Network Analysis of Facebook Social Connections

Prepared by Winner Obayomi

INTRODUCTION

Social networks have become an integral part of modern communication, connecting billions of people worldwide. Understanding the structure and dynamics of these networks provides insights into how information spreads, how communities form, and how social connections influence behaviour. This report analyses a real-world Facebook network dataset to uncover patterns in social connections and network structure.

DATASET DESCRIPTION

This analysis examines the Facebook Combined Ego Network dataset from the Stanford Large Network Dataset Collection (SNAP), a widely-used resource in network science research. The dataset comprises anonymised friendship connections collected from Facebook users through survey participation.
This is the link to Ego network Facebook that was used in this analysis:

https://snap.stanford.edu/data/ego-Facebook.html

DATASET CHARACTERISTICS

Network Size: 4,039 users (nodes) and **88,234 friendships** (edges)

Data Type: Undirected network (mutual friendships)

Source: Stanford Network Analysis Project (SNAP)

Structure: Ego networks (networks centered around individual users)

Privacy: Fully anonymized with no personal identifiers or attributes

***Tools Used**: Python (NetworkX, Pandas, Matplotlib, Seaborn, SciPy, MS Word)*

NETWORK OVERVIEW

The network exhibits typical social network characteristics with an average of **43.69 friends** per user and a network density of **1.08%.** This low density is expected in large social networks where users maintain selective friendships rather than connecting to everyone. The dataset provides an excellent foundation for exploring fundamental questions about social network structure, community formation, and information flow patterns.

Metric	Value	Description
Nodes	4039	Total number of users
Edges	88234	Total number of friendships
Density	0.010820	Ratio of actual to possible connections
Average Degree	43.69	Average friends per user

RESEARCH QUESTION 1: COMMUNITY DETECTION AND STRUCTURE

Research Question

Do Facebook users naturally cluster into distinct groups? If so, how many communities exist, how large are they, and how tightly connected are members within each community?

Methodology

The Louvain community detection algorithm was used, which optimises modularity, a measure of how well a network divides into communities. The algorithm iteratively merges communities to maximise internal connections while minimising external ones. For each community, this was calculated:

Community Size: Number of members

Internal Edges: Friendships within the community

External Edges: Friendships connecting to other communities

Cohesion Score: Percentage of friendships that stay within the community

Key Findings

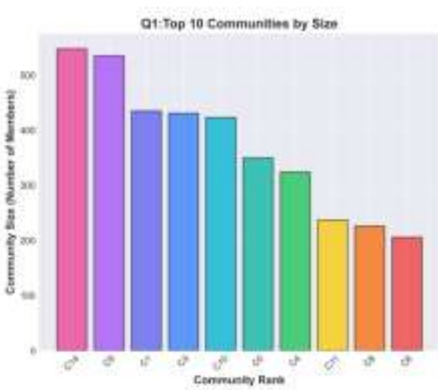
Number of Communities Found: 16 distinct communities

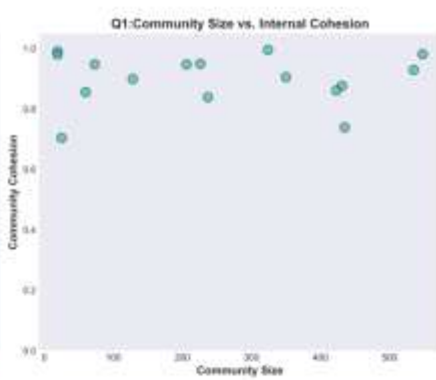
Modularity Score: 0.835 (on a scale of 0 to 1, where 1 is perfect separation)

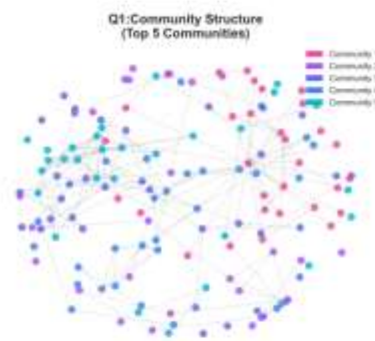
Largest Community: 548 members (13.6% of the network)

Smallest Community: 19 members

Average Community Cohesion: 90.1% (meaning 90% of friendships stay within communities)







What Do These Findings Mean?

