of 'community'. The precise definition of a community is in turn driven by the motivation behind employing community detection: different motivations lead to different definitions, none of them is the best, but each of them is potentially useful for a specific goal [16]. Here, we adopt the clustering approach of finding groups of nodes that are 'similar' or 'close'. In the context of complex networks, similarity or closeness between nodes is often described by the number of links or the weights on the links in weighted networks. A cluster is then a group of nodes that is densely connected internally, while sparsely connected to the nodes of other groups. Figure 2.1 exemplifies the definition.

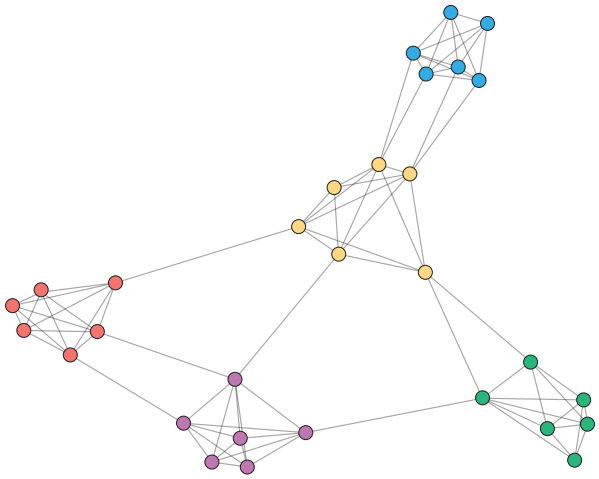


Figure 2.1: A network with 5 densely connected clusters of nodes. The network is drawn using the force-directed graph visualization algorithm in Matlab [59].

Initially, the research on clustering in complex networks focused on modularity optimization, initiated by Newman and Girvan [9, 60, 61]. The concept of modularity is naturally linked to the definition of the clustering problem: nodes that share more links than expected are considered to be part of one group, a cluster. While exact optimization of modularity is computationally intractable [4], many approximate algorithms were shown to achieve high modularity for some well-known real-world networks. Many heuristic modularity maximization algorithms do not require the number of clusters as prior knowledge; they discover the number of clusters in the network during the optimization.

Another popular and accurate type of clustering is spectral clustering. In spectral clustering, nodes are assigned to clusters based on the values of their corresponding

components in a subset of the eigenvectors of a matrix representation of the network, such as the Laplacian. However, most spectral clustering methods require the knowledge of the number of clusters. In real-world networks, the actual number of clusters is typically unknown, requiring researchers to explore a range of possible values.

Recently, a spectral clustering algorithm based on the non-backtracking matrix was shown to achieve optimal clustering results for graphs generated by the stochastic block model [62]. Additionally, Krzakala *et al.* [62] conjectured that the number of real eigenvalues of the non-backtracking matrix **H** that are separated from the bulk of the eigenvalues is an estimate of the number of clusters in the network. Our contribution consists of the assessment of the accuracy of the non-backtracking method in comparison with three modularity-based methods for estimating the number of clusters in a network. We find that the non-backtracking method outperforms the three other methods for all benchmark graphs considered in this work.

This Chapter is organized as follows. In Section 2.2, we provide a short description of several clustering techniques for complex networks, focusing on modularity optimization methods and spectral clustering. We describe the literature related to this work in Section 2.3. In Section 2.4, we describe the non-backtracking method and three modularity-based methods for detecting the number of clusters in a network. We compare the four detection methods on benchmark graphs in several simulation experiments and describe the experimental results in Section 2.5. We conclude our research and reflect upon our findings in Section 2.6.

2.2. CLUSTERING IN COMPLEX NETWORKS

2.2.1. MODULARITY MAXIMIZATION

The modularity of a graph is a measure for the quality of a given partition of a network based on the number of links between nodes belonging to the same cluster [4]. The modularity measure m proposed by Newman and Girvan [9] is defined as the difference between the actual number of intra-cluster links and the expected number of intra-cluster links if links were to be placed at random. The expected number of links between node i and node j if links are placed randomly is $\frac{k_i k_j}{2L-1}$, or, when L is large, approximately $\frac{k_i k_j}{2L}$. The modularity m is calculated as

$$m = \sum_{i=1}^{N} \sum_{j=1}^{N} \left(a_{ij} - \frac{k_i k_j}{2L} \right) 1_{\{i \text{ and } j \text{ belong to the same cluster}\}}. \tag{2.1}$$

A modularity close to 1 indicates a strong modular structure, while a modularity of 0 indicates that the partition is not better than random. Maximizing the modularity for a number of clusters c larger than two is equivalent to the maximum cut problem, which is NP-hard [60]. However, Van Mieghem *et al.* [63] find that it is possible to derive an upper bound on the modularity measure m for any graph given the true clustering:

$$m \leq 1 - \frac{1}{c} - \frac{L_{\text{inter}}}{L}, \tag{2.2}$$

with c the true number of clusters in the graph and L_{inter} the total number of inter-cluster links in the true clustering of the network. Additionally, the $N \times N$ modularity matrix **M**

of a network is defined with elements $m_{ij} = a_{ij} - \frac{k_i k_j}{2L}$. If we define the $N \times c$ community matrix $\mathbf{S}$ with elements s_{ig} to be equal to 1 if node i is in cluster g and 0 otherwise, we can express the modularity m [4] in terms of these matrices:

$$m = \frac{1}{2L} \sum_{g=1}^{c} \sum_{i=1}^{N} \sum_{j=1}^{N} s_{ig} m_{ij} s_{jg} = \frac{\text{trace}\left(\mathbf{S}^T \mathbf{M} \mathbf{S}\right)}{2L}. \tag{2.3}$$

Several heuristic algorithms that approximately maximize modularity have been proposed. We consider the popular Louvain method [10] and Newman's iterative bisection algorithm [61], because they have been shown to achieve high modularity for several real-world networks.

2.2.2. SPECTRAL CLUSTERING

In spectral clustering, nodes are assigned to clusters based on the values of their corresponding components in one or more of the eigenvectors of a matrix representation of the network. The Laplacian is the most popular matrix representation for spectral clustering [6]. Fiedler showed that the eigenvector corresponding to the second smallest eigenvalue μ_{N-1} of the Laplacian $\mathbf{Q}$ can be used to obtain a bipartition of a graph into two equivalent parts [4]. Here, *equivalent* means equivalent with respect to the second smallest eigenvalue's eigenvector, the Fiedler eigenvector. If a disconnected network consists of c connected components, then the Laplacian $\mathbf{Q}$ will have c eigenvalues that are equal to zero. The eigenvectors corresponding to these eigenvalues are the trivial all-ones eigenvectors of the connected components, with entry 1 for nodes that are part of the corresponding component and entry 0 for nodes that are not part of the component. The c eigenvectors map the nodes of one connected component onto a single point on one of the axes in a c-dimensional space. The c eigenvectors can then be used to detect the component membership of the nodes.

The idea of spectral clustering with the Laplacian $\mathbf{Q}$ is that if the graph consists of c weakly linked subgraphs (*e.g.*, a network with community structure), the smallest $c-1$ non-zero eigenvalues will still be relatively close to zero. An eigenvector belonging to one of the c smallest eigenvalues no longer maps the nodes of one subgraph onto a single point on one of the axes, but rather to a small cloud of points that are still relatively close to each other [6]. Any clustering algorithm that can identify clusters shaped as clouds of points in metric space given the true number of clusters c, such as the *k-means* algorithm, can detect the original cluster memberships in the network. To what extent the clusters are considered weakly-linked and to what extent spectral clustering works, we will discuss in the context of detecting the number of clusters with the maximum eigengap property in Section 2.2.3. For matrix representations other than the Laplacian $\mathbf{Q}$, spectral clustering also works [6].

2.2.3. MAXIMUM EIGENGAP

Consider the $N \times N$ adjacency matrix $\mathbf{A}$ of a graph G with c equally-sized clusters of $N_g = N/c$ nodes. The nodes can always be rearranged such that the nodes are ordered according to the cluster memberships. The $N_g \times N_g$ adjacency matrices $\mathbf{A}_g$ of the cluster subgraphs are then located on the diagonal of the adjacency matrix $\mathbf{A}$ and contain only

intra-cluster links:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{B}_{12} & \dots & \mathbf{B}_{1c} \\ \mathbf{B}_{12}^T & \mathbf{A}_2 & \dots & \mathbf{B}_{2c} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{B}_{1c}^T & \mathbf{B}_{2c}^T & \dots & \mathbf{A}_c \end{bmatrix}. \tag{2.4}$$

We first consider the case where the c clusters are disconnected and the adjacency matrix $\mathbf{A}$ in (2.4) is block diagonal with all blocks $\mathbf{B}_{ij} = \mathbf{O}$. The eigenvalues of the matrix $\mathbf{A}$ are the union of the sets of eigenvalues of the blocks $\mathbf{A}_g$ and each block $\mathbf{A}_g$ has N_g eigenvalues $\lambda_1(\mathbf{A}_g) \geq \lambda_2(\mathbf{A}_g) \geq \cdots \geq \lambda_{N_g}(\mathbf{A}_g)$. Let the degree distribution of the subgraph g have expectation $E[D_g]$ and variance $\mathrm{Var}[D_g]$. The Perron-Frobenius eigenvalue $\lambda_1(\mathbf{A}_g)$ of each subgraph g must satisfy:

$$\lambda_1(\mathbf{A}_g) \geq E[D_g] \sqrt{1 + \frac{\mathrm{Var}[D_g]}{(E[D_g])^2}}, \tag{2.5}$$

as derived in [4]. If the subgraph g has no community structure itself, then the largest eigenvalue $\lambda_1(\mathbf{A}_g)$ is substantially larger than the other $N_g - 1$ eigenvalues [64, 65]. An exception to this statement would be when the expected degree $E[D_g]$ of nodes in the subgraph g is relatively low. However, in general we have $\lambda_1(\mathbf{A}_g) \gg \lambda_2(\mathbf{A}_g)$ for subgraphs without a community structure [65]. If indeed $\lambda_1(\mathbf{A}_g) \gg \lambda_2(\mathbf{A}_g)$ and the degree distributions of the subgraphs g are similar, then the c largest eigenvalues of $\mathbf{A}$ are substantially larger than the other $N - c$ eigenvalues and we observe a large gap. Finding the largest gap among the sorted eigenvalues of $\mathbf{A}$, the maximum eigengap identifies the number of clusters c in this case.

Adding the inter-community links to the graph G can be considered as a perturbation of the adjacency matrix $\mathbf{A}$, such that $\mathbf{A}(\zeta) = \mathbf{A} + \zeta\mathbf{B}$ is the adjacency matrix with inter-community links for a value ζ that is small enough [65, 66]. The symmetric matrix $\zeta\mathbf{B}$ contains only the inter-community links in the blocks $\mathbf{B}_{ij}$ from (2.4) and has zeros $\mathbf{O}$ on the diagonal. When the perturbed eigenvalues $\lambda(\zeta)$ of $\mathbf{A}(\zeta)$ are close to the eigenvalues λ of $\mathbf{A}$, the maximum eigengap is still a good estimator of the number of clusters c. Similarly, when the perturbed eigenvectors $\mathbf{x}(\zeta)$ are close to the eigenvectors $\mathbf{x}$ of $\mathbf{A}$, spectral clustering with these eigenvectors still works. If ζ is sufficiently small, then the i-th perturbed eigenvalue $\lambda_i(\zeta)$ can be approximated by the first-order approximation of $\lambda_i(\zeta)$,

$$\lambda_i(\zeta) \approx \lambda_i + \zeta \mathbf{x}_i^T \mathbf{B} \mathbf{x}_i, \tag{2.6}$$

and also for the corresponding eigenvector $\mathbf{x}_i(\zeta)$,

$$\mathbf{x}_i(\zeta) \approx \mathbf{x}_i + \zeta \sum_{j \neq i}^{N} \frac{\mathbf{x}_j^T \mathbf{B} \mathbf{x}_i}{\lambda_i - \lambda_j} \mathbf{x}_j, \tag{2.7}$$

as derived in [67]. In the clustering problem, we assume $\lambda_i \gg \lambda_j$ for the c largest eigenvalues $\lambda_1(\mathbf{A}_g)$ of $\mathbf{A}$. Wu *et al.* [66] therefore propose to write the approximations for the eigenvectors $\mathbf{x}_i(\zeta)$ of the c largest eigenvalues as

$$\mathbf{x}_i(\zeta) \approx \mathbf{x}_i + \zeta \sum_{j \neq i}^{c} \frac{\mathbf{x}_j^T \mathbf{B} \mathbf{x}_i}{\lambda_i - \lambda_j} \mathbf{x}_j + \frac{\zeta}{\lambda_i} \mathbf{B} \mathbf{x}_i, \tag{2.8}$$

where the last term of the approximation error is smaller for large λ_i. Denote by $\mathbf{X}_{\setminus\{i\}}$ the $N \times (N-1)$ matrix of all eigenvectors $\mathbf{x}_j$ except $\mathbf{x}_i$. The approximations are close to the actual $\lambda_i(\zeta)$ and $\mathbf{x}_i(\zeta)$ when

$$|\lambda_i - \lambda_{i+1}| - |\zeta| \left\| \mathbf{x}_i^T \mathbf{B} \mathbf{x}_i \right\|_2 - |\zeta| \left\| \mathbf{X}_{\{i\}}^T \mathbf{B} \mathbf{X}_{\setminus\{i\}} \right\|_2 > 0, \tag{2.9}$$

and

$$|\zeta| \left\| \mathbf{X}_{\{i\}}^T \mathbf{B} \mathbf{x}_i \right\|_2 < \frac{1}{2} \left(|\lambda_i - \lambda_{i+1}| - |\zeta| \left\| \mathbf{x}_i^T \mathbf{B} \mathbf{x}_i \right\|_2 - |\zeta| \left\| \mathbf{X}_{\{i\}}^T \mathbf{B} \mathbf{X}_{\setminus\{i\}} \right\|_2 \right), \tag{2.10}$$

as described in [66, 68].

2.2.4. STOCHASTIC BLOCK MODEL

The stochastic block model (SBM) is a generative random graph model in which the link probabilities depend on cluster memberships. A realization of the model therefore results in a graph with a community structure. Clustering algorithms are often evaluated on SBM graphs. The hardest case for cluster detection is the symmetric SBM (SSBM), where there are only two different link probabilities: p_{in} for two nodes that belong to the same cluster, p_{out} for two nodes that belong to different clusters. When the intra-cluster link probability is larger than the inter-cluster probability, $p_{\text{in}} > p_{\text{out}}$, we obtain graphs that look similar to the graph in Figure 2.1. In SSBM graphs with clusters of equal size, the clusters are not detectable based on the node degrees alone, because the expected degree is the same for each node in the graph, irrespective of its cluster membership.

We consider sparse, assortative SSBMs with c clusters of equal size. The SSBM is said to be sparse and assortative if $p_{\text{in}} = b_{\text{in}}/N$ and $p_{\text{out}} = b_{\text{out}}/N$ for two constants $b_{\text{in}} > b_{\text{out}} > 0$ that do not depend on the number of nodes N. In case of equally-sized clusters, the expected average degree $E[D]$ is a weighted average of the constants:

$$E[D] = \frac{b_{\text{in}} + (c-1) b_{\text{out}}}{c}. \tag{2.11}$$

Decelle *et al.* [69] show that there is a regime where no algorithm can detect the clusters because the block structure is not apparent enough in the thermodynamic limit $N \to \infty$. When the difference $|b_{\text{in}} - b_{\text{out}}|$ is larger than the detectability threshold,

$$|b_{\text{in}} - b_{\text{out}}| > c\sqrt{E[D]}, \tag{2.12}$$

it is theoretically possible to detect the clusters by some algorithm. However, for most algorithms the difference between b_{in} and b_{out} must be much larger than the detectability threshold $c\sqrt{E[D]}$ to detect the clusters in the sparse case [62]. Decelle *et al.* [69] also show that the detectability limit marks a phase transition from the undetectable state to the theoretically detectable state.

2.3. RELATED WORK

There is not yet a single, generally accepted method that determines the number of clusters in any given network [6, 7, 13]. In the absence of a golden standard, there have been a few works that—at least partly—address the problem of detecting the number of clusters. In many of the papers, the clustering methods are evaluated on real-world networks

by comparing the results with the presumed underlying clustering or with a benchmark random graph model.

Shen and Cheng [11] discuss the detection of the number of clusters for spectral clustering using the maximum eigengap property based on several different matrix representations of the network and, additionally, based on the covariance matrix of the node degrees. They evaluate the estimators on a random graph model proposed by Lancichinetti *et al.* [20] with clusters of different sizes and heterogeneous node degrees. Shen and Cheng [11] conclude that in graphs with heterogeneous cluster sizes and heterogeneous node distributions, the maximum eigengaps of the normalized Laplacian and the covariance matrix provide the best estimates of the number of clusters because the two matrices both correct for heterogeneous node degrees. At the time of their analysis [11], the non-backtracking matrix had not been proposed for spectral clustering yet. Shea and Macker [70] try and formalize the selection of the number of clusters based on the eigengap property of the normalized Laplacian by combining the eigengap property with statistics of random cuts of the graph.

Another approach to the detection of the number of clusters consists of fitting a statistical parametric model to the graph and including the number of clusters as one of the parameters to be estimated. The stochastic block model is often explicitly assumed as the true underlying model of the graph and then the most likely number of clusters is estimated using maximum likelihood [13–15]. Given that the true clustering is unobserved, multiple values for the number of clusters are evaluated based on some criterion, before the estimated number of clusters is found. Alternatively, the problem of finding the number of clusters can also be cast in a Bayesian framework [71, 72], in which one formulates a prior distribution on both the number of clusters and the cluster memberships and then finds the *a posteriori* most likely number of clusters after learning from some network data. The disadvantage of Bayesian approaches is the computational burden of the required data sampling, which Decelle *et al.* [69] overcome by proposing a belief propagation algorithm that runs in polynomial time. They argue that their method is asymptotically exact for the stochastic block model, also in the sparse case.

Krzakala *et al.* [62] introduced the non-backtracking matrix of Hashimoto [73] in a spectral clustering algorithm for networks. They conjectured that a spectral clustering algorithm based on the non-backtracking matrix can detect clusters all the way down to the detectability limit of the stochastic block model in the sparse case. Krzakala *et al.* [62] remark that the number of eigenvalues located outside a circle around the origin in the complex plane seems to be a good estimator of the number of clusters in the graph. Given that the bulk of the eigenvalues of the non-backtracking matrix are contained within a circle around the origin of the complex plane, the number of eigenvalues outside that circle yields a clear and unambiguous decision rule for the choice of the number of clusters.

2.4. DETECTING THE NUMBER OF CLUSTERS

We compare 4 different methods to detect the number of clusters in a network. We compare the estimate obtained by counting the eigenvalues outside the circle in the spectrum of the non-backtracking matrix $\mathbf{H}$ with 3 different methods based on the concept of modularity.

2.4.1. DETECTION METHODS

SPECTRUM OF THE NON-BACKTRACKING MATRIX **H**

The non-backtracking matrix **H** is based on the idea of a non-backtracking walk: a walk that does not turn around and goes back to its starting point immediately after the first step. Two directed links are backtracking if they both connect the same pair of nodes but in opposite directions. Two directed links are non-backtracking if one follows after the other and does not loop back to the starting node of the first link. For example, the two links $e_e = i \to j$ and $e_l = i \leftarrow j$ are backtracking, while the two links $e_e = i \to j$ and $e_l = j \to k$ are non-backtracking for $k \neq i$. In a non-backtracking walk, no two consecutive links are backtracking. For illustrating non-backtracking walks, consider the example of the directed graph in Figure 2.2.

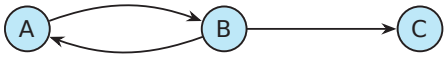


Figure 2.2: A directed graph with 3 nodes and 3 lexicographically ordered directed links $e_1 = A \to B$, $e_2 = A \leftarrow B$ and $e_3 = B \to C$

A walk over the links $e_1 = A \to B$ and $e_2 = A \leftarrow B$ immediately returns to its starting node and the walk is therefore backtracking. A walk over the links $e_1 = A \to B$ and $e_3 = B \to C$ does not immediately return to its beginning node and is therefore non-backtracking. The latter is a non-backtracking walk of length 2: a non-backtracking walk consisting of a succession of 2 links. For undirected graphs, each undirected link $\{i, j\}$ is considered twice, once for the direction $i \to j$ and once for the direction $j \to i$, resulting in $2L$ bi-directional links, illustrated by Figure 2.3.

Figure 2.3: Representing an undirected link in graph G by two bi-directional links.

The non-backtracking matrix **H** is then a $2L \times 2L$ matrix with elements

$$h_{el} = \begin{cases} 1 & \text{if } e_e = i \to j,\ e_l = j \to k \text{ and } i \neq k \\ 0 & \text{otherwise} \end{cases}, \tag{2.13}$$

where $e, l \in \{1, \ldots, 2L\}$ the set of all bi-directional links and $i, j, k \in \{1, \ldots, N\}$. Two consecutive links $e_e = i \to j$ and $e_l = j \to k$ are non-backtracking if $i \neq k$. If $i = k$, then the link e_e connects the same two nodes as the link e_l but in opposite direction and the link pair (e_e, e_l) is backtracking. A non-backtracking walk is a walk in which every pair of consecutive links is non-backtracking. Similar to the adjacency matrix **A**, raising the non-backtracking matrix **H** to the k-th power counts the number of non-backtracking walks of length $k+1$ on the graph. We describe the non-backtracking matrix **H** in more detail in Appendix A.1.

Krzakala *et al.* [62] conjectured that the number of real eigenvalues of the non-backtracking matrix **H** that are separated from the bulk of the eigenvalues indicates the number of clusters in the network. The bulk of the eigenvalues are located in a circle around

the origin of the complex plane with radius the square root of the largest eigenvalue, therefore the number of eigenvalues located outside of the circle indicates the number of clusters. Figure 2.4 shows the eigenvalues of the matrix $\mathbf{H}$ in the complex plane for a graph on $N = 1000$ nodes generated by an SSBM with 3 clusters. The largest eigenvalue is approximately equal to the expected degree of 7, and there are exactly 3 real eigenvalues located outside the circle. The spectrum of the matrix $\mathbf{H}$ indeed shows that the number of clusters $c = 3$, which is in agreement with the conjecture.

Let $|\lambda_1(\mathbf{H})| \geq \cdots \geq |\lambda_{2L}(\mathbf{H})|$ denote the eigenvalues of the non-backtracking matrix $\mathbf{H}$, sorted in descending order according to the modulus. The eigenvalues $\lambda_l(\mathbf{H}) \in \mathbb{C}$ are the solutions to the characteristic polynomial (A.1) in Appendix A.1.2. The largest eigenvalue $\lambda_1(\mathbf{H})$ is a real, non-negative number [4]. The estimate $\hat{c}$ of the number of clusters is the number of real eigenvalues $\lambda_l(\mathbf{H})$ larger than the square root $\sqrt{\lambda_1(\mathbf{H})}$ of the largest eigenvalue including the largest eigenvalue $\lambda_1(\mathbf{H})$ itself:

$$\hat{c} = \sum_{l=1}^{2L} 1_{\left\{\mathrm{Re}(\lambda_l(\mathbf{H})) > \sqrt{\lambda_1(\mathbf{H})} \wedge \mathrm{Im}(\lambda_l(\mathbf{H})) = 0\right\}}. \tag{2.14}$$

For finding the estimate $\hat{c}$ of the number of clusters, not all $2L$ eigenvalues have to be computed. One first has to find the largest root $\lambda_1(\mathbf{H})$ of the characteristic polynomial and compute its square root $\sqrt{\lambda_1(\mathbf{H})}$. Next, find all real roots larger than $\sqrt{\lambda_1(\mathbf{H})}$ with a numerical procedure. The number of such roots is the estimate $\hat{c}$. Moreover, it is even more efficient to obtain the estimate $\hat{c}$ by applying the above iterative procedure to the $2N \times 2N$ block matrix $\mathbf{H}^*$ specified in (A.2), Appendix A.1.2. The size of the square non-symmetric matrix $\mathbf{H}^*$ is of order $O\left(N^2\right)$. We discuss the computational complexity of the non-backtracking method in Appendix A.1.4.

LOUVAIN METHOD

The Louvain method is a popular heuristic method aimed at maximizing the modularity of a graph [10]. The method starts out with N clusters, one for every node. The clusters are then iteratively merged in a two-stage procedure such that the modularity m increases in every iteration, until the modularity cannot be improved upon anymore. However, there is no guarantee that the final result is a global modularity maximum. An estimate of the number of clusters is obtained from the clustering results.

One iteration in the Louvain method consists of two stages. In stage one, each node i is considered sequentially and possibly multiple times. Blondel *et al.* [10] compute the resulting gain in modularity from moving node i to the cluster g of some neighboring node j as:

$$\Delta m = \left[\frac{\Sigma_{\text{in}} + 2\Sigma_{l:s_{lg}=1} w_{il}}{2L} - \left(\frac{\Sigma_{\text{tot}} + k_i}{2L}\right)^2\right] - \left[\frac{\Sigma_{\text{in}}}{2L} - \left(\frac{\Sigma_{\text{tot}}}{2L}\right)^2 - \left(\frac{k_i}{2L}\right)^2\right], \tag{2.15}$$

where $\mathbf{W}$ is the weighted adjacency matrix of the graph with as elements the weights w_{ij}, while $2L$ is defined as $2L = \sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}$ is the sum of the weights of all links in the graph, $k_i = \sum_{l=1}^{N} w_{il}$ is the sum of the weights of the links incident to node i, Σ_{in} is the sum of the weights of the intra-cluster links in cluster g, Σ_{tot} is the sum of the weights of all links incident to one of the nodes in cluster g. Node i is moved to cluster g of

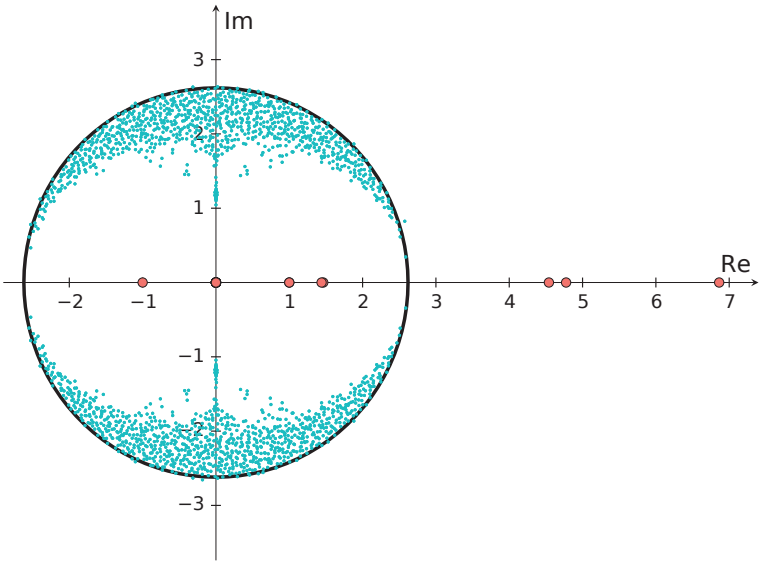


Figure 2.4: The spectrum of the non-backtracking matrix **H** in the complex plane for a graph with $N = 1000$ nodes generated by an SSBM with $c = 3$ clusters of approximately equal size. The real eigenvalues are indicated by red dots and the eigenvalues with a non-zero imaginary part are indicated by blue dots. The circle containing the bulk of the eigenvalues is centered around the origin and has a radius equal to the square root of the largest real eigenvalue. There are 3 real eigenvalues located outside the circle.

the neighboring node j for which the modularity gain is most positive. If the resulting gain in modularity is not larger than some small threshold or even negative for all neighboring nodes j, then the node i remains in its original cluster. In the first iteration, the weighted adjacency matrix **W** is simply the unweighted adjacency matrix **A**, and from the second iteration onward, we have the weighted adjacency matrix **W** constructed in the second stage of the previous iteration. The cluster re-assignment procedure is repeated until there is no node anymore for which there is a positive gain in the modularity m achievable.

In the second stage of an iteration, a new weighted graph is constructed in which each node g represents a cluster g resulting from the first stage. In the new graph, the weight on the link from node g to node h is the sum of the weights of all inter-cluster links between cluster g and cluster h in the graph of stage 1. The intra-cluster links in the graph of stage 1 lead to self-loops in the new graph, such that the new graph has the same modularity m as the graph in stage 1. The newly constructed graph in stage 2 is the input for stage 1 of the next iteration. The described iterations are repeated until there is no positive gain in the modularity m achievable anymore.

NEWMAN'S ITERATED BISECTION

To maximize the modularity m, Newman [61] proposes to make recursive splits according to the leading eigenvector of the modularity matrix **M**, which is inspired by the approach of Fiedler [4]. In the case of $c = 2$ clusters, clustering is equivalent to choosing a vector **y** with elements $+1$ and -1 that indicate the cluster membership. The vector **y**

can be written as a linear combination of the orthogonal eigenvectors $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_N$ of the modularity matrix $\mathbf{M}$, $\mathbf{y} = \sum_{j=1}^{N} \beta_j \mathbf{w}_j$, with coefficients $\beta_j = \mathbf{y}^T \mathbf{w}_j$. Invoking the orthogonality of the eigenvectors, the modularity m is written as:

$$m = \frac{1}{4L} \mathbf{y}^T \mathbf{M} \mathbf{y} = \frac{1}{4L} \sum_{j=1}^{N} \beta_j^2 \lambda_j(\mathbf{M}). \tag{2.16}$$

Maximizing the modularity m is equivalent to choosing the vector $\mathbf{y}$ with cluster memberships proportional to the eigenvectors corresponding to a few of the largest eigenvalues. Newman [61] proposes to perform the bisection by maximizing the term for the most positive eigenvalue: $\beta_1 = \mathbf{y}^T \mathbf{w}_1$. Since the elements y_j of $\mathbf{y}$ only take two possible values, the coefficient β_1 is maximized for $y_j = -1$ if $(\mathbf{w}_1)_j < 0$ and $y_j = +1$ if $(\mathbf{w}_1)_j \geq 0$.

For the case of $c > 2$ clusters, the network is first split in two and then the procedure is repeated on each of the resulting subgraphs separately. However, simply applying the same procedure on the block components of the modularity matrix $\mathbf{M}$ corresponding to the subgraphs is not correct, because the block components do not contain inter-cluster links and the modularity m would change if the inter-cluster links are disregarded. Instead, Newman [61] proposes to write the modularity matrix $\mathbf{M}_g$ of a cluster g as:

$$\mathbf{M}_g = m_{ij} - \left(\sum_{k \in g} m_{ik} \right) \delta_{ij}, \tag{2.17}$$

with Kronecker delta δ_{ij} equal to 1 if $i = j$, and 0 otherwise. In the iterated bisection algorithm of Newman [61], the subgraphs are iteratively split in two until further improvements in the modularity are no longer possible, indicated by an absence of positive eigenvalues of the modularity matrix $\mathbf{M}_g$ of the subgraph. The stopping criterion is evaluated for each subgraph separately, therefore the resulting number of clusters can be any number greater than or equal to 1 and is not necessarily a multiple of 2. The algorithm provides an estimate of the number of clusters and the corresponding clustering results too.

EIGENGAP MODULARITY MATRIX $\mathbf{M}$

We estimate the number of clusters c by the maximum eigengap of the modularity matrix $\mathbf{M}$ as described for the adjacency matrix $\mathbf{A}$ in Section 2.2.3. Let the eigenvalues of the modularity matrix $\mathbf{M}$ be sorted in descending order: $\lambda_1(\mathbf{M}) \geq \lambda_2(\mathbf{M}) \geq \cdots \geq \lambda_N(\mathbf{M})$. The eigenvalues of the modularity matrix $\mathbf{M}$ and the adjacency matrix $\mathbf{A}$ are interlaced:

$$\lambda_1(\mathbf{A}) \geq \lambda_1(\mathbf{M}) \geq \lambda_2(\mathbf{A}) \geq \lambda_2(\mathbf{M}) \geq \cdots \geq \lambda_N(\mathbf{A}) \geq \lambda_N(\mathbf{M}), \tag{2.18}$$

as described in [4]. The maximum eigengap property then maximizes the difference $\lambda_{i-1}(\mathbf{M}) - \lambda_i(\mathbf{M})$ in the sequence of N eigenvalues as

$$\hat{c} = \arg\max_i \left(\lambda_{i-1}(\mathbf{M}) - \lambda_i(\mathbf{M}) \right), \qquad i = 2, \ldots, N, \tag{2.19}$$

with $\hat{c}$ the resulting estimate of the number of clusters c.

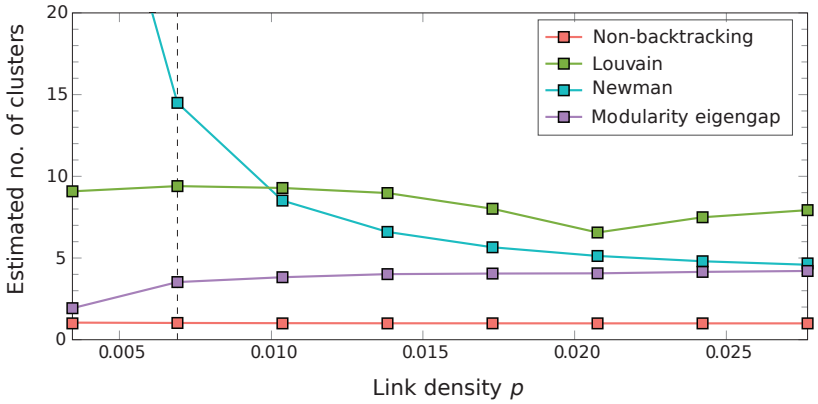


Figure 2.5: The estimated number of clusters of the largest connected component for ER graphs $G_p(N)$ with $N = 1000$ nodes and different link densities p. Each point in the plots represents an average over 10^4 realizations. The vertical dashed line indicates the connectivity threshold p_c for an ER graph with $N = 1000$ nodes.

2.5. RESULTS

2.5.1. ERDŐS-RÉNYI GRAPHS

We evaluate the cluster detection methods on Erdős-Rényi (ER) random graphs with different link densities in Figure 2.5. An ER random graph features no clustering structure in expectation, but in individual realizations some clustering structure might be present due to randomness. The non-backtracking method finds on average one single cluster in the ER graph for each of the link density values. The Louvain method and the modularity eigengap method find a higher, constant number of clusters on average, while Newman's iterated bisection method finds a lower number of clusters as the link density increases. The non-backtracking method detects 1 cluster in absence of a clustering structure, which is a convenient property for an estimator of the number of clusters.

2.5.2. RING OF CLIQUES

We evaluate the cluster detection methods on a ring of c cliques, similar to the evaluation method of Fortunato and Barthélemy [74]. Each clique is the complete graph K_{N_g} with N_g nodes, and we connect neighboring cliques by a single link. This setup is a seemingly easy clustering problem, but Fortunato and Barthélemy [74] illustrate the resolution limit of modularity optimization methods by showing that for a ring of $c > \sqrt{L}$ cliques, the modularity m is maximized for $c/2$ clusters that consist of 2 cliques each. We choose a set-up where each cluster has relatively many nodes and links, therefore we have the opposite: $c < \sqrt{L}$. The complete graph K_{N_g} has the largest possible spectral gap between the two largest eigenvalues.

Figure 2.6 shows the estimated number of clusters for our experiment. Intuitively, a cluster detection method is expected to find as many clusters as there are cliques. The non-backtracking method estimates the correct number of clusters (cliques) in almost every instance, while Newman's iterated bisection and the modularity eigengap method

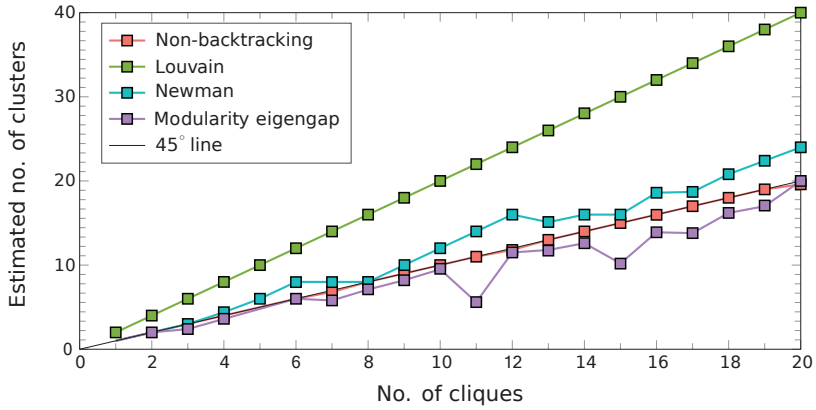


Figure 2.6: The estimated number of clusters for a ring of c cliques, where each clique is the complete graph K_{100} with $N_g = 100$ nodes. The entire graph has $N = N_g \cdot c = 100c$ nodes and two neighboring cliques in the ring are connected by a single link. The 45° line maps the true number of clusters (cliques) to the estimated number of clusters on a 1:1 scale, the estimates of an ideal estimator would be positioned close to the 45° line. For each point in the graph, we randomize the order of the nodes 20 times and average over these 20 estimations.

are close on average. Surprisingly, the Louvain method finds exactly twice the number of cliques in every instance. Inspecting the Louvain clustering results reveals that the method subdivides each clique into two smaller clusters. The clustering results of the Louvain method are therefore almost equivalent to the partition where each clique is a cluster, but the detected number of clusters is not the intuitively expected number.

2.5.3. THE NUMBER OF CLUSTERS IN SSBMS

CLUSTERING IN THE EIGENVECTORS OF THE MODULARITY MATRIX $\mathbf{M}$

First, we inspect if and how the clustering pattern appears in the eigenvector components of the modularity matrix for a network generated by an SSBM with $N = 1000$ nodes and average degree $E[D] = 7$. Figure 2.7 shows the components of the first eigenvector $\mathbf{w}_1$ versus the components of the second eigenvector $\mathbf{w}_2$ for two graphs generated by SSBMs with $c = 3$ clusters. For the network in Figure 2.7(a), the difference $b_{\text{in}} - b_{\text{out}} = 19$ is well above the detectability threshold in (2.12) of 7.94. The objects corresponding to the nodes of a single cluster are clearly cluttered and they are well separated from the objects corresponding to the nodes of other clusters. For the graph in Figure 2.7(a), a spectral clustering algorithm based on the first two eigenvectors would successfully detect the cluster memberships of the majority of the nodes. For the graph in Figure 2.7(b) the difference $b_{\text{in}} - b_{\text{out}} = 11$ is closer to the detectability threshold and the objects corresponding to the nodes of the different clusters now show significant overlap. Even though a spectral clustering algorithm would now classify fewer nodes correctly, there are still three (overlapping) clouds of points visible. The detection of the number of clusters through the maximum eigengap is potentially still possible.

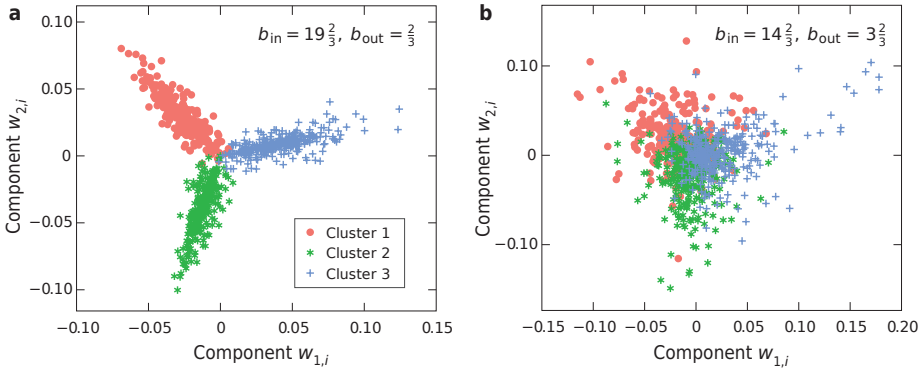


Figure 2.7: Scatter plots of the components of the second eigenvector $\mathbf{w}_2$ versus the first eigenvector $\mathbf{w}_1$ of the modularity matrix for two different graphs, both with $N = 1000$ nodes and generated by SSBMs with $c = 3$ clusters with (a) $b_{in} = 19\frac{2}{3}$, $b_{out} = \frac{2}{3}$ and (b) $b_{in} = 14\frac{1}{3}$, $b_{out} = 3\frac{1}{3}$. The parameters b_{in} and b_{out} are chosen such that the expected average degree $E[D] = 7$ in both networks and that they only differ through the difference $b_{in} - b_{out}$, for which the theoretical detectability threshold in (2.12) is 7.94. The colors and shapes of the objects indicate the true cluster memberships.

CLUSTERS OF EQUAL SIZE

Figure 2.8 shows the estimated number of clusters for networks generated by SSBMs with 2, 3, 4, 10, and 20 equally-sized clusters. In each panel, from left to right the number of intra-cluster links increases with respect to the number of inter-cluster links while keeping the average degree constant. The contrast of the clusters becomes stronger as the difference $b_{in} - b_{out}$ increases and the detection of clusters becomes easier. On the left side of the detectability threshold, all methods appear to have their own default guess for the number of clusters. The non-backtracking method and the modularity eigengap find the correct number of clusters already slightly above the detectability threshold for graphs with a lower number of clusters in Figures 2.8(a-c). Newman's iterated bisection method finds the correct number of clusters, but much higher above the detectability threshold than the first two methods. The Louvain method does not find the correct number of clusters. However, inspection of the actual clustering results from the Louvain method reveals that the true clusters are found, but they are subdivided into two or more clusters, similar to the ring of cliques in Figure 2.6. For the SSBMs with a higher number of clusters c in Figures 2.8(d-e), the results are similar to the results for the low number of clusters. The difference is that Newman's iterated bisection does not find the correct number of clusters and the modularity eigengap needs the difference $b_{in} - b_{out}$ to be larger than the non-backtracking method before it detects the correct number of clusters c.

Figure 2.9 shows the simulated densities of the estimators for the case in Figure 2.8(c) where the number of clusters $c = 4$. The three different values $b_{in} - b_{out} = \{6, 14, 23\}$ correspond to the left, center, and right of the interval in Figure 2.8(c), respectively. The estimates of the non-backtracking method are narrowly distributed around the correct number of clusters for the two cases with the highest contrast of the clusters. Although the modularity eigengap method on average finds a number close to the true number

of clusters for the case $b_{in} - b_{out} = 14$, most estimates deviate significantly from the true number and it seems a coincidence that the average estimated number is close to the true number of clusters. Combining the three cases in Figure 2.9, the distribution of the non-backtracking method is overall the most accurate.

IMBALANCED CLUSTERS

Figure 2.10 shows the estimated number of clusters for a similar experiment as in Figure 2.8(c), but with 4 clusters of heterogeneous size. The size of one cluster, cluster 1, is set to deviate from the sizes of the other clusters to assess the impact of imbalancedness. For example, in Figure 2.10(d) the first cluster contains $N_1 = 400$ nodes and the other three clusters contain $N_2 = N_3 = N_4 = 200$ nodes each. In Figure 2.8(c), cluster 1 and the other three clusters have 250 nodes each. How far the size N_1 deviates from the value 250 indicates the degree of imbalancedness. Already when the clusters are slightly imbalanced, Figures 2.10(b-c), the modularity eigengap fails to detect the right number of clusters c. The graphs in Figure 2.10(a) are recognized as graphs with 3 clusters by the detection methods and the imbalancedness appears to be too strong. In Figure 2.10(e), the modularity eigengap method seemingly detects the number of clusters perfectly. However, when comparing Figure 2.10(e) with the other figures, it seems more likely that the imbalancedness is too strong and that none of the methods can detect the clustering structure. For the graphs generated by the SSBMs of Figures 2.10(a,e) one could also argue that the graphs actually have 3 clusters and 1 cluster respectively. A hard definition of the concept of a cluster would be required to make stronger statements.

Table 2.1: Detected number of clusters and numbers of nodes and links for 5 real-world networks.

Network	N	L	***Non-backtracking***	***Louvain***	***Newman***	***Modularity eigengap***
Zachary karate	34	78	2	6	4	1
Political books	105	441	3	4	4	2
Facebook	347	2,519	8	26	18	2
NetSci	1,589	2,742	23	401	300	2
ArXiv Hep-Ph	9,877	25,988	83	446	68	2

2.5.4. EVALUATION ON REAL-WORLD NETWORKS

We evaluate the 4 methods for detecting the number of clusters c on several real-world networks commonly used in community detection. Table 2.1 shows the detected number of clusters for Zachary's karate club network [75], a network of political books sold by Amazon[1], a social circle network from Facebook [76], a co-authorship network for publications in Network Science [77] and a collaboration network of the Arxiv High Energy Physics Theory category [78]. The detected number of clusters c differs significantly across the four different methods. The definition of a cluster indeed appears to depend on the method. It is difficult to make statements about the validity of the methods based

[1] Unpublished, obtained from http://www-personal.umich.edu/~mejn/netdata/.

on these results since the ground truth is not known. The results of the modularity eigengap are not useful, since the detected number of clusters does not seem to depend much on the network structure. Potentially the modularity eigengap method fails because the clusters in real networks are often not of equal sizes since we know from the simulation experiment in Section 2.5.3 that the modularity eigengap does not work well for clusters of different sizes. The non-backtracking method finds a lower number of clusters c than the Louvain method and Newman's iterated bisection for most networks.

2.6. CONCLUSION

This work considers the detection of the number of clusters in a graph. Many clustering methods require the number of clusters as an input and their results depend on the chosen number of clusters. Partly also due to the lack of a clear definition of the concept of a cluster, the precise number of clusters in a graph is debatable.

We have compared the estimates based on the non-backtracking matrix with several estimators based on the concept of modularity. We find that the number of eigenvalues of the non-backtracking matrix located outside a circle in the complex plane is an excellent estimator of the number of clusters in sparse graphs where the clusters are not distinguishable based on differences in the node degrees alone. For graphs without a clustering structure, the non-backtracking method detects one single cluster, which is a convenient property for an estimator of the number of clusters. We also find that the detection based on the maximum eigengap of the modularity matrix performs similarly to the non-backtracking method for equally-sized clusters, but the performance of the modularity eigengap method breaks down already when the clusters are slightly imbalanced. The estimates of the non-backtracking method are narrowly distributed around the true number of clusters for the benchmark graphs. In conclusion, the method based on the eigenvalues of the non-backtracking matrix indeed yields an unambiguous decision rule for the choice of the number of clusters.

The non-backtracking matrix is, however, an asymmmetric matrix and many of its eigenvalues are complex, making the non-backtracking method spectrally and conceptually difficult. The computational complexity can be reduced by calculating the spectrum of $\mathbf{H}^*$ in (A.2) from a quadratic eigenvalue equation based on the adjacency matrix $\mathbf{A}$. Moreover, for obtaining the estimate of the number of clusters, only the few largest, real eigenvalues have to be computed, reducing the computational complexity further. The computation time of only a few large eigenvalues of the non-backtracking matrix scales approximately linearly with the number of nodes N in the network.

The problem of detecting the number of clusters in a graph remains a difficult problem because of the lack of a hard definition of the concept of a cluster. However, we find that when loosely defining a cluster as a group of densely connected nodes, the number of eigenvalues of the non-backtracking located outside a circle in the complex plane is a good estimator of the number of clusters.

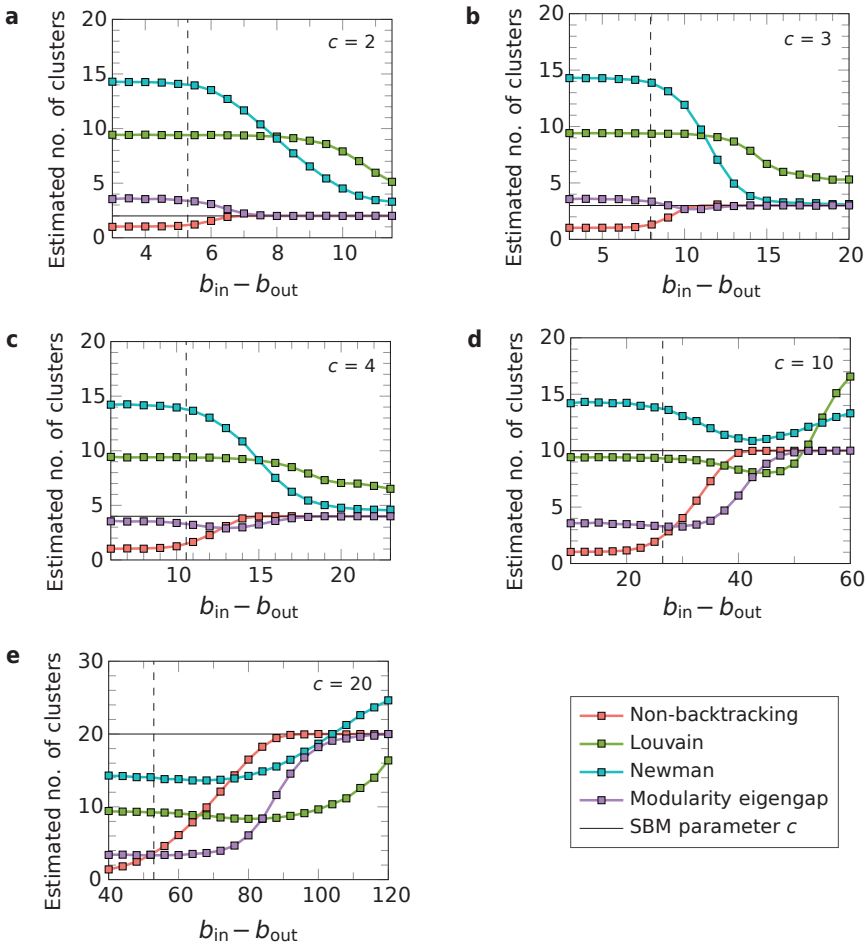


Figure 2.8: The estimated number of clusters for graphs generated by SSBMs of $N = 1000$ nodes and c clusters with (a) $c = 2$, (b) $c = 3$, (c) $c = 4$, (d) $c = 10$ and (e) $c = 20$. For an SSBM with a given number of clusters c, we vary the difference $b_{in} - b_{out}$ while keeping the expected average degree constant at $E[D] = 7$. Each point in the plots represents an average over 10^4 realizations. The vertical dashed line indicates the theoretical detectability limit of the SSBM.

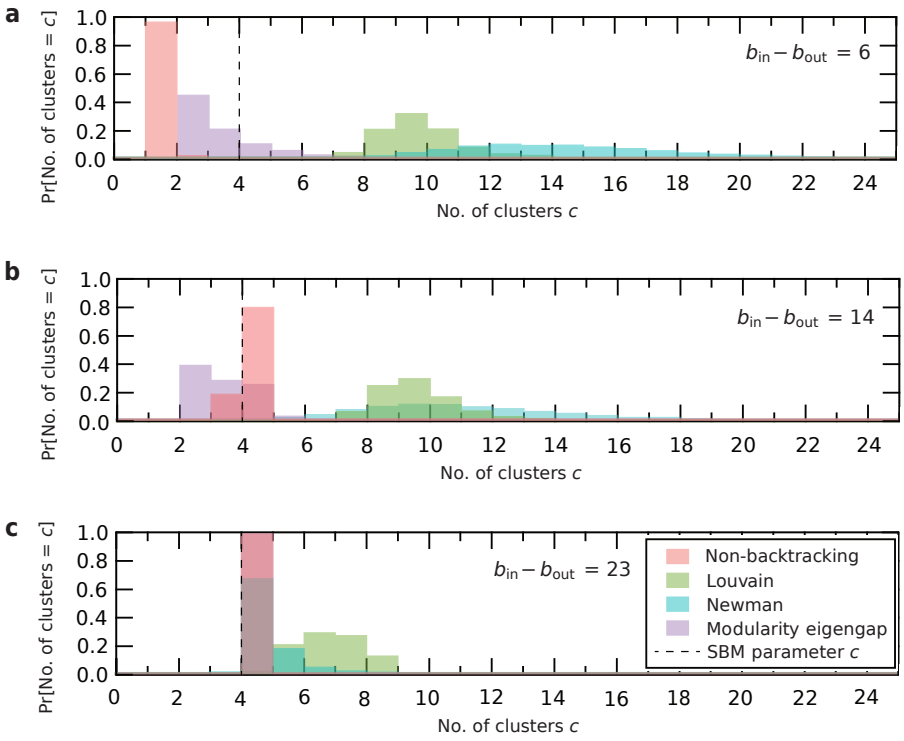


Figure 2.9: Histograms of the estimates of the number of clusters c by the four different methods for graphs generated by SSBMs with $c = 4$ clusters. The values in (a) $b_{in} - b_{out} = 6$, (b) $b_{in} - b_{out} = 14$ and (c) $b_{in} - b_{out} = 23$ correspond to the left (a), center (b) and right (c) of the interval of the experiment in Figure 2.8(c). The estimated probabilities are obtained by evaluating the methods on 10^4 simulated SSBM graphs with $N = 1000$ nodes. In panel (a), none of the methods performs well since the difference $b_{in} - b_{out}$ is below the theoretical detectability limit in this case. In panel (c), both the non-backtracking method and the modularity eigengap detect the right number of clusters for each of the 10^4 networks, therefore the red and purple bars overlap completely.

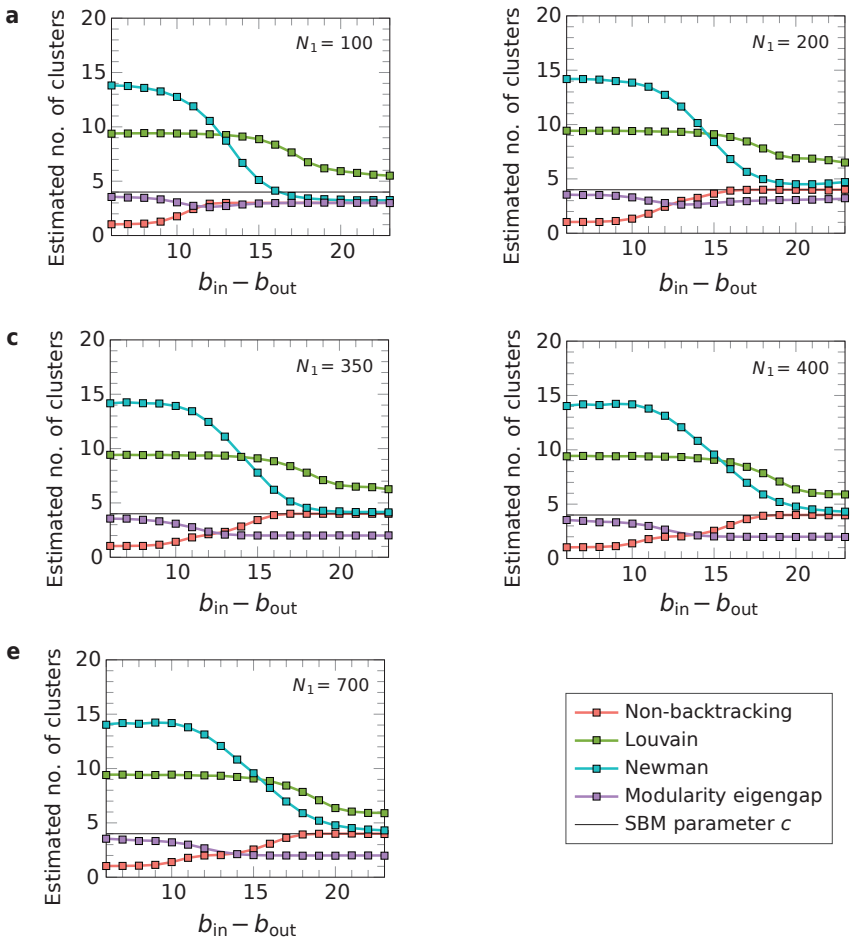


Figure 2.10: The estimated number of clusters for graphs generated by SSBMs of $N = 1000$ nodes subdivided into 4 imbalanced clusters. The size N_1 of cluster 1 is set to deviate from the sizes of the other three approximately equally sized clusters, with (a) $N_1 = 100$, (b) $N_1 = 200$, (c) $N_1 = 350$, (d) $N_1 = 400$ and (e) $N_1 = 700$. The deviation of the size N_1 from the value 250 indicates the degree of imbalancedness. We vary the difference $b_{in} - b_{out}$ exactly like in Figure 2.8(c) for comparability, but because of the imbalanced cluster sizes the average degree $E[D]$ is not equal to 7 anymore. Each point in the plots represents an average over 10^4 realizations.

II

Hyperbolic Network Representations

3

HYPERBOLIC GRAPHS IN $d+1$ DIMENSIONS

They think that intelligence is about noticing things that are relevant;
in a complex world, intelligence consists of ignoring things that are irrelevant.

Nassim Nicholas Taleb

We consider random hyperbolic graphs in hyperbolic spaces of any dimension $d+1 \geq 2$. We introduce a rescaling of the model parameters that casts the random hyperbolic graph model of any dimension to a unified mathematical framework, leaving the degree distribution invariant with respect to the dimension. Unlike the degree distribution, clustering does depend on the dimension, decreasing to 0 at $d \to \infty$. In this Chapter, we analyze all of the other limiting regimes of the model, and we describe our software package that generates random hyperbolic graphs and their limits in hyperbolic spaces of any dimension.

This Chapter is based on G. Budel, M. Kitsak, R. Aldecoa, K. Zuev and D. Krioukov (2024), Random hyperbolic graphs in $d+1$ dimensions, arXiv preprint arXiv:2010.12303 [79].

3.1. INTRODUCTION

Random hyperbolic graphs (RHGs) [21, 80] are a latent space network model [81, 82], in which the latent space is the hyperbolic disk $\mathbb{H}^2$: nodes are random points on the disk, while connections between nodes are established with distance-dependent probabilities. RHGs reproduce many structural properties of real networks including sparsity, self-similarity, power-law degree distribution, strong clustering, small worldness, and community structure [21–25]. They are also exponential random graphs with just two sufficient statistics—the number of links and the sum of their hyperbolic lengths [83]. Using the RHG as a null model, one can map real networks to hyperbolic spaces [28–30], the applications of which include routing and navigation [28, 31–36], link prediction [29, 38–44], network scaling [22, 25, 35], semantic analysis [50, 84–86], and many others [45]. Refer to Appendix B.1 for the main definitions and degree distribution of the 2-dimensional RHG.

While 2-dimensional RHGs have been studied extensively in the literature, their higher-dimensional generalizations have not received as much attention. Here, we aim to fill in this gap by offering a systematic analysis of the RHG model in a hyperbolic space $\mathbb{H}^{d+1}$ of arbitrary dimensionality $d+1 \geq 2$. Apart from purely academic interest, our work is inspired by several practical questions. Hyperbolic spaces expand exponentially for any dimension $d+1 \geq 2$. Thus, intuitively, RHGs should have similar topological properties regardless of the dimensionality of their latent hyperbolic spaces. We aim to verify this intuition here. Second, the dimensionality of the latent space has been shown to affect the accuracy of graph embedding tasks [87]. Finally, a recent study suggests that the dimensionality of the hyperbolic space allows one to generate more realistic and diverse community structures [88].

Spatial graphs in hyperbolic spaces comprise an active area of research and many related results were obtained recently. RHGs are equivalent to geometric inhomogeneous random graphs (GIRGs), as mentioned in [21] and formalized in [89]. This GIRG formulation is followed in [83, 90–92], where the small-world, clustering coefficient, and other properties of GIRGs are analyzed for any dimension d. Yang and Rideout [93] follow the hyperbolic formulation and provide rigorous mathematical derivations of the degree distributions and degree correlations of the high-dimensional RHG model. The related popularity-similarity optimization (PSO) model has recently been extended to arbitrary dimensionality $d+1 \geq 2$ in [94]. Whereas the RHG is a static network model, the $(d+1)$-dimensional PSO model is a growing network model in hyperbolic space possessing similar structural properties.

In this Chapter, we conduct a systematic analysis of the structural properties of RHGs and their limiting regimes. The organization is as follows: in Section 3.2, we define the RHG model in the hyperbolic space $\mathbb{H}^{d+1}$ of any dimension $d+1 \geq 2$. We propose a rescaling of the model parameters that renders the degree distribution invariant with respect to d in Section 3.3, focusing on the three connectivity regimes—cold, critical, and hot—in the model, Section 3.4. In Section 3.5, we analyze the limiting regimes of the model when its parameters tend to their extreme values. Section 3.6 introduces our software package that generates RHGs and their limits for any d, generalizing the $d = 1$ generator in [95]. The concluding remarks are in Section 3.7.

