Information Diffusion on Multiplex Networks

Raman Ebrahimi* and Atefeh Mollabagher†

Abstract

This work investigates information diffusion in multiplex networks, where nodes interact across multiple layers representing different types of connections. Using a model that includes varying transmission probabilities both between layers and in each layer separately, we analyze diffusion patterns in empirical data and compare the results for multiplex and single-layer models. Our results show that multiplex networks can either accelerate or hinder information spread, depending on their structure and layer interdependence. We identify crucial nodes and layers that optimize diffusion, providing insights valuable for improving strategies in information dissemination, viral marketing, and epidemic control within complex networked systems.

1 Introduction

The study of information diffusion has garnered significant attention in recent years (Singh et al., 2019; Li et al., 2017), driven by the ubiquity of social media platforms and the rapid dissemination of information across various networks. Information diffusion pertains to the process through which information, ideas, behaviors, or innovations spread through a population (online or not) over time. This phenomenon is critical to understanding social dynamics, marketing strategies, epidemic spread, and the propagation of technological innovations (Newman, 2002; Castellano et al., 2009; Ji et al., 2023). In the realm of network science, multiplex networks—networks consisting of multiple layers representing different types of connections among the same set of nodes—offer a rich framework for modeling and analyzing complex systems of interconnected entities.

In multiplex networks, each layer can represent a distinct type of interaction, such as friendships, professional relationships, or co-authorships. This layered structure captures the multifaceted nature of real-world interactions more effectively than single-layer networks. Consequently, studying information diffusion in multiplex networks provides deeper insights into the mechanisms governing the spread of information and highlights the interplay between different types of relationships.

Despite the potential of multiplex networks to more accurately represent complex systems, research on information diffusion within such networks is still in its nascent stages. Previous studies have predominantly focused on single-layer networks, which

*(PID: A69028430)
†(PID: A69027897)

may oversimplify the diffusion process and overlook critical inter-layer interactions. By leveraging the layered structure of multiplex networks, researchers can uncover how different layers influence the spread of information and identify key factors that accelerate or hinder diffusion across the network.

In this work, we focus on exploring the effects of adding a second layer of interconnections to an existing layer, creating a two-layered multiplex, and comparing the multiplex with the single-layer network. We identify some notions of the most important nodes in the multilayer setting and compare them to the similar notions in the single-layer setting. All the material and the code for this paper are available in the project's GitHub repository.

1.1 Motivating Example

Consider a scenario involving the spread of a research article within a scientific community. In this community, researchers are connected through multiple social media platforms, each serving a distinct purpose in their professional and personal lives. These platforms can be represented as separate layers in a multiplex network.

LinkedIn Layer: This layer captures professional interactions where researchers share their professional achievements, publications, and industry-related news. LinkedIn is primarily used for professional networking and career development. When a researcher publishes a new article, they are likely to post about it on LinkedIn to inform their professional contacts, including colleagues, collaborators, and industry professionals.

Twitter Layer: This layer represents a more diverse interaction platform where researchers can engage with a broader audience, including both professional peers and the general public. On Twitter, researchers might share links to their new articles, engage in discussions about their work, and participate in trending conversations. The audience that a person on Twitter can reach has some overlap with connections on LinkedIn; however, this overlap varies depending on the content shared on both platforms.

Instagram Layer: This layer encompasses social interactions primarily focused on personal life, visual content, and casual updates. Researchers typically do not share their professional work on Instagram, as the platform is more suited for personal and lifestyle content. However, they might share glimpses of their daily routines, conference travels, or behind-the-scenes moments from their research activities, indirectly contributing to their professional presence.

This multiplex structure illustrates how information diffusion is influenced by the interplay between different social media platforms. A single-layer network model would fail to capture the nuanced pathways through which the article spreads, potentially missing critical dynamics that drive the diffusion process. By examining this scenario through the lens of a multiplex network, researchers can gain a comprehensive understanding of how scientific knowledge propagates and identify strategies to facilitate or control the spread of information across complex social systems.

2 Related Work

Information diffusion in multiplex networks has garnered significant research interest due to its complex nature and practical implications in various domains, including social influence, opinion dynamics, and epidemic spreading. This section reviews key contributions in the field, emphasizing the identification of influential spreaders, the role of network layers, and the dynamics of information diffusion.