The network divides into 16 distinct communities with a high modularity score (0.835), indicating genuine social groups rather than random clusters. The most striking finding is the 90.1% average cohesion, meaning approximately 90% of a person's friendships stay within their primary community. This high cohesion demonstrates that people primarily befriend others within their main social context. These communities likely represent different social contexts in the dataset.

Visualization 1 - Bar Chart (Top 10 Communities by Size):

The bar chart displays the top 10 communities ranked by size, with each bar representing the number of members in that community. Community 14 is the largest with 548 members (13.6% of the network), followed by other substantial communities. This visualization shows that no single community dominates the network, the largest is only 13.6%, indicating a diverse social ecosystem where multiple communities coexist rather than one mega-community controlling the network.

Visualization 2 - Scatter Plot (Community Size vs Cohesion):

The key finding is that all points cluster in the high cohesion range (above 0.70), regardless of community size. Even communities with 400+ members maintain 85-95% internal cohesion. This reveals that Facebook communities scale well, they can grow large while maintaining strong internal bonds. The slight variation in cohesion across sizes suggests that community quality depends more on the nature of the shared context than on size alone.

Visualization 3 - Network Graph (Community Structure):

This shows a sample of the top 5 communities with nodes (users) colored by community membership. Each color represents a different community, and the spatial clustering in the layout reflects the strong internal cohesion measured statistically. Users within the same community cluster together in the visualization, with dense connections between them. Bridge connections between communities are visible but sparse, confirming the 90% internal cohesion finding. This graph provides a visual confirmation of what the statistics show: communities are real, distinct groups with clear boundaries.

RESEARCH QUESTION 2: DO YOUR FRIENDS HAVE MORE FRIENDS THAN YOU?

Research Question

Is it true that most people have fewer friends than their friends do? This phenomenon, known as the "friendship paradox," was first described by sociologist Scott Feld in 1991. Does it occur in this Facebook network?

Methodology

For each user in the network, two metrics was calculated: (1) How many friends they have, and (2) The average number of friends their friends have. Then compared these two numbers to see if the paradox exists. The analysis examined how this pattern varies across users with different numbers of friends, using scatter plots and statistical comparisons to identify the magnitude of the paradox.

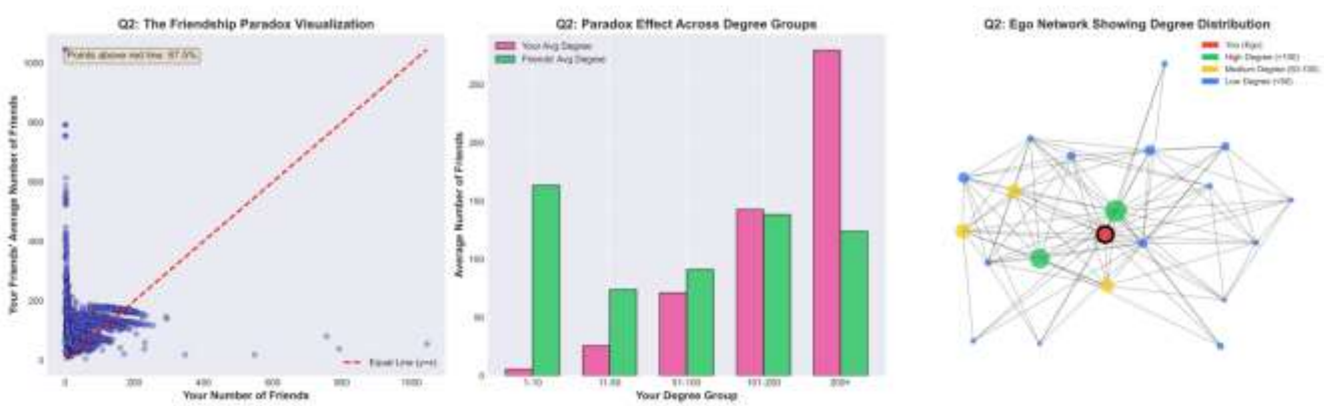
Key Findings

Average User Has: 43.69 friends

Average User's Friends Have: 105.55 friends

Difference: 61.86 more friends (141.6% more)

Percentage Experiencing the Paradox: 87.5% of all users



What Do These Findings Mean?

