Investigating the modeling assumptions of alert-driven attack graphs

A cognitive load-based quantification approach of interpretability in attack graphs

Vlad-Mihai Constantinescu

Supervisors: Sicco Verwer, Azqa Nadeem

EEMCS, Delft University of Technology, The Netherlands

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
June 25, 2023

Name of the student: Vlad-Mihai Constantinescu
Final project course: CSE3000 Research Project
Thesis committee: Sicco Verwer, Azqa Nadeem, Asterios Katsifodimos

Abstract

The interpretability of an attack graph is a key principle as it reflects the difficulty of a specialist to take insights into attacker strategies. However, the quantification of interpretability is considered to be a subjective manner and complex attack graphs can be challenging to read and interpret. In this research paper, we propose a new metric for quantifying the interpretability of attack graphs, aiming for comparable results between attack graphs regardless of the chosen drawing configuration or generation method. We address the gap in existing metrics by combining elements from the theory of cognitive chunks of information and user-experience-related fields to measure interpretability in terms of cognitive load. Our metric leverages Gestalt principles to formalize the quantification of interpretability based on cognitive overload. Compared to a similar approach, the proposed metric reveals a high level of similarity with the baseline, however, qualitative analysis revealed the proposed metric eliminates certain discrepancies with the expert's opinion that the baseline metric presented. Furthermore, a use case of the metric is presented and we evaluate our metric by comparing attack graphs generated using different methods, such as deterministic finite automaton (S-PDFA), Markov chain, and suffix tree. Finally, further work is proposed toward the goal of completing the metric by incorporating the remaining Gestalt principles.

Keywords: attack graphs, interpretability, cognitive load, cybersecurity, network security

1 Introduction

Security operations centers (SOC) have a significant responsibility of conducting alert investigations, primarily to enhance their reactive defense capabilities. To assist in their mission, attack graphs (AG) are commonly utilized for visual analytics and forensic analysis to represent attacker strategies. However, when summarizing complex attacks in a network, the corresponding AG can become hard to read and interpret by the specialist.

The concept of interpretability is important as the purpose of an attack graph is to provide important insights regarding the intention of malicious parties. This is done by summarizing multiple alerts, generated by Intrusion Detection Systems (IDSes), into a more human-readable, "interpretable" data visualization, reducing the time required by specialists to analyze such incidents and facilitating the forensic analysis of past attacks [7]. As the main goal of an attack graph is to provide vital insights into an attack, the interpretability problem is briefly tackled within the generation phase of various attack graph generation tools as presented in [8] and [10].

Making a clear distinction between the interpretability of the output attack graph and the interpretability of the model used to generate the attack graph, a metric used to quantify the interpretability of such attack graph can be beneficial even from the generation phase of attack graphs. By evaluating the interpretability of the output AGs resulting from multiple generation methods (such as using an S-PDFA: *suffix-based probabilistic deterministic finite automaton*, a suffix tree, or a Markov chain), a specialist can select the generation alternative that produces the most interpretable results for the specific task.

However, there is a lack of a metric based on which we can quantify the interpretability of an attack graph such that we achieve comparable results between different attack graphs. As has been stated by [8], metrics such as AIC, BIC, and Perplexity yield arbitrary values when applied to models trained with different parameters, rendering any comparison devoid of meaningful interpretation.

Having in mind the knowledge gap presented above, the question we address in this paper is: **"How can the interpretability of attack graphs be quantified?"**. As a response, this paper is aiming to use a cognitive approach to quantify the interpretability of an attack graph by following the theory of cognitive chunks of information and formalizing an interpretability metric based on cognitive overload as described by Gestalt's Principles. Finally, we want to test our new metric by evaluating attack graphs resulting from different generation methods: S-PDFA, Markov chain, and suffix tree.

This paper aims to propose a new metric, which is combining elements from the theory of cognitive chunks of information with concepts from user-experience-related field to measure the interpretability of attack graphs in the form of cognitive load. This newly proposed metric showcased good and comparable results according to expert's opinion when tested against a metric with a similar approach.

The remainder of this paper is organized as follows. Section 2 gives an overview of related work toward the interpretability of attack graphs. Section 3 describes the methodology used in our research. Section 4 describes the adopted experimental setup. Section 5 provides the results of the experiments. Section 6 is composed of discussions on our findings. Section 7 presents our conclusion and proposed future work.

