

Node and Edge Removal on Complex Networks in Labor Market Research and their Influence on Centrality Measures

Meetkumar Pravinbhai Mangroliya ¹, Jens Dörpinghaus ², and Robert Rockenfeller¹

Abstract: This research examines the impact of node and edge removal strategies on centrality measures within complex networks. Investigating random, scale-free, and small-world networks, various removal approaches, including targeted and random removal, are evaluated. The study assesses their influence on centrality metrics such as degree, betweenness, closeness, and eigenvector centrality on random networks and networks from educational research describing longitudinal data in labor market-related topics in social networks. The findings contribute insights applicable across domains. In social network analysis, an understanding of key actors is beneficial for the development of targeted interventions or marketing strategies. Historical network analyses benefit from the discernment of pivotal nodes or connections, which elucidate information flow or influential figures across different periods. Such applications underscore the significance of the research in optimizing network performance in diverse contexts.

Keywords: Centrality Measures, Complex networks, Labor Market Networks, Social Network Analysis

1 Introduction

Understanding the intricate dynamics of complex networks is crucial for numerous real-world applications, ranging from social network analysis to historical reconstructions. This study investigates the impact of node and edge removal strategies on centrality measures within complex networks, covering diverse network models. These models include random networks, typified by their lack of inherent structure or organization [ER59; Gi59]; scale-free networks, notable for the prevalence of a power-law degree distribution, where most nodes possess few connections while a few nodes (hubs) are highly interconnected [BA99]; and small-world networks, distinguished by their capacity to exhibit both high clustering (dense connections between nearby nodes) and short average path lengths, facilitating efficient information propagation, akin to real-world systems [WS98]. In addition, this study examines labor market networks, particularly online social network data from YouTube, to analyze how removal strategies affect centrality measures. Such networks are crucial for understanding labor market dynamics, including information flow, influential actors, and structural resilience in response to node and edge removal. This research aims to shed light on the resilience and adaptability of these networks, providing insights applicable across domains.

¹ University of Koblenz, Germany,
meetmangroliya987@buni-koblenz.de,  <https://orcid.org/0009-0003-3727-9527>

² Federal Institute for Vocational Education and Training (BIBB), Germany, University of Koblenz, Germany,
doerpinghaus@uni-koblenz.de,  <https://orcid.org/0000-0003-0245-7752>

Centrality measures in network analysis, including degree, betweenness, closeness, and eigenvector centrality, are vital for assessing node importance [Bo05; Ne10]. These metrics provide insights into information flow facilitation, communication path control, and network structure. To gauge network resilience to disruptions, these removal strategies are crucial [AJB00; Ne10]. This study explores targeted and random removal approaches to understand their impact on network topology. Targeted removal selectively eliminates nodes based on criteria, while random removal involves random node or edge elimination. Evaluating these strategies is essential for assessing network resilience in coping with disruptions.

The study explores the practical implications of these findings, particularly in social network analysis. Identifying key influencers or central nodes in social networks is crucial for understanding information diffusion, opinion formation, and targeted interventions. Historical analyses benefit by revealing pivotal nodes or connections that played significant roles in historical events, offering insights into information flow and historical significance. This research advances our understanding of network behavior under stress and optimizes network performance in real-world applications. For labor market networks; understanding the role of central players within these networks can prevent fragmentation and maintain robust communication and operational efficiency, aiding in the development of strategic interventions to enhance network stability and performance.

This paper is divided into five sections. Following the introduction, we provide an overview of related work. In the third section, we describe the methodology utilized. Subsequently, in the fourth section, we discuss the experimental results of the study. Finally, in the last section, we conclude and discuss the outcomes of the study.

2 Related Work

Networks exhibit both resilience and fragility, responding differently to various perturbations like node and edge removal. This technique allows for studying network dynamics and resilience, enhanced by predictive models from machine learning and computational methods, which predict the impacts of such removals, aiding proactive network management and optimization strategies through mathematical models, optimization algorithms, and simulations.

