Facebook Ego Network
A Graph Anaysis

Roidis Dimitrios

Contents

1 Introduction

Ego network analysis focuses on the immediate social structure surrounding a central individual, known as the ego. This structure consists of the ego, their direct connections (friends), and the ties between them. Unlike global network analysis, which looks at the entire graph, ego networks provide a more focused lens for understanding localized behavior, influence patterns, and personal connectivity.

The study of ego networks is particularly valuable in social network analysis, recommendation systems, and targeted marketing. For instance, analyzing a user's local structure can reveal influential peers, tightly-knit friend groups, or potential for community outreach. Additionally, comparing attribute similarities (such as interests or demographics) between the ego and its alters can offer insights into homophily — the principle that people tend to associate with others who are similar to themselves.

In this report, we analyze the Facebook Ego Network dataset, examining not only the raw graph structure itself but also the intuitions derived from metrics such as centrality, community development and similarity among users, always keeping in mind that we are talking about and undirected graph.

2 Dataset and Preprocessing

The analysis is based on the Facebook Ego Network dataset accessible via snap.stanford.edu, a collection of anonymized social network data capturing the ego-centric view of multiple users. The dataset comprises three core components:

- Ego Features: Binary feature vectors representing user attributes such as locale, education or work.
- Edge Lists: Unweighted and undirected edges capturing friendship links among users in each ego network.
- Circle Memberships (not utilized in this analysis): Annotated community affiliations manually curated by users.

For this analysis, one ego node was selected for in-depth investigation. The ego's graph was constructed using the edge list and then visualized utilizing netowrkx library(Figure 1), and the corresponding feature matrix was loaded to compute similarity-based metrics. The graph was treated as undirected, as is consistent with the nature of Facebook friendships.

The preprocessing steps included:

- Parsing the edge list and constructing the ego-centric subgraph using NetworkX.
- Mapping feature matrices to individual nodes for similarity analysis.
- Building the graph.