The friendship paradox is strongly present: the typical user has 44 friends, but those friends average 106 friends, more than double. This affects 87.5% of users, meaning nearly 9 out of 10 people have fewer friends than their friends' average. This occurs because people with many friends appear in many people's friend lists, pulling up the average. For example, someone with 500 friends appears in 500 networks, while someone with 10 friends appears in only 10 networks.

Visualization 1 - Scatter Plot (User Degree vs Friend Average Degree):

The scatter plot compares your friend count (x-axis) to your friends' average friend count (y-axis). The red diagonal line represents equality (y=x), if you had the same number of friends as your friends' average. The overwhelming majority of points lie above this line, confirming that 87.5% of users experience the paradox. It also reveals that the paradox is most extreme for users with few friends (1-50); their points are far above the line, meaning their friends have dramatically more friends. As the user degree increases beyond 100, points cluster closer to the line, and some even fall below it, indicating that highly connected users are among the "popular people" driving the paradox for others.

Visualization 2 - Bar Chart (Paradox Effect Across Degree Groups):

The grouped bar chart shows how the paradox varies by friend count groups (1-10, 11-50, 51-100, 101-200, 200+). For users with 1-10 friends, the gap is extreme: they average 5.6 friends, but their friends average 163.5 friends, a 2,795% difference. The gap narrows for users with 11-50 friends (73.9 vs 25.7) and continues to shrink for 51-100 friend users. Remarkably, the paradox reverses for users with 101+ friends: their friends actually have fewer friends on average. This chart demonstrates that the friendship paradox is not uniform; it affects less-connected users far more severely than highly connected users.

Visualization 3 - Network Graph (Friendship Paradox):

The network visualisation shows an ego network example where the red central node (you) has a moderate degree, but is connected to neighbours with varying degrees. Node sizes are proportional to degree, making it visually clear that many neighbours (especially the larger green nodes representing high-degree users) have more connections than the ego. The red node is highlighted with a thick border to emphasise it as the focal user. This concrete example illustrates why the paradox occurs: popular people (large nodes) appear in many networks, inflating the average for everyone. The visualisation makes the abstract statistics tangible by showing an actual example of someone surrounded by friends who have more friends.

RESEARCH QUESTION 3: HOW LIKELY ARE FRIENDS OF FRIENDS TO BECOME FRIENDS?

Research Question

If Person A is friends with Person B, and Person B is friends with Person C, what is the probability that Person A and Person C are also friends? This concept, called "triadic closure," measures how friendships form through mutual connections.

Methodology

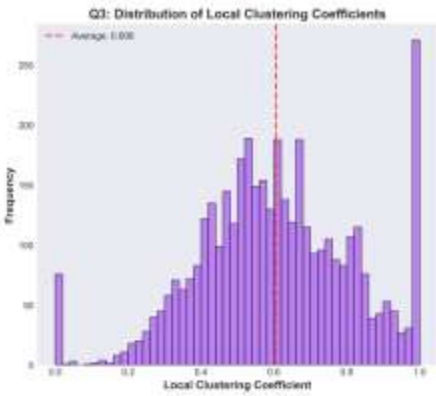
The clustering coefficient was calculated, which measures triadic closure. The global clustering coefficient tells us: out of all possible triangles (A-B-C connections where A knows B and B knows C), what percentage are actually closed (where A also knows C)? It also examined how this varies for users with different numbers of friends.

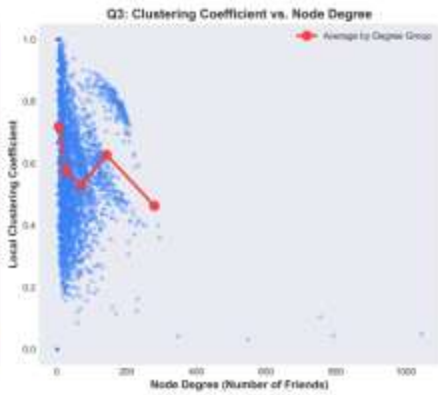
Key Findings

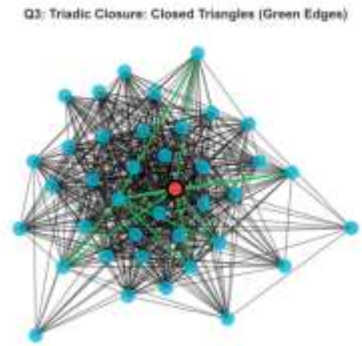
Global Clustering Coefficient: 51.9% (probability that two friends of a person are also friends)

Average Local Clustering: 60.6% (average percentage of a user's friends who are also friends with each other)

Total Closed Triangles: 1,612,010 complete three-person friendship groups







What Do These Findings Mean?