Researchers have proposed and explored diverse methods for removing nodes and edges, encompassing strategies like random removal [AB02; AJB00], as well as centrality-based removals such as degree centrality and betweenness centrality [JM20; Wa18]. A prevalent attack strategy involves pinpointing critical nodes based on metrics like degree or other centrality measures and systematically removing them in descending order of importance until the network either disconnects or loses essential properties Holme et al. [Ho02]. Albert's investigation into the fragmentation of random and scale-free networks employed two distinct node removal strategies: targeted removal (attacks) and random removal (errors) [AJB00]. In the research [SJ15], the authors systematically eliminated genes from a

biological network to examine the consequences on both network connectivity and behavior. Additionally, in the study [Be20], the authors performed a comparative analysis to evaluate the effects of diverse link removal strategies on various real-world complex networks. Additionally, numerous studies have examined the impact of removal strategies on complex networks across various scientific domains [CAD18; Ga05; LM01; YNM17; ZL13]. These methodologies provide a comprehensive understanding of how the removal of nodes and edges influences the overall structure and functionality of networks.

Moreover, studying the impact of node and removal on different structures and connectivity is not without challenges. The interplay between local and global network structure presents complexities in predicting network behavior Holme et al. [Ho02]. Furthermore, quantifying the impact of removal strategies on network robustness requires sophisticated algorithms and computational resources [SK11]. Moreover, the ethical implications of removal actions in real-world networks, such as social networks, introduce ethical considerations [BB12]. Our research aims to navigate these challenges and contribute to a more comprehensive understanding of network connectivity and structural changes due to node and edge removal.

3 Method

In our analysis, we employ a systematic methodology to generate three types of complex networks and investigate the influence of node and edge removal strategies on centrality measures within these networks. We focus on four fundamental centrality measures—degree centrality, betweenness centrality, closeness centrality, and eigenvector centrality—due to their distinct and complementary perspectives on node importance within networks. To compute centrality measures, we utilize the NetworkX library in Python [HSS08], which offers efficient implementations of various centrality algorithms. These measures provide valuable insights into the structural significance of nodes within complex networks.

Generation of Complex Networks: A random network was generated using the Erdős-Rényi (ER) graph generator from the NetworkX library [HSS08]. The network parameters were set to include 100 nodes (n) and an edge probability of 0.2 (p) [ER59; Gi59]. This edge probability influences the network's density, with higher values resulting in denser networks and lower values yielding sparser ones. Additionally, Python's random state function was used to establish a consistent random seed ($\text{seed} = 1$) for reproducible network generation.

The Barabási-Albert (BA) model is widely used for generating scale-free networks [BA99]. This model employs preferential attachment, where new nodes connect to existing nodes with high degrees, resulting in a power-law degree distribution. To create a scale-free network using this model, we specified the number of nodes (n) and the parameter for preferential attachment (m). The parameter m dictates the number of edges added for each new node introduced, influencing the network's structure and mimicking real-world scenarios such as scientific collaborations or social interactions.

The Watts-Strogatz (WS) model is employed to generate small-world networks, characterized by both a small average shortest path length and a high clustering coefficient [WS98]. This model begins with a regular ring lattice where each node is connected to its k nearest neighbors. Then, with a probability p , each link connected to a clockwise neighbor is rewired to a randomly chosen node, resulting in small-world characteristics. Varying p allows us to produce graphs with different levels of small-worldness, ranging from a regular lattice ($p = 0$) to a random graph with minimum connectivity constraints ($p = 1$).

Real-World Network: This dataset comprises videos from YouTube channels classified under occupational Group 3, covering occupations in construction, architecture, surveying, and technical building services. The distribution of video shares from different channels within this group, for example Job Market and Career Advice Channels, e.g., Stellenmarkt-direkt.de with 2.76%, Educational Channels, e.g., MariusEbert 6.6%, Dr. Marius Ebert 0.42%, Business and Company Profiles, e.g., FirmenABC 0.38%, Construction and Building Associations and Industry-Specific Content, e.g., Schweizerischer Baumeisterverband, Style Dach - Aus Liebe zum Handwerk, Zentralverband Sanitär Heizung Klima - ZVSHK or Media and Broadcasting.

Categories for all 271,255 videos include “People & Blogs” (26.1716%), “Education” (21.6263%) and “Science & Technology” (18.9351%). In summary, this dataset includes videos covering a wide range of topics related to occupational Group 3, such as job markets, education, business, construction, and other relevant aspects.