3.2. RANDOM HYPERBOLIC GRAPH MODEL IN $d+1$ DIMENSIONS

A random hyperbolic graph (RHG) in the $(d+1)$-dimensional hyperbolic space $\mathbb{H}^{d+1}$ on N nodes is defined as follows. Each of the N nodes of an RHG corresponds to a point $\mathbf{x}$ in $\mathbb{H}^{d+1}$. The points are the components of the multivariate random variable $\mathbf{X} = (X_0, X_1, \ldots, X_d)$, which are independently and identically distributed according to a distribution $F_{\mathbf{X}}(\mathbf{x})$ prescribed by the model. The links between node pairs in an RHG are also independently and identically distributed with a distribution $F_{\mathbf{A}}(d)$, which is a function of the distances d in $\mathbb{H}^{d+1}$. The connection probability p_{ij} between nodes i and j is $p_{ij} = F_{\mathbf{A}}\left(d\left(\mathbf{x}_i, \mathbf{x}_j\right)\right)$, where $d\left(\mathbf{x}_i, \mathbf{x}_j\right)$ is the distance between the points $\mathbf{x}_i$ and $\mathbf{x}_j$. The $(d+1)$-dimensional RHG model, as a result, is fully defined by its latent space $\mathbb{H}^{d+1}$, coordinate distribution $F_{\mathbf{X}}(\mathbf{x})$, and the connection probability distribution $F_{\mathbf{A}}(d)$.

To justify our choices for the distributions $F_{\mathbf{X}}(\mathbf{x})$ and $F_{\mathbf{A}}(d)$, we first recall the definition of the d-dimensional hyperbolic space and its basic geometric properties. To this end, we consider the upper sheet of the $(d+1)$-dimensional hyperboloid of curvature $K = -\zeta^2$

$$x_0^2 - x_1^2 - \cdots - x_{d+1}^2 = \frac{1}{\zeta^2},\ x_0 > 0, \tag{3.1}$$

in the $(d+2)$-dimensional Minkowski space with metric

$$\mathrm{d}s^2 = -\mathrm{d}x_0^2 + \mathrm{d}x_1^2 + \cdots + \mathrm{d}x_{d+1}^2. \tag{3.2}$$

The spherical coordinate system on the hyperboloid $(r, \theta_1, \ldots, \theta_d)$ is defined by

$$\begin{aligned}
x_0 &= \frac{1}{\zeta}\cosh\zeta r,\\
x_1 &= \frac{1}{\zeta}\sinh\zeta r\cos\theta_1,\\
x_2 &= \frac{1}{\zeta}\sinh\zeta r\sin\theta_1\cos\theta_2,\\
&\vdots\\
x_d &= \frac{1}{\zeta}\sinh\zeta r\sin\theta_1\ldots\sin\theta_{d-1}\cos\theta_d,\\
x_{d+1} &= \frac{1}{\zeta}\sinh\zeta r\sin\theta_1\ldots\sin\theta_{d-1}\sin\theta_d,
\end{aligned} \tag{3.3}$$

where $r > 0$ is the radial coordinate and $(\theta_1, \ldots, \theta_d)$ are the standard angular coordinates on the unit d-dimensional sphere $\mathbb{S}^d$.

The coordinate transformation in (3.3) yields the spherical coordinate metric in the $(d+1)$ -dimensional hyperbolic space $\mathbb{H}^{d+1}$

$$\mathrm{d}s^2 = \mathrm{d}r^2 + \frac{1}{\zeta^2}\sinh^2(\zeta r)\,\mathrm{d}\Omega_d^2, \tag{3.4a}$$

$$\mathrm{d}\Omega_d^2 = \mathrm{d}\theta_1^2 + \sin^2(\theta_1)\mathrm{d}\theta_2^2 + \cdots + \sin^2(\theta_1)\ldots\sin^2(\theta_{d-1})\mathrm{d}\theta_d^2, \tag{3.4b}$$

resulting in the volume element in $\mathbb{H}^{d+1}$:

$$\mathrm{d}V = \left(\frac{1}{\zeta}\sinh\zeta r\right)^d \mathrm{d}r \prod_{k=1}^{d} \sin^{d-k}(\theta_k)\mathrm{d}\theta_k. \tag{3.5}$$

The distance d_{ij} between two points i and j in $\mathbb{H}^{d+1}$ is given by the hyperbolic law of cosines:

$$\cosh\zeta d_{ij} = \cosh\zeta r_i \cosh\zeta r_j - \sinh\zeta r_i \sinh\zeta r_j \cos\Delta\theta_{ij}, \tag{3.6}$$

where $\Delta\theta_{ij}$ is the angle between points i and j:

$$\begin{aligned}\cos(\Delta\theta_{ij}) = &\cos\theta_{i,1}\cos\theta_{j,1} + \sin\theta_{i,1}\sin\theta_{j,1}\cos\theta_{i,2}\cos\theta_{j,2} + \dots \\ &+ \sin\theta_{i,1}\sin\theta_{j,1}\dots\sin\theta_{i,d-1}\sin\theta_{j,d-1}\cos\theta_{i,d}\cos\theta_{j,d} \\ &+ \sin\theta_{i,1}\sin\theta_{j,1}\dots\sin\theta_{i,d-1}\sin\theta_{j,d-1}\sin\theta_{i,d}\sin\theta_{j,d},\end{aligned} \tag{3.7}$$

where $(\theta_{i,1},\dots,\theta_{i,d})$ and $(\theta_{j,1},\dots,\theta_{j,d})$ are the coordinates of points i and j on $\mathbb{S}^d$.

For sufficiently large values of ζr_i and ζr_j, the hyperbolic law of cosines in (3.6) is closely approximated by

$$\tilde{d}_{ij} = r_i + r_j + \frac{2}{\zeta}\ln\left(\sin(\Delta\theta_{ij}/2)\right), \tag{3.8}$$

such that $d_{ij} \approx \tilde{d}_{ij}$, like in the 2-dimensional case (B.6). The hyperbolic ball $\mathbb{B}^{d+1}$ of radius $R > 0$ is defined as the set of points with radial coordinates

$$r \in [0, R]. \tag{3.9}$$

The nodes of an RHG are points in $\mathbb{B}^{d+1}$ selected independently at random with multivariate density $f_{\mathbf{X}}(\mathbf{x}) \equiv f_{X_0}(r) f_{X_1}(\theta_1)\dots f_{X_d}(\theta_d)$, where

$$f_{X_0}(r) = [\sinh(\alpha r)]^d / C_d, \qquad \alpha \geq \zeta/2, \tag{3.10a}$$

$$C_d \equiv \int_0^R [\sinh(\alpha r)]^d \,\mathrm{d}r, \tag{3.10b}$$

$$f_{X_k}(\theta_k) = [\sin(\theta_k)]^{d-k} / I_{d,k}, \qquad k = 1,\dots,d, \tag{3.10c}$$

$$I_{d,k} \equiv \int_0^\pi [\sin(\theta_k)]^{d-k}\,\mathrm{d}\theta_k = \sqrt{\pi}\frac{\Gamma[\frac{d-k+1}{2}]}{\Gamma\left[1+\frac{d-k}{2}\right]}, \qquad k < d, \tag{3.10d}$$

$$I_{d,d} \equiv 2\pi. \tag{3.10e}$$

In other words, the points are uniformly distributed on the unit sphere $\mathbb{S}^d$ with respect to their angular coordinates. In the special case of $\alpha = \zeta$, the points are also uniformly distributed in $\mathbb{B}^{d+1}$.

Pairs of nodes i and j are connected independently with connection probability

$$p_{ij} = F_{\mathbf{A}}(d_{ij}) = \frac{1}{1+\exp\left(\zeta\left(d_{ij}-\mu\right)/2T\right)}, \tag{3.11}$$

where $\mu > 0$ and $T > 0$ are model parameters, and d_{ij} is the distance between points i and j in $\mathbb{B}^{d+1}$, given by (3.6). We refer to parameters T and μ as temperature and chemical potential, respectively, using the analogy with the Fermi-Dirac statistics. We note that the factors of 2 and ζ in (3.11) are to agree with the 2-dimensional RHG that corresponds to $d = 1$, Appendix B.1.

Thus, an RHG is formed in a three-step network generation process:

1. Randomly select N points in $\mathbb{B}^{d+1}$ with probability density functions $f_{\mathbf{X}}(\mathbf{x})$ in (3.10).
2. Calculate distances in $\mathbb{B}^{d+1}$ between all node pairs (i, j) using (3.6).
3. Connect node pairs (i, j) independently at random with distance-dependent connection probabilities $p_{ij} = F_{\mathbf{A}}(d_{ij})$, prescribed by (3.11).

Taken together, $(d+1)$-dimensional RHGs are fully defined by the following 6 parameters: the properties of the hyperbolic ball, R and ζ; the number of nodes N; the radial component of node distribution α; the chemical potential μ and the temperature T. Figure 3.1 visualizes a subgraph of an RHG for $d = 2$.

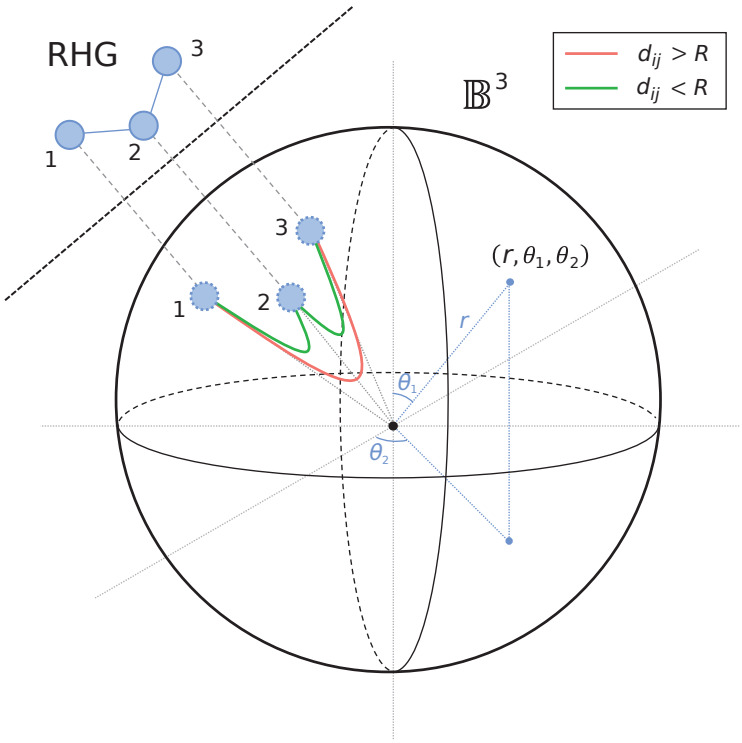


Figure 3.1: A visualization of an RHG for $d = 2$ in $\mathbb{B}^3$. The spherical coordinates (r, θ_1, θ_2) of a point in $\mathbb{B}^3$ are shown. A subgraph of 3 nodes is shown with geodesics between the nodes. Node pairs (1,2) and (2,3) are connected with high probability because $d_{12} < R$ and $d_{23} < R$ (green), while node pair (1,3) is disconnected with high probability because $d_{23} > R$ (red).

Only the four parameters $\{N, \alpha, T, R\}$ are independent, however. It follows from (3.6) that the negative curvature ζ is merely a rescaling parameter for the distances d_{ij}, and can be absorbed into the radial coordinates r by the appropriate rescaling. The chemical potential μ controls the expected number of links and the sparsity of the resulting graphs. We demonstrate below that the sparsity requirement uniquely determines R and μ as a function of the number of nodes N, $R = R(N)$, $\mu = \mu(N)$.

3.3. DEGREE DISTRIBUTION AND CLUSTERING COEFFICIENT

The structural properties of an RHG can be computed by making use of the hidden variable framework of Boguñá and Pastor-Satorras [96], by treating the node coordinates as hidden variables. We begin by calculating the expected degree of a node l which is located at a point $\mathbf{x}_l = \left(r_l, \theta_{1,l}, \ldots, \theta_{d,l}\right)$,

$$E[D\,|\,\mathbf{X} = \mathbf{x}_l] = (N-1)\int \frac{\mathrm{d}\mathbf{x}_k f_{\mathbf{X}_k}(\mathbf{x}_k)}{1 + e^{\frac{\zeta(d_{lk}-\mu)}{2T}}}. \tag{3.12}$$

The symmetry in the angular distribution of points ensures that the expected degree of the node depends only on its radial coordinate r_l and not on its angular coordinates, $E[D\,|\,\mathbf{X} = \mathbf{x}] = E[D\,|\,\mathbf{X} = (r_l, 0, ..., 0)] \equiv E[D\,|\,X_0 = r_l]$. This allows us to integrate out d angular coordinates in (3.12).

We also note that the choice of the radial coordinate distribution given by (3.10a) with $\alpha \geq \zeta/2$ results in most of the nodes having large radial coordinates, $r_i \approx R \gg 1$. This fact allows us to approximate the hyperbolic distances using (3.8):

$$E[D\,|\,X_0 = r_l] = \int_0^R \int_0^\pi \frac{(N-1) f_{X_0}(r)\mathrm{d}r\, f_{X_1}(\theta_1)\mathrm{d}\theta_1}{1 + \left[e^{\zeta(r+r_l-\mu)}\left(\sin\left(\frac{\theta_1}{2}\right)\right)^2\right]^{\frac{1}{2T}}}. \tag{3.13}$$

To further simplify our calculations, we perform the following change of variables,

$$\mathfrak{X}_0 \equiv \frac{d\zeta}{2} X_0, \tag{3.14}$$

and we rescale the model parameters,

$$\mathscr{R} = \frac{d\zeta}{2} R \tag{3.15a}$$

$$\mathfrak{m} = \frac{d\zeta}{2} \mu, \tag{3.15b}$$

$$\tau = d\,T, \tag{3.15c}$$

$$a = \frac{2\alpha}{\zeta}. \tag{3.15d}$$

such that the transformed radial coordinate $\mathfrak{r} = \frac{1}{2} d\zeta r$ follows the distribution $F_{\mathfrak{X}_0}(\mathfrak{r})$ with probability density function

$$f_{\mathfrak{X}_0}(\mathfrak{r}) = \frac{a}{C_d}\left[\sinh\left(\frac{a\mathfrak{r}}{d}\right)\right]^d, \qquad 0 \leq \mathfrak{r} \leq \mathscr{R}, \tag{3.16}$$

where C_d is defined in (3.10b). In terms of the transformed variable and the rescaled parameters, the connection probability is

$$p_{ij} = \frac{1}{1 + e^{\frac{\mathfrak{r}_i + \mathfrak{r}_j - \mathfrak{m}}{\mathfrak{T}}} \left[\sin\left(\frac{\Delta\theta_{ij}}{2}\right)\right]^{\frac{d}{\mathfrak{T}}}}, \tag{3.17}$$

while (3.13) now reads

$$E[D \mid \mathfrak{X}_0 = \mathfrak{r}_I] = \int_0^{\mathscr{R}} \int_0^{\pi} \frac{(N-1) f_{\mathfrak{X}_0}(\mathfrak{r}) \mathrm{d}\mathfrak{r} f_{\Theta_1}(\theta_1) \mathrm{d}\theta_1}{1 + e^{\frac{\mathfrak{r} + \mathfrak{r}_I - \mathfrak{m}}{\mathfrak{T}}} \left[\sin\left(\frac{\theta_1}{2}\right)\right]^{\frac{d}{\mathfrak{T}}}}. \tag{3.18}$$

In our analysis, we operate with $\mathscr{R} \gg 1$ and $a > 1$. In this case, most nodes are characterized by large $\mathfrak{r}$ values and the probability density $f_{\mathfrak{X}_0}(\mathfrak{r})$ is well-approximated as

$$\tilde{f}_{\mathfrak{X}_0}(\mathfrak{r}) = a e^{a(\mathfrak{r} - \mathscr{R})}, \tag{3.19}$$

such that $f_{\mathfrak{X}_0}(\mathfrak{r}) \approx \tilde{f}_{\mathfrak{X}_0}(\mathfrak{r})$ for $\mathfrak{r} \in [0, \mathscr{R}]$. The function $\tilde{f}_{\mathfrak{X}_0}(\mathfrak{r})$ in (3.19) is only approximately a proper probability density function under the assumptions of $\mathscr{R} \gg 1$ and $a > 1$, as

$$\int_0^{\mathscr{R}} \tilde{f}_{\mathfrak{X}_0}(\mathfrak{r}) d\mathfrak{r} = 1 - e^{-a\mathscr{R}} \approx 1. \tag{3.20}$$

The expected degree of an arbitrary node in the graph is given by

$$E[D] = \int_0^{\mathscr{R}} \mathrm{d}\mathfrak{r} f_{\mathfrak{X}_0}(\mathfrak{r}) E[D \mid \mathfrak{X}_0 = \mathfrak{r}], \tag{3.21}$$

and the degree distribution density of an RHG can be expressed as

$$\Pr[D = k] = \int_0^{\mathscr{R}} \mathrm{d}\mathfrak{r} f_{\mathfrak{X}_0}(\mathfrak{r}) \Pr[D = k \mid \mathfrak{X}_0 = \mathfrak{r}], \tag{3.22}$$

where $\Pr[D = k \mid \mathfrak{X}_0 = \mathfrak{r}]$ is the conditional probability that a node with radial coordinate $\mathfrak{r}$ has degree k.

Like in the 2-dimensional case in Appendix B.1, for sparse graphs, the density $\Pr[D = k \mid \mathfrak{X}_0 = \mathfrak{r}]$ can be replaced by its Poisson approximation:

$$\Pr[D = k \mid \mathfrak{X}_0 = \mathfrak{r}] \approx \frac{1}{k!} e^{-\lambda(\mathfrak{r})} [\lambda(\mathfrak{r})]^k, \tag{3.23}$$

with the rate $\lambda(\mathfrak{r}) = E[D \mid \mathfrak{X}_0 = \mathfrak{r}]$, see Boguñá and Pastor-Satorras [96]. The resulting degree distribution $\Pr[D = k]$ is approximately a mixed Poisson distribution with the probability density

$$\Pr[D = k] \approx \frac{1}{k!} \int_0^{\mathscr{R}} \mathrm{d}\mathfrak{r}\, e^{-\lambda(\mathfrak{r})} [\lambda(\mathfrak{r})]^k f_{\mathfrak{X}_0}(\mathfrak{r}), \tag{3.24}$$

with mixing parameter $\lambda(\mathfrak{r}) = E[D \mid \mathfrak{X}_0 = \mathfrak{r}]$.

The clustering coefficient $c_G(i)$ of a node i in graph G with degree $k_i > 1$ is defined as the ratio of the number of connected neighbors y over the maximum possible number of connected neighbors

$$c_G(i) = \frac{2y}{k_i(k_i-1)}, \tag{3.25}$$

as defined in [97]. Since nodes with degrees $k = 0$ and $k = 1$ do not form triangles, their clustering coefficients are undefined. The graph clustering coefficient c_G is then the average of the node clustering coefficients $c_G(i)$ over all nodes with degree $k > 1$,

$$c_G = \frac{1}{N_{k>1}} \sum_{i=1}^{N} c_G(i), \tag{3.26}$$

where $N_{k>1}$ is the number of nodes with degree $k > 1$. The hidden variable framework [96] allows for expressing the graph clustering coefficient of the RHG as a multiple integral over triples of hidden variables. Apart from special cases of the RHG, however, these integrals do not have simple closed-form solutions [98]. Therefore, we restrict ourselves to studying the clustering coefficients of RHGs numerically in this work.

3.4. CONNECTIVITY REGIMES OF THE RHG

Depending on the value of the rescaled temperature $\tau = dT$, there exist three distinct regimes of the RHG: (i) cold regime ($\tau < 1$), (ii) critical regime ($\tau = 1$), and (iii) hot regime ($\tau > 1$). We provide detailed analyses of the properties of RHGs in these regimes below, and summarize our findings in Figure 3.16.

3.4.1. COLD REGIME, $\tau < 1$

Since the inner integral in (3.18) does not have a closed-form solution, to estimate $E[D \,|\, \mathfrak{X}_0 = \mathfrak{r}_l]$ we need to employ several approximations. We note that most nodes have large radial coordinates, $e^{\mathfrak{r}+\mathfrak{r}_l-\mathfrak{m}} \gg 1$, and the dominant contribution to the inner integral in (3.18) comes from small θ_1 values. This allows us to estimate the integral by replacing $\sin(\theta_1)$ and $\sin(\theta_1/2)$ with the leading Taylor series terms, as $\sin(x) = x + \mathcal{O}\left(x^3\right)$, resulting in

$$E[D \,|\, \mathfrak{X}_0 = \mathfrak{r}_l] \approx \frac{(N-1)\pi^d}{d I_{d,1}} \int_0^{\mathscr{R}} \mathrm{d}\mathfrak{r}\, \tilde{f}_{\mathfrak{X}_0}(\mathfrak{r})\, {}_2F_1\left(1, \tau, 1+\tau, -[u_{\max}(\mathfrak{r}_l, \mathfrak{r}, \mathfrak{m})]^{\frac{1}{\tau}}\right), \tag{3.27}$$

with

$$u_{\max}(\mathfrak{r}_l, \mathfrak{r}, \mathfrak{m}) \equiv \left(\frac{\pi}{2}\right)^d e^{\mathfrak{r}_l+\mathfrak{r}-\mathfrak{m}}, \tag{3.28}$$

and where ${}_2F_1$ is the Gauss hypergeometric function, and

$$I_{d,1} \equiv \int_0^{\pi} \sin^{d-1}(\theta_1)\mathrm{d}\theta_1 = \sqrt{\pi}\,\Gamma\left[\frac{d}{2}\right] / \Gamma\left[\frac{d+1}{2}\right], \tag{3.29}$$

for $d > 1$, and $I_{1,1} = 2\pi$. In the $\tau < 1$ regime, the hypergeometric function in (3.27) can be approximated as

$${}_2F_1\left(1, \tau, 1+\tau, -[u_{\max}(\mathfrak{r}_l, \mathfrak{r}, \mathfrak{m})]^{\frac{1}{\tau}}\right) \approx [u_{\max}(\mathfrak{r}_l, \mathfrak{r}, \mathfrak{m})]^{-1} \frac{\pi\tau}{\sin(\pi\tau)}, \tag{3.30}$$

and $E[D\,|\,\mathfrak{X}_0=\mathfrak{r}_l]$ and $E[D]$ are then approximately given by:

$$E[D] \approx (N-1)\frac{2^d}{dI_{d,1}}\frac{\pi\tau}{\sin(\pi\tau)}E\left[e^{-\mathfrak{X}_0}\right]^2 e^{\mathfrak{m}}, \tag{3.31a}$$

$$E[D\,|\,\mathfrak{X}_0=\mathfrak{r}] \approx \frac{E[D]}{E\left[e^{-\mathfrak{X}_0}\right]}e^{-\mathfrak{r}}, \tag{3.31b}$$

where $E\left[e^{-\mathfrak{X}_0}\right] \equiv \int_0^{\mathscr{R}} \mathrm{d}\mathfrak{r} f_{\mathfrak{X}_0}(\mathfrak{r})e^{-\mathfrak{r}} \approx \int_0^{\mathscr{R}} \mathrm{d}\mathfrak{r}\tilde{f}_{\mathfrak{X}_0}(\mathfrak{r})e^{-\mathfrak{r}}$, and an explicit expression for $E\left[e^{-\mathfrak{X}_0}\right]$ follows from (3.16),

$$E\left[e^{-\mathfrak{X}_0}\right] \approx \frac{a}{a-1}\left(e^{-\mathscr{R}}-e^{-a\mathscr{R}}\right). \tag{3.32}$$

Next, we discuss the choice for the rescaled chemical potential $\mathfrak{m}$. To do so, we discuss the leading order behavior of $E[D]$ in the thermodynamic limit. Since $a>1$, we neglect the second term in (3.32) to obtain

$$E[D\,|\,\mathfrak{X}_0=\mathfrak{r}] \sim Ne^{\mathfrak{m}-\mathfrak{r}-\mathscr{R}}, \tag{3.33a}$$

$$E[D] \sim Ne^{\mathfrak{m}-2\mathscr{R}}. \tag{3.33b}$$

Henceforth, we write $f(x)\sim g(x)$ when $\lim_{x\to\infty}\frac{f(x)}{g(x)}=K$, where $K>0$ is a constant.

We note that $E[D\,|\,\mathfrak{X}_0=\mathfrak{r}]$ decreases exponentially as a function of $\mathfrak{r}$ with the largest expected degree $E[D_{\max}]$ corresponding to $\mathfrak{r}=0$, while the smallest expected degree $E[D_{\min}]$ corresponds to $\mathfrak{r}=\mathscr{R}$. By demanding that the largest and smallest expected degrees scale as

$$E[D_{\max}] = E[D\,|\,\mathfrak{X}_0=0] \sim N, \tag{3.34a}$$

$$E[D_{\min}] = E[D\,|\,\mathfrak{X}_0=\mathscr{R}] \sim 1, \tag{3.34b}$$

we obtain $\mathscr{R}\sim\ln N$ and $\mathfrak{m}=\mathscr{R}+\lambda$, where λ is an arbitrary constant.

First, we note that the scaling for $\mathscr{R}$ is consistent with our initial assumption of $\mathscr{R}\gg 1$ for large graphs. We also note that the exact value of the parameter λ is not important as long as it is independent of the number of nodes N. To be consistent with the original $\mathbb{H}^2$ formulation, Appendix B.1, we set $\lambda=0$, obtaining

$$\mathfrak{m}=\mathscr{R}=\ln\left(\frac{N}{\nu}\right), \tag{3.35}$$

where ν is the parameter controlling the expected degree of the RHG. Applying the scaling relationships (3.35) to (3.31), we obtain

$$E[D] \approx \frac{\nu 2^d}{dI_{d,1}}\left(\frac{a}{a-1}\right)^2\frac{\pi\tau}{\sin(\pi\tau)}\left[1-2\left(\frac{N}{\nu}\right)^{1-a}+\left(\frac{N}{\nu}\right)^{2(1-a)}\right], \tag{3.36}$$

and

$$E[D\,|\,\mathfrak{X}_0=\mathfrak{r}] \approx \frac{N}{\nu}\left(\frac{a-1}{a}\right)E[D]\,e^{-\mathfrak{r}}\sum_{\ell=0}^{\infty}\left(\frac{\nu}{N}\right)^{\ell(a-1)}. \tag{3.37}$$

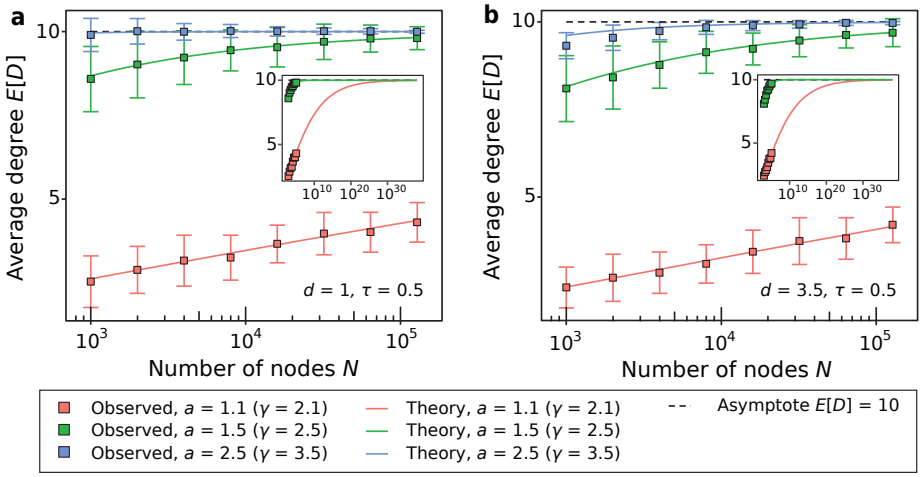


Figure 3.2: The average degree $E[D]$ as a function of the number of nodes N for RHGs in the cold regime ($\tau = 0.5$) with (a) $d = 1$ and (b) $d = 3$. Each panel includes the results for (red) $a = 1.1$ ($\gamma = 2.1$), (green) $a = 1.5$ ($\gamma = 2.5$), and (blue) $a = 2.5$ ($\gamma = 3.5$). Each point is the average of 100 simulations and the error bars display standard deviations. The solid lines are the theoretical values for average degree $E[D]$ prescribed by (3.18) and (3.21), and the dashed line is the thermodynamic limit of (3.36). The insets in (a) and (b) correspond to extended domains of N values for the cases $a = 1.1$ and $a = 1.5$. Note that the $a = 1.1$ case converges to the asymptotic value at a much slower rate compared to the $a = 1.5$ and $a = 2.5$ cases.

As seen from (3.36) and Figure 3.2, RHGs are sparse in the cold regime ($\tau < 1$). We call graphs sparse if their expected degree converges to a finite constant in the thermodynamic limit. The slow convergence to the asymptotic value of $E[D] = 10$ in the $a = 1.1$ case is due to the breakdown of the $\sin(x) \approx x$ approximation in (3.27) for small a values. Indeed, at small a values, a larger fraction of nodes is characterized by small $\mathfrak{r}$ values, for which the $e^{(\mathfrak{r}+\mathfrak{r}_l-\mathfrak{m})} \gg 1$ assumption fails.

Finally, using (3.23) and (3.24) we obtain the Pareto-mixed Poisson distribution, which is a power law

$$\Pr[D = k] \approx a\kappa_0^a \frac{\Gamma[k-a,\kappa_0]}{\Gamma[k+1]} \sim k^{-(a+1)}, \tag{3.38}$$

where $\Gamma[s,x]$ is the upper incomplete gamma function, and $\kappa_0 \equiv \left(\frac{a-1}{a}\right) E[D]$, as confirmed by simulations in Figure 3.3.

Hence, the cold regime corresponds to sparse power-law graphs, $\Pr[D = k] \sim k^{-\gamma}$, with $\gamma = a+1 \in (2,\infty)$. We note that the degree distribution is a power law if it takes the form of $\Pr[D = k] = \ell(k)k^{-\gamma}$, where $\ell(k)$ is a slowly varying function, *i.e.*, a function that varies slowly at infinity, see [99]. Any function converging to a constant is slowly varying, and, in the case of (3.38), $\ell(k) \to a\kappa_0^a$ as $k \to \infty$. The degree distribution is called scale-free if $\gamma \in (2,3)$.

The special case of $a = 1$ ($\gamma = 2$) is also well defined. In this case the expressions for $E[D]$ and $E[D|\mathfrak{X}_0 = \mathfrak{r}]$ given by (3.31) remain valid but $E\left[e^{-\mathfrak{X}_0}\right]$ is now given by

$$E\left[e^{-\mathfrak{X}_0}\right] \approx \mathscr{R}e^{-\mathscr{R}}. \tag{3.39}$$

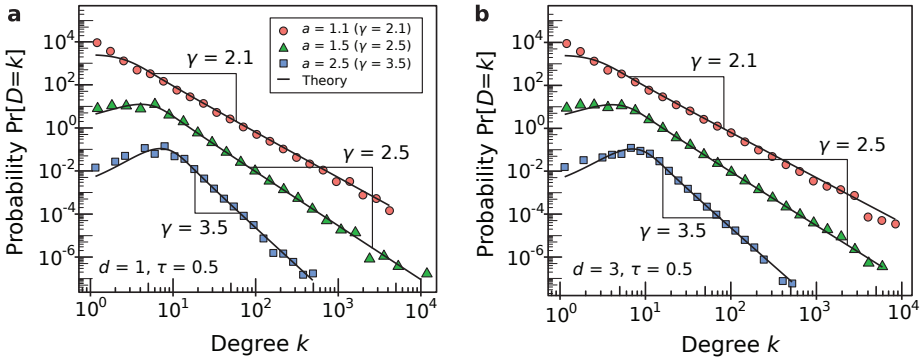


Figure 3.3: The degree distribution densities for RHGs with (a) $d = 1$ and (b) $d = 3$ at $\tau = 0.5$ and $N = 1000 \cdot 2^7$. Each panel includes the densities for (red) $a = 1.1$ ($\gamma = 2.1$), (green) $a = 1.5$ ($\gamma = 2.5$), and (blue) $a = 2.5$ ($\gamma = 3.5$). The probabilities $\Pr[D = k]$ are obtained from a single network realization. The degree distribution densities are binned logarithmically to suppress noise at large k values. The solid lines are the theoretical values for $\Pr[D = k]$ prescribed by (3.38). For visibility, the probabilities corresponding to $\gamma = 2.5$ and $\gamma = 2.1$ are multiplied by 10^2 and 10^4, respectively. The scaling constant $\nu = 10 \times \frac{dI_{d,1}}{2^d}\left(\frac{a-1}{a}\right)^2 \frac{\sin(\pi\tau)}{\pi\tau}$, corresponding to $E[D] = 10$ in the thermodynamic limit.

The scaling of $\mathscr{R} = \mathfrak{m} = \ln(N/\nu)$ does not lead to the desired calibration of node degrees in the $a = 1$ case, $E[D_{\max}] \sim N$ and $E[D_{\min}] \sim 1$. Instead, the proper scaling for $a = 1$ is

$$\mathscr{R} = \ln(N/\nu), \tag{3.40a}$$
$$\mathfrak{m} = \mathscr{R} - \ln\mathscr{R}, \tag{3.40b}$$

resulting in the approximations

$$E[D] \approx \frac{\nu 2^d}{dI_{d,1}} \frac{\pi\tau}{\sin(\pi\tau)} \ln\left(\frac{N}{\nu}\right), \tag{3.41a}$$
$$E[D \,|\, \mathfrak{X}_0 = \mathfrak{r}] \approx \frac{N}{\nu} \frac{E[D]}{\ln(N/\nu)} e^{-\mathfrak{r}}. \tag{3.41b}$$

In other words, the $a = 1$ ($\gamma = 2$) case corresponds to graphs with $E[D] \sim \ln(N/\nu)$, as confirmed by Figure 3.5. The degree distribution approximately matches a power-law with $\gamma = 2$, as shown in Figure 3.6. The phenomenon that RHGs are no longer sparse in the $a = 1$ ($\gamma = 2$) case is not specific to the model but is a general property of all scale-free network models with $\Pr[D = k] \sim k^{-2}$.

Here, we do not attempt to obtain analytical expressions of the clustering coefficient for $d \geq 1$, as its computation is quite involved already in the $d = 1$ case [26, 100], but we study the graph clustering coefficient numerically as functions of both the dimension and the network size, as shown in Figure 3.4. In the cold regime $\tau < 1$, we observe that clustering is non-vanishing in the thermodynamic limit, similar to the $d = 1$ case. The graph clustering coefficient decreases as a function of dimension d. This property of the clustering coefficient has been observed in other spatial graph models as well,

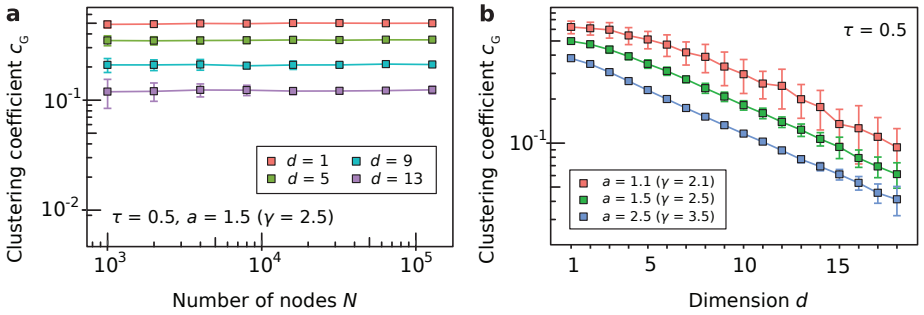


Figure 3.4: The graph clustering coefficient c_G of nodes with degree $k > 1$ as a function of the number of nodes N (a) and as a function of the dimension d (b) at $\tau = 0.5$. Panel (a) includes results for (red) $d = 1$, (green) $d = 5$, (blue) $d = 9$ and (purple) $d = 13$, for $a = 1.5$ ($\gamma = 2.5$), while panel (b) includes results for (red) $a = 1.1$ ($\gamma = 2.1$), (green) $a = 1.5$ ($\gamma = 2.5$), and (blue) $a = 2.5$ ($\gamma = 3.5$), for $N = 10^4$. Each point is the average of 100 simulations and the error bars display standard deviations. The scaling constant $\nu = 10 \times \frac{dI_{d,1}}{2^d} \left(\frac{a-1}{a}\right)^2 \frac{\sin(\pi\tau)}{\pi\tau}$, corresponding to $E[D] = 10$ in the thermodynamic limit.

particularly, in GIRGS [91]. This is because the angular distance distribution between two random points on a d-sphere [101] approaches the Dirac delta function centered at $\pi/2$. As a result, the role of a node's angular coordinates in the hyperbolic distances diminishes, and the network becomes 'less geometric' and more similar to the hypersoft configurational model (HSCM), Section 3.5.6.

GRAPH PROPERTY PERSPECTIVE, $\tau < 1$

From a graph property viewpoint, the RHG is instrumental in generating synthetic networks with desired properties. It follows from our analysis that, in the cold regime ($\tau < 1$), RHGs are characterized by a power-law degree distribution, $\Pr[D = k] \sim k^{-\gamma}$, where the exponent $\gamma \in [2, \infty)$ is a function of RHG temperature τ and node density parameter a. The radius $\mathscr{R}$ of the hyperbolic ball $\mathbb{B}^{d+1}$, on the other hand, controls the average degree $E[D]$ and the sparsity of the resulting graphs. Relying on these results, we can redefine the RHG model in terms of the parameters $\{N, \gamma, \tau, E[D]\}$.

To generate an RHG in the cold regime with desired graph properties for $\gamma > 2$, one sets the node density parameter a, the chemical potential $\mathfrak{m}$, and radius $\mathscr{R}$ of $\mathbb{B}^{d+1}$ to

$$a = \gamma - 1, \tag{3.42a}$$
$$\mathfrak{m} = \mathscr{R} = \ln(N/\nu), \tag{3.42b}$$

where ν is the solution of (3.36), now taking the form of

$$E[D] = \frac{\nu 2^d}{dI_{d,1}} \left(\frac{\gamma-1}{\gamma-2}\right)^2 \frac{\pi\tau}{\sin(\pi\tau)} \left[1 - 2\left(\frac{N}{\nu}\right)^{2-\gamma} + \left(\frac{N}{\nu}\right)^{2(2-\gamma)}\right]. \tag{3.43}$$

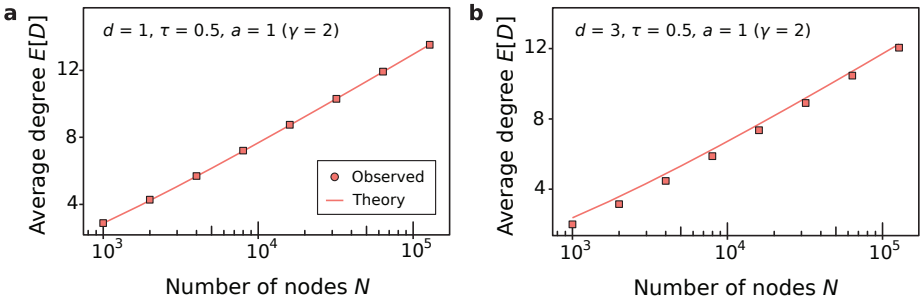


Figure 3.5: The average degree $E[D]$ of RHGs with (a) $d = 1$ and (b) $d = 3$ in the case of $a = 1$ at $\tau = 0.5$. Each point is the average of 100 simulations and the error bars display standard deviations (not visible for most points). To avoid fluctuations associated with large degree nodes, we have imposed a cutoff in the radial coordinate distribution, removing nodes with $\mathfrak{r} \leq \mathfrak{r}_{\text{cut}}$, where $\mathfrak{r}_{\text{cut}}$ is defined such that $E[D\,|\,\mathfrak{X}_0 = \mathfrak{r}_{\text{cut}}] = \sqrt{N}$. The solid lines are the theoretical values for $E[D]$ prescribed by (3.41a), corrected for the radial coordinate cutoff. The scaling constant is set to $\nu = 10 \times \frac{dI_{d,1}}{2^d} \frac{\sin(\pi\tau)}{\pi\tau}$.

When $\gamma = 2$ in the cold regime, one must set

$$a = \gamma - 1 = 1, \tag{3.44a}$$

$$\mathscr{R} = \ln(N/\nu), \tag{3.44b}$$

$$\mathfrak{m} = \mathscr{R} - \ln\mathscr{R}, \tag{3.44c}$$

and ν is obtained as the solution of (3.41a), which now takes the form of

$$E[D] = \frac{\nu 2^d}{dI_{d,1}} \frac{\pi\tau}{\sin(\pi\tau)} \ln\left(\frac{N}{\nu}\right). \tag{3.45}$$

3.4.2. CRITICAL REGIME, $\tau = 1$

In the critical regime ($\tau = 1$), equations (3.27) and (3.21) can be approximated as:

$$E[D] \approx \frac{(N-1)2^d}{dI_{d,1}} e^{\mathfrak{m}} \left[d\ln\left(\frac{\pi}{2}\right) E\left[e^{-\mathfrak{X}_0}\right]^2 + 2E\left[e^{-\mathfrak{X}_0}\right] E\left[\mathfrak{X}_0 e^{-\mathfrak{X}_0}\right] - \mathfrak{m} E\left[e^{-\mathfrak{X}_0}\right]^2 \right], \tag{3.46a}$$

$$E[D\,|\,\mathfrak{X}_0 = \mathfrak{r}] \approx \frac{(N-1)2^d}{dI_{d,1}} e^{\mathfrak{m}-\mathfrak{r}} \left[d\ln\left(\frac{\pi}{2}\right) E\left[e^{-\mathfrak{X}_0}\right] + (\mathfrak{r}-\mathfrak{m}) E\left[e^{-\mathfrak{X}_0}\right] + E\left[\mathfrak{X}_0 e^{-\mathfrak{X}_0}\right] \right], \tag{3.46b}$$

where $E\left[\mathfrak{X}_0 e^{-\mathfrak{X}_0}\right] \equiv \int_0^{\mathscr{R}} d\mathfrak{r} f_{\mathfrak{X}_0}(\mathfrak{r}) e^{-\mathfrak{r}}$ is given by

$$E\left[\mathfrak{X}_0 e^{-\mathfrak{X}_0}\right] = \left(\frac{a}{a-1}\right) \left[\left(\mathscr{R} - \frac{1}{a-1}\right) e^{-\mathscr{R}} + \frac{1}{a-1} e^{-a\mathscr{R}} \right]. \tag{3.47}$$

Given that $a > 1$, we drop the second terms in (3.32) and (3.47) to obtain:

$$E[D\,|\,\mathfrak{X}_0 = \mathfrak{r}] \approx \frac{(N-1)2^d}{dI_{d,1}} \left(\frac{a}{a-1}\right) \left(d\log\left(\frac{\pi}{2}\right) - \frac{1}{a-1} + \mathscr{R} - \mathfrak{m} + \mathfrak{r} \right) e^{\mathfrak{m}-\mathscr{R}-\mathfrak{r}}. \tag{3.48}$$

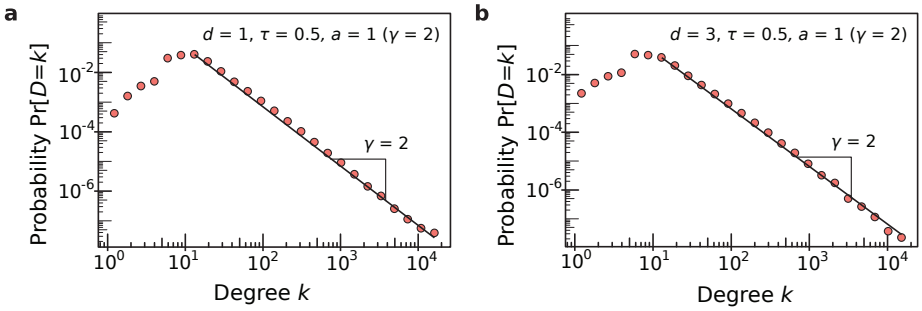


Figure 3.6: The degree distribution densities for RHGs with (a) $d = 1$ and (b) $d = 3$ in the case of $a = 1$ at $\tau = 0.5$ and $N = 1000 \cdot 2^7$. The probabilities $\Pr[D = k]$ are obtained from a single network realization. The densities are binned logarithmically to suppress noise at large k values. To avoid fluctuations associated with large degree nodes, we have imposed a cutoff in the radial coordinate distribution, removing nodes with $\mathfrak{r} \leq \mathfrak{r}_{\text{cut}}$, where $\mathfrak{r}_{\text{cut}}$ is defined such that $E[D \,|\, \mathfrak{X}_0 = \mathfrak{r}_{\text{cut}}] = \sqrt{N}$. The solid lines are the theoretical values for $\Pr[D = k]$ based on a slope of $\gamma = 2$. The scaling constant is set to $\nu = 10 \times \frac{dI_{d,1}}{2^d} \frac{\sin(\pi\tau)}{\pi\tau}$.

Similar to the $\tau < 1$ regime, we demand $E[D_{\max}] \sim N$ and $E[D_{\min}] \sim 1$ to obtain the scaling relationships for $\mathfrak{m}$ and $\mathscr{R}$. For $E[D_{\max}] = E[D \,|\, \mathfrak{X}_0 = 0]$ and $E[D_{\min}] = E[D \,|\, \mathfrak{X}_0 = \mathscr{R}]$, we have

$$E[D \,|\, \mathfrak{X}_0 = 0] \sim N\left(d\log\left(\frac{\pi}{2}\right) - \frac{1}{a-1} + \mathscr{R} - \mathfrak{m}\right) e^{\mathfrak{m}-\mathscr{R}}, \tag{3.49a}$$

$$E[D \,|\, \mathfrak{X}_0 = \mathscr{R}] \sim N\left(d\log\left(\frac{\pi}{2}\right) - \frac{1}{a-1} + 2\mathscr{R} - \mathfrak{m}\right) e^{\mathfrak{m}-2\mathscr{R}}. \tag{3.49b}$$

The scaling of $E[D \,|\, \mathfrak{X}_0 = 0] \sim N$ and $E[D \,|\, \mathfrak{X}_0 = \mathscr{R}] \sim 1$ are achieved when $\mathfrak{m} = \mathscr{R}$ and $\mathscr{R}e^{-\mathscr{R}} \sim \frac{1}{N}$. Analogous to the cold regime, we set $\mathscr{R}^{-1}e^{\mathscr{R}} = \frac{N}{\nu}$, obtaining

$$\mathfrak{m} = \mathscr{R} = -W_{-1}\left(-\frac{\nu}{N}\right), \tag{3.50}$$

where $W_{-1}(x)$ is the W_{-1} branch of the Lambert W function.

Invoking the scaling in (3.50), we obtain

$$E[D] \approx \frac{2^d}{dI_{d,1}}\left(\frac{a}{a-1}\right)^2\left[1 - \left(d\log\left(\frac{\pi}{2}\right) - \frac{2}{a-1}\right)\left(W_{-1}\left(-\frac{\nu}{N}\right)\right)^{-1}\right], \tag{3.51a}$$

$$E[D \,|\, \mathfrak{X}_0 = \mathfrak{r}] \approx \frac{N2^d}{dI_{d,1}}\left(\frac{a}{a-1}\right)\left(d\log\left(\frac{\pi}{2}\right) - \frac{1}{a-1} + \mathfrak{r}\right) e^{-\mathfrak{r}}. \tag{3.51b}$$

Hence, the critical regime corresponds to sparse graphs in the thermodynamic limit, as confirmed by simulations in Figure 3.7. Note that the convergence to the $N \to \infty$ asymptote of $E[D] = 10$ is slower than in the cold regime, likely due to relatively large subleading terms in (3.51a).

We observe that, in the critical regime, the degree distribution of the RHG seems to follow a power-law with the same exponent as in the cold regime:

$$\begin{aligned} \Pr[D = k] &\sim k^{-\gamma}, \\ \gamma &= a + 1, \end{aligned} \tag{3.52}$$

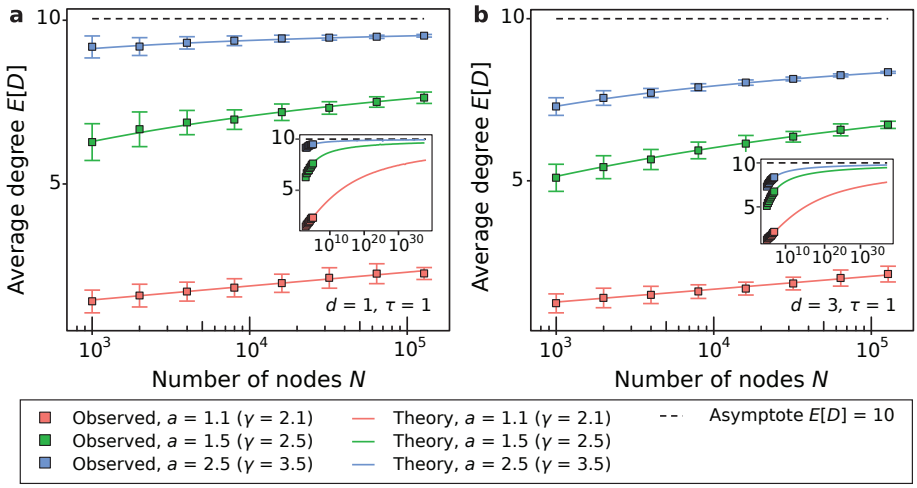


Figure 3.7: The average degree $E[D]$ as a function of the number of nodes N for RHGs in the critical regime ($\tau = 1$) with (a) $d = 1$ and (b) $d = 3$. Each panel includes the results for (red) $a = 1.1$ ($\gamma = 2.1$), (green) $a = 1.5$ ($\gamma = 2.5$), and (blue) $a = 2.5$ ($\gamma = 3.5$). Each point is the average of 100 simulations and the error bars display standard deviations. The solid lines are the theoretical values for $E[D]$ prescribed by (3.18) and (3.21) and the dashed line is the thermodynamic limit of (3.51a). The insets in (a) and (b) correspond to extended domains of N values. Note that in the critical regime, all 3 cases converge to the asymptotic value at a much slower rate than in the cold regime, while the $a = 1.1$ case has not yet converged even within the extended domain of N values.

as seen from Figure 3.8. This is the case since the tail of $\Pr[D = k]$ is dominated by nodes with small $\mathfrak{r}$ values. In the critical regime, $E[D \mid \mathfrak{X}_0 = \mathfrak{r}] \sim e^{-\mathfrak{r}}$ for $\mathfrak{r}$ values close to 0, similar to the cold regime, resulting in the same degree distribution exponent γ.

To investigate the $a = 1$ ($\gamma = 2$) case of the critical regime, we need to re-examine the scaling of $E[D_{\max}]$ and $E[D_{\min}]$. To this end, we plug $\tau = 1$ and $a = 1$ into (3.18), arriving at the leading-order approximations:

$$E[D \mid \mathfrak{X}_0 = \mathscr{R}] \approx \frac{2^d \left(\frac{\pi}{2}\right)^d}{I_{d,1}} N, \tag{3.53a}$$

$$E[D \mid \mathfrak{X}_0 = 0] \approx \frac{3 \times 2^{d-1} \left(\frac{\pi}{2}\right)^d}{I_{d,1}} N \mathscr{R}^2 e^{\mathfrak{m}-2\mathscr{R}}. \tag{3.53b}$$

It is seen from (3.53) that the desired scalings of $E[D_{\max}] = E[D \mid \mathfrak{X}_0 = 0] \sim N$ and $E[D_{\min}] = E[D \mid \mathfrak{X}_0 = \mathscr{R}] \sim 1$ are obtained if we set $\mathscr{R} = \ln(N/\nu)$, and $\mathfrak{m} = \mathscr{R} - 2\ln\mathscr{R}$. Then,

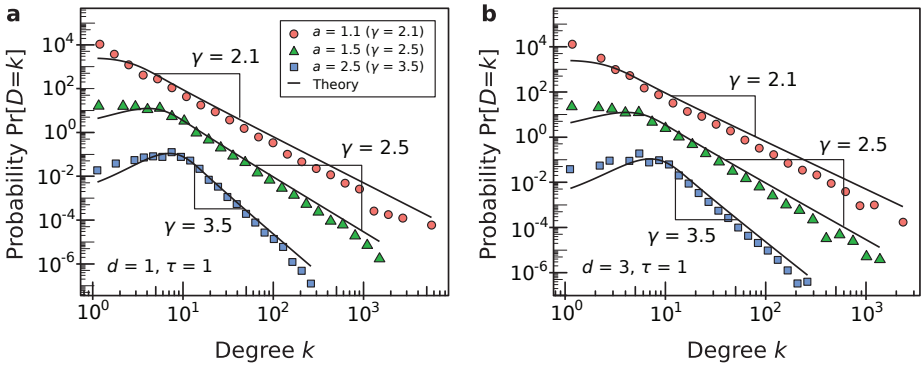


Figure 3.8: The degree distribution densities for RHGs with (a) $d = 1$ and (b) $d = 3$ at $\tau = 1$ and $N = 1000 \cdot 2^7$. Each panel includes the degree distributions for (red) $a = 1.1$ ($\gamma = 2.1$), (green) $a = 1.5$ ($\gamma = 2.5$), and (blue) $a = 2.5$ ($\gamma = 3.5$). The probabilities $\Pr[D = k]$ are obtained from a single network realization. The degree distributions are binned logarithmically to suppress noise at large k values. For visibility, the probabilities corresponding to $\gamma = 2.5$ and $\gamma = 2.1$ are multiplied by 10^2 and 10^4, respectively. The solid lines are the theoretical values for $\Pr[D = k]$ prescribed by (3.38). The scaling constant ν is chosen such that $\nu = 10 \times \frac{dI_{d,1}}{2^d}\left(\frac{a-1}{a}\right)^2$, corresponding to $E[D] = 10$ in the thermodynamic limit.

$$E[D \mid \mathfrak{X}_0 = \mathfrak{r}] \approx \frac{2^d}{dI_{d,1}} \frac{N}{[\ln(N/\nu)]^2} e^{-\mathfrak{r}} \times \left[\mathrm{Li}_2\left[-\left[\ln\left(\frac{N}{\nu}\right)\right]^2 \left(\frac{\pi}{2}\right)^d e^{\mathfrak{r}-\mathscr{R}}\right] - \mathrm{Li}_2\left[-\left[\ln\left(\frac{N}{\nu}\right)\right]^2 \left(\frac{\pi}{2}\right)^d e^{\mathfrak{r}}\right]\right], \tag{3.54a}$$

$$E[D] \approx \frac{\nu 2^d}{dI_{d,1}} \frac{1}{[\ln(N/\nu)]^2} \left[2\mathrm{Li}_3\left[-\left[\ln\left(\frac{N}{\nu}\right)\right]^2 \left(\frac{\pi}{2}\right)^d\right] - \mathrm{Li}_3\left[-\left[\ln\left(\frac{N}{\nu}\right)\right]^2 \left(\frac{\pi}{2}\right)^d e^{-\mathscr{R}}\right] - \mathrm{Li}_3\left[-\left[\ln\left(\frac{N}{\nu}\right)\right]^2 \left(\frac{\pi}{2}\right)^d e^{\mathscr{R}}\right]\right], \tag{3.54b}$$

and

$$E[D] \sim \frac{\nu 2^d}{dI_{d,1}} \ln\left(\frac{N}{\nu}\right), \tag{3.55}$$

where $\mathrm{Li}_s(x)$ is the s-th order polylogarithm function. Like in the cold regime, the $a = 1$ case in the critical regime corresponds to graphs with $E[D] \sim \ln(N/\nu)$, as confirmed by Figure 3.9. We note that there is a disagreement of theoretical and simulated values for $d > 1$ in Figure 3.9(b), likely caused by the breakdown of the approximation of $[\sin(\theta_1/2)]^d$ in (3.18) and imperfect control of $\mathscr{R}$ and $\mathfrak{m}$ at $a = 1, \tau = 1$. The density of the degree distribution for $a = 1$ in the critical regime is shown in Figure 3.10.

Unlike the cold regime, Figure 3.11(a) shows that the graph clustering coefficient c_G decreases logarithmically with the number of nodes N in the critical regime, like in the $d = 1$ case [100]. Similar to the cold regime, we observe that the graph clustering coefficient decreases as the dimension d increases, Figure 3.11(b).

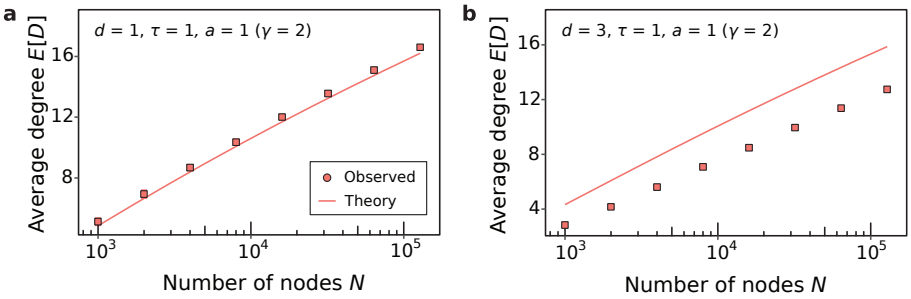


Figure 3.9: The average degree $E[D]$ as a function of the number of nodes N for RHGs with (a) $d = 1$ and (b) $d = 3$ for the case of $a = 1$ at $\tau = 1$. Each point is the average of 100 simulations and the error bars display standard deviations. To avoid fluctuations associated with large degree nodes, we have imposed a cutoff in the radial coordinate distribution, removing nodes with $r \leq r_{\text{cut}}$, where r_{cut} is defined like in the cold regime at $\tau = 0.5$, as (3.54a) cannot be solved exactly. The solid lines are the theoretical values for $E[D]$ prescribed by (3.54b), corrected for the radial coordinate cutoff. The scaling constant is set to $\nu = 10 \times \frac{dI_{d,1}}{2^d}$.

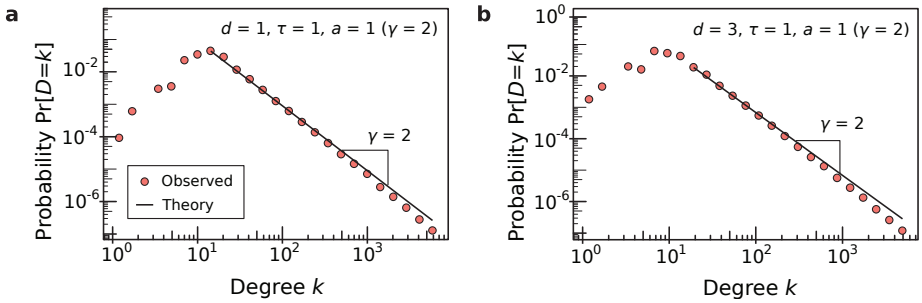


Figure 3.10: The densities of the degree distribution for RHGs with (a) $d = 1$ and (b) $d = 3$ in the case of $a = 1$ at $\tau = 1$ and $N = 1000 \cdot 2^7$. The probabilities $\Pr[D = k]$ are obtained from a single network realization. The degree distributions are binned logarithmically to suppress noise at large k values. To avoid fluctuations associated with large degree nodes, we have imposed a cutoff in the radial coordinate distribution, removing nodes with $r \leq r_{\text{cut}}$, where r_{cut} is defined like in the cold regime at $\tau = 0.5$, as (3.54a) cannot be solved exactly. The solid lines are the theoretical values for $\Pr[D = k]$ based on a slope of $\gamma = 2$. The scaling constant is set to $\nu = 10 \times \frac{dI_{d,1}}{2^d}$.

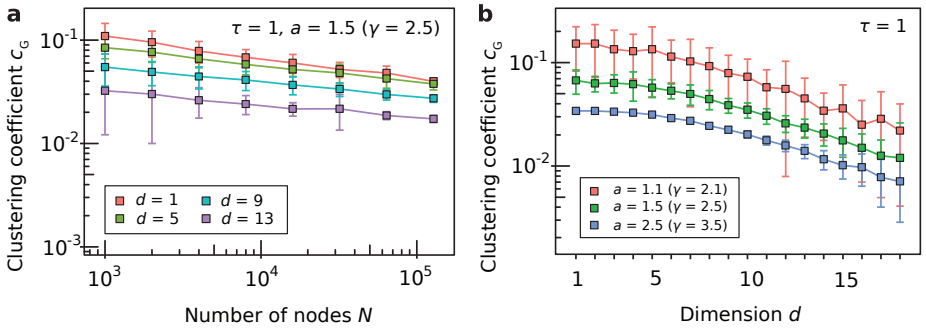


Figure 3.11: The graph clustering coefficient c_G of nodes with degree $k > 1$ (a) as a function of the number of nodes N and (b) as a function of dimension d at $\tau = 1$. Panel (a) includes results for (red) $d = 1$, (green) $d = 5$, (blue) $d = 9$ and (purple) $d = 13$, for $a = 1.5$ ($\gamma = 2.5$), while panel (b) includes results for (red) $a = 1.1$ ($\gamma = 2.1$), (green) $a = 1.5$ ($\gamma = 2.5$), and (blue) $a = 2.5$ ($\gamma = 3.5$), for $N = 10^4$. Each point is the average of 100 simulations and the error bars display standard deviations. The scaling constant ν is chosen such that $\nu = 10 \times \frac{dI_{d,1}}{2^d}\left(\frac{a-1}{a}\right)^2$, corresponding to $E[D] = 10$ in the thermodynamic limit.

