



UNIVERSIDAD
POLITÉCNICA
DE YUCATÁN



Social Network Analysis

Social Networks Analysis

Student:

Can Montero Christian Javier

Professor:

Gamboa Angulo Didier Omar

Abstract:

This report analyzes user interactions within a specific Reddit thread from the subreddit 'r/KGATLW'. Utilizing network analysis, the study maps these interactions into a directed graph, revealing a network with moderate clustering and distinct community structures. Key metrics such as centrality identify influential users, while link prediction techniques forecast potential future interactions. The findings provide insights into community dynamics, offering valuable information for enhancing engagement and understanding social interactions on Reddit.

August 3, 2024

I. INTRODUCTION

Online forums have taken on an important role in the digital age in determining how people communicate, exchange information, and create communities centred around common interests. One of the biggest and most varied online communities, Reddit, has thousands of subreddits, each a microcosm of conversation on subjects ranging from the common to the obscure. One of these is the subreddit r/KGATLW, which is devoted to King Gizzard & the Lizard Wizard's fans and followers.

The band, known for its eclectic style and prolific output, has cultivated a passionate and engaged fanbase. Discussions within this subreddit often extend beyond simple song discussions, delving into detailed exchanges about concert experiences, interpretations of lyrics, and anticipations of new releases. The selected thread for this network analysis captures a vibrant snapshot of such interactions, where users not only share information but also form social bonds over common musical interests.

This report aims to dissect the structural and relational dynamics within this specific thread, applying network analysis to map out and understand the patterns of user interactions. By examining how users respond to and engage with each other, we can gain insights into the underlying social fabric of the subreddit. Such an analysis is not only relevant for understanding this particular community but also serves as a case study for similar dynamics across other digital forums. The findings aim to highlight the roles of key users, the pathways of information flow, and the overall cohesion within the community, providing a comprehensive overview of how digital spaces facilitate complex social interactions in the age of online communication.

II. MAPPING PROCESS

The process of mapping the network of user interactions from the chosen Reddit thread involves the following steps:

- **Data Collection:** Utilizing praw, a Python library for accessing Reddit's API, the script authenticates and accesses the target post. It then recursively fetches all top-level comments and their replies, capturing user interactions as data points.
- **Data Processing:** This raw data is filtered and cleaned to exclude any comments that have been deleted or removed, ensuring the analysis is based on genuine user interactions. The cleaned data includes details about the comment authors and the hierarchical structure of replies.

- **Network Construction:** With the data prepared, a directed graph is constructed using networkx. In this graph, nodes represent individual Reddit users, and directed edges represent the direction of communication, in other words, who replies to whom. This directed approach is crucial for understanding the flow of information and influence.

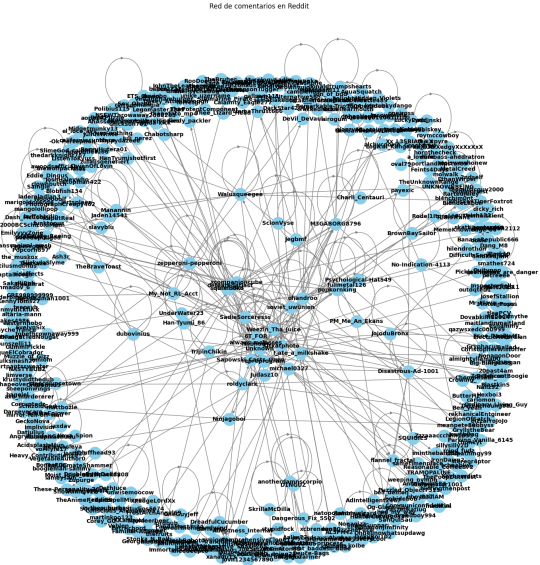


Fig. 1. Reddit's comments network.

III. NETWORK CHARACTERISTICS AND VISUALIZATION

The network constructed from the Reddit comments exhibits several interesting characteristics:

A. Basic Characteristics

- **Network Type:** As a DiGraph, the network aptly represents the directional nature of conversations where replies can distinctly point from one user to another.
- **Global Clustering Coefficient:** A coefficient of 0.583 indicates a moderate tendency among users to cluster into tightly-knit groups, likely around specific topics or threads within the subreddit.
- **Degree Distribution:** The analysis of degree distribution reveals how users are interconnected. Most users have a low degree, suggesting they interact with few others, while a smaller number of highly connected individuals (hubs) have interactions spanning a broad array of users.

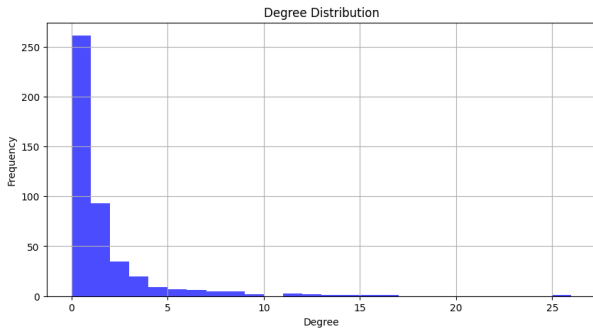


Fig. 2. Degree distribution plot.

