P1. Networks Theory

Juan Fernandez Cruz
Data 7A
Universidad Politecnica de Yucatan
Ucu, Yucatan
Email: 2109061@upy.edu.mx

Didier Gamboa
Universidad Politecnica de Yucatan
Ucu, Mexico
Email: didier.gamboa@upy.edu.mx

I. Abstract

This study provides a comprehensive analysis of an Ego network from the social platform Letterboxd, which is centered around a particular user and includes direct and indirect connections within a film enthusiast community. Through methods of social network analysis, including measures of centrality, community detection, and degree distribution, we explore the structural and relational dynamics within the network. The findings illuminate the influence patterns, community structure, and connectivity within the network, offering insights into social interaction dynamics on a niche social media platform. This research contributes to the understanding of digital social networks by highlighting how specific interests and interactions shape the network structure and user influence within Letterboxd.

II. Introduction

Letterboxd, a social platform for film enthusiasts, offers a unique environment for exploring social interactions through film reviews and user connections. Unlike typical social networks that focus on a broad range of topics, Letterboxd specifically caters to those who cherish cinematic experiences, providing a space to share and discover movie reviews and manage personal film collections. The dataset under examination focuses on the Ego network of a particular user, encapsulating their direct connections and interactions within the community. This network comprises intricate connections involving both 'following' and 'followers' relationships across several users, each with a distinct set of interactions. For example, the primary user in this analysis follows and is followed by multiple other users, creating a dense network of over 50 unique connections, highlighting the interconnected nature of social interactions on Letterboxd. This analysis delves into a network that includes not just the primary user's activities but also those of their immediate connections, revealing patterns and influences in film preferences and interaction dynamics.

A. Network Scope and Computational Considerations

In the study of Ego networks, the degree of depth is crucial as it often dictates the complexity and the computational feasibility of the analysis. Typically, the depth of these networks can expand exponentially, which means that even a small increase in depth can lead to a significant growth in the size of the network. This exponential growth can pose considerable computational challenges, especially when dealing with very large datasets.

For this particular analysis of the Letterboxd Ego network, the network maintains a manageable size by limiting the depth to two levels. This constraint ensures that while the network remains computationally tractable, it still retains sufficient complexity to serve as a representative example of larger networks. By focusing on a depth-two network, the analysis captures not only the direct interactions of the primary user but also the interactions of their immediate connections. This approach provides a comprehensive view of the user's social influence and interaction dynamics within the Letterboxd community, while also showcasing the analytical techniques that could be scaled to larger, more complex networks. This balance makes the current study a valuable model for understanding both the potential and limitations of social network analysis in a highly interconnected digital environment.

B. Importance and Applications:

Analyzing the Ego network of a Letterboxd user can provide profound insights into the personal and communal aspects of movie-watching behavior. For the individual, it offers a mirror reflecting their social influence and engagement level within the community, potentially guiding them in enhancing their interaction or exploring new film genres. Academically, such analysis contributes to understanding how cultural products are consumed and recommended in digital spaces. Practically, insights drawn from this network can aid in improving platform features such as personalized recommendations, community engagement strategies, and targeted marketing campaigns. Moreover, the analysis can serve as a foundational study for sociologists and cultural studies scholars interested in the digital consumption of media, demonstrating how social interactions around cultural preferences are structured in an online setting.

C. Network Classification

The network depicted represents a typical Ego network, centered around a specific user within the Letterboxd community. This type of network is characterized by its radial structure, where the central node (the ego) is directly connected to various other nodes (alters), which may or may not be interconnected. In this particular case, the Ego network exhibits both direct and indirect connections, highlighting not