Identifying key nodes that maximize information spread is critical for understanding and controlling diffusion processes. (Liu et al., 2023) focuses on finding influential spreaders in multiplex networks with asymmetric interactions, highlighting the importance of considering interaction asymmetry in diffusion models. Similarly, (Zhou et al., 2023) presents a method to rank influential spreaders by weighting the influence of different layers, emphasizing that not all layers contribute equally to the diffusion process. (Rahmede et al., 2017) further explored the centralities of nodes and the influences of layers, emphasizing the importance of layer-specific centrality measures in large multiplex networks.

Various dynamic models have been developed to understand how information spreads across multiplex networks. The majority-vote dynamics model, as explored by (Choi and Goh, 2019), examines opinion dynamics in networks with two layers, highlighting how inter-layer interactions affect consensus formation. The Deffuant model, applied to one-dimensional multiplex networks by (Shang, 2015), provides insights into opinion formation through repeated averaging processes among agents. Additionally, (Alvarez-Zuzek et al., 2017) studied epidemic spreading influenced by opinion exchanges on vaccination, demonstrating the interplay between opinion dynamics and disease spread on a multiplex network model. (Wang et al., 2021) tries to design an extended influence spreading model to simulate the influence diffusion process in multiplex networks, based on which a memetic algorithm is developed to find the seeds that are influential in all network layers.

Maximizing the spread of influence through networks is another crucial aspect of information diffusion. (Amato et al., 2017) propose a model of opinion competition where individuals are organized according to two different structures in two layers.

We aim to continue this line of work by looking at the multiplex networks and finding the most important nodes by using centrality measures.

3 Dataset Description

We used two real-world datasets in this study. Acquiring multiplex network real-world data is a hard task in general for multiple reasons. For starters, not everyone has their social media accounts under the same name; therefore finding their accounts on various platforms could be impossible. Researchers mostly use the APIs that platforms offer for collecting data, but this brings in privacy concerns, the data collected from one platform will be anonymized, and we cannot use that data to collect the same nodes on another social media.

We were able to find two studies that managed to collect multiplex social media data (Sapiezynski et al., 2019) and (Dickison et al., 2016) where we specifically use the multiplatform social media data (Celli et al., 2010).

3.1 Copenhagen Study Dataset

This data represents a multiplex (temporal) network that connects a population of more than 700 university students over a period of four weeks. The layers are a “calls” network, an “SMS” network, and a “Facebook friendship” network. The first two layers are weighted directed networks, and the last one is a binary (unweighted and undirected) network.

The data consists of 924 calls between 536 nodes for the “calls” network, and 1303 texts between 568 nodes in the “SMS network”. We do not use the weights in calls and SMS networks and we assume that the probability of sharing information with a node in all layers is independent of the weights of the connections between them.

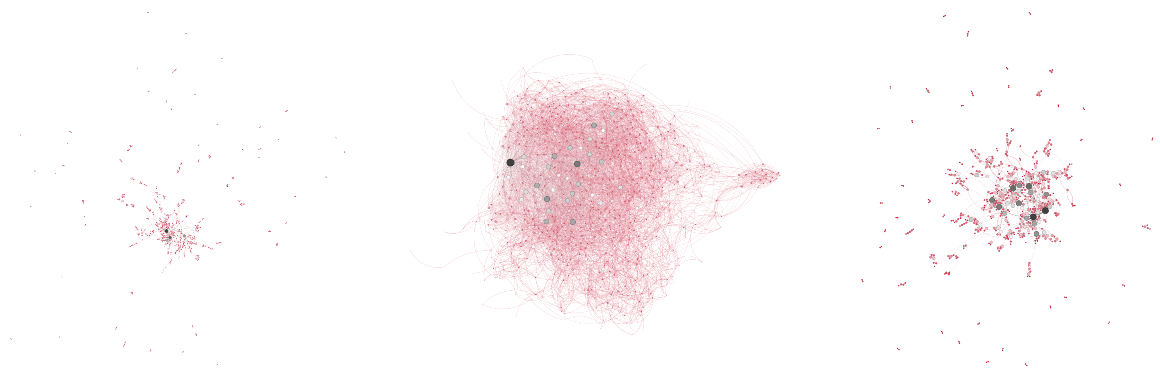


Figure 1: Layers visualization for comparison: Calls layer (left), Facebook Friendship layer (middle), and SMS layer (right)

As we can see in Figure 1, the layers have a major difference in their structure, and the Facebook layer is much more connected than the other two layers. Even though this is a property of online and in-person social dynamics, it could be problematic for our study since Facebook layer is dominant and the multiplex could be looked at as a perturbation of the Facebook layers. This perspective could help us in many ways in both mathematical and real-world simulations of multiplex networks of social interactions over the hybrid in-person and online platforms.