Removal Strategies: In our analysis, we consider two primary node removal strategies: targeted and random. Targeted node removal involves strategically removing nodes based on specific attributes, such as high degree [AJB00]. This approach allows us to evaluate the network’s response to the intentional removal of highly connected nodes or less-connected nodes. Conversely, random node removal involves removing nodes at random, mimicking stochastic node failures or removals. Additionally, we explore edge removal strategies, focusing on the deliberate elimination of specific or random connections within the network. Random edge removal serves as a benchmark for assessing the network’s resilience to random disturbances and offers insights into its robustness when faced with unpredictable edge failures [Be20].

Error Measures: To quantify the impact of node and edge removal strategies on network structures, we employ a fundamental centrality error measure. Centrality error quantifies the disparity in centrality scores before and after removal actions [Bo05]. These error measure provides valuable insights into how these removal strategies affect network resilience and the relative importance of nodes within the network [AB02; CSN09; Es05; Ne18]. This comprehensive methodology enables us to systematically analyze the effects of node and

edge removal strategies on centrality measures within these networks, contributing to a deeper understanding of network dynamics and resilience.

4 Experimental Results

This section is divided into three subsections, high-degree node removal and random node removal and random edge removal strategy. Each removal strategy is applied to a random network, a scale-free network, and a small-world network, all generated using the same parameters as described in the methodology section. The analysis includes box plots that show the absolute differences in degree centrality, betweenness centrality, closeness centrality, and eigenvector centrality. These differences are shown for both before and after applying the removal strategy, with removal percentages ranging from 5% to 25%.

4.1 High-degree node removal strategy

The analysis of an Erdős-Rényi (ER) random network after high-degree node removal, see Figure 1, indicates that the differences in degree centrality, closeness centrality and eigenvector centrality are higher in comparison to betweenness centrality and the differences increase as the percentage of node removal increases from 5% to 25%. The variation in degree centrality differences indicates the influence of these high-degree nodes in sustaining the network's overall connectivity. Simultaneously, closeness centrality experiences considerable shifts due to high-degree node removal. These differences highlight the network's sensitivity to alterations in the distribution of node importance and raise questions about the stability of random networks when subjected to the removal of highly connected nodes.

In the analysis of a scale-free network after high-degree node removal, a different response was observed, see Figure 2. The decreasing betweenness centrality differences with increasing node removal percentages signify the network's vulnerability to disconnection. In scale-free networks, high-degree nodes act as vital connectors between clusters, facilitating communication and information flow. Their removal disrupts network structure, reducing crucial hubs and leading to increased path lengths and decreased network efficiency. This fragmentation is reflected in declining betweenness centrality differences, underscoring the pivotal role of high-degree nodes in maintaining network connectivity and the consequences of their removal in scale-free networks.

Furthermore, as node removal percentages vary, degree centrality differences increase, reflecting the network's complex response to removal. Closeness centrality differences peak at 15 percent removal before stabilizing, while eigenvector centrality differences generally remain stable. These findings highlight the intricate dynamics of centrality measures in scale-free networks under node removal scenarios.

The box plots in Figure 3 illustrate centrality measure differences before and after high-degree node removal in a small-world network. Degree centrality differences consistently increase