2 Related Work

This section aims to describe previously conducted work that helped in this paper. Section 2.1 describes the Gestalt principles as concepts used in the user experience field. Section 2.2 provides a summary of a previously proposed interpretability quantification approach based on the theory of cognitive chunks of information. Section 2.3 summarizes a previously proposed metric that is proven to impact the readability of an attack graph. Section 2.4 summarizes a previously proposed complexity metric based on the planarity of a graph.

2.1 Gestalt Principles

The Gestalt principles, also known as the laws of perceptual organization, are a set of principles that describe how humans perceive and make sense of visual stimuli.

Their implementation can significantly enhance both the visual appeal and functionality of a design, making it more user-friendly according to [2].

The 7 Gestalt principles are:

- Similarity: similar elements are visually grouped.
- Continuity: the human eye will follow the smoothest path.
- Closure: when confronted with an incomplete or partially obscured stimulus, our minds tend to fill in the missing information and perceive the object as a complete whole.
- Proximity: how close elements are to one another
- Figure: the human mind will distinguish between the objects it considers to be in the foreground of an image and the background.
- Symmetry: the human mind perceives symmetrical figures as more stable and organized than asymmetrical ones.
- Common-fate: objects that follow the same pattern are perceived as belonging together.

2.2 FAIXID Framework

FAIXID [6] emphasizes the limited capacity of humans to process information, with the average human being able to process 7 ± 2 pieces of information. According to it, the effectiveness of an explanation is dependent on the number of cognitive chunks or informational elements that the recipient must process to comprehend it, with the addition that "interaction among cognitive chunks ... complicates the explanation"[6, p.24]

$$E = \frac{1}{N_c} + (1 - I_n)$$

Where E = explainability, N_c = the number of cognitive chunks, and I_n = interaction.

2.3 Link density-based metric

The readability of link-based representations of graphs is significantly influenced by two main factors: the number of nodes and the density of connections between them. [3]

To support this statement, a link-density metric was proposed in [3]:

$$d = \sqrt{\frac{l}{n^2}}$$

where n = the number of nodes and l = the number of links.

The findings demonstrate that as link density increases, readability decreases.

2.4 Planarity-based complexity metric

By definition, a planar graph is a type of graph that can be projected to a two-dimensional plane in such a way that its edges intersect only at their endpoints. [5]

A planarity-based complexity metric was proposed in [5], defined as the minimum number of edges to achieve a planar graph, a problem also known as Maximal Planarization.

3 Methodology

All the attack graphs that will be discussed in this paper are generated using SAGE (Intrusion alert-driven attack graph extractor), a tool capable of condensing alerts into "alert-driven" attack graphs without prior knowledge about the network.[9]

The attack graphs were generated using two datasets containing intrusion alerts captured through the Collegiate Penetration Testing Competition (CPTC) from the years 2018 (CPTC-2018) and 2017 (CPTC-2017).

By following the definition of interpretability presented as explainability in the FAIXID framework [6], a variation of this approach is proposed as a baseline for our new metric. This variation is presented in section 3.1. Furthermore, a formalization of the FAIXID framework using Gestalt principles is described in section 3.2. In section 3.3 we propose an addition to the baseline metric, constructing our newly proposed metric, and in section 3.4 we aim to evaluate the proposed metric against the presented baseline.

3.1 Baseline

The chosen baseline for this metric is a variation of FAIXID that uses a pattern-based clustering algorithm to generate cognitive chunks inside an attack graph and betweenness centrality as a measure of interdependency.

The clustering algorithm is based on the hypothesis that popular attack patterns can be found inside an attack path due to the methodology of attackers. The algorithm takes all possible pairs $(Node_a, Node_b)$ of attack nodes and computes the probability of $Node_a$ to be preceded by $Node_b$:

$$P(Node_a, Node_b) = \frac{EdgeCount(Node_a, Node_b)}{Outgoing(Node_b)}$$

At the beginning of the algorithm, every node dictates its own cognitive chunks. At every step, we find the most popular pair of nodes $(Node_a, Node_b)$ in our pattern list and merge them by merging the chunk that $Node_b$ is part of to the chunk of $Node_a$ as long as $Node_a$ was not already merged using a pattern of higher probability.