However, we wanted to look into a more equal setting where layers had equal levels of connectivity. For this reason, we used the multi-platform dataset of Friendfeed and Twitter.

3.2 Friendfeed and Twitter Dataset

This anonymized dataset has been obtained starting from Friendfeed, a social media aggregator (Magnani and Rossi, 2011). In this system, while users can directly post messages and comment on other messages, much like on Facebook and other similar online social networks, they can also register their accounts on other systems. The original data acquisition consists of 322,967 users who registered at least one service outside Friendfeed, with a total number of 1,587,273 services. From these, two multilayer networks were retrieved, one with users who registered exactly one Twitter account and whose Twitter account was associated to exactly one Friendfeed account (ff-tw) and one smaller dataset with an additional YouTube layer (ff-tw-yt).

We will work with the latter, but we will exclude the YouTube layer since the connections of that layer are only around 4% of all connections. The other two layers each form around 50% of the rest of the connections, meaning that we have a multiplex network with two layers that have similar connectivity. The network has 6,407 nodes and 74,862 edges in total. The Twitter layer with 5,702 nodes and 42,324 edges, and Friendfeed layer has 5,540 nodes with 31,921 edges.

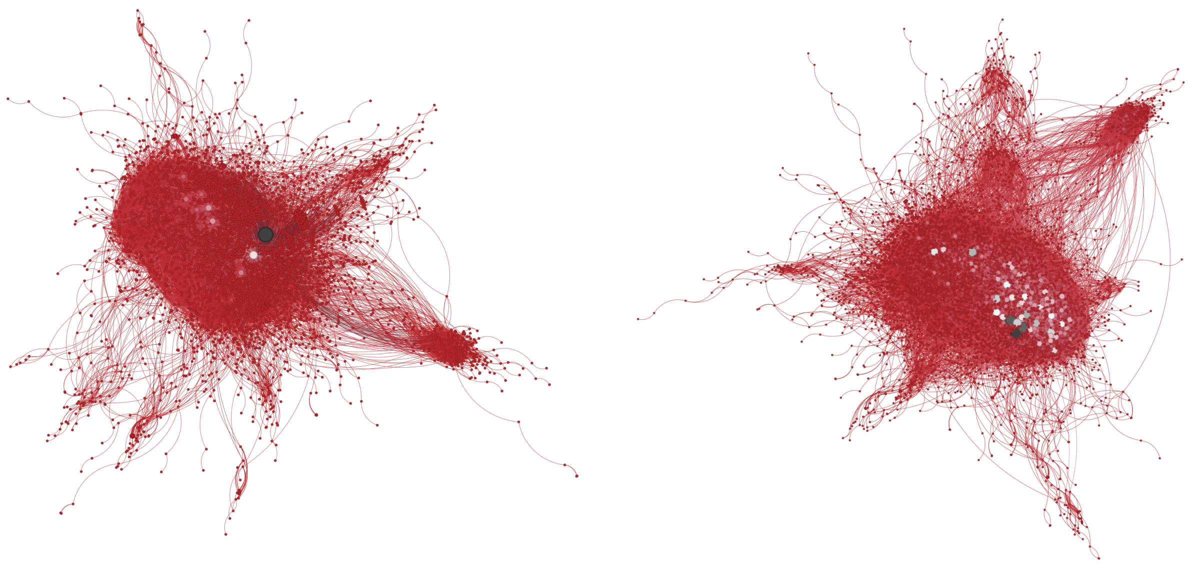


Figure 2: Layers visualization for comparison: Friendfeed layer (left) and Twitter layer (right)

As seen in the figures, this dataset is more dense and has more connections and nodes which will help more for making general comments and testing the model.

4 Simulations and Numerical Experiments

4.1 Code Explanation and Assumptions

For simulating the spread of information within a network, when a node is informed, it can inform each of its neighbors with probability p. The goal is to find the average fraction of nodes that will be informed over some iterations.