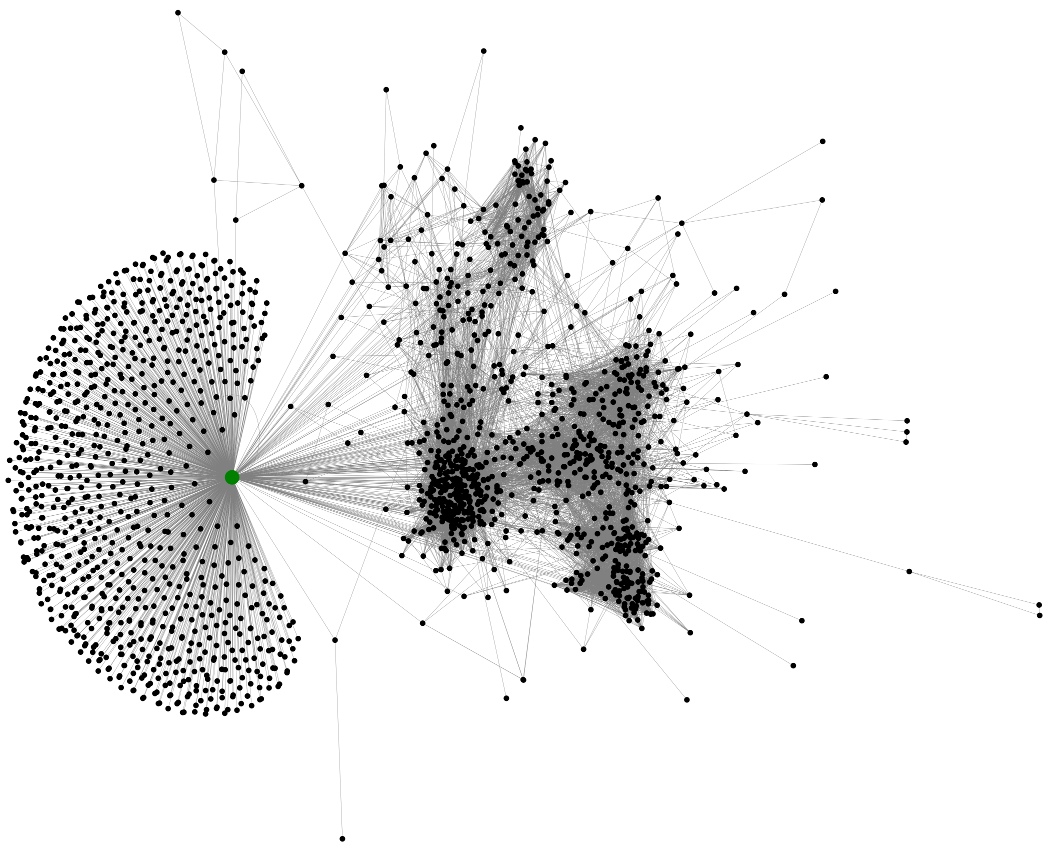

Figure 1: Ego Graph for user with ID 107

On a first look it becomes apparent that clusters seem to have been formed. Graph is divided into two primary sections: Friends of user 107 who are not friends themselves on the left side of figure 1, and those with edges connecting them meaning they are friends on the right side. On top of that a deeper clustering seems to have formed on the right side which indicates possible communities and individuals with common features(interests, background etc.). Graph consists of 1903 nodes, 27794 edges.

3 Intuitions from Centrality Measures and other metrics

To better understand the structural properties and roles of individuals within the ego network, we computed several centrality metrics utilizing the networkx library: degree centrality, betweenness centrality, closeness centrality, and eigenvector centrality, along with connected components and average clustering coefficient.

Degree centrality measures the number of immediate connections a node has. In the ego network, nodes with high degree centrality represent individuals who are directly connected to many others in the ego's social circle. These may correspond to highly active users or individuals with a broad social presence.

Betweenness centrality captures the extent to which a node lies on paths between other nodes. Nodes with high betweenness act as bridges or gatekeepers of information flow within the network. Their removal could significantly disrupt connectivity.

Closeness centrality assesses how close a node is to all others in terms of shortest paths. A high closeness score suggests that a node can efficiently interact with all others in the network—possibly representing well-positioned users in terms of influence or reach.

Eigenvector centrality considers not just the number of connections but also the importance of those connections. Nodes scoring high here are connected to other influential or well-connected nodes. This reflects prestige or embeddedness within the social structure.

Collectively, these metrics provide a multifaceted view of individual influence and network cohesion(table 1). For instance, in the analyzed ego network, we observed that nodes with high eigenvector centrality often overlap with those high in degree centrality, indicating that socially active users tend to be embedded within influential substructures. However, nodes with high betweenness centrality may not always be highly connected themselves, but still play critical bridging roles across different parts of the network.

Table 1 summarizes the top 5 nodes across four major centrality metrics: degree, betweenness, closeness, and eigenvector centrality. Node **107** stands out as it ranks first across the three most structurally informative metrics: degree, betweenness, and closeness centrality. This suggests that node 107 not only has the highest number of direct connections (degree) but also plays a pivotal role in connecting different parts of the network (betweenness), while being relatively close to all

Rank	Degree Centrality	Betweenness Centrality	Closeness Centrality	Eigenvector Centrality
1	**107 (0.5499)**	**107 (0.7662)**	**107 (0.6718)**	1888 (**0.1166**)
2	1888 (**0.1330**)	1086 (**0.0227**)	483 (**0.4778**)	1800 (**0.1123**)
3	1800 (**0.1283**)	483 (**0.0196**)	917 (**0.4753**)	1663 (**0.1095**)
4	1663 (**0.1230**)	917 (**0.0195**)	637 (**0.4718**)	1352 (**0.1085**)
5	1352 (**0.1225**)	925 (**0.0109**)	606 (**0.4692**)	1431 (**0.1048**)

Table 1: Top 5 Nodes by Centrality Measures in the Ego Network

other nodes in terms of shortest path distance (closeness). In nowadays social terms, this probably means that user with id 107 could be an influencer.

Interestingly, node 107 does not appear in the top 5 for eigenvector centrality, which emphasizes connections to highly connected nodes rather than just raw connectivity. Instead, node 1888 ranks highest by eigenvector centrality, indicating its close association with other influential nodes, even if it does not have the highest number of direct links or bridging power. This contrast highlights how different centrality metrics capture unique aspects of node influence and importance within the ego network.

Additionally, the ego network consists of a single connected component, indicating that all nodes are reachable from one another, either directly or through intermediate connections. This structural property implies a cohesive network where no subgroup is entirely isolated from the rest.

We found only one connected component meaning that there are no isolated nodes, or better, there is always a path from a node to reach any node in the graph. The Average Clustering Coefficient is 0.282 and indicates that 28.2% of a nodes' neighbours are connected to each other and possibly reflects the tendency of friends of a person to also be friends with each other in the ego network.