GRAPH PROPERTY PERSPECTIVE, $\tau = 1$

To generate an RHG in the critical regime with desired graph properties for $\gamma > 2$, one must set

$$a = \gamma - 1, \tag{3.56a}$$

$$\mathfrak{m} = \mathscr{R} = -W_{-1}\left(-\frac{\nu}{N}\right), \tag{3.56b}$$

where ν is determined by (3.51a), which now takes the form of

$$E[D] = \frac{2^d}{dI_{d,1}}\left(\frac{\gamma-1}{\gamma-2}\right)^2\left[1-\left(d\log\left(\frac{\pi}{2}\right)-\frac{2}{\gamma-2}\right)\left(W_{-1}\left(-\frac{\nu}{N}\right)\right)^{-1}\right]. \tag{3.57}$$

When $\gamma = 2$ in the critical regime, one sets

$$\mathscr{R} = \ln(N/\nu), \tag{3.58a}$$

$$\mathfrak{m} = \mathscr{R} - 2\ln\mathscr{R}, \tag{3.58b}$$

while ν is obtained by solving (3.54b), now taking the form of

$$E[D] = \frac{\nu 2^d}{dI_{d,1}}\frac{1}{[\ln(N/\nu)]^2}\left[2\mathrm{Li}_3\left[-\left[\ln\left(\frac{N}{\nu}\right)\right]^2\left(\frac{\pi}{2}\right)^d\right]\right. \tag{3.59}$$
$$\left. -\mathrm{Li}_3\left[-\left[\ln\left(\frac{N}{\nu}\right)\right]^2\left(\frac{\pi}{2}\right)^d e^{-\mathscr{R}}\right]-\mathrm{Li}_3\left[-\left[\ln\left(\frac{N}{\nu}\right)\right]^2\left(\frac{\pi}{2}\right)^d e^{\mathscr{R}}\right]\right],$$

3.4.3. HOT REGIME, $\tau > 1$

In the $\tau > 1$ case, equations (3.18) and (3.21) can be approximated as

$$E[D] \approx (N-1)\,\mathfrak{I}(d,\tau)\,e^{\mathfrak{m}/\tau}\,E\left[e^{-\mathfrak{X}_0/\tau}\right]^2, \tag{3.60a}$$

$$E[D \mid \mathfrak{X}_0 = \mathfrak{r}] \approx \frac{E[D]}{E\left[e^{-\mathfrak{X}_0/\tau}\right]}\,e^{-\mathfrak{r}/\tau}, \tag{3.60b}$$

where

$$\mathfrak{I}(d,\tau) \equiv \frac{1}{I_{d,1}}\int_0^\pi \frac{\sin^{d-1}\theta_1 \mathrm{d}\theta_1}{\sin\left(\frac{\theta}{2}\right)^{d/\tau}}, \tag{3.61}$$

and

$$E\left[e^{-\mathfrak{X}_0/\tau}\right] \equiv \int_0^{\mathscr{R}} f_{\mathfrak{X}_0}(\mathfrak{r})e^{-\mathfrak{r}/\tau}\mathrm{d}\mathfrak{r} \approx \frac{a\tau}{a\tau-1}\left(e^{-\mathscr{R}/\tau} - e^{-a/\mathscr{R}}\right). \tag{3.62}$$

Note that the approximation for $E\left[e^{-\mathfrak{X}_0/\tau}\right]$ given by (3.62) is valid for all values of a and τ since $a\tau > 1$.

Similar to the $\tau < 1$ regime, we demand $E[D_{\max}] \sim N$ and $E[D_{\min}] \sim 1$ to obtain the scaling relationships for $\mathfrak{m}$ and $\mathscr{R}$:

$$\mathfrak{m} = \mathscr{R} = \tau \ln(N/\nu). \tag{3.63}$$

In combination with (3.60a), this scaling leads to

$$E[D] \approx \nu\,\mathfrak{I}(d,\tau)\left(\frac{a\tau}{a\tau-1}\right)^2, \tag{3.64a}$$

$$E[D \mid \mathfrak{X}_0 = \mathfrak{r}] \approx \frac{N}{\nu}\left(\frac{a\tau-1}{a\tau}\right)E[D]\,e^{-\mathfrak{r}/\tau}, \tag{3.64b}$$

$$\Pr[D=k] \approx a\tau\,(E[D \mid \mathfrak{X}_0 = \mathscr{R}])^{a\tau}\,\frac{\Gamma[k-a\tau]}{\Gamma[k+1]} \sim k^{-a\tau-1}, \tag{3.64c}$$

in the thermodynamic limit, as confirmed by Figure 3.12.

Similar to the cold and critical regimes, RHGs in the hot regime are sparse and have power-law degree distributions, $\Pr[D=k] \sim k^{-\gamma}$. Different from the cold and critical regimes, degree distribution exponent $\gamma = a\tau + 1$ in the hot regime depends on both a and τ, as confirmed by Figure 3.13.

In the hot regime, we observe that the graph clustering coefficient decays with N as $n^{-\sigma}$, where the value of the exponent σ depends on both d and τ, Figure 3.14. It is already known from [100] that the scaling of the clustering coefficient with N depends on τ in the $d = 1$ case. Here, we observe that it also depends on the dimension $d \geq 1$. Your sentence is mostly accurate but could be slightly rephrased for clarity and conciseness. Here is a corrected version:

As discussed in Section 3.4.1, the angular distance distribution between two random points on a d-dimensional sphere converges to the Dirac delta function centered at $\pi/2$ as $d \to \infty$ [101], causing the graph clustering coefficient to tend to 0. Figure 3.14(b) confirms that clustering in the hot regime decreases slowly with dimension d.

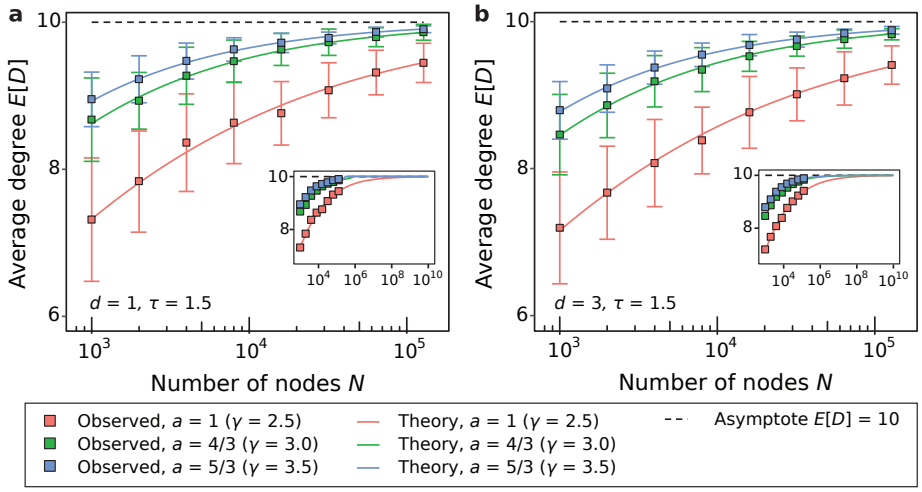


Figure 3.12: The average degree $E[D]$ as a function of the number of nodes N for RHGs in the hot regime ($\tau = 1.5$) with (a) $d = 1$ and (b) $d = 3$. Each panel includes the results for (red) $a = 1$ ($\gamma = 2.5$), (green) $a = \frac{4}{3}$ ($\gamma = 3.0$), and (blue) $a = \frac{5}{3}$ ($\gamma = 3.5$). Each point is the average of 100 simulations and the error bars display standard deviations. The solid lines are the theoretical values for $E[D]$ prescribed by (3.18) and (3.21) and the dashed line is the thermodynamic limit for $E[D]$ given by (3.64a). The insets in (a) and (b) correspond to an extended domain of N values.

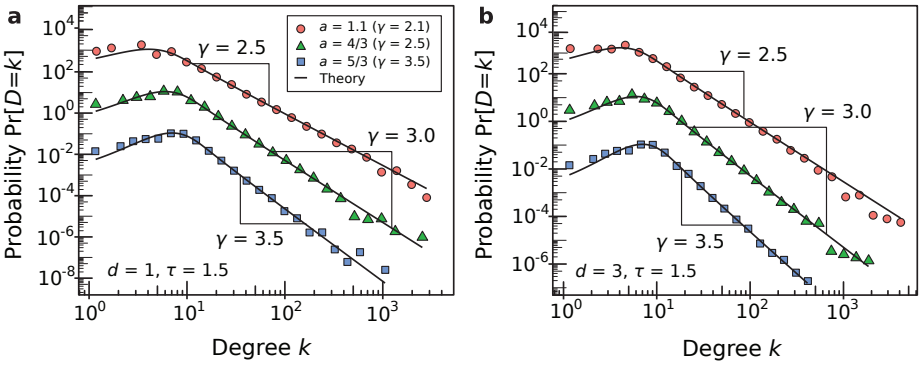


Figure 3.13: The degree distribution densities for RHGs with (a) $d = 1$ and (b) $d = 3$ at $\tau = 1.5$ and $N = 1000 \cdot 2^7$. Each panel includes the degree distributions for (red) $a = 1$ ($\gamma = 2.5$), (green) $a = \frac{4}{3}$ ($\gamma = 3.0$), and (blue) $a = \frac{5}{3}$ ($\gamma = 3.5$). The degree distributions are binned logarithmically to suppress noise at large k values. The solid lines are the theoretical values for $\Pr[D = k]$ prescribed by (3.64c). For visibility, the probabilities corresponding to $\gamma = 3.0$ and $\gamma = 2.5$ are multiplied by 10^2 and 10^4, respectively. The scaling constant ν is chosen such that $\nu = 10 \times \frac{1}{\mathfrak{I}(d,\tau)} \left(\frac{a\tau-1}{a\tau}\right)^2$, corresponding to $E[D] = 10$ in the thermodynamic limit.

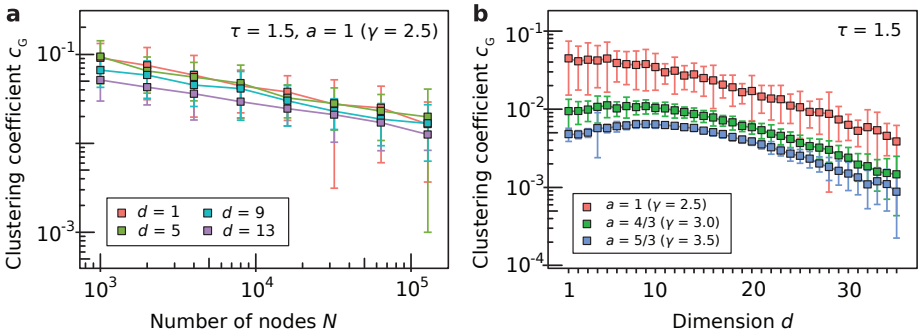


Figure 3.14: The graph clustering coefficient c_G (a) as a function of the number of nodes N and (b) as a function of dimension d at $\tau = 1.5$. Panel (a) includes results for (red) $d = 1$, (green) $d = 5$, (blue) $d = 9$ and (purple) $d = 13$, for $a = 1$ ($\gamma = 2.5$), while panel (b) includes results for (red) $a = 1$ ($\gamma = 2.5$), (green) $a = \frac{4}{3}$ ($\gamma = 3.0$), and (blue) $a = \frac{5}{3}$ ($\gamma = 3.5$), for $N = 10^4$. Each point is the average of 100 simulations and the error bars display standard deviations. The scaling constant ν is chosen such that $\nu = 10 \times \frac{1}{\mathfrak{I}(d,\tau)} \left(\frac{a\tau-1}{a\tau}\right)^2$, corresponding to $E[D] = 10$ in the thermodynamic limit.

GRAPH PROPERTY PERSPECTIVE, $\tau > 1$

To generate an RHG in the hot regime with a given average degree $E[D]$ and power-law exponent $\gamma > 2$, one needs to set

$$a = \frac{1}{\tau}(\gamma - 1), \tag{3.65a}$$

$$\mathfrak{m} = \mathscr{R} = \tau \ln(N/\nu), \tag{3.65b}$$

where ν is given by

$$E[D] = \nu \mathfrak{I}(d,\tau) \left(\frac{\gamma-1}{\gamma-2}\right)^2, \tag{3.66}$$

and $\mathfrak{I}(d,\tau)$ is given by (3.61). Because we require $a > 1$, in this regime the power-law exponent is bounded $\gamma - 1 > \tau$. Since $\tau > 1$, a power-law exponent $\gamma = 2$ is not possible in the hot regime.

In summary, we find that under the proper change of variables prescribed by (3.15d), RHGs of any dimensionality can be naturally described by three regimes based on the rescaled temperature τ, the cold regime ($0 \leq \tau < 1$), the critical regime ($\tau = 1$), and the hot regime ($\tau > 1$). In each of these regimes, RHGs of any dimensionality $d \geq 1$ exhibit similar topological properties with respect to the node degrees, Figure 3.16. Our approximations of $E[D|\mathfrak{X}_0 = \mathfrak{r}]$ work well in each of the three regimes, see Figure B.2, but there is a small constant bias in the hot regime. The approximations break down for nodes close to the center of $\mathbb{B}^{d+1}$ (values of $\mathfrak{r}$ close to 0) across all three regimes. Consistent with the $d = 1$ case [21], the graph clustering coefficient in RHGs seems to approach a constant in the thermodynamic limit when $\tau < 1$, decreases polynomially as the RHG size increases when $\tau > 1$, and decreases logarithmically as a function of network size in the critical regime when $\tau = 1$. We find that the graph clustering coefficient in the RHG model decreases monotonically as a function of dimension d in each of the cold, critical,

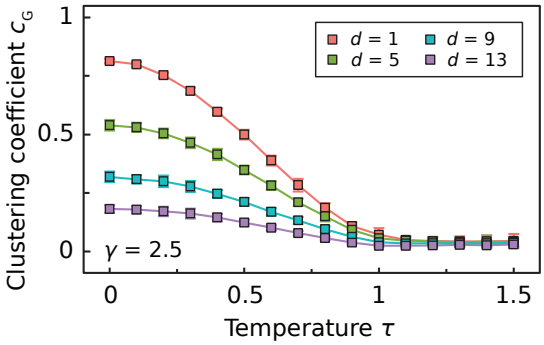


Figure 3.15: The graph clustering coefficient c_G of nodes with degree $k > 1$ as a function of temperature τ for different dimensions d. Each point is the average of 100 simulations and the error bars display standard deviations (not visible for most points). The scaling constant ν is chosen such that $E[D] = 10$ in the thermodynamic limit, and the number of nodes $N = 10^4$.

and hot regimes, consistent with findings for the GIRG model. Finally, Figure 3.15 shows that for any dimension, the graph clustering coefficient is a decreasing function of τ.

3.5. LIMITING CASES OF THE RHG MODEL

In this section, we analyze several important parameter limits of the RHG and show that they correspond to well-known graph ensembles.

3.5.1. THE $\tau \to 0$ LIMIT IN THE COLD REGIME

The case of $\tau = 0$ is well-defined as the $\tau \to 0$ limit of the cold regime. The $T \to 0$ limit of the connection probability function in (3.11) is the step function

$$p_{ij} = \Theta\left(\mu - d_{ij}\right), \tag{3.67}$$

such that connections are established deterministically between node pairs separated by distances smaller than μ. In this case, we have $\frac{\pi\tau}{\sin(\pi\tau)} \to 1$ in (3.36), leading to

$$E[D] \approx \frac{\nu 2^d}{d I_{d,1}} \left(\frac{a}{a-1}\right)^2 \left[1 - 2\left(\frac{N}{\nu}\right)^{1-a} + \left(\frac{N}{\nu}\right)^{2(1-a)}\right], \tag{3.68a}$$

$$E[D \,|\, \mathfrak{X}_0 = \mathfrak{r}] \approx \frac{N}{\nu}\left(\frac{a-1}{a}\right) E[D]\, e^{-\mathfrak{r}} \sum_{\ell=0}^{\infty} \left(\frac{\nu}{N}\right)^{\ell(a-1)}. \tag{3.68b}$$

The resulting graphs are sparse and are characterized by a power-law degree distribution $\Pr[D = k] \sim k^{-\gamma}$, $\gamma = a + 1$, similar to the $0 < \tau < 1$ case.

<table>
<tr><th>τ \ a</th><th>$a = 1$</th><th>$a \in (1, \infty)$</th><th>$a \to \infty$</th></tr>
<tr><td>$\tau \to 0$</td><td rowspan="2">$\mathscr{R} = \ln(N/\nu), \mathfrak{m} = \mathscr{R} - \ln \mathscr{R}$
$E[D] \sim O(\ln(N/\nu))$
$\gamma = 2$
Section 3.4.1</td><td rowspan="2">$\mathscr{R} = \ln(N/\nu), \mathfrak{m} = \mathscr{R}$
$E[D] \sim O(1)$
$\gamma = a + 1$
Section 3.4.1</td><td>Spherical Random Geometric Graphs (SpRGG)
$p_{ij} = \Theta(\theta_c - \Delta\theta)$
Section 3.5.3</td></tr>
<tr><td>$\tau \in (0,1)$</td><td rowspan="3">Spherical Soft Random Geometric Graphs (SpSoRGG)
$p_{ij} = f(\Delta\theta)$
Section 3.5.2</td></tr>
<tr><td>$\tau = 1$</td><td>$\mathscr{R} = \ln(N/\nu), \mathfrak{m} = \mathscr{R} - 2\ln \mathscr{R}$
$E[D] \sim O(\ln(N/\nu))$
$\gamma = 2$
Section 3.4.2</td><td>$\mathscr{R} = -W_{-1}(-\nu/N), \mathfrak{m} = \mathscr{R}$
$E[D] \sim O(1)$
$\gamma = a + 1$
Section 3.4.2</td></tr>
<tr><td>$\tau \in (1, \infty)$</td><td colspan="2">$\mathscr{R} = \tau \ln(N/\nu), \mathfrak{m} = \mathscr{R}$
$E[D] \sim O(1)$
$\gamma = a\tau + 1$
Section 3.4.3</td></tr>
<tr><td>$\tau \to \infty, \lim_{\tau\to\infty} \zeta/\tau = \lambda$</td><td colspan="3">Hyper Soft Configuration Model (HSCM), $\gamma = \frac{2\alpha}{d\lambda} + 1$
Section 3.5.4</td></tr>
<tr><td>$\tau \to \infty, \lim_{\tau\to\infty} \mathfrak{m}/\tau = \lambda$</td><td colspan="3">Erdős-Rényi (ER) graph, $p = \frac{1}{1+e^{-\lambda}}$, $\lambda \equiv \lim_{\tau\to\infty} \mu(\tau)/\tau$
Section 3.5.5</td></tr>
</table>

Figure 3.16: Graph properties in terms of the rescaled model parameters in the different RHG regimes.

3.5.2. THE $a \to \infty$ LIMIT: SPHERICAL SOFT RANDOM GEOMETRIC GRAPHS

In this limit, the density of the radial coordinate distribution (3.16) degenerates to

$$f_{\mathfrak{X}_0}(\mathfrak{r}) \to \delta(\mathfrak{r} - \mathscr{R}). \tag{3.69}$$

As a result, all nodes are placed at the boundary of the hyperbolic ball $\mathbb{B}^{d+1}$ with $\mathfrak{r}_i = \mathscr{R}$. Even though the distances between nodes are still hyperbolic, they are fully determined by the angles on $\mathbb{S}^d$:

$$\zeta d_{ij} = \cosh^{-1}\left[\cosh\left(\frac{2\mathscr{R}}{d}\right)^2 - \sinh\left(\frac{2\mathscr{R}}{d}\right)^2 \sin\left(\Delta\theta_{ij}\right)\right].$$

Hence, the connection probability p_{ij} is fully determined by the angles $\Delta\theta_{ij}$:

$$p_{ij} = \frac{1}{1+\exp\left(\frac{\mathfrak{d}(\Delta\theta_{ij})-\mathfrak{m}}{\tau}\right)}, \tag{3.70}$$

where $\mathfrak{d}\left(\Delta\theta_{ij}\right) \equiv \frac{d\zeta}{2} d_{ij}$. In the $a \to \infty$ regime, all nodes are effectively placed at the surface of the unit sphere $\mathbb{S}^d$, and connections are made with distance-dependent probabilities on the sphere. Hence, RHGs are soft RGGs on $\mathbb{S}^d$ in the $a \to \infty$ limit.

3.5.3. THE $a \to \infty$, $\tau \to 0$ LIMIT: SPHERICAL RANDOM GEOMETRIC GRAPHS

If $a \to \infty$ and $\tau \to 0$, the connection probabilities in (3.17) become

$$p_{ij} = \Theta(\theta_c - \Delta\theta_{ij}), \tag{3.71}$$

where θ_c is the solution to the equation $\mathfrak{d}(\theta_c) = \mathfrak{m}$. Thus, in this limit, the RHG becomes the sharp random geometric graph on $\mathbb{S}^d$ (SpRGG).

The expected degree of the SpRGG equals the expected number of nodes that fall within an angle θ_c of the $\theta_1 = 0, \ldots, \theta_d = 0$ point,

$$E[D] = (N-1)\mathfrak{p}, \tag{3.72}$$

where the volume of the $(d-1)$-dimensional sphere of radius θ_c in $\mathbb{S}^d$ is

$$\mathfrak{p} = \frac{\int_0^{\theta_c} [\sin(\theta)]^{d-1} \mathrm{d}\theta}{\int_0^{\pi} [\sin(\theta)]^{d-1} \mathrm{d}\theta}. \tag{3.73}$$

The degree distribution is thus binomial,

$$\Pr[D = k] = \mathrm{Bin}[N-1, \mathfrak{p}](k), \tag{3.74}$$

converging to the Poisson distribution with mean $E[D]$ if θ_c is such that $N\mathfrak{p} \to E[D]$. Since the Poisson distribution is the $\gamma \to \infty$ limit of the Pareto-mixed Poisson distribution (3.38), we refer to this regime as the $\gamma \to \infty$ case in Figure 3.16.

3.5.4. THE $\zeta \to \infty$, $\tau \to \infty$ LIMIT: THE HYPERSOFT CONFIGURATION MODEL

In the $\zeta \to \infty$ limit, the hyperbolic distances in (3.6) degenerate to

$$d_{ij} = r_i + r_j, \tag{3.75}$$

so that the angular coordinates of nodes are ignored in this limit. Further, if τ also tends to infinity, $\tau \to \infty$, but such that $\lim_{\zeta\to\infty} \frac{\zeta}{\tau} = \lambda > 0$, where λ is a constant, then the connection probability in (3.11) simplifies to

$$p_{ij} = \frac{1}{1 + e^{\omega_i} e^{\omega_j}}, \tag{3.76}$$

which is the connection probability in the hypersoft configuration model (HSCM) [102]. Here, $\omega_i = \frac{d\lambda}{2}\left(r_i - \frac{\mu}{2}\right)$ are the Lagrange multipliers controlling the expected node degrees. The Lagrange multipliers are the random variables $\mathfrak{Z} = \frac{d\lambda}{2}\left(X_0 - \frac{\mu}{2}\right)$, with the probability density function

$$f_{\mathfrak{Z}}(\omega) \approx \frac{2\alpha}{d\lambda} e^{\alpha\left(\frac{\mu}{2} - R\right)} e^{\frac{2\alpha}{d\lambda}\omega}, \quad \omega \in \left(-\frac{d\lambda\mu}{4}, \frac{d\lambda}{2}\left(R - \frac{\mu}{2}\right)\right). \tag{3.77}$$

The expected degrees in the HSCM are approximated by

$$E[D \,|\, \mathfrak{Z} = \omega_i] \approx (N-1) E\left[e^{-\mathfrak{Z}}\right] e^{-\omega_i}, \tag{3.78a}$$

$$E[D] \approx (N-1)\left(E\left[e^{-\mathfrak{Z}}\right]\right)^2, \tag{3.78b}$$

where $E\left[e^{-\mathfrak{Z}}\right] \equiv \int \mathrm{d}\omega f_{\mathfrak{Z}}(\omega) e^{-\omega}$. By demanding that $E\left[D \,\middle|\, \mathfrak{Z} = -\frac{d\lambda\mu}{4}\right] \sim N$, corresponding to $X_0 = 0$, and $E\left[D \,\middle|\, \mathfrak{Z} = \frac{d\lambda}{2}\left(R - \frac{\mu}{2}\right)\right] \sim 1$, corresponding to $X_0 = R$, we obtain $R = \frac{2}{d\lambda}\ln(N)$, while $\mu = R$ in the case of $\frac{2\alpha}{d\lambda} > 1$ and $\mu = \frac{2\alpha}{d\lambda} R$ in the case of $\frac{2\alpha}{d\lambda} < 1$.

In both cases, the expected node degree $E[D \,|\, \mathfrak{Z} = \omega_i] \sim e^{-\omega_i}$, and graphs produced by the HSCM are sparse, while the conditional probability $\Pr[D = k \,|\, \mathfrak{Z} = \omega]$ is well-approximated by the Poisson distribution:

$$\Pr[D = k \,|\, \mathfrak{Z} = \omega] \approx \frac{1}{k!} e^{-E[D \,|\, \mathfrak{Z}=\omega]} \left(E[D \,|\, \mathfrak{Z} = \omega]\right)^k, \tag{3.79}$$

see [102]. The resulting degree distribution is a mixed Poisson distribution with density

$$\Pr[D = k] \approx \frac{1}{k!} \int_{-\frac{d\lambda\mu}{4}}^{\frac{d\lambda}{2}\left(R - \frac{\mu}{2}\right)} e^{-E[D \,|\, \mathfrak{Z}=\omega]} \left(E[D \,|\, \mathfrak{Z} = \omega]\right)^k f_{\mathfrak{Z}}(\omega) \mathrm{d}\omega, \tag{3.80}$$

where $E[D \,|\, \mathfrak{Z} = \omega]$ is the mixing parameter. Using (3.23) and (3.24), we obtain

$$\Pr[D = k] \approx (\gamma - 1) \kappa_0^{\gamma-1} \frac{\Gamma[k + 1 - \gamma, e^{\frac{d\lambda}{2}\left(R - \frac{\mu}{2}\right)} \kappa_0]}{\Gamma[k+1]} \sim k^{-\gamma}, \tag{3.81}$$

where $\gamma = \frac{2\alpha}{d\lambda} + 1$ and $\kappa_0 \equiv e^{\frac{d\lambda}{2}\left(\frac{\mu}{2} - R\right)} \frac{E[D]}{E[e^{-\mathfrak{Z}}]}$.

Thus, in the $\zeta \to \infty$, $\tau \to \infty$, $\zeta/\tau \to \lambda$ limit the RHG model degenerates to the HSCM with a scale-free degree distribution with exponent $\gamma = \frac{2\alpha}{d\lambda} + 1$.

3.5.5. THE $\tau \to \infty$ LIMIT: THE ERDŐS-RÉNYI GRAPH

The limit of $\tau \to \infty$ and finite ζ is the most degenerate case. Indeed, in this regime connection probabilities p_{ij} become independent of the hyperbolic distances d_{ij} between the nodes:

$$p_{ij} = \lim_{\tau\to\infty} \frac{1}{1+e^{-\mu(\tau)/\tau}}. \tag{3.82}$$

It is seen from (3.82) that the connection probabilities are non-trivial only in the case $\mu(\tau) \sim \tau$. In this case, the connection probabilities are constant:

$$p_{ij} = p = \lim_{\tau\to\infty} \frac{1}{1+e^{-\lambda}}, \tag{3.83}$$

where $\lambda \equiv \lim_{\tau\to\infty} \frac{\mu(\tau)}{\tau}$. By varying $\lambda \in (-\infty,\infty)$ one can tune the connection probabilities $p \in (0,1)$ of the resulting ER graphs.

One can also check that the ER limit can be obtained either as the $\gamma \to \infty$ ($\alpha \to \infty$ or $\lambda \to 0$) limit of the HSCM or as the $\tau \to \infty$ limit of the SpSoRGG.

3.5.6. THE $d \to \infty$ LIMIT

As the dimension d increases, the angular distance distribution between two random points on a d-sphere approaches the Dirac delta function centered at $\pi/2$ [101]. As a result, the role of a node's angular coordinates in the hyperbolic distances diminishes, and the network becomes more similar to the HSCM. The distances between nodes now depend only on their radial coordinates,

$$d_{ij} \approx \mathfrak{r}_i + \mathfrak{r}_j - \frac{1}{\zeta}\ln 2. \tag{3.84}$$

In the $d \gg 1$ regime, we estimate the expected degree of a node with radial coordinate $\mathfrak{r}_\ell$ using this approximation as

$$E[D\,|\,\mathfrak{X}_0 = \mathfrak{r}] \approx (N-1)\,e^{-a\mathscr{R}} \int_1^{e^{a\mathscr{R}}} \frac{\mathrm{d}\xi}{1+\left(\frac{1}{2}\right)^{\frac{d}{2\tau}} e^{\frac{\mathfrak{r}-\mathfrak{m}}{\tau}} \xi^{\frac{1}{a\tau}}}. \tag{3.85}$$

For any positive T, we have $\tau \to \infty$ as $d \to \infty$ and, hence, we are in the hot regime, $\tau > 1$. In this regime, we set $\mathfrak{m} = \mathscr{R} = \tau \ln(N/\nu)$, resulting in

$$E[D\,|\,\mathfrak{X}_0 = \mathfrak{r}] \approx N 2^{\frac{d}{2\tau}} e^{-\frac{\mathfrak{r}}{\tau}}, \tag{3.86a}$$

$$E[D] \approx \nu 2^{\frac{d}{2\tau}}, \tag{3.86b}$$

$$\Pr[D=k] \sim k^{-\gamma}, \quad \gamma = a\tau + 1. \tag{3.86c}$$

Thus, high-dimensional RHGs are akin to low-dimensional RHGs in the hot regime. As seen from (3.17), high-dimensional RHGs are sparse graphs with power-law degree distributions and a temperature-dependent exponent γ. High-dimensional RHGs are well-defined in the $d \to \infty$ limit. In this case, RHGs are sparse graphs with $E[D] \approx \nu 2^{\frac{1}{2\tau}}$. In the $d \to \infty$ limit, $\gamma \to \infty$ implies that RHGs are no longer described by a power-law

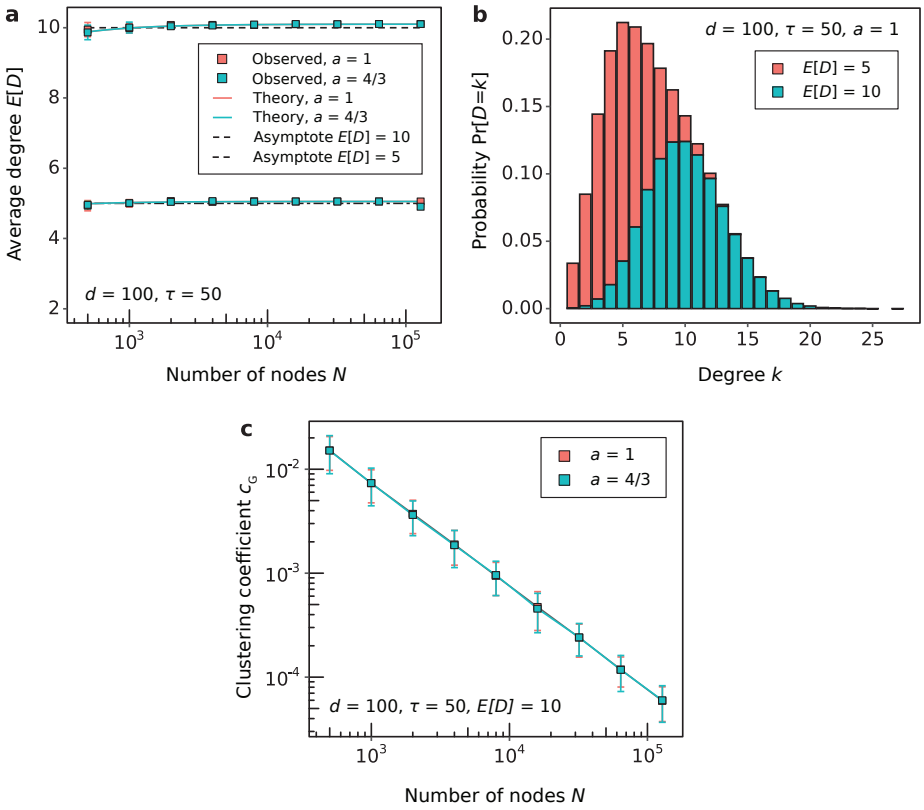


Figure 3.17: Average degree $E[D]$ (a) as a function of network size N, density $\Pr[D=k]$ of the degree distribution (b), and graph clustering coefficient c_G (c) of nodes with degree $k>1$ as a function of network size N for RHGs with $d=100$ at $\tau=50$. Panels (a) and (c) include results for (red) $a=1$ and (blue) $a=4/3$, panel (b) includes results for (red) $E[D]=5$ and (blue) $E[D]=10$ at $a=1$ and $N=1000\cdot 2^7$, while $E[D]=10$ in panel (c). Each point in panels (a) and (c) is the average of 20 simulations and the error bars display standard deviations, while probabilities $\Pr[D=k]$ in panel (b) are obtained from a single network realization for each different value of $E[D]$. The scaling constant ν is chosen such that $\nu = E[D]\,2^{-\frac{1}{2T}}$.

degree distribution. This is indeed the case, as all nodes are located at the boundary of $\mathbb{B}^{d+1}$, $\mathfrak{r}=\mathscr{R}$, leading to a Poisson degree distribution

$$\Pr[D=k]=\Pr[D\,|\,\mathfrak{X}_0=\mathscr{R}]\approx\frac{1}{k!}e^{-E[D]}\big(E[D]\big)^k, \tag{3.87}$$

where $E[D]=E[D\,|\,\mathfrak{X}_0=\mathscr{R}]\approx\nu 2^{\frac{1}{2T}}$.

GRAPH PROPERTY PERSPECTIVE, LIMITING CASES

Figure 3.18 summarizes the properties of the RHG and its limiting cases in the (τ,γ) phase space. Within the (τ,γ) phase space, all the RHG temperature regimes condense into the heterogeneous ($2\leq\gamma<\infty$) soft-geometric ($0\leq\tau<\infty$) state. The sharp-geometric limit ($\tau\to 0$) of this state is well-defined and is obtained by taking the $\tau\to 0$ limit of (3.43).

In this case, to generate RHGs with desired average degree $E[D]$ and power-law distribution exponent $\gamma > 2$, one needs to set $\mathfrak{m} = \mathscr{R} = \ln(N/\nu)$, where ν is given by

$$E[D] = \frac{\nu 2^d}{d I_{d,1}} \left(\frac{\gamma-1}{\gamma-2}\right)^2 \left[1 - 2\left(\frac{N}{\nu}\right)^{2-\gamma} + \left(\frac{N}{\nu}\right)^{2(2-\gamma)}\right]. \tag{3.88}$$

By setting $\gamma \to \infty$ ($a \to \infty$) in the RHG, one arrives at Spherical Soft Random Geometric Graphs (SpSoRGG). Here nodes are placed at the boundary of the $\mathbb{B}^{d+1}$ ball, and connections are established with probabilities dependent on distances between the nodes on its $\mathbb{S}^d$ boundary, see Section 3.5.2. Since SpSoRGGs are characterized by the Poisson degree distribution, we refer to them as the homogeneous ($\gamma \to \infty$) soft-geometric limit of the RHG. The expected degree of the SpSoRGG can be obtained by taking the $\gamma \to \infty$ limit of the RHG in the cold, critical, or hot regimes, depending on the τ value. In other words, to generate an SpSoRGG with prescribed τ and $E[D]$, one needs to set $\mathfrak{m}$, $\mathscr{R}$, and ν as follows

$$0 < \tau < 1 : \mathfrak{m} = \mathscr{R} = \ln(N/\nu),\ E[D] = \frac{\nu 2^d}{d I_{d,1}} \frac{\pi\tau}{\sin(\pi\tau)}; \tag{3.89a}$$

$$\tau = 1 : \mathfrak{m} = \mathscr{R} = -W_{-1}(\nu/N),\ E[D] = \frac{\nu 2^d}{d I_{d,1}}; \tag{3.89b}$$

$$\tau > 1 : \mathfrak{m} = \mathscr{R} = \tau \ln(N/\nu),\ E[D] = \nu \mathfrak{I}(d,\tau). \tag{3.89c}$$

By taking the $\tau \to 0$ limit of the Spherical Soft RGG, we arrive at the Spherical Sharp RGG, or simply Spherical Random Geometric Graph (SpRGG). Similar to its soft counterpart, nodes in the SpRGG are placed at the $\mathbb{B}^{d+1}$ boundary but connections are established deterministically between nodes separated by distances smaller than the threshold, Section 3.5.2. Another possibility to arrive at the SpRGG is by taking the $\gamma \to \infty$ limit of the Sharp RHG. One can generate Spherical Sharp RGGs with desired average degree $E[D]$ by setting $\mathfrak{m} = \mathscr{R} = \ln(N/\nu)$, and selecting ν from

$$E[D] = \frac{\nu 2^d}{d I_{d,1}}. \tag{3.90}$$

While both the hypersoft configuration model (HSCM) and the Erdős-Rényi (ER) model are the $\tau \to \infty$ limits of the RHG, they belong to two distinct classes, as seen from the graph property perspective.

The HSCM belongs to the non-geometric ($\tau \to \infty$) heterogeneous ($2 \le \gamma < \infty$) case and is a $\tau \to \infty, \zeta \to \infty$ limit of the RHG. To build RHGs with desired average degree $E[D]$ and a power-law degree distribution exponent $\gamma > 2$, one sets $\mu = R = \frac{2}{d\lambda} \ln\left(\frac{N}{\nu}\right)$, where ν is the solution of

$$E[D] = \nu \left(\frac{\gamma-1}{\gamma-2}\right)^2 \left[1 - \left(\frac{\nu}{n}\right)^{\gamma-2}\right]^2. \tag{3.91}$$

The ER model, on the other hand, belongs to the non-geometric ($\tau \to \infty$) homogeneous ($\gamma \to \infty$) state and is a $\gamma \to \infty$ limit of the HSCM. Alternatively, the ER model can also be attained as the $\tau \to \infty, \zeta \to \infty$ limit of the SpSoRGG.

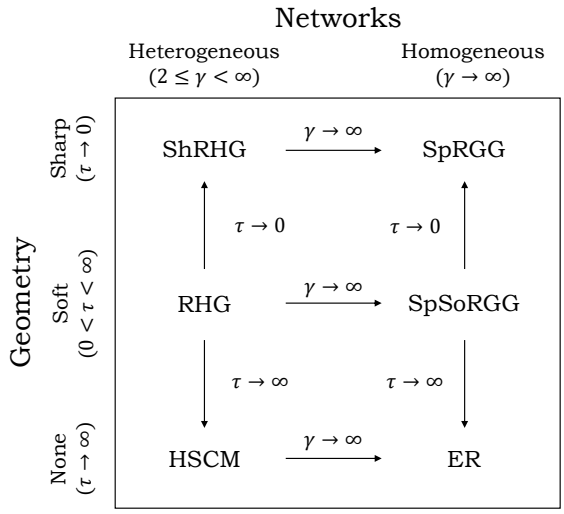


Figure 3.18: Limiting regimes of the RHG, graph property perspective.

In the $d \to \infty$ limit, RHGs have a Poisson degree distribution, and one can generate RHGs with a desired expected degree by selecting ν from

$$E[D] = \nu 2^{\frac{1}{2T}}. \tag{3.92}$$

3.6. HYPERBOLIC GRAPH GENERATOR IN $d+1$ DIMENSIONS

We conclude by presenting a software package that generates RHGs of arbitrary dimensionality, to be specified by the user. The generator covers the cold ($\tau < 1$), critical ($\tau = 1$) and hot ($\tau > 1$) regimes. The software package and detailed instructions on how to compile and use it are available from the Bitbucket repository [103].

The RHG generator can operate in two different modes: hybrid and model-based. In hybrid mode, the user provides the average degree $E[D]$, power-law exponent γ, rescaled temperature τ and dimension d. The equations (3.18) and (3.21) are solved for the rescaled radius $\mathscr{R}$ that yields the desired $E[D]$ using the bisection method. The triple integral that is found by combining (3.18) and (3.21) is evaluated numerically using Monte Carlo integration with importance sampling with the GNU Scientific Library (GSL) [104]. In model-based mode, the user directly provides the model parameters a, τ, $\mathscr{R}$ (or ν), and d. We expect the model-based mode to be instrumental for research purposes.

3.7. SUMMARY

In this Chapter, we have generalized random hyperbolic graphs (RHGs) to an arbitrary number of dimensions. In doing so, we have found a rescaling of the model parameters given by (3.15) that allows us to reduce RHGs of arbitrary dimensionality to a single

mathematical framework. Summarized in Figure 3.16, our results indicate that RHGs exhibit similar connectivity properties, irrespective of the dimension d of the latent space. At the same time, higher dimensional realizations of the RHG model differ from the original $d = 1$ RHG model for other topological properties.

One such property is the clustering coefficient. We find that the degree-dependent and graph clustering coefficients behave differently depending on the temperature regime. In the cold regime, $0 \leq \tau < 1$, RHGs are characterized by non-vanishing graph and degree-dependent clustering coefficients. In the hot regime, $\tau > 1$, the graph clustering coefficient becomes size-dependent and vanishes in the thermodynamic limit. These observations are expected and have been previously studied in the special $d = 1$ case [21, 26, 98]. The critical temperature of $\tau = 1$ corresponds to a continuous phase transition, which has been shown in $d = 1$ to be topological, characterized by diverging entropy and atypical finite-size scaling behavior of the clustering coefficient [100]. RHGs of arbitrary dimensionality allow us to study the behavior of the clustering coefficient as a function of the dimension d. To this end, we observe that in all three regimes, the clustering coefficient decreases as a function of d. This observation is consistent with the $d \to \infty$ limit, which is akin to the hypersoft configuration model. In general, we observe that the degree-dependent clustering coefficient depends on both the dimension d and the temperature τ, as seen in our numerical experiments. This observation is in line with another work proposing to use the density of cycles to estimate the network dimensionality [91]. Yet it remains an open question what exactly is different between two RHGs of different dimensionalities where the clustering coefficient is matched by selecting the appropriate temperature values.

Higher-dimensional RHGs may be instrumental in graph embedding tasks. Indeed, the dimensionality of the latent space has been shown to impact the accuracy of many network inference tasks, including link prediction, clustering, and node classification [87]. One of the standard mapping approaches is maximum likelihood estimation, finding node coordinates of the network of interest by maximizing the likelihood that the network was generated as an RHG in the latent space. The likelihood function in the case of $\mathbb{H}^2$ is extremely non-convex to the node coordinates [39], making standard learning tools, like stochastic gradient descent, inefficient. Raising the dimensionality of the latent space $\mathbb{H}^{d+1}$ may lift some of the local maxima of the likelihood function, potentially leading to faster and more accurate graph embedding algorithms [100].

4

COMPLEMENTARITY IN COMPLEX NETWORKS

It doesn't matter how beautiful your theory is,
if it doesn't agree with experiment, it's wrong.

Richard P. Feynman

In many networks—including networks of protein-protein interactions, interdisciplinary collaboration networks, and semantic networks—connections are established between nodes with complementary rather than similar properties. While complementarity is abundant in networks, we lack the mathematical intuition and quantitative methods to study the complementarity mechanisms in these systems. In this Chapter, we close this gap by providing a rigorous definition of complementarity and developing a geometric framework for modeling complementarity-driven networks. We demonstrate the utility of the complementarity-based framework by learning the geometric representations of several real systems. Complementarity not only offers novel practical analytical methods but also enhances our intuition about formation mechanisms in networks on a broader scale, calling for a careful re-evaluation of existing similarity-inspired methods.

This Chapter is based on G. Budel and M. Kitsak (2024), Complementarity in Complex Networks, *arXiv preprint arXiv:2003.06665* [105].

4.1. INTRODUCTION

Complementarity plays an important role in many real networks, including networks of molecular interactions, interdisciplinary scientific collaborations, semantic networks, and production networks. Similar protein molecules are not guaranteed to interact. Instead, interacting proteins seem to exhibit complementarity at several levels, including electrostatic complementarity [106], hydrophobic mismatch [107–109], and shape complementarity [110, 111]. Collaboration teams, be it scientific collaborators, advisory boards, or military units, greatly benefit from combining experts who complement each other's knowledge and skills. Semantic networks form another domain where complementarity could play an important role. Indeed, our language would be rather bleak if we used only similar words together in a sentence. Comparisons, generalizations, and analogies are only a few of many examples of elements that make colorful representations in modern language. All these examples indicate that we construct sentences with words that are complementary rather than similar in meaning. Finally, recent work finds that complementarity plays a key role in the formation of production networks since companies are especially similar to their close competitors but not to their trading partners [112].

Intuitively, complementarity is when one object complements another one by contributing properties or features lacking in the other object. Unfortunately, our understanding of complementarity and its mechanisms in network formation and dynamics does not go far beyond this general definition. We lack both an intuitive understanding of complementarity and the methods to quantify complementarity mechanisms in networks. In this work, we aim to fill in both gaps by proposing a principled complementarity framework and developing methods to learn complementarity representations of real networks.

In the existing literature, complementarity-based networks are routinely analyzed with similarity-inspired methods. This came to be because the mechanisms of similarity are much better understood. Indeed, similarity plays a key role in the formation of social interactions—the more similar individuals are, the higher the chance of their interaction. Social interactions are relatively easy to measure and document, and, historically, the first studies of social networks can be traced back to the beginning of the 20th century. Much later, in the early 2000s, the availability of large-scale social networks fueled the rise of Network Science and Data Science. As a result, the majority of existing network analysis and inference techniques have been either developed for or inspired by social networks.

Another reason for the wide acceptance of similarity methods is, arguably, their geometric interpretation. Network nodes can be viewed as points in a certain latent space, such that their similarity is a function of the distance between the points. The smaller the distance, the larger the similarity between the nodes. These geometric frameworks can be traced back to sociology in the 1970s [81]. It is the similarity interpretation of distances in the latent space that leads to a large array of network analysis methods, including link prediction [29, 44, 84, 113–120], soft community detection and clustering [24, 121–123], network navigation [28, 32, 33, 35], and search [124–126].

The success of network embeddings in representations of social networks—where node similarity is recognized as one of the key mechanisms for link formation—can be

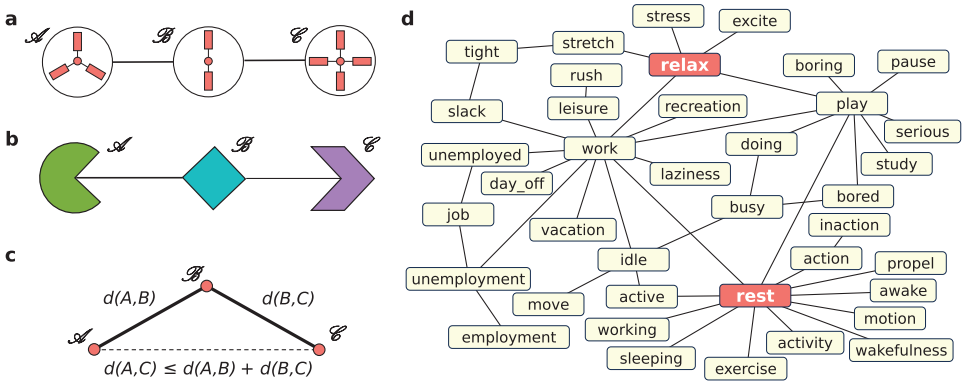


Figure 4.1: (a) Similarity-based toy network. Nodes are depicted as propellers with a variable number of blades, all nodes are similar since they are all propellers. Similarity is transitive: if A is similar to B, and B is similar to C, then A is also similar to C. (b) Complementarity-based toy network: the complementarity of two nodes is shown by matching interfaces. Complementarity is not transitive: if A is complementary to B (there are matching interfaces) and B is complementary to C (there are matching interfaces), A is not guaranteed to be complementary to C (there are no matching interfaces). (c) Embedding the toy network into a metric space imposes constraints on the distances due to the triangle inequality. (d) A subgraph of the 'Antonym' semantic network. Note that the rest-relax and unemployed-unemployment node pairs share 2 common neighbors each and, hence, can be incorrectly interpreted as being connected if similarity-based methods are applied to the network.

attributed to the agreement between the transitive property of similarity and the metric property of the latent spaces used in network embeddings. Indeed, consider the similarity-shaped toy network in Figure 4.1(a). If node A is similar to node B and B is similar to C, by transitivity, A is expected to be similar to C. A generic network embedding, Figure 4.1(c), would result in a small distance $d(A,B)$ since nodes A and B are connected. Likewise, $d(B,C)$ is also expected to be small since nodes B and C are connected. Then $d(A,C)$ must be small by the triangle inequality: $d(A,C) \leq d(A,B) + d(B,C)$, implying that nodes A and C are similar and, therefore, may be connected.

Network embedding methods are not readily applicable to complementarity-driven systems due to the non-transitivity of complementarity. Imagine that the same toy network is now formed by principles of complementarity, Figure 4.1(b). The complementarity of A and B and of B and C does not imply the complementarity of A and C. The complementarity network in Figure 4.1(b), however, is identical to its similarity counterpart in Figure 4.1(a), resulting in the same embedding as in Figure 4.1(c). Consequently, a small distance $d(A,C)$ in the embedding of the complementarity network immediately leads to inference errors. If the goal of the network embedding is to predict missing links, for instance, a small distance $d(A,C)$ implies a false positive link candidate between A and C. If, on the other hand, the problem at hand is cluster analysis, A, B, and C may end up in the same cluster of nodes in $\mathscr{M}$, implying that they all have similar characteristics, which is not the case for a complementarity-based network. As a practical illustration, we examine a subgraph of the 'Antonym' semantic network derived from ConceptNet [127], which can be regarded as a colloquial complementarity network. Here, nodes are words, and links are established between words with opposite meanings, Fig-

ure 4.1(d). The 'rest-relax' word pair are synonyms, in the 'Antonym' network these two words share two antonyms: 'work' and 'play'. Being synonyms, the words 'rest' and 'relax' are not connected in the 'Antonym' network. If a similarity-based link prediction algorithm is indiscriminately applied to the 'Antonym' network, however, the 'rest-relax' pair can be falsely interpreted as a missing link candidate, contrary to our intuition.

The demonstrated non-transitivity of complementarity principles prompts a revision of existing network-based and embedding-based approaches, aiming to adapt them for use with complementarity-driven networks, which we do in the following sections of this Chapter. In Section 4.2, we define the principled complementarity framework, quantifying the complementarity between nodes as their ability to execute certain functions or tasks. In Section 4.3, we analyze the properties of the principled complementarity framework to deduce a minimal practical framework for learning representations of real systems, while we define the ensuing graph model of the minimal framework in Section 4.4. In Section 4.5, we use the minimal complementarity framework to learn the complementarity representations of five real networks. We conclude this Chapter with the summary and outlook in Section 4.6.

The idea that the complementarity principle is different from similarity is quickly getting traction in the scientific community. Therefore, before proceeding further, we find it important to pause and acknowledge other concurrent efforts to study complementarity mechanisms in networks. Shortly after the initial work of Kovács *et al.* [128], which highlights the role of complementarity in protein-protein interaction (PPI) networks and the first version of Budel and Kitsak [105], Mattson *et al.* [112] discovered the prominent role of complementarity in economic production networks. Another recent work by Talaga and Nowak [129] proposes to quantify the relative role of similarity and complementarity in a network through densities of triangle and quadrangle subgraphs. While these works are invaluable in developing our intuition about complementarity in real networks, we lack the rigorous mathematical frameworks that would allow us to model synthetic complementarity networks and learn complementarity representations of real networks.

4.2. TOWARDS A PRINCIPLED COMPLEMENTARITY FRAMEWORK

To define complementarity from first principles, we refer to the Oxford English Dictionary, which asserts that 'two people or things that are complementary are different but together form a useful or attractive combination of skills, qualities or physical features'.

This definition implies that complementarity can be defined if each network node is characterized by at least two different skill types or features that might be able to complement each other. These features are intrinsic properties of network nodes that may or may not be readily observable. In the context of scientific collaborations, the features are the expertise of researchers in different disciplines. In the context of molecular interactions, the features might correspond to the chemical properties of molecules of interest.

To define the complementarity between two nodes i and j, we need to quantify the extent to which i and j form a useful combination of features or skills. One way to define such usefulness is through tasks or functions that nodes i and j can jointly execute. Depending on the system of interest, the task can be either a scientific problem in the case of a scientific collaboration network, or a biological function that two molecules can ex-

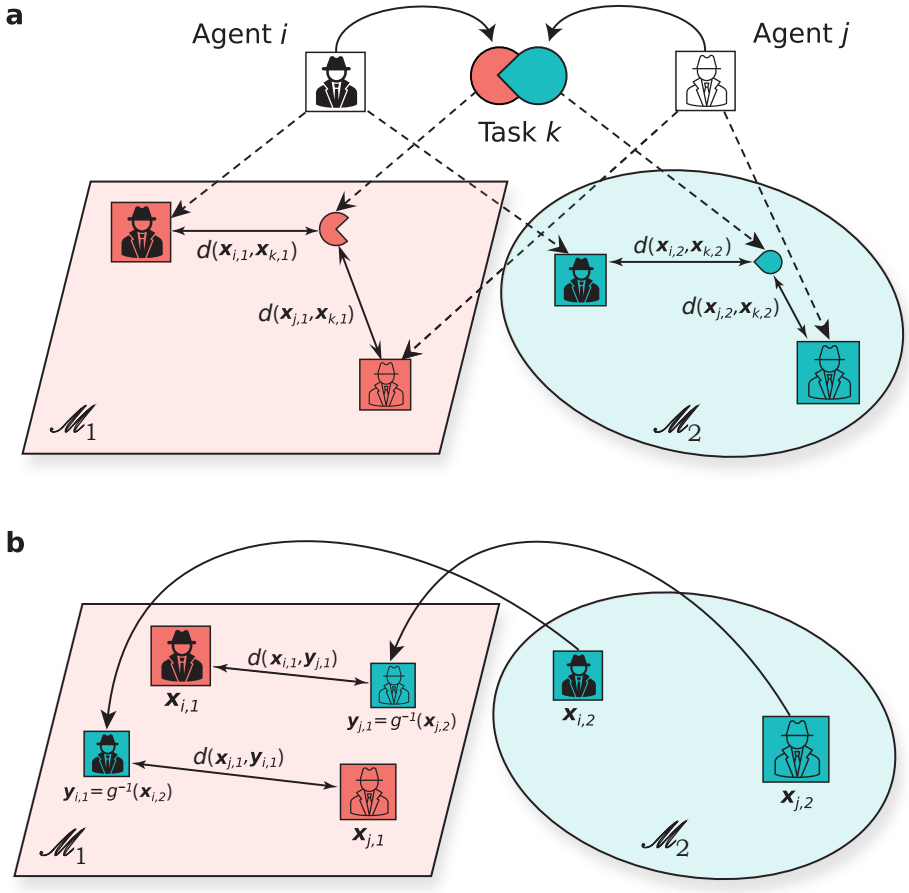


Figure 4.2: (a) The principled complementarity framework. Agents i and j complement each other in solving an interdisciplinary task k. The two disciplines related to the task k are represented as the latent metric spaces $\mathscr{M}_1$ and $\mathscr{M}_2$. Task k consists of two independent disciplinary parts that are represented as points $\mathbf{x}_{k,1}$ and $\mathbf{x}_{k,2}$ in the latent spaces $\mathscr{M}_1$ and $\mathscr{M}_2$, respectively. Likewise, the skill sets of agents i and j are the points $\{\mathbf{x}_{i,1},\mathbf{x}_{i,2}\}$ and $\{\mathbf{x}_{j,1},\mathbf{x}_{j,2}\}$, respectively. Agent i has a higher probability to solve the subtask in $\mathscr{M}_1$ if $d(\mathbf{x}_{i,1},\mathbf{x}_{k,1}) < d(\mathbf{x}_{j,1},\mathbf{x}_{k,1})$. Similarly, agent j has a higher probability to solve the subtask in $\mathscr{M}_2$ if $d(\mathbf{x}_{j,2},\mathbf{x}_{k,2}) < d(\mathbf{x}_{i,2},\mathbf{x}_{k,2})$. (b) The minimal complementarity framework. The minimal complementarity framework is a special case of the principled complementarity framework when task coordinates in the latent spaces $\mathscr{M}_1$ and $\mathscr{M}_2$ are related by an injective continuous map $g : \mathscr{M}_1 \to \mathscr{M}_2$ for $k = 1,\ldots,T$. In this case, the agent coordinates in $\mathscr{M}_2$ are effectively mapped to $\mathscr{M}_1$, $\mathbf{y}_{i,1} = g^{-1}(\mathbf{x}_{i,2})$ for $i = 1,\ldots,N$, see (4.6). As a result, the complementarity between agents i and j is quantified by the distances $d(\mathbf{x}_{i,1},\mathbf{y}_{j,1})$ and $d(\mathbf{x}_{j,1},\mathbf{y}_{i,1})$, the smaller the distances, the higher the complementarity.

ecute by forming a physical interaction in the case of a PPI network. In our framework, tasks mediate the complementarity between agents and are also characterized by two features each.

All aspects considered, to introduce a complementarity-based graph G, we consider N nodes, each of which is characterized by two skills $\mathbf{x}_{i,1}$ and $\mathbf{x}_{i,2}$, $i = 1, \ldots, N$, and T tasks that are also characterized by two skills each, $\mathbf{x}_{k,1}$ and $\mathbf{x}_{k,2}$, $k = 1, \ldots, T$.

To quantify the complementarity between any two nodes i and j, we need to be able to compare their skills. To do so, we postulate that skills are nothing else but points in two latent spaces $\mathscr{M}_1$ and $\mathscr{M}_2$. Therefore, distances can be defined between features of the same type but not necessarily between features of different types. In the case of scientific collaboration, for instance, it is straightforward to compare two skill sets within the same discipline but it is not for two skill sets from different disciplines. Thus, in agreement with the similarity intuition, we postulate that the distance $d(\mathbf{x}_{i,a}, \mathbf{x}_{j,a})$ between two nodes i and j quantifies the similarity between the features $\mathbf{x}_{i,a}$ and $\mathbf{x}_{j,a}$ in the space $\mathscr{M}_a$. The smaller the distance $d(\mathbf{x}_{i,a}, \mathbf{x}_{j,a})$, the larger the similarity between i and j with respect to skill a.

We are now in a good position to define complementarity. For a task k, define a graph G_k with adjacency matrix $\mathbf{A}_k$ that indicates if two nodes jointly execute the task k. The complementarity between any two nodes i and j with respect to task k is the probability $\Pr[(\mathbf{A}_k)_{i,j} = 1]$ that the two agents i and j can jointly execute k. In the most basic setting, one can think of the task k as consisting of two independent parts 1 and 2 corresponding to $\mathscr{M}_1$ and $\mathscr{M}_2$, respectively. Then, task k can be executed either as a result of node i executing part 1 and node j executing part 2, or vice versa:

$$\begin{aligned}\Pr[(\mathbf{A}_k)_{i,j} = 1] = 1 - [1 - F_1(d(\mathbf{x}_{i,1}, \mathbf{x}_{k,1}))F_2(d(\mathbf{x}_{j,2}, \mathbf{x}_{k,2}))] \\ \cdot [1 - F_1(d(\mathbf{x}_{j,1}, \mathbf{x}_{k,1}))F_2(d(\mathbf{x}_{i,2}, \mathbf{x}_{k,2}))]. \end{aligned} \tag{4.1}$$

Here, $F_a(d(\mathbf{x}_{i,a}, \mathbf{x}_{k,a}))$ is the probability that node i can independently execute part a of task k, for $a = 1, 2$. Our physical intuition suggests that the connection probabilities $F_1(d)$ and $F_2(d)$ are decreasing functions of the distances d, since the smaller the distance, the more similar the agent's expertise $\mathbf{x}_{i,a}$ is to the task requirements $\mathbf{x}_{k,a}$ in $\mathscr{M}_a$.