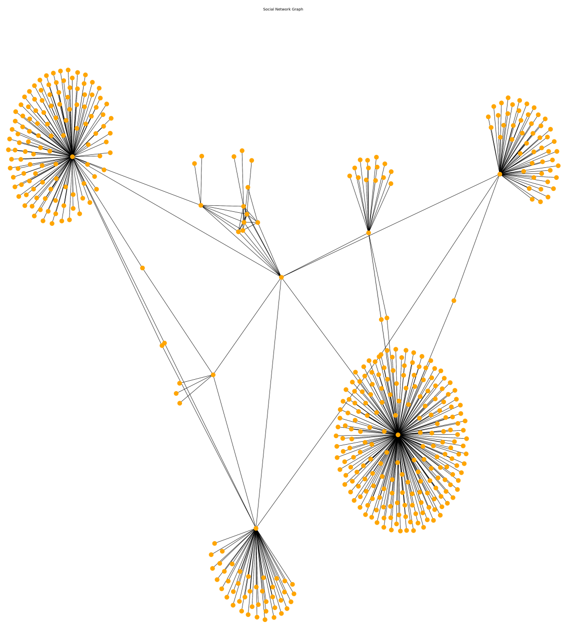

Fig. 1. Letterboxd Ego Network Graph

only relationships where the ego is a direct participant but also where the ego's connections interact among themselves.

From the visual representation in Fig. 1. it is evident that the network is undirected, indicating that the relationships are mutual—both parties acknowledge the connection, typical of social interactions on platforms like Letterboxd where users follow each other. The structure also suggests a level of modularity, with distinct clusters potentially representing different social circles or communities within the larger network. These clusters could delineate groupings based on shared interests in specific movie genres or directorial styles, reflecting the social and interest-based affiliations that naturally emerge in a community-centric platform like Letterboxd.

The network's architecture, while seemingly simple, can become complex very quickly as the degrees of separation increase. The exponential growth of connections at each successive level away from the ego illustrates the dense web of interactions typical in digital social networks. This visualization serves as a practical example for understanding how individual user interactions contribute to the broader dynamics of social connectivity and influence within a specialized community.

III. Network Characteristics

A. Size and Links

The **size** of a network refers to the number of nodes (or vertices), which in this context is 458. This measure indicates the total number of users within the analyzed Ego network. The **number of links** (or edges), totaling 974, represents the connections between these nodes. These metrics are fundamental for understanding the scale and connectivity of the network.

B. Average Path Length

The **average path length** is a measure of the average number of steps along the shortest paths for all possible pairs of network nodes. It is calculated using the formula:

$$L = \frac{1}{n(n-1)} \sum_{i \neq j} d(n_i, n_j)$$

where n is the number of nodes and $d(n_i, n_j)$ is the shortest path length between nodes n_i and n_j. In this network, the average path length is 3.298, suggesting that it takes about three steps to connect any two members, on average. This metric is crucial for understanding the 'small-world' nature of a network, which can significantly affect information spread and accessibility across the network.

C. Clustering Coefficient

The **clustering coefficient** measures the degree to which nodes in a network tend to cluster together. It is defined as:

$$C = \frac{3 \times \text{number of triangles in the network}}{\text{number of connected triples of nodes}}$$

The value in this network is 0.0176, indicating a low level of clustering. This low coefficient suggests that users' followers are not strongly interconnected, a typical characteristic of social networks where individual circles may remain distinct.

D. Average Distance and Diameter

The **average distance** in a network is similar to the average path length, providing an overall indication of the network's 'tightness'. The **diameter** of a network is the longest shortest path between any two nodes. In this case, both the average distance and the network diameter are crucial for understanding the maximum extents of spread and interaction within the network. The diameter is 4, meaning the farthest distance between any two nodes is four steps.

E. Eccentricity

The **eccentricity** of a node is the greatest distance between that node and any other node in the network. For most nodes, the eccentricity is 3, and for certain peripheral nodes, it is 4. This variation highlights the nodes that are more central versus those on the periphery.

F. Radius and Periphery

The **radius** of a network is the minimum eccentricity of any node. Here, the radius is 2, indicating that the minimum distance to reach all other nodes from the most central node is two steps. The **periphery** of a network includes all nodes whose eccentricity is equal to the network's diameter, listed in this analysis as numerous nodes with an eccentricity of 4. These nodes represent the 'outermost' reaches of the network.