We assumed that for an informed node in one network, its corresponding node in the other network will also be informed (similar to linked social media accounts like Instagram and Facebook). Each node makes its neighbors informed by different probabilities, p and q, in each layer, respectively.

The methods we consider for selecting initially informed nodes include completely random selection, highest betweenness centrality, highest eigenvector centrality, and highest performance determined by brute force search.

For both betweenness and eigenvector centralities, we sorted the nodes based on their centrality measures and selected the top n most central ones. In the brute force search method, to identify the n nodes with the highest performance, we spread the information from each node in isolation and calculated the number of informed nodes at the end of the diffusion process [1]. The node with best performance was selected and stored. For selecting the second best node by performance, we repeated the process, starting with the first node already chosen and informed. After selecting the two best nodes, the third one was chosen by considering the cooperation with the first two best nodes. This procedure continued iteratively until the best n nodes were selected.

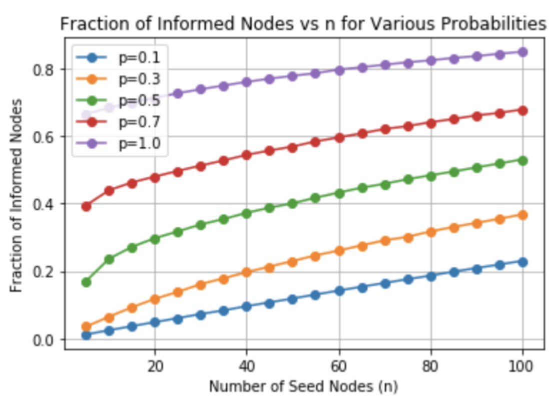
4.2 Results

In this section, all results are related to the Copenhagen Study Dataset, but as functions are reproducible, they easily can be applied to the other dataset. [2]

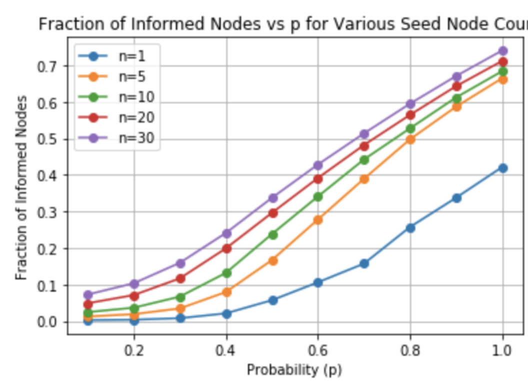
4.2.1 Single-layer Networks

Figure 3a shows fraction of informed nodes versus different number of seed nodes (which is the random number of informed nodes in the initialization). It can be seen that, as the probability of informing the neighbors by each node increases, the fraction of informed nodes for each n, increases.

Figure 3b indicated that for each specific probability, if the number of initially informed nodes is higher, the fraction of informed nodes is greater.



(a) Fraction of Informed Nodes vs. Different Number of Seed Nodes for Various Probabilities (Calls Network)



(b) Fraction of Informed Nodes vs. Probability for Different Number of Seed Nodes (Calls Network)

Figure 3: Comparison of Informed Nodes for Different Parameters in the Calls Network

[1]We use this definition as a measure of performance throughout the paper.

[2]Due to the delayed access to the Friendfeed-Twitter dataset, we were unable to complete this task within the given timeframe.

Same as Calls network, for the SMS and Facebook friendship network, with higher probability of sharing the information, less number of random nodes is needed to have higher fraction of informed nodes after iterations (Figure 4). Based on figure 4, we can conclude that Facebook friendship network is more dense, as with less number of initially informed nodes, it has higher fraction of informed nodes, compared to SMS layer. This was expected since we have the visualization of layers, but for larger networks which we cannot visualize, this can be even more informative.