To represent the interdependency metric, we are using a partial dependency analysis on key nodes that are generated based on the betweenness centrality:

$$I_n = \frac{\sum_{n}^{keynodes} \frac{\#n_{outgoing}}{(\#nodes-1)\cdot\#teams}}{\#key_{nodes}}$$

The above equation can be described as the average of the connectivity between key nodes, per attacker team.

A key node inside a cluster is a node with the highest betweenness centrality and the connectivity between a key code and the rest of the cluster is calculated as the ratio between the number of outgoing edges from the key node and the number of non-key nodes in the cluster. This ratio is divided by the number of attacker teams presented in the graph.

This process is applied to every cluster and the average of interdependency inside clusters is considered as the global interdependency, which is used to calculate the interpretability of a graph.

A key difference between this metric and the traditional FAIXID framework is the definition used for interdependency. In this variation, we are analyzing the interaction between nodes inside a cluster, while the traditional FAIXID method is analyzing the interaction between the cognitive chunks.

However, by examining from Table 1 the average number of clusters discovered in datasets CPTC-2018 and CPTC-2017 by the clustering algorithm, we can observe a limited number of clusters were created, which would result in an overall small interdependency using the traditional FAIXID framework.

Dataset	Avg. number of clusters
CPTC-2018	2.342
CPTC-2017	2.785

Table 1: Average number of cognitive chunks discovered by the proposed clustering algorithm in datasets CPTC-2018 and CPTC-2017.

3.2 Baseline generalized to Gestalt principles

By following the FAIXID framework, it can be observed that interpretability is composed of two components that we can formalize with respect to Gestalt's principles: Similarity and Connection.

The first component under discussion is quantifying the number of cognitive chunks: $\frac{1}{N_c}$. We can argue this respects the Similarity Principle as the proposed chunking mechanism is considering the observed probability of nodes appearing together.

The second component quantifies the interdependency of nodes inside the graph. We can argue this respects the Connection principle since we are measuring the impact of central nodes, determined by betweenness centrality, upon other nodes.

3.3 Continuity principle proposal

Considering the above formalization, this paper proposes the addition of a metric following the Continuity principle: planarity. As stated in [3]: "The traditional node-link representation suffers from link overlapping", which can lead to visual clutter and increased complexity.

Adapting to the information above, we define planarity as:

$$P = \frac{1}{e_{removed} + 1}$$

where $e_{removed}$ is the minimum number of edges to be removed from the graph to obtain a planar graph.

It can be observed the planarity is in a normalized form between 0 and 1 since the minimum value of $e_{removed}$ by definition is 0 in the case of a planar graph. Since 0 is a lower bound for $e_{removed}$, we have: $lim_{e\to 0}\frac{1}{e+1} = \frac{1}{1} = 1$. Additionally, we have $lim_{e\to\infty}\frac{1}{e+1} = 0$, concluding that $P \in [0,\ 1]$.

With the addition of planarity and considering that every component of the metric is normalized between 0 and 1, the new metric for interpretability proposed by this paper (in a normalized form) is:

$$I = \frac{\frac{1}{N_c} + (1 - I_n) + P}{3}$$

Maximal planarization is an NP-hard problem, making planarity unachievable as we have to test all the combinations of edges if they describe a planar graph and take the maximum number of edges that describes such a graph. The solution proposed in [5] is to random sample the generated combinations.

In this paper we are considering a subtractive method to generate graphs instead of the additive method proposed in [5], resulting in the following algorithm to calculate the planarity of a graph:

- Take n from 0 to the number of edges in the graph.
- Take 1000 samples of combinations of n edges.
- Remove the one by one the sampled combinations of edges from the graph and test the planarity.
- Stop when finding the first planar graph and the corresponding n is the minimum number of edges to achieve a planar graph.

To test the planarity of a graph we use the planarity test implementation from the NetworkX library.[4]

Now we can formalize our hypothesis: "The interpretability of an attack graph is dependent on the planarity of the graph". The higher the planarity, the higher the interpretability.

As stated by [3], the link density of a graph, is defined as:

$$d = \sqrt{\frac{l}{n^2}}$$

where n = the number of nodes and l = the number of links, is affecting proportionally the interpretability of a graph. By proving that as planarity decreases, it increases the link density, we can, by transitivity, confirm our hypothesis. So our new hypothesis becomes: "The planarity of a graph decreases with the increase of link density."