Assuming independence between the available tasks, the complementarity between agents i and j is the probability $p_{ij} \equiv \Pr[a_{ij} = 1]$ they can co-execute at least one task:

$$p_{ij} = 1 - \prod_k (1 - \Pr[(\mathbf{A}_k)_{i,j} = 1]). \tag{4.2}$$

In other words, the agents are connected in G with the probability that they are connected in at least one of the graphs G_k, $k = 1, \ldots, T$. Equations (4.1) and (4.2) serve as a foundation for the general framework for modeling and learning complementarity representations in real networks, which we summarize in Figure 4.2(a). From the modeling perspective, the complementarity framework can be used to generate both bipartite (when both agents and tasks are present) and conventional networks (when only agents are considered). To generate a complementarity-based synthetic network, one needs to define the latent spaces $\mathscr{M}_1$ and $\mathscr{M}_2$, the task execution probability functions $F_1(d)$ and $F_2(d)$, and the mechanism of distributing nodes and tasks in the two latent spaces. Since the connections between the nodes are functions of the distances

between node positions in the latent spaces, and the network links are established independently of one another, the complementarity framework belongs to the class of network models with hidden variables [96, 130]—with node positions serving the roles of hidden variables—allowing for analytical treatment.

From a learning perspective, obtaining a complementarity representation of a real network includes learning the characteristics of the latent spaces $\mathscr{M}_1$ and $\mathscr{M}_2$, the task execution probability functions $F_1(d)$ and $F_2(d)$, as well as the coordinates of the agents and tasks in the two latent spaces from the observed adjacency matrix $\mathbf{A}$ of the network of interest.

4.3. MINIMAL COMPLEMENTARITY FRAMEWORK

From practical considerations, learning the principled complementarity representation may prove suboptimal. One reason is that in many systems, tasks are either poorly defined, as in the case of molecular interaction networks, or not observable, such as in the case of semantic networks. Another reason is that the principled complementarity framework includes too many 'degrees of freedom', which may lead to overfitting when learning the representations of real systems. Therefore, our next step is to simplify the framework that covers both agents and tasks to a more practical one including only agents.

When the connection probabilities in the task graphs G_k are small, $\Pr\left[(\mathbf{A}_k)_{i,j}=1\right] \ll 1$, the largest contribution to the connection probability p_{ij} (4.2) in the complementarity graph comes from the linear terms in $\Pr\left[(\mathbf{A}_k)_{i,j}=1\right]$, resulting in

$$p_{ij} \approx \sum_k \Pr\left[(\mathbf{A}_k)_{i,j}=1\right]. \tag{4.3}$$

If the number of tasks T is large, we may replace individual tasks with a certain mean-field, treating them as drawn independently at random from a joint distribution with pdf $f_T(\mathbf{X}_1,\mathbf{X}_2)$. In this scenario, we can replace the sums over individual tasks in (4.3) with integrals over the volumes of $\mathscr{M}_1$ and $\mathscr{M}_2$, thereby obtaining, to the leading order,

$$p_{ij} = \int_{\mathscr{M}_1}\int_{\mathscr{M}_2} \mathrm{d}\mathbf{X}_1\mathrm{d}\mathbf{X}_2 f_T(\mathbf{X}_1,\mathbf{X}_2)\left[F_1\left(d\left(\mathbf{x}_{i,1},\mathbf{X}_1\right)\right)F_2\left(d\left(\mathbf{x}_{j,2},\mathbf{X}_2\right)\right) + i \leftrightarrow j\right], \tag{4.4}$$

where the second term represented by $i \leftrightarrow j$ is identical to the first term but with subscripts i and j swapped. The next simplification can be made by assuming a particular functional form for $f_T(\mathbf{X}_1,\mathbf{X}_2)$. To this end, the simplest case is that of uncorrelated task coordinates, $f_T(\mathbf{X}_1,\mathbf{X}_2) = f_1(\mathbf{X}_1)f_2(\mathbf{X}_2)$. This choice results in a trivial degenerate model, where the collaboration probability of any two agents i and j, to the leading order, is proportional to the product of their individual productivities.

Empirical studies suggest that a more interesting scenario arises when there is a strong correlation between the task coordinates in $\mathscr{M}_1$ and $\mathscr{M}_2$ [131–133]. In such cases, the simplest choice for $f_T(\mathbf{X}_1,\mathbf{X}_2)$ is established through a deterministic injective continuous function $g:\mathscr{M}_1 \to \mathscr{M}_2$ that maps skills between the two spaces, such that $\mathbf{X}_2 = g(\mathbf{X}_1)$. Consequently, the task coordinates $\mathbf{x}_{k,1}$ in $\mathscr{M}_1$ uniquely determine the task coordinates $\mathbf{x}_{k,2}$ in $\mathscr{M}_2$, and the pdf of the task distribution is given by

$$f_T(\mathbf{X}_1,\mathbf{X}_2) = f(\mathbf{X}_1)\delta\left(\mathbf{X}_2 - g(\mathbf{X}_1)\right), \tag{4.5}$$

where δ is the multi-dimensional Dirac delta function.

The strong correlation between task coordinates provides an effective mapping between skills in $\mathscr{M}_1$ and $\mathscr{M}_2$. Indeed, for every point $\mathbf{y}$ in $\mathscr{M}_1$, there is a corresponding unique point $\mathbf{x} = g(\mathbf{y})$ in $\mathscr{M}_2$, and vice versa. Using this property and the explicit form of f_T given by (4.5), we integrate out $\mathbf{X}_2$, obtaining the collaboration probability p_{ij} as a function of type 1 skills in $\mathscr{M}_1$, $\mathbf{x}_{i,1}$ and $\mathbf{x}_{j,1}$, and the images of type 2 skills in $\mathscr{M}_1$, $\mathbf{y}_{i,1} \equiv g^{-1}(\mathbf{x}_{i,2})$ and $\mathbf{y}_{j,1} \equiv g^{-1}(\mathbf{x}_{j,2})$:

$$p_{ij} = \int_{\mathscr{M}_1} d\mathbf{X}_1 f(\mathbf{X}_1) \left[F_1\left(d(\mathbf{x}_{i,1}, \mathbf{X}_1)\right) F_2\left(d(g(\mathbf{y}_{j,1}), g(\mathbf{X}_1))\right) + i \leftrightarrow j \right], \tag{4.6}$$

see Figure 4.2(b). Finally, if the tasks are distributed uniformly in $\mathscr{M}_1$, we get

$$p_{ij} = \frac{1}{\mathrm{vol}_{\mathscr{M}_1}} \int_{\mathscr{M}_1} d\mathbf{X}_1 \left[F_1\left(d(\mathbf{x}_{i,1}, \mathbf{X}_1)\right) F_2\left(d(g(\mathbf{y}_{j,1}), g(\mathbf{X}_1))\right) + i \leftrightarrow j \right], \tag{4.7}$$

where $\mathrm{vol}_{\mathscr{M}_1}$ is the volume of $\mathscr{M}_1$.

At this point, we can make another important observation for (4.7). The collaboration between agents i and j is likely if either $\mathbf{x}_{i,1}$ is close to $\mathbf{y}_{j,1}$, or $\mathbf{x}_{j,1}$ is close to $\mathbf{y}_{i,1}$, or both, see Figure 4.2(b). This is the case since $F_1(d)$ and $F_2(d)$ are both decreasing functions of d. Then, the largest contributions to the first term in (4.7) are those when both $d(\mathbf{x}_{i,1}, \mathbf{X}_1)$ and $d(g(\mathbf{y}_{j,1}), g(\mathbf{X}_1))$ are small at the same time. This is the case when $\mathbf{x}_{i,1} = \mathbf{y}_{j,1}$. By the same argument, the contributions to the second integral in (4.7) are maximized when $\mathbf{x}_{j,1} = \mathbf{y}_{i,1}$.

In summary, we have made two observations. The first one is that the strong correlation between task skills in $\mathscr{M}_1$ and $\mathscr{M}_2$ provides an effective mapping between the two spaces, allowing us to project type 2 skills from $\mathscr{M}_2$ onto $\mathscr{M}_1$ and vice versa. This means that one can reduce the complementarity framework to a single latent space. If this space is $\mathscr{M}_1$, type 2 skills in it are the images of type 2 skills from $\mathscr{M}_2$, $\mathbf{y}_1 = g^{-1}(\mathbf{x}_2)$. A second conclusion is that the complementarity between two agents i and j, quantified by their collaboration probability, is maximized when either node i's type 1 skill, $\mathbf{x}_{i,1}$, is close to node j's type 2 skill, $\mathbf{y}_{j,1}$, or when node i's type 2 skill, $\mathbf{y}_{i,1}$, is close to node j's type 1 skill, $\mathbf{x}_{j,1}$, or when both are close.

These findings allow us to guess a simpler form of the collaboration probability. Without loss of generality, we consider the projection of the generalized complementarity framework onto the latent space $\mathscr{M}_1$. Since there is only one latent space after the projection, we can simplify the notation by dropping the second subscripts of node coordinates indicating the type. Each agent i is characterized by two features or skills in a single latent space $\mathscr{M}$, which we refer to as $\mathbf{x}_i$ and $\mathbf{y}_i$. Consistent with the first observation that strong correlations between task features provide an effective mapping between the two manifolds, we postulate that the features of both types are represented by points in the same metric space $\mathscr{M}$: the features $\mathbf{x}_i$ are native in the space, while the features $\mathbf{y}_i$ are images projected by the tasks from the second space.

Satisfying the second observation, we postulate that the complementarity between any two agents i and j is a function of the distances between their skills of different types, $d(\mathbf{x}_i, \mathbf{y}_j)$ and $d(\mathbf{y}_i, \mathbf{x}_j)$, in $\mathscr{M}$:

$$p_{ij} = 1 - \left[1 - F_{\mathbf{A}}(d(\mathbf{x}_i, \mathbf{y}_j))\right] \left[1 - F_{\mathbf{A}}(d(\mathbf{y}_i, \mathbf{x}_j))\right], \tag{4.8}$$

where $F_{\mathbf{A}} : R^+ \to [0,1]$ is a probability distribution that is a decreasing function. Equation (4.8) prescribes that the complementarity probability p_{ij} is maximized when $\mathbf{x}_i$ is close to $\mathbf{y}_j$, $\mathbf{y}_i$ is close to $\mathbf{x}_j$, or both. The complementarity probability p_{ij} in (4.8) can be interpreted as the union of two probabilities: the complementarity between two agents is possible due to either the complementarity of skills $\mathbf{x}_i$ and $\mathbf{y}_j$, $\mathbf{y}_i$ and $\mathbf{x}_j$, or both.

A very important observation that we can make based on (4.8) is that complementarity can also model similarity in the special case where the two points of a given node are the same, $\mathbf{x}_i = \mathbf{y}_i$. If this is the case for a single node i, we can think of this node as self-complementary. In the context of molecular interactions, we are dealing with a molecule capable of interacting with itself. In the context of interdisciplinary collaborations, a self-complementary agent possesses sufficient skills in both domains to solve tasks independently. At the same time, self-complementary nodes can interact with other nodes that are complementary to them. If all network nodes are self-complementary, the complementarity framework reduces to a similarity framework. In this case, every node is effectively characterized by a single point in a space $\mathscr{M}$, since $\mathbf{x}_i = \mathbf{y}_i$ for $i = 1, \ldots, N$, and the complementarity between any two nodes i and j in (4.8) reduces to a function of the distance between the corresponding points, $p_{ij} = 2F_{\mathbf{A}}\big(d(\mathbf{x}_i,\mathbf{x}_j)\big) - F_{\mathbf{A}}\big(d(\mathbf{x}_i,\mathbf{x}_j)\big)^2$.

We note that the proposed minimal complementarity framework is consistent with the non-transitivity of complementarity: A being complementary to B, and B being complementary to C, does not imply that A is complementary to C. Indeed, within the minimal complementarity framework the formation of the A-B-C toy wedge network, Figure 4.3(a), is possible in two cases: (i) points $\mathbf{x}_A$, $\mathbf{y}_B$, and $\mathbf{x}_C$ are close to each other, as in Figure 4.3(b), or (ii) point $\mathbf{x}_A$ is close to point $\mathbf{y}_B$, and $\mathbf{x}_B$ is close to $\mathbf{y}_C$, as in Figure 4.3(c). Neither case creates constraints on the distances $d\left(\mathbf{x}_A,\mathbf{y}_C\right)$ and $d\left(\mathbf{y}_A,\mathbf{x}_C\right)$, relevant for the formation of the A-C link.

This observation is consistent with a recent result by Kovács *et al.* [128] in the prediction of protein-protein interactions, where the authors argued that the principles of complementarity suppress paths of even length between interacting proteins while simultaneously promoting the appearance of paths of odd length. To complete this comparison, consider the most likely geometric configurations that lead to an $\ell = 3$ path, Figure 4.3(d). As seen from Figures 4.3(e-h), two out of the eight possible configurations imply the appearance of a link between the endpoints of the $\ell = 3$ path. In other words, while the minimal complementarity framework does not suppress connections between the endpoints of $\ell = 3$ paths, it also does not strongly impose them. We revisit this question in Section 4.4, where we assess the densities of 3- and 4-cycles in a synthetic network generated from the minimal complementarity framework.

In our analysis, we use point representations for similarity-based fields: agents are defined as points, and similarities are quantified as distances between the points: the shorter the distance, the higher the similarity. While the distance-based definition of similarity is popular in social sciences, physics, and network science, there exists an alternative vector representation of similarity, which is more common in computer science. In a vector representation, each network node i is represented as a vector $\mathbf{v}_i$, and the similarity between any two nodes i and j is quantified by the inner product of the two vectors $\mathbf{v}_i \cdot \mathbf{v}_j$. We chose distance-based representations because they are identical for all metric spaces $\mathscr{M}$. While vector representations are simple in a vector space, for

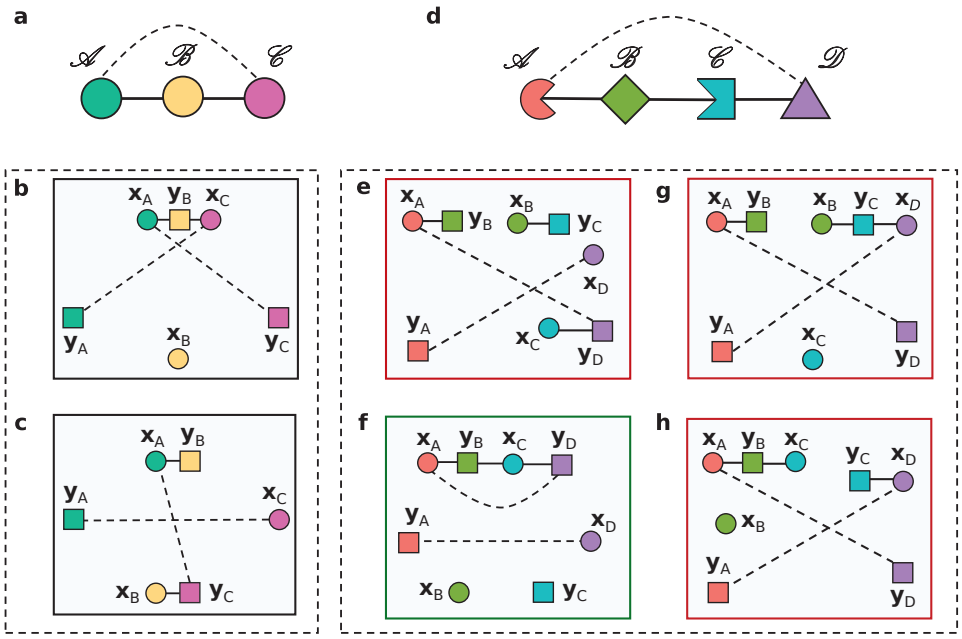


Figure 4.3: The minimal complementarity framework does not support the triangle closure rule and may support the diamond closure rule. (a) The triangle closure rule suggests that node A is connected to node C in the toy network, forming a triangle. Four complementarity representations may lead to the wedge toy network in (a). Two representations are shown in (b-c), and the other two are obtained from (b-c) by swapping the features of each node, $\mathbf{x}_i \leftrightarrow \mathbf{y}_i$. The links between node pairs (A,B) and (B,C) are possible due to latent-geometric proximity of points $\mathbf{x}_A$, $\mathbf{y}_B$, and $\mathbf{x}_C$ or latent geometric proximity of $\mathbf{x}_A$ and $\mathbf{y}_B$, and $\mathbf{x}_B$ and $\mathbf{y}_C$. Neither configuration imposes constraints on distances between point pairs $\mathbf{x}_A$ $\mathbf{y}_C$, and $\mathbf{y}_A$ and $\mathbf{x}_C$, which are relevant to the formation of the A-C tie. (d) The diamond closure rule suggests that node A is connected to D completing the diamond in the A-B-C-D toy network. There are 8 latent-geometric configurations corresponding to the toy network in d. Four configurations are shown in (e-h), and the remaining four are obtained from the configurations in (e-h) by swapping the features of each node, $\mathbf{x}_i \leftrightarrow \mathbf{y}_i$. Note that only two out of the possible eight configurations, (f) and its counterpart, imply the A-D connection.

example, *Euclidean* space, extending these representations to a non-vector space, such as the hyperbolic space, is not trivial.

It should be straightforward, starting with the same first principles, to derive the complementarity framework using vector representations. Here we conjecture that similar to the point representation, vector-based complementarity frameworks will characterize each node i by at least two vectors, which we call $\mathbf{u}_i$ and $\mathbf{v}_i$, such that the complementarity of two nodes i and j is a function of the inner products of vectors of different types,

$$p_{ij} = F_{\mathbf{A}}\left(\mathbf{u}_i \cdot \mathbf{v}_j; \mathbf{v}_i \cdot \mathbf{u}_j\right). \tag{4.9}$$

By examining (4.9), we realize that some network embedding methods previously developed in the machine learning and natural language processing (NLP) communities do represent network nodes or vectors by two vectors each. The most basic is, arguably, the singular value decomposition (SVD), which can be used to factorize the network adjacency matrix $\mathbf{A}$ as $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, where $\mathbf{\Sigma}$ is the diagonal matrix with the singular values of $\mathbf{A}$ on the diagonal, and $\mathbf{U}$ and $\mathbf{V}$ can be regarded as two vector representations for each network node. The Global Vectors for Word Representation (GloVe) unsupervised NLP approach represents each word i by two vectors $\mathbf{u}_i$ and $\mathbf{v}_i$, referred to as the target and context vectors, respectively [134]. GloVe quantifies the frequency of finding word i in the context of word j by a function of the inner product of the corresponding vectors, $\mathbf{u}_i \cdot \mathbf{v}_j$. Recently, the GloVe method has been adapted for network embedding, known as Global Vectors for Node Representations (GVNR) [135]. Since SVD and GVNR also use a dual-vector representation per node, they have the capacity to effectively learn complementary representations. Consequently, we examine the framework of complementarity in conjunction with the methods of SVD and GVNR.

In summary, the minimal complementarity framework can be used in both modeling synthetic complementarity networks and learning complementarity representations of real networks. We discuss them in the following sections. Complementary representations can be considered as a specific arrangement that combines two or more similarity representations. Indeed, we define complementarity as the ability of agents to co-execute interdisciplinary tasks in two distinct similarity-based spaces. It is the manifestation of tasks in the corresponding similarity-based fields that connects the agents and defines the nature of complementarity.

4.4. SYNTHETIC COMPLEMENTARITY-BASED GRAPHS

The minimal complementarity framework can be realized for any latent space $\mathscr{M}$ and any decreasing connection probability function $F_{\mathbf{A}}(d)$. We expect that for optimal results, both $\mathscr{M}$ and $F_{\mathbf{A}}(d)$ should be customized for, or learned from, the specific graph G under consideration. For brevity, we leave the task of determining $\mathscr{M}$ and $F_{\mathbf{A}}(d)$ for future work and focus on the example of using the hyperbolic disk, $\mathscr{M} = \mathbb{H}^2$, and the case of $F_{\mathbf{A}}(d)$ given by the sigmoid shape. We refer to the resulting model as the Complementarity Random Hyperbolic Graph (CRHG).

The CRHG is inspired by the success of the similarity-based RHG [21], which uses a representation of one point per node for modeling the properties of real similarity-driven networks, see Appendix B.1. As argued in Chapter 3, RHGs reproduce many

topological properties of real networks, such as sparsity, a scale-free degree distribution, strong local clustering, a hierarchical structure, and self-similarity [21–24]. In RHGs, nodes are sprinkled as points into a hyperbolic disk $\mathbb{H}^2$, and pairwise connections are established independently at random with probabilities that are decreasing functions of the distances between the corresponding points.

Following the minimal complementarity framework, we demand that every node of the CRHG is represented by two points in $\mathbb{H}^2$, and connections between nodes are established independently according to (4.8). In more precise terms, each node i is characterized by two points $\mathbf{x}_i = (r_{i,x}, \theta_{i,x})$ and $\mathbf{y}_i = (r_{i,y}, \theta_{i,y})$ in $\mathbb{H}^2$, and the connection probability

$$p_{ij} = F_{\mathbf{A}}(d(\mathbf{x}_i, \mathbf{y}_j)) + F_{\mathbf{A}}(d(\mathbf{x}_j, \mathbf{y}_i)) - F_{\mathbf{A}}(d(\mathbf{x}_i, \mathbf{y}_j))F_{\mathbf{A}}(d(\mathbf{x}_j, \mathbf{y}_i)), \tag{4.10}$$

4

where $d(\mathbf{x}_i, \mathbf{y}_j)$ is the distance between points $\mathbf{x}_i$ and $\mathbf{y}_j$ in $\mathbb{H}^2$,

$$\cosh d(\mathbf{x}_i, \mathbf{y}_j) = \cosh r_{i,1} \cosh r_{j,2} - \sinh r_{i,1} \sinh r_{j,2} \cos \Delta\theta_{ij}, \tag{4.11a}$$

$$\Delta\theta_{ij} = \pi - |\pi - |\theta_{i,1} - \theta_{j,2}||, \tag{4.11b}$$

where $\sinh x$ and $\cosh x$ are the hyperbolic sine and cosine functions [136].

We distinguish two special cases of the CRHG. In the case where the two points $\mathbf{x}_i$ and $\mathbf{y}_j$ are selected independently for each node i, we have the pure complementarity model. On the other hand, if the two points per node are equal to each other, $\mathbf{x}_i = \mathbf{y}_j$, the CRHG reduces to a similarity model that is very close to the standard RHG, with the only difference that each link is established with two independent attempts. In a more general setting, a certain mixture CRHG model is obtained if every node $i = 1, \ldots, N$ with probability $p \in (0,1)$ has two independent points $(r_{i,1}, \theta_{i,1})$ and $(r_{i,2}, \theta_{i,2})$, and with probability $1-p$ has identical points $r_{i,1} = r_{i,2}$ and $\theta_{i,1} = \theta_{i,2}$.

Below, we consider the pure CRHG, where $p = 0$, and postpone the analysis of the general $p \neq 0$ case for future work. Inspired by the RHG, we assume that the connection probabilities in (4.10) are of the sigmoid shape:

$$F_{\mathbf{A}}(d) = \frac{1}{1 + e^{\frac{d-R}{2T}}}, \tag{4.12}$$

for a certain distance d, and where $T \in (0,1)$ is the temperature parameter controlling the relevance of long-distance connections, and $R > 0$ is the radius of the hyperbolic disk. Small values of T result in graphs where most connections are established at distances of at most R. As T increases, links are more likely to form over distances $d > R$.

The node coordinates are drawn independently at random in $\mathbb{H}^2$. The coordinates of a node are the random variables $X_{0,t}$ for the radial coordinate r_t and $X_{1,t}$ for the polar coordinate θ_t, with $t \in \{1,2\}$, which are assumed to be independently distributed. The random variables $X_{0,t}$ and $X_{1,t}$ each follow the same distributions as X_0 and X_1 from the regular hyperbolic graph, Appendix B. Following the definition in [21], the radial coordinate r defined by X_0 follows the distribution

$$F_{X_0}(r) = \Pr[X_0 \leq r] = \frac{\cosh(\alpha r) - 1}{\cosh(\alpha R) - 1}, \quad 0 \leq r \leq R, \tag{4.13}$$

for a parameter $\alpha > \frac{1}{2}$ controlling the node density. The angular coordinate defined by X_1 is uniformly distributed $X_1 \overset{d}{=} U[0, 2\pi]$, with distribution function

$$F_{X_1}(\theta) = \Pr[X_1 \leq \theta] = \frac{\theta}{2\pi}, \quad 0 \leq \theta < 2\pi, \tag{4.14}$$

and probability density function $f_{X_1}(\theta) = \frac{dF_{X_1}(\theta)}{d\theta} = \frac{1}{2\pi}$. Using the hidden variable formalism [96], we find that the density $\Pr[D = k]$ of the CRHG degree distribution follows a power law, $\Pr[D = k] \sim k^{-\gamma}$, with $\gamma = 2\alpha + 1$. Indeed, the expected degree of a node characterized by the coordinates $\{(r_{i,x}, \theta_{i,x}), (r_{i,y}, \theta_{i,y})\}$ is approximated, to the leading order, by

$$\begin{aligned} E\big[D | X_{0,i} = r_{i,x}, X_{1,i} = \theta_{i,x}, Y_{0,i} = r_{i,y}, Y_{1,i} = \theta_{i,y}\big] \\ \approx E\big[D^{\mathrm{RHG}} | X_0 = r_{i,x}\big] + E\big[D^{\mathrm{RHG}} | X_0 = r_{i,y}\big], \end{aligned} \tag{4.15}$$

where $E[D^{\mathrm{RHG}} | X_0 = r]$ is the expected degree of a node with radial coordinate r in the standard RHG (B.30), Appendix B. We will further discuss the derivation of (4.15) in Chapter 5. Then, the degree distribution of the CRHG is the convolution of the degree distributions of two RHGs:

$$\Pr[D = k] = \sum_{k'=0}^{k} \Pr[D^{\mathrm{RHG}} = k - k'] \Pr[D^{\mathrm{RHG}} = k'] \sim k^{-\gamma}, \tag{4.16}$$

where $\gamma = 2\alpha + 1$.

Let M_{ij} be the number of common neighbors between nodes i and j. The expected number of common neighbors $E\big[M_{ij}\big]$ between two nodes with known coordinates can then be expressed, to the leading order, as

$$\begin{aligned} E\big[M_{ij}\big] \approx E\big[M_{ij}^{\mathrm{RHG}} | X_{0,i} = r_{i,x}, X_{1,i} = \theta_{i,x}, X_{0,j} = r_{j,x}, X_{1,j} = \theta_{j,x}\big] \\ + E\big[M_{ij}^{\mathrm{RHG}} | Y_{0,i} = r_{i,y}, Y_{1,i} = \theta_{i,y}, Y_{0,j} = r_{j,y}, Y_{1,j} = \theta_{j,y}\big], \end{aligned} \tag{4.17}$$

where $E\big[M_{ij}^{\mathrm{RHG}} | X_{0,i} = r_{i,x}, X_{1,i} = \theta_{i,x}, X_{0,j} = r_{j,x}, X_{1,j} = \theta_{j,x}\big]$ is the expected number of common neighbors of nodes with coordinates $(r_{i,x}, \theta_{i,x})$ and $(r_{j,x}, \theta_{j,x})$ in a regular RHG. Then, the probability $\Pr[M_{ij} = m]$ of two randomly chosen nodes in the CRHG to have exactly m common neighbors is the convolution of those in the RHG model,

$$\Pr\big[M_{ij} = m\big] = \sum_{m'=0}^{m} \Pr\big[M_{ij}^{\mathrm{RHG}} = m - m'\big] \Pr\big[M_{ij}^{\mathrm{RHG}} = m'\big] \sim m^{-\eta}, \tag{4.18}$$

since $\Pr\big[M_{ij}^{\mathrm{RHG}} = m\big] \sim m^{-\eta}$, where the exponent $\eta = \eta(\gamma, T) > 2$, as has been documented in both empirical [137, 138] and theoretical [139] studies.

When developing the complementarity framework in Section 4.3, we hypothesized that it would inhibit the emergence of odd-length cycles and, conversely, encourage the formation of even-length cycles. In line with our predictions, we found that the graph clustering coefficient diminishes as a function of the number of nodes N. Furthermore,

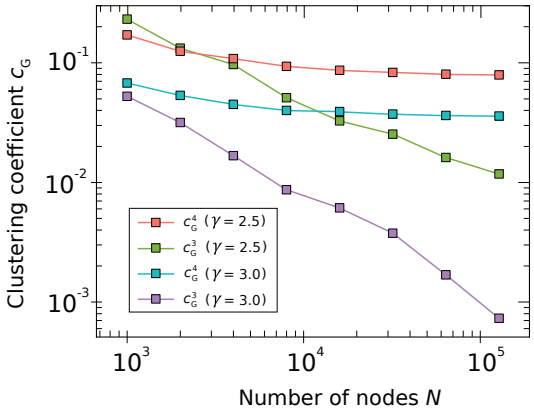


Figure 4.4: The graph clustering coefficient c_G^3 based on 3-cycles and the bipartite clustering coefficient c_G^4 based on 4-cycles as a function of the number of nodes N for synthetically generated CRHGs. We set the parameters $\alpha = 0.75$ ($\gamma = 2.5$) and $\alpha = 1.0$ ($\gamma = 3.0$), $E[D] = 20$, and $T = 0.5$. CRHGs are characterized by vanishing densities of 3-cycles and non-vanishing densities of 4-cycles.

in the CRHG model, as illustrated in Figure 4.4, the bipartite clustering coefficient remains constant with respect to N.

To assess how well the CRHG reproduces the topological properties of real complementarity-driven networks, we consider the 'Antonym' network, the human [140, 141] and yeast [142] protein-protein interaction (PPI) networks, the Messel shell food web [143], and a social network from the website `hamsterster.com` [144], see Table C.1 and Appendix C.1. We measure the density $\Pr[D = k]$ of the degree distribution, the degree-based clustering coefficient $c_G^3(i)$ of nodes i with degree k, and its bipartite counterpart $c_G^4(i)$ [139] for these networks. We compare the properties of the real networks to their degree-preserved randomizations, as well as synthetic CRHGs and RHGs with the same average degree $E[D]$, Figure 4.5. Refer to Van Mieghem *et al.* [145] for details on the rewiring procedure. We observe that both the CRHG (complementarity) and RHG (similarity) can adequately model the degree distributions of real networks, Figure 4.5(a,d,g,j,m). All of the real-world networks considered here have small clustering coefficients that are comparable to those of their corresponding degree-preserving randomizations and CRHGs, Figure 4.5(b,e,h,k,n). Similarity-based RHGs, on the other hand, tend to over-inflate with triangles, resulting in substantially higher clustering coefficients. Both the RHG and CRHG models are capable of generating networks with significant densities of 4-cycles, as measured by $c_G^4(i)$, Figure 4.5(c,f,i,l,o).

4.5. COMPLEMENTARITY REPRESENTATIONS OF REAL NETWORKS

After establishing that the CRHG model generates synthetic networks with topological properties similar to real complementarity-based systems, we next use the CRHG to learn hyperbolic complementarity representations of these real systems.

In a hyperbolic complementarity representation, every node $i = 1, \ldots, N$ from a graph G of interest is characterized by two points $\mathbf{x}_i = (r_{i,x}, \theta_{i,x})$ and $\mathbf{y}_i = (r_{i,y}, \theta_{i,y})$ in hyperbolic

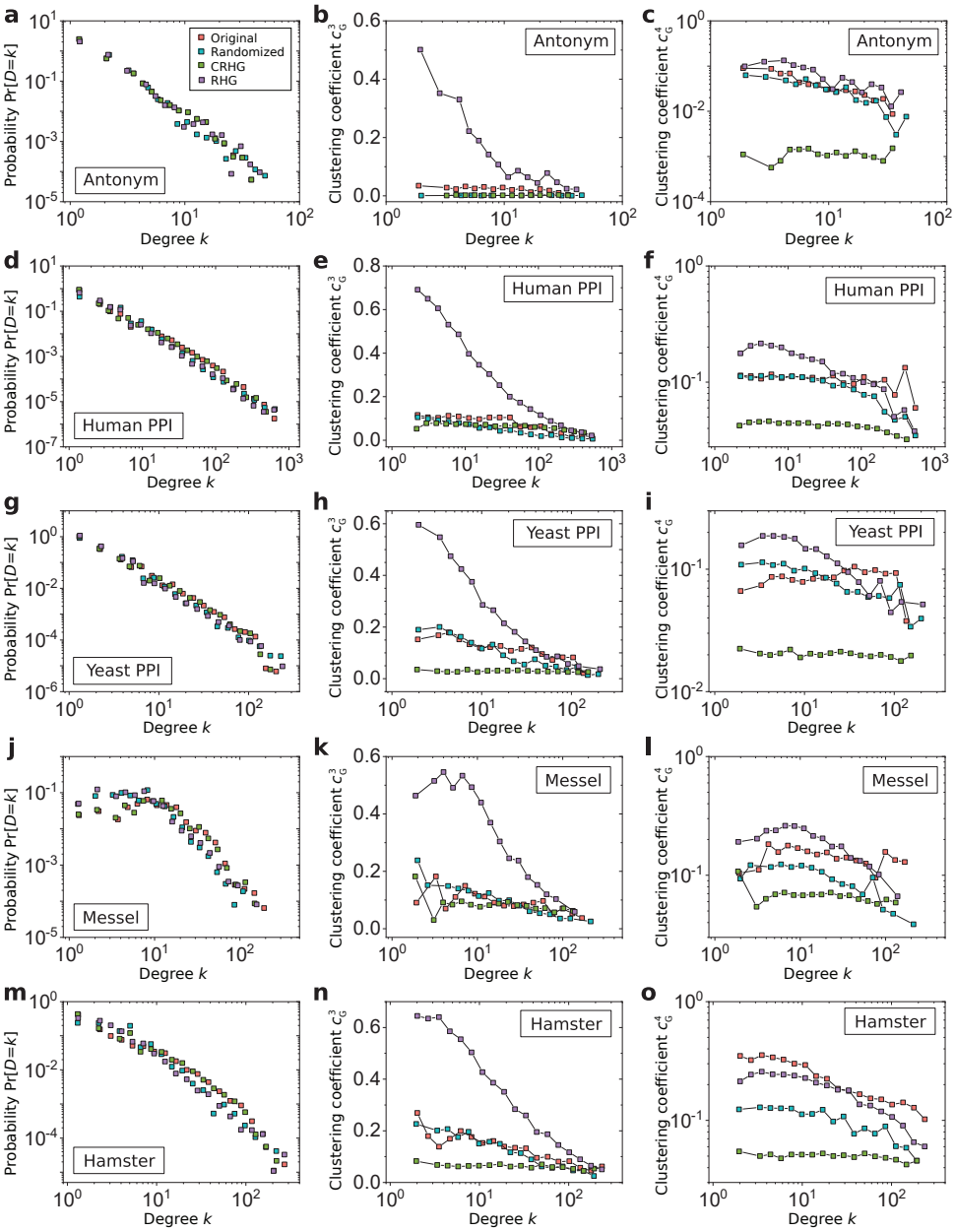


Figure 4.5: Topological properties of several complementarity-driven networks. Shown are the degree distribution densities $\Pr[D = k]$, the clustering coefficient $c_G^3(i)$ and bipartite clustering coefficient $c_G^4(i)$ of nodes i with degree k for (a-c) the 'Antonym', (d-f) human PPI, (g-i) yeast PPI, (j-l) Messel, and (m-o) `hamsterster.com` networks. The topological properties of the real networks (red) are compared with those of their degree-preserving randomizations (blue), CRHGs (green), and RHGs (purple).

disk $\mathbb{H}^2$. To learn complementarity representations, we employ maximum likelihood estimation. Given a graph G with an adjacency matrix $\mathbf{A}$, we aim to find $\{\mathbf{x}_i\}$ and $\{\mathbf{y}_i\}$ for $i = 1, \ldots, N$, maximizing the likelihood $\mathscr{L}$ that G was generated from the CRHG model with a given node assignment. Due to the independence of links in the CRHG, the thought likelihood is given by

$$\mathscr{L}\left(\mathbf{A} \,|\, \{(r_{i,x}, \theta_{i,x}), (r_{i,y}, \theta_{i,y})\}, \mathscr{P}\right) = \prod_{i<j} \left[q p_{ij}\right]^{a_{ij}} \left[1 - q p_{ij}\right]^{1-a_{ij}}, \tag{4.19}$$

where $\mathscr{P}$ is the set of model parameters, $\mathbf{A}$ is the adjacency matrix with values a_{ij}, p_{ij} are the connection probabilities prescribed by (4.10), and $q \in [0,1]$ is the probability that a link is present. We set $q = 1$ when learning the representations of fully observed networks. If on the other hand, a fraction λ of network links are unobserved, we set $q = 1-\lambda$. Our complementarity learning algorithm, which we call the Complementarity HyperLink (CHL) embedder, is identical to the HyperLink embedder [29], modulo the different likelihood function $\mathscr{L}$ given in (4.19), Appendix C.2.

Using the CHL embedder, we have obtained complementarity representations for the five complementarity-based systems under consideration. To evaluate the quality of these representations and the accuracy of the CHL embedder in link prediction, we conduct experiments by removing a fraction of $1 - q = 0.5$ links from each network, uniformly at random. This process creates a truncated version of the original network. Subsequently, we learn the complementarity representations of these truncated networks using the CHL embedder. To predict which links are missing, we rank all possible link candidates (unconnected node pairs) based on the sum of the complementarity distances between their corresponding points in the complementarity space:

$$\text{rank}_{ij} = d\left(\mathbf{x}_i, \mathbf{y}_j\right) + d\left(\mathbf{y}_i, \mathbf{x}_j\right), \tag{4.20}$$

such that the smaller the rank, the higher the chance of a missing link. Figure 4.6 compares the link prediction results of CHL with those of other embedding methods. For comparison, we used the ordinary HyperLink Embedder (HL), Global Vectors for Node Representations (GVNR), Singular Value Decomposition (SVD), the `node2vec` [117] (N2VEC) and DEEPWALK [146] (DWALK) network embedding methods. As depicted in Figure 4.6, the link prediction accuracy of CHL is significantly higher than that of the similarity-based embedding methods HL, N2VEC, and DWALK. Furthermore, the accuracy of CHL surpasses that of GVNR and exceeds that of SVD in three out of five networks.

The notable accuracy of the SVD method can be attributed to its utilization of a dual-vector representation for each node, aligning with our hypothesis in Section 4.3 regarding the necessity of such a representation within the complementarity framework. Surprisingly, the GVNR embedding method, which also employs a dual-vector representation per node and thus holds promise for effectively representing complementarity-based systems, exhibited only baseline performance. This underperformance of GVNR may be ascribed to its reliance on random walks for augmenting the network's topology. Such random walks could dilute the distinct complementarity structure inherent to the network by promoting the formation of closed triangles, thereby averaging out the nuanced relationships that define complementarity.

To gain a more comprehensive understanding, we also compared the link prediction accuracy of CHL to representative non-embedding link prediction techniques, Figure 4.7. Our findings show that CHL outperforms all examined similarity-based methods, including the Resource Allocation (RA) [147] and the Adamic-Adar (AA) [148] methods. This result is expected since similarity-based link prediction methods promote triangle closure, which contradicts the complementarity principle. Moreover, in line with expectations, non-similarity-based methods, including the baseline preferential attachment (PA) method [149], demonstrate improved performance over similarity-based approaches across the five complementarity-based networks analyzed. Although CHL is also competitive when compared to non-similarity link prediction methods, it generally achieves the second-highest accuracy in our experiments, falling behind the $L3$ [128] method and, occasionally, the SPM [150] or Katz [151] methods.

The effectiveness of complementarity representations in predicting missing links, particularly in comparison to similarity-based methods, validates the accuracy of these representations. While there still is potential for further improvement in the complementarity-based link prediction method, we leave this problem for future work. Instead, we now shift our focus to the complementarity representation of the 'Antonym' semantic network to delve deeper into its complementarity structure and better understand the workings of the complementarity embedding method.

Rank \ Seed	***Bad***	***Free***	***Man***	***Real***
1	Wrong	Master	Men	Actual
2	Ugly	Freedom	He	Genuine
3	Incorrect	Free man	Guy	Truth
4	Unpropitious	Essential	Person	Authentic
5	Composite	Escape	Male	Nothing
6	Show	Costlessly	King	Fact
7	Uncover	Unshackle	Fellow	Sincere
8	Brat	Steal	Wife	Rigidity
9	Mole	Receive	Chick	Reality
10	Beneficial	Lord	Mistress	Farm

Table 4.1: Seed words (top row) and the words in their hyperbolic vicinity in the CHL representation of the 'Antonym' network. The words are ranked based on their similarity distances to the corresponding seed word, $d\left[\mathbf{x}_{\text{seed}},\mathbf{x}\right]$ and $d\left[\mathbf{y}_{\text{seed}},\mathbf{y}\right]$.

Figure 4.8 displays the complementarity representation of the 'Antonym' semantic network. While links in the 'Antonym' network are established between words with opposite meanings, we can use it to not only find new antonyms, as we demonstrated in the link prediction exercise but also to find synonyms. Indeed, consider two words i and j characterized by points $\left(\mathbf{x}_i,\mathbf{y}_i\right)$ and $\left(\mathbf{x}_j,\mathbf{y}_j\right)$, respectively. While distances between the points of different types—$d\left(\mathbf{x}_i,\mathbf{y}_j\right)$ and $d\left(\mathbf{x}_j,\mathbf{y}_i\right)$—quantify complementarity, distances between points of the same type—$d\left(\mathbf{x}_i,\mathbf{x}_j\right)$ and $d\left(\mathbf{y}_j,\mathbf{y}_i\right)$—quantify the similarities, according to the minimal complementarity framework. Using the above intuition, we iden-

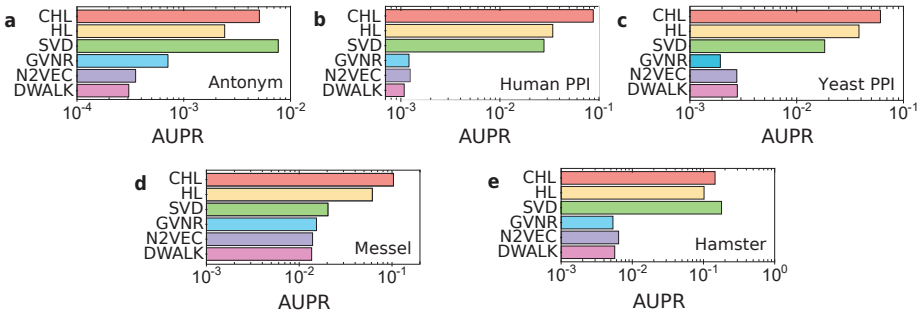


Figure 4.6: Embedding qualities as quantified by link prediction accuracy, measured by AUPR. Link prediction results obtained with the Complementarity HyperLink (CHL) method are compared to the ordinary HyperLink (HL) method, singular value decomposition (SVD), `node2vec` (N2VEC), and DEEPWALK (DWALK) in (a) the 'Antonym' semantic network (Antonym), (b) human protein-protein interaction network (Human PPI), (c) *S. cerevisiae* protein-protein interaction network (Yeast PPI), (d) Messel food web (Messel), and (e) `hamsterster.com` social network (Hamster). In all experiments, the fraction of removed links $1-q=0.5$. In all *Euclidean* embedding methods, we set the latent space dimensionality to $d=256$. We employed 100 epochs for the N2VEC and DWALK embeddings, and 10 epochs for the GVNR embeddings. For link prediction using the SVD, we truncated the resulting decomposition vectors to a dimension $d=256$.

tified the words closest to the general words 'bad', 'free', 'man', and 'real', finding that most of these words are indeed close in meaning to the seed words, see Table 4.1. When marked on the complementarity map of the network, as shown in Figure 4.8(a), each word group is represented by several geometric clusters. Unlike similarity networks, where typically only one similarity cluster is expected per group of similar nodes, complementarity representations can exhibit multiple such clusters. This phenomenon is particularly evident in the case of words near the word 'free', which possesses two meanings: 'no longer confined or imprisoned' and 'without cost or payment'. These two distinct meanings are represented by two points for the word 'free', located around the 6 and 1 o'clock positions, respectively, as illustrated in Figures 4.8(b-c). In the minimal complementarity framework, there can be as many as two main clusters for each group, corresponding to distinct properties of the nodes. Given that each node is depicted by two points and only one point is necessary to connect the node to others, the second point might be found outside the primary clusters. For instance, the word 'escape', which is associated with the seed word 'free', has one of its points in the 6 o'clock cluster, highlighting its similarity to 'free'. Meanwhile, the other point representing 'escape' is situated around 9 o'clock on the map, indicating its connection to concepts like 'away' and 'travel'.

4.6. DISCUSSION

We have developed two frameworks to quantify complementarity mechanisms in networks. The first, the principled complementarity framework, defines complementarity as a phenomenon emerging from the collaboration of agents with related skills. Within the principled framework, collaborating agents bring different but related skills that complement each other in their actions. Although inspired by scientific collaborations and

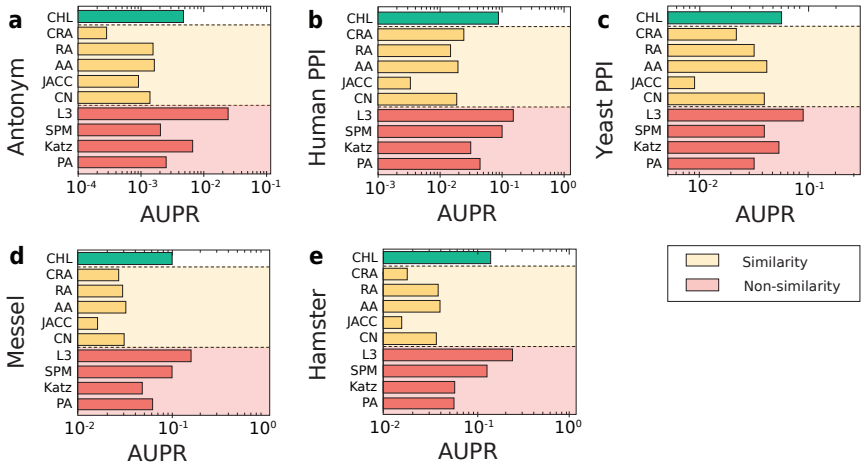


Figure 4.7: Link prediction results obtained with the Complementarity HyperLink (CHL) method in comparison with non-embedding link prediction methods. We conduct link prediction experiments in (a) the 'Antonym' semantic network (Antonym), (b) the human protein-protein interaction network (Human PPI), (c) *S. cerevisiae* protein-protein interaction network (Yeast PPI), (d) Messel food web (Messel), and (e) the `hamsterster.com` social network (Hamster) for a fraction $1-q=0.5$ of removed links. The considered link prediction methods are the (CHL) Complementarity HyperLink (our method), Cannistraci Resource Allocation (CRA) [152], Resource Allocation (RA) [147], Adamic Adar (AA) [148] index, Jaccard index (JACC) [153], the number of Common Neighbors (CN) [154], the $L3$ method [128], the Structural Perturbation Method (SPM) [150], the Katz index [151] with parameter $\beta=0.1$ (Katz), and the Preferential Attachment score (PA) [149],

team formation challenges, this concept can be easily adapted to other domains with minimal adjustments, including systems biology, ecological networks, and natural language processing. We anticipate that this principled framework will serve as a foundational basis for developing network complementarity models tailored to specific problems and systems of interest. Building on the principled framework, we have devised a practical minimal complementarity framework that facilitates the generation of synthetic complementarity-based networks and the learning of complementarity representations of real networks. We believe that the insights from our work will help us optimize existing and develop new analysis methods in the Science of Science, where interdisciplinary collaborations are of great importance, as well as in BioMedicine, which depends on understanding networks of molecular interactions to understand human diseases.

By comparing our complementarity framework to existing learning approaches, we found that the Singular Value Decomposition and the Global Vectors for Node Representations method are potentially applicable to complementarity-driven systems. Our results indicate, however, that additional improvements are required for these methods to optimize their predictive power in complementarity-based systems.

Since the complementarity perspective views each network node as represented by more than one point per node, an important question arises if complementarity can be embedded into a higher-dimensional similarity representation. Indeed, consider a minimal complementarity framework where each node corresponds to two points in a d-dimensional latent space. Consider the minimal complementarity framework, where

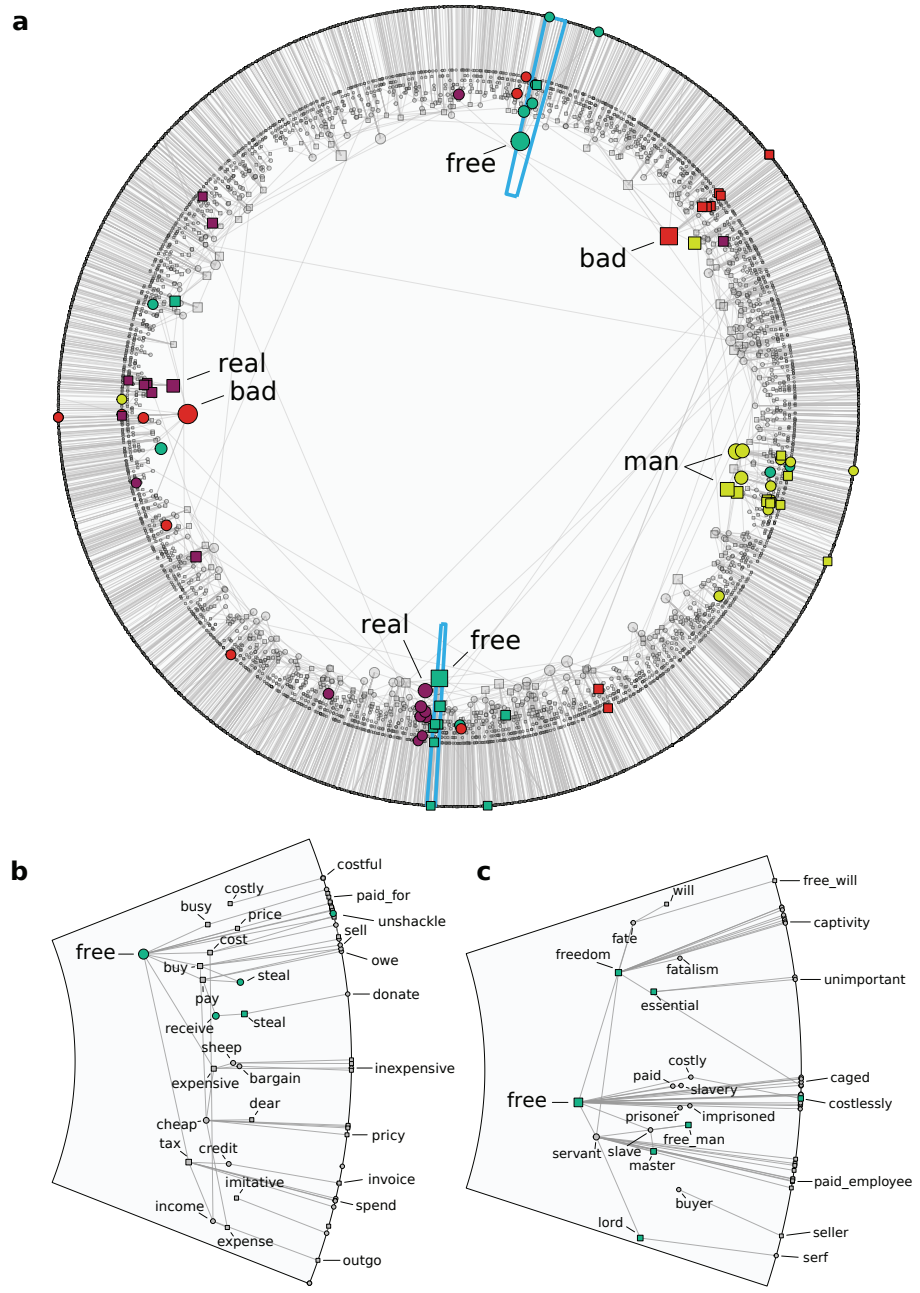


Figure 4.8: Complementarity representation of the 'Antonym' semantic network in a 2-dimensional hyperbolic disk. Each word is represented by two points shown with squares and circles. Larger node sizes correspond to more general words, as quantified by the number of their antonyms. Colored are the four groups of nodes closest to the words 'bad', 'free', 'man', and 'real', see Table 4.1. For every node pair (i, j), we draw only one link, consistent with the geometric definition of complementarity. In doing so, we connect either points $\mathbf{x}_i$ and $\mathbf{y}_j$, or $\mathbf{x}_j$ and $\mathbf{y}_i$, selecting the closest pair. (b-c) We zoom into the two regions surrounding the word 'free'. The angular node coordinates are stretched out to improve visibility.

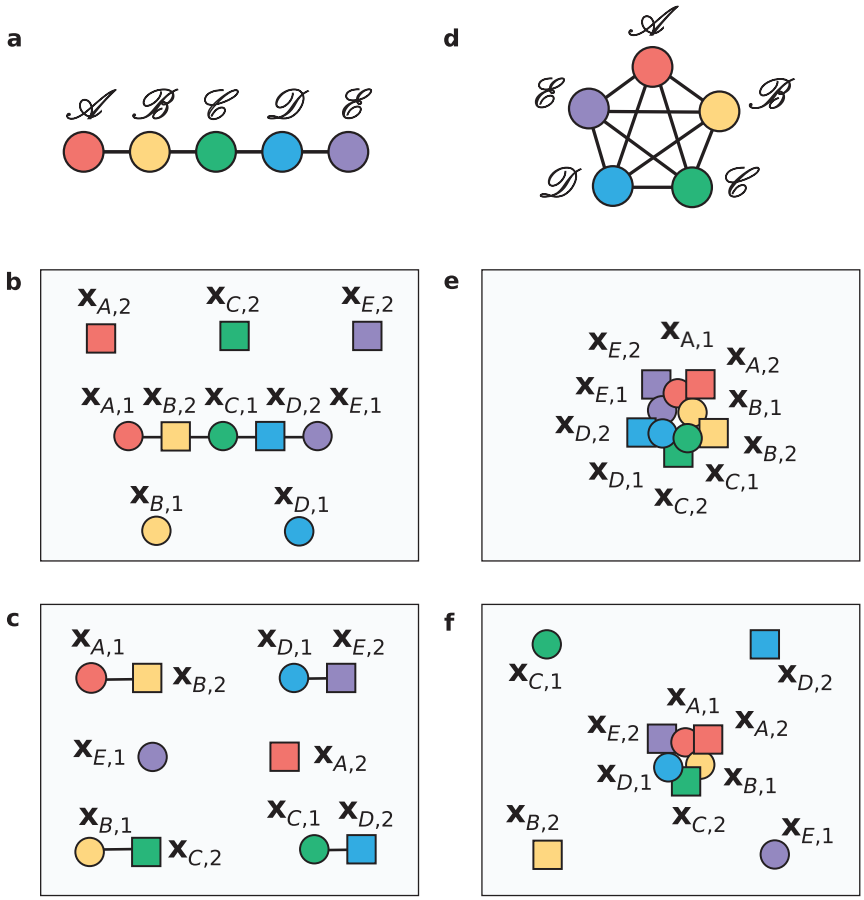


Figure 4.9: Paths and communities in complementarity-driven networks. Panel (a) showcases a toy network comprised of a chain of 5 nodes, while panels (b-c) present two potential spatial configurations that could give rise to this network. In configuration (b), the chain of connections observed in the network is formed by arranging complementary points into a geometric trajectory within the latent space. Configuration (c) illustrates how the chain of connections seen in (a) could emerge from pairwise proximities between points in the latent space. Unlike complementarity-driven networks, similarity-driven networks are limited to trajectory-based alignment as their sole mechanism. Panel (d) displays another toy network consisting of a 5-node clique, with panels (e-f) depicting two spatial configurations that could lead to such a network. Configuration (e) shows that the clique network can result from the complete clustering of all points in the latent space. Meanwhile, configuration (f) demonstrates that the clique network seen in (d) could also be produced from partially clustered points in the latent space. This is in contrast to similarity-driven networks, where the formation of a clique necessitates the complete clustering of all points in the latent space, showcasing the unique flexibility of complementarity-driven networks in modeling complex structures.

each node is represented by two points within a d-dimensional latent space. Formally, each node is then characterized by $2d$ features or vector components, and the connection probability between nodes is a function of $4d$ variables. This raises the question: is it possible to view the complementarity framework as a similarity framework, where each node corresponds to a single vector in a $2d$-dimensional latent space, rather than a d-dimensional space? Our current answer is *no*, at least not from the viewpoint of similarity. In the realm of similarity, the proximity of two vectors—or nodes, depending on the representation—dictates their similarity level, with maximal similarity achieved when two vectors are parallel or two points coincide. However, complementarity operates on a different principle: two nodes can be maximally complementary to each other even if there is alignment with respect to just one of their vectors or points.

How do we know if the system at hand is driven by similarity or by complementarity? One method to assess the relative presence of the two mechanisms is by measuring the relative densities of triangles and quadrangles in the network of interest [129]. The authors proposing this method confirm that the majority of studied biological networks are predominantly complementarity-driven, while social networks may have a significant presence of both similarity and complementarity, as quantified by densities of triangles and quadrangles, respectively. However, a broader conclusion can be drawn: purely similarity-driven or complementarity-driven systems do not exist in isolation. Fortunately, our findings suggest that similarity might be considered a special case of complementarity. Specifically, our minimal complementarity framework accommodates both complementarity and similarity principles. The similarity aspect is encapsulated in the special scenario where each node is represented by two identical points in the space, illustrating how our framework can seamlessly integrate the principles of both complementarity and similarity.

The complementarity perspective not only paves new pathways for analyzing systems driven by complementarity but also poses a challenge to the traditional methodologies of network science, which were primarily developed for social networks and subsequently applied to various other network categories. Concepts such as shortest paths and communities, derived from similarity-based networks, have become staples in the toolkit for studying complementarity-driven networks. Network community detection, for instance, is extensively utilized to identify disease and functional modules in biological networks [155–157], as well as to delineate scientific communities within collaboration networks [6, 158]. Similarly, shortest paths are frequently employed to measure separations between specific network modules [159, 160].

Expanding on the concept of the shortest path, which is typically visualized as a specific discrete trajectory within the latent space of a network, as depicted in Figure 4.9(a), we find that similar trajectories can also exist within a complementarity framework. In this context, a sequence of connections could arise from the spatial arrangement of complementary points forming a geometric trajectory, illustrated in Figure 4.9(b). However, such an alignment, while capable of creating a network chain, is not the only way to achieve connectivity. An alternative scenario involves a series of pairwise, yet distinct, proximities between corresponding points in the latent space, as shown in Figure 4.9(c). This highlights the versatility of complementarity frameworks in representing network structures through various forms of spatial organization.

Another example is the concept of the network community. Traditionally defined as a group of nodes that are densely connected internally and sparsely connected with external nodes, as discussed in Chapter 2. This definition implies that, within social sciences, network communities are typically visualized as clusters of similar node points that are geographically localized in the network, as shown in Figure 4.9(d-f). However, within a complementarity-driven system, the notions of similarity among nodes and the formation of a relatively dense subgraph are distinctly separate concepts. This distinction arises because similar nodes are not necessarily expected to connect, and connected nodes may not always be similar. As illustrated in the complementarity representation of the 'Antonym' semantic network in Figure 4.8, similar nodes might form multiple localized clusters in the latent space, with each cluster representing a different feature. In contrast, densely connected subgraphs, traditionally identified as communities in network science, can be represented in a complementarity framework as either fully, as in Figure 4.9(e), or partially localized, as in Figure 4.9(f). This distinction underscores the unique challenges and opportunities in identifying and understanding community structures within complementarity-driven systems.

In summary, we hope that the complementarity concept introduced in our work will serve to not only sharpen our intuition but also expand the toolkit of quantitative methods available for network analysis. These advancements aim to provide a more nuanced understanding of network dynamics, particularly in complementarity-driven systems, and offer innovative approaches for dissecting complex network structures and behaviors.

5

PROPERTIES OF COMPLEMENTARITY HYPERBOLIC GRAPHS

There are 10^{11} *stars in the galaxy. That used to be a huge number.*
But it's only a hundred billion. It's less than the [United States] national deficit!
We used to call them astronomical numbers.
Now we should call them economical numbers.