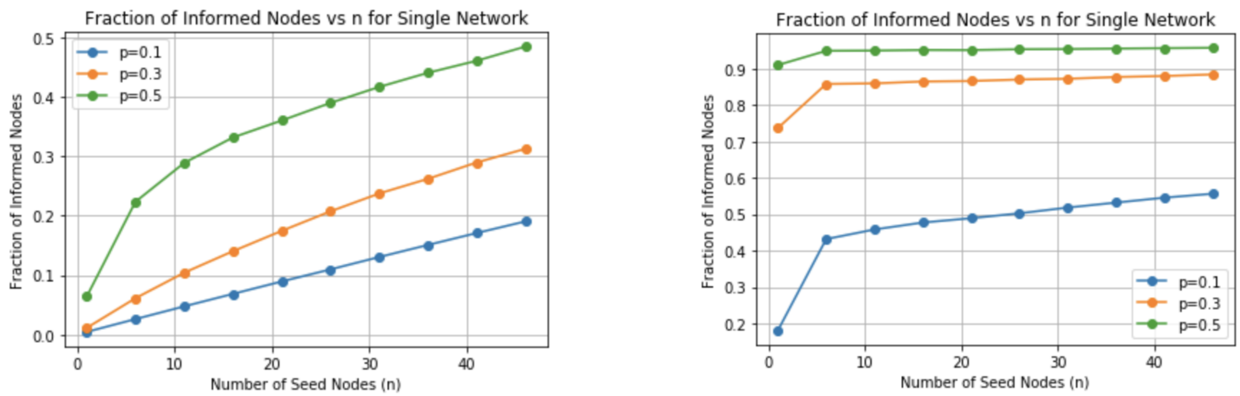


Figure 4: Fraction of Informed Nodes vs. Different Number of Seed Nodes for Various Probabilities: SMS Network (left) Facebook Friendship Network (right)

4.2.2 Multiplex Networks

Compared to figure 3 and 4, while the random nodes are selected in multiplex network, better performance can be achieved with smaller number of initially informed nodes (figure 5 which is related to multiplex network of SMS and Facebook friendship layers). Also, in figure 5, it can be seen that nearly all the nodes are informed even with probability of 0.3 of sharing the information for each node in both layers. It is not exactly 100% as there exist some isolated nodes in the network.

Assuming that we are aiming to inform 80% of the nodes (success), while the probability of sharing information by each node is 0.2, we would need 210 initially informed nodes for SMS Network, 60 for Facebook network, and 10 when we have the Multiplex network. (figure 6)

This result is interesting because even though Facebook layer is dominant in the multiplex, the difference between multiplex and single-layer is not negligible.

4.2.3 Notions of Centrality

In figure 7, it is obvious that with same p, Facebook network is more dense, as less number of most central nodes, achieves better performance. Also, considering figure 4, when number of selected nodes is not many, selecting based on betweenness centrality has better performance than random method, while we see the opposite when the number of initially

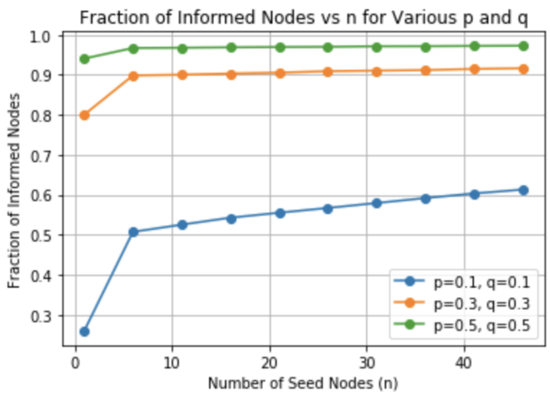


Figure 5: Fraction of Informed Nodes vs. Different Number of Seed Nodes for Various Probabilities p (SMS layer) and q (Facebook friendship layer)

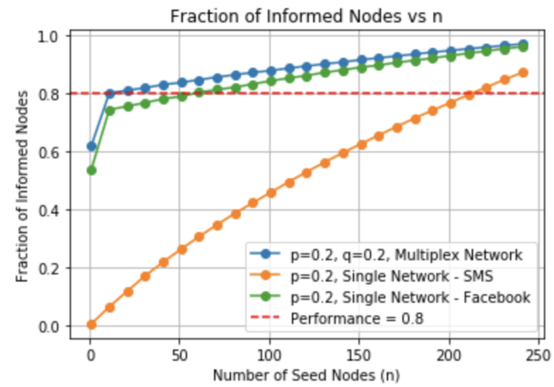
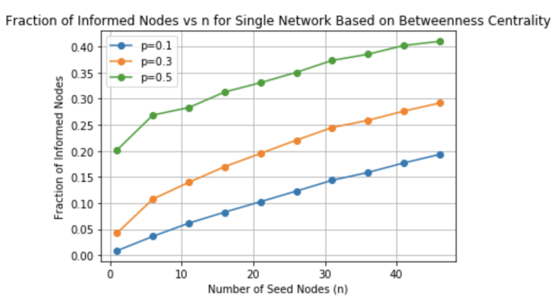
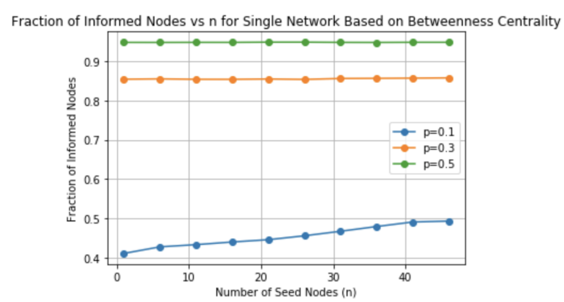