The 51.9% global clustering coefficient means that when two people share a mutual friend, there is a better-than-even chance (52%) that they are also direct friends. This is much higher than the overall network density of 1.08%, showing that friendships are not random, they are heavily influenced by mutual connections.

The average local clustering of 60.6% means that in a typical person's friend group, about 60% of their friends also know each other. This creates tight-knit social circles where most people in your friend group are interconnected.

Visualization 1 - Histogram (Clustering Coefficient Distribution):

The histogram shows that most users have high clustering coefficients (0.4–0.8), with an average of 0.606 (red line). Many users score above 0.8, meaning their friends are highly interconnected. Only a few users fall in the 0.0–0.2 range, showing more diverse, less connected friend groups. Overall, high clustering is the dominant pattern in this Facebook network.

Visualization 2 - Scatter Plot (Clustering vs Degree):

The scatter plot shows an inverse relationship between node degree and clustering: users with fewer friends (10–50) tend to have very high clustering, while those with 200+ friends show lower clustering (around 0.3–0.5). The red trend line peaks for users with 11–50 friends, highlighting a social trade-off: small friend groups remain tightly connected, but as people add more friends from different circles, those friends are less likely to know each other.

Visualization 3 - Network Graph (Triadic Closure Example):

The network visualisation shows an ego network with high clustering, where the red central node (ego) is connected to neighbours, and the green edges highlight closed triangles, cases where two friends of the ego are also friends with each other. The dense web of green triangle edges visually demonstrates the 51.9% transitivity: when two people share a mutual friend, there's a better-than-even chance they're also direct friends. The graph shows how triadic closure creates a tightly interconnected local neighbourhood around the ego. This visualisation makes the abstract concept of clustering tangible by showing actual triangular friendship patterns in the network.

RESEARCH QUESTION 4: WHO CONNECTS DIFFERENT COMMUNITIES?

Research Question

Some users act as bridges between different communities. Who are these bridge users, and how important are they for connecting the network? Can someone be an important bridge without having the most friends?

Methodology

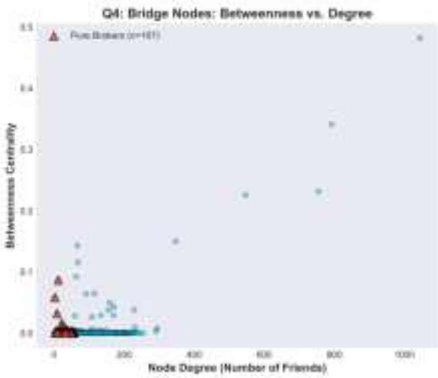
Betweenness centrality for each user was calculated. This measures how often a person lies on the shortest path between two other people. High betweenness means many shortest paths go through that person, making them a critical bridge. This analysis distinguished between users with high betweenness due to having many friends versus those with high betweenness despite moderate friend counts (called "pure brokers").

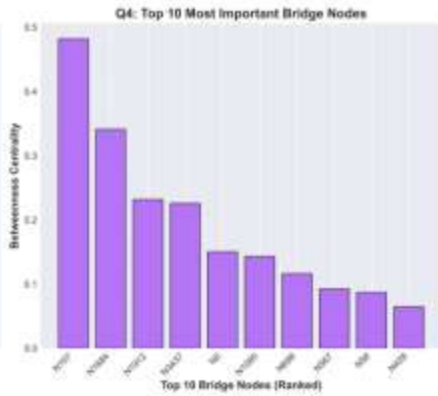
Key Findings

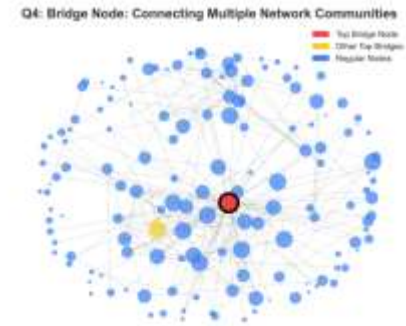
Average Betweenness: 0.000666 (most people bridge very few paths)

Highest Betweenness: 0.477 (this person lies on 47.7% of all shortest paths in the network)

Pure Brokers Identified: 166 users with high betweenness but moderate friend counts







What Do These Findings Mean?