B. Centrality Measures

- **Degree Centrality:** Identifies the most active users within the network, those who frequently engage with many others.
- **Betweenness Centrality:** Identifies users who control the flow of information in the network, acting as bridges or connectors between different groups of users.

IV. VISUALIZATION AND LINK PREDICTION

A. Visualization

The network visualization is crafted to enhance understanding of the community structure:

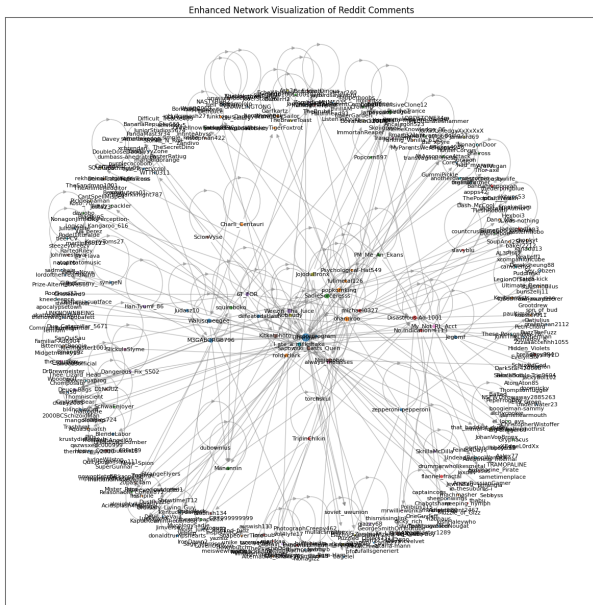


Fig. 3. Link prediction visualizer.

Each node's size is proportional to its degree centrality, highlighting more influential users, and colors differentiate the various communities or clusters within the network, indicating groups of users who interact more frequently with each other.

B. Link Prediction

- **Jaccard Coefficient:** Measures the similarity between the sets of neighbors of two nodes, predicting a link if they share a high proportion of common neighbors.
- **Preferential Attachment:** Assumes that users with many connections are more likely to create new links, thus nodes with higher degrees are more likely to connect.
- **Resource Allocation Index:** A measure that quantifies the amount of resource two nodes share through their common neighbors, useful for identifying potential new interactions in tightly-knit communities.

V. RESULTS

While specific results of link prediction algorithms were not included initially, applying these methods would likely highlight potential connections between active users who share common community ties but have not yet interacted directly. This could be particularly useful for community moderators to foster engagement or for marketers aiming to target influential users within the subreddit.

This comprehensive analysis provides a snapshot of the community dynamics within a Reddit thread, offering insights that could be leveraged for various purposes, including community management, sociological research, or targeted marketing strategies within similar online platforms.

VI. CONCLUSION

The network analysis of the Reddit thread in the 'r/KGATLW' subreddit has highlighted the complex dynamics and structures of online communities. This analysis revealed:

- **Community Structure:** Users cluster around shared interests, which is indicative of the subreddit's role as a gathering place for like-minded individuals.
- **Influential Users:** Centrality measures identified key users who significantly influence community interactions and information flow.

- **Link Prediction:** Techniques used predicted potential future connections that could enhance the network's interconnectedness and foster community growth.

These insights provide valuable perspectives for community management, demonstrating the importance of network science in understanding and enhancing digital social ecosystems. The visual and quantitative analysis offers a foundation for strategic engagement and community development, highlighting the role of influential users and potential areas for network strengthening. This approach underscores the evolving significance of network analytics in managing online communities effectively.

REFERENCES

- SNAP: Signed network datasets: Bitcoin Alpha web of trust network. (n.d.). Stanford.edu. Retrieved July 7, 2024, from <http://snap.stanford.edu/data/soc-sign-bitcoin-alpha.html>
- Newman, M. (2018). *Networks: The empirical study of Networks*. Oxford University Press.
- Zafarani, R., Abbasi, M. A., & Liu, H. (2014).
- Barabási, A.-L. (2016). *Network Science*.
- Menczer, F., Fortunato, S., & Davis, C. (2020). *A First Course in Network Science*.
- S. Kumar, F. Spezzano, V.S. Subrahmanian, C. Faloutsos. *Edge Weight Prediction in Weighted Signed Networks*. IEEE International Conference on Data Mining (ICDM), 2016.
- Alright gizzers, lets hear it. (n.d.). Reddit.Com. Retrieved August 3, 2024, from https://www.reddit.com/r/KGATLW/comments/vftri5/alright_gizzers_lets_hear_it/
- Tanner, G. (2019, January 5). Scraping Reddit data. Towards Data Science. <https://towardsdatascience.com/scraping-reddit-data-1c0af3040768>