3.4 Metric evaluation

The main problem this approach wants to solve is to generate comparable results between attack graphs resulting from various generation strategies and environments. In this sense, we test the rankings generated by our metric against the baseline using as inputs the generated attack graphs from SAGE when run against CPTC-2017 and CPTC-2018 datasets.

Kendall rank correlation coefficient [1] is proposed as the method of comparison between the two rankings generated for each dataset:

$$\tau = \frac{(P - Q)}{\sqrt{(P + Q + T) * (P + Q + U)}}$$

Where P = the number of concordant pairs, Q = the number of discordant pairs, T = the number of ties only in the first-ranking list, and U = the number of ties only in the second-ranking list. It is to be observed that we are using the imple-

mentation from the SciPy python library[1] for the tau-b variant of the Kendall rank correlation to account for ties.

Other evaluation methods were considered such as Rank Biased Overlap (RBO) [12], however, the Kendall rank correlation coefficient was selected as our goal for this experiment is to quantify the number of inversions in the two rankings, regardless of the ranking positions of the items relative to the rest of the items in the list. This is because, when comparing attack graphs, we sample two graphs randomly from the list regardless of their ranking positions.

Furthermore, for each dataset, we sample 10 discordant pairs (pairs of attack graphs in which orders are reversed in the ranking generated by our metric compared to the one generated by the baseline) and test whether the new ordering coincides with the specialist's opinion.

The preliminary results we expect are a high similarity between the rankings generated by the proposed metric and the baseline, with the discordant pairs being in an order that is in concordance with specialist opinion.

4 Experimental Setup

This section presents the experimental setup adopted in this research to consolidate and confirm the hypothesis that interpretability increases with planarity-based complexity.

To test the hypothesis, we test on graphs with 3 different node-based sizes (10 vertices, 20 vertices, 30 vertices), and different 9 different edges-based sizes (10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% of the maximum number of edges possible in the graph) for each node-based size.

The selected sizes of 10, 20, and 30 vertices can be explained by looking at the average node-count-based size of attack graphs generated by SAGE using the datasets CPTC-2018 and CPTC-2019 in Table 2

Dataset	Avg. size	Min. size	Max. size
CPTC-2018	17.1	3	32
CPTC-2017	17.7	3	48

Table 2: Size represented by the number of nodes in attack graphs generated by SAGE using CPTC-2018 and CPTC-2017 datasets.

The maximum number of edges in a directed graph is calculated as:

$$Nr_{edges} = Nr_{nodes} * (Nr_{nodes} - 1)$$

where Nr_{nodes} = number of nodes and Nr_{edges} = number of edges.

To mitigate any bias arising from the selected data, we decided to generate random directed graphs using the following algorithm:

- Generate a fully connected graph of the selected size
- At random eliminate edges one by one until achieving the proposed link-density

[1]Kendall rank correlation coefficient from SciPy https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.kendalltau.html

For each graph configuration, 50 samples will be generated. For each sample, we will measure planarity as defined above.

It is expected the planarity to follow a negative trend while the link density increases.

5 Results

This section presents the results gathered during the research. In section 5.1 can be observed the results of the experiment testing the hypothesis of interpretability increasing with planarity-based complexity. In section 5.2 we present the results of the evaluation of the proposed metric against the presented baseline metric.

5.1 Planarity proof results

The primary goal of this experiment is to confirm or deny the decrease in planarity in relation to the increase in link density. For each graph node-based size, we plot the planarity in relation to the link density. Furthermore, we are drawing the exponential trend line in the dotted line to be easier to follow the trend.

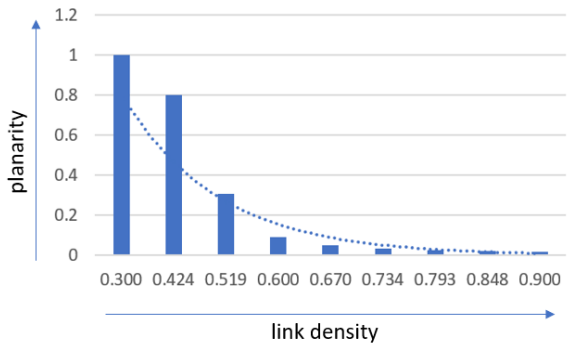


Figure 1: Average planarity in relation to link density in graphs with 10 nodes. The exponential trend line is visible as a blue dotted line.