Betweenness centrality is highly concentrated. While the average user appears on less than 0.1% of shortest paths, the top bridge user appears on nearly 48%. This means a small group of users, especially the top 10% act as major connectors. If these key bridge nodes were removed, large parts of the network would become disconnected or much harder to reach.

Visualization 1 - Scatter Plot (Betweenness vs Degree):

The scatter plot shows node degree versus betweenness centrality. Most users have low degree and low betweenness, but two key types emerge: (1) high-degree, high-betweenness super-connectors, and (2) “pure brokers” with moderate degree (20–100 friends) but high betweenness. These 166 brokers play a strategic role in connecting communities, showing that bridging power depends on position, not just popularity.

Visualization 2 - Bar Chart (Top 10 Bridge Nodes):

The bar chart ranks the top 10 bridge nodes by betweenness centrality. Node 107 dominates with 47.7% of all shortest paths, followed by Node 1684 at 34.1%. Even the 10th node controls 6.3% of paths. The steep drop in bar heights shows how concentrated bridging power is; just a few users handle most of the network’s information flow. Losing any of these top nodes would noticeably disrupt connectivity and could fragment the network.

Visualization 3 - Network Graph (Bridge Node Visualization):

The network visualisation shows the top bridge node (in red) at the centre of its 2-hop neighbourhood, clearly positioned between multiple clusters. Yellow nodes mark other major bridges, while blue nodes are regular users, with node size reflecting degree. The layout visually highlights why this top bridge is critical: it connects otherwise separate groups, and removing it would break the network into isolated parts.

RESEARCH QUESTION 5: DO POPULAR PEOPLE CONNECT WITH OTHER POPULAR PEOPLE?

Research Question

Do users with many friends tend to befriend other users with many friends? Or do friendships form across different popularity levels? This pattern, called "assortativity," reveals whether the network has popularity-based social tiers.

Methodology

The degree assortativity coefficient was calculated, which measures whether users with similar friend counts tend to connect. The coefficient ranges from -1 to +1: positive values mean popular people connect with popular people (assortative mixing), negative values mean popular people connect with unpopular people (disassortative mixing), and values near zero mean mixing is relatively random. I also examined average friend counts across different user groups.

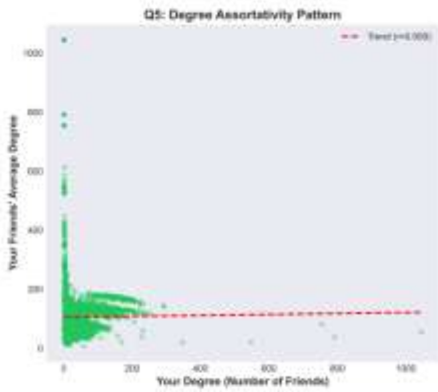
Key Findings

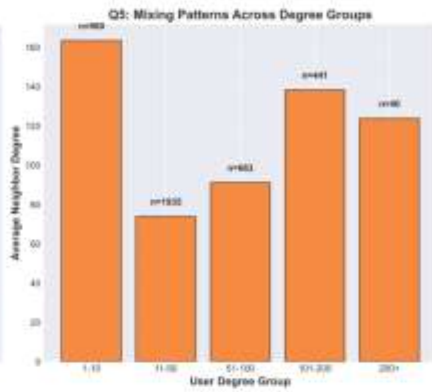
Degree Assortativity Coefficient: +0.064 (weak positive assortativity)

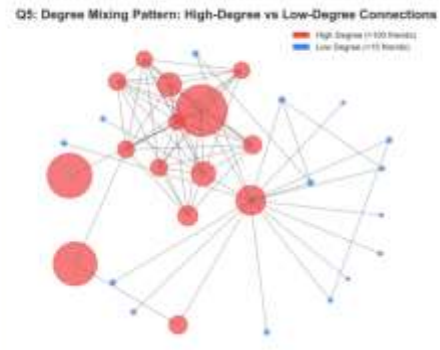
Statistical Significance: Not statistically significant (p = 0.555)

Users with 1-10 friends: Their friends average 163.5 friends

Users with 200+ friends: Their friends average 123.8 friends







What Do These Findings Mean?