G. Center

The **center** of a network consists of all nodes whose eccentricity is equal to the radius. In this network, the node 'fer_nwn' is in the center, signifying its pivotal role in connectivity within the network.

IV. Centrality Measures

Centrality measures are crucial for identifying the most influential nodes within a network. They provide insights into the roles of individual nodes in terms of network connectivity and the flow of information. We focus on three primary measures: Degree Centrality, Eigenvector Centrality, and Betweenness Centrality.

A. Degree Centrality

Degree Centrality measures the number of direct connections a node has. It is an indicator of the node's activity and potential to influence its immediate neighbors. The formula for Degree Centrality is:

$$C_D(v) = \frac{\deg(v)}{n-1}$$

where $\deg(v)$ is the degree of vertex v, and n is the total number of nodes in the network. Nodes with high degree centrality can significantly influence the network's dynamics by affecting their immediate neighbors.

1) Interpretation in the Dataset: For instance, 'fumilayo' has a Degree Centrality of 0.5426695842450766, indicating a very high level of connectivity compared to other nodes, making it one of the most central and influential nodes in the network.

B. Eigenvector Centrality

Eigenvector Centrality measures a node's influence based on the centrality of its neighbors. It extends the concept of degree centrality by not only considering the number of connections but also the quality (centrality) of these connections. The mathematical expression is:

$$\mathbf{A}\mathbf{x} = \lambda \mathbf{x}$$

where $\mathbf{A}$ is the adjacency matrix of the network, λ is the largest eigenvalue, and $\mathbf{x}$ is the eigenvector of λ. High eigenvector centrality means a node is connected to many well-connected nodes.

1) Interpretation in the Dataset: 'gerardo_tri', with an Eigenvector Centrality of 0.7058618602401349, is significantly influential, indicating connections to other highly connected nodes, enhancing its overall impact on the network.

C. Betweenness Centrality

Betweenness Centrality quantifies how often a node acts as a bridge along the shortest path between two other nodes. This measure shows the node's capacity to facilitate communication between diverse parts of the network. The formula is:

$$C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

where σ_{st} is the total number of shortest paths from node s to node t and $\sigma_{st}(v)$ is the number of those paths that pass through v.

The high Betweenness Centrality of 'fer_nwn' at 0.49825553124239863 shows its crucial role in mediating the flow of information across the network, acting as a key connector among various nodes.

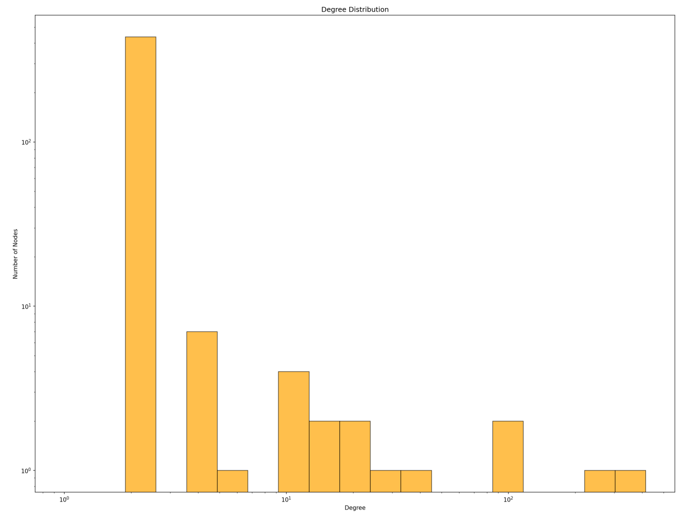

Fig. 2. Degree Distribution for the Ego Network

V. Degree Distribution

Degree distribution is a key property of networks that describes the probability distribution of the degrees across the entire network. This measure is essential for understanding the connectivity patterns among the nodes, where each point on the distribution curve represents the proportion of nodes in the network having a specific degree.