4 Community Detection

Understanding the modular structure of social networks is crucial for analyzing how users cluster based on shared characteristics, interactions, or features. In this project, we applied three widely used community detection algorithms to the ego network uitilizing community and networkx libraries: Louvain, Spectral Clustering, and Greedy Modularity. Each method identifies groupings of nodes (communities) that are more densely connected internally than externally.

The Louvain algorithm is a modularity-based greedy optimization method that seeks to maximize the density of links inside communities compared to links between communities. It efficiently discovers hierarchical communities and is known for producing high modularity scores.

Similar in spirit to Louvain, Greedy Modularity attempts to optimize modularity by repeatedly joining node pairs or communities that result in the highest increase in modularity. Though computationally more intensive for large graphs, it offers a valuable alternative perspective to uncover communities in a network

Spectral clustering on the other hand, is a graph partitioning technique that leverages the eigenvectors of the Laplacian matrix of the network. By embedding nodes into a low-dimensional space defined by the smallest eigenvalues and then applying k-means clustering, this method captures global structural features that local methods might miss. It is particularly suitable for finding balanced or non-convex communities.

Using the Greedy Modularity algorithm, we partitioned the ego network into 8 distinct communities. This method aims to optimize modularity by iteratively merging communities that maximize local modularity gains. The results indicate a strongly modular structure, with community sizes showing a highly skewed distribution: [1422, 392, 56, 20, 5, 3, 3, 2]. This skew suggests the presence of a giant core community encompassing a large portion of the network, alongside several much smaller peripheral groups. Such a structure is common in ego networks, where the ego and its close connections form a central component, while smaller clusters represent niche or more isolated groups.

To better understand the structure and semantics of the largest community (1422 nodes), we analyzed the most frequent anonymized user attributes. These attributes hint at possible social or demographic factors binding the community together:

Feature ID	**Description**	**Count**
0	Birthday; anonymized feature 376	981
221	Education; type; anonymized feature 54	775
223	Education; with; id; anonymized feature 539	695
349	Locale; anonymized feature 278	686
266	Hometown; id; anonymized feature 558	629
265	Gender; anonymized feature 78	341

Table 2: Greedy Modularity - Most Common Anonymized Features in the Largest Community

Indeed, friends on social networks, tend to have studied together, live in the same broader areas or be born in the same hometown. Birthday feature here, is not of much importance, as it is not realistic for these communities to consist of individuals with same birthday. Even though the features are anonymized, we supose this feature is probably a boolean value, indicating whether a user has added birthday in the profile or not.

After Greedy Modularity, Louvain, and Spectral Clustering algorithms were applied to the ego network, all three methods consistently revealed a dominant large community, followed by significantly smaller clusters. This consistency across algorithms strengthens confidence in the presence of strong community cohesion around certain user groups. Spectral Clustering, though based on an entirely different mathematical principle—using eigenvalues of the graph Laplacian—also yielded a comparable largest cluster, both in node count and in dominant user features. This convergence suggests that the ego network contains well-defined, robust communities, and that these are stable across modularity-based and spectral-based methods. Such agreement is particularly notable given the differences in algorithmic approaches and reinforces the credibility of the detected community structure.

For a more intuitive approach to communities, Figure 2 overlays the Louvain community structure on the ego network previously visualized in Figure 1. While the network topology remains unchanged, the added color segmentation highlights the modular decomposition of the graph. Each color represents a distinct community identified by the Louvain algorithm. This enriched visualization illustrates how the algorithm groups nodes into tightly-knit clusters with dense internal connections and sparser links between clusters. Notably, the ego node sits at the intersection of several communities, underscoring its role as a central connector across social circles. Compared to the uncolored version, this rendering offers clearer visual evidence of community boundaries and supports earlier observations derived from centrality and clustering metrics. The spatial compactness of color-coded clusters suggests the presence of meaningful social subdivisions within the ego's extended network.

