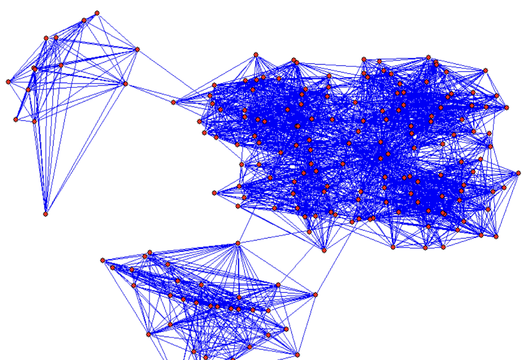
Analysis of the relationship between Structural Holes and Influencers over the Last.fm social network

CSE 533, Social Network Analysis

Socially Unsocial

Rucha Sanghvi	AU2040012
Dhruvam Bhalodiya	AU2040041
Harshal Trivedi	AU2040214
Tej Thakar	AU2040262



Ahmedabad University

Contents

1 | Abstract

In order to comprehend the structure and behavior of complex systems, such as social networks and information networks, useful methods include network analysis and graph theory. This research study investigates structural holes, which are gaps or voids between nodes in a network, and its ramifications in a variety of domains. Furthermore, we investigate the concepts of direct and indirect effects of nodes in directed graphs, taking into consideration elements such as edge weights, outdegrees, and all feasible pathways. In order to assist researchers in better comprehending and creating applications for these methodologies, we want to explain the ideas of structural gaps and effects on network analysis through this research. The whole research study revolves around the basic ideation of structural holes, direct influence, and indirect influence of the whole network. And the conclusions and results confront the scenarios of roles of indirect influence over direct influence and how the nodes with more structural hole score may or may not be proper influencers even though it is the best intermediator to two or more different communities in the network.

2 | Introduction

Influence is a strong force that may shape our attitudes, behaviors, and decisions. It can come from a multitude of sources, such as individuals, groups, or societal norms. Understanding the importance of influence is crucial for individuals, organizations, and society at large. Powerful people and organizations may have an impact on politics, public opinion, and cultural trends.

Now in this report, we have done all the analysis on Last.fm users. Last.fm is a music streaming and social networking website that allows users to listen to music, create personal profiles, and connect with other music lovers. Now, in Last.fm there is a feature which is called scrobble. Scrobble is nothing but the information of the user and their listening records. The information contains the username, tracks heard by the user, album heard by the user, timestamp of when the user has heard the track, friends list of the user, geographical location of the user based on the country name, and many more. The process of getting the data from the scrobble is called scrobbling.

The motivation for studying the temporal influence of Last.fm lies in the growing importance of music streaming services and their impact on music consumption patterns. This report is particularly interested in how the listening habits of Last.fm users change over time and how this temporal effect affects the popularity and influence of specific artists, genres, and songs. Massive amounts of user listening data allow us to identify patterns and trends that would be challenging to identify using more traditional methods and based on the sharing of songs/tracks we can identify which user is the most influential among all.

The overall report is an analysis based on Last.fm. Here the analysis is purely based on how one user can influence the other user based on their shared songs/tracks. This is revised on two broader aspects of influence: Direct and Indirect influence. Now based on those two terms, this report targets how the indirect influential role also plays a major role in finding the better influencer rather than just having a corps of direct influence and a number of friends. This report also works on the idea of how structural holes may or may not be the best influencer based on their structural hole score. So based on all of this, the paper concludes with the main discussion of an influential analysis of this kind of song-sharing platform.

3 | Related Work

3.1 | *Temporal Influence over last.fm Social Network*

In this paper they justify the existence of the social influence by considering the temporal behavior of Last.fm users. In order to clearly distinguish between friends sharing the same interest, especially since Last.fm recommends friends based on similarity of taste, they separated the timeless effect of similar taste from the temporal impulses of immediately listening to the same artist after a friend. They measured strong increase of listening to a completely new artist in a few hours period after a friend compared to non-friends representing a simple trend or external influence. In their experiment to eliminate network independent elements of taste, they improved collaborative filtering and trend based methods by blending with simple time aware recommendations based on the influence of friends.

3.2 | *Unbiased metrics of friend's influence in multi-level networks*

They demonstrate that the currently existing metrics of friends' influence are biased by the presence of highly popular items in the data, and as a result can lead to an illusion of friends influence where there is none. They correct for this bias and develop three metrics that allow to distinguish the influence of friends from the effects of item popularity, and apply the metrics on real datasets. They use a simple network model based on the influence of friends and preferential attachment to illustrate the performance of our metrics at different levels of friends' influence.