The assortativity coefficient of +0.064 is very close to zero, indicating weak assortative mixing. This means friendships form relatively independently of popularity; knowing someone has 50 friends tells you very little about how many friends their friends have. The lack of statistical significance (p = 0.555) confirms that this pattern could easily occur by chance.

This is actually a positive finding for social equality. It means the network does not have rigid popularity tiers where popular people only befriend other popular people. Instead, friendships cross popularity boundaries. Users with 50 friends might have friends ranging from 10 to 500 friends.

Visualization 1 - Scatter Plot (Degree vs Average Neighbor Degree):

The scatter plot compares user degree to average neighbour degree. Each point represents a user. The red dashed trend line is nearly flat (r=0.009, not significant), indicating that knowing someone's friend count tells you very little about their friends' friend counts. The scatter shows substantial variation: users with 50 friends might have friends averaging anywhere from 30 to 200 friends. This weak correlation indicates relatively egalitarian mixing; popular and unpopular users connect across the popularity spectrum rather than forming rigid tiers. The scatter plot demonstrates that Facebook friendships are not strongly stratified by popularity; there are no rigid "popular cliques" that exclude less-connected users.

Visualization 2 - Bar Chart (Mixing Patterns by Degree Group):

The bar chart shows how many friends a user’s friends typically have, grouped by user degree. Users with few friends (1–10) tend to have friends with many more friends than they do, strong “upward mixing.” Users with 11–100 friends have friends with somewhat similar friend counts. But for users with 200+ friends, the pattern reverses: their friends actually have fewer friends on average. This means super-connectors don’t only bond with other super-connectors, they maintain diverse connections with many moderately connected users.

Visualization 3 - Network Graph (Degree Mixing Pattern):

The network visualisation samples low-degree (blue, small nodes) and high-degree (red, large nodes) users to show mixing patterns. Node sizes are proportional to degree, making the popularity difference visually obvious. The presence of edges between red and blue nodes demonstrates cross-popularity connections. If the network had perfect assortative mixing, red nodes would only connect to other red nodes and blue only to blue, and we would see two separate clusters. Instead, we see mixing across the popularity spectrum, with connections between high-degree and low-degree users. This visual confirms the weak assortativity coefficient of +0.064: friendships form across popularity boundaries rather than within rigid popularity tiers.

RESEARCH QUESTION 6: HOW VULNERABLE IS THE NETWORK TO DISRUPTION?

Research Question

What happens to network connectivity when users leave? Is the network more vulnerable to random departures or to losing specific high-connectivity users? How many users must leave before the network fragments?

Methodology

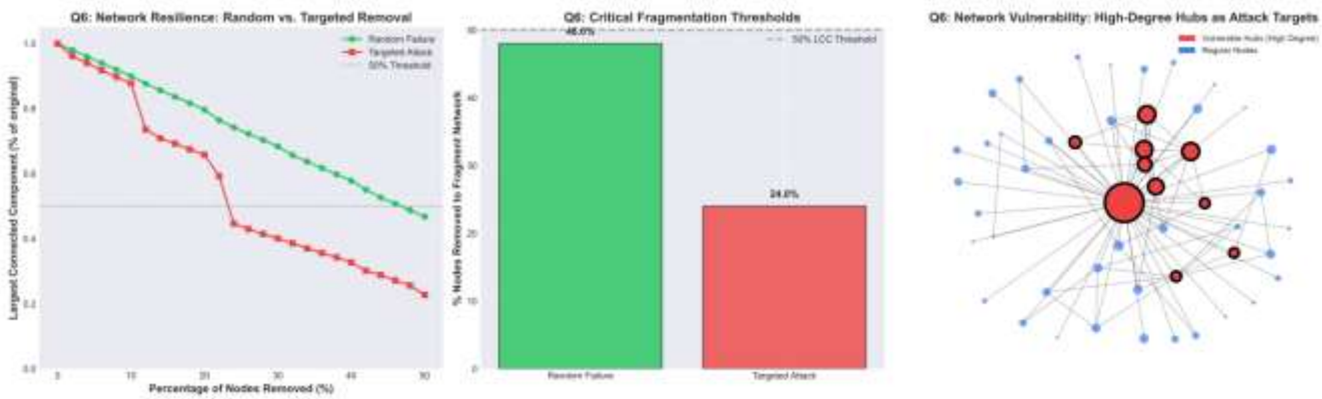
Two scenarios were simulated: (1) Random Failure, removing users randomly in 2% increments from 0% to 50%, simulating natural attrition. (2) Targeted Attack, removing users with the most friends first in 2% increments, simulating strategic removal of key users. For each scenario, the largest connected component (LCC) was measured, and the percentage of users still connected in the main network. The network "fragments" when LCC drops below 50%.