Richard P. Feynman

In this chapter, we study the topological properties of the Complementarity Random Hyperbolic Graph (CRHG). Not all intuitions gained from studying similarity-driven social networks also apply to networks driven by complementarity. We describe and illustrate CRHGs and study their topological properties. Just like with standard hyperbolic graphs, we find that we can construct complementarity hyperbolic graphs that are sparse and have a power-law degree distribution. CRHGs, however, are dominated by motifs of quadrangles rather than triangles, resulting in a lower-than-expected clustering coefficient and a lower effective graph resistance. A good understanding of the properties of CRHGs might help us design better statistical methods for real networks in which the link formation is driven by complementarity.

5.1. INTRODUCTION

Many of the statistical methods commonly used in Network Science are rooted in social network analysis [3]. Social networks are, in turn, for a large part driven by the principle of similarity: the tendency of individuals to connect to similar individuals. The study of social networks therefore naturally gave rise to well-studied phenomena like scale-free degree distributions [149], degree assortativity [161], the small-world property and strong local clustering of nodes [97], latent network geometry [21] and community structure [158]. Based on these properties, network scientists developed a wide range of tools that successfully helped us understand the governing mechanisms of link formation in many real networks.

There are, however, real networks in which the links were shown to establish not between similar nodes, but between complementary nodes. These complementarity-driven networks include protein-protein interaction networks [128], company-level production networks [112], and semantic networks [162]. In Chapter 4, we have described our geometric framework for embedding complementarity networks based on the Complementarity Random Hyperbolic Graph (CRHG) model. Here, we study the topological properties of the CRHG.

Consider, for example, interdisciplinary scientific collaboration networks, where the nodes are individuals and individuals connect if they together write an interdisciplinary scientific paper. Two individuals can collaborate if they together possess the skills to successfully write a paper; if the skill set of one author complements the skill set of the other. A biologist and a statistician together might write a brilliant paper at the intersection of Biology and Statistics, while neither two biologists nor two statisticians alone would likely be equally successful at this task. A complementary relationship can be conceptualized as two individuals (or nodes) possessing complementary properties, which together contribute to the formation of a link between them. Complementary relations lead to quadrangles in the network, rather than triangles like in social networks [128, 129]. The node properties that constitute the formations of complementarity links are potentially unobserved and, in that case, exclusively manifest through which other nodes a node forms links with.

In this chapter, we aim to further study the topological properties of the CRHG. Understanding the properties of a complementarity-driven network model will help us design better statistical methods for real-world networks in which the connections are not well explained by similarity alone.

5.2. COMPLEMENTARITY RANDOM HYPERBOLIC GRAPHS

The complementarity random hyperbolic graph (CRHG), defined in Section 4.4, is part of the class of random graph models with hidden variables proposed in [96]. In CRHGs, each node is described by more than one hidden variable, each of which is a coordinate in its hidden space $\mathscr{M}$. First, we describe the framework for graphs with hidden variables, how the framework can be adjusted to describe complementarity and, finally, we describe the connection probability function that incorporates the complementarity of two nodes.

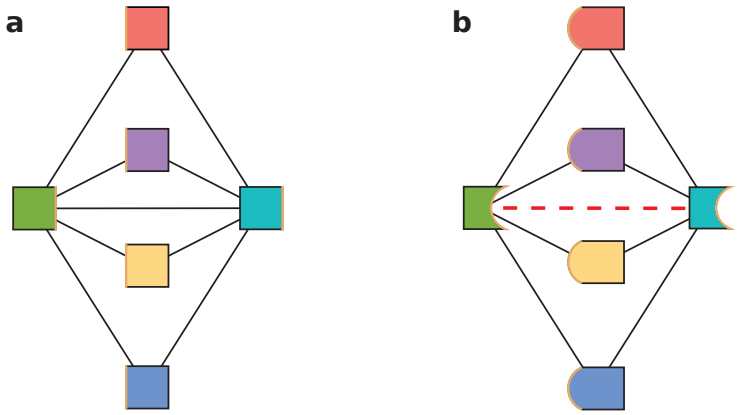


Figure 5.1: Toy example graphs based on (a) similarity and (b) complementarity link formation principles. Nodes might form a connection if the shapes of the edges indicated in orange are compatible. In panel (a), all nodes have the same shape (similarity), leading to multiple triangles. In panel (b), nodes only connect if their shapes are complementary, leading to an absence of triangles. Consequently, in panel (a), the blue and green nodes are connected, whereas in panel (b), they are not, as indicated by the red dashed line. This example is inspired by Kovács *et al.* [128].

5.2.1. HIDDEN REPRESENTATION

Let the hidden space $\mathscr{M}$ of a graph G with N nodes be a d-dimensional metric space in which each node i is represented by a d-dimensional point $\mathbf{x}_i \in \mathscr{M}$. Each element in $\mathbf{x}_i$ can be interpreted as a (distinct) feature or property of the node. The node features are in general unobserved, but a feature can also be an observable node property, such as the degree k_i. Often, the probability that two nodes i and j are connected in G is a decreasing function of the distance $d(\mathbf{x}_i,\mathbf{x}_j)$ in $\mathscr{M}$, where the distance function $d(\mathbf{x},\mathbf{y})$ is defined by the properties of the chosen metric space. For example, if $\mathscr{M}$ is the Euclidean space, then the distance $d(\mathbf{x}_i,\mathbf{x}_j)$ is the Euclidean distance between $\mathbf{x}_i$ and $\mathbf{x}_j$. Graphs with a connection probability function that is a decreasing function of the distance in the hidden space generally have large clustering coefficients [26, 27, 163]. A hidden space like the one described above can be interpreted as a 'similarity space'.

5.2.2. COMPLEMENTARITY REPRESENTATION

A complementarity graph G with N nodes can be created in a single d-dimensional hidden space $\mathscr{M}$ when each node is represented by $m > 1$ points in $\mathscr{M}$. The d-dimensional points $\mathbf{x}_1,\ldots,\mathbf{x}_m$ are random variables and each of the m points can be interpreted as a distinct set of features or properties of the node. Two nodes are said to be complementary if they have at least one pair of complementary points. We define two points $\mathbf{x}_k$ and $\mathbf{x}_l$, $k,l \in \{1,\ldots,m\}$ and $k \neq l$, to be complementary when their respective distance $d(\mathbf{x}_k,\mathbf{x}_l)$ in the space $\mathscr{M}$ is small. A point of type k for a node $i \in \{1,\ldots,N\}$ is denoted as $\mathbf{x}_{i,k}$. We define the complementarity of two nodes as follows.

For two nodes $i,j \in \{1,\ldots,N\}$ the distance $d(\mathbf{x}_{i,k},\mathbf{x}_{j,l})$, $k \neq l$, is also a measure of the extent of the complementarity of the two nodes: the lower the distance $d(\mathbf{x}_{i,k},\mathbf{x}_{j,l})$, the higher the complementarity of nodes i and j with respect to features k and l. The dis-

tance $d(\mathbf{x}_{i,k}, \mathbf{x}_{j,k})$ between their features of the same type k is a measure of the similarity of the two nodes i and j.

The simplest case of a complementarity random graph is where each node is represented by $m = 2$ points $\mathbf{x}_1, \mathbf{x}_2 \in \mathscr{M}$, or $\mathbf{x}, \mathbf{y} \in \mathscr{M}$ for the convenience of notation, such that there are only two points $\mathbf{x}$ and $\mathbf{y}$. The simple case is the graph model underlying the minimal complementarity framework, Section 4.3.

5.2.3. CONNECTION PROBABILITY

In a complementarity random graph, the connection probabilities depend only on the distances between points of different types in the latent space $\mathscr{M}$. We define the connection probability for the case with 2 points $\mathbf{x}$ and $\mathbf{y}$ per node, the probability p_{ij} that two nodes with coordinates $(\mathbf{X}_i, \mathbf{Y}_i) = (\mathbf{x}_i, \mathbf{y}_i)$ and $(\mathbf{X}_j, \mathbf{Y}_j) = (\mathbf{x}_j, \mathbf{y}_j)$ are connected,

$$p_{ij} \equiv \Pr\left[a_{ij} = 1 \middle| (\mathbf{X}_i, \mathbf{Y}_i) = (\mathbf{x}_i, \mathbf{y}_i), (\mathbf{X}_j, \mathbf{Y}_j) = (\mathbf{x}_j, \mathbf{y}_j)\right]. \tag{5.1}$$

The connection probability p_{ij} for nodes i and j is a function of the two distances $d(\mathbf{x}_i, \mathbf{y}_j)$ and $d(\mathbf{x}_j, \mathbf{y}_i)$, which are each plugged into a similarity-based connection probability distribution $F_{\mathbf{A}}(d)$ that is a decreasing function of the distance d, for $d \geq 0$. The connection probability

$$p_{ij} = 1 - \left[1 - F_{\mathbf{A}}\left(d(\mathbf{x}_i, \mathbf{y}_j)\right)\right] \cdot \left[1 - F_{\mathbf{A}}\left((\mathbf{x}_j, \mathbf{y}_i)\right)\right]. \tag{5.2}$$

The probability $1 - F_{\mathbf{A}}\left((\mathbf{x}_i, \mathbf{y}_j)\right)$ is the complement of the probability that the nodes are connected based solely on the features $\mathbf{x}_i$ and $\mathbf{y}_j$, thus that the nodes are not connected. The product $\left[1-F_{\mathbf{A}}\left((\mathbf{x}_i, \mathbf{y}_j)\right)\right] \cdot \left[1-F_{\mathbf{A}}\left((\mathbf{x}_j, \mathbf{y}_i)\right)\right]$ is then the probability of the intersection of the two nodes being not connected through either feature pairs $(\mathbf{x}_i, \mathbf{y}_j)$ and $(\mathbf{x}_j, \mathbf{y}_i)$. The probability (5.2) is the complement of the probability of that intersection. To see how the connection probability p_{ij} introduces complementarity into the hidden space, observe that the connection probability p_{ij} is large when either one of $F_{\mathbf{A}}\left((\mathbf{x}_i, \mathbf{y}_j)\right)$ and $F_{\mathbf{A}}\left((\mathbf{x}_j, \mathbf{y}_i)\right)$ is large, which is when either one of their two feature pairs of opposite types are close in the hidden space.

5.2.4. HYPERBOLIC GRAPHS

As demonstrated in Section 4.4, when the latent space is the two-dimensional hyperbolic disk, $\mathbb{H}^2$, as in [21], hyperbolic graphs based on complementarity exhibit the fundamental properties observed in regular hyperbolic graphs, which are prevalent in large real-world networks: a power-law degree distribution and a sparse average degree.

On the hyperbolic disk $\mathbb{H}^2$ with radius R and (negative) curvature $K = -\zeta^2$, for $\zeta > 0$, a point $\mathbf{x}$ is defined as $\mathbf{x} = (r, \theta)$, where the radial coordinate $r \in [0, R]$ is the distance to the origin of the space and the polar coordinate $\theta \in [0, 2\pi)$ is the angle with respect to the origin. The negative curvature ζ of the space is merely a scaling constant and we set $\zeta = 1$ to shorten the notation. In the random hyperbolic graph (RHG) model, the two coordinates of a node are random variables X_0 and X_1, which are assumed to be independently distributed. Following the definition in [21], the radial coordinate defined by X_0 has the distribution (4.13). The corresponding probability density function $f_{X_0}(r) = \frac{dF_{X_0}(r)}{dr} = \frac{\alpha \sinh(\alpha r)}{\cosh(\alpha R)-1}$, with $\sinh(x)$ the hyperbolic sine function. The angular

coordinate defined by X_1 is uniformly distributed $X_1 \overset{d}{=} U[0,2\pi]$ with distribution (4.14) and probability density function $f_{X_1}(\theta) = \frac{dF_{X_1}(\theta)}{d\theta} = \frac{1}{2\pi}$. The probability density function of the variable $\mathbf{X} = (X_0, X_1)$ is the product of the two individual probability density functions $f_{\mathbf{X}}(\mathbf{x}) = f_{X_0}(r) \cdot f_{X_1}(\theta)$ by the independence of the two variables. The radial coordinate X_0 directly influences the degree of a node, the closer to the origin of the space a node is, the higher its degree. The angle between the polar coordinates in X_1 of two nodes and their radial coordinates in X_0 determine the distance between the nodes.

The two points $\mathbf{x}_i$ and $\mathbf{x}_j$ form a hyperbolic triangle with the origin of the space, therefore the distance between $\mathbf{x}_i$ and $\mathbf{x}_j$ in the hyperbolic disk with curvature $K = -\zeta^2$ is computed through the hyperbolic law of cosines:

$$\cosh\zeta d(\mathbf{x}_i,\mathbf{x}_j) = \cosh\zeta r_i \cosh\zeta r_j - \sinh\zeta r_i \sinh\zeta r_j \cos\Delta\theta_{ij}, \tag{5.3}$$

with $\Delta\theta_{ij} = \pi - |\pi - |\theta_i - \theta_j||$ the angle between the polar coordinates θ_i and θ_j [21]. As we set $\zeta = 1$, the distance

$$d(\mathbf{x}_i,\mathbf{x}_j) = \operatorname{arccosh}\left(\cosh r_i \cosh r_j - \sinh r_i \sinh r_j \cos\Delta\theta_{ij}\right), \tag{5.4}$$

where $\operatorname{arccosh}(x)$ is the inverse hyperbolic cosine function [136]. The distance is well-approximated as $d(\mathbf{x}_i,\mathbf{x}_j) \approx r_i + r_j + 2\ln\left(\sin\left(\Delta\theta_{ij}/2\right)\right)$ when both r_i and r_j are sufficiently large, as first proposed in [21, 28] and briefly shown in Appendix B.1.1. The approximation holds well for most node pairs given the distribution (4.13).

The connection probability distribution in the RHG model is the Fermi-Dirac distribution

$$F_{\mathbf{A}}\left(d\left(\mathbf{x}_i,\mathbf{x}_j\right)\right) = \frac{1}{1+e^{\frac{d(\mathbf{x}_i,\mathbf{x}_j)-R}{2T}}}, \tag{5.5}$$

such that $F_{\mathbf{A}}\left(d\left(\mathbf{x}_i,\mathbf{x}_j\right)\right) \in [0,1]$, where the temperature $T \in (0,1)$ is a parameter that determines the size of the effect that distance has on the connection probability. Comparing (5.5) to the traditional Fermi-Dirac formulation, the parameter T is the temperature, the radius R is the chemical potential μ, while the Boltzman constant $k_B = 2$ in this case.

In complementarity graphs in the hyperbolic disk, for each node i, both the points $\mathbf{X}_i$ and $\mathbf{Y}_i$ follow the distributions (4.13) and (4.14). The connection probability function in (5.5) is plugged into the complementarity connection probability function (5.2).

5.3. BASIC PROPERTIES OF CRHGS

5.3.1. DEGREE DISTRIBUTION

The probability density function $\Pr[D = k]$ of the degree of an arbitrary node in a CRHG is found by integrating the probability density function $\Pr[D = k | \mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y}]$ of the degree distribution of a node with known coordinates $\mathbf{x} = (r_x, \theta_x)$ and $\mathbf{y} = (r_y, \theta_y)$ over all possible coordinates, by the law of total probability [3],

$$\begin{aligned}\Pr[D = k] = \int_0^R dr_x \int_0^R dr_y \int_0^{2\pi} d\theta_y \int_0^{2\pi} d\theta_x \, f_{X_0}(r_x) f_{Y_0}(r_y) f_{X_1}(\theta_x) f_{Y_1}(\theta_y) \\ \times \Pr[D = k | \mathbf{X} = (r_x,\theta_x), \mathbf{Y} = (r_y,\theta_y)],\end{aligned} \tag{5.6}$$

following the framework of [96]. Already for RHGs with a single coordinate per node, there exists no straightforward closed-form solution for the probability density function $\Pr[D = k]$ of the degree distribution. Also for CRHGs, we have to employ several approximations to obtain an expression for the probability density function $\Pr[D = k]$.

First, because the distribution of the polar coordinates in $\mathbf{X}$ and $\mathbf{Y}$ is uniform, the probability density of the degree distribution of a node does not depend on its angular coordinates. Therefore, we set the polar coordinates $\theta_x = 0$ and $\theta_y = 0$ for the convenience of calculation,

$$\Pr[D = k \,|\, \mathbf{X} = (r_x, \theta_x), \mathbf{Y} = (r_y, \theta_y)] = \Pr[D = k \,|\, X_0 = r_x, Y_0 = r_y], \tag{5.7}$$

where we have omitted $X_1 = 0$ and $Y_1 = 0$ from (5.7) to shorten the notation.

Next, Boguñá and Pastor-Satorras [96] have shown for sparse graphs with hidden variables that the distribution of the degree D for a node with known coordinates $\mathbf{x}$ and $\mathbf{y}$ is closely approximated by the Poisson distribution with rate $\lambda = E[D \,|\, \mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y}]$, for known $\mathbf{x}$ and $\mathbf{y}$. Hence, the corresponding Poisson approximation of the probability density function of the degree of a node with known radial coordinates r_x and r_y is

$$\Pr[D = k \,|\, X_0 = r_x, Y_0 = r_y] \approx \frac{1}{k!} \left[\lambda(r_x, r_y)\right]^k e^{-\lambda(r_x, r_y)}, \tag{5.8}$$

where $\lambda(r_x, r_y) = E[D \,|\, X_0 = r_x, Y_0 = r_y]$. We first derive approximations of $E[D \,|\, X_0 = r_x, Y_0 = r_y]$ and $E[D]$ to show that CRHGs are sparse, such that the Poisson approximation of the degree (5.8) is justified.

We approximate $E[D \,|\, X_0 = r_x, Y_0 = r_y]$ and plug the approximation into (5.8) in order to obtain the corresponding approximation of $\Pr[D = k]$ in (5.6). The connection probability p_{ij} is a function of the coordinates $(\mathbf{x}_i, \mathbf{y}_i)$ and $(\mathbf{x}_j, \mathbf{y}_j)$,

$$p_{ij} = p(\mathbf{x}_i, \mathbf{y}_i; \mathbf{x}_j, \mathbf{y}_j) = p(\theta_{i,x}, r_{i,x}, \theta_{i,y}, r_{i,y}; \theta_{j,x}, r_{j,x}, \theta_{j,y}, r_{j,y}). \tag{5.9}$$

We find $E[D \,|\, X_0 = r_x, Y_0 = r_y] = E[D_i \,|\, X_{1,i} = r_{i,x}, Y_{1,i} = r_{i,y}]$ by integrating the connection probability (5.9) given the coordinates of a node i over all possible coordinates for nodes $j \neq i$, while setting $\theta_{i,x} = \theta_{i,y} = 0$,

$$\begin{aligned} &E[D_i \,|\, X_{0,i} = r_{i,x}, Y_{0,i} = r_{i,y}] \\ &\quad = (N-1) \left[\int_0^R dr_{j,x} f_{X_0}(r_{j,x}) \int_0^R dr_{j,y} f_{Y_0}(r_{j,y}) \int_0^{2\pi} d\theta_{j,x} f_{X_1}(\theta_{j,x}) \right. \\ &\qquad\qquad \left. \times \int_0^{2\pi} d\theta_{j,y} f_{Y_1}(\theta_{j,y}) \, p(r_{i,x}, r_{i,y}; \theta_{j,x}, r_{j,x}, \theta_{j,y}, r_{j,y}) \right]. \end{aligned} \tag{5.10}$$

Plugging in the connection probability p_{ij} (5.2), we have

$$\begin{aligned} &E[D_i \,|\, X_{0,i} = r_{i,x}, Y_{0,i} = r_{i,y}] \\ &\quad = (N-1) \left[\int_0^R dr_{j,x} f_{X_0}(r_{j,x}) \int_0^R dr_{j,y} f_{Y_0}(r_{j,y}) \int_0^{2\pi} d\theta_{j,x} f_{X_1}(\theta_{j,x}) \right. \\ &\qquad \times \int_0^{2\pi} d\theta_{j,y} f_{Y_1}(\theta_{j,y}) \left[F_{\mathbf{A}}\left(d\left(r_{i,x}; \theta_{j,y}, r_{j,y}\right)\right) + F_{\mathbf{A}}\left(d\left(r_{i,y}; \theta_{j,x}, r_{j,x}\right)\right) \right. \\ &\qquad\qquad \left. \left. - F_{\mathbf{A}}\left(d\left(r_{i,x}; \theta_{j,y}, r_{j,y}\right)\right) F_{\mathbf{A}}\left(d\left(r_{i,y}; \theta_{j,x}, r_{j,x}\right)\right) \right] \right]. \end{aligned} \tag{5.11}$$

For most nodes, the radial coordinates of X_0 are very close to the maximum value R as defined by the distribution (4.13). Given that the angular coordinates in X_1 are uniformly distributed, the leading-order terms of the hyperbolic distance (5.3) given in the approximation (B.6) will therefore be close to $2R$ for the majority of the node pairs (i,j). The term $e^{\frac{d(\mathbf{x}_i,\mathbf{x}_j)-R}{2T}}$ in the Fermi-Dirac probability (5.5) is then exponentially large, such that the connection probability is exponentially small. For those node pairs, the product $F_{\mathbf{A}}\left(d\left(r_{i,x};\theta_{j,y},r_{j,y}\right)\right)F_{\mathbf{A}}\left(d\left(r_{i,y};\theta_{j,x},r_{j,x}\right)\right)$ is extremely small and contributes very little to the average degree in (5.11). We approximate the average degree of a node with known coordinates by omitting the product inside the multiple integral:

$$\begin{aligned}
&E[D_i \mid X_{0,i}=r_{i,x}, Y_{0,i}=r_{i,y}] \\
&\quad\approx (N-1)\left[\int_0^R dr_{j,x} f_{X_0}(r_{j,x})\int_0^R dr_{j,y} f_{Y_0}(r_{j,y})\int_0^{2\pi} d\theta_{j,x} f_{X_1}(\theta_{j,x})\right. \\
&\qquad\left.\times\int_0^{2\pi} d\theta_{j,y} f_{Y_1}(\theta_{j,y})\left[F_{\mathbf{A}}\left(d\left(r_{i,x};\theta_{j,y},r_{j,y}\right)\right)+F_{\mathbf{A}}\left(d\left(r_{i,y};\theta_{j,x},r_{j,x}\right)\right)\right]\right].
\end{aligned} \tag{5.12}$$

By rearranging terms, the multiple integral splits into two independent double integrals:

$$\begin{aligned}
&E[D_i \mid X_{0,i}=r_{i,x}, Y_{0,i}=r_{i,y}] \\
&\quad\approx (N-1)\int_0^R dr_{j,y} f_{Y_0}(r_{j,y})\int_0^{2\pi} d\theta_{j,y} f_{Y_1}(\theta_{j,y}) F_{\mathbf{A}}\left(d\left(r_{i,x};\theta_{j,y},r_{j,y}\right)\right) \\
&\qquad+(N-1)\int_0^R dr_{j,x} f_{X_0}(r_{j,x})\int_0^{2\pi} d\theta_{j,x} f_{X_1}(\theta_{j,x}) F_{\mathbf{A}}\left(d\left(r_{i,y};\theta_{j,x},r_{j,x}\right)\right).
\end{aligned} \tag{5.13}$$

The two remaining double integrals are each the average degree in the RHG model for a node with known coordinates r_x and r_y, respectively:

$$E[D \mid X_0=r_x, Y_0=r_y] \approx E[D^{\mathrm{RHG}} \mid X_0=r_x]+E[D^{\mathrm{RHG}} \mid X_0=r_y]. \tag{5.14}$$

Plugging in the approximation of $E[D^{\mathrm{RHG}} \mid X_0=r]$,

$$E[D \mid X_0=r_x, Y_0=r_y] \approx 2(N-1)\frac{T}{\sin(\pi T)} e^{-R/2} E\left[e^{-X_0/2}\right]\left(e^{-r_x/2}+e^{-r_y/2}\right), \tag{5.15}$$

as derived in Appendix B.1.2, where the expectation

$$E\left[e^{-X_0/2}\right]=\frac{\alpha}{\cosh(\alpha R)-1}\left[\frac{1}{2\alpha-1}\left(e^{(\alpha-\frac{1}{2})R}-1\right)+\frac{1}{2\alpha+1}\left(e^{-(\alpha+\frac{1}{2})R}-1\right)\right]. \tag{5.16}$$

The expected degree $E[D]$ of an arbitrary node is found by integrating (5.14) over the radial coordinates r_x and r_y,

$$E[D]=\int_0^R\int_0^R f_{X_0}(r_x) f_{Y_0}(r_y) E[D \mid X_0=r_x, Y_0=r_y]\, dr_y\, dr_x, \tag{5.17}$$

by the law of total expectation [3]. By plugging (5.14) into (5.17), we find the average degree $E[D]$ of an arbitrary node:

$$E[D]\approx\int_0^R\int_0^R f_{X_0}(r_x) f_{Y_0}(r_y)\left(E[D^{\mathrm{RHG}} \mid X_0=r_x]+E[D^{\mathrm{RHG}} \mid X_0=r_y]\right) dr_y\, dr_x. \tag{5.18}$$

5

Again by rearranging terms, the integral splits into two independent integrals

$$E[D] \approx \int_0^R f_{X_0}(r_x) E[D^{\mathrm{RHG}} \mid X_0 = r_x]\, dr_x + \int_0^R f_{Y_0}(r_y) E[D^{\mathrm{RHG}} \mid Y_0 = r_y]\, dr_y \tag{5.19}$$

$$= 2\int_0^R f_{X_0}(r) E[D^{\mathrm{RHG}} \mid X_0 = r]\, dr, \tag{5.20}$$

twice the average degree of an arbitrary node in the RHG model:

$$E[D] \approx 2E[D^{\mathrm{RHG}}]. \tag{5.21}$$

Plugging in the approximation of $E[D^{\mathrm{RHG}}]$,

$$E[D] \approx (N-1)\frac{4T}{\sin(\pi T)} e^{R/2} \left(E\left[e^{-X_0/2}\right]\right)^2, \tag{5.22}$$

as derived in Appendix B.1.2 and with $E\left[e^{-X_0/2}\right]$ given by (5.16). Inserting the leading-order approximation of $E\left[e^{-X_0/2}\right]$ given by (B.28), under the assumption that $R \gg 1$, yields

$$E[D] \approx (N-1)\left(\frac{2\alpha}{2\alpha-1}\right)^2 \frac{4T}{\sin(\pi T)} e^{-R/2}. \tag{5.23}$$

Equation (5.23) reveals that we can make CRHGs sparse if we set $R = 2\ln(N-1) + b \approx 2\ln(N) + b$, for any constant b independent of N, which justifies the Poisson approximation in equation (5.8). Following [21], we set

$$R = 2\ln\left(\frac{N}{\nu}\right), \tag{5.24}$$

for some scaling constant $\nu > 0$ that controls the resulting average degree $E[D]$. By the definition (5.24), we indeed have that the radius $R \to \infty$ when $N \to \infty$. Substituting the radius R (5.24) into the expression of the average degree $E[D]$ in (5.23), we obtain a final leading-order approximation,

$$E[D] \approx 4\nu\left(\frac{2\alpha}{2\alpha-1}\right)^2 \frac{T}{\sin(\pi T)}. \tag{5.25}$$

The corresponding approximation for $E[D \mid X_0 = r_x, Y_0 = r_y]$ is

$$E[D \mid X_0 = r_x, Y_0 = r_y] \approx 2\nu\left(\frac{2\alpha}{2\alpha-1}\right) \frac{T}{\sin(\pi T)} e^{R/2}\left(e^{-r_x/2} + e^{-r_y/2}\right). \tag{5.26}$$

The equation for the average degree $E[D]$ in (5.25) also allows for generating CRHGs with desired average degree $E[D]$ and power-law exponent γ in the limit of $N \to \infty$. For given average degree $E[D]$, power-law exponent γ and temperature T, the parameter $\alpha = \frac{1}{2}(\gamma - 1)$, while equation (5.25) should be solved for ν, after which the radius R is set according to (5.24).

Since $E[D \mid X_0, Y_0]$ is the sum of two Poisson-distributed random variables with identical rates λ, we find that the conditional distribution of the degree $D \mid X_0, Y_0$ is also Poisson distributed with the rate $2E[D \mid X_0 = r]$, two times the expected degree of a node

with known radial coordinates in the regular hyperbolic graph. We find the probability density $\Pr[D=k]$ of the degree k by plugging $\lambda = 2E[D^{\text{RHG}} \mid X_0 = r]$ into (5.8):

$$\Pr[D=k] \approx \frac{1}{k!}\int_0^R f_{X_0}(r)\left[2ce^{-r/2}\right]^k e^{-2ce^{-r/2}}\,dr, \tag{5.27}$$

for a constant $c \equiv 2\nu\left(\frac{2\alpha}{2\alpha-1}\right)\frac{T}{\sin(\pi T)}e^{R/2}$. Following the same reasoning as for regular hyperbolic graphs in Appendix B.1.3, we find the approximation of the degree distribution density

$$\Pr[D=k] \approx (2\alpha)\,c^{2\alpha}\frac{\Gamma[k-2\alpha,c]}{\Gamma[k+1]} \sim k^{-\gamma}, \tag{5.28}$$

which is approximately a power law with negative slope $\gamma = 2\alpha+1$ for large values of k, identical to regular hyperbolic graphs.

Figure 5.2 shows that our approximations are close to the simulated values for both the average degree $E[D]$ and density of the degree distribution $\Pr[D=k]$. The convergence of the leading-order approximation of the average degree $E[D]$ to the target average degree is slower than for regular RHGs and not yet fully realized in Figure 5.2(a), likely originating from dropping the product term in the connection probability. The approximation for $\Pr[D=k]$ closely resembles the simulated values, already for a single realization of the random graph, but we do observe a fat tail, which is characteristic for power-law distributions. Grouping the observations in bins of logarithmically increasing size could suppress the noise observed in the tail.

5.4. PROPERTIES COMPLEMENTARITY VS. SIMILARITY

We compare complementarity and similarity hyperbolic graphs across several graph metrics. The results obtained in this section are computed numerically. For complementarity graphs, we study the case where both radial coordinates are set to the same value, $X_0 = Y_0 = r$, in order to direct our focus on the complementarity aspect and not on the slow convergence of the variables X_0 and Y_0.

5.4.1. CLUSTERING COEFFICIENT

As the node clustering coefficient $c_G(i)$ (3.25) measures the local density of connections around node i, the graph clustering coefficient c_G is also called the *average local* clustering coefficient. An alternative definition of the clustering coefficient is given by Luce and Duncan [164], while popularized by Newman [165]. They define the clustering coefficient $\check{c}_G$ as 3 times the total number t of undirected triangles in G divided by the number of connected triples,

$$\check{c}_G = \frac{3t}{\sum_{i=1}^N \frac{k_i(k_i-1)}{2}} = \frac{6t}{\sum_{i=1}^N k_i(k_i-1)}. \tag{5.29}$$

The clustering coefficient $\check{c}_G$ is also called the transitivity coefficient, or *global* clustering coefficient. We evaluate both clustering coefficients for CRHGs and RHGs and compare them to the coefficients after degree-preserving rewiring. The clustering coefficients after degree-preserving rewiring serve as a measure of the clustering we could expect solely based on the degrees of the nodes.

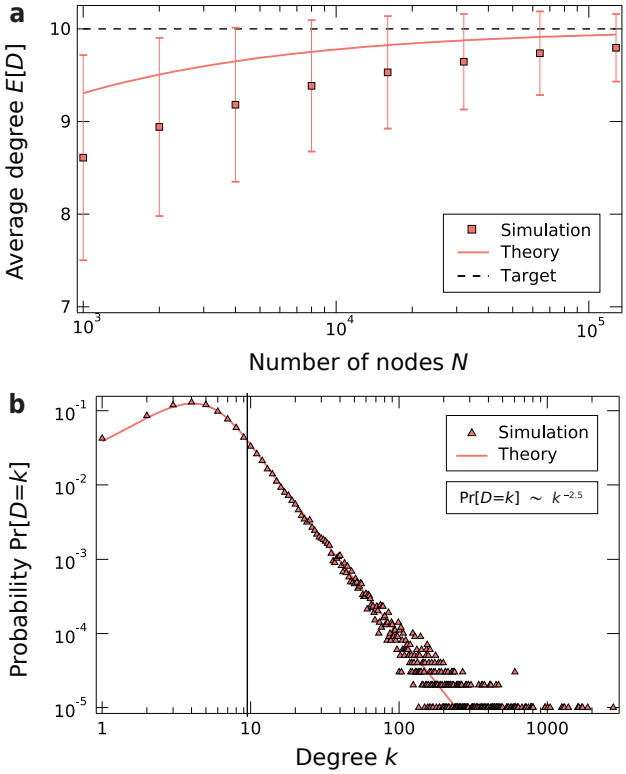


Figure 5.2: Average degree $E[D]$ and density $\Pr[D = k]$ of the degree distribution. The average degree $E[D]$ (a) and the density $\Pr[D = k]$ of the degree distribution (b) of CRHGs in the case where $X_0 = Y_0 = r$. Panel (a) shows the observed average degree $E[D]$ for a target average degree $E[D] = 10$ according to (5.25), with $N = \{10^3 \cdot 2^0, 10^3 \cdot 2^1, \ldots, 10^3 \cdot 2^7\}$, the parameter $\alpha = 0.75$ ($\gamma = 2.5$) and the temperature $T = 0.5$. Each simulated value is the average of 10^4 realizations, while the error bars display the standard deviation. The theory line is given by equation (5.22). Panel (b) shows the observed density $\Pr[D = k]$ of the degree distribution for $\alpha = 0.75$ (target $\gamma = 2.5$), where the number of nodes $N = 10^5$, the temperature $T = 0.5$ and the target expected degree $E[D] = 10$. The probabilities $\Pr[D = k]$ are computed from a single realization of the random graph, therefore the realized average degree $E[D] = 9.42$ deviates slightly from the target value. The negative slope $\gamma = 2.5$ is estimated based on the linear part of the tail in the binned density with 20 bins of equal logarithmic width.

Figure 5.3 shows the clustering coefficients c_G and $\check{c}_G$ of CRHGs and RHGs as a function of the temperature T, also after degree-preserving rewiring. We observe that CRHGs are just as clustered locally as could be expected based on the degrees. Globally, CRHGs are slightly more clustered than expected. In contrast, RHGs are significantly more clustered than expected based on the degrees, both locally and globally. Both clustering coefficients are a decreasing function of the temperature T for RHGs, as was known already from [21]. For CRHGs, the relation between the clustering coefficients and the temperature T is almost completely flat. Interestingly, CRHGs are exactly just as clustered locally as RHGs are after rewiring.

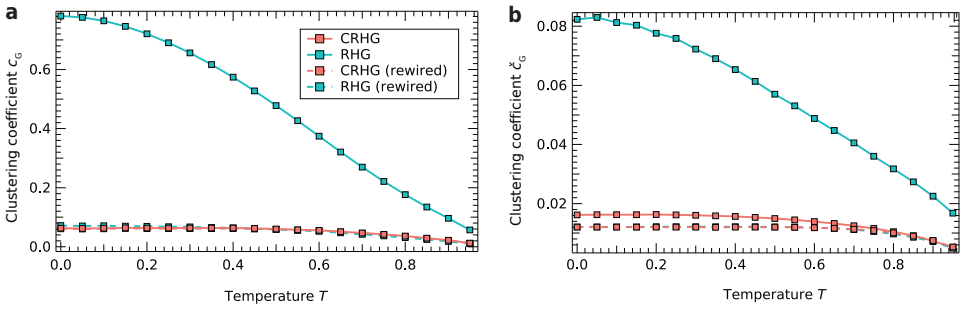


Figure 5.3: Clustering coefficients c_G and $\check{c}_G$ as a function of temperature T. Average local clustering coefficient c_G (a) and global clustering coefficient $\check{c}_G$ (b) before and after degree-preserving rewiring of CRHGs and RHGs as a function of temperature T. The number of nodes $N = 10^4$, the expected degree $E[D] = 10$ and the parameter $\alpha = 0.75$ ($\gamma = 2.5$). The lines for the rewired networks and the CRHG line overlap in panel (a), while in panel (b) only the lines for the rewired networks overlap. Each point is the average of 10^4 simulations. For the rewiring, we select two links at random and randomly connect two nodes not connected before, in total for $4 \times 2L$ link pairs (drawn with repetition).

5

5.4.2. BIPARTITE CLUSTERING COEFFICIENT

Whereas the clustering coefficient c_G measures the average local density of triangles in a graph, the clustering coefficient c_G^4 for bipartite networks from Zhang *et al.* [166] analogously measures the average local density of 4-cycles or quadrangles. As first remarked by Kovács *et al.* [128], triangles are not likely to occur in a complementarity-driven graph, as it is unlikely that in a triplet of nodes each node is simultaneously complementary to both of the other two nodes. Instead, one might expect to observe quadrangles, similar to the example in Figure 5.1(b). The bipartite clustering coefficient $c_G^4(i)$ of a node i is then defined as the fraction of closed cycles of length 4 containing node i to the total number of node quadruples that might form a 4-cycle with i. Let $\mathcal{N}(i)$ be the set of neighbors of node i, such that a node j is an element of $\mathcal{N}(i)$ only if $a_{ij} = 1$. The bipartite clustering coefficient $c_G^4(i)$ is then computed as

$$c_G^4(i) = \frac{\sum_{j=1}^{N} \sum_{l=1, l\neq j}^{N} a_{ij} a_{il} |\mathcal{N}(j) \cap \mathcal{N}(l)|}{\sum_{j=1}^{N} \sum_{l=1, l\neq j}^{N} a_{ij} a_{il} |\mathcal{N}(j) \cup \mathcal{N}(l)|}, \tag{5.30}$$

which is the fraction of neighbors of i that have a common neighbor other than i. The graph bipartite clustering coefficient c_G^4 is the average of $c_G^4(i)$ over all nodes,

$$c_G^4 = \frac{1}{N} \sum_{i=1}^{N} c_G^4(i). \tag{5.31}$$

The coefficient $c_G^4 \in [0, 1]$ and the maximum value of 1 is attained for the complete bipartite graph, which is a bipartite graph where each node is connected to all other nodes in the other of the two sets of nodes.

Figure 5.4 shows the bipartite clustering coefficient c_G^4 for CRHGs and RHGs as a function of the temperature T, also after degree-preserving rewiring. We observe that CRHGs and RHGs both contain on average more quadrangles than could be expected

based on the degrees. The density of quadrangles in similarity graphs is significantly higher than in CRHGs, explained by the fact that any two triangles together also form a quadrangle and that similarity graphs contain many triangles, as we saw in Figure 5.1. Even though we observe a clear difference between CRHGs and RHGs in terms of the bipartite clustering coefficient c_G^4 in Figure 5.4, the difference is only in the magnitude of the coefficient, and leveraging c_G^4 for classifying graphs as either complementarity- or similarity-driven remains difficult.

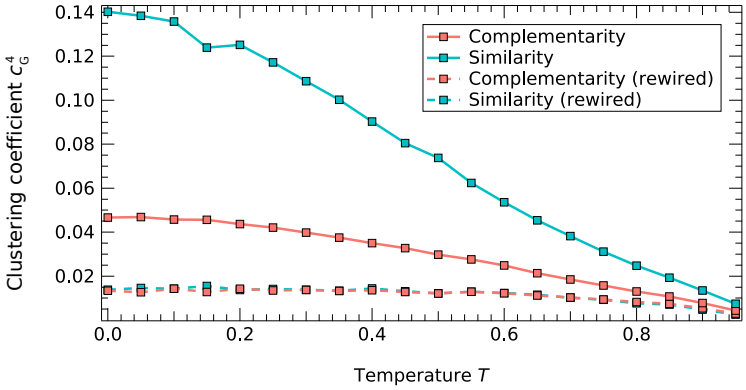


Figure 5.4: Bipartite clustering coefficient c_G^4 as a function of temperature T. The bipartite clustering coefficient c_G^4 of CRHGs and RHGs as a function of temperature T. For all graphs, the number of nodes $N = 10^4$, the expected degree $E[D] = 10$, and the power-law exponent in the degree distribution $\gamma = 2.5$. Each point is the average of $5 \cdot 10^3$ simulations.

5.4.3. SIMILARITY AND COMPLEMENTARITY COEFFICIENTS

In a recent work, Talaga and Nowak [129] propose a rigorous framework for measuring the density of triangles and quadrangles in a graph relative to the expected density based on the node degrees. They note that the clustering coefficient only considers connected triples where the focal node i is located in the middle position, while one should also count triples where the node i is located at an outer end of the path. In addition, they argue that true complementarity results in a strong quadrangle, which is a quadrangle where the chord or diagonal links are not present, a ring graph with $N = 4$ nodes, while the bipartite clustering coefficient c_G^4 considers all quadrangles if the graph is not strictly bipartite. They propose measures for the densities of triangles and strong quadrangles that include all 2-hop and 3-hop paths, also the paths where node i is located at an outer end of those paths.

The local clustering coefficient $c_G(i)$ in (3.25) is the fraction of triangles in G that include node i to the number of connected triples of nodes with node i located in the middle position. Talaga and Nowak [129] name these wedge triples and the total number of wedge triples $t_G^W(i) = k_i(k_i - 1)$. For distinction, we denote the similarity coefficient of

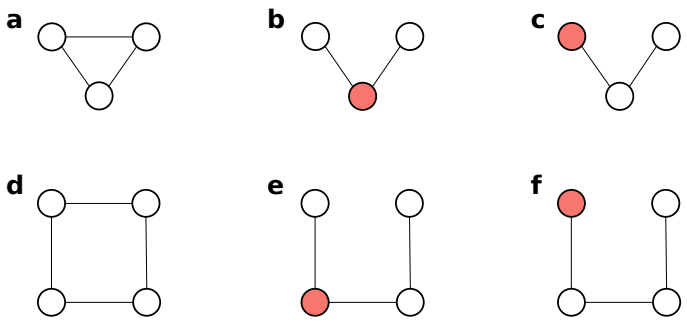


Figure 5.5: Patterns of node triples and quadruples considered for computation of the complementarity and similarity coefficients. Triangle closure, panel (a), and strong quadrangle closure, panel (d). A wedge triple (b) has the focal node i in the middle position and a head triple (c) has the focal node i in the first position. A wedge quadruple (e) has the focal node i in the second position and a head quadruple (f) has the focal node i in the first position.

node i based on wedge triples $\tilde{s}_G^W(i) \equiv c_G(i)$, such that

$$\tilde{s}_G^W(i) = \frac{2y}{t_G^W(i)} = \frac{2y}{k_i(k_i-1)}, \tag{5.32}$$

with the superscript W referring to 'wedge'. Analogously, the number of head triples, connected triples where node i is located at an outer end, $t_G^H(i) = \sum_{j=1}^N a_{ij}(k_j-1)$. The similarity coefficient of node i based on head triples

$$\tilde{s}_G^H(i) = \frac{2y}{t_G^H(i)} = \frac{2y}{\sum_{j=1}^N a_{ij}(k_j-1)}, \tag{5.33}$$

where y is again the number of triangles containing node i and with the superscript H referring to 'head'. The coefficient $\tilde{s}_G^H(i) \in [0,1]$. The structural similarity coefficient $\tilde{s}_G(i)$ of node i is then the average of $\tilde{s}_G^W(i)$ and $\tilde{s}_G^H(i)$ weighted by the corresponding numbers of triples,

$$\tilde{s}_G(i) = \frac{t_G^W(i)\tilde{s}_G^W(i) + t_G^H(i)\tilde{s}_G^H(i)}{t_G^W(i) + t_G^H(i)}. \tag{5.34}$$

The value of $\tilde{s}_G(i)$ is bounded between those of $\tilde{s}_G^W(i)$ and $\tilde{s}_G^H(i)$. The graph similarity coefficient $\tilde{s}_G$ is the average of $\tilde{s}_G^W(i)$ over all nodes,

$$\tilde{s}_G = \frac{1}{N}\sum_{i=1}^N \tilde{s}_G(i). \tag{5.35}$$

The structural complementarity coefficient $\tilde{c}_G(i)$ of node i is computed in a similar way as the structural similarity coefficient in (5.32), (5.33) and (5.34), but based on the fraction of strong quadrangles to the number of connected quadruples of nodes. We compute the wedge and head quadrangular clustering coefficients, which measure the density of strong quadrangles containing node i to the total numbers of wedge and head

quadruples. A wedge quadruple for node i is a 3-hop path with node i in the second position and the number of such quadruples is denoted by $q_G^W(i)$. The quadrangular clustering coefficient of node i based on wedge quadruples with node i in the second position

$$\tilde{c}_G^W(i) = \frac{2z}{q_G^W(i)} = \frac{2z}{\sum_{j=1}^N a_{ij}\left[(k_i-1)(k_j-1) - |\mathcal{N}(i)\cap\mathcal{N}(j)|\right]}, \tag{5.36}$$

where z is the number of strong quadrangles containing node i. The quantity $|\mathcal{N}(i)\cap\mathcal{N}(j)|$ is the number of common neighbors between nodes i and j. A head quadruple is a connected quadruple of nodes with node i in the first position and we name $q_G^H(i)$ the total number of such quadruples. The complementarity coefficient based on head quadruples

$$\tilde{c}_G^H(i) = \frac{2z}{q_G^H(i)} = \frac{2z}{\sum_{j\neq i} a_{ij}\sum_{h\neq i,j} a_{jk}(k_h - 1 - a_{ih})}, \tag{5.37}$$

where z is again the number of strong quadrangles containing node i. The structural complementarity coefficient $\tilde{c}_G(i)$ of node i is then the weighted average of $\tilde{c}_G^W(i)$ and $\tilde{c}_G^H(i)$,

$$\tilde{c}_G(i) = \frac{q_G^W(i)\tilde{c}_G^W(i) + q_G^H(i)\tilde{c}_G^H(i)}{q_G^W(i) + q_G^H(i)}. \tag{5.38}$$

The graph complementarity coefficient $\tilde{c}_G$ is the average of $\tilde{c}_G^W(i)$ over all nodes,

$$\tilde{c}_G = \frac{1}{N}\sum_{i=1}^N \tilde{c}_G(i). \tag{5.39}$$

The absolute quantities of $\tilde{s}_G$ and $\tilde{c}_G$ are incomparable and neither are the values $\tilde{c}_G$ and $\tilde{c}_H$ for two different graphs G and H. The coefficients should be standardized, or calibrated, by comparing the coefficients of a graph G to B different sampled random graphs with the same number of nodes and approximately the same degrees as G has [129]. We sample B graphs using the configuration model with the degrees of G as the expected degrees and compute a calibrated score $\hat{x}_G$ from each coefficient x_G by computing the average log ratio with the coefficients x_{G_b} of the $b = 1, \ldots, B$ sampled random graphs G_b,

$$\hat{x}_G = \frac{1}{B}\sum_{b=1}^B \log\left(\frac{x_G}{x_{G_b}}\right). \tag{5.40}$$

A value of the calibrated score $\hat{x}_G$ of 0 indicates that the coefficient x_G is equal to the value that could be expected based on the degrees in the graph G.

Figure 5.6 shows the calibrated similarity and complementarity scores for CRHGs and RHGs. The similarity score perfectly separates similarity and CRHGs for all values of T, similar to what we saw in Figure 5.3(a). RHGs contain significantly more triangles than expected, whereas complementary graphs contain a number of triangles that is approximately what would be expected by chance. The complementarity score separates the two graph types for temperature values $T < 0.5$, which is when connections are formed over smaller distances in the latent space. CRHGs contain more strong quadrangles than expected, while RHGs contain fewer such quadrangles than expected. For $T \geq 0.5$, the two graph types both contain more quadrangles than expected and no conclusion can be drawn based on the complementarity score.

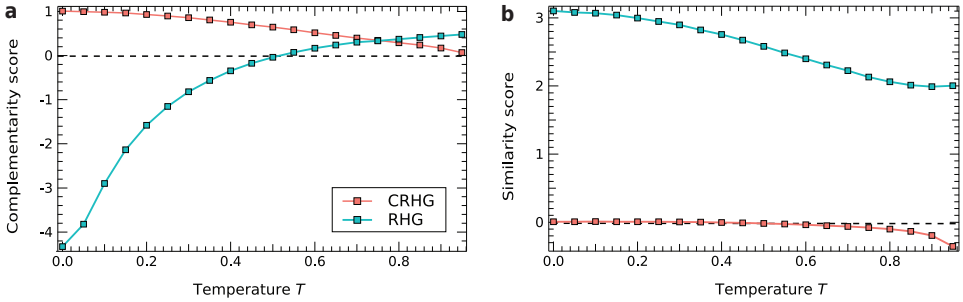


Figure 5.6: Calibrated complementarity and similarity scores as a function of temperature T. Calibrated complementarity (a) and similarity (b) scores for CRHGs and RHGs as a function of the temperature T. The number of nodes $N = 5 \cdot 10^3$, the expected degree $E[D] = 10$ and the parameter $\alpha = 0.75$ ($\gamma = 2.5$). The number of sampled random graphs for calibration $B = 100$. Each point is the average of 10^4 simulations.

5.4.4. EFFECTIVE GRAPH RESISTANCE

The effective graph resistance R_G is a fundamental graph metric that measures the difficulty of transport in a graph, defined as the sum of all pairwise effective resistances in the graph [167]. The lower the effective graph resistance R_G, the easier it is for a flow to move through the network. The effective resistance ω_{ij} for two nodes i and j takes into account all paths that exist between i and j, therefore Ellens *et al.* [167] argue that graphs with a lower effective graph resistance are more robust.

We compute the effective resistance matrix $\mathbf{\Omega}$ through the definition in terms of the pseudoinverse of the Laplacian $\mathbf{Q}^\dagger$. For a connected graph G, the pseudo-inverse of the Laplacian $\mathbf{Q}^\dagger$ is defined as

$$\mathbf{Q}^\dagger = \sum_{k=1}^{N-1} \frac{1}{\mu_k} \mathbf{z}_k \mathbf{z}_k^T, \tag{5.41}$$

where μ_k is the k-th eigenvalue of the Laplacian $\mathbf{Q}$ such that $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_{N-1} \geq \mu_N$ and $\mathbf{z}_k$ is the corresponding $N \times 1$ eigenvector. The pseudoinverse $\mathbf{Q}^\dagger$ gives

$$\mathbf{Q}\mathbf{Q}^\dagger = \mathbf{Q}^\dagger \mathbf{Q} = \mathbf{I} - \frac{1}{N}\mathbf{J}, \tag{5.42}$$

where $\mathbf{J}$ is the $N \times N$ all-one matrix. The $N \times N$ effective resistance matrix $\mathbf{\Omega}$ is then defined in terms of $\mathbf{Q}^\dagger$ as

$$\mathbf{\Omega} = \mathbf{u}\boldsymbol{\zeta}^T + \boldsymbol{\zeta}\mathbf{u}^T - 2\mathbf{Q}^\dagger, \tag{5.43}$$

where $\mathbf{u}$ is the $N \times 1$ all-one vector and $\boldsymbol{\zeta} = (q^\dagger_{11}, q^\dagger_{22}, \ldots, q^\dagger_{NN})^T$ the $N \times 1$ vector with the diagonal elements of $\mathbf{Q}^\dagger$. The element $(\mathbf{\Omega})_{ij} = \omega_{ij}$ is the effective resistance between nodes i and j. Finally, the effective graph resistance

$$R_G = \frac{1}{2}\mathbf{u}^T \mathbf{\Omega} \mathbf{u}. \tag{5.44}$$

Figure 5.7 shows the effective graph resistance R_G of CRHGs and RHGs as a function of the temperature T. CRHGs have a lower graph resistance than RHGs when they are matched by average degree. The lower effective graph resistance R_G suggests that CRHGs

are on average more robust than RHGs for a given value of T. For both types of graphs, the effective graph resistance R_G is fast increasing when the temperature T approaches 1.

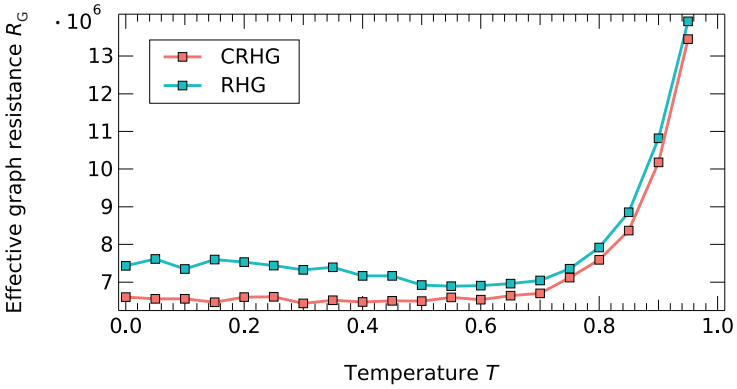


Figure 5.7: Effective graph resistance R_G as a function of temperature T. The effective graph resistance R_G of CRHGs and RHGs as a function of temperature T. For all graphs, the number of nodes $N = 5 \cdot 10^3$, the expected degree $E[D] = 10$ and the power-law exponent in the degree distribution $\gamma = 2.5$. Each point is the average of 100 simulations.

5

5.4.5. PAIRWISE METRICS

We inspect the connection probability $\Pr[a_{ij} = 1]$ of CRHGs and RHGs as a function of several pairwise metrics. Metric 1 is the effective resistance ω_{ij}, the element (i, j) of the effective resistance matrix $\mathbf{\Omega}$ defined in (5.43). Metrics 2-6 aim to quantify the similarity of the neighbor sets of two nodes and were shown to be good predictors of missing links in real similarity-driven networks. For metrics 2-6, we follow the definitions in [29]. For metric 7, the preferential attachment index, we follow the definition in [168] and we follow the definition in [128] for metric 8, the L_3 index. Let $\mathcal{N}(i)$ again be the set of neighbors of node i. We consider the following pairwise metrics:

1. Effective resistance $\omega_{ij} = (\mathbf{\Omega})_{ij}$,
2. Jaccard similarity neighbor sets $J(i, j) = \dfrac{|\mathcal{N}(i) \cap \mathcal{N}(j)|}{|\mathcal{N}(i) \cup \mathcal{N}(j)|}$,
3. Number of common neighbors $CN(i, j) = |\mathcal{N}(i) \cap \mathcal{N}(j)|$,
4. Adamic-Adar index $AA(i, j) = \sum_{h \in \mathcal{N}(i) \cap \mathcal{N}(j)} \frac{1}{\ln k_h}$,
5. Resource Allocation index $RA(i, j) = \sum_{h \in \mathcal{N}(i) \cap \mathcal{N}(j)} \frac{1}{k_h}$,
6. Cannistraci Resource Allocation index $CRA(i, j) = \sum_{h \in \mathcal{N}(i) \cap \mathcal{N}(j)} \frac{|\mathcal{N}(i) \cap \mathcal{N}(j) \cap \mathcal{N}(h)|}{k_h}$,
7. Preferential Attachment index $PA(i, j) = k_i \cdot k_j$,

8. Paths of length 3 index $L_3(i,j) = \sum_{h,l \in \mathcal{N}(i) \cap \mathcal{N}(j)} \frac{a_{hl}}{\sqrt{k_h k_l}}$.

Figure 5.11 shows the empirical connection probability as a function of each of the 8 pairwise metrics for CRHGs and RHGs separately. The largest difference between CRHGs and RHGs occurs for the effective resistance ω_{ij}. Whereas the connection probability in similarity graphs is a decreasing function of ω_{ij}, for CRHGs the line is almost entirely flat. This relation is explained when the graph is dominated by quadrangles rather than by triangles. We indeed saw that RHGs contain many triangles, while CRHGs contain many quadrangles. To illustrate the explanation for the flat relation between $\Pr[a_{ij} = 1]$ and ω_{ij}, consider a simple example: the effective resistances of two connected nodes and of two nodes that are not connected but that are located at opposite sides of a quadrangle are identical, as shown in Figure 5.8. If in both graphs (a) and (b) in Figure 5.8

Figure 5.8: Example of two connection patterns with equal effective resistance. Example of two nodes i and j that are (a) connected or (b) not connected and located at opposite sides of a quadrangle. If each link has a resistance R, in both cases the effective resistance $\omega_{ij} = R$. Thinking of the two graphs as electrical circuits, in graph (a) we simply have one link with resistance R. In graph (b), we have two series connections of length 2 in parallel, each with resistance $2R$, that together have effective resistance R.

we would add the same additional paths from i to j of at least length 3 (paths of smaller length would form triangles, which we do not have), the effective resistance ω_{ij} would not change, see Figure 5.9. Thus, if nodes i and j are part of a quadrangle, the effective resistance ω_{ij} is not able to explain the probability $\Pr[a_{ij} = 1]$ of the existence of a link between them.

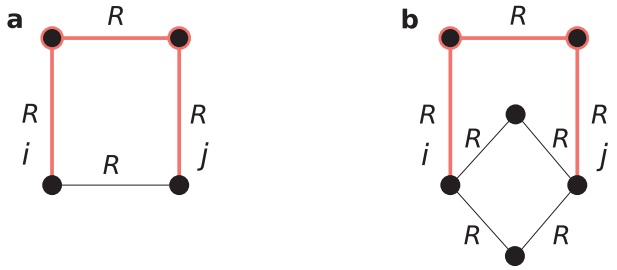


Figure 5.9: Example of two connection patterns with equal effective resistance. Again the example of Figure 5.8, but with an additional path from i to j of length 3 (highlighted in red). The effective resistances ω_{ij} are still identical between the graphs in (a) and (b), although ω_{ij} is no longer equal to R (in fact, now $\omega_{ij} = \frac{3}{4}R$ in both cases).

Second, the connection probability in CRHGs is a decreasing function of the Jaccard similarity of the neighbor sets of a node pair for CRHGs: the more neighbors two nodes

have in common, the less likely they are to connect. This relation is explained by the absence of triangles in complementarity graphs. For RHGs, the relation between the connection probability and the Jaccard similarity is not straightforward to classify but for higher values of the Jaccard similarity the function is increasing.

For the remaining pairwise metrics, the connection probability is an increasing function of the metric for both CRHGs and RHGs. Except for the effective resistance and the Jaccard similarity, the pairwise metrics considered here cannot distinguish between CRHGs and RHGs.

5.4.6. DEGREE ASSORTATIVITY

The concept of complementarity, defined in the latent metric space of a network and therefore not directly observed, could materialize as a disassortativity between observable node properties. We inspect the degree assortativity ρ_D, which measures the tendency of nodes to connect preferentially to nodes with similar degree [161]. Newman [169] defines the degree assortativity ρ_D as the Pearson correlation coefficient between the degrees of connected nodes. Van Mieghem *et al.* [145] find the following graph-theoretic expression of the degree assortativity

$$\rho_D = 1 - \frac{\sum_{i\sim j}\left(k_i - k_j\right)^2}{\sum_{j=1}^{N} k_j^3 - \frac{1}{2L}\left(\sum_{j=1}^{N} k_j^2\right)^2}, \tag{5.45}$$

where $i \sim j$ indicates that nodes i and j are direct neighbors.

Figure 5.10 shows the degree assortativity ρ_D of CRHGs and RHGs as a function of the temperature T. We first observe that CRHGs and RHGs have a degree assortativity ρ_D very close to 0. We also observe that CRHGs and RHGs do not differ through the degree assortativity ρ_D on average. The degree assortativity ρ_D is not a metric that distinguishes CRHGs and RHGs.

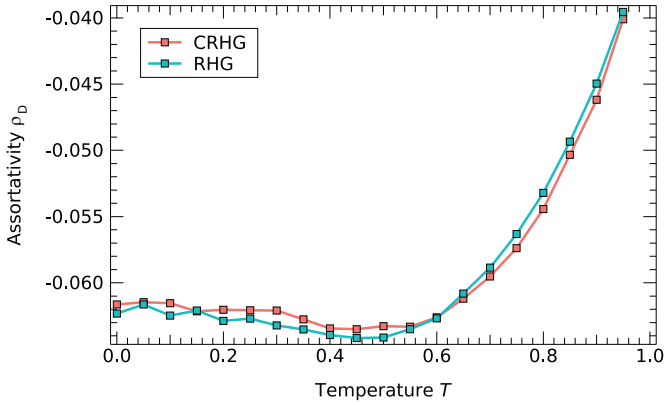


Figure 5.10: Degree assortativity ρ_D as a function of temperature T. The degree assortativity ρ_D of CRHGs and RHGs as a function of temperature T. For all graphs, the number of nodes $N = 10^4$, the expected degree $E[D] = 10$, and the power-law exponent in the degree distribution $\gamma = 2.5$. Each point is the average of 10^4 simulations.

5.5. CONCLUSION

We have studied the topological properties of the CRHG from Chapter 4. We are able to construct CRHGs that are sparse and we find that they have approximately a power-law degree distribution, like similarity-driven RHGs. We find that the topology of the CRHG is dominated by quadrangles, in contrast to the RHG, which is dominated by triangles. This finding agrees with earlier studies of real complementarity-driven networks. Because of the quadrangular patterns, we find that in CRHGs the effective resistance is not related to the probability of link existence. Overall, CRHGs have a lower effective graph resistance than regular hyperbolic graphs, which indicates that they might be more robust. Both CRHGs and RHGs show almost no degree assortativity. A good understanding of the topological properties of the CRHG model might help us design better statistical methods for real networks in which the link formation is driven by complementarity.