3.3 | *To what extent homophily and influencer networks explain song popularity*

Here they determine how musical homophily can be used to predict song popularity. The study is based on an extensive dataset from the last.fm online music platform from which we can extract social links between listeners and their listening patterns. To quantify the importance of networks in the spreading of songs that eventually determines their popularity, they use musical homophily to design a predictive influence parameter and show that its inclusion in state-of-the-art machine learning models enhances predictions of song popularity. The influence parameter improves the prediction precision (TP/(TP+FN)) by about 50 percentage from 0.14 to 0.21, indicating that the social component in the spreading of music plays at least as significant a role as the artist's popularity or the impact of the genre.

4 | Model, Approach and Methods

4.1 | Dataset

Here we have fetched two types of data.

1. User-Friends Data
   The dataset consists of in total 91,224 relations between users. The dataset gave us the information about connection between two users.

2. User Track List (Quarter Year)
   The dataset consists of data of each track listened by every user within the time period of Jan 2021 to Apr 2021.

We then in order to get the track list of one user, their friend list and track list of their friends merged both datasets based on tracks.

4.2 | Threshold Detection

There are many ways of getting to the threshold for a user-user interaction. We used the method of double averaging the quarter data firstly monthwise and after quarterly. The threshold we obtained from averaging monthwise ranged between 5000 to 7000 minutes. And quarterly was 5230 minutes.

4.3 | Direct Influence

Our approach in order to find the direct influence was to perform Breadth First Search (BFS) for the given user. This give the broadness of the graph. The reason of knowing the direct influence is to know how much influence a person has or the connectivity a user has.

Think of the graph as a binary search tree. Each parent node is connected to many different child nodes. The calculation of the immediate child nodes a particular user has which has listened to the particular track in the decided threshold of time gives us the direct influence of that user.

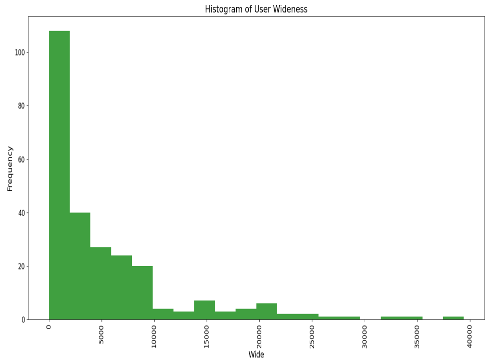


Figure 4.1: Direct Influence vs Frequency

4.4 | Indirect Influence

Indirect influence can be calculated by performing the Depth First Search (DFS) of the given user. This gives us the depth of a graph. Reason for performing DFS is to know the reach of a user in a community. The more the reach the more the indirect influence is and the more the community growth is. Here we did the DFS till 2 depths due to computational constraints.

Think of the graph as a binary search tree. Each parent node is connected to many different child nodes and after that in turn is connected to a leaf node. Summing the no. of nodes till the leaf node a particular user is connected which has listened to the particular track in the decided threshold of time gives us the indirect influence of that user.

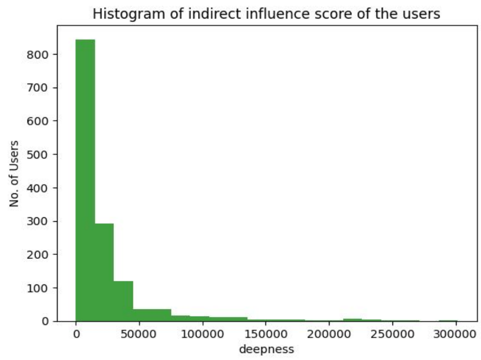


Figure 4.2: Indirect Influence vs Frequency

4.5 | Structural Holes

The structural hole score is a social network analysis metric that assesses how well an individual or organization bridges gaps or "structural holes" in a network with other individuals or organizations. A high structural hole score indicates that an individual or organization is important in a network by connecting previously isolated areas.

Now in this case the whole structural hole score will refractor the premise of influence by mentioning how the node with a high structural hole score has minimal influence even though it acts as an intermediator between two groups/organizations. The results which are mentioned in the graphs will suggest the prospect of how structural holes can have an impact and what is the relationship between influence and structural holes. Now the two graphs (Fig 5.1 and 5.2) are based on the comparison of direct and indirect influence which is based on deepness and wideness of the graph and from this we can conclude that though, generally nodes with higher value of structural holes are predicted to be an influencer, It is not always necessary that an influencer is a structural hole. And structural holes can and cannot be an influencer.