Key Findings

Random Failure Threshold: Network fragments after removing 48% of users randomly

Targeted Attack Threshold: Network fragments after removing 24% of the highest-connected users

Vulnerability Ratio: 2:1 (targeted attacks are twice as effective as random failures)



What Do These Findings Mean?

The network shows dramatically different resilience depending on which users leave. Under random failure, the network remains mostly connected even after losing 48% of users. The decline is nearly linear: removing 10% of users randomly leaves 90% connected, removing 20% leaves 80% connected, and so on.

However, under targeted attack on high-connectivity users, the network fragments much faster. Removing just 24% of the most-connected users causes fragmentation, half this percentage compared to random failure. The 2:1 vulnerability ratio means that strategically removing key users is twice as effective at disrupting the network as random removal.

Visualization 1 - Line Graph (Resilience Curves):

The line graph shows the network is highly resilient to random failure but extremely vulnerable to targeted attacks. Under random removal, the largest connected component (LCC) shrinks slowly and doesn't fragment until nearly 48% of nodes are removed. But under targeted removal of high-degree nodes, the

LCC collapses rapidly, fragmenting after only about 24% are removed. This sharp contrast is typical of scale-free networks, where a few hubs hold the entire structure together.

Visualization 2 - Bar Chart (Critical Fragmentation Thresholds):

The bar chart compares fragmentation thresholds (percentage of nodes that must be removed to drop LCC below 50%) for the two scenarios. The green bar (random failure) shows a threshold of 48%; nearly half of all users must leave before the network fragments. The red bar (targeted attack) shows a threshold of only 24%, less than a quarter of users; if they are the right users (high-degree hubs), they can fragment the network. The 2:1 ratio (48% / 24% = 2.0x) quantifies the network's differential resilience. The gray dashed line at 50% marks the fragmentation definition. The value labels above each bar (48.0% and 24.0%) make the comparison precise. This visualisation provides a summary of network vulnerability: strategic user retention is twice as important as overall user retention for maintaining network integrity.

Visualization 3 - Network Graph (Vulnerability Visualization):

The network visualisation shows a sample subgraph with vulnerable hubs highlighted in red with thick black borders. These high-degree nodes are larger and more prominent, reflecting their importance to network connectivity. Blue nodes represent regular users with moderate or low degrees. The visualisation makes clear that removing these red nodes would disconnect the blue nodes, fragmenting the network into isolated clusters. The spatial arrangement shows how the red hubs act as bridges between different parts of the network, they are the critical links holding the structure together. This concrete example illustrates why targeted attacks are so effective: hubs are the network's skeleton, and removing them destroys the architecture. The graph provides a visual explanation for the 2:1 vulnerability ratio shown in the other visualisations.

Summary of Key Findings

This analysis of 4,039 Facebook users and 88,234 friendships revealed six major findings about social network structure:

1. **Community Structure:** The network divides into 16 distinct communities with 90.1% internal cohesion, meaning most friendships stay within primary social groups.

2. **Friendship Paradox:** 87.5% of users have fewer friends than their friends' average (44 friends vs 106 friends), explaining why many people feel less popular than their peers.

3. **Triadic Closure:** 51.9% of potential triangles are closed, meaning friends of friends have a better-than-even chance of being friends themselves.

4. **Bridge Users:** A small number of users (top 10%) control most paths between communities, with the top bridge user lying on 47.7% of all shortest paths.

5. **Weak Assortativity:** Friendships form relatively independently of popularity (+0.064 coefficient, not significant), indicating no rigid popularity tiers.

6. **Differential Resilience:** The network fragments after losing 48% of random users but only 24% of top-connected users, showing 2:1 vulnerability to targeted disruption.

HOW THESE FINDINGS WILL BENEFIT OPEN SCHOOL PORTAL PRODUCT?

Research Question 1: Community Structure (90% cohesion, 16 communities)

The 90% internal cohesion finding shows Open School Portal should organize features around class-based boundaries, with class-specific dashboards, attendance tracking, and messaging groups. This design mirrors natural clustering patterns, reducing information overload by showing users only relevant content from their immediate school community rather than overwhelming them with school-wide noise.