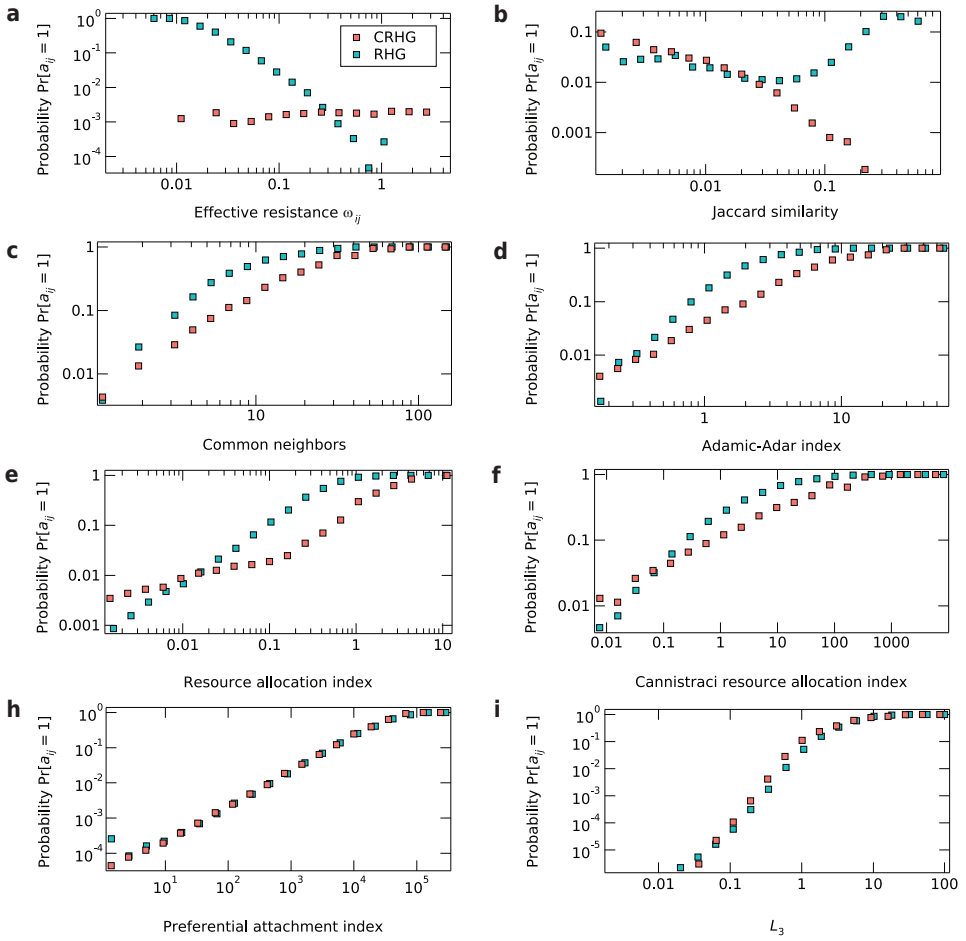


Figure 5.11: Connection probability Pr[$a_{ij} = 1$] as a function of different pairwise metrics. Histograms of the connection probability Pr[$a_{ij} = 1$] as a function of different pairwise metrics for CRHGs and RHGs. In each panel, the empirical connection probabilities and metrics are computed from a single simulated network per network type, with $N = 5 \cdot 10^3$, $E[D] = 10$, $T = 0.5$ and $\gamma = 2.5$. In a given histogram, each bin has equal logarithmic width.

III

SEMANTIC NETWORKS

6

TOPOLOGICAL PROPERTIES AND ORGANIZING PRINCIPLES OF SEMANTIC NETWORKS

We acquire languages in one way, and only one way: by understanding messages, not through explicit teaching aimed at conscious learning.

Stephen D. Krashen

Interpreting natural language is an increasingly important task in computer algorithms due to the growing availability of unstructured textual data. Natural Language Processing (NLP) applications rely on semantic networks for structured knowledge representation. We study the properties of semantic networks from ConceptNet, defined by 7 semantic relations from 11 different languages. We find that semantic networks have universal basic properties: they are sparse and highly clustered, and many exhibit power-law degree distributions. Our findings show that the majority of the considered networks are scale-free. Some networks exhibit language-specific properties determined by grammatical rules, for example, networks from highly inflected languages, such as e.g. Latin, German, French, and Spanish, show peaks in the degree distribution that deviate from a power law. Depending on the semantic relation type and the language, the link formation in semantic networks is guided by different connection principles. In some networks, the links are similarity-based, while in others, the links are more complementarity-based. Finally, we demonstrate how knowledge of similarity and complementarity in semantic networks can improve NLP algorithms in missing link inference.

This Chapter is based on G. Budel, Y. Jin, P. Van Mieghem and M. Kitsak (2023), Topological properties and organizing principles of semantic networks, Scientific Reports 13(1), 11728 [162].

6.1. INTRODUCTION

Due to the explosive increase in the availability of digital content, the demand for computers to efficiently handle textual data has never been greater. Large amounts of data and improved computing power have enabled a vast amount of research on Natural Language Processing (NLP). The goal of NLP is to allow computer programs to interpret and process unstructured text. In computers, text is represented as a string, while in reality, human language is much richer than just a string. People relate text to various concepts based on previously acquired knowledge. To effectively interpret the meaning of a text, a computer must have access to a considerable knowledge base related to the domain of the topic [46].

Semantic networks can represent human knowledge in computers, as first proposed by Quillian in the 1960s [47, 170]. 'Semantic' means 'relating to meaning in language or logic', and a semantic network is a graph representation of structured knowledge. Such networks are composed of nodes, which represent concepts (*e.g.*, words or phrases), and links, which represent semantic relations between the nodes [48, 171]. The links are tuples of the format *(source, semantic relation, destination)* that encode knowledge. For example, the information that a car has wheels is represented as *(car, has, wheels)*. Figure 6.1 shows a toy example of a semantic network as the subgraph with the neighborhood around the node *car*.

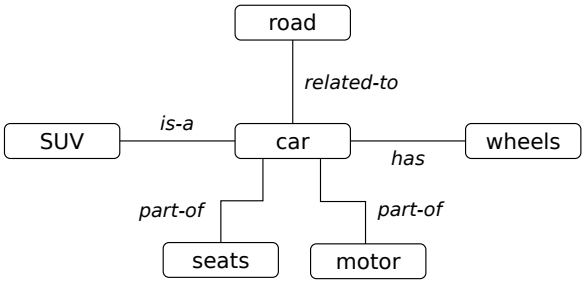


Figure 6.1: Toy example of a semantic network with six concepts and five semantic relations of four different types.

The past two decades have witnessed a rise in the importance of NLP applications [172–174]. For instance, Google introduced Google Knowledge Graph to enhance their search engine results [49]. A knowledge graph is a specific type of semantic network, in which the relation types are more explicit [175, 176]. Voice assistants and digital intelligence services, such as Apple Siri [52] and IBM Watson [53], use semantic networks as a knowledge base for retrieving information [177, 178]. As a result, machines can process information in raw text, comprehend unstructured user input, and achieve the goal of communicating with users, all up to a certain extent. Recently, OpenAI made a great leap forward in user-computer interaction with InstructGPT, better known to the general public as ChatGPT [55].

Language is a complex system with diverse grammatical rules. To grasp the meaning of a sentence, humans leverage their natural understanding of language and concepts in contexts. Language is still poorly understood from a computational perspective and,

hence, it is difficult for computers to utilize similar strategies. Namely, machines operate under unambiguous instructions that are strictly predefined and structured by humans. Though we can argue that human languages are structured by grammar, these grammatical rules often prove to be ambiguous [179]. After all, in computer languages, there are no synonyms, namesakes, or tones that can lead to misinterpretation [180]. Computers rely on external tools to enable the processing of the structure and meaning of texts.

In this paper, we conduct systematic analyses of the topological properties of semantic networks. Our work is motivated by the following purposes:

- Understand fundamental formation principles of semantic networks.
  In many social networks, the links between nodes are driven by similarity [181–184]. The more similar two nodes are in terms of common neighbors, the more likely they are connected. Thanks to the intensive study of similarity-based networks, many successful tools of data analysis and machine learning were developed, such as link prediction [185] and community detection [186]. These tools may not work well for semantic networks, because words in a sentence do not necessarily co-occur because of similarity. Sometimes, two words are used in conjunction because they have complementary features. Therefore, we study the principles that drive the formation of links in semantic networks.

- Document language-specific features.
  Languages vary greatly between cultures and across time [187]. Two languages that originate from two different language families can differ in many types of features since they are based on different rules. It is natural to conjecture that there exist diverse structures in semantic networks for different languages.

- Better inform NLP methods.
  Although there have been numerous real-world NLP applications across various domains, existing NLP technologies still have limitations [188]. For example, processing texts from a language where single words or phrases can convey more than one meaning is difficult for computers [189, 190]. Existing, successful algorithms built on top of semantic networks are usually domain-specific, and designing algorithms for broader applications remains an open problem. To design better language models that can handle challenges such as language ambiguity, we first need to gain a better understanding of the topological properties of semantic networks.

Previous studies on semantic networks focused on a few basic properties and relied on multiple datasets with mixed semantic relations, which we discuss in detail in Section 6.2. Therefore, it is difficult to compare the results within one study and between two different studies. To our knowledge, there has been no systematic and comprehensive analysis of the topological properties of semantic networks at the semantic relation level.

To sum up, the main objective of this paper is to understand the structure of semantic networks. Specifically, we first study the general topological properties of semantic networks from a single language with distinct semantic relation types. Second, we compare semantic networks with the same relation type between different languages to find

language-specific patterns. In addition, we investigate the roles of similarity and complementarity in the link formation principles in semantic networks.

The main contributions of this paper include:

1. We study the topological properties of seven English semantic networks, each network defined by a different semantic relation (*e.g.*, 'Is-A' and 'Has-A'). We show that all networks possess high sparsity and many possess a power-law degree distribution. In addition, we find that most networks have a high average clustering coefficient, while some networks show the opposite.

2. We extend the study of the topological properties of semantic networks to ten other languages. We find non-trivial structural patterns in networks from languages that have many grammatical inflections. Due to the natural structure of the grammar in these languages, words have many distinct inflected forms, which leads to peaks in the density of the degree distribution and results in deviations from a power law. We find this feature not only in inflecting languages but also in Finnish, which is classified as agglutinating.

3. We study the organizing principles of 50 semantic networks defined by different semantic relations in different languages. We quantify the structural similarity and complementarity of semantic networks by counting the relative densities of triangles and quadrangles in the graphs, following a recent work by Talaga and Nowak [129]. Hereby, we show to what extent these networks are similarity- or complementarity-based. We find that the connection principles in semantic networks are mostly related to the type of semantic relation, not the language origin.

This Chapter is organized as follows. In Section 6.2, we provide a brief overview of the previous work on the properties of semantic networks. In Section 6.3, we study the general topological properties of seven English semantic networks. We compare the properties of semantic networks between 11 different languages in Section 6.4, while Section 6.5 deals with the fundamental connection principles in semantic networks. We measure and compare the structural similarity and complementarity in the networks in this study and we discuss the patterns that arise. Finally, in Section 6.6 we summarize our conclusions and findings and give recommendations for future research.

6.2. RELATED WORK

Due to the growing interest in semantic networks, related studies were carried out in a wide range of different fields. Based on our scope, we focus on two main aspects in each work: the topological properties that were analyzed in the study and the dataset that was used in the analysis (i), and the universal and language-specific patterns which were found and discussed (ii).

The majority of semantic network literature is centered around three link types: co-occurrence, association, and semantic relation. In a co-occurrence network, sets of words that co-occur in a phrase, sentence, or text form a link. For association networks, participants in a cognitive-linguistic experiment are given a word and asked to give the first word that they think of. There are several association datasets, one example

is the University of South Florida Free Association Norms [191]. Semantic relations are relations defined by professionals like lexicographers, typical examples are synonyms, antonyms, hypernyms, and homonyms. The specific instances of the semantic relations are also defined by the lexicographers or extracted computationally from text corpora.

In 2001, Ferrer-i-Cancho and Sole [192] studied undirected co-occurrence graphs constructed from the British National Corpus dataset [193]. They measured the average distance between two words and observed the small-world property, which was found in many real-world networks [97]. Motter *et al.* [194] have analyzed an undirected conceptual network constructed from an English Thesaurus dictionary [195]. They focused on three properties: sparsity (small average degree), average shortest path length, and clustering. That same year, Sigman and Cecchi [196] studied undirected lexical networks extracted from the noun subset of WordNet [197], where the nodes are sets of noun synonyms. They grouped networks by three semantic relations: antonymy, hypernymy, and meronymy. A detailed analysis of characteristic length (the median minimal distance between pairs of nodes), degree distributions, and clustering of these networks was provided. Semantic networks were also found to possess the small-world property of sparse connectivity, short average path length, and strong local clustering [194, 196].

Later, Steyvers and Tenenbaum [56] performed statistical analyses on 3 kinds of semantic networks: word associations [191], WordNet and Roget's Thesaurus [198]. Apart from the above-mentioned network properties, they also considered network connectedness and diameter. They pointed out that the small-world property may originate from the scale-free organization of the network, which exists in a variety of real-world systems [3, 199, 200].

As for patterns across different languages, Ferrer-i-Cancho *et al.* [201] built syntactic dependency networks from corpora (collections of sentences) for three European languages: Czech, German, and Romanian. They showed that networks from different languages have many non-trivial topological properties in common, such as the small-world property, a power-law degree distribution, and disassortative mixing [161].

Existing studies have identified some general network properties in semantic networks such as the small-world property and power-law degree distributions. However, the datasets used in these studies are often different, sometimes even within the same study, rendering direct comparison of results difficult. Some used associative networks generated from experiments and some studied thesauri that were manually created by linguists. In addition, most of the research performed consists of coarse-grained statistical analyses. Specifically, different semantic relations were sometimes treated as identical and the subset of included nodes was often limited (*e.g.*, only words and no phrases, or only nouns). Further, there are only very few studies on semantic networks from languages other than English.

Therefore, our analyses focus on semantic networks with different semantic relations (link types) from a single dataset. We consider networks defined by a specific link type, make these networks undirected and unweighted, and compare the structural properties between networks with different link types. In addition, we apply similar analyses to semantic networks with the same link type across different languages. Furthermore, we investigate the roles that similarity and complementarity play in the formation of links in semantic networks.

6.3. GENERAL PROPERTIES OF SEMANTIC NETWORKS

To gain an understanding of the structure of semantic networks, we first study the general topological properties of English semantic networks. The dataset that we use throughout this study is ConceptNet [127], Appendix D.1. We list the semantic relations that define the networks in this study in Table D.1. An overview of the semantic networks under consideration is given by Table D.2. In this Section, we compute various topological properties of these networks related to connectedness, degree, assortative mixing, and clustering. The overall descriptive statistics of the semantic networks are also listed in Table D.2.

6.3.1. CONNECTEDNESS

We measure the connectedness of a network by the size of the largest connected component and the size distribution of all connected components. The component size distributions of the English semantic networks are shown in Figure D.4. Table D.4 lists the sizes of the largest connected components (LCCs) in absolute as well as relative numbers. The same statistics are computed after degree-preserving random rewiring of the links for comparison [161]. The purpose of random rewiring is to estimate the value of the graph metric that could be expected by chance, solely based on the node degrees (see Appendix D.3.1 for details on the rewiring process).

Based on the percentages of nodes in the LCC, all seven semantic networks are not fully connected. The networks 'Is-A', 'Related-To', and 'Union' are almost fully connected given that their LCCs contain over 90% of nodes. Networks 'Has-A', 'Part-Of', 'Antonym', and 'Synonym' are largely disconnected, with the percentages of nodes in their LCCs ranging from 22% to 65%. Most of the rewired networks are more connected than the corresponding original networks, especially networks 'Antonym' and 'Synonym'. In other words, the majority of our semantic networks are less connected than what could be expected by chance. For networks 'Related-To' and 'Union', the percentage of nodes in the LCC remains almost unchanged, while the 'Is-A' network is more connected than expected.

6.3.2. DEGREE DISTRIBUTION

Figure 6.2 shows that the densities $\Pr[D = k]$ of the degree distributions of our seven English semantic networks all appear to follow power laws in the tail visually. A more rigorous framework for assessing power laws was proposed by Voitalov *et al.* [99], who consider networks to have a power-law degree distribution if $\Pr[D = k] = \ell(k)k^{-\gamma}$ for a slowly varying function $\ell(k)$, see Appendix D.3.1. Figure 6.2 includes the estimates $\hat{\gamma}$ based on the slopes of the densities $\Pr[D = k]$ on a log-log scale, along with the three consistent estimators from the framework of Voitalov *et al.* [99, 202]. According to these estimators, the degree sequences of 5 out of the 7 networks are power laws. The degree sequences of the 'Synonym' and 'Antonym' networks are hardly power laws because at least one of the $\hat{\gamma} > 5$ and therefore the estimated exponents are not listed.

For most networks, the estimated exponent $\hat{\gamma}$ lies between 2 and 3. Therefore, most semantic networks are scale-free, except for the 'Synonym' and 'Antonym' networks. In the literature, semantic networks were also found to be highly heterogeneous [56, 57]. Moreover, the word frequencies in several modern languages were found to follow power

laws [203]. In Section 6.4, we will see that the degree distributions of the 'Synonym' and 'Antonym' networks from most considered languages are hardly power laws or not power laws.

The heterogeneity in the degree distribution seems natural for networks such as the 'Is-A' network: there are many specific or unique words with a small degree that connect to only a few other words, while there are also a few general words that connect to almost anything, resulting in a large degree. Examples of general words with a large degree are 'plant' and 'person', while specific words like 'neotectonic' and 'cofinance' have a small degree. Our results show that many semantic networks have power-law degree distributions, like many other types of real-world networks [204–206].

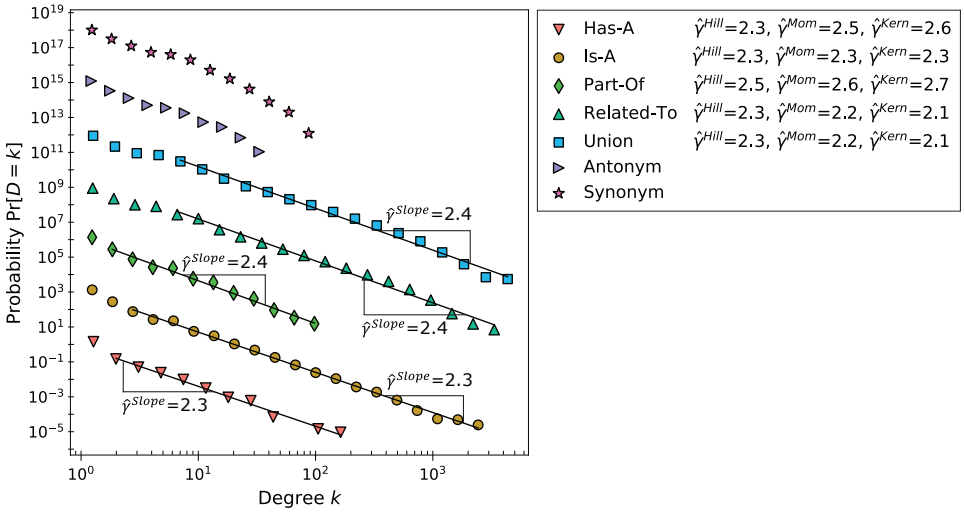


Figure 6.2: Degree distribution densities $\Pr[D = k]$ of the LCCs of the seven English semantic networks. The data is scaled by powers of 1000 to better visualize the power law in each density. The corresponding estimated power-law exponents $\hat{\gamma}$ are shown if there is a power law, $\Pr[D = k] \approx \ell(k)k^{-\gamma}$. The degree sequences of the networks 'Antonym' and 'Synonym' were estimated to be hardly power-law because at least one of the $\hat{\gamma} > 5$. The data are logarithmically binned to suppress noise in the tails of the distributions, see Appendix D.3.1 for details on how the power-law densities are processed and the power-law exponent estimation procedures.

6.3.3. DEGREE ASSORTATIVITY

Several measures have been established to quantify the degree assortativity in a network, such as the degree correlation coefficient ρ_D and the Average Nearest Neighbor Degree (ANND) [161]. Figure 6.3 shows the average nearest neighbor degree as a function of the degree k for four selected networks and their values after random rewiring as well as the degree correlation coefficient ρ_D. Refer to Figure D.2 for the ANND plots of all networks. The randomized networks have no degree-degree correlation and, as a result, the function ANND does not vary with k. The rewired networks serve as a reference for the expected ANND values when the links are distributed at random.

We find that most semantic networks are disassortative as their ANND is a decreasing function of the degree k and the degree correlation coefficient ρ_D is negative. These

networks are 'Has-A', 'Part-Of', 'Is-A', 'Related-To', and 'Union'. In disassortative networks, nodes with larger degrees (general words) tend to connect to nodes with smaller degrees (less general words). This is not surprising. Indeed, if we use these relations in a sentence, then we often relate specific words to more general words. For example, we say 'horse racing is a sport', in which 'horse racing' is a very specific phrase while 'sport' is more general.

On the other hand, network 'Synonym' is assortative as the function ANND increases in the degree k. This indicates that large-degree nodes (general words) connect to nodes that have similar degree (words with the same generality). The same applies to network 'Antonym'. Although the degree correlation is not very pronounced and reflected by the small correlation coefficient $\rho_D = -0.005$, we still see a slight upward trend in the curve of the ANND.

The function ANND of a rewired network is not degree-dependent anymore, shown by the blue curves in Figure 6.3. The curve is almost flat for 'Synonym' and 'Related-To'. At a larger degree k, the curve may drop slightly, as for nodes with large degrees, there are not enough nodes of equal degree to connect to.

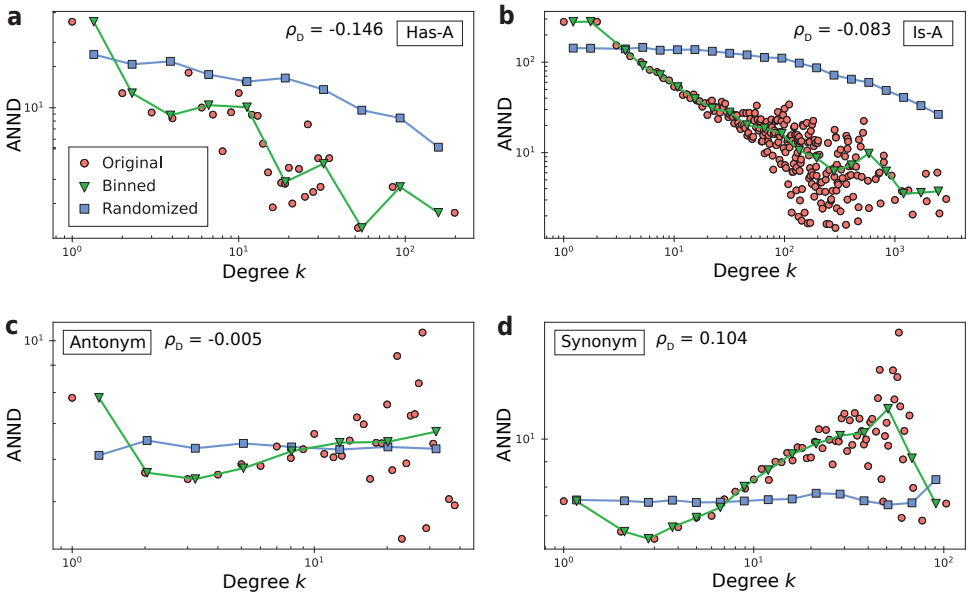


Figure 6.3: Average nearest neighbor degree (ANND) as a function of degree k and degree correlation coefficient ρ_D of four English semantic networks. Panels correspond to the networks (a) 'Has-A', (b) 'Is-A', (c) 'Antonym', and (d) 'Synonym'. See Figure D.2 for the results of all seven networks. Data points in circles are the average ANND of nodes with degree k in a network, triangles represent the data after logarithmic binning, and squares are the average ANND of nodes with degree k in the rewired network. The data are logarithmically binned to better visualize the data.

6.3.4. CLUSTERING COEFFICIENT

In networks such as social networks, the neighbors of a node are likely to be connected as well, which is called clustering [97, 165]. If a person has a group of friends, there is a high chance that these friends also know each other. These networks are characterized by many triangular connections.

The clustering coefficient $c_G(i)$ of a node i is defined in (3.25). Figure 6.4 shows the average clustering coefficient $c_G(i)$ of nodes i with degree k of four English networks. Refer to Figure D.3 for the clustering coefficients of all seven networks. All networks have small clustering coefficients in absolute terms, which, in combination with the small average degree $E[D]$, indicates a local tree-like structure. We find that the networks 'Part-Of', 'Antonym', and 'Synonym' have substantially larger clustering coefficients than their rewired counterparts: there are more triangles in these networks than expected by chance. On the other hand, the network 'Has-A' has lower clustering coefficients $c_G(i)$ than the randomized network, therefore it seems that the 'Has-A' network is organized in a different way than the other networks. As for the networks 'Is-A', 'Related-To', and 'Union', the clustering coefficients $c_G(i)$ are similar to their corresponding rewired networks.

In summary, we find that English semantic networks have power-law degree distributions and most are scale-free, which coincides with the results in previous studies [56, 57]. Besides, semantic networks with different link types show different levels of degree assortativity and average local clustering. Most works in the literature have identified high clustering coefficients in semantic networks [56, 57, 194, 196]. This encourages us to further investigate the organizing principles of these semantic networks, which we will discuss explicitly in Section 6.5.

6.4. LANGUAGE-SPECIFIC PROPERTIES

In this Section, we consider semantic networks from 10 languages other than English contained within ConceptNet: French, Italian, German, Spanish, Russian, Portuguese, Dutch, Japanese, Finnish, and Chinese. We group the 11 languages based on their language families and we again study the topological properties of 7 different semantic relations per language. Finally, we observe peculiarities in the degree distribution densities of the 'Related-To' networks in some languages, which we link to rules of grammar in those languages, as we explain in Section 6.4.4.

6.4.1. LANGUAGE CLASSIFICATION

In linguistics, languages can be partitioned in multiple different ways. Mainly, there are two kinds of language classifications: genetic and typological.

The genetic classification assorts languages according to their level of diachronic relatedness, where languages are categorized into the same family if they evolved from the same root language [207]. An example is the Indo-European family, which includes the Germanic, Balto-Slavic, and Italic languages [208].

One popular typological classification distinguishes isolating, agglutinating, and inflecting languages. It groups languages in accordance with their morphological word formation styles. A morph or morpheme (the Greek word μορφή means 'outer shape,

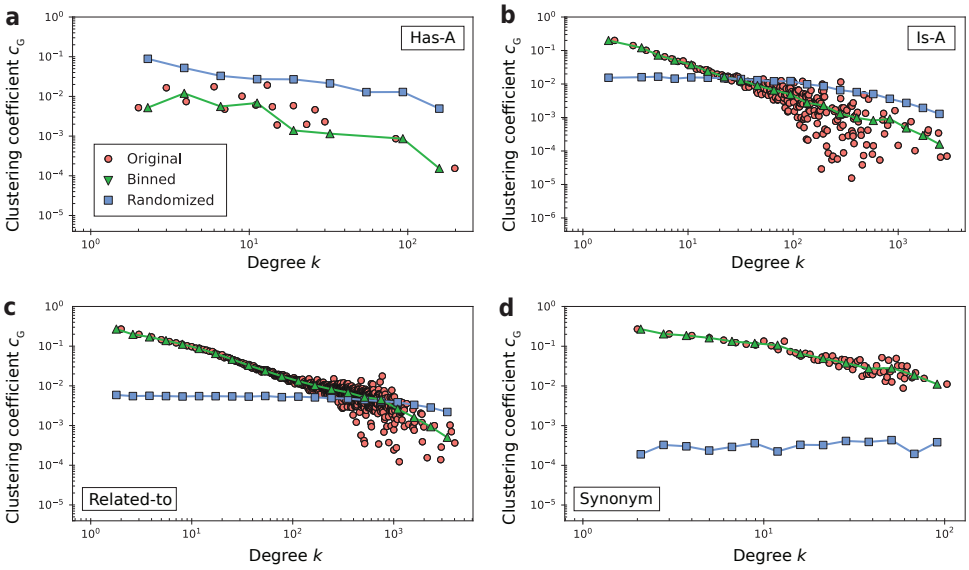


Figure 6.4: The average clustering coefficient $c_G(i)$ of nodes i with degree k of four English semantic networks. Panels correspond to the networks (a) 'Has-A', (b) 'Is-A', (c) 'Antonym', and (d) 'Synonym'. See Figure D.3 for the results of all seven networks. Data points in circles are the original average local clustering coefficients of nodes i with degree k, triangles represent data after logarithmic binning, and squares show the average clustering coefficients of nodes i with degree k (logarithmically binned) in the randomized networks. The data are logarithmically binned to better visualize the data.

6

appearance' of which the English 'form' is derived) is the basic unit of a word, such as a stem or an affix [209]. For instance, the word 'undoubtedly' consists of three morphs: 'un-', 'doubted', and '-ly'. In an isolating language, like Mandarin Chinese, each word contains only a single morph [207]. In contrast, words from an agglutinating language can be divided into morphs with distinctive grammatical categories like tense, person, and gender. In an inflecting language, however, there is no exact match between morphs and grammatical categories [207]. A word changes its form based on different grammar rules. Most Indo-European languages belong to the inflecting category.

Based on these two types of classifications, we have selected 11 languages to cover different language types, Table D.5. Typologically, Chinese is an isolating language, while Japanese and Finnish are agglutinating languages. The rest of the languages under consideration (8 out of 11) belong to the inflecting category. Genetically, French, Italian, Spanish, and Portuguese belong to the Italic family, while English, German, and Dutch are Germanic. Russian is a Balto-Slavic language, Japanese is Transeurasian, Chinese is Sino-Tibetan and Finnish belongs to the Uralic family. We mainly refer to the typological classification throughout our analyses.

6.4.2. OVERVIEW OF SEMANTIC NETWORKS

For every language, we construct seven undirected semantic networks with the link types 'Has-A', 'Part-Of', 'Is-A', 'Related-To', 'Union', 'Antonym', and 'Synonym'. Due to missing

data in ConceptNet, only three languages have the 'Has-A' relation. For these languages, the 'Union' network is the union of three link types: 'Part-Of', 'Is-A', and 'Related-To'. In this Section, we provide an overview of the numbers of nodes N and numbers of links L of the semantic networks. Again, we restrict our study to the LCCs of these networks.

Regarding the numbers of nodes N, the networks 'Related-To' and 'Union' are generally the largest networks in a language, with the French 'Union' network being the absolute largest with $N = 1,296,622$, as denoted in Table D.6. Nevertheless, there are many small networks with size $N < 100$, particularly for the 'Part-Of' and 'Synonym' networks.

Similar to the English semantic networks, we observe that most networks with more than 100 nodes are sparse. All networks have an average degree between 1 and 6, which is small compared to the network size. Consider the Dutch 'Is-A' network for example, where a node has about 5 connections on average, which is only 2.45% of the 191 nodes in the whole network. Table D.7 lists the average degree $E[D]$ of all our semantic networks.

6.4.3. DEGREE DISTRIBUTION

Many of the semantic networks in the 11 languages have degree distributions that are approximately power laws. We estimate the power-law exponents only for networks with size $N > 1000$ because we require a sufficient number of observations to estimate the power-law exponent γ. Table D.8 lists the estimated power-law exponents $\hat{\gamma}$ using the same 4 methods as in Figure 6.2 for each semantic network. Refer to Section D.3.1 for details on these estimation procedures.

We find that many networks have power laws in their degree distributions and many of those networks are scale-free ($2 < \hat{\gamma} < 3$). The Chinese 'Related-To' network even has a power-law exponent $\hat{\gamma} < 2$. The degree distributions of all 'Synonym' and 'Antonym' networks are hardly or not power laws, however. The likely reason for this is that nodes in these networks generally have smaller degrees than in other networks. As a result, the slope of the degree distribution is steeper and therefore not classified as a power law. This is not unexpected, as for a given word there are only a certain number of synonyms or antonyms and, therefore, there are not many nodes with high degrees. Another interesting finding is that the densities of the degree distributions of the 'Related-To' and 'Union' networks for French, Spanish, Portuguese, and Finnish show notable deviations from a straight line in the log-log plot, which we discuss in-depth in the next Section.

6.4.4. LANGUAGE INFLECTION

In some languages, the densities of the degree distributions of the 'Related-To' and 'Union' networks show deviations from a straight line on a log-log scale. An example is the Spanish 'Related-To' network in Figure 6.5(a), where we observe a peak in the tail of the distribution. To find the cause of the anomaly in the degree distribution, we inspect the words with a degree k located in the peak, referred to as peak words, and their neighbors. Table D.9 lists a few examples of the peak words, which are almost all verbs and have similar spellings. The links adjacent to these nodes with higher-than-expected degrees might be the result of grammatical inflections of the same root words since Spanish is a highly inflected language. We observe a similar anomaly in the degree distributions of French, Portuguese, and Finnish 'Related-To' and 'Union' networks. In Table D.6 we

saw that the network 'Union' is mostly composed of 'Related-To' in these four languages, therefore we restrict the analysis to the 'Related-To' networks.

Two common types of language inflection are conjugation, the inflection of verbs, and declension, the inflection of nouns. The past tense of the verb 'to sleep' is 'slept', an example of conjugation in English. The plural form of the noun 'man' is 'men', an example of declension. The languages Spanish, Portuguese, and French are much richer in conjugations than English, while Finnish is rich in declensions.

PART-OF-SPEECH TAGS

In the ConceptNet dataset, only part of the nodes is part-of-speech (POS) tagged with one of four types: verb, noun, adjective, and adverb. For French, Spanish, and Portuguese, the percentage of verbs in the peak is larger than in the LCC, while for Finnish the percentage of nouns in the peak is larger than in the LCC, see Table D.10. Remarkably, 100% of the Portuguese peak words are verbs. Most neighbors of the peak words are verbs for Spanish (97%), Portuguese (99%), and French (87%), while most neighbors of the peak words are nouns for Finnish (90%), Table D.11. This strengthens our belief that the abnormal number of nodes with a certain degree k is related to language inflection in these four languages.

MERGING OF WORD INFLECTIONS

To investigate whether the peaks in the degree distribution densities are indeed related to word inflections, we leverage the 'Form-Of' relation type in ConceptNet, which connects two words A and B if A is an inflected form of B, or B is the root word of A [210]. We merge each node and its neighbors from the 'Form-Of' network (its inflected forms) into a single node in the 'Related-To' network, as depicted in Figure D.5. Figure 6.5 shows the degree distribution densities of the 'Related-To' networks before and after node merging. The range of the anomalous peak in the density of the degree distribution is highlighted in yellow. In each panel, the number of grammatical variations m coincides with the center of the peak. As seen in Figure 6.5(a), the peak completely disappears in the Spanish 'Related-To' after node merging, thus the peak is described entirely by connections due to word inflections. We also observe significant reductions in the heights of the peaks for Portuguese and Finnish 'Related-To' networks. However, for the French 'Related-To' network, there is only a slight reduction in height after merging, which we believe is likely due to poor coverage in the French 'Form-Of' network with only 17% of words in the peak. In contrast, the Spanish 'Form-Of' network covers 97% of the Spanish peak words, while for Portuguese and Finnish, approximately 50% of the peak words are covered, Table D.12.

THE NUMBER OF INFLECTIONS

In a language, the number of distinct conjugations of regular verbs is determined by the number of different pronouns and the number of verb tenses, which are grammatical time references [211]. In Spanish, there are 6 pronouns and 9 simple verb tenses, resulting in at most $m = 6 \times 9 = 54$ distinct verb conjugations [212, 213]. Table 6.1 exemplifies these 54 different conjugations for the verb 'amar', which means 'to love'. There are also irregular verbs that follow different, idiosyncratic grammatical rules, but the majority of

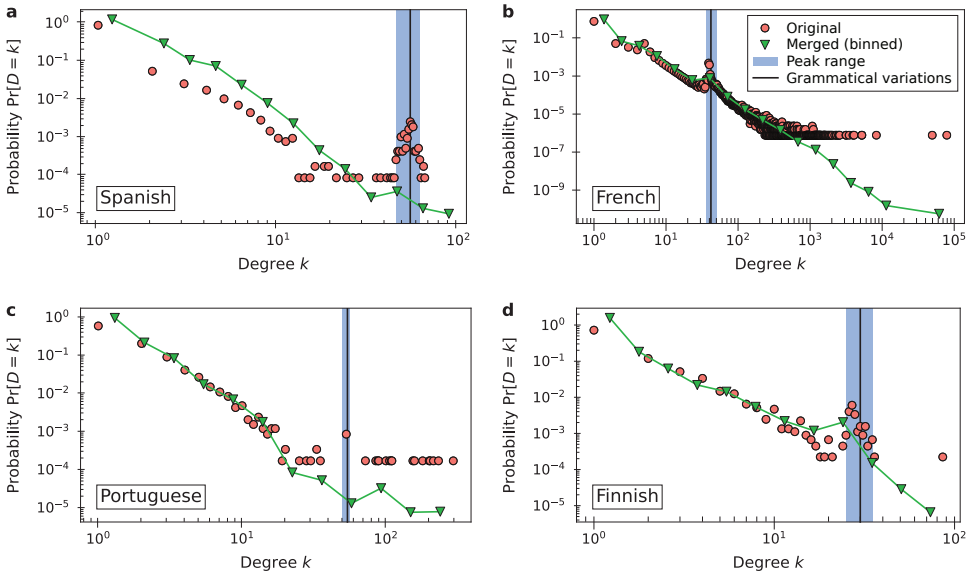


Figure 6.5: Densities Pr[$D = k$] of the degree distributions of the 'Related-To' networks before and after node merging of inflected forms in (a) Spanish, (b) French, (c) Portuguese and (d) Finnish. The logarithmically binned densities after node merging are shown in green. The peaks are highlighted in blue. The vertical black lines indicate the number of grammatical variations m for the relevant grammatical rule in the respective language. In each panel, the number of grammatical variations m coincides with the center of the peak.

verbs in Spanish are classified as regular, like in most inflecting languages. The number of grammatical variations $m = 54$ coincides with the center of the peak in Figure 6.5(a).

Like Spanish, Portuguese has $m = 54$ distinct conjugated verb forms [214]. In French, there are $m = 6 \times 7 = 42$ distinct verb conjugations [215]. In Finnish, there are in total 15 noun declensions or cases with distinct spelling, each having singular and plural forms, resulting in $m = 30$ different cases of a Finnish noun [216]. Table D.13 lists the number of grammatical variations m in French, Spanish, Portuguese, and Finnish, along with the minimum and maximum degree $k_{\min}$ and $k_{\max}$ where the peak starts and ends. By Figure 6.5 we confirm that the number of grammatical variations m coincides with or is close to the center of the peak.

In summary, we observe anomalies in the degree distributions of 'Related-To' networks from the inflecting languages Spanish, French, and Portuguese and the agglutinating Finnish. Because of grammatical structures, root words in these languages share many links with their inflected forms, resulting in more nodes with a certain degree than expected. While Finnish is typologically classified as agglutinating, it still has many noun declensions, suggesting that the agglutinating and inflecting language categories are not mutually exclusive.

Tense \ Pron.	**Yo** (I)	**Tú** (You)	**Él/Ella/Usted** (He/She)	**Nosotros** (We)	**Vosotros** (You)	**Ellos/Ellas/ Ustedes** (They)
Pres. ind.	amo	amas	ama	amamos	amáis	aman
Imp. ind.	amaba	amabas	amaba	amábamos	amabais	amaban
Pret. ind.	amé	amaste	amó	amamos	amasteis	amaron
Fut. ind.	amaré	amarás	amará	amaremos	amaréis	amarán
Cond. ind.	amaría	amarías	amaría	amaríamos	amaríais	amarían
Pres. subj.	ame	ames	ame	amemos	améis	amen
Imp. subj. 1	amara	amaras	amara	amáramos	amarais	amaran
Imp. subj. 2	amase	amases	amase	amásemos	amaseis	amasen
Fut. subj.	amare	amares	amare	amáremos	amareis	amaren

Table 6.1: Verb conjugation table for the Spanish verb 'amar' (to love). The 6 pronouns and 9 verb tenses result in a maximum of 54 different conjugated forms. The 9 verb tenses are: present indicative (pres. ind.), imperfect indicative (imp. ind.), preterite indicative (pret. ind.), future indicative (fut. ind.), conditional indicative (cond. ind.), present subjunctive (pres. subj.), imperfect subjunctive 1 (imp. subj. 1), imperfect subjunctive 2 (imp. subj. 2), and future subjunctive (fut. subj.).

6.5. COMPLEMENTARITY AND SIMILARITY IN SEMANTIC NETWORKS

Although we have identified several universal characteristics of semantic networks, we also observe notable differences in some of their properties. The clustering coefficient in some semantic networks, for instance, is greater than expected by chance, while in other semantic networks, *e.g.*, the English 'Has-A' network, the clustering coefficient is smaller than expected by chance.

We hypothesize that these semantic networks are organized according to different principles. It is commonly known that one such principle is similarity: all factors being equal, similar nodes are more likely to be connected. Similarity is believed to play a leading role in the formation of ties in social networks and lies at the heart of many network inference methods. At the same time, recent works indicate that many networks may be organized predominantly according to the complementarity principle, which dictates that interactions are preferentially formed between nodes with complementary properties. Complementarity has been argued to play a significant role in protein-protein interaction networks [128] and production networks [112]. In addition, we have proposed a geometric complementarity framework for modeling and learning complementarity representations of real networks, Chapter 4.

This Section aims to assess the relative roles of complementarity and similarity mechanisms in different semantic networks. We utilize the method by Talaga and Nowak [129]. The method assesses the relative roles of the two principles by measuring the relative densities of triangular and quadrangle motifs in the networks. Intuitively, the transitivity of similarity—A similar to B and B similar to C implies A similar to C—results in a high density of triangles [182, 217, 218], Figure D.6. The non-transitivity of complementarity, on the other hand, suppresses the appearance of triangles but enables the appearance of quadrangles in networks [128, 219].

We measure and compare the density of triangles and quadrangles with the structural similarity and complementarity coefficients using the framework of Talaga and

Nowak [129]. After computing the densities of triangles and quadrangles, the framework assesses their significance by comparing the densities to those of graphs with matching degree distributions generated by the configuration model, see Appendix D.7.1 for a summary. As a result of the assessment, the network of interest is quantified by two normalized structural coefficients corresponding to complementarity and similarity.

Figure 6.6 depicts the relative roles of complementarity and similarity in 50 semantic networks. We observe that semantic networks are clustered together according to semantic relation types and not their language families, indicating that specific types of semantic relations matter more for the organizing principles of a semantic network rather than its language.

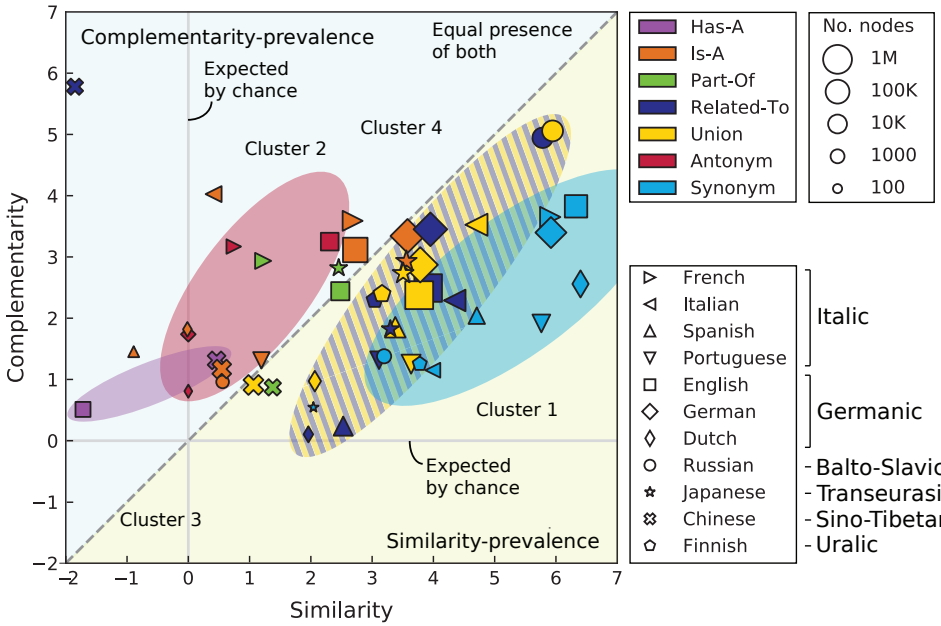


Figure 6.6: Calibrated average structural coefficients of the LCCs of 50 semantic networks from 11 different languages. Languages that belong to the same family are marked with similar shapes: triangles represent Italic, quadrilaterals represent Germanic, circles represent Balto-Slavic, a star represents Transeurasian, a cross represents Sino-Tibetan, and a pentagon represents Uralic. The marker size scales logarithmically with the number of nodes N in the network and is further adjusted for visibility. The grey lines at $x = 0$ and $y = 0$ indicate the expected coefficients based on the configuration model (see Appendix D.7.1). The dashed line at $y = x$ indicates that the structural similarity and complementarity coefficients are equal. Networks in the upper left area (shaded in blue) are more complementarity-based, while networks in the lower right area (shaded in yellow) are more similarity-based. We highlight four clusters of networks using different colors.

Based on the calibrated complementarity and similarity values, we can categorize most semantic networks as (i) predominantly complementarity-based, (ii) predominantly similarity-based, and (iii) networks where both complementarity and similarity are substantially present.

We observe four clustering patterns in Figure 6.6.

- Cluster 1 (light blue): the 'Synonym' networks are characterized by stronger sim-

ilarity than complementarity values. This observation is hardly surprising since 'Synonym' networks link words with similar meanings. Since similarity is transitive, the 'Synonym' networks contain significant numbers of connected node triples, leading to large clustering coefficients.

- Cluster 2 (red): the 'Antonym' networks, as observed in Figure 6.6, belong to the upper triangle of the scatter plot plane, hence complementarity is more prevalent in these networks than similarity. This observation is as expected, as antonyms are word pairs with opposite meanings that complement each other. In Chapter 4, we have learned a geometric representation of the English 'Antonym' network, demonstrating that antonyms indeed complement each other.

  More surprisingly, some 'Antonym' networks are characterized by substantial similarity values, implying the presence of triangle motifs. This is the case since there are instances of three or more words that have opposite meanings to all other words in the group. One example is the triple of words (*man*, *woman*, *girl*). Each pair of words in the triple is opposite in meaning along a certain dimension, here either gender or age.

- Cluster 3 (purple): the 'Has-A' networks show more complementarity than similarity. Intuitively, the words in 'Has-A' networks complement one another. For instance, 'a *house* has a *roof*' describes a complementary relation, and these two objects are not similar to one another.

- Cluster 4 (dark blue/yellow): Most of the 'Related-To' and 'Union' networks show more similarity, except for Chinese. As defined in the 'Related-To' relation, words are connected if there is any sort of positive relationship between them, therefore triangles are easily formed. One exception to that rule is that the Chinese 'Related-To' network (dark blue cross) shows the strongest complementarity of all networks and lower-than-expected similarity. We find that a possible explanation is that the Chinese language has many measure words that are connected to a wide range of nouns. Measure words, also known as numeral classifiers, are used in combination with numerals to describe the quantity of things [220, 221]. For example, the English 'one apple' translates to 一个苹果 in Chinese, where the measure word 个 (gè) must be added as a unit of measurement between the number 'one', 一 (yī), and the noun 'apple', 苹果 (píng guǒ). The Chinese measure word 个 can be loosely translated to English as 'unit(s) of', as in 'one unit of apple'. This grammatical construct is comparable to the phrase 'one box of apples' in English, where 'box' serves as a measure word but, contrary to Chinese, measure words are rare in English. In the Chinese 'Related-To' network, there are many measure words that are connected to multiple nouns, and these nouns may have no connection with each other at all. Most of the measure words are not connected either. Hence, the pairings of measure words and nouns lead to quadrangles, a likely explanation for why the Chinese 'Related-To' network shows the highest complementarity.

6.6. DISCUSSION

In summary, we have conducted an exploratory analysis of the topological properties of semantic networks with 7 distinct semantic relations arising from 11 different languages. We have identified both universal and unique characteristics of these networks.

We find that semantic networks are sparse and that many are characterized by power-law degree distributions. We also find that many semantic networks are scale-free. We observe two different patterns of degree-degree mixing in these networks, some networks are assortative, while some are disassortative. In addition, we find that most networks are more clustered than expected.

On the other hand, some semantic networks—the 'Related-To' networks in French, Spanish, Portuguese, and Finnish—possess unique features that can be explained by rules of grammatical inflection. Because of the grammar in these languages, words have many conjugations or declensions. We have related the anomalous peaks in their degree distributions to language inflections. Notably, we found the effect due to word inflection not only in inflecting languages but also in Finnish, which is an agglutinating language.

We have also quantified the relative roles of complementarity and similarity principles in semantic networks. The proportions of similarity and complementarity in networks differ depending on the semantic relation type. For example, the 'Synonym' networks exhibit stronger similarity, while the links in the 'Antonym' network are primarily driven by complementarity. In addition, the Chinese 'Related-To' network has the highest structural complementarity coefficient of all networks, which we have attributed to a unique grammatical phenomenon in Chinese: measure words.

Through the analysis of the topological properties of semantic networks, we found that complementarity may play an important role in their formation. Since most of the state-of-the-art network inference methods are either built on or inspired by the similarity principle, we call for a careful re-evaluation of these methods when it comes to inference tasks on semantic networks. One basic example is the prediction of missing links. In a seminal work, Kovács *et al.* [128] demonstrated that protein interactions should be predicted with complementarity-tailored methods. We expect that similar methods might be in place for semantic networks. Instead of using the triangle closure principle, one might benefit from the methods based on quadrangle closure, Figure 6.7.

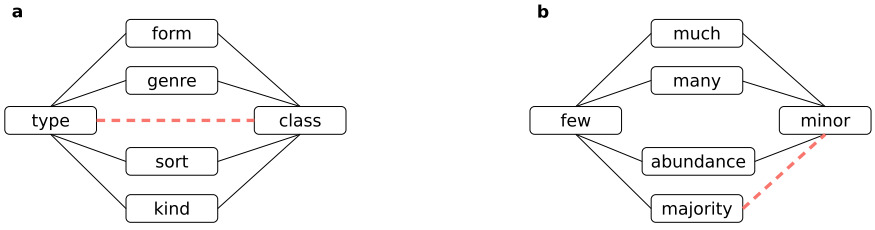


Figure 6.7: Examples of similarity and complementarity in semantic networks. (a) Similarity: triangle closure in the 'Synonym' network. (b) Complementarity: quadrangle closure in the 'Antonym' network.

It is not as easy to illustrate quadrangle closure in network embedding methods or NLP methods in general. A plethora of methods use multiple modules and parameters in learning tasks and can, in principle, be better optimized for the complementarity struc-

ture of semantic networks, see for example [50, 51]. In Chapter 4, we have proposed a complementarity embedding method and illustrated it using the 'Antonym' semantic network of ConceptNet.

Recent groundbreaking advances in large language models are attributed to the multi-head attention mechanism of the Transformer, which uses ideas consistent with the complementarity principle [54]. We advocate that a better understanding of statistical mechanisms underlying semantic networks can help us improve NLP methods even further.

7

CONCLUSION

Risk happens in the future.

Nassim Nicholas Taleb

This dissertation explores several representations of complex networks, focusing on the leading mechanisms that govern the formation of links. The findings of this thesis improve our understanding of the leading connection principles in networks, those of similarity and complementarity. We have studied networks with community structures driven by similarity, and graphs that have hidden representations in hyperbolic spaces of arbitrary dimensionality. Additionally, we have developed a framework for embedding complementarity-driven networks into hyperbolic space.

7.1. MAIN CONTRIBUTIONS

In Chapter 2, we have demonstrated that the spectrum of the non-backtracking matrix provides an excellent estimator of the number of clusters in a network. When the clusters are of approximately equal size, the number of eigenvalues separated from the bulk accurately indicates the number of clusters in the network. In addition, we have shown that the non-backtracking method for estimating the number of clusters has an attractive computational complexity: the computation time scales approximately linearly with the number of nodes in the network.

In Chapter 3, we have extended the 2-dimensional random hyperbolic graph model to the hyperbolic space of arbitrary dimensionality. Our rescaling of the model parameters casts the random hyperbolic graph model of any dimension to a unified mathematical framework, such that the degree distribution is invariant to the dimensionality of the space. We have systematically analyzed the properties of hyperbolic graphs with arbitrary dimensionality of the hidden space for each of the different connectivity regimes of the model and their limiting cases. This work serves the Network Science community

by providing a reference of the properties of hyperbolic graphs, which can be useful for embedding real networks into hyperbolic spaces of arbitrary dimensions.

We have described the newly identified connection principle of complementarity in Chapter 4. We have argued how many of the methods within Network Science are based on the connection principle of similarity, either explicitly or implicitly. We have remarked that there is a class of real-world networks in which the links are driven by principles of complementarity rather than similarity. Consequently, we have proposed a framework for embedding these complementarity-driven networks into hyperbolic spaces, aiding inference tasks on real networks that are driven by complementarity. We have complemented these results by further investigating the topological properties of the complementarity random hyperbolic graph in Chapter 5. A good understanding of both the similarity and complementarity connection principles is fundamental for gaining a full understanding of the geometric and topological properties of complex networks.

In Chapter 6, we have systematically studied the topological properties of semantic networks. We have found both general and language-specific structural properties of semantic networks from 11 different languages. We have shown that the connection principles of similarity and complementarity both play a role in their link formation, and we have discussed how a better understanding of these organizing principles benefits applications of natural language processing.

7.2. DIRECTIONS FOR FUTURE RESEARCH

In Chapter 2, we studied the non-backtracking matrix within the scope of network clustering. However, deriving fundamental results proved challenging, primarily due to the matrix's inherent asymmetry. We observed that employing the non-backtracking matrix in clustering shows potential for specific types of network models, yet formalizing these findings remains an open problem. As advancements in the understanding of asymmetric matrices emerge, it might be worthwhile to revisit the spectrum of the non-backtracking matrix.

We have identified the complementarity connection principle as an alternative to the connection principle of similarity, Chapter 4. Many of the methods within Network Science assume similarity between connected nodes, either explicitly or implicitly. Similarity-based embedding methods are often indiscriminately applied to networks driven by complementarity. It might be useful to re-evaluate inference tasks on networks driven by complementarity, which are currently reliant on embedding methods based on similarity. Furthermore, although our minimal complementarity embedding framework has been demonstrated as a viable solution for embedding networks driven by complementarity, the fine-tuning of this embedding method is far from complete, presenting further opportunities for improvement.

We have shown that both similarity and complementarity play significant roles in the link formation of semantic networks, as discussed in Chapter 6. While we have provided an illustrative example in link prediction for semantic networks, other inference tasks based on semantic networks should be revisited with their proper organizing principles in mind. Particularly, the exploration of recent developments in Transformer-based large language models from the perspective of complementarity might offer an opportunity to

improve the interpretability of these models.

A

APPENDIX TO CHAPTER 2

A.1. THE NON-BACKTRACKING MATRIX H

A.1.1. DEFINITION

The non-backtracking matrix **H** is defined in (2.13). Starting from an undirected graph G, the matrix **H** is defined for each pair of bi-directional links in the bi-directional graph representation of G. Each of the L undirected links in G is represented by two bi-directional links, one in each direction, hence there are $2L$ links in the bi-directional representation, see Figure 2.3. The two bi-directional links that are created from a single undirected link in G are backtracking by construction.

Raising the matrix **H** to the k-th power and taking the element (e, l) counts the number of non-backtracking walks of length $k+1$ from the link e_e to the link e_l. Similarly, raising the adjacency matrix **A** to the k-th power and taking the element (i, j) counts all walks of length k from node i to node j. Krzakala *et al.* [62] and Bordenave *et al.* [222] have argued that counting only the non-backtracking walks rather than all possible walks is more informative on the structure of the graph in specific applications such as clustering.

A.1.2. SPECTRUM

The non-backtracking matrix was first mentioned by Hashimoto in the context of the Ihara-Bass zeta function [73]. The eigenvalues of the non-backtracking matrix **H** are the reciprocal of the poles of the Ihara-Bass zeta function. Stark and Terras [223] found and proved the surprising relation between the eigenvalues of the non-backtracking matrix **H** and the adjacency matrix **A** and the diagonal matrix **Δ** with the degrees k_i on the diagonal. Following [224], the characteristic polynomial of the non-backtracking matrix **H** is:

$$\det(\mathbf{I}_{2L} - z\mathbf{H}) = (1-z^2)^{(L-N)} \det(\mathbf{I}_N - \mathbf{A}z + (\mathbf{\Delta} - \mathbf{I}_N)z^2), \tag{A.1}$$

for complex values $z \in \mathbb{C}$ and where $\mathbf{I}_{2L}$ is the $2L \times 2L$ identity matrix and $\mathbf{I}_N$ the $N \times N$ identity matrix. The equation (A.1) has $2(L-N)$ simple zeros that are equal to $z = \pm 1$.

The other $2N$ eigenvalues are given by the quadratic eigenvalue equation on the right-hand side of (A.1). Angel *et al.* [224] remark that the $2N$ non-trivial eigenvalues can be obtained from the eigenvalue equation of a block matrix $\mathbf{H}^*$ constructed as follows

$$\mathbf{H}^* = \begin{bmatrix} \mathbf{A} & \mathbf{I}-\mathbf{\Delta} \\ \mathbf{I} & \mathbf{O} \end{bmatrix}. \tag{A.2}$$

Computing the $2N$ eigenvalues of the $2N \times 2N$ block matrix $\mathbf{H}^*$ rather than all $2L$ eigenvalues of the $2L \times 2L$ non-backtracking matrix $\mathbf{H}$ reduces the number of required computations significantly for most graphs.

A.1.3. SYMMETRY

The non-backtracking matrix $\mathbf{H}$ is an asymmetric matrix by definition: if a walk over two directed links e_e and e_l is non-backtracking, then a walk in the opposite direction is not possible because of the directions of the links. If the element (e,l) equals 1, then the element (l,e) must always equal 0. Still, the non-backtracking matrix $\mathbf{H}$ features certain symmetry. The matrix $\mathbf{H}$ is defined for the pairs of bi-directional links that are created from the undirected links in the graph G. Bordenave *et al.* [222] note that if the link e_e followed by the link e_l is non-backtracking, then the same links but in opposite directions and opposite order are also non-backtracking. It turns out that the transpose $\mathbf{H}^T$ of the matrix $\mathbf{H}$ is the non-backtracking matrix of the same graph G but with opposite directions in the bi-directional graph representation. After all, the chosen directions in the bi-directional graph representation are arbitrary (see Figure 2.3), and turning them around does not change the structure of the graph G. There exists a permutation matrix $\mathbf{P}$ that relabels the bi-directional links in the bi-directional graph representation such that all directions are reverted:

$$\mathbf{HP} = \mathbf{PH}^T, \tag{A.3}$$

where $\mathbf{HP}$ is a symmetric matrix. Bordenave *et al.* [222] find that while the eigenvalues of $\mathbf{HP}$ are strongly related to the node degrees in the graph G, the bulk of the eigenvalues of $\mathbf{H}$ are not.

A.1.4. COMPUTATIONAL COMPLEXITY

We discuss the computational complexity of detecting the number of clusters c with the non-backtracking method as defined in (2.14). The size of the square non-symmetric matrix $\mathbf{H}^*$ in (A.2) is of order $O(N^2)$ and calculating all $2N$ eigenvalues of $\mathbf{H}^*$ takes an order of $O(N^3)$ operations, for example with the QR algorithm for non-symmetric matrices [225]. For dense networks, calculating the eigenvalues of $\mathbf{H}^*$ offers a large reduction in computational complexity as compared to calculating all eigenvalues of $\mathbf{H}$ in an order of $O(L^3)$ operations. It is even more efficient to apply the iterative procedure described in Section 2.4.1 to the matrix $\mathbf{H}^*$. Alternatively, it suffices to compute the k largest eigenvalues in absolute value for a small number $k << N$ if the number k is larger than the number of clusters c. An efficient way to approximate the $k << N$ largest eigenvalues of an $N \times N$ (or $2N \times 2N$) sparse matrix is with ARPACK, a collection of algorithms mainly based on the Arnoldi process [225]. ARPACK is used for eigenvalue computations in popular software packages. It is difficult to determine the exact complexity of ARPACK, but

in some cases, it is possible to obtain the k eigenvalues in an order of as low as $O(N)$ operations for a sparse matrix [226], which offers a great reduction in computational complexity for the non-backtracking method. Whether the condition $k > c$ holds has to be checked in retrospect, but it is safest to choose a number k that is much higher than the number of clusters c one expects to find.