Figure 6: Comparison of Fraction of Informed Nodes vs. Different Number of Seed Nodes for SMS Network, Facebook Friendship Network, Multiplex Network

informed nodes is higher, so random selecting could be better in that scenario[3]. The reason behind that could be existence of some nodes that are not connected to other parts of the graph, so they do not have high centralities. As a result they can not be selected by centrality method, while random way may select them.



(a) Fraction of Informed Nodes vs. Different Number of Nodes for Various Probabilities Based on Betweenness Centrality (SMS Network)



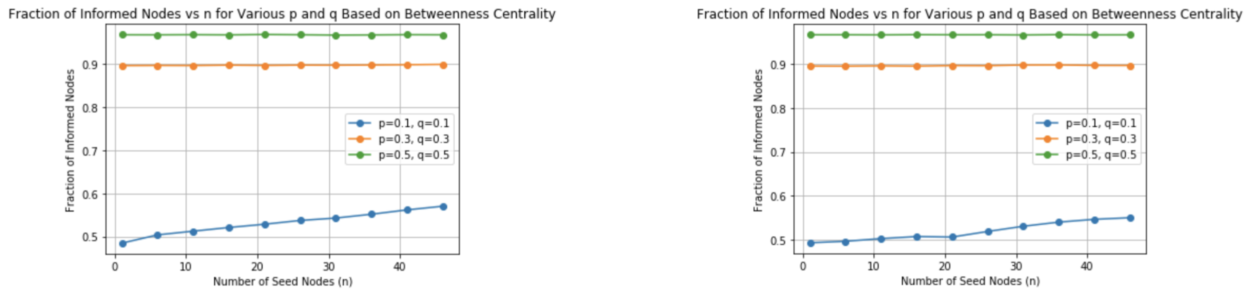
(b) Fraction of Informed Nodes vs. Different Number of Nodes for Various Probabilities Based on Betweenness Centrality (Facebook Network)

Figure 7: Comparison of Informed Nodes for Different Parameters in the SMS and Facebook Networks Based on Betweenness Centrality

When the nodes are selected based on the highest betweenness centrality in the Facebook layer, compared to SMS layer, fraction of informed nodes is higher (figure 8). Considering the single network plots, it can be seen that even with less than 10 initially informed nodes, multiplex network has higher informed nodes at the end, while for both single networks the result is weaker. For example, in figure 7a, for SMS network with p equals to 0.1, with 50 nodes, around 20 percent of nodes are informed at the end, while the similar parameter in figure 8a is near 60 percent. In addition, like single networks, in multiplex one, when the number of n is lower, choosing based on betweenness centrality lead to

[3]We must note that we are assuming all the nodes in the network are accessible and we can seed them with information, which is not necessarily the case in the real-world.

better performance, while this is the opposite case, meaning when n is large, random choosing is better.

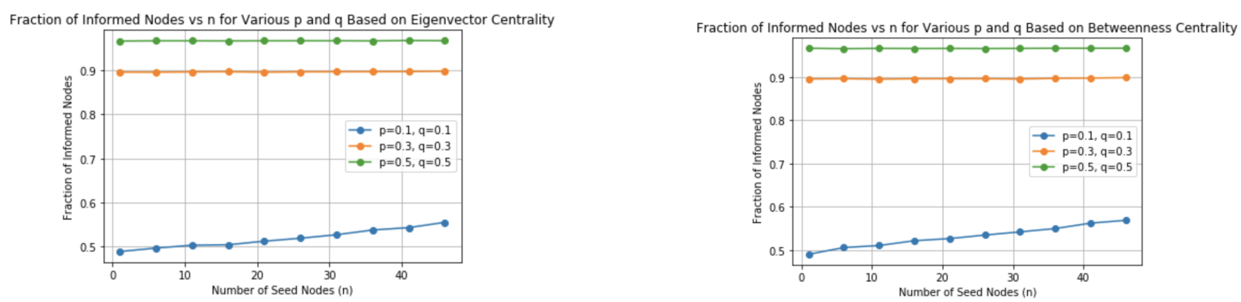


(a) Fraction of Informed Nodes in Multiplex Network vs. n for Various p Based on Highest Betweenness Centrality in SMS Layer