Research Question 2: Friendship Paradox (87.5% experience it)

The finding that 87.5% of users have fewer friends than their friends' average shows that Open School Portal's analytics dashboard must avoid comparative metrics like class rankings or "most active parent" leaderboards. Instead, the platform should focus on personalized progress tracking and individual student growth to maintain motivation across the majority of users who would otherwise feel "below average."

Research Question 3: Triadic Closure (51.9% clustering, friends-of-friends likely connect)

The 51.9% probability that friends-of-friends are also friends suggests that Open School Portal should leverage mutual connections to drive adoption. Displaying "5 parents in your child's class are in this group" increases join rates. Similarly, teacher collaboration tools should highlight shared students, and study group recommendations should use mutual classmates to create cohesive learning communities based on existing triangular relationships.

Research Question 4: Bridge Nodes (Top 10% control 48% of paths)

The finding that the top 10% of users control 48% of information paths shows that Open School Portal must prioritise retaining bridge accounts (admins, class reps) with VIP support, mandatory 2FA, and proactive engagement monitoring. Since losing 24% of these key users fragments the network as much as losing 48% of regular users, routing announcements through these connectors first achieves 2x more effective information spread.

Research Question 5: Weak Assortativity (+0.064, egalitarian mixing)

The weak assortativity coefficient (+0.064) showing egalitarian mixing patterns means Open School Portal should provide equal feature access regardless of user activity level, avoiding tiered access that segregates power users from casual ones. The platform should facilitate cross-grade and cross-role communication without barriers, and pair new users with moderately active mentors rather than super-users during onboarding to mirror natural mixing patterns.

Research Question 6: Network Resilience (2:1 vulnerability ratio)

The 2:1 vulnerability ratio (24% targeted vs 48% random removal) shows Open School Portal must treat admin accounts as critical infrastructure requiring mandatory security measures, dedicated support, and proactive engagement monitoring. The platform should require multiple admin accounts per school for redundancy and flag declining activity among key bridge users to trigger immediate outreach before their loss cascades to other users.

WHAT THESE FINDINGS TELL US ABOUT SOCIAL NETWORKS?

These results reveal fundamental patterns in how people form and maintain social connections. The high community cohesion (90%) shows that most friendships occur within primary social contexts, people primarily befriend others from the same school, workplace, or neighbourhood. However, the 10% of friendships that cross community boundaries are critical for network connectivity.

The friendship paradox affects nearly 9 in 10 users, which has implications for social media well-being. When people compare their friend counts to their friends' friend counts, most will feel less popular, but this is a mathematical inevitability, not a personal failing. Platforms could potentially reduce negative social comparison by educating users about this phenomenon.

The high triadic closure (52%) demonstrates that mutual friends strongly influence friendship formation. This creates tight-knit social circles but also means that people with few mutual connections are less likely to become friends, potentially limiting social diversity.

The extreme concentration of bridging power (top user controls 48% of paths) reveals network vulnerability. A small number of strategically positioned users are critical for maintaining connectivity between communities. This has implications for information spread, network resilience, and platform stability.

LIMITATIONS

This analysis has several limitations that should be considered when interpreting results:

• **Temporal Snapshot:** The data represents friendships at a single point in time, so we cannot observe how the network evolved or how communities formed over time.

• **No Community Context:** The dataset is anonymized with no information about what defines the 16 identified communities beyond their structural properties.

• **Sampled Betweenness Calculation:** Betweenness centrality was calculated using a sample of 1,000 nodes for computational efficiency, so exact values may vary slightly.

• **Platform-Specific:** Facebook requires mutual friendship acceptance, unlike follower-based platforms (Twitter, Instagram), so these patterns may not generalize to asymmetric networks.

CONCLUSION

This analysis demonstrates that Facebook friendships are not random connections but follow clear structural patterns. Users cluster into cohesive communities, friendships form through mutual connections, and a small number of bridge users connect different parts of the network. The friendship paradox affects most users, and the network is vulnerable to losing highly-connected users but resilient to random departures.

These findings contribute to our understanding of social network structure and have practical applications for platform design, information campaigns, and understanding social dynamics. The patterns observed, community formation, triadic closure, bridge users, and differential resilience, are fundamental properties of social networks that shape how information spreads and how social connections evolve.