Sunil Thapa
W0779269

Analysis of Social Structure and Information Dynamics in the Auburn University Facebook Network

Introduction and Data Description

The social network – socfb-Auburn dataset represents the social friendship network of Auburn University as captured in September 2005, collected by Traud, Mucha and Porter. This dataset is a part of the extensive Facebook 100 collection. The network is modelled as an undirected, unweighted graph where nodes represent individuals and edges represent a mutual Facebook connection or friendship.

Social networks of this magnitude often exhibit the hairball effect (Fig. 1), where extreme density obscures structural patterns and renders complex algorithms computationally prohibitive. The full Auburn network contains 973,918 edges, 18,448 nodes, with an average degree of 105.59, making visualization and simulation difficult.

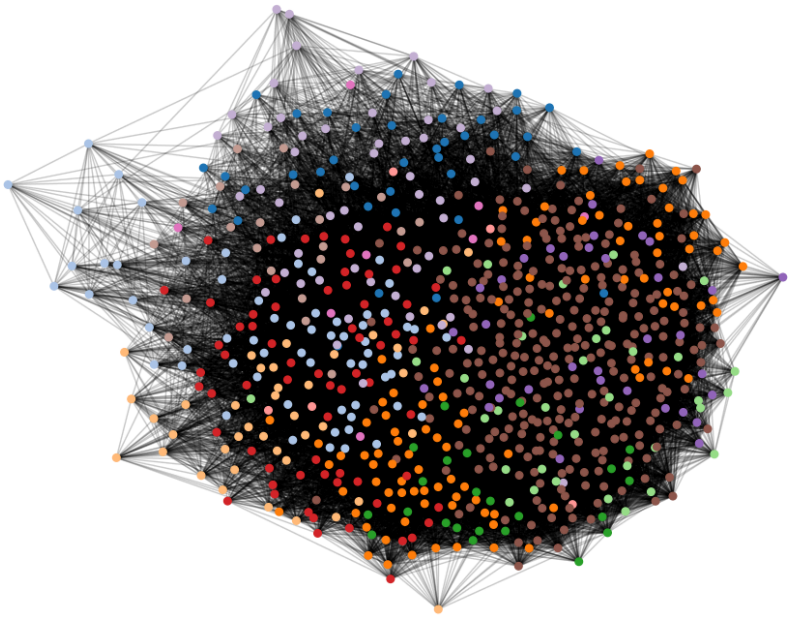

Fig. 1: Original Dataset – Hairball Effect

To address this, this analysis utilizes a k-core decomposition (k=95) to filter the network. By retaining only nodes with at least 95 connections, we isolate the "Elite Core" or social backbone of the university. This approach effectively removes peripheral noise, allowing for a focused analysis of the community structures and dynamics among the network's most active participants.

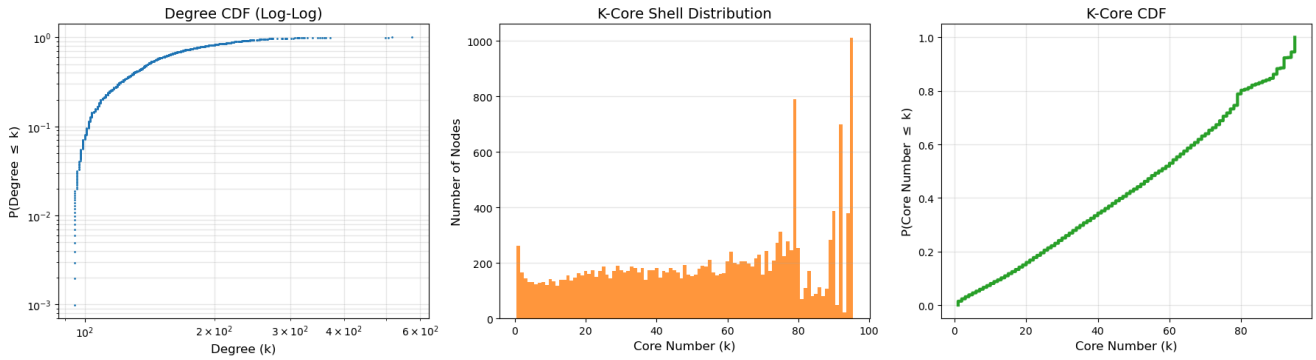

Fig. 2: K-Core Distribution

The contrast between the full network and the elite core highlights the structural stratification of the university:

- Full Network: N = 18,448 nodes, M = 973,918 edges. The global network is sparse (Density ~0.0057).
- Elite Core (k=95): N = 1,011 nodes, M = 78,470 edges. The core is highly dense (~0.15), indicating that the social backbone is tightly interconnected.
- Average Degree: The average individual in the full graph has 106 friends; in the elite core, this jumps to over 155, reflecting the hyper-sociability of this subgroup.

Fig. 3: Visualization of Auburn K-95 Elite Core, colored by Modularity Class

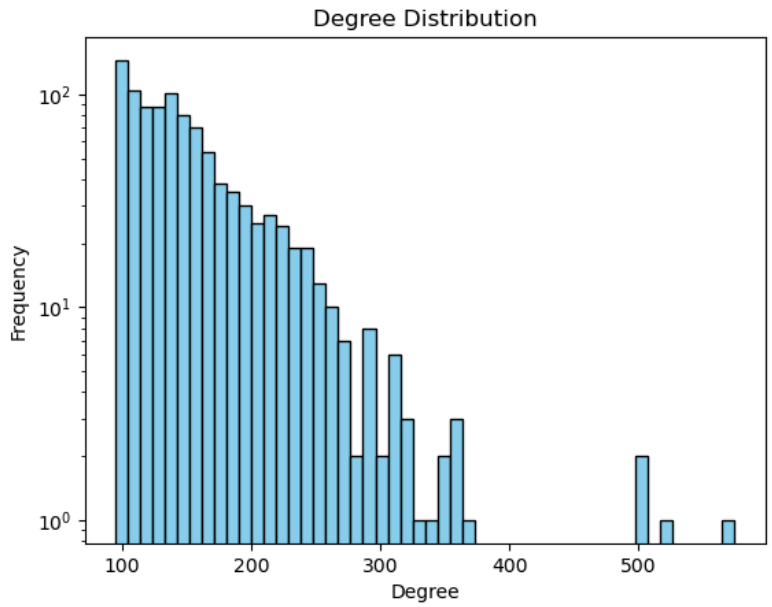

Fig. 4: Degree Distribution

Measure of Structural Importance (Centrality)

To identify the key factors within the elite core, we calculated and compared Degree Centrality (popularity), Eigenvector Centrality (influence), and Betweenness Centrality.

The table below summarizes the highest-ranking nodes across the three metrics:

Rank	Node ID	Degree	Eigenvector	Betweenness
1	**5471**	**0.5683**	**0.0985**	**0.0214**
2	10076	0.5119	0.0981	0.0138
3	3211	0.5010	0.0959	0.0113
4	14496	0.4931	0.0886	0.0107
5	16368	0.3673	0.0737	0.0051

The centrality analysis reveals a highly stratified "monarchy" structure. Node 5471 ranks first across all three metrics, acting as the "Super-Hub" of the network. The fact that the top nodes are nearly identical across metrics indicates that in this elite group, popularity equals influence.