From Figure 1 we can see a downward trend of planarity in relation to the increase in link density. The minimum difference between each link density considered in the experiment is 0.059 with no outliers or local upward trends detected.

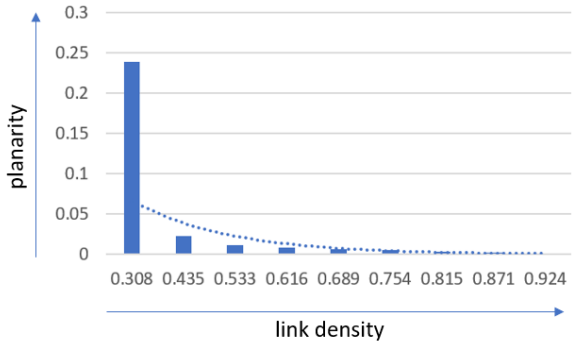


Figure 2: Average planarity in relation to link density in graphs with 20 nodes. The exponential trend line is visible as a blue dotted line.

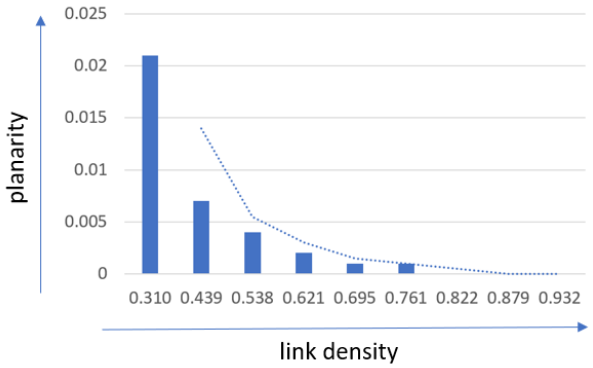


Figure 3: Average planarity in relation to link density in graphs with 30 nodes. The exponential trend line is visible as a blue dotted line.

It can be observed from Figures 2 and 3 the trend is maintained with the increase of node-based graph size. This downward trend can be explained by analyzing what happens to the two metrics: link density and planarity with the increase in the number of edges, considering the number of nodes constant. By definition, we know link density increases with the number of links presented in the attack graph. On the other hand, the total number of edges in the graph is an upper bound of the number of edges we can eliminate to obtain a planar graph. This means that planarity increases while the number of edges decreases.

5.2 Metric evaluation results

As it can be observed in Table 3, and confirming our expectancies, the values of Kendall Coefficient for both CPTC-2018 and CPTC-2017 datasets are close to 1, showing a high correlation between the rankings denoted by our proposed and baseline metrics.

Dataset	Kendall Coefficient
CPTC-2018	0.931
CPTC-2017	0.869

Table 3: Kendall Corelance Coefficient in datasets CPTC-2018 and CPTC-2017.

As specified in section 3.4, for each dataset 10 samples were taken among all the inversions. Each of the samples was tested against the specialist's opinion and considered "Accepted" if according to the specialist's opinion, the new order dictated by our metric is valid, or "Denied" if the new order is either considered invalid or the pair of attack graphs are too similar in order to create a clear order.

From Table 4 it can be concluded that the number of accepted inversions overcomes the number of denied inversions.

Dataset	Accepted Inversions	Denied Inversions
CPTC-2018	8	2
CPTC-2017	7	3

Table 4: Inversions sampled among the rankings generated by the baseline and proposed metrics in CPTC-2018 and CPTC-2017 datasets.

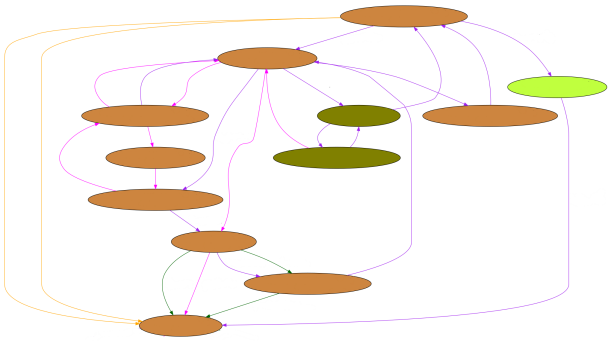
An observation worth to be mentioned is that 100% of the denied inversions have been denied because the pair of attack graphs were too similar to be evaluated in this scope and 0 samples were denied as a result of a discrepancy between the metric and the specialist's opinion.