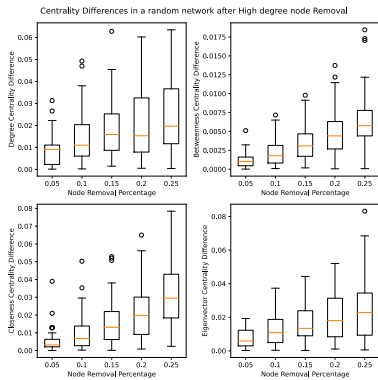


Fig. 1: Erdős-Rényi Random Network

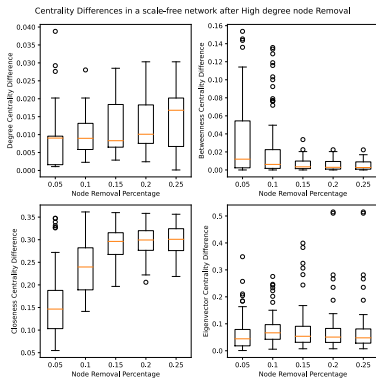


Fig. 2: Barabási-Albert Scale-free Network

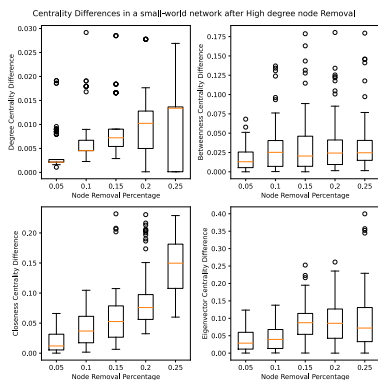


Fig. 3: Watts-Strogatz Small-world Network

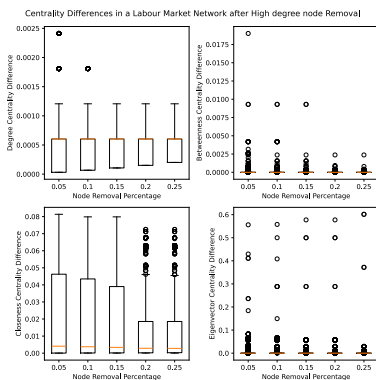


Fig. 4: Labor Market Network

with higher percentages of node removal, reflecting the network’s response to removal. Conversely, betweenness centrality differences remain relatively stable, indicating network resilience. Closeness centrality differences gradually increase until 20 percent removal, then rise more significantly, highlighting network sensitivity to disruptions. Eigenvector centrality differences also show higher values but remain relatively stable after any more than 15% of high-degree node removal, demonstrating the network’s capacity to maintain its overall structure despite the loss of key nodes.

In Figure 4, the box plots illustrate the impact of targeted removal of nodes with the highest degrees on centrality measures within the labor market network. Notably, this strategic removal strategy results in minimal and stable absolute differences across centrality measures, showcasing the network’s robustness against selective perturbations. Moreover, the consistent and minimal absolute differences observed in centrality measures following high-degree node removal highlight the network’s structural stability. This resilience suggests

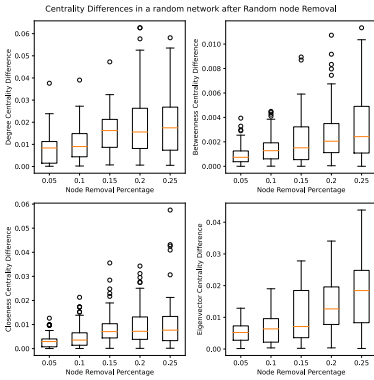


Fig. 5: Erdős–Rényi Random Network

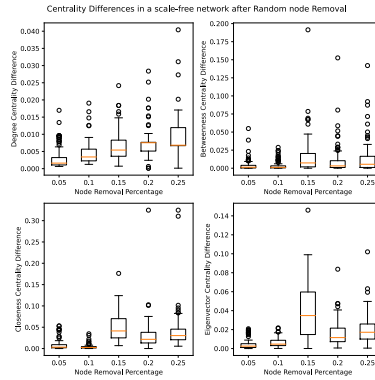


Fig. 6: Barabási-Albert Scale-free Network

that despite the intentional removal of key nodes, the labor market network maintains its overall cohesion and connectivity. Such stability is vital for ensuring the continued flow of information and opportunities within the labor market ecosystem.

4.2 Random node removal strategy

Random node removal is the third approach we employed to assess the impact of node removal on centrality measures in these networks. Changes in centrality measures such as betweenness centrality or closeness centrality can be the indicator of the network's structural integrity. A significant decrease in these measures after node removal could imply that key nodes responsible for maintaining network cohesion and efficiency have been removed, potentially leading to fragmentation. In the analysis of random node removal within a random network, see Figure 5, we observed that as the percentage of nodes removed increased, the degree centrality differences increased. In contrast, betweenness centrality differences were relatively minor, which indicates that random node removal is more likely to disrupt the network's connectivity than its communication paths. Closeness and eigenvector centrality differences moderately increased but not as high as the degree centrality changes. This indicates that random node removal can disrupt the network's topology and make it more difficult for information to flow through the network, which can lead to decreases in closeness and eigenvector centrality.

Random networks exhibit resilience to the random node removal strategy due to their inherent redundancy, characterized by the presence of multiple alternative paths connecting nodes. This redundancy impedes the complete disruption of network connectivity despite random node removal. However, such removal can still significantly affect network resilience by reducing the number of connected components, thereby increasing vulnerability to subsequent disruptions.