Figure A.1 shows a numerical evaluation of the scaling of the computation time of the non-backtracking method. We measure the computation time T of estimating the number of clusters c by calculating the $k = 10$ largest eigenvalues with the ARPACK implementation of MATLAB [59]. The non-backtracking method is evaluated on graphs of exponentially increasing sizes generated by an SSBM with the number of clusters $c = 3$ and parameters $b_{\text{in}} = 10$ and $b_{\text{out}} = 1$, which is slightly above the detectability threshold in (2.12). Figure A.1 shows the logarithm of the computation time T plotted versus the logarithm of the number of nodes N. Regression lines are fitted to 3 different sections. In the leftmost section, the computation time T scales slower than linearly with the network size N. In the middle and rightmost sections, the computation time T scales slightly faster than linearly with the network size N. Overall, the scaling of the computation time T is close to linear. The scaling of the computation time of detecting the number of clusters with the non-backtracking method is therefore quite favorable.

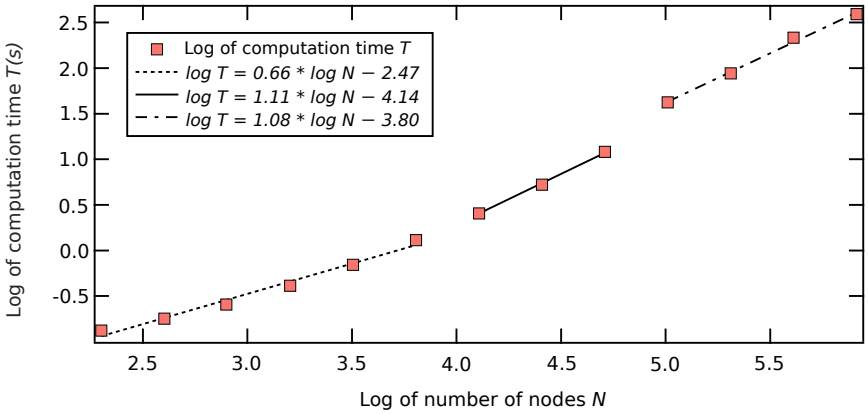


Figure A.1: Measured computation times T in seconds for evaluating the non-backtracking method on graphs generated by an SSBM with an exponentially increasing number of nodes N, from $N = 200$ to $N = 200 \cdot 2^{12}$. The SSBM graphs are generated with the number of clusters $c = 3$ and parameters $b_{\text{in}} = 10$ and $b_{\text{out}} = 1$. The $k = 10$ largest eigenvalues are calculated from the $2N \times 2N$ matrix $\mathbf{H}^*$ (A.2) with ARPACK. For each network size N, we average the computation time over 20 instances. Regression lines are shown for three different sections.

B

APPENDIX TO CHAPTER 3

B.1. HYPERBOLIC GRAPHS IN 2 DIMENSIONS

Here we briefly illustrate the main steps for derivations of the approximations of the average degree $E[D]$ and density $\Pr[D=k]$ of the degree distribution of an arbitrary node in a 2-dimensional hyperbolic graph, as proposed in [21].

B.1.1. APPROXIMATION HYPERBOLIC LAW OF COSINES

The hyperbolic law of cosines for the distance $d(\mathbf{x}_i,\mathbf{x}_j)$ between two points $\mathbf{x}_i = (r_i,\theta_i)$ and $\mathbf{x}_j = (r_j,\theta_j)$ in a hyperbolic disk with curvature $K = -\zeta^2$ given in (3.6),

$$\cosh\zeta d(\mathbf{x}_i,\mathbf{x}_j) = \cosh\zeta r_i \cosh\zeta r_j - \sinh\zeta r_i \sinh\zeta r_j \cos\Delta\theta_{ij}, \tag{B.1}$$

with $\Delta\theta_{ij} = \pi - |\pi - |\theta_i - \theta_j||$, can be approximated well with a simpler expression when both r_i and r_j are sufficiently large. The functions $\sinh(x) = \frac{1}{2}[e^x - e^{-x}]$ and $\cosh(x) = \frac{1}{2}[e^x + e^{-x}]$ are dominated by the term e^x when x is large, namely $\sinh(x) = \frac{1}{2}e^x + O(e^{-x})$ and $\cosh(x) = \frac{1}{2}e^x + O(e^{-x})$. Invoking the approximations $\sinh(x) \approx \frac{1}{2}e^x$ and $\cosh(x) \approx \frac{1}{2}e^x$ in (5.3), provided that r_i and r_j are sufficiently large, we have

$$e^{\zeta d(\mathbf{x}_i,\mathbf{x}_j)} \approx \frac{1}{2}e^{\zeta r_i}e^{\zeta r_j} - \frac{1}{2}e^{\zeta r_i}e^{\zeta r_j}\cos\Delta\theta_{ij} \tag{B.2}$$

$$\approx \frac{1}{2}e^{\zeta r_i + r_j)}\left(1 - \cos\Delta\theta_{ij}\right). \tag{B.3}$$

Substituting the trigonometric identity $1 - \cos\Delta\theta_{ij} = 2\sin^2(\Delta\theta_{ij}/2)$,

$$e^{\zeta d(\mathbf{x}_i,\mathbf{x}_j)} \approx e^{\zeta(r_i+r_j)}\sin^2\left(\frac{\Delta\theta_{ij}}{2}\right), \tag{B.4}$$

and

$$\zeta d(\mathbf{x}_i,\mathbf{x}_j) \approx \zeta(r_i + r_j) + 2\ln\left(\sin\left(\frac{\Delta\theta_{ij}}{2}\right)\right). \tag{B.5}$$

Finally,

$$d(\mathbf{x}_i,\mathbf{x}_j) \approx r_i + r_j + \frac{2}{\zeta}\ln\left(\sin\left(\frac{\Delta\theta_{ij}}{2}\right)\right). \tag{B.6}$$

B

Figure B.1 shows the true distance (5.3) and its approximation (B.6) for two points at an angle $\Delta\theta_{ij} = 0.1$, where point i is located at the boundary of the space. Only when the radial coordinate r_j is very small relative to R, the approximation breaks down. The approximation breakdown only occurs for a few nodes in the graph, as most radial coordinates will be close to the boundary of the disk, as specified by the radial coordinate distribution (4.13). The slight uptick in the hyperbolic distance for low values of r_j is due to the curvature of the hyperbolic disk: nodes near the center of the space are close to almost every point in the space.

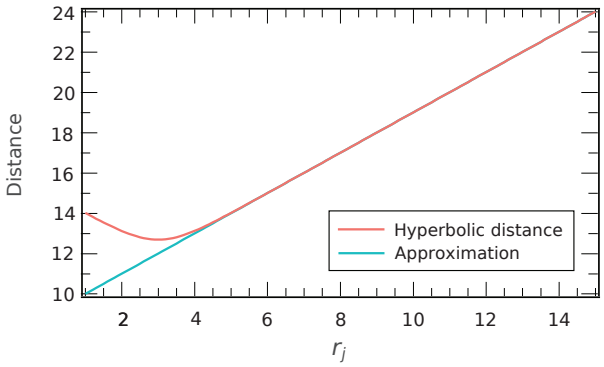


Figure B.1: Evaluation of the hyperbolic distance approximation. Hyperbolic distance and its approximation between two points i and j at an angle $\Delta\theta_{ij} = 0.1$, as a function of r_j, for $r_i = 15$, in a hyperbolic disk with $R = 15$ and $\zeta = 1$.

B.1.2. AVERAGE DEGREE

We find the average degree $E[D_i|X_{0,i} = r_i, X_{1,i} = \theta_i]$ of a node i with known coordinates $\mathbf{X}_i = (r_i,\theta_i)$ by integrating the connection probability density function (5.5) over all possible coordinates for other nodes j,

$$E[D_i|X_{0,i} = r_i, X_{1,i} = \theta_i] = (N-1)\int dr_j f_{X_0}(r_j)\int d\theta_j f_{X_1}(\theta_j) f(d(r_i,\theta_i;r_j,\theta_j)) \tag{B.7}$$

$$= (N-1)\int dr_j f_{X_0}(r_j)\int d\theta_j f_{X_1}(\theta_j)\frac{1}{1+e^{\frac{1}{2T}(d(r_i,\theta_i;r_j,\theta_j)-R)}}, \tag{B.8}$$

with $d(r_i,\theta_i;r_j,\theta_j)$ as defined in (5.3). By the symmetry and uniformity of the distribution of X_1, the expectation $E[D_i|X_{0,i} = r_i, X_{1,i} = \theta_i]$ does not depend on the coordinate $X_{1,i} = \theta_i$ and we set $\theta_i = 0$ for convenience of calculation. At $\theta_i = 0$, the angle $\Delta\theta_{ij} =$

θ_j . We compute $E[D_i|X_{0,i}=r_i, X_{1,i}=0] = E[D_i|X_{0,i}=r_i]$ and plug in $f_{X_1}(\theta_j)=\frac{1}{2\pi}$,

$$E[D_i|X_{0,i}=r_i] = \frac{N-1}{2\pi}\int_0^R dr_j f_{X_0}(r_j)\int_0^{2\pi} d\theta_j \frac{1}{1+e^{\frac{1}{2T}(d(r_i,0;r_j,\theta_j)-R)}}. \tag{B.9}$$

The integrand in the inner integral depends on θ_j only through the term $\cos(\theta_j)$ in $d(r_i,0;r_j,\theta_j)$, therefore the integrand is symmetric around π in the interval $\theta_j \in [0,2\pi]$. We simplify the inner integral from $[0,2\pi]$ by taking twice the integral from $[0,\pi]$,

$$E[D_i|X_{0,i}=r_i] = \frac{N-1}{\pi}\int_0^R dr_j f_{X_0}(r_j)\int_0^{\pi} d\theta_j \frac{1}{1+e^{\frac{1}{2T}(d(r_i,0;r_j,\theta_j)-R)}}. \tag{B.10}$$

Still, the resulting double integral has no closed-form solution and we have to employ several approximations to obtain the average degree [21]. First, we approximate the hyperbolic distance as in (B.6) while simultaneously setting $\zeta=1$,

$$E[D_i|X_{0,i}=r_i] \approx \frac{N-1}{\pi}\int_0^R dr_j f_{X_0}(r_j)\int_0^{\pi} d\theta_j \frac{1}{1+e^{\frac{1}{2T}(r_i+r_j+2\ln(\sin(\theta_j/2))-R)}} \tag{B.11}$$

$$= \frac{N-1}{\pi}\int_0^R dr_j f_{X_0}(r_j)\int_0^{\pi} d\theta_j \frac{1}{1+\left[e^{\frac{1}{2}(r_i+r_j-R)}\sin(\theta_j/2)\right]^{\frac{1}{T}}}. \tag{B.12}$$

Second, by the Taylor series expansion $\sin(x) = x + O(x^3)$, we observe that the most significant contributions to the inner integral come from θ_j values close to 0. As the term $O(x^3)$ will be very small for θ_j values close to 0, we approximate $\sin(\theta_j/2)\approx\theta_j/2$,

$$E[D_i|X_{0,i}=r_i] \approx \frac{N-1}{\pi}\int_0^R dr_j f_{X_0}(r_j)\int_0^{\pi} d\theta_j \frac{1}{1+\left[e^{\frac{1}{2}(r_i+r_j-R)}\left(\frac{\theta_j}{2}\right)\right]^{\frac{1}{T}}}. \tag{B.13}$$

Next, we perform a change-of-variable substitution $u = e^{\frac{1}{2}(r_i+r_j-R)}\left(\frac{\theta_j}{2}\right)$, such that $du = \frac{1}{2}e^{\frac{1}{2}(r_i+r_j-R)}d\theta_j$,

$$E[D_i|X_{0,i}=r_i] \approx \frac{2(N-1)}{\pi}e^{(R-r_i)/2}\int_0^R dr_j f_{X_0}(r_j)e^{-r_j/2}\int_0^{\frac{1}{2}e^{\frac{1}{2}(r_i+r_j-R)}} \frac{du}{1+u^{\frac{1}{T}}}. \tag{B.14}$$

For most node pairs, r_i+r_j-R will be approximately equal to R, as each of the coordinates r_i and r_j are close to R. We require $R\sim\ln N$, therefore $R\to\infty$ when $N\to\infty$. For $T<1$, the function $(1+u^{\frac{1}{T}})^{-1}$ is fast decreasing and function values for large u contribute very little. We approximate the upper limit of the inner integral as $\frac{1}{2}e^{\frac{1}{2}(r_i+r_j-R)}\approx\infty$,

$$E[D_i|X_{0,i}=r_i] \approx \frac{2(N-1)}{\pi}e^{(R-r_i)/2}\int_0^R dr_j f_{X_0}(r_j)e^{-r_j/2}\int_0^{\infty} \frac{du}{1+u^{\frac{1}{T}}}. \tag{B.15}$$

We observe that the integral over r_j is now independent of u and that

$$\int_0^R dr_j f_{X_0}(r_j)e^{-r_j/2} = E\left[e^{-X_0/2}\right], \tag{B.16}$$

for which the exact evaluation is denoted in (5.16), leading to

$$E[D_i|X_{0,i}=r_i] \approx \frac{2(N-1)}{\pi} e^{(R-r_i)/2} E\left[e^{-X_0/2}\right] \int_0^\infty \frac{du}{1+u^{\frac{1}{T}}}. \tag{B.17}$$

For $T<1$, the integral over u has a well-defined closed-form solution

$$\int_0^\infty \frac{du}{1+u^{\frac{1}{T}}} = \frac{\pi T}{\sin(\pi T)}, \tag{B.18}$$

which gives us the final approximate expression for the average degree of a node with known radial coordinate

$$E[D_i|X_{0,i}=r_i] \approx 2(N-1)\frac{T}{\sin(\pi T)} E\left[e^{-X_0/2}\right] e^{(R-r_i)/2}. \tag{B.19}$$

The corresponding approximation of the average degree of an arbitrary node in the RHG model is

$$E[D_i] = \int_0^R dr_i f_{X_0}(r_i) E[D_i|X_{0,i}=r_i] \tag{B.20}$$

$$\approx 2(N-1)\frac{T}{\sin(\pi T)} E\left[e^{-X_0/2}\right] \int_0^R dr_i f_{X_0}(r_i) e^{(R-r_i)/2} \tag{B.21}$$

$$\approx 2\nu e^R \left(E\left[e^{-X_0/2}\right]\right)^2 \frac{T}{\sin(\pi T)}. \tag{B.22}$$

To obtain a leading-order approximation of (B.19), we approximate $f_{X_0}(r)$ in (B.16) using that

$$f_{X_0}(r) = \frac{\frac{\alpha}{2}[e^{\alpha r} - e^{-\alpha r}]}{\frac{1}{2}[e^{\alpha R} - e^{-\alpha R}] - 1} = \frac{\frac{\alpha}{2} e^{\alpha r} + O(e^{-\alpha r})}{\frac{1}{2} e^{\alpha R} + O(e^{-\alpha R}) - 1}, \tag{B.23}$$

allowing us to drop the terms $O(e^{-\alpha r})$ and $O(e^{-\alpha R})$, since they are both close to 0 for $\alpha > \frac{1}{2}$,

$$f_{X_0}(r) \approx \frac{\frac{\alpha}{2} e^{\alpha r}}{\frac{1}{2} e^{\alpha R} - 1}. \tag{B.24}$$

Also, since R is large $\frac{1}{2} e^{\alpha R} - 1 \approx \frac{1}{2} e^{\alpha R}$, arriving at an approximation of the density function

$$\tilde{f}_{X_0}(r) = \alpha e^{\alpha(r-R)}, \tag{B.25}$$

such that $f_{X_0}(r) \approx \tilde{f}_{X_0}(r)$ for $r \in [0,R]$. Given that integrating $\tilde{f}_{X_0}(r)$ over the entire domain of X_0 yields

$$\int_0^R \tilde{f}_{X_0}(r)\, dr = 1 - e^{-\alpha R}, \tag{B.26}$$

the density function $\tilde{f}_{X_0}(r)$ is only normalized approximately, as in general $e^{-\alpha R} > 0$, but $e^{-\alpha R} \to 0$ as $R \to \infty$ for $\alpha > \frac{1}{2}$. The expectation $E\left[e^{-X_0/2}\right]$ is approximated by substituting the approximate probability density function $\tilde{f}_{X_0}(r)$ for $f_{X_0}(r)$ in (B.16),

$$E\left[e^{-X_0/2}\right] \approx \int_0^R \tilde{f}_{X_0}(r) e^{-r/2}\, dr = \left(\frac{2\alpha}{2\alpha - 1}\right)\left[e^{-R/2} - e^{-\alpha R}\right]. \tag{B.27}$$

We further approximate $E\left[e^{-X_0/2}\right]$ to the leading order by again invoking $e^{-\alpha R} \approx 0$, therefore

$$E\left[e^{-X_0/2}\right] \approx \left(\frac{2\alpha}{2\alpha-1}\right) e^{-R/2}, \tag{B.28}$$

which we plug into (B.22) to obtain a leading-order approximation of the average degree in the RHG model

$$E[D_i] \approx 2\nu \left(\frac{2\alpha}{2\alpha-1}\right)^2 \frac{T}{\sin(\pi T)}. \tag{B.29}$$

The average degree $E[D_i|X_{0,i} = r_i]$ of a node with known radial coordinate is approximated in a similar way,

$$E[D_i|X_{0,i} = r_i] \approx 2\nu \left(\frac{2\alpha}{2\alpha-1}\right) \frac{T}{\sin(\pi T)} e^{(R-r_i)/2}, \tag{B.30}$$

such that the average degree approximations are then related as

$$E[D_i|X_{0,i} = r_i] \approx \frac{E[D_i]}{E\left[e^{-X_0/2}\right]} e^{-r_i/2}. \tag{B.31}$$

B.1.3. DEGREE DISTRIBUTION

The density $\Pr[D = k]$ of the degree distribution in the RHG model is found by integrating the density $\Pr[D = k|X_{0,i} = r_i]$ of the degree distribution of a node with known radial coordinate over all possible radial coordinates, by the law of total probability [3],

$$\Pr[D = k] = \int_0^R dr_i f_{X_0}(r_i) \Pr[D = k|X_{0,i} = r_i]. \tag{B.32}$$

In general, it is difficult to find a closed-form expression for the probability density function $\Pr[D = k|X_{0,i} = r_i]$. For sparse graphs, Boguñá and Pastor-Satorras [96] propose that $\Pr[D = k]$ in (B.32) is approximated sufficiently well by replacing the density $\Pr[D = k|X_{0,i} = r_i]$ by its Poisson approximation,

$$\Pr[D = k|X_{0,i} = r_i] \approx \frac{1}{k!} [\lambda(r_i)]^k e^{-\lambda(r_i)}, \tag{B.33}$$

with the rate $\lambda(r_i) = E[D_i|X_{0,i} = r_i]$, which is the average degree of a node with known radial coordinate in (B.30). We define the constant $c = 2\nu\left(\frac{2\alpha}{2\alpha-1}\right)\frac{T}{\sin(\pi T)} e^{R/2}$ to shorten the notation and we substitute the approximate probability density function $\tilde{f}_{X_0}(r)$ for $f_{X_0}(r)$ in (B.32),

$$\Pr[D_i = k] \approx \frac{1}{k!} \int_0^R \tilde{f}_{X_0}(r_i) [\lambda(r_i)]^k e^{-\lambda(r_i)} dr_i \tag{B.34}$$

$$= \frac{\alpha}{k!} \int_0^R dr_i e^{\alpha(r_i-R)} \left[ce^{(R-r_i)/2}\right]^k e^{-ce^{(R-r_i)/2}}. \tag{B.35}$$

The integral (B.35) is evaluated with a change-of-variable substitution $t = ce^{(R-r_i)/2}$, $dt = -\frac{1}{2}ce^{(R-r_i)/2}dr_i$ and $dr_i = -2t^{-1}dt$. We have $e^{\alpha(r_i-R)} = c^{2\alpha}t^{-2\alpha}$, and $t = ce^{R/2}$ when $r_i =$

0, while $t = c$ when $r_i = R$. Hence,

$$\Pr[D_i = k] \approx -\frac{2\alpha}{k!} c^{2\alpha} \int_{ce^{R/2}}^{c} dt\, t^{k-2\alpha-1} e^{-t} \tag{B.36}$$

$$= \frac{2\alpha}{k!} c^{2\alpha} \int_{c}^{ce^{R/2}} dt\, t^{k-2\alpha-1} e^{-t}, \tag{B.37}$$

where we have flipped the sign of the integral in the last step. Since $ce^{R/2} \to \infty$ as $R \to \infty$ and since the integrand is a fast decreasing function, we approximate the upper bound of the integral in (B.37) as $ce^{R/2} \approx \infty$ and

$$\Pr[D_i = k] \approx \frac{2\alpha}{k!} c^{2\alpha} \int_{c}^{\infty} dt\, t^{k-2\alpha-1} e^{-t}. \tag{B.38}$$

Recognizing that the resulting integral is the definition of the upper incomplete gamma function $\Gamma[k-2\alpha, c]$, we arrive at

$$\Pr[D_i = k] \approx \frac{2\alpha}{k!} c^{2\alpha} \Gamma[k-2\alpha, c] \tag{B.39}$$

$$= (2\alpha)\, c^{2\alpha} \frac{\Gamma[k-2\alpha, c]}{\Gamma[k+1]}. \tag{B.40}$$

From here we conclude that if the constant c is relatively small, that for large k the density $\Pr[D_i = k]$ of the degree distribution scales approximately as a power-law

$$\Pr[D_i = k] \sim k^{-\gamma} \tag{B.41}$$

with exponent $\gamma = 2\alpha + 1$, controlled by the parameter $\alpha > \frac{1}{2}$, such that $\gamma \in (2, \infty)$.

B.1.4. AVERAGE DEGREE $E[D \mid \mathfrak{X}_0 = \mathfrak{r}]$

Figure B.2 shows the theoretical values and their approximations of the average degree $E[D \mid \mathfrak{X}_0 = \mathfrak{r}]$ for different values of $\mathfrak{r} \in [0, \mathscr{R}]$. The approximations of $E[D \mid \mathfrak{X}_0 = \mathfrak{r}]$ are close to the theoretical values in each of the three regimes, but there is a small constant bias in the hot regime. The approximations break down for nodes close to the center of $\mathbb{B}^{d+1}$ (values of $\mathfrak{r}$ close to 0) across all three regimes.

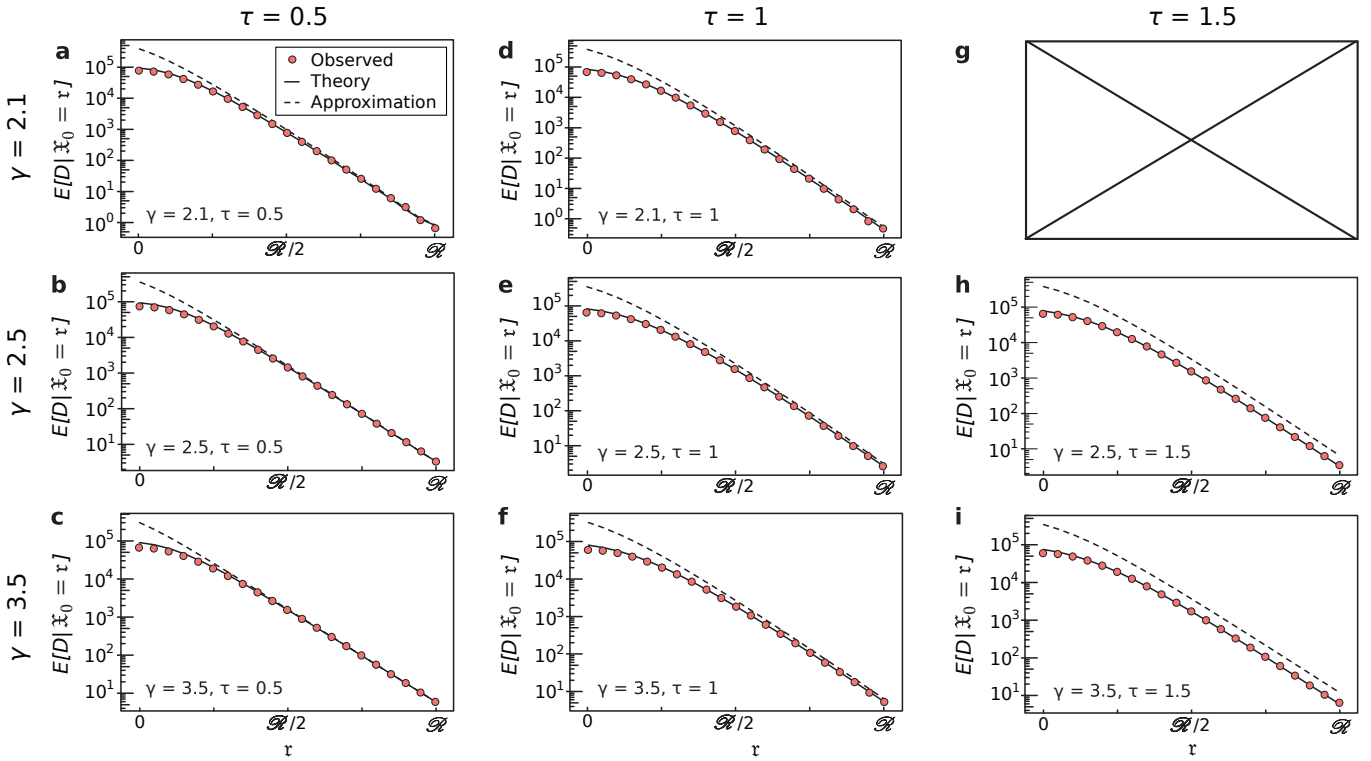


Figure B.2: The average degree $E[D|\mathfrak{X}_0 = \mathfrak{r}]$ for $\mathfrak{r} \in [0, \mathscr{R}]$ and $N = 10^5$, $d = 3$ and $E[D] = 10$. Panels (a-c) correspond to the cold regime at $\tau = 0.5$, panels (d-f) correspond to the critical regime at $\tau = 1$, and panels (g-i) correspond to the hot regime at $\tau = 1.5$. Further, $\gamma = 2.1$ in panels (a,d,g), $\gamma = 2.5$ in panels (b,e,h), and $\gamma = 3.5$ in panels (c,f,i). The combination $\gamma = 2.1$, $\tau = 1.5$ in panel (g) implies $a < 1$, which is not possible in our framework. Each point is the average of 100 simulations. The solid lines are the theoretical values for $E[D|\mathfrak{X}_0 = \mathfrak{r}]$ prescribed by Eq. (3.18) and the dashed lines are the same values but with the approximations $\sinh x \approx x$.

C

APPENDIX TO CHAPTER 4

C.1. COMPLEMENTARITY-BASED NETWORKS

In our work, we employ five networks where complementarity plays an important role. Here we provide basic description of these networks. The basic topological properties of these networks are described in Table C.1.

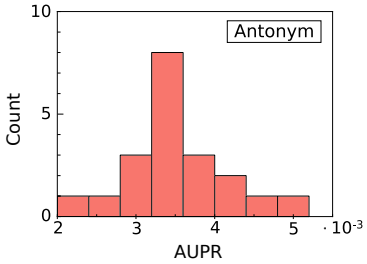
The 'Antonym' semantic network is a network of words and short phrases, where links are established between words or short phrases if those are antonyms. The network is a subset of semantic relationships of the ConceptNet database [127], Appendix D.1.

Human PPI and Yeast PPI are networks of protein-protein interactions. In both networks, the nodes are specific proteins, and the links represent interactions between these proteins. The Human PPI network is constructed by the Human Reference Protein Interactome Mapping Project (HuRi) [141]. The Yeast PPI network is the network of protein-protein interactions of *S. cerevisiae* commonly known as baker's yeast. We obtain the Yeast PPI network from the BioGrid interaction database [227].

Messel is a foodweb network of the Messel Shale [143]. The nodes are organisms or taxa, and the links are feeding relationships between the organisms. Finally, Hamster is the social network consisting of friendship ties between the users of the website `hamsterster.com`. The network is obtained from the KONECT database [144].

Network	N	L	$E[D]$
Antonyms	5,912	7,986	2.70
Human PPI	6,480	31,576	9.75
Yeast PPI	3,488	11,386	6.53
Messel	701	6,395	18.25
Hamster	1,788	12,476	13,96

Table C.1: Basic properties of the largest connected components of the considered complementarity-based networks.



C

Figure C.1: Distribution of link prediction AUPR scores for 20 independently learned complementarity representations of the 'Antonym' semantic network.

C.2. COMPLEMENTARITY HYPERLINK EMBEDDER

We learn complementarity representations by repurposing the HyperLink (HL) embedder [29]. The HL embedder is designed for embedding similarity-driven networks into a 2-dimensional hyperbolic disk $\mathbb{H}^2$ using maximum likelihood estimation. In more precise terms, the HL embedder aims to find node coordinates in $\mathbb{H}^2$ by maximizing the likelihood that the network of interest is generated as a Random Hyperbolic Graph. The HL embedder is freely available from the repository [228].

The Complementarity HyperLink (CHL) embedder finds node coordinates in $\mathbb{H}^2$ by maximizing the likelihood that the network of interest is generated by the Complementarity Random Hyperbolic Graph (CRHG) model. The CHL embedder is almost identical to the HL embedder save for the function it is maximizing. To facilitate the embedding of a trimmed network, with links missing at rate $1-q$ uniformly at random, the CHL embedder is looking for node coordinate assignment $\{\mathbf{x}_i\}$ and $\{\mathbf{y}_i\}$ for every node i in the network, $i = 1, \ldots, N$, maximizing the posterior probability

$$\mathscr{L}\left(\{\mathbf{x}_i, \mathbf{y}_i\}|\mathbf{A}, \mathscr{P}\right) = \frac{\mathscr{L}\left(\mathbf{A}|\{\mathbf{x}_i, \mathbf{y}_i\}, \mathscr{P}\right) \Pr\left[\{\mathbf{x}_i\}\right] \Pr\left[\{\mathbf{y}_i\}\right]}{\mathscr{L}\left(\mathbf{A}|\mathscr{P}\right)}, \tag{C.1}$$

where $\mathscr{P}$ are the model parameters, $\mathbf{A}$ is the adjacency matrix with elements a_{ij}, and $\mathscr{L}\left(\{\mathbf{x}_i, \mathbf{y}_i\}|\mathbf{A}, \mathscr{P}\right)$ is given by (C.1). $\Pr[\{\mathbf{x}_i\}]$ and $\Pr[\{\mathbf{y}_i\}]$ are the prior probabilities of latent coordinates given by the model, which we assume to be independent.

Other than the posterior probability, the CHL embedder is nearly identical to the HL embedder. As a result, the CHL embedder has the same computational complexity of $O\left(N^2\right)$, with N the number of nodes.

To test the stability of our complementarity representations, we obtained 20 independent complementarity embeddings of incomplete complementarity networks and evaluated their accuracies in link prediction experiments. As seen from Figure C.1, all independent complementarity representations result in similar link prediction scores narrowly distributed around the mean of AUPR $= 3.5 \cdot 10^{-3}$.

C.3. HIGHER-ORDER GENERALIZATIONS

The complementarity framework proposed in Chapter 4 assumes that each network node is mapped to two points in the latent space, each point corresponding to a unique feature

or characteristic of a node. Here, we sketch a possible generalization of the complementarity framework to the case of an arbitrary number of features per node.

We formulate such a generalization using the terminology of a collaboration network. Suppose we have N nodes, $i = 1, \ldots, N$, each of which is characterized by M different skills, $\mathbf{x}_{i,m}$, $m = 1, \ldots, M$. Any two nodes i and j are connected in the complementarity network if they jointly engage in at least one task k, $k = 1, \ldots, K$. Depending on the context, tasks can be either explicit, *e.g.*, scientific publications or problems in a collaboration network, or implicit (unobserved), *e.g.*, biological functions in molecular interaction networks. Thus, we introduce a collection of tasks, each of which is characterized by a set $\{\mathbf{y}_k\}$ of features. The basic building block is the probability $\Pr(\mathbf{x}_{i,m}, \mathbf{y}_k)$ that node i engages in task k using its skill $\mathbf{x}_{i,m}$. The probability for node i to engage in task k through any of its tasks independently is

$$q_{ik} = 1 - \prod_{m=1}^{M} \left[1 - \Pr(\mathbf{x}_{i,m}, \mathbf{y}_k)\right], \tag{C.2}$$

and the probability for any two nodes i and j to interact in the network is the probability that any two nodes co-engage in at least one task:

$$p_{ij} = 1 - \prod_{k=1}^{K} (1 - q_{ik} q_{jk}). \tag{C.3}$$

In case the probability $q_{ik} \ll 1$, the connection probability in (C.3) can be approximated to the leading order as

$$p_{ij} = \sum_{m=1}^{M} \sum_{n=1}^{M} g^{mn}(\mathbf{x}_{i,m}, \mathbf{x}_{j,n}), \tag{C.4}$$

$$g^{mn}(\mathbf{x}_{i,m}, \mathbf{x}_{j,n}) = \sum_{k=1}^{K} \Pr(\mathbf{x}_{i,m}, \mathbf{y}_k) \Pr(\mathbf{x}_{j,n}, \mathbf{y}_k). \tag{C.5}$$

As seen from (C.5), the general framework contains both similarity and complementarity components. In the $M = 1$ case, each network node has only one feature and the connection probability p_{ij} depends on the mutual match between features $\mathbf{x}_i^1$ and $\mathbf{x}_j^1$. The case of $M = 2$ contains both similarity, g^{11} and g^{22}, and complementarity components, g^{12} and g^{21}. We note that the functions $g^{mn}(\mathbf{x}, \mathbf{y})$ comprising the minimal complementarity framework in (C.4) and (C.5) are not necessarily geometric, and their functional form should be learned from the network of interest.

It is established that ordinary geometric networks, including random geometric graphs and random hyperbolic graphs, are characterized by a strong clustering coefficient due to the constraints imposed by the triangle inequality. Recent results from Krioukov [229] imply that strong clustering should be a sufficient condition for network geometricity. What are the network geometricity conditions for complementarity-driven networks remains an open question.

C.4. COMPLEMENTARITY AND BIPARTITE NETWORKS

Several parallels can be drawn between the complementarity-driven networks and bipartite (multipartite) networks. In a bipartite network, nodes are split into two classes

or domains, and links are possible only between nodes of different classes. The principled complementarity framework operates with agents and tasks: agents that are able to collaborate on tasks are complementary to each other. Therefore, synthetic networks generated by the principled complementarity framework are bipartite. In a one-mode projection, nodes of one domain of the bipartite network are removed. Nodes of the other domain are connected if they share at least one common neighbor in the bipartite network. Therefore, complementarity links between the agents in the principled complementarity framework can be viewed as one-mode projections of the full network onto the domain of agents. Coordinates of agents and tasks in the principled complementarity framework can be viewed as hidden variables, such that connections between agents and tasks are established independently with probabilities that are functions of the corresponding hidden variables. Therefore, the principled complementarity framework belongs to the class of bipartite networks with hidden variables [130].

Furthermore, one can think of a reduced complementarity framework if each agent is represented by one point either in $\mathscr{M}_1$ or $\mathscr{M}_2$ but not in both. The reduced complementarity framework may be instrumental in modeling, for example, scientific collaborations where each scientist has exactly one domain-specific expertise. If this is the case, the reduced complementarity model also belongs to the class of bipartite networks in latent spaces [139].

D

APPENDIX TO CHAPTER 6

D.1. DATA

ConceptNet [127] is a multilingual database in the form of a semantic network where the nodes are words and phrases from natural language. The links indicate in total 34 different semantic relations. The 'knowledge' is collected from a variety of resources, including crowdsourced resources, expert-created resources, and games with a purpose [127]. We study the semantic networks from ConceptNet as it is one of the richest semantic network resources available.

We study the networks belonging to in total 7 link types (relations), 6 of which are directly contained within ConceptNet. They are the 'Has-A', 'Part-Of', 'Is-A', 'Related-To', 'Antonym', and 'Synonym' relations. These are the relation types in ConceptNet that we deem the most meaningful and that also have a sufficient amount of data. In addition, we define an additional link type 'Union', which is the set union of the nodes and links of four networks: 'Has-A', 'Part-Of', 'Is-A', and 'Related-To'. The purpose of adding this link type is to treat all four relations equally and to evaluate how the structure of the whole network is different from the individual ones. The definitions of the six selected relations from ConceptNet and related examples are outlined in Table D.1. The links of some networks are directed, *i.e.*, of the 'Has-A', 'Part-Of', and 'Is-A' networks, but we treat all networks as undirected for simplicity of the analysis and also comparability. We remove nodes with phrases consisting of more than 5 words, as we deem these to be artifacts of the automated data extraction in ConceptNet.

D.2. SEMANTIC RELATIONS

Relation	*Description*	*Directed*	*Example*	*Creation*
Has-A	B belongs to A, either as an inherent part or due to a social construct of possession. Has-A is often the reverse of Part-Of.	Yes	bird → wing	Manual + Automatic
Part-Of	A is a part of B. This is the *part_meronym* relation in WordNet.	Yes	gearshift → car	Manual + Automatic
Is-A	A is a subtype or a specific instance of B; every A is a B. This can include specific instances; the distinction between subtypes and instances is often blurry in language. This is the *hyponym* relation in WordNet.	Yes	car → vehicle	Manual + Automatic
Related-To	The most general relation. There is some positive relationship between A and B, but ConceptNet can't determine what that relationship is based on the data.	No	learn ↔ erudition	Manual + Automatic
Antonym	A and B are opposites in some relevant way, such as being opposite ends of a scale, or fundamentally similar things with a key difference between them. Counterintuitively, two concepts must be quite similar before people consider them antonyms. This is the *antonym* relation in WordNet.	No	black ↔ white	Automatic
Synonym	A and B have very similar meanings. They may be translations of each other in different languages. This is the *synonym* relation in WordNet.	No	sunlight ↔ sunshine	Automatic

Table D.1: Definition of the six relations and related information from ConceptNet [210].

D.3. METHODS

D.3.1. DEGREE-PRESERVING NETWORK REWIRING

Degree-preserving network rewiring randomly rewires the links between nodes without changing the node degrees. To preserve the degrees of all nodes, we randomly select 1

link pair (4 nodes) and swap the endpoints of these 2 links. Figure D.1 illustrates the rewiring method. To make sure that all links are likely to be rewired at least once, we repeat the random selection of links for T times, where we choose $T = 4L$, four times the number of links. The pseudocode is provided in Algorithm 1.

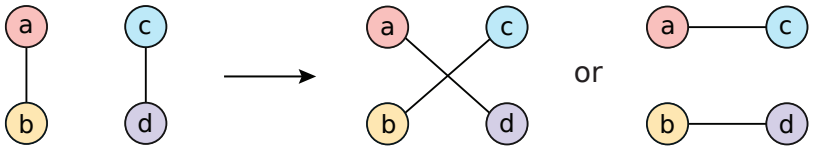


Figure D.1: Illustration of degree-preserving rewiring. By randomly swapping the endpoints of two links (a, b) and (c, d), new links can be constructed without changing the node degrees.

Algorithm 1: Degree-preserving network rewiring

```
Data: a list of links
Result: a rewired network
E ← a list of links;
T ← 4L ;                          /* all links are rewired at least once */
while T ≠ 0 do
    (a, b) and (c, d) ← randomly pick 2 links from E;
    n ← |set(a, b, c, d)| ;        /* number of unique nodes in 2 links */
    if n < 4 then
        continue
    else
        a and c ← randomly select one node from each link ;
        (c, b) and (a, d) ← swap the two selected nodes;
        if (c, b) ∈ E or (a, d) ∈ E then
            continue
        else
            E ← update the list of links with the 2 rewired links (c, b) and (a, d);
            T ← T − 1
        end
    end
end
```

LOGARITHMIC BINNING

To suppress noise at larger values of the degree k in the density $\Pr[D = k]$ of the degree distribution (the tail), we group the data in bins of equal logarithmic width. In linear binning, every bin has the same linear width $w = k_{i+1} - k_i$, while in logarithmic binning, the bins have constant logarithmic width b, where $b = \log(k_{i+1}) - \log(k_i)$ [230]. Thus, the linear bin width of a logarithmic bin, $w_i = k_{i+1} - k_i = k_i(e^b - 1)$, is proportional to k_i. The sizes of the logarithmic bins grow exponentially. Therefore, the number of observations x in a bin is equal to the density $f(k)$ in that bin times the width w of the bin.

SIMPLE POWER-LAW EXPONENT ESTIMATION

A common method for estimating the power-law exponent γ in $\Pr[D = k] \approx ck^{-\gamma}$ is to measure the slope of $\log(\Pr[D = k])$ versus $\log(k)$. Since the probability density $\Pr[D = k]$ of the degree is proportional to $k^{-\gamma}$, the number of observations $x \propto \Pr[D = k] \times w \propto k^{1-\gamma}$. Regressing $\log(x)$ against $\log(k)$ yields a slope equal to $1-\gamma$. To estimate γ, normalization of the number of observations x is required. Due to the increasing bin width, a bin might contain more than one value of k. The sum of all observations within a bin is x. To ensure that the total area under the density is equal to 1, the number of observations x is normalized by the linear width of the bin, thereby converting x to the number of observations per unit of the bin width, $(x/w) \propto k^{-\gamma}$. As a result, regressing the normalized logarithmic bin counts $\log(x/w)$ against the logarithmic degree $\log(k)$ yields a slope of $-\gamma$ [230]. We base our slope estimate of γ on the linear part of the tail in the density $\Pr[D = k]$, which we determine by inspection for each network.

D

CONSISTENT POWER-LAW EXPONENT ESTIMATORS

A more rigorous approach to power-law degree distributions is provided by Voitalov *et al.* [99], here we summarize the main conclusions for the convenience of the reader. They consider a degree distribution to be a power law if the probability density function is a member of the class of regularly varying functions: $\Pr[D = k] = \ell(k)k^{-\gamma}$, where $\ell(k)$ is a slowly varying function. The function $\ell(k)$ is called slowly varying if

$$\lim_{k\to\infty} \frac{\ell(ak)}{\ell(k)} = 1, \tag{D.1}$$

for any $a > 0$. This definition corresponds with a perfect power law in the tail of the distribution. Voitalov *et al.* [99] propose to use three different consistent estimators of the power-law exponent: the Hill, moments, and kernel estimators. We use the software package of [202] to obtain these estimates for the degree sequences of our networks. These estimators do not estimate γ directly, but rather the extreme value index

$$\xi = \frac{1}{\gamma - 1}. \tag{D.2}$$

We adopt the following rule of thumb from Voitalov *et al.* [99] to distinguish power-law distributions: the distribution is a power law if $\hat{\xi} > 1/4$ for all three estimators, corresponding with $\hat{\gamma} < 5$. In addition, they call a distribution *hardly power-law* if all $\hat{\xi} > 0$, but at least one $\hat{\xi} \leq 1/4$ ($\hat{\gamma} > 5$). If any $\hat{\xi} < 0$, the distribution is not a power law.

D.4. ENGLISH SEMANTIC NETWORKS

D.4.1. OVERVIEW

In Table D.2, we list basic descriptive statistics of the seven English semantic networks: the number of nodes N, the number of links L, the maximum degree $k_{\max}$ and the average degree $E[D]$.

Based on the number of nodes, the network 'Has-A' is the smallest ($N = 7,503$), and the network 'Union' is the largest ($N = 677,426$). The number of links L ranges from 5,421 to 1,819,646. Relative to the network sizes, all 7 networks have a small average

degree $E[D]$, ranging from 1.45 to 5.43. For instance, in the network 'Part-Of', a node has, on average, connections to only 2 (0.02%) out of the total of 11,839 nodes. In other words, the number of links is of the same order as the number of nodes, which indicates that semantic networks are sparse.

Network	***Has-A***	***Is-A***	***Part-Of***	***Related-To***	***Union***	***Antonym***	***Synonym***
N	7,503	152,538	11,839	592,816	677,426	16,867	166,922
L	5,421	220,589	12,003	1,610,452	1,819,646	14,371	155,048
$k_{\max}$	372	2913	116	4025	5263	38	103
$E[D]$	1.45	2.89	2.03	5.43	5.37	1.70	1.86

Table D.2: Basic statistics of the seven English semantic networks extracted from ConceptNet.

D

D.4.2. DESCRIPTIVE STATISTICS

Network	***Has-A***	***Is-A***	***Part-Of***	***Related-To***	***Union***	***Antonym***	***Synonym***
N	1,664	140,024	7,562	571,079	650,079	5,912	53,279
L	1,842	213,319	9,212	1,598,548	1,804,666	7,986	80,668
$k_{\max}$	198	2913	116	4025	5263	38	103
$E[D]$	2.21	3.05	2.44	5.60	5.55	2.70	3.03
ANND	33.6	242	14.1	170	219	6.77	7.13
$\hookrightarrow$ *rewired*	23.3	142	10.8	145	173	6.25	7.51
c_G	$2.2\cdot10^{-3}$	$5.7\cdot10^{-2}$	$4.6\cdot10^{-2}$	$1.0\cdot10^{-1}$	$1.0\cdot10^{-1}$	$1.5\cdot10^{-2}$	$1.1\cdot10^{-1}$
$\hookrightarrow$ *rewired*	$1.8\cdot10^{-2}$	$6.3\cdot10^{-3}$	$2.1\cdot10^{-3}$	$3.3\cdot10^{-3}$	$3.7\cdot10^{-3}$	$7.3\cdot10^{-4}$	$1.5\cdot10^{-4}$
$\hat{\gamma}^{Slope}$	2.3	2.3	2.4	2.4	2.4	×	×
$\hat{\gamma}^{Hill}$	2.3	2.3	2.5	2.3	2.3	×	×
$\hat{\gamma}^{Mom}$	2.5	2.3	2.6	2.2	2.2	×	×
$\hat{\gamma}^{Kern}$	2.6	2.3	2.7	2.1	2.1	×	×

Table D.3: Descriptive statistics of the LCCs of the seven English semantic networks extracted from ConceptNet. A cross (×) indicates the degree sequence of the corresponding network is hardly or no power-law.

Table D.3 shows the overall descriptive statistics of the largest connected components (LCCs) the English semantic networks: the number of nodes N, the number of links L, the maximum degree $k_{\max}$, the average degree $E[D]$, the average nearest neighbor degree (ANND), the graph clustering coefficient c_G and the estimated power-law exponents $\hat{\gamma}$. We rewire all semantic networks using the methods described before, after which the same statistics are calculated for the rewired networks.

For networks obtained by degree-preserving rewiring, only the ANND and the graph clustering coefficient c_G change. The average nearest neighbor degree ANND becomes smaller for all randomized semantic networks, except for the 'Synonym' network.

All networks except the 'Has-A' network have a remarkably larger graph clustering coefficient c_G (at least by an order of magnitude) than the randomized networks. Because in random networks links are randomly distributed, there are fewer triangles. On the contrary, the randomized networks of 'Has-A' exhibit a clustering coefficient more than seven times larger than their original networks.

In summary, we find universal properties in semantic networks across different languages: degree distribution, degree assortativity, clustering, sparsity, and connectedness. Most semantic networks have power-law degree distributions and most of them are scale-free networks. There are two types of degree mixing patterns in semantic networks: assortative and disassortative. Most networks have higher average clustering coefficients than expected by chance, except for one network, the network 'Has-A', which has a lower clustering coefficient. All semantic networks have high sparsity. Most networks have a single connected component containing the majority of the nodes, except for the network 'Has-A', which is more fragmented.

D.4.3. LARGEST CONNECTED COMPONENTS

Table D.4 shows the absolute and relative sizes of the LCCs of the seven English semantic networks and their rewired counterparts.

Network	N	***Size LCC***	***Percentage***
Has-A	7,503	1,664	22.18%
↪ *rewired*		2,416 ± 35	(32.20 ± 0.47)%
Is-A	152,538	140,024	91.80%
↪ *rewired*		127,258 ± 73	(83.43 ± 0.05)%
Part-Of	11,839	7,562	63.87%
↪ *rewired*		7,993 ± 53	(67.51 ± 0.45)%
Related-To	592,816	571,079	96.33%
↪ *rewired*		570,012 ± 116	(96.15 ± 0.02)%
Union	677,426	650,079	95.96%
↪ *rewired*		650,474 ± 182	(95.77 ± 0.03)%
Antonym	16,867	5,912	35.05%
↪ *rewired*		8,845 ± 59	(52.44 ± 0.35)%
Synonym	166,922	53,279	31.92%
↪ *rewired*		103,466 ± 142	(61.98 ± 0.09)%

Table D.4: Number of nodes in the LCCs of the seven English networks in the original and rewired networks. The LCC sizes of the rewired networks are each the average over 10 rewiring realizations with the standard deviation shown.

D.4.4. AVERAGE NEAREST NEIGHBOR DEGREE

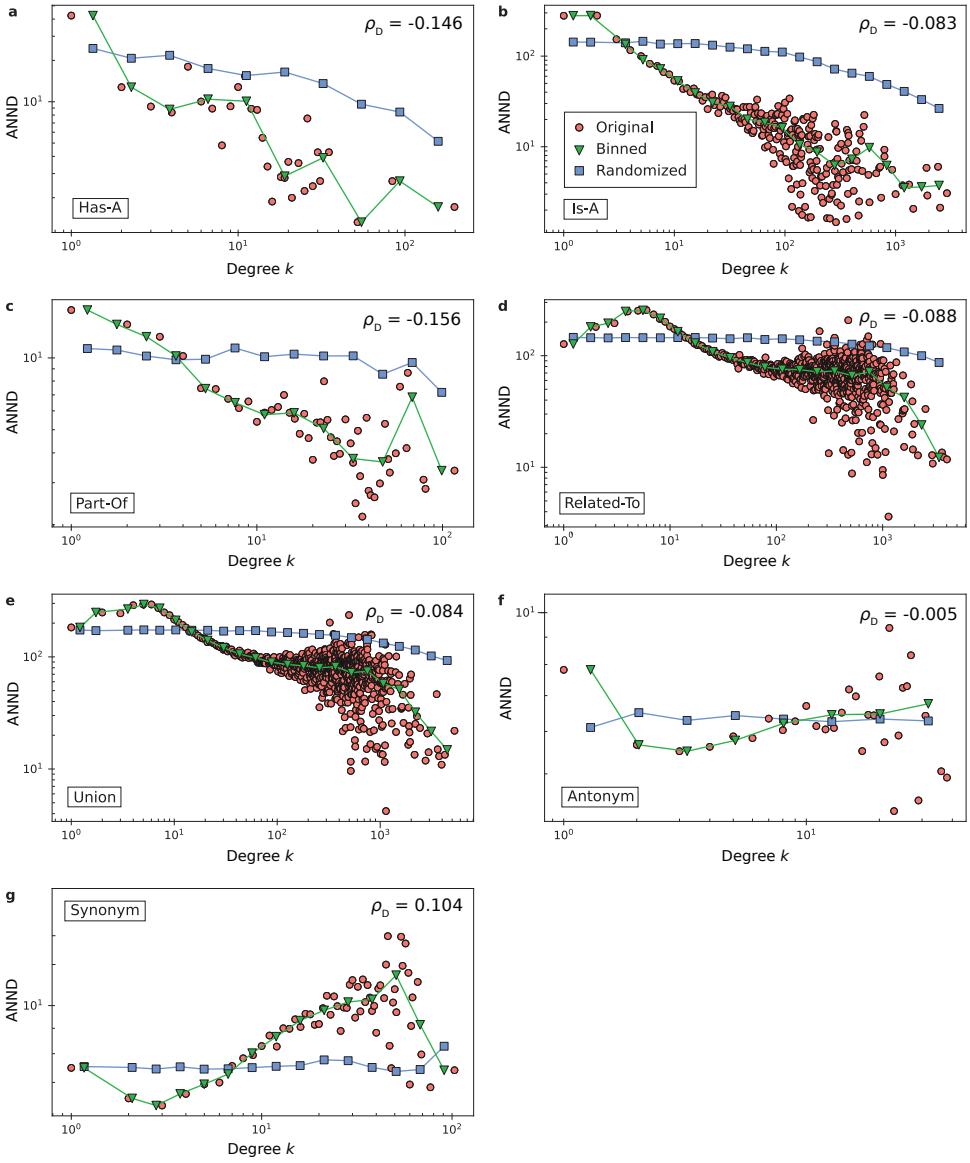


Figure D.2: Average nearest neighbor degree (ANND) as a function of the degree k and degree assortativity ρ_D of the seven English semantic networks. Panels correspond to networks (a) 'Has-A', (b) 'Is-A', (c) 'Part-Of', (d) 'Related-To', (e) 'Union', (f) 'Antonym', and (g) 'Synonym'. The circles are the original average ANND of nodes with degree k in a network, triangles represent the data after logarithmic binning, and squares are the average ANND of nodes with degree k in the randomized networks.

D.4.5. CLUSTERING COEFFICIENT

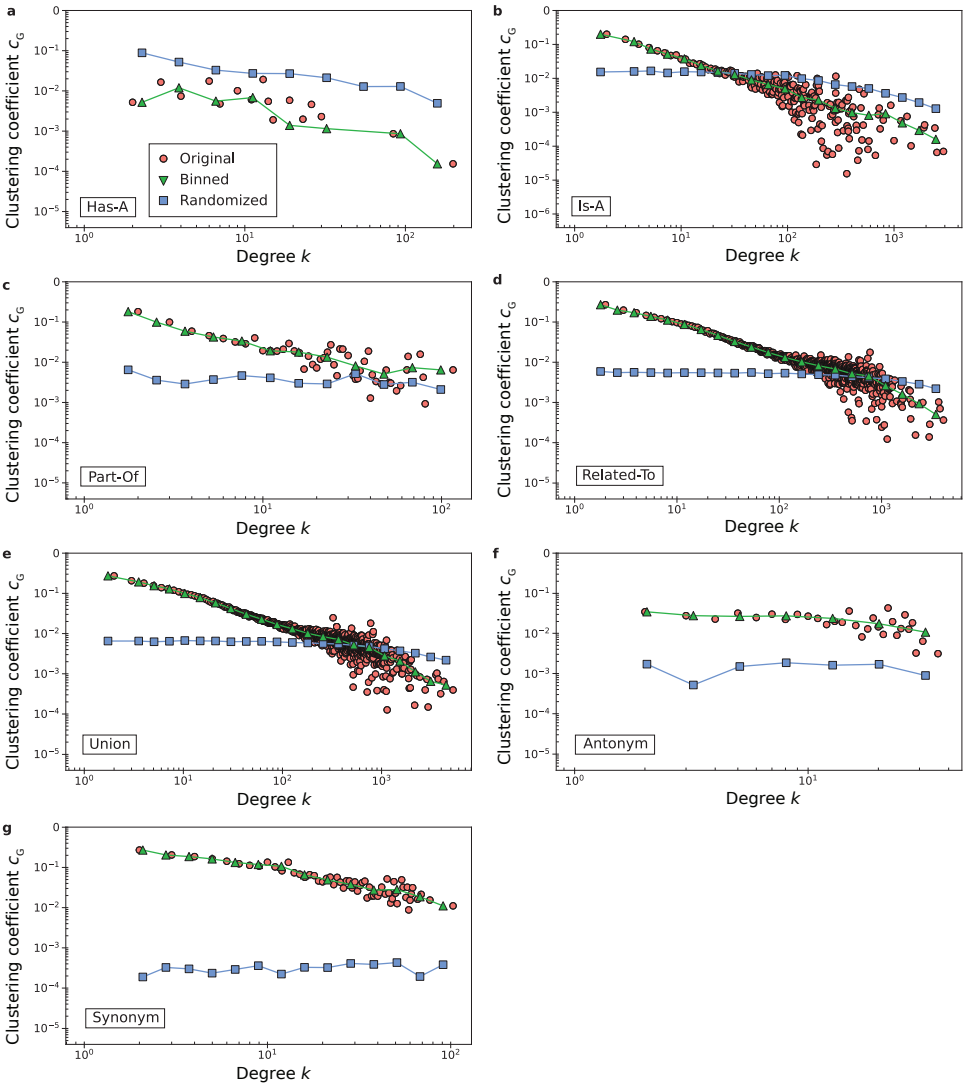


Figure D.3: The average clustering coefficient $c_G(i)$ of nodes i with degree k of the seven English semantic networks. Panels correspond to the networks (a) 'Has-A', (b) 'Is-A', (c) 'Part-Of', (d) 'Related-To', (e) 'Union', (f) 'Antonym', and (g) 'Synonym'. The circle data points are the original average local clustering coefficients of nodes, triangles represent data after logarithmic binning, and squares show the average clustering coefficient of nodes with degree k in the randomized networks (logarithmically binned).

D.4.6. CONNECTED COMPONENT SIZE DISTRIBUTION

Figure D.4 shows the size distribution of the connected components for each semantic network.

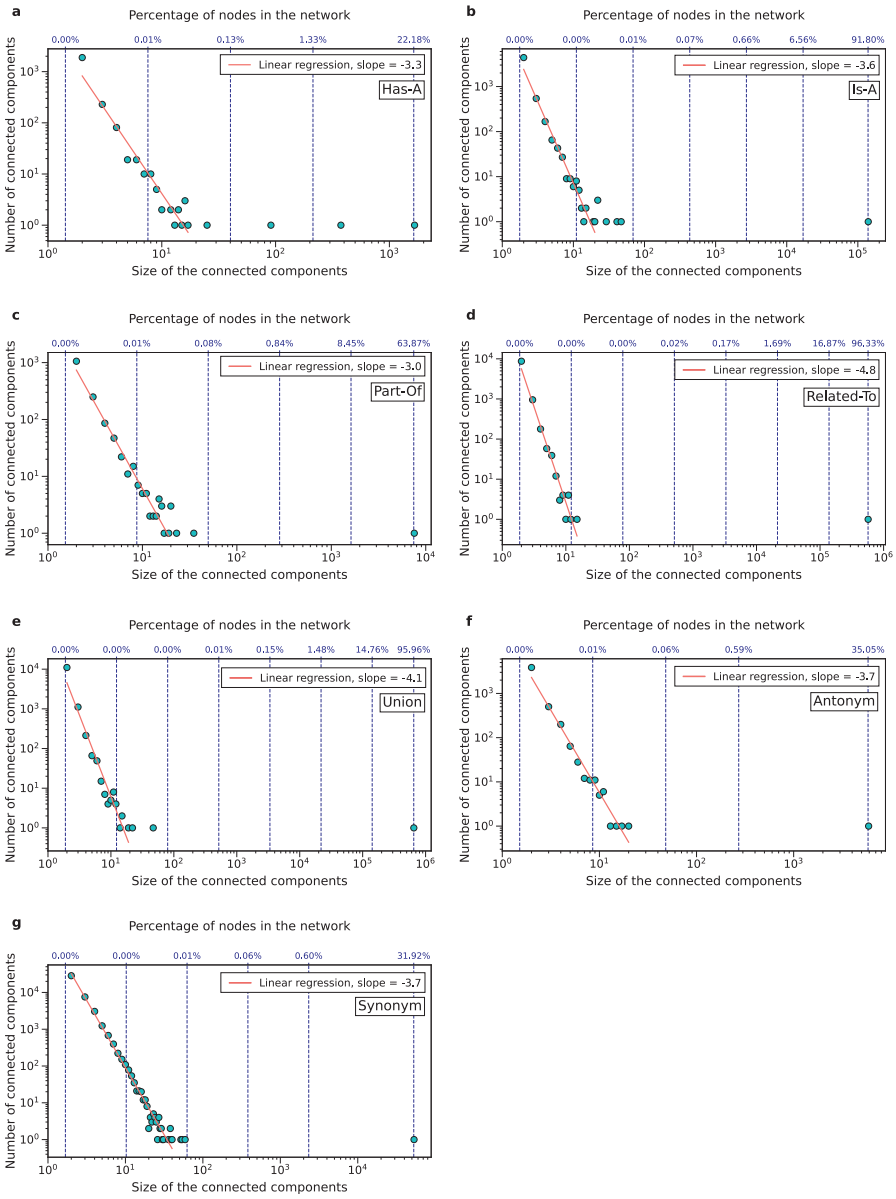


Figure D.4: Size distributions of connected components of the seven English semantic networks. Panels correspond to the networks (a) 'Has-A', (b) 'Is-A', (c) 'Part-Of', (d) 'Related-To', (e) 'Union', (f) 'Antonym', and (g) 'Synonym'. The dashed lines indicate the percentages of nodes in the connected components relative to the total. The regression line is fitted to the bulk of the data points.