6 Discussion

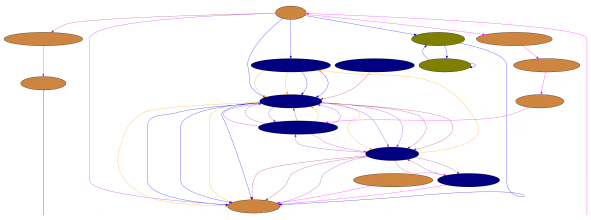
6.1 Manual analysis on inversions

By manually analyzing two attack graphs generated by SAGE[9] using the dataset CPTC-2018, we discover a discrepancy between the expert's opinion and the baseline metric. In contradiction with the expert's opinion, the baseline interpretability metric shows that graph 'A' is more interpretable than graph 'B'. On the other hand, we can observe that the newly proposed metric is expressing a bigger value for the interpretability of graph 'A'.

- Graph A:
  - Baseline interpretability: 0.303
  - Baseline + continuity interpretability: 0.535



- Graph B:
  - Baseline interpretability: 0.377
  - Baseline + continuity interpretability: 0.418



This is a result of the difference in the planarity of the graphs as graph 'A' presents a planarity of 1 and graph 'B' present a planarity of 0.5.

The distinction in planarity can be observed through the high ratio between the number of links and the number of vertices in the dark blue cluster in graph B'. This ratio indicates a significant link density.

6.2 Use case

As a proposed use case, we use the new metric to evaluate the interpretability of attack graphs resulting from the S-PDFA, a Markov chain, and a suffix tree.

SAGE is based on FlexFringe [11] automaton learning framework. To generate the attack graphs, we are running SAGE with different parameters as follows: To use an S-PDFA to generate the attack graphs we can use the default parameters of SAGE. We can create a Markov chain by utilizing a statistical test threshold set to an extremely low value (or even negative). With the Markovian parameter set to 1, performing a likelihood-ratio analysis gives rise to the formation of a Markov chain.[11] On the other hand, we can disable all merges to obtain a suffix tree.

Calculating the interpretability of attack graphs generated by SAGE with the proposed sets of parameters, it can be observed in Table 5 that on average, attack graphs resulting from a suffix tree are more interpretable than the ones resulting from the S-PDFA or a Markov Chain. This is in concordance with the expert's opinion as in a suffix tree merges are not allowed, thus, every attack path is separate from the rest, being easier to follow by the specialist. However, attack graphs resulting from this method are larger in size.

Dataset	S-PDFA	Markov Chain	Suffix Tree
CPTC-2018	0.727	0.743	0.790
CPTC-2017	0.679	0.669	0.682

Table 5: Average interpretability of attack graphs generated using SAGE against datasets CPTC-2018 and CPTC-2017 resulting from the S-PDFA, a Markov chain, and a suffix tree.

7 Conclusions and Future Work

In this paper, we have proposed a new metric to quantify the interpretability of an attack graph, a variation of the FAIXID framework that we generalized to 2 of the 7 Gestalt principles: Similarity and Connection, to which we have inserted the Continuity principles in the form of maximum planarity.

$$I = \frac{\frac{1}{N_c} + (1 - I_n) + P}{3}$$

As shown above, the rankings denoted by the proposed and the baseline metrics have a high level of correlation, which can be interpreted as similar results when it comes to pairwise comparison.

Furthermore, we have shown that, by sampling the changes in rankings induced by our new metric in comparison to the baseline, we can observe a big number of changes that are, according to the expert's opinion, valid and can be considered an improvement from the baseline metric.

On the other hand, the question of quantifying a subjective metric such as interpretability is not closed. As presented earlier, there are cases where our new metric fails to rank correctly and creates errors that were not in the rankings induced by the baseline. However, according to our results, the correct changes from the baseline ranking are outnumbering the invalid changes in the same ranking.

Since our metric takes into consideration only 3 out of the 7 Gestalt principles, we propose work on the integration of other user-experience-related principles. The "Closed Shape" principle can be considered a good start as it can be used to quantify the amount of connectivity between the node clusters that we treat as cognitive chunks, information that is not covered by our metric.