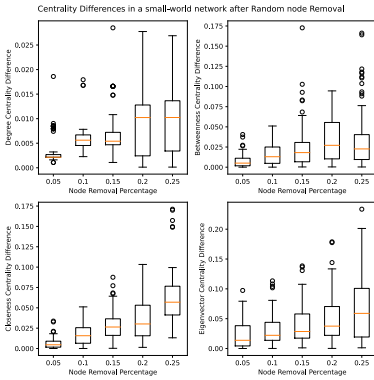


Fig. 7: Watts-Strogatz Small-world Network

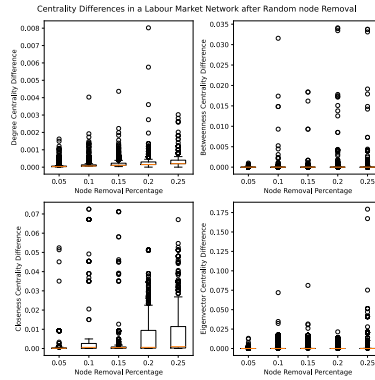


Fig. 8: Labor Market Network

In our study of random node removal within a scale-free network, as illustrated in Figure 6, we found that increasing the proportion of removed nodes had a minimal effect on degree centrality differences. This highlights the robustness of scale-free networks, attributed to their highly connected hubs. However, other centrality measures like betweenness, closeness, and eigenvector centrality showed a distinct pattern. Initially, these centrality differences increased until 15% of nodes were removed, after which they began to decrease. This suggests that the removal of random nodes disrupted the network's connectivity beyond a certain threshold, reducing its ability to facilitate efficient paths and information flow. This decline may indicate potential network fragmentation, with centrality measures focusing predominantly on the largest component of the network.

The impact of random node removal in a small-world network was explored in Figure 7. As the proportion of nodes removed increased, there was a noticeable rise in degree centrality differences, albeit remaining relatively modest throughout. This indicates that similar to scale-free networks, removing random nodes had a limited structural impact on the small-world network. However, for centralities like betweenness, closeness, and eigenvector centrality, increasing removal percentages showed escalating differences. This progressive increase suggests a significant effect on network centralities, highlighting the network's sensitivity to such removal and its potential to disrupt connectivity and alter centrality measures. Overall, these observations emphasize the substantial impact of random node removal on small-world networks.

In Figure 8, the box plots depict the effects of random node removal on centrality measures within the labor market network. Despite resulting in minimal absolute differences in centrality measures, this strategy exhibits a higher frequency of outliers compared to targeted high-degree node removal. Such variability suggests occasional disruptions to network cohesion, albeit to a lesser extent than random edge removal.