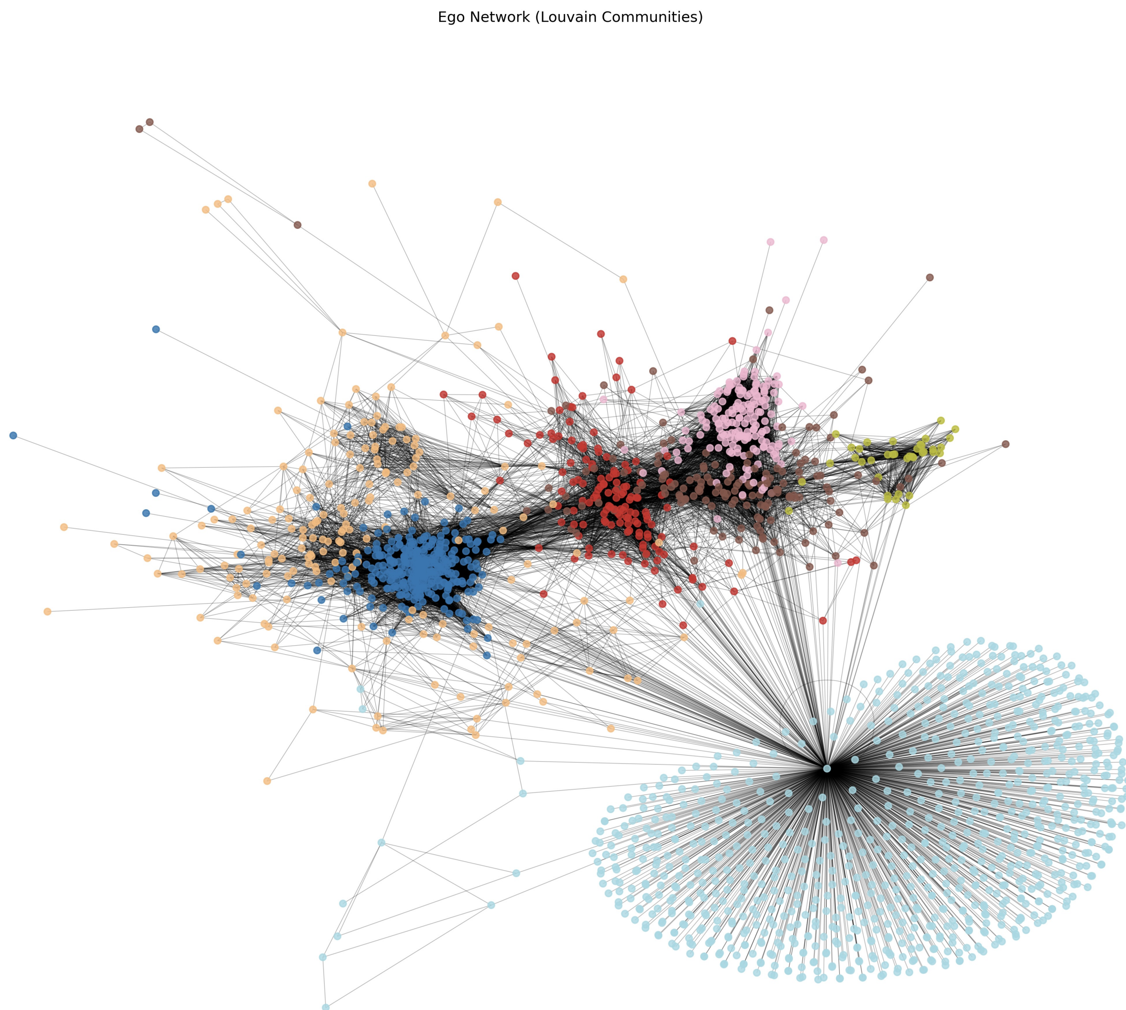

Figure 2: Ego Network Graph - Louvain Communities Colored

5 Conclusion

In this analysis, we explored the structure and characteristics of a Facebook ego network using a variety of graph-theoretic tools. We began by constructing the ego-centric subgraph, then investigated its connectivity, clustering tendencies, and prominent nodes through multiple centrality measures. Notably, one node (ID 107) consistently ranked highest across all centrality metrics, indicating its central and influential role within the network.

Community detection provided further insight into the underlying structure of the ego network. Algorithms such as Greedy Modularity, Louvain, and Spectral Clustering identified densely connected user groups, often aligned with shared anonymized attributes like education type or hometown. Interestingly, despite the methodological differences, all three community detection approaches yielded similar partitions, reinforcing the reliability of the identified community structure.

Visualizations enriched our understanding by highlighting the positioning and cohesion of these communities within the broader network. The application of color-coded clusters revealed clear modular organization and localized connectivity patterns.

Overall, this analysis demonstrates how combining network metrics, structural properties, and community detection methods can yield a nuanced view of social networks. These findings lay the groundwork for future tasks such as role classification, anomaly detection, or link prediction in online social graphs.