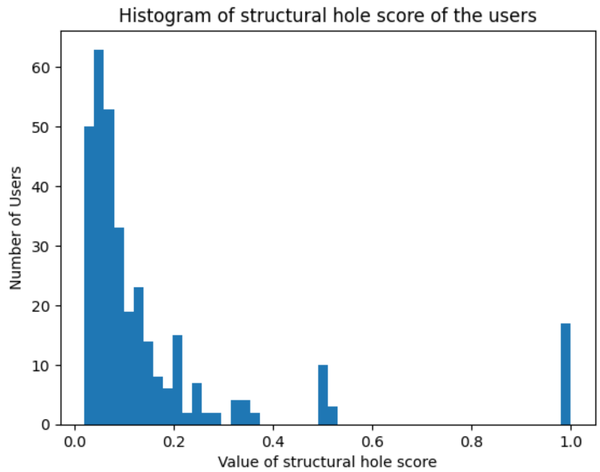


Figure 4.3: Structural Holes vs Frequency

5 Results

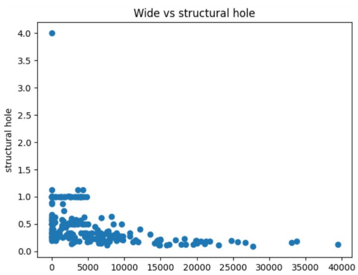


Figure 5.1: Direct Influence vs Structural Holes

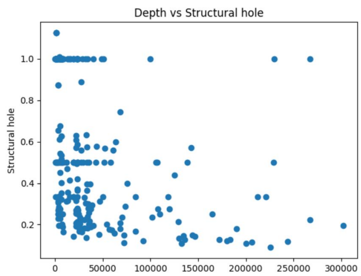


Figure 5.2: Indirect Influence vs Structural Holes

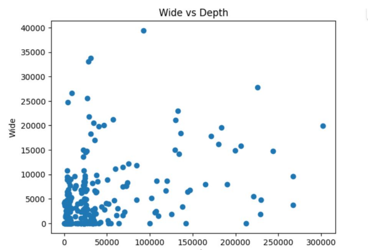


Figure 5.3: Direct Influence vs Indirect Influence

We found intriguing correlations between influencer and structural holes, as well as between wideness and depth of influence, based on the analysis of the graph derived from the CSV file containing user-friend associations with edge weights.

First, we discovered that nodes with high centrality metrics, like brokerage, constraint, and effective size, does not frequently display traits of network influencers. These nodes are not likely to frequently fill structural holes, linking nodes or groups that would otherwise be isolated, and they aid in the distribution of resources and the flow of information. As a result of their optimal locations for information propagation and influence, nodes with high structural hole metrics are unlikely to be more influential in the network as seen in the above figures.

Additionally, we found a significant correlation between the deepness of effect and the breadth of influence using metrics like outdegrees, edge weights, and all feasible pathways. High outdegree nodes can impact more nodes in the network because they have a big number of outgoing edges or connections, which

indicates a significant number of connections. As their effect can spread through stronger interconnections in the network, nodes with greater edge weights, which signify stronger connections or relationships, are more likely to have a wider influence. Additionally, nodes that are connected to other nodes through many paths and are a member of more paths tend to have a stronger influence since their influence can go along different paths within the network. But having more out-going edges cannot always mean that they can reach to many amongst the network.
Our analysis revealed the significant role of indirect influence in the network in addition to the relationships between influencers, structural holes, wideness, and depth of influence, emphasising that having a larger network in terms of followers or connections does not always imply having more influence.

6 | Conclusion

In this study, using a dataset generated from a CSV file comprising user, friend, and edge weight information, we examined the correlations between influencers, structural gaps, wideness, and depth of influence in networks. Our research showed that indirect influence, which can and cannot occur only when nodes with high betweenness centrality serve as bridges along the shortest paths, can have a considerable impact on how the network distributes information or allocates resources. Additionally, we found that the depth of influence—as determined by elements like edge weights and all viable paths—is a more accurate predictor of influence than the breadth of the network, as assessed by the number of connections or followers.
Thus, influencer is not always likey to be a structural hole, but a structural hole can be an influencer.
To gain a more thorough understanding of the intricate dynamics of influence in networks, additional research is required to validate and generalise our findings in different datasets and contexts. It is crucial to keep in mind that our findings are based on the specific dataset and study context. Nevertheless, our research highlights the significance of taking into account indirect influence and the depth of influence in addition to the wideness of the network when analysing complex networks and offers insightful information about the relationship between influencer, structural holes, wideness, and deepness of influence in networks.

7 | Future Work

To confirm the robustness and generalizability of the findings, additional research might be done to validate them using various datasets and circumstances. This could involve examining the relationship between influencers, structural gaps, wideness, and depth of influence in various contexts utilising a variety of datasets from various domains, such as social networks, communication networks, or organisational networks. Additionally, taking into account regional information and user information can provide more information regarding influence in various regions and among various age groups.

8 | References

1. Pálovics, Róbert Benczúr, András A, Temporal influence over the Last.fm social network, (2013)
2. Vidmer, Alexandre Medo, Matúš Zhang, Yi-Cheng, Unbiased metrics of friends' influence in multi-level networks, EPJ Data Science, (2015)
3. Niklas Reisza, Vito D. P. Servedioa, Stefan Thurner, To what extent homophily and influencer networks explain song popularity (2022)