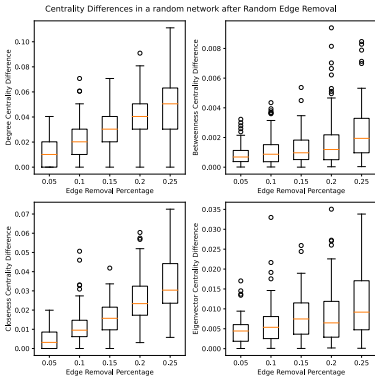


Fig. 9: Erdős-Rényi Random Network

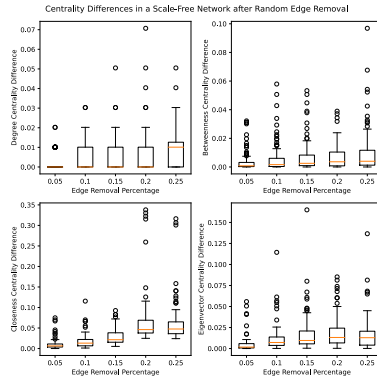


Fig. 10: Barabási-Albert Scale-free Network

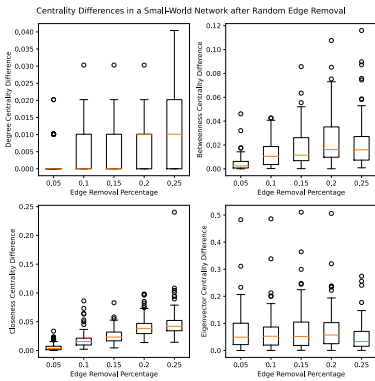


Fig. 11: Watts-Strogatz Small-world Network

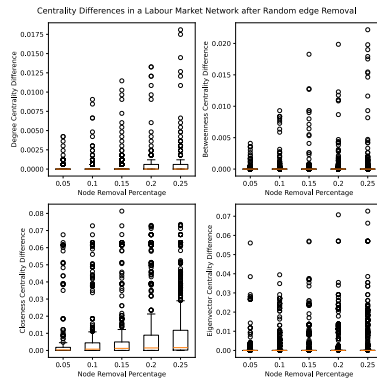


Fig. 12: Labor Market Network

4.3 Random edge removal strategy

In this analysis, we examined the effects of random edge removal on these four centrality measures in complex networks. In a random network, see Figure 9, we observed a noticeable difference in degree and closeness centrality as the percentage of random edge removal increased. This difference became more pronounced with higher removal percentages, indicating that network connectivity was affected. On the other hand, the difference in betweenness centrality and eigenvector centrality were relatively small, showcasing the network's resilience in terms of information flow. This indicates that, despite the disruptions caused by edge removal, the network maintains its capacity for efficient information transmission.

The impact of random edge removal in a scale-free network is presented in Figure 10. Degree centrality differences remained relatively small even at higher removal percentages,

indicating the network's resilience in terms of connectivity. The differences in betweenness and eigenvector centrality were higher compared to degree centrality differences but still relatively minor, suggesting a slight susceptibility of high-degree nodes to random edge removal and a potential disruption of critical information pathways. Closeness centrality differences were more pronounced, indicating a reduction in the network's clustering. The scale-free nature of this network, characterized by a few highly connected hubs, contributed to its robustness against random edge removal. This resilience against edge removal demonstrates the network's ability to maintain its structure and information flow even when subjected to significant disruptions.

The impact of random edge removal on a small-world network was presented in Figure 11. The difference in degree centrality stayed minor, showing that nodes stayed connected even when higher percentages of edges were randomly removed. However, the increasing difference in betweenness centrality with a higher edge removal percentage indicates a more substantial impact on the ability of the network's information flow. Similarly, the higher closeness centrality difference at higher removal percentages suggests a reduction in network clustering. In contrast, eigenvector centrality differences were the highest among all centrality measures, indicating a more significant impact on the influence of nodes within the network. While a small-world network's structure provides efficient information flow, its higher degree of interconnectedness makes it more vulnerable to disruptions caused by random edge removal.

Furthermore, the box plots in Figure 12 demonstrate the outcomes of random edge removal on centrality measures in the labor market network. While this strategy yields minimal absolute differences similar to random node removal, its impact on network cohesion is slightly more pronounced. The presence of outliers indicates occasional disturbances to network structure, potentially affecting the flow of information and opportunities within the labor market ecosystem.

The above observation highlights the importance of evaluating network structure and centrality measures when assessing the impact of various removal strategies. Targeted node removal strategies, such as removing high-degree nodes, result in substantial disruptions in centrality measures like degree, betweenness, and closeness. This leads to a notable decrease in these centrality values, suggesting potential fragmentation and compromising the network's ability to maintain efficient communication paths and structural integrity.

5 Discussion and Outlook

This analysis reveals a stark contrast in network resilience depending on the removal strategy employed. Random networks exhibit minimal impact compared to the significant disruption observed when targeting high-degree nodes or removing random edges. These findings highlight the critical role of highly connected nodes and the random structure of a network in maintaining resilience. The observed decrease in degree centrality values following

random node removal underscores the substantial influence of connectivity on a node's resilience within the random networks.

Scale-free networks, characterized by a few highly connected hubs, rely heavily on these hubs for efficient communication and overall structural integrity. While demonstrating robustness against random node removal, these networks exhibit vulnerability when high-degree nodes, particularly hubs, are specifically targeted. The removal of hubs significantly impacts centrality measures, emphasizing their indispensability in safeguarding network integrity. Small-world networks, known for their balance between redundancy and efficiency, display a moderate positive correlation among centrality measures. However, the removal of high-degree nodes, which often act as bridges between distinct communities, can significantly disrupt network connectivity and structure. Interestingly, the varying degrees of connectivity between nodes and the removed node in small-world networks highlight the inherent adaptability of these systems. This adaptability allows small-world networks to maintain their overall structure even after node removal.