In the provided degree distribution graph for the ego network, most nodes exhibit relatively few connections, while a few nodes have a significantly higher number of connections. This characteristic distribution is typical for social networks, often indicative of a scale-free network where few nodes, known as hubs, are crucial for connecting various parts of the network.

1) Characteristics of an Ego Network: The graph distinctly shows:

- A prominent spike at one or a few nodes, suggesting the presence of an ego or central node with a very high degree. This node acts as a primary connector or influencer within the network.
- A majority of nodes have a low degree, indicating that most users in this network are connected to only a few other nodes.

This distribution is characteristic of an ego network, where the ego often has connections to all other alters, forming a star-like configuration. The high degree value (e.g., 416) represents the ego, which is significantly connected, while the multitude of low degrees (mostly 2s) represents alters that interact mainly through the ego.

2) Use of Logarithmic Scale in Plotting: The logarithmic scale is utilized on the axes to enhance the visualization of data that spans several orders of magnitude. This scale is particularly useful when the network exhibits a wide variance in node degrees. Using a logarithmic scale:

- Helps to compress the scale but maintains the order, facilitating a clearer observation of the distribution among nodes with both high and low degrees.
- Makes it easier to discern structural features of the network, such as the presence of hubs and the extent of connectivity variance, which are less apparent on a linear scale.

This plotting method underscores the hierarchical or non-uniform connectivity often observed in real-world networks, especially in those based on social interactions and communications.

VI. Community Detection

Community detection refers to the process of identifying groups of nodes in a network that are more densely connected internally than with the rest of the network. These groups, or communities, often represent underlying structures or shared properties among the nodes.

1) Calculation Methods: One common method for community detection is the **modularity optimization**, which quantifies the density of links inside communities as compared to links between communities. The modularity Q is defined as:

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j)$$

where A_{ij} represents the edge weight between nodes i and j, k_i and k_j are the sum of the weights of the edges attached to nodes i and j, m is the sum of all edge weights in the network, and δ is the Kronecker delta function that is 1 if i and j are in the same community and 0 otherwise.

2) Analysis Based on the Provided Network Diagram: The provided visualization illustrates the community structure within an ego network, where different colors represent different communities:

- The central node (ego) connects with multiple distinct communities, signifying its role as a hub across diverse groups.
- Each color-coded community cluster likely shares common attributes or interactions that are more frequent within the group than with nodes outside the group.
- The presence of multiple communities connected by a single node highlights the ego's pivotal role in bridging diverse social groups or interest areas.

This configuration underscores the complexity and multi-dimensional nature of the relationships in the network, providing valuable insights into how information or influence might flow within and across these groups.

By examining the community structure, we can infer potential areas of influence, sub-group dynamics, and the network's overall cohesion and segmentation, which are critical for understanding social structures and designing targeted interventions.