(b) Fraction of Informed Nodes in Multiplex Network vs. n for Various p Based on Highest Betweenness Centrality in Facebook Layer

Figure 8: Comparison of Informed Nodes for Different Parameters in the Multiplex Network Based on Betweenness Centrality

Plots related to choosing nodes based on highest eigenvector centrality for signle networks are available on the project's GitHub repository. Figure 9, shows that choosing nodes based on eigenvector centrality is somehow similar to choosing based on betweenness centrality, although it has slightly worse performance.



(a) Fraction of Informed Nodes in Multiplex Network vs. n for Various p Based on Highest Eigenvector Centrality in SMS Layer

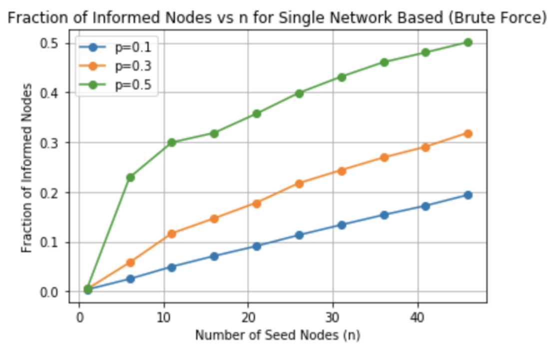
(b) Fraction of Informed Nodes in Multiplex Network vs. n for Various p Based on Highest Eigenvector Centrality in Facebook Layer

Figure 9: Comparison of Informed Nodes for Different Parameters in the Multiplex Network Based on Eigenvector Centrality

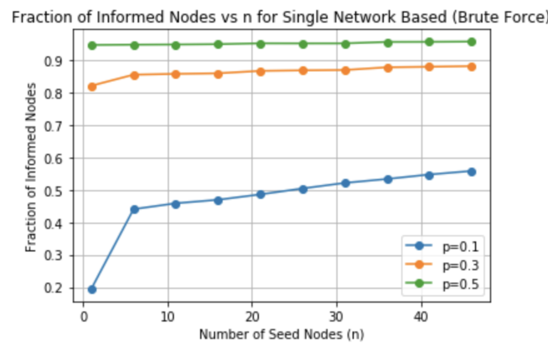
By choosing the nodes with brute force search method, the results are almost similar to random selection for single networks (figure 10). Even for multiplex network with choosing nodes based on brute force search, the performance is similar to random selection (figure 11). In multiplex network, the nodes selected by brute force search method, may not be the best in one layer, but they are the best considering both layers.

5 Future Works and Conclusion

The study of information diffusion on multiplex networks is a burgeoning field with numerous opportunities for further exploration. Building on this project, several promising



(a) Fraction of Informed Nodes in SMS Signle Network vs. n for Various p Based on Brute Force Search Centrality



(b) Fraction of Informed Nodes in Facebook Network vs. n for Various p Based on Brute Force Search

Figure 10: Comparison of Informed Nodes for Different Parameters in the Single Networks Based on Brute Force Search

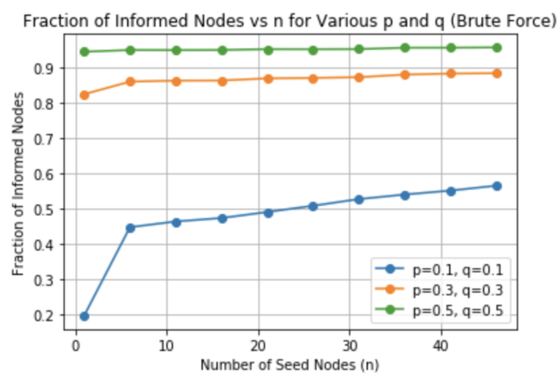


Figure 11: Fraction of Informed Nodes in Multiplex Network vs. n for Various p Based on Brute Force Search in Multiplex Network

directions for future work are outlined below:

The mathematical exploration of diffusion processes is a critical avenue for future research. One area of focus is the development and analysis of mathematical models that capture the dynamics of information diffusion across different layers of multiplex networks. This can involve deriving closed-form solutions or approximations for the spread of information and identifying critical parameters that influence diffusion rates, or performance as we defined in this work.