In general, we find that labor market networks form (online) social networks like YouTube follow the patterns found in two random networks. They remain relatively stable, but show a high degree of variability. Further research is needed to investigate the topology of these networks. For labor market research, however, we can summarize that these networks demonstrate a stability comparable to other social networks, providing a solid foundation for further exploration and application of these findings in real-world scenarios.

References

- [AB02] Albert, R.; Barabási, A.-L.: Statistical mechanics of complex networks. *Reviews of Modern Physics* 74/1, pp. 47–77, 2002.
- [AJB00] Albert, R.; Jeong, H.; Barabasi, A.: Error and attack tolerance of complex networks. *Nature*, 2000.
- [BA99] Barabási, A. L.; Albert, R.: Emergence of scaling in random networks. *Science* 286/5439, pp. 509–512, 1999.
- [BB12] Baltar, F.; Brunet, I.: Unintended consequences and new challenges of online social systems: Methods, challenges, and applications. *Journal of Information Technology & Politics* 9/1, pp. 1–20, 2012.
- [Be20] Bellingeri, M. et al.: A comparative analysis of link removal strategies in real complex weighted networks. *Scientific Reports* 10/1, p. 3911, 2020.
- [Bo05] Borgatti, S. P.: Centrality and network flow. *Social Networks* 27/, pp. 55–71, 2005.
- [CAD18] Caldu-Primo, J. L.; Alvarez-Buylla, E. R.; Davila-Velderrain, J.: Structural robustness of mammalian transcription factor networks reveals plasticity across development. *Scientific Reports* 8/1, pp. 1–15, 2018.

- [CSN09] Clauset, A.; Shalizi, C. R.; Newman, M. E. J.: Power-law distributions in empirical data. *SIAM review* 51/4, pp. 661–703, 2009.
- [ER59] Erdős, P.; Rényi, A.: On random graphs I. *Publicationes Mathematicae* 6/, pp. 290–297, 1959.
- [Es05] Estrada, E.: Subgraph centrality in complex networks. *Physical Review E* 71/5, p. 056103, 2005.
- [Ga05] Gallos, L. K. et al.: Stability and topology of scale-free networks under attack and defense strategies. *Physical Review Letters* 94/18, p. 188701, 2005.
- [Gi59] Gilbert, E. N.: Random Graphs. *The Annals of Mathematical Statistics* 30/4, pp. 1141–1144, 1959.
- [Ho02] Holme, P. et al.: Attack Vulnerability of Complex Networks. *Physical review. E, Statistical, nonlinear, and soft matter physics* 65/, p. 056109, 2002.
- [HSS08] Hagberg, A. A.; Schult, D. A.; Swart, P. J.: Exploring network structure, dynamics, and function using NetworkX. *Proceedings of the 7th Python in Science Conference (SciPy 2008)* 11/, pp. 11–15, 2008.
- [JM20] Jenelius, E.; Mattsson, L.-G.: Resilience of Transport Systems. In. 2020.
- [LM01] Latora, V.; Marchiori, M.: Efficient Behavior of Small-World Networks. *Physical Review Letters* 87/19, p. 198701, 2001.
- [Ne10] Newman, M. E. J.: *Networks: An Introduction*. Oxford University Press, 2010.
- [Ne18] Newman, M. E.: *Networks*. Oxford University Press, 2018.
- [SJ15] Smith, J.; Johnson, L.: Edge removal and network reconstruction in biological networks. *PLoS ONE* 10/7, e0131180, 2015.
- [SK11] Schneider, C. N.; Koutsopoulos, I.: Mitigation of malicious attacks on networks. *Proceedings of the ACM SIGMETRICS joint international conference on Measurement and modeling of computer systems*/, 2011.
- [Wa18] Wandelt, S. et al.: A comparative analysis of approaches to network-dismantling. *Scientific Reports* 8/, p. 13513, 2018.
- [WS98] Watts, D. J.; Strogatz, S. H.: Collective dynamics of 'small-world' networks. *Nature* 393/6684, pp. 440–442, 1998.
- [YNM17] Yang, Y.; Nishikawa, T.; Motter, A. E.: Small vulnerable sets determine large network cascades in power grids. *Science* 358/, 2017.
- [ZL13] Zanin, M.; Lillo, F.: Modelling the air transport with complex networks: A short review. *The European Physical Journal Special Topics* 215/, pp. 5–21, 2013.