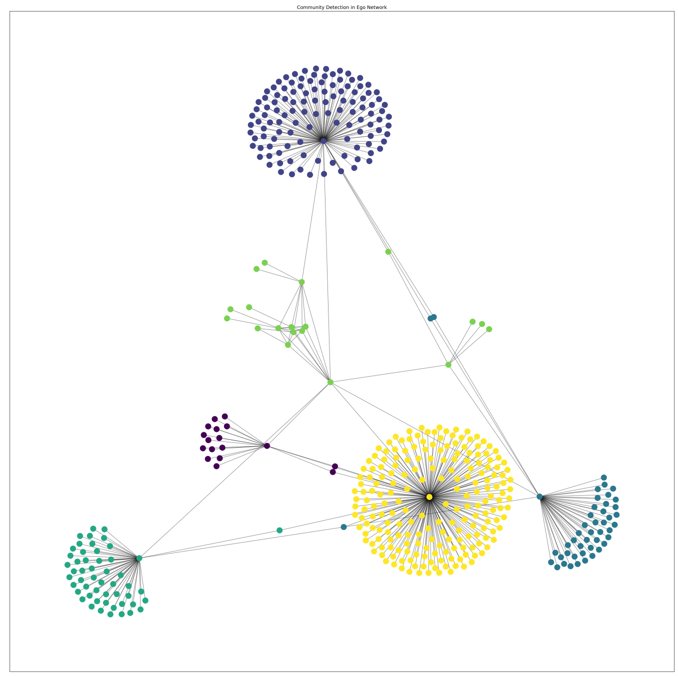

Fig. 3. Community Detection

3) Overview of Detected Communities: Community detection within the network has revealed several clusters that indicate the presence of subgroups within the larger network. These communities are identified by distinct labels, where each node is assigned to a community that represents a closely-knit group of nodes. This analysis is based on the connectivity patterns among the nodes, suggesting that nodes within the same community interact more frequently or share similar properties.

4) Interpretation of Community Assignments: The output of the community detection process assigns a community number to each node. Here are some observations:

- **Community 0** includes nodes that are central but relatively isolated in terms of direct connections to the majority of other clusters, except through specific bridge nodes.
- **Community 1** and **Community 2** represent dense clusters that may signify groups of users with high levels of interaction or shared interests.
- **Community 3** appears to contain peripheral nodes that, while less connected, play a critical role in connecting otherwise disparate sections of the network.
- **Community 4** and **Community 5** include nodes that have significant roles in network connectivity, often acting as bridges or hubs among various communities.

5) Significance of Communities: The identified communities are significant as they:

- Provide insights into the structural organization of the network.
- Highlight potential areas of influence where information or behaviors might spread more effectively within the

network.
- Help in understanding how individual nodes (or users) are grouped, which can be essential for targeted marketing, information dissemination, or network optimization strategies.

VII. Conclusion

This paper has delved into the intricate structure of an Ego network within the Letterboxd community, employing a variety of social network analysis techniques to uncover the depth and complexity of social interactions. The analysis has revealed a network characterized by a 'small-world' nature, significant clustering, and prominent community structures, which are influenced by both the central user and their diverse connections. This network exhibits typical properties of social networks, including scale-free characteristics and modular community structures, reflecting the nuanced ways in which film enthusiasts interact online.

The centrality measures applied in the study pinpoint key influencers and bridge-builders within the network, enhancing our understanding of information flow and social influence on Letterboxd. Moreover, the community detection results have uncovered several subgroups within the network, each with distinct interaction patterns, indicating different spheres of interest or interaction styles among users.

Overall, the findings underscore the importance of network analytical approaches in uncovering the latent structures within social media platforms, particularly those like Letterboxd, which cater to niche interests. These insights not only enrich our understanding of social dynamics in digital spaces but also provide valuable implications for enhancing user engagement and personalization strategies on social platforms. The methodology and findings from this study can serve as a benchmark for further research on similar networks, particularly in understanding the role of central users in influencing community formation and interaction dynamics in digital social environments.

References

[1] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*, 1st ed. Cambridge, UK: Cambridge University Press, 1994.
[2] M. E. J. Newman, "The structure and function of complex networks," *SIAM Review*, vol. 45, no. 2, pp. 167–256, Jun. 2003.
[3] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, pp. 440–442, Jun. 1998.
[4] A. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, Oct. 1999.
[5] J. Scott, *Social Network Analysis*, 3rd ed. London, UK: SAGE Publications Ltd, 2012.
[6] L. C. Freeman, "Centrality in social networks conceptual clarification," *Social Networks*, vol. 1, no. 3, pp. 215–239, 1978-1979.
[7] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proceedings of the National Academy of Sciences*, vol. 99, no. 12, pp. 7821–7826, Jun. 2002.
[8] J. Fer, "Social Network Analysis of Letterboxd," GitHub repository, 2024. [Online]. Available: https://github.com/juanjuanjuanfer/SNA-of-Letterboxd