Introducing new notions of centrality can be another important direction for future work. Developing new centrality measures that account for the influence of nodes within specific layers of a multiplex network is key. Understanding the impact of network topology on information diffusion is another promising research direction. Future work can study how different topological features of multiplex networks, such as community structure, clustering, and degree distribution, influence the diffusion process. Identifying topological characteristics that enhance or inhibit information spread can inform strategies for optimizing network structures to facilitate diffusion or prevent it.

All in all, we aim to continue this project to achieve more intriguing results in either of these directions.

References

Alvarez-Zuzek, L. G., La Rocca, C. E., Iglesias, J. R., and Braunstein, L. A. (2017). Epidemic spreading in multiplex networks influenced by opinion exchanges on vaccination. *PloS one*, 12(11):e0186492.

Amato, R., Kouvaris, N. E., Miguel, M. S., and Díaz-Guilera, A. (2017). Opinion competition dynamics on multiplex networks. *New Journal of Physics*, 19(12):123019.

Castellano, C., Fortunato, S., and Loreto, V. (2009). Statistical physics of social dynamics. *Reviews of modern physics*, 81(2):591.

Celli, F., Lascio, F. M. L. D., Magnani, M., Pacelli, B., and Rossi, L. (2010). Social Network Data and Practices: the case of Friendfeed. In *International Conference on Social Computing, Behavioral Modeling and Prediction*, Lecture Notes in Computer Science. Springer Berlin Heidelberg.

Choi, J. and Goh, K.-I. (2019). Majority-vote dynamics on multiplex networks with two layers. *New Journal of Physics*, 21(3):035005.

Dickison, M. E., Magnani, M., and Rossi, L. (2016). *Multilayer Social Networks*. Cambridge University Press.

Ji, P., Ye, J., Mu, Y., Lin, W., Tian, Y., Hens, C., Perc, M., Tang, Y., Sun, J., and Kurths, J. (2023). Signal propagation in complex networks. *Physics reports*, 1017:1–96.

Li, M., Wang, X., Gao, K., and Zhang, S. (2017). A survey on information diffusion in online social networks: Models and methods. *Information*, 8(4):118.

Liu, Y., Zeng, Q., Pan, L., and Tang, M. (2023). Identify influential spreaders in asymmetrically interacting multiplex networks. *IEEE Transactions on Network Science and Engineering*, 10(4):2201–2211.

Magnani, M. and Rossi, L. (2011). The ml-model for multi-layer social networks. In *2011 International Conference on Advances in Social Networks Analysis and Mining*, pages 5–12.

Newman, M. E. (2002). Spread of epidemic disease on networks. *Physical review E*, 66(1):016128.

Rahmede, C., Iacovacci, J., Arenas, A., and Bianconi, G. (2017). Centralities of nodes and influences of layers in large multiplex networks. *Journal of Complex Networks*, 6(5):733–752.

Sapiezynski, P., Stopczynski, A., Lassen, D. D., and Lehmann, S. (2019). Interaction data from the copenhagen networks study. *Scientific Data*, 6(1):315.

Shang, Y. (2015). Deffuant model of opinion formation in one-dimensional multiplex networks. *Journal of Physics A: Mathematical and Theoretical*, 48(39):395101.

Singh, S. S., Singh, K., Kumar, A., Shakya, H. K., and Biswas, B. (2019). A survey on information diffusion models in social networks. In *Advanced Informatics for Computing Research: Second International Conference, ICAICR 2018, Shimla, India, July 14–15, 2018, Revised Selected Papers, Part II 2*, pages 426–439. Springer.

Wang, S., Liu, J., and Jin, Y. (2021). Finding influential nodes in multiplex networks using a memetic algorithm. *IEEE Transactions on Cybernetics*, 51(2):900–912.

Zhou, X., Bouyer, A., Maleki, M., Mohammadi, M., and Arasteh, B. (2023). Identifying top influential spreaders based on the influence weight of layers in multiplex networks. *Chaos, Solitons Fractals*, 173:113769.