In conclusion, as far as the presented evidence goes, we consider cognitive load as a direct way of quantifying the interpretability of an attack graph, thus, we can retrieve comparable results between the interpretability of attack graphs resulting from different generation methods using the metric showcased above.

8 Responsible Research

All resources are open source and available on GitHub [2].

The reproduction and regeneration of all the attack graphs presented in this paper are possible since SAGE is a deterministic algorithm, and the datasets used, CPTC-2017 and CPTC-2018, are publicly accessible. The methodology for generating attack graphs using different approaches is detailed in section 6.2. However, it is worth noting that in order to replicate the attack graphs as depicted in this paper, it is necessary to employ the latest version of FlexFringe, which should be available by June 25th, 2023, due to the ongoing development of the software.

As we are using random sampling of the combinations of edges for the maximum planarity problem, running experiments involving planarity as a metric might produce different results than the ones presented in the paper.

The analysis was approached with a focus on minimizing bias in the results, achieved through the establishment of explicit research objectives. Nevertheless, it is important to note that the absence of bias cannot be fully guaranteed due to the subjective nature of the manual analysis that was conducted.

9 Acknowledgements

I express my gratitude to Dr. Ir. Sicco Verwer and Ph.D. Candidate Azqa Nadeem for the delightful collaboration we had. Their feedback and guidance have been crucial in ensuring the progress of this research in the right direction.

I would also like to extend my appreciation to my esteemed peers, Jegor Zelenjak, Ioan Oprea, Alexandru Dumitriu, and Senne Van den Broeck, whose valuable insights have contributed to this work.

References

[1] Hervé Abdi. The kendall rank correlation coefficient. *Encyclopedia of Measurement and Statistics. Sage, Thousand Oaks, CA*, pages 508–510, 2007.

[2] Cameron Chapman. Exploring the gestalt principles of design. *Toptal*.

[3] M. Ghoniem, J.-D. Fekete, and P. Castagliola. A comparison of the readability of graphs using node-link and matrix-based representations. In *IEEE Symposium on Information Visualization*, pages 17–24, 2004.

[2] https://github.com/Kheoss/AGIAS

[4] Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. Exploring network structure, dynamics, and function using networkx. In Gaël Varoquaux, Travis Vaught, and Jarrod Millman, editors, *Proceedings of the 7th Python in Science Conference*, pages 11 – 15, Pasadena, CA USA, 2008.

[5] Sebastian Kortler, Matthias Kreimeyer, and Udo Lindemann. A planarity-based complexity metric. *DS 58-6: Proceedings of ICED 09, the 17th International Conference on Engineering Design*, 6, 01 2009.

[6] Hong Liu, Chen Zhong, Awny Alnusair, and Sheikh Rabiul Islam. Faixid: A framework for enhancing ai explainability of intrusion detection results using data cleaning techniques. *Journal of Network and Systems Management*, 29, 10 2021.

[7] Azqa Nadeem, Sicco Verwer, Stephen Moskal, and Shanchieh Jay Yang. Enabling visual analytics via alert-driven attack graphs. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, CCS '21, page 2420–2422, New York, NY, USA, 2021. Association for Computing Machinery.

[8] Azqa Nadeem, Sicco Verwer, Stephen Moskal, and Shanchieh Jay Yang. Alert-driven attack graph generation using s-pdfa. *IEEE Transactions on Dependable and Secure Computing*, 19(2):731–746, 2022.

[9] Azqa Nadeem, Sicco Verwer, and Shanchieh Jay Yang. Sage: Intrusion alert-driven attack graph extractor. In *Symposium on Visualization for Cyber Security (Vizec)*. IEEE, 2021.

[10] O. Sheyner, J. Haines, S. Jha, R. Lippmann, and J.M. Wing. Automated generation and analysis of attack graphs. In *Proceedings 2002 IEEE Symposium on Security and Privacy*, pages 273–284, 2002.

[11] Sicco Verwer and Christian Hammerschmidt. Flexfringe: Modeling software behavior by learning probabilistic automata, 03 2022.

[12] William Webber, Alistair Moffat, and Justin Zobel. A similarity measure for indefinite rankings. *ACM Trans. Inf. Syst.*, 28(4), nov 2010.