D.5. MULTILINGUAL SEMANTIC NETWORKS

D.5.1. DESCRIPTIVE STATISTICS

This Section shows the descriptive statistics of semantic networks from the eleven languages. Each property is compared among the seven networks for the eleven languages.

LANGUAGE CLASSIFICATIONS

Table D.5 shows the typological and genetic classifications of the eleven considered languages.

Genetic \ Typological	***Inflecting***	***Isolating***	***Agglutinating***
Italic	Spanish, French, Italian, Portuguese		
Germanic	English, Dutch, German		
Balto-Slavic	Russian		
Transeurasian			Japanese
Sino-Tibetan		Chinese	
Uralic			Finnish

Table D.5: Genetic and typological language classifications of the eleven languages in this study.

DESCRIPTIVE STATISTICS

Table D.6 shows the number of nodes of each semantic network in the eleven different languages. A blank element in the table indicates that the network does not exist, *i.e.*, a relation is not available in that language.

Network	***Has-A***	***Is-A***	***Part-Of***	***Related-To***	***Union***	***Antonym***	***Synonym***
English	1,664	140,024	7,562	571,079	650,079	5,912	53,279
French		17,519	2,832	1,289,083	1,296,622	1,361	20,144
Italian		2,663	9	36,295	46,468	13	1,580
German		113,301	5	100,737	172,147	187	43,072
Spanish		255	11	12,094	22,861	15	3,491
Russian		557	3	20,268	25,887	12	1,148
Portuguese		3,341	15	5,929	11,426	17	6,421
Dutch		191	53	303	1,418	111	11,964
Japanese	38	40,256	7,230	7,200	43,286	20	230
Finnish		76	12	4,483	6,958	24	1,569
Chinese	6,355	10,073	3,417	3,163	17,128	4	17

Table D.6: Number of nodes N in the LCCs of the semantic networks from the eleven different languages extracted from ConceptNet. A blank element indicates the corresponding network is not available. The 'Union' network is the union of four networks ('Has-A', 'Is-A', 'Part-Of', and 'Related-To'). Because we display the sizes of the LCCs, for the 'Union' networks in some languages the number of nodes exceeds the sum of the sizes of its four constituent networks.

AVERAGE DEGREE

Network	***Has-A***	***Is-A***	***Part-Of***	***Related-To***	***Union***	***Antonym***	***Synonym***
English	2.21	3.05	2.44	5.60	5.55	2.70	3.03
French		2.64	2.51	3.44	3.46	2.45	2.81
Italian		2.86	2.89	2.20	2.27	1.85	2.54
German		2.75	1.60	4.77	4.53	2.16	3.57
Spanish		2.45	2.73	2.13	2.13	1.87	2.57
Russian		2.23	1.33	4.14	3.88	1.83	2.26
Portuguese		2.24	2.67	2.49	2.65	2.00	2.84
Dutch		4.68	4.98	2.30	2.69	2.11	3.53
Japanese	2.89	4.42	4.11	4.34	4.79	2.00	2.73
Finnish		1.97	1.83	2.30	2.26	1.92	2.24
Chinese	3.58	3.02	3.36	4.06	3.78	1.50	2.24

Table D.7: Average degree $E[D]$ in the LCCs of the semantic networks from the eleven different languages extracted from ConceptNet. A blank element indicates the corresponding network is unavailable.

POWER-LAW EXPONENTS

Table D.8 lists the estimated power-law exponents $\hat{\gamma}$ for each semantic network in the eleven languages. We consider a network to not have a power-law degree distribution if it is not or hardly power-law according to the method of Voitalov *et al.* [99].

Network	$\hat{\gamma}$	***Has-A***	***Is-A***	***Part-Of***	***Related-To***	***Union***	***Antonym***	***Synonym***
English	$\hat{\gamma}^{Slope}$	2.3	2.3	2.4	2.4	2.4	×	×
	$\hat{\gamma}^{Hill}$	2.3	2.3	2.5	2.3	2.3	×	×
	$\hat{\gamma}^{Mom}$	2.5	2.3	2.6	2.2	2.2	×	×
	$\hat{\gamma}^{Kern}$	2.6	2.3	2.7	2.1	2.1	×	×
French	$\hat{\gamma}^{Slope}$		2.4	2.3	×	×	2.7	3.1
	$\hat{\gamma}^{Hill}$		2.5	2.5	×	×	3.3	3.6
	$\hat{\gamma}^{Mom}$		2.5	2.6	×	×	4.5	3.9
	$\hat{\gamma}^{Kern}$		2.6	2.6	×	×	4.2	4.8
Italian	$\hat{\gamma}^{Slope}$		2.3		2.6	2.6		×
	$\hat{\gamma}^{Hill}$		2.8		2.4	2.4		×
	$\hat{\gamma}^{Mom}$		2.2		2.4	2.5		×
	$\hat{\gamma}^{Kern}$		2.3		2.5	2.6		×
German	$\hat{\gamma}^{Slope}$		2.5		2.6	2.5		3.1
	$\hat{\gamma}^{Hill}$		2.2		2.7	2.9		3.6
	$\hat{\gamma}^{Mom}$		3.3		2.7	2.9		3.7
	$\hat{\gamma}^{Kern}$		2.3		2.9	2.6		3.9
Spanish	$\hat{\gamma}^{Slope}$				×	×		×
	$\hat{\gamma}^{Hill}$				×	×		×
	$\hat{\gamma}^{Mom}$				×	×		×
	$\hat{\gamma}^{Kern}$				×	×		×
Russian	$\hat{\gamma}^{Slope}$				×	×		×
	$\hat{\gamma}^{Hill}$				×	×		×
	$\hat{\gamma}^{Mom}$				×	×		×
	$\hat{\gamma}^{Kern}$				×	×		×
Portuguese	$\hat{\gamma}^{Slope}$		2.6		2.4	2.5		×
	$\hat{\gamma}^{Hill}$		2.8		2.6	2.8		×
	$\hat{\gamma}^{Mom}$		2.6		2.1	2.4		×
	$\hat{\gamma}^{Kern}$		2.6		2.9	2.7		×
Dutch	$\hat{\gamma}^{Slope}$					2.2		×
	$\hat{\gamma}^{Hill}$					2.8		×
	$\hat{\gamma}^{Mom}$					3.1		×
	$\hat{\gamma}^{Kern}$					3.5		×
Japanese	$\hat{\gamma}^{Slope}$		2.4	2.3	2.2	2.3		
	$\hat{\gamma}^{Hill}$		2.6	2.9	4.9	2.6		
	$\hat{\gamma}^{Mom}$		2.6	2.9	2.4	2.7		
	$\hat{\gamma}^{Kern}$		2.6	2.6	2.6	2.6		
Finnish	$\hat{\gamma}^{Slope}$				×	×		×
	$\hat{\gamma}^{Hill}$				×	×		×
	$\hat{\gamma}^{Mom}$				×	×		×
	$\hat{\gamma}^{Kern}$				×	×		×
Chinese	$\hat{\gamma}^{Slope}$	2.5	2.3	2.7	1.9	2.3		
	$\hat{\gamma}^{Hill}$	3.4	2.4	2.3	2.7	4.3		
	$\hat{\gamma}^{Mom}$	3.8	2.4	2.4	1.9	2.4		
	$\hat{\gamma}^{Kern}$	2.7	2.3	2.3	2.3	2.5		

Table D.8: Estimated power-law exponents $\hat{\gamma}$ for the LCCs of the semantic networks in different languages. A blank element indicates the corresponding network is either unavailable or the number of nodes $N < 1000$. A cross (×) indicates that the degree sequence of that network is not or hardly power-law.

D.6. DEGREE DISTRIBUTION PEAKS

D.6.1. SPANISH 'RELATED-TO' PEAK WORDS

Peak word	***Translation***	***Neighbors***
cenar	to dine	cená, cenábamos, cenáculo, cenáis, cenáramos, ...
viajar	to travel	viaja, viajaba, viajabais, viajaban, viajabas, viajad, ...
pasear	to walk	pasea, paseaba, paseabais, paseaban, paseabas, ...
reparar	to repair	repararais, repararan, repararas, repararеis, ...
comparar	to compare	comprar, comparaba, comparabais, comparaban, ...

Table D.9: Examples of words in the peak and their neighoring words in the Spanish 'Related-To' network.

D.6.2. POS TAGS OF PEAK WORDS

Percentage (%)	**French**		**Spanish**		**Portuguese**		**Finnish**	
	LCC	Peak	LCC	Peak	LCC	Peak	LCC	Peak
POS tagged	98.71	98.66	92.72	77.84	67.60	60.00	81.37	64.13
Verb	68.90	89.97	87.62	98.44	32.56	100.00	11.40	11.36
Noun	19.21	7.14	9.20	1.56	51.96	0	77.96	84.09
Adjective	11.53	2.75	2.89	0	14.60	0	7.17	4.55
Adverb	0.36	0.15	0.29	0	0.88	0	3.47	0

Table D.10: Percentages of POS tags among peak words and in the LCCs of the 'Related-To' networks of four inflecting languages.

D.6.3. POS TAGS OF PEAK WORD NEIGHBORS

Percentage (%)	**French**		**Spanish**		**Portuguese**		**Finnish**	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
POS tagged	97.39	0.88	96.96	1.73	97.74	0.75	93.72	4.45
Verb	87.26	25.85	97.24	2.59	99.23	0.94	3.86	14.64
Noun	9.34	20.08	2.07	2.15	0.77	0.94	89.67	26.50

Table D.11: The mean and Standard Deviation (SD) percentage of verbs and nouns in the neighbors of peak words of the LCC of network 'Related-To' in four inflecting languages.

D.6.4. NODE MERGING PROCEDURE

The network 'Form-Of' is extracted in the same way as the other networks. Then, we merge words in the 'Related-To' network with their inflected forms according to 'Form-Of' and treat them as single nodes. Next, we calculate the number of nodes with degree k in the new 'Related-To' network. Finally, we plot the densities $\Pr[D = k]$ of the degree distributions of French, Spanish, Portuguese, and Finnish networks.

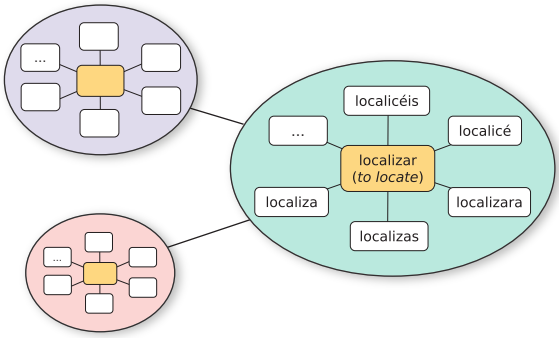


Figure D.5: Illustration of the merging of words and their inflected forms in the 'Related-To' network through the 'Form-Of' network. After merging a root word and its neighbors, all words in a circle are seen as a single word.

D

D.6.5. COVERAGE OF PEAK WORDS IN THE 'FORM-OF' NETWORK

Nodes	French	Spanish	Portuguese	Finnish
Peak words	33.72%	100%	60.00%	91.30%
Neighbors of peak words	17.38%	97.76%	55.47%	45.08%

Table D.12: The percentages of peak words and their neighbors from the LCCs of the 'Related-To' networks that are covered in the 'Form-Of' network.

D.6.6. THE NUMBER OF GRAMMATICAL VARIATIONS

Language	m	$k_{\min}$	$k_{\max}$
French	42	36	51
Spanish	54	45	61
Portuguese	54	53	53
Finnish	30	25	35

Table D.13: The maximum number of grammatical variations m for the grammatical rule of interest in French, Spanish, Portuguese, and Finnish. The minimum and maximum degree $k_{\min}$ and $k_{\max}$ where the peak starts and ends in the densities of the degree distributions of the 'Related-To' networks are included for comparison.

D.7. SIMILARITY AND COMPLEMENTARITY COEFFICIENTS

For the convenience of the reader, here we summarize the main components of the framework for computing structural similarity and complementarity coefficients by Talaga and Nowak [129].

STRUCTURAL COEFFICIENTS

Similarity-based networks are rich in triangles because of the triangle closure principle. The clustering coefficient is a classic measure of the density of triangles in a network.

However, we cannot simply compare the number of triangles and quadrangles between two networks, because these networks have different sizes and degree distributions. We need to reliably calculate the statistics of triangles and quadrangles of a network to quantify similarity and complementarity. To this end, we rely on recent work on detecting motifs of complementarity [129]. The structural similarity coefficient is a weighted average of two clustering coefficients based on head and wedge triples, Figures D.6(b-c). Analogous to the clustering coefficient, we can use structural complementarity measures based on quadrangle closure rules, Figure D.6(d). Similarly, the structural complementarity coefficient is a weighted average of two coefficients based on head and wedge quadrangles, Figures D.6(f-e). Here, we summarize the main components of the procedures for calculating the structural similarity coefficient and complementarity coefficient of a network G.

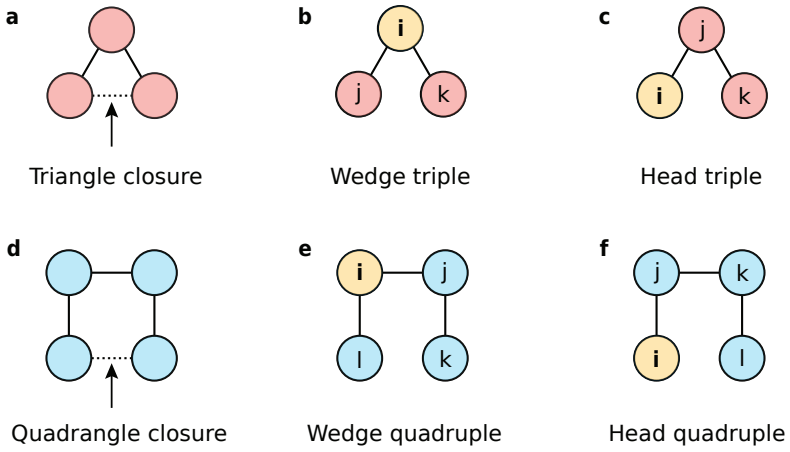


Figure D.6: Quadrangle and quadruples in comparison with triangle and triples. Wedge and head triples (or quadruples) are different with respect to the location of the node i. In a wedge triple (b), node i is located in the middle, while in a head triple (c), node i is located at the head. Similarly, in a wedge quadruple (e), node i is located at the second location, while node i is located at the head of a head quadruple (f).

STRUCTURAL SIMILARITY COEFFICIENT

The structural similarity coefficient s_i generalizes the local clustering and closure coefficients. The local clustering coefficient s_i^W of a node i is the classic clustering coefficient. It is defined as the fraction of triples centered at i which can be closed to form a triangle,

$$s_i^W = \frac{2T_i}{t_i^W} = \frac{\sum_{j,k} a_{ij} a_{ik} a_{jk}}{k_i (k_i - 1)}, \tag{D.3}$$

where T_i is the number of triangles including i and t_i^W is the number of wedge triples, Figure D.6(b), or 2-paths with node i in the middle, *e.g.*, (j, i, k). The definition of the local closure coefficient [231] is given as follows

$$s_i^H = \frac{2T_i}{t_i^H} = \frac{\sum_{j,k} a_{ij} a_{ik} a_{jk}}{\sum_j a_{ij} (k_j - 1)}, \tag{D.4}$$

where t_i^H is the number of head triples, Figure D.6(c), *i.e.*, 2-paths starting from node i, such as (i, j, k). Both s_i^W and s_i^H are bounded in the range $[0, 1]$, but they capture different parts of the spectrum of similarity-driven structures [129].

Combining the weighted average of these two coefficients results in a more comprehensive measure of local structure, the *structural similarity coefficient* [129], which captures the full spectrum of structural similarity. It is defined as

$$s_i = \frac{4T_i}{t_i^W + t_i^H} = \frac{t_i^W s_i^W + t_i^H s_i^H}{t_i^W + t_i^H}. \tag{D.5}$$

The coefficient $s_i = 1$ only if node i is in a fully connected network.

The structural similarity coefficient of a whole network G is then the average over all nodes

$$s(G) = \frac{1}{N}\sum_{i=1}^{N} s_i. \tag{D.6}$$

D

STRUCTURAL COMPLEMENTARITY COEFFICIENT

Analogously, the local quadruples clustering coefficient at node i is defined as the fraction of closed quadruples with i at the second position [129]

$$c_i^W = \frac{2Q_i}{q_i^W} = \frac{\sum_{j\neq i} a_{ij} \sum_{k\neq i,j} a_{ik}\left(1-a_{jk}\right) \sum_{l\neq i,j,k} a_{kl} a_{jl}\left(1-a_{il}\right)}{\sum_j a_{ij}\left[\left(k_i-1\right)\left(k_j-1\right)-n_{ij}\right]}, \tag{D.7}$$

where Q_i represents the number of quadrangles contain that node i and q_i^W is the number of wedge quadruples Figure D.6(e), or 3-paths with i at the second node, *e.g.*, (l, i, j, k). Similarly, the local quadruples closure coefficient of a node i calculates the percentage of closed quadruples beginning at i

$$c_i^H = \frac{2Q_i}{q_i^H} = \frac{\sum_{j\neq i} a_{ij} \sum_{k\neq i,j} a_{ik}\left(1-a_{jk}\right) \sum_{l\neq i,j,k} a_{kl} a_{jl}\left(1-a_{il}\right)}{\sum_{j\neq i} a_{ij} \sum_{k\neq i,j} a_{jk}\left(k_k-1-a_{ik}\right)}, \tag{D.8}$$

where q_i^H is the number of head quadruples originating from node i, Figure D.6(f).

Finally, the *structural complementarity coefficient* is constructed as the weighted average of the local quadruples clustering and closure coefficients [129]

$$c_i = \frac{4Q_i}{q_i^W + q_i^H} = \frac{q_i^W c_i^W + q_i^H c_i^H}{q_i^W + q_i^H}. \tag{D.9}$$

The structural complementarity coefficient $c_i \in [0, 1]$, which is proven to be a more general measure than using only c_i^W or c_i^H [129]. The maximum $c_i = 1$ happens only if node i belongs to a fully connected bipartite graph. In a bipartite graph, nodes are divided into two groups, and connections are only formed between groups but not within the same group.

The structural complementarity coefficient of a whole network G is then the average of all nodes:

$$c(G) = \frac{1}{N}\sum_{i=1}^{N} c_i. \tag{D.10}$$

D.7.1. CALIBRATION

This Section presents the calibration procedure using the configuration model used to normalize the structural coefficients of semantic networks.

UNDIRECTED BINARY CONFIGURATION MODEL

We utilize the Undirected Binary Configuration Model (UBCM) [232] to calibrate the structural coefficients. The UBCM generates a maximum entropy probability distribution over the nodes of an undirected and unweighted network with the constraints of an expected degree sequence. The resulting maximum entropy distributions are maximally unbiased with respect to any other property [233].

CALIBRATION OF STRUCTURAL COEFFICIENTS

First of all, we calculate one structural coefficient (similarity or complementarity) of a given network G. We denote this coefficient as $x(G)$. Second, we sample R randomized copies G_i's of the given network from the configuration model. Then, we calculate the structural coefficient $x(G_i)$ for each sampled network. At last, we take the average log ratio of $x(G)$ and $x(G_i)$'s. As a result, the calibrated coefficient $\mathscr{C}_G(x)$ based on R samples from the configuration model is obtained as follows [129]

$$\mathscr{C}(x)_G = \frac{1}{R}\sum_{i=1}^{R} \log \frac{x(G)}{x(G_i)}. \tag{D.11}$$

The calibrated structural coefficient can be less than, equal to, or larger than zero. Consider the calibrated structural similarity coefficient $\mathscr{C}_G(s)$ for example:

- $\mathscr{C}_G(s) < 0$, the structural similarity coefficient $s(G)$ is smaller than $s(G_i)$ of random networks.
- $\mathscr{C}_G(s) = 0$, the structural similarity coefficient is comparable to the ones in random networks.
- $\mathscr{C}_G(s) > 0$, the structural similarity coefficient is larger than in random networks.

We do not compute the structural coefficients for networks that have less than 100 nodes, because likely there exist no triangles or quadrangles in the sampled networks, and the structural coefficient $x(G_i) = 0$, in that case. When $x(G) = 0$ or $x(G_i) = 0$, the calibrated coefficient $\mathscr{C}(x)_G$ (D.11) is undefined.

Table D.14 lists the procedures for computing the structural similarity and complementarity coefficients to quantify the density of triangles and quadrangles in a network G, respectively.

Since the runtime of the algorithm depends on the size of a network and the choice of the number of randomized networks R, we do not compute the structural coefficients for the two largest networks, the French 'Related-To' and 'Union' networks with $N > 1,200,000$ each, as the computation time is infeasible for the resources we have at hand. We use $R = 500$ for most networks, while for the remaining two largest networks, the English 'Related-To' and 'Union', we set $R = 100$ to avoid the long computation time.

Procedure	Structural coefficients	Network G	
		Similarity ($\triangle$)	**Complementarity** ($\square$)
Step 1	Wedge triple/quadruple	s_i^W (D.3)	c_i^W (D.7)
	Head triple/quadruple	s_i^H (D.4)	c_i^H (D.8)
Step 2	Node-wise	s_i (D.5)	c_i (D.9)
Step 3	Network-wise	$s(G) = \frac{1}{N}\sum_{i=1}^{N} s_i$ (D.6)	$c(G) = \frac{1}{N}\sum_{i=1}^{N} c_i$ (D.10)
Step 4	Calibrated Network-wise	$\mathscr{C}(s)_G = \frac{1}{R}\sum_{i=1}^{R} \log\frac{s(G)}{s(G_i)}$	$\mathscr{C}(c)_G = \frac{1}{R}\sum_{i=1}^{R} \log\frac{c(G)}{c(G_i)}$

Table D.14: The procedure of calculating the structural similarity coefficient and complementarity coefficient of a network G. The calibrated structural coefficients in step 4 are obtained by taking the average log ratio of a network-wise coefficient over the coefficients of the sampled networks G_i (D.11).

BIBLIOGRAPHY

[1] A.-L. Barabási. Network science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1987):20120375, 2013.

[2] M. Newman. *Networks*. Oxford university press, 2018.

[3] P. Van Mieghem. *Performance analysis of complex networks and systems*. Cambridge University Press, Cambridge, United Kingdom, 1st edition, 2014.

[4] P. Van Mieghem. *Graph spectra for complex networks*. Cambridge University Press, 2010.

[5] P. Van Mieghem, J. Omic, and R. Kooij. Virus spread in networks. *IEEE/ACM Transactions on Networking*, 17(1):1–14, 2 2009.

[6] S. Fortunato. Community detection in graphs. *Physics Reports*, 486(3-5):75–174, 2010.

[7] S. Fortunato and M. E. J. Newman. 20 years of network community detection. *Nature Physics*, 18(8):848–850, 8 2022.

[8] I. Jokić and P. Van Mieghem. Linear clustering process on networks. *IEEE Transactions on Network Science and Engineering*, pages 1–10, 2023.

[9] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69(2):26113, 2004.

[10] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, 2008.

[11] H.-W. Shen and X.-Q. Cheng. Spectral methods for the detection of network community structure: A comparative analysis. *Journal of Statistical Mechanics: Theory and Experiment*, 2010(10):P10020, 2010.

[12] M. E. J. Newman. Fast algorithm for detecting community structure in networks. *Physical Review E*, 69(6 2), 2004.

[13] M. E. J. Newman and G. Reinert. Estimating the number of communities in a network. *Physical Review Letters*, 117(7):78301, 2016.

[14] K. Chen and J. Lei. Network cross-validation for determining the number of communities in network data. *Journal of the American Statistical Association*, 113(521):241–251, 2018.

[15] E. Côme and P. Latouche. Model selection and clustering in stochastic block models based on the exact integrated complete data likelihood. *Statistical Modelling*, 15(6):564–589, 2015.

[16] M. T. Schaub, J.-C. Delvenne, M. Rosvall, and R. Lambiotte. The many facets of community detection in complex networks. *Applied Network Science*, 2(1):1–13, 2017.

[17] P. W. Holland, K. B. Laskey, and S. Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109–137, 1983.

[18] E. Abbe and C. Sandon. Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 670–688. IEEE, 10 2015.

[19] B. Karrer and M. E. J. Newman. Stochastic blockmodels and community structure in networks. *Physical Review E*, 83(1):016107, 1 2011.

[20] A. Lancichinetti, S. Fortunato, and F. Radicchi. Benchmark graphs for testing community detection algorithms. *Physical Review E*, 78(4):46110, 2008.

[21] D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat, and M. Boguná. Hyperbolic geometry of complex networks. *Physical Review E*, 82(3):36106, 2010.

[22] M. Á. Serrano, D. Krioukov, and M. Boguñá. Self-similarity of complex networks and hidden metric spaces. *Physical Review Letters*, 100(7):78701, 2 2008.

[23] F. Papadopoulos, M. Kitsak, M. Á. Serrano, M. Boguñá, and D. Krioukov. Popularity versus similarity in growing networks. *Nature*, 489(7417):537–540, 9 2012.

[24] K. Zuev, M. Boguñá, G. Bianconi, and D. Krioukov. Emergence of soft communities from geometric preferential attachment. *Scientific Reports*, 5(1):9421, 4 2015.

[25] M. Zheng, G. García-Pérez, M. Boguñá, and M. Á. Serrano. Scaling up real networks by geometric branching growth. *Proceedings of the National Academy of Sciences*, 118(21), 5 2021.

[26] E. Candellero and N. Fountoulakis. Clustering and the hyperbolic geometry of complex networks. *Internet Mathematics*, 12(1-2):2–53, 3 2016.

[27] L. Gugelmann, K. Panagiotou, and U. Peter. Random hyperbolic graphs: degree sequence and clustering. In *Automata, Languages, and Programming: 39th International Colloquium, ICALP 2012, Warwick, UK, July 9-13, 2012, Proceedings, Part II 39*, 5 2012.

[28] M. Boguná, F. Papadopoulos, and D. Krioukov. Sustaining the Internet with hyperbolic mapping. *Nature Communications*, 1(1):1–8, 2010.

[29] M. Kitsak, I. Voitalov, and D. Krioukov. Link prediction with hyperbolic geometry. *Physical Review Research*, 2(4):043113, 10 2020.

[30] G. García-Pérez, A. Allard, M. Á. Serrano, and M. Boguñá. Mercator: Uncovering faithful hyperbolic embeddings of complex networks. *New Journal of Physics*, 21(12):123033, 2019.

[31] M. Boguñá, D. Krioukov, and K. C. Claffy. Navigability of complex networks. *Nature Physics*, 5(1):74–80, 2009.

[32] A. Gulyás, J. J. Bíró, A. Kőrösi, G. Rétvári, and D. Krioukov. Navigable networks as Nash equilibria of navigation games. *Nature Communications*, 6(1):7651, 7 2015.

[33] I. Voitalov, R. Aldecoa, L. Wang, and D. Krioukov. Geohyperbolic routing and addressing schemes. *ACM SIGCOMM Computer Communication Review*, 47(3):11–18, 9 2017.

[34] E. Ortiz, M. Starnini, and M. Á. Serrano. Navigability of temporal networks in hyperbolic space. *Scientific Reports*, 7(1):15054, 2017.

[35] G. García-Pérez, M. Boguñá, and M. Á. Serrano. Multiscale unfolding of real networks by geometric renormalization. *Nature Physics*, 14(6):583–589, 6 2018.

[36] A. Muscoloni and C. V. Cannistraci. Navigability evaluation of complex networks by greedy routing efficiency. *Proceedings of the National Academy of Sciences*, 116(5):1468–1469, 2019.

[37] M. Kitsak *et al.* Finding shortest and nearly shortest path nodes in large substantially incomplete networks by hyperbolic mapping. *Nature Communications*, 14(1):186, 1 2023.

[38] M. Á. Serrano, M. Boguñá, and F. Sagués. Uncovering the hidden geometry behind metabolic networks. *Molecular BioSystems*, 8(3):843, 2012.

[39] F. Papadopoulos, C. Psomas, and D. Krioukov. Network mapping by replaying hyperbolic growth. *IEEE/ACM Transactions on Networking*, 23(1):198–211, 2015.

[40] F. Papadopoulos, R. Aldecoa, and D. Krioukov. Network geometry inference using common neighbors. *Physical Review E*, 92(2):22807, 2015.

[41] A. Muscoloni, J. M. Thomas, S. Ciucci, G. Bianconi, and C. V. Cannistraci. Machine learning meets complex networks via coalescent embedding in the hyperbolic space. *Nature Communications*, 8(1):1615, 12 2017.

[42] A. Muscoloni and C. V. Cannistraci. Leveraging the nonuniform PSO network model as a benchmark for performance evaluation in community detection and link prediction. *New Journal of Physics*, 20(6), 2018.

[43] A. Muscoloni and C. V. Cannistraci. Minimum curvilinear automata with similarity attachment for network embedding and link prediction in the hyperbolic space. *arXiv preprint arXiv:1802.01183*, 2018.

[44] G. García-Pérez, R. Aliakbarisani, A. Ghasemi, and M. Á. Serrano. Precision as a measure of predictability of missing links in real networks. *Physical Review E*, 101(5):52318, 5 2020.

[45] M. Boguñá, I. Bonamassa, M. De Domenico, S. Havlin, D. Krioukov, and M. Á. Serrano. Network geometry. *Nature Reviews Physics*, 3(2):114–135, 2 2021.

[46] E. Cambria and B. White. Jumping NLP curves: A review of natural language processing research. *IEEE Computational Intelligence Magazine*, 9(2):48–57, 2014.

[47] M. R. Quillian. Word concepts: A theory and simulation of some basic semantic capabilities. *Behavioral Sciences*, 12(5):410–430, 1967.

[48] J. F. Sowa. Semantic networks. *Encyclopedia of Cognitive Science*, 2012.

[49] A. Singhal. Introducing the Knowledge Graph: Things, not strings. Google Blog https://blog.google/products/search/introducing-knowledge-graph-things-not, 5 2012.

[50] M. Nickel and D. Kiela. Poincaré embeddings for learning hierarchical representations. *Advances in Neural Information Processing Systems*, 5 2017.

[51] M. Le, S. Roller, L. Papaxanthos, D. Kiela, and M. Nickel. Inferring concept hierarchies from text corpora via hyperbolic embeddings. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3231–3241, Stroudsburg, PA, USA, 2019. Association for Computational Linguistics.

[52] V. Kepuska and G. Bohouta. Next-generation of virtual personal assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home). In *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*, pages 99–103, 2018.

[53] R. High. The era of cognitive systems: An inside look at IBM Watson and how it works. *IBM Corporation, Redbooks*, 1:16, 2012.

[54] A. Vaswani *et al.* Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 2017.

[55] L. Ouyang *et al.* Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35:27730–27744, 2022.

[56] M. Steyvers and J. B. Tenenbaum. The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science*, 29(1):41–78, 2005.

[57] J. Borge-Holthoefer and A. Arenas. Semantic networks: Structure and dynamics. *Entropy*, 12(5):1264–1302, 2010.

[58] G. Budel and P. Van Mieghem. Detecting the number of clusters in a network. *Journal of Complex Networks*, 8(6), 3 2021.

[59] The MathWorks Inc. MATLAB version 9.5.0 (R2018b), 2017.

[60] M. E. J. Newman. Finding community structure in networks using the eigenvectors of matrices. *Physical Review E*, 74(3):36104, 2006.

[61] M. E. J. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582, 2006.

[62] F. Krzakala *et al.* Spectral redemption in clustering sparse networks. *Proceedings of the National Academy of Sciences*, 110(52):20935–20940, 2013.

[63] P. Van Mieghem, X. Ge, P. Schumm, S. Trajanovski, and H. Wang. Spectral graph analysis of modularity and assortativity. *Physical Review E*, 82(5):56113, 2010.

[64] S. Chauhan, M. Girvan, and E. Ott. Spectral properties of networks with community structure. *Physical Review E*, 80(5):56114, 2009.

[65] F. Chung, L. Lu, and V. Vu. Eigenvalues of random power law graphs. *Annals of Combinatorics*, 7(1):21–33, 2003.

[66] L. Wu, X. Ying, X. Wu, and Z.-H. Zhou. Line orthogonality in adjacency eigenspace with application to community partition. In *Twenty-Second International Joint Conference on Artificial Intelligence*, 2011.

[67] J. H. Wilkinson. *The algebraic eigenvalue problem*. Oxford University Press, 1965.

[68] G. W. Stewart and J.-G. Sun. *Matrix perturbation theory*. Associated Press, 1990.

[69] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová. Inference and phase transitions in the detection of modules in sparse networks. *Physical Review Letters*, 107(6):65701, 2011.

[70] J. M. Shea and J. P. Macker. Automatic selection of number of clusters in networks using relative eigenvalue quality. In *MILCOM 2013-2013 IEEE Military Communications Conference*, pages 131–136, 2013.

[71] M. A. Riolo, G. T. Cantwell, G. Reinert, and M. E. J. Newman. Efficient method for estimating the number of communities in a network. *Physical Review E*, 96(3):32310, 2017.

[72] T. P. Peixoto. Nonparametric Bayesian inference of the microcanonical stochastic block model. *Physical Review E*, 95(1):12317, 2017.

[73] K.-i. Hashimoto. Zeta functions of finite graphs and representations of p-adic groups. In *Automorphic forms and geometry of arithmetic varieties*, pages 211–280. Elsevier, 1989.

[74] S. Fortunato and M. Barthelemy. Resolution limit in community detection. *Proceedings of the National Academy of Sciences*, 104(1):36–41, 2007.

[75] W. W. Zachary. An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, 33(4):452–473, 1977.

[76] J. Leskovec and J. J. Mcauley. Learning to discover social circles in ego networks. In *Advances in Neural Information Processing Systems*, 2012.

[77] M. E. J. Newman. Finding community structure in networks using the eigenvectors of matrices. *Physical Review E*, 74(3):36104, 2006.

[78] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data*, 1(1), 2007.

[79] G. Budel, M. Kitsak, R. Aldecoa, K. Zuev, and D. Krioukov. Random hyperbolic graphs in d+1 dimensions. *arXiv preprint arXiv:2010.12303*, 2024.

[80] D. Krioukov, F. Papadopoulos, A. Vahdat, and M. Boguñá. Curvature and temperature of complex networks. *Physical Review E*, 80(3):35101, 9 2009.

[81] D. D. McFarland and D. J. Brown. Social distance as a metric: A systematic introduction to smallest space analysis. In *Bonds of pluralism: form and substance of urban social networks*, volume 6, pages 213–252. ed. E.O. Laumann, New York: Wiley, New York, 1973.

[82] P. D. Hoff, A. E. Raftery, and M. S. Handcock. Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460):1090–1098, 2002.

[83] M. Boguñá, D. Krioukov, P. Almagro, and M. Á. Serrano. Small worlds and clustering in spatial networks. *Physical Review Research*, 2(2):23040, 4 2020.

[84] M. Nickel and D. Kiela. Learning continuous hierarchies in the Lorentz model of hyperbolic geometry. In *35th International Conference on Machine Learning PMLR, 2018*, 2018.

[85] B. Dhingra, C. Shallue, M. Norouzi, A. Dai, and G. Dahl. Embedding text in hyperbolic spaces. In *Proceedings of the Twelfth Workshop on Graph-Based Methods for Natural Language Processing (TextGraphs-12)*, pages 59–69, Stroudsburg, PA, USA, 2018. Association for Computational Linguistics.

[86] A. T. Ifrea, G. Bécigneul, and O. E. Ganea. Poincaré Glove: Hyperbolic word embeddings. In *7th International Conference on Learning Representations (ICLR) 2019*, 2019.

[87] W. Gu, A. Tandon, Y.-Y. Ahn, and F. Radicchi. Principled approach to the selection of the embedding dimension of networks. *Nature Communications*, 12(1):3772, 12 2021.

[88] B. Désy, P. Desrosiers, and A. Allard. Dimension matters when modeling network communities in hyperbolic spaces. *arXiv preprint arXiv:2209.09201*, 2022.

[89] K. Bringmann, R. Keusch, and J. Lengler. Geometric inhomogeneous random graphs. *Theoretical Computer Science*, 760:35–54, 2 2019.

[90] M. Boguñá and D. Krioukov. Navigating ultrasmall worlds in ultrashort time. *Physical Review Letters*, 102(5):58701, 2 2009.

[91] P. Almagro, M. Boguñá, and M. Á. Serrano. Detecting the ultra low dimensionality of real networks. *Nature Communications*, 13(1):6096, 10 2022.

[92] T. Friedrich, A. Göbel, M. Katzmann, and L. Schiller. Cliques in high-dimensional geometric inhomogeneous random graphs. *arXiv preprint arXiv:2302.04113*, 2023.

[93] W. Yang and D. Rideout. High dimensional hyperbolic geometry of complex networks. *Mathematics*, 8(11):1861, 10 2020.

[94] B. Kovács, S. G. Balogh, and G. Palla. Generalised popularity-similarity optimisation model for growing hyperbolic networks beyond two dimensions. *Scientific Reports*, 12(1):968, 1 2022.

[95] R. Aldecoa, C. Orsini, and D. Krioukov. Hyperbolic graph generator. *Computer Physics Communications*, 196:492–496, 11 2015.

[96] M. Boguñá and R. Pastor-Satorras. Class of correlated random networks with hidden variables. *Physical Review E*, 68(3):36112, 2003.

[97] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 6 1998.

[98] N. Fountoulakis, P. van der Hoorn, T. Müller, and M. Schepers. Clustering in a hyperbolic model of complex networks. *Electronic Journal of Probability*, 26, 1 2021.

[99] I. Voitalov, P. van der Hoorn, R. van der Hofstad, and D. Krioukov. Scale-free networks well done. *Physical Review Research*, 1(3):33034, 10 2019.

[100] J. van der Kolk, M. Á. Serrano, and M. Boguñá. An anomalous topological phase transition in spatial random graphs. *Communications Physics*, 5(1):245, 10 2022.

[101] R. D. Lord. The distribution of distance in a hypersphere. *The Annals of Mathematical Statistics*, 25(4):794–798, 1954.

[102] P. van der Hoorn, G. Lippner, and D. Krioukov. Sparse maximum-entropy random graphs with a given power-law degree distribution. *Journal of Statistical Physics*, 173(3-4):806–844, 11 2018.

[103] G. Budel and M. Kitsak. The RHG generator. https://bitbucket.org/gbudel/rhg-generator/, 2022.

[104] M. Galassi and others. GNU scientific library reference manual (3rd ed.), 2021.

[105] G. Budel and M. Kitsak. Complementarity in complex networks. *arXiv preprint arXiv:2003.06665*, 2024.

[106] A. J. McCoy, V. Chandana Epa, and P. M. Colman. Electrostatic complementarity at protein/protein interfaces. *Journal of Molecular Biology*, 268(2):570–584, 5 1997.

[107] B. A. Lewis and D. M. Engelman. Lipid bilayer thickness varies linearly with acyl chain length in fluid phosphatidylcholine vesicles. *Journal of Molecular Biology*, 166(2):211–217, 5 1983.

[108] Y. Yano and K. Matsuzaki. Measurement of thermodynamic parameters for hydrophobic mismatch 1: Self-association of a transmembrane helix. *Biochemistry*, 45(10):3370–3378, 3 2006.

[109] A. V. Botelho, T. Huber, T. P. Sakmar, and M. F. Brown. Curvature and hydrophobic forces drive oligomerization and modulate activity of rhodopsin in membranes. *Biophysical Journal*, 91(12):4464–4477, 12 2006.

[110] Q. Zhang, M. Sanner, and A. J. Olson. Shape complementarity of protein-protein complexes at multiple resolutions. *Proteins: Structure, Function, and Bioinformatics*, 75(2):453–467, 5 2009.

[111] Y. Li, X. Zhang, and D. Cao. The role of shape complementarity in the protein-protein interactions. *Scientific Reports*, 3(1):3271, 11 2013.

[112] C. E. S. Mattsson *et al.* Functional structure in production networks. *Frontiers in Big Data*, 4, 5 2021.

[113] C. V. Cannistraci, G. Alanis-Lobato, and T. Ravasi. Minimum curvilinearity to enhance topological prediction of protein interactions by network embedding. *Bioinformatics*, 29(13):i199–i209, 7 2013.

[114] B. Yang, W.-t. Yih, X. He, J. Gao, and L. Deng. Embedding entities and relations for learning and inference in knowledge bases. *3rd Proceedings of the International Conference on Learning Representations (ICLR) 2015*, 12 2015.

[115] H. Xiao, M. Huang, and X. Zhu. From one point to a manifold: Knowledge graph embedding for precise link prediction. *International Joint Conferences on Artificial Intelligence (IJCAI)*, 12 2015.

[116] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei. LINE. In *Proceedings of the 24th International Conference on World Wide Web*, pages 1067–1077, Republic and Canton of Geneva, Switzerland, 5 2015. International World Wide Web Conferences Steering Committee.

[117] A. Grover and J. Leskovec. node2vec. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 855–864, New York, NY, USA, 8 2016. ACM.

[118] L. Zhu, D. Guo, J. Yin, G. V. Steeg, and A. Galstyan. Scalable temporal latent space inference for link prediction in dynamic social networks. *IEEE Transactions on Knowledge and Data Engineering*, 28(10):2765–2777, 10 2016.

[119] S. M. Kazemi and D. Poole. SimplE embedding for link prediction in knowledge graphs. *Advances in Neural Information Processing Systems*, 2 2018.

[120] M. Brew Anthony; Salter-Townshend. A latent space mapping for link prediction. *Neural Information Processing Systems*, pages 1–7, 2010.

[121] M. E. J. Newman and T. P. Peixoto. Generalized communities in networks. *Physical Review Letters*, 115(8), 2015.

[122] L. Yang, X. Cao, D. Jin, X. Wang, and D. Meng. A unified semi-supervised community detection framework using latent space graph regularization. *IEEE Transactions on Cybernetics*, 45(11):2585–2598, 11 2015.

[123] D. K. Sewell and Y. Chen. Latent space approaches to community detection in dynamic networks. *Bayesian Analysis*, 12(2), 6 2017.

[124] J. Kleinberg. Complex networks and decentralized search algorithms. In *Proceedings of the International Congress of Mathematicians Madrid, August 22–30, 2006*, pages 1019–1044. European Mathematical Society Publishing House, Zuerich, Switzerland, 2006.

[125] S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Shenker. A scalable content-addressable network. *ACM SIGCOMM Computer Communication Review*, 31(4):161–172, 10 2001.

[126] J. Risson and T. Moors. Survey of research towards robust peer-to-peer networks: Search methods. *Computer Networks*, 50(17):3485–3521, 12 2006.

[127] R. Speer, J. Chin, and C. Havasi. ConceptNet 5.5: An open multilingual graph of general knowledge. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, AAAI'17, pages 4444–4451. AAAI Press, 2017.

[128] I. A. Kovács *et al.* Network-based prediction of protein interactions. *Nature Communications*, 10(1):1240, 12 2019.

[129] S. Talaga and A. Nowak. Structural measures of similarity and complementarity in complex networks. *Scientific Reports*, 12(1):16580, 2022.

[130] M. Kitsak and D. Krioukov. Hidden variables in bipartite networks. *Physical Review E*, 84(2):26114, 8 2011.

[131] R. Sinatra, P. Deville, M. Szell, D. Wang, and A.-L. Barabási. A century of physics. *Nature Physics*, 11(10):791–796, 10 2015.

[132] R. K. Pan, S. Sinha, K. Kaski, and J. Saramäki. The evolution of interdisciplinarity in physics research. *Scientific Reports*, 2(1):551, 8 2012.

[133] A. J. Gates, Q. Ke, O. Varol, and A.-L. Barabási. Nature's reach: narrow work has broad impact. *Nature*, 575(7781):32–34, 11 2019.

[134] J. Pennington, R. Socher, and C. Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Stroudsburg, PA, USA, 2014. Association for Computational Linguistics.

[135] R. Brochier, A. Guille, and J. Velcin. Global vectors for node representations. In *WWW '19: The World Wide Web Conference*, pages 2587–2593, New York, NY, USA, 5 2019. ACM.

[136] F. W. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark. *NIST handbook of mathematical functions*. Cambridge University Press, Cambridge, United Kingdom, 1st edition, 2010.

[137] A. Iamnitchi, M. Ripeanu, and I. Foster. Small-world file-sharing communities. In *IEEE INFOCOM 2004*, volume 2, pages 952–963. IEEE, 2004.

[138] E. Burgos, H. Ceva, L. Hernández, R. P. J. Perazzo, M. Devoto, and D. Medan. Two classes of bipartite networks: Nested biological and social systems. *Physical Review E*, 78(4):46113, 10 2008.

[139] M. Kitsak, F. Papadopoulos, and D. Krioukov. Latent geometry of bipartite networks. *Physical Review E*, 95(3):32309, 3 2017.

[140] J. Das and H. Yu. HINT: High-quality protein interactomes and their applications in understanding human disease. *BMC Systems Biology*, 6(1):92, 2012.

[141] K. Luck *et al.* A reference map of the human binary protein interactome. *Nature*, 580(7803):402–408, 4 2020.

[142] H. Yu *et al.* High-quality binary protein interaction map of the yeast interactome network. *Science*, 322(5898):104–110, 10 2008.

[143] J. A. Dunne, R. J. Williams, and N. D. Martinez. Food-web structure and network theory: The role of connectance and size. *Proceedings of the National Academy of Sciences*, 99(20):12917–12922, 10 2002.

[144] J. Kunegis. KONECT. In *WWW 2013 Companion - Proceedings of the 22nd International Conference on World Wide Web*, 2013.

[145] P. Van Mieghem, H. Wang, X. Ge, S. Tang, and F. A. Kuipers. Influence of assortativity and degree-preserving rewiring on the spectra of networks. *The European Physical Journal B*, 76(4):643–652, 8 2010.

[146] B. Perozzi, R. Al-Rfou, and S. Skiena. DeepWalk. In *Proceedings 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 701–710, New York, NY, USA, 8 2014. ACM.

[147] T. Zhou, L. Lü, and Y.-C. Zhang. Predicting missing links via local information. *The European Physical Journal B*, 71(4):623–630, 10 2009.

[148] L. A. Adamic and E. Adar. Friends and neighbors on the Web. *Social Networks*, 25(3):211–230, 7 2003.

[149] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 10 1999.

[150] L. Lü, L. Pan, T. Zhou, Y.-C. Zhang, and H. E. Stanley. Toward link predictability of complex networks. *Proceedings of the National Academy of Sciences*, 112(8):2325–2330, 2 2015.

[151] L. Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18(1):39–43, 3 1953.

[152] C. V. Cannistraci, G. Alanis-Lobato, and T. Ravasi. From link-prediction in brain connectomes and protein interactomes to the local-community-paradigm in complex networks. *Scientific Reports*, 3(1):1613, 4 2013.

[153] P. Jaccard. Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bulletin de la Société Vaudoise des Sciences Naturelles*, 37(Jan. 1901):547–579, 1901.

[154] D. Liben-Nowell and J. Kleinberg. The link prediction problem for social networks. In *Proceedings of the Twelfth International Conference on Information and Knowledge Management*, pages 556–559, New York, NY, USA, 11 2003. ACM.

[155] S. D. Ghiassian, J. Menche, and A.-L. Barabási. A DIseAse MOdule Detection (DIAMOnD) algorithm derived from a systematic analysis of connectivity patterns of disease proteins in the human interactome. *PLOS Computational Biology*, 11(4):e1004120, 4 2015.

[156] Y.-Y. Ahn, J. P. Bagrow, and S. Lehmann. Link communities reveal multiscale complexity in networks. *Nature*, 466(7307):761–764, 8 2010.

[157] H. Mahmoud, F. Masulli, S. Rovetta, and G. Russo. Community detection in protein-protein interaction networks using spectral and graph approaches. In *Lecture Notes Computer Science*, volume 8452 LNBI, pages 62–75. Springer Verlag, 2014.

[158] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, 6 2002.

[159] J. Menche *et al.* Uncovering disease-disease relationships through the incomplete interactome. *Science*, 347(6224):1257601, 2 2015.

[160] A. R. Sonawane, S. T. Weiss, K. Glass, and A. Sharma. Network medicine in the age of biomedical big data. *Frontiers in Genetics*, 10, 4 2019.

[161] R. Noldus and P. Van Mieghem. Assortativity in complex networks. *Journal of Complex Networks*, 3(4):507–542, 12 2015.

[162] G. Budel, Y. Jin, P. Van Mieghem, and M. Kitsak. Topological properties and organizing principles of semantic networks. *Scientific Reports*, 13(1):11728, 7 2023.

[163] J. Dall and M. Christensen. Random geometric graphs. *Physical Review E*, 66(1):016121, 7 2002.

[164] R. D. Luce and A. D. Perry. A method of matrix analysis of group structure. *Psychometrika*, 14(2):95–116, 6 1949.

[165] M. E. J. Newman. Clustering and preferential attachment in growing networks. *Physical Review E*, 64(2):25102, 2001.

[166] P. Zhang, J. Wang, X. Li, M. Li, Z. Di, and Y. Fan. Clustering coefficient and community structure of bipartite networks. *Physica A: Statistical Mechanics and its Applications*, 387(27):6869–6875, 12 2008.

[167] W. Ellens, F. Spieksma, P. Van Mieghem, A. Jamakovic, and R. Kooij. Effective graph resistance. *Linear Algebra and its Applications*, 435(10):2491–2506, 11 2011.

[168] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1):47–97, 1 2002.

[169] M. E. J. Newman. Mixing patterns in networks. *Physical Review E*, 67(2):026126, 2 2003.

[170] M. R. Quillian. The teachable language comprehender: A simulation program and theory of language. *Communications of the ACM*, 12(8):459–476, 1969.

[171] H. B. de Barros Pereira *et al.* Systematic review of the "semantic network" definitions. *Expert Systems with Applications*, page 118455, 2022.

[172] J. F. Sowa. *Principles of semantic networks: Explorations in the representation of knowledge*. Morgan Kaufmann, 2014.

[173] S. Peters and H. E. Shrobe. Using semantic networks for knowledge representation in an intelligent environment. In *Proceedings of the First IEEE International Conference on Pervasive Computing and Communications, 2003 (PerCom 2003)*, pages 323–329, 2003.

[174] A.-B. M. Salem and M. Alfonse. Ontology versus semantic networks for medical knowledge representation. *Recent Advances In Computer Engineering*, pages 769–774, 2008.

[175] R. Popping. Knowledge graphs and network text analysis. *Social Science Information*, 42(1):91–106, 2003.

[176] D. Fensel *et al.* Introduction: What is a knowledge graph? In *Knowledge Graphs*, pages 1–10. Springer, 2020.

[177] J. Piskorski and R. Yangarber. Information extraction: Past, present and future. In *Multi-source, multilingual information extraction and summarization*, pages 23–49. Springer, 2013.

[178] F. Shi, L. Chen, J. Han, and P. Childs. A data-driven text mining and semantic network analysis for design information retrieval. *Journal of Mechanical Design*, 139(11), 2017.

[179] P. Resnik. Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language. *Journal of Artificial Intelligence Research*, 11:95–130, 1999.

[180] A. Harris. Human languages vs. programming languages. Medium https://medium.com/@anaharris/human-languages-vs-programming-languages-c89410f13252, 11 2018.

[181] M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, pages 415–444, 2001.

[182] G. Kossinets and D. J. Watts. Origins of homophily in an evolving social network. *American Journal of Sociology*, 115(2):405–450, 2009.

[183] D. R. Schaefer, J. M. Light, R. A. Fabes, L. D. Hanish, and C. L. Martin. Fundamental principles of network formation among preschool children. *Social Networks*, 32(1):61–71, 2010.

[184] T. A. B. Snijders. Statistical models for social networks. *Annual Review of Sociology*, 37:131–153, 2011.

[185] M. A. Hasan and M. J. Zaki. A survey of link prediction in social networks. In *Social Network Data Analytics*, pages 243–275. Springer, 2011.

[186] F. D. Zarandi and M. K. Rafsanjani. Community detection in complex networks using structural similarity. *Physica A: Statistical Mechanics and its Applications*, 503:882–891, 2018.

[187] N. Evans and S. C. Levinson. The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, 32(5):429–448, 2009.

[188] D. Khurana, A. Koli, K. Khatter, and S. Singh. Natural language processing: State of the art, current trends and challenges. *Multimedia Tools and Applications*, pages 1–32, 2022.

[189] H. M. Alfawareh and S. Jusoh. Resolving ambiguous entity through context knowledge and fuzzy approach. *International Journal of Computational Science and Engineering*, 3(1):410–422, 2011.

[190] S. Jusoh. A study on NLP applications and ambiguity problems. *Journal of Theoretical and Applied Information Technology*, 96(6), 2018.

[191] D. L. Nelson, C. L. McEvoy, and T. A. Schreiber. The University of South Florida word association, rhyme, and word fragment norms. http://w3.usf.edu/FreeAssociation, 1998.

[192] R. Ferrer-i Cancho and R. V. Solé. The small world of human language. *Proceedings of the Royal Society B: Biological Sciences*, 268(1482):2261–2265, 2001.

[193] BNC Consortium. British National Corpus, XML edition. http://hdl.handle.net/20.500.12024/2554, 2007.

[194] A. E. Motter, A. P. S. De Moura, Y.-C. Lai, and P. Dasgupta. Topology of the conceptual network of language. *Physical Review E*, 65(6):65102, 2002.

[195] G. Ward. *Moby thesaurus list*. Quality Classics, 2015.

[196] M. Sigman and G. A. Cecchi. Global organization of the WordNet lexicon. *Proceedings of the National Academy of Sciences*, 99(3):1742–1747, 2002.

[197] G. A. Miller. WordNet: A lexical database for English. *Communications of the ACM*, 38(11):39–41, 1995.

[198] P. M. Roget. *Roget's thesaurus of English words and phrases*. TY Crowell Company, 1911.

[199] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[200] S. H. Strogatz. Exploring complex networks. *Nature*, 410(6825):268–276, 2001.

[201] R. Ferrer-i Cancho, R. V. Solé, and R. Köhler. Patterns in syntactic dependency networks. *Physical Review E*, 69(5):51915, 2004.

[202] I. Voitalov. Tail index estimation for degree sequences of complex networks. https://github.com/ivanvoitalov/tailestimation, 2019.

[203] A. M. Petersen, J. N. Tenenbaum, S. Havlin, H. E. Stanley, and M. Perc. Languages cool as they expand: Allometric scaling and the decreasing need for new words. *Scientific Reports*, 2(1):943, 2012.

[204] L. A. Adamic and B. A. Huberman. Power-law distribution of the world wide web. *Science*, 287(5461):2115, 2000.

[205] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási. The large-scale organization of metabolic networks. *Nature*, 407(6804):651–654, 2000.

[206] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. *ACM SIGCOMM Computer Communication Review*, 29(4):251–262, 1999.

[207] J. Lyons. Language classification. Encyclopædia Britannica https://www.britannica.com/science/linguistics/Other-relationships, 2022.

[208] D. M. Eberhard, G. F. Simons, and C. D. Fennig. Ethnologue: Languages of the world. SIL International https://www.ethnologue.com, 2022.

[209] M. Haspelmath. The morph as a minimal linguistic form. *Morphology*, 30(2):117–134, 2020.

[210] R. Speer. Relations in ConceptNet5. ConceptNet 5 Wiki https://github.com/commonsense/conceptnet5/wiki/Relations, 6 2019.

[211] B. Comrie. *Aspect: An introduction to the study of verbal aspect and related problems*, volume 2. Cambridge University Press, 1976.

[212] C. Kendris and T. Kendris. *501 Spanish verbs*. Barrons Educational Series, 2020.

[213] F. J. Vare. Your all-in-one guide to the 18 Spanish tenses and moods. Enux Education, FluentU https://www.fluentu.com/blog/spanish/spanish-tenses/, 7 2022.

[214] J. J. Nitti and M. J. Ferreira. *501 Portuguese verbs*. Simon and Schuster, 2015.

[215] L. K. Lawless. *The everything French verb book: A handy reference for mastering verb conjugation*. Simon and Schuster, 2005.

[216] F. Karlsson. *Finnish: A comprehensive grammar*. Routledge, 2017.

[217] M. T. Rivera, S. B. Soderstrom, and B. Uzzi. Dynamics of dyads in social networks: Assortative, relational, and proximity mechanisms. *Annual Review of Sociology*, 36:91–115, 2010.

[218] A. Asikainen, G. Iñiguez, J. Ureña-Carrión, K. Kaski, and M. Kivelä. Cumulative effects of triadic closure and homophily in social networks. *Science Advances*, 6(19):eaax7310, 2020.

[219] M. Jia, B. Gabrys, and K. Musial. Measuring quadrangle formation in complex networks. *IEEE Transactions on Network Science and Engineering*, 9(2):538–551, 2021.

[220] J. H. Y. Tai. Chinese classifier systems and human categorization. In honor of William S.-Y. Wang: Interdisciplinary studies on language and language change, Pyramid Press Taipei, 1994.

[221] L. L.-S. Cheng and R. Sybesma. Yi-wan tang, yi-ge tang: Classifiers and massifiers. *Tsing Hua Journal of Chinese Studies*, 28(3):385–412, 1998.

[222] C. Bordenave, M. Lelarge, and L. Massoulié. Non-backtracking spectrum of random graphs: Community detection and non-regular Ramanujan graphs. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 1347–1357, 2015.

[223] H. M. Stark and A. A. Terras. Zeta functions of finite graphs and coverings. *Advances in Mathematics*, 121(1):124–165, 1996.

[224] O. Angel, J. Friedman, and S. Hoory. The non-backtracking spectrum of the universal cover of a graph. *Transactions of the American Mathematical Society*, 367(6):4287–4318, 2015.

[225] G. H. Golub and C. F. Van Loan. *Matrix computations*. Johns Hopkins University Press, 1996.

[226] R. B. Lehoucq, D. C. Sorensen, and C. Yang. *ARPACK user's guide: Solution of large-scale eigenvalue problems with implicitly restarted Arnoldi methods*. SIAM, 1998.

[227] C. Stark, B.-J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz, and M. Tyers. BioGRID: a general repository for interaction datasets. *Nucleic Acids Research*, 34(suppl_1):D535–D539, 2006.

[228] M. Kitsak and I. Voitalov. HyperLink. https://bitbucket.org/dk-lab/2020_code_hyperlink/, 2020.

[229] D. Krioukov. Clustering implies geometry in networks. *Physical Review Letters*, 116(20), 2016.

[230] E. White, B. Enquist, and J. Green. On estimating the exponent of power-law frequency distributions. *Ecology*, 89:905–912, 9 2008.

[231] H. Yin, A. R. Benson, and J. Leskovec. The local closure coefficient: A new perspective on network clustering. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, pages 303–311, 2019.

[232] N. Vallarano *et al.* Fast and scalable likelihood maximization for exponential random graph models with local constraints. *Scientific Reports*, 11(1):1–33, 2021.

[233] T. Squartini, R. Mastrandrea, and D. Garlaschelli. Unbiased sampling of network ensembles. *New Journal of Physics*, 17(2):23052, 2015.

LIST OF PUBLICATIONS

PEER-REVIEWED JOURNAL PUBLICATIONS

7. **Budel, G.**, F. Frasincar and D. Boekestijn (2024). DBHC: Discrete Bayesian HMM clustering. *International Journal of Machine Learning & Cybernetics*, 1–16.

6. **Budel, G.**, Y. Jin, P. Van Mieghem and M. Kitsak (2023). Topological properties and organizing principles of semantic networks. *Scientific Reports* 13(1), 11728.

5. **Budel, G.** and P. Van Mieghem (2021). Detecting the number of clusters in a network. *Journal of Complex Networks*, 8(6), cnaa047.

PEER-REVIEWED CONFERENCE PUBLICATIONS

4. **Budel, G.**, L. Hoogenboom, W. Kastrop, N. Reniers and F. Frasincar (2018). Predicting user flight preferences in an airline E-shop. In: Mikkonen T., Klamma R., Hernández J. (eds) Web Engineering. ICWE 2018. Lecture Notes in Computer Science, vol 10845. Springer, Cham.

UNDER SUBMISSION

3. **Budel, G.**, M. Kitsak, R. Aldecoa, K. Zuev and D. Krioukov (2024). Random Hyperbolic Graphs in $d+1$ Dimensions. *arXiv preprint arXiv:2010.12303*.

WORKING PAPERS

2. **Budel, G.**, P. Van Mieghem and M. Kitsak (2024). Complementarity random graph model.

1. **Budel, G.** and M. Kitsak (2024). Complementarity in complex networks. *arXiv preprint arXiv:2003.06665*.