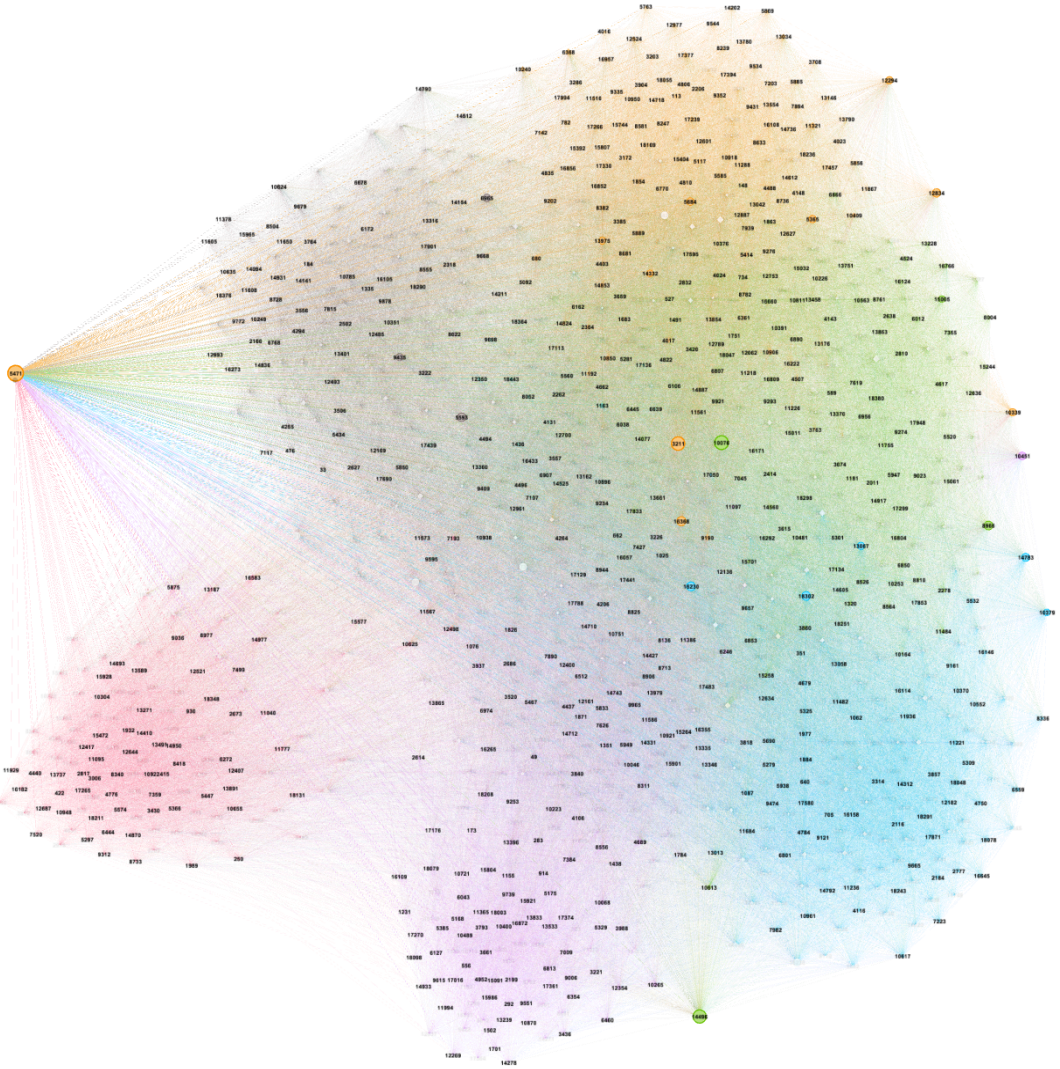

Fig. 5: Node 5471 Connectivity

Large-Scale Structure (Communities)

We applied the Louvain Algorithm to detect communities and calculated Assortativity to understand mixing patterns.

- Modularity Score: 0.3846
- Number of Communities: 7
- Assortativity Coefficient: 0.0127

The Louvain algorithm detected distinct communities within the core. Given the university context, the largest clusters likely represent major Greek organizations or large residential dorms. The Modularity score indicates a moderate separation; the groups are distinct but have significant cross-connections, preventing total isolation.

The network exhibits neutral/disassortative mixing. Despite the presence of a "popular crowd," the near-zero score suggests that high-degree nodes do not just form an exclusive club. Instead, the "social hubs" (like Node 5471) connect with a wide variety of peers across the network. This

"Hub-and-Spoke" model is typical of social networks, where popular leaders act as broadcasters to the general population rather than isolating themselves.

Modelling and Dynamics: SIR Spreading Process

To analyze how centrality influences information diffusion, we simulated an SIR (Susceptible Infected Recovered) spreading process (β=0.05, γ=0.1) under two distinct scenarios:

1. Hub Start: The “infection” (rumour) starts at the Super-Hub (Node 5471).
2. Random Start: The infection starts at a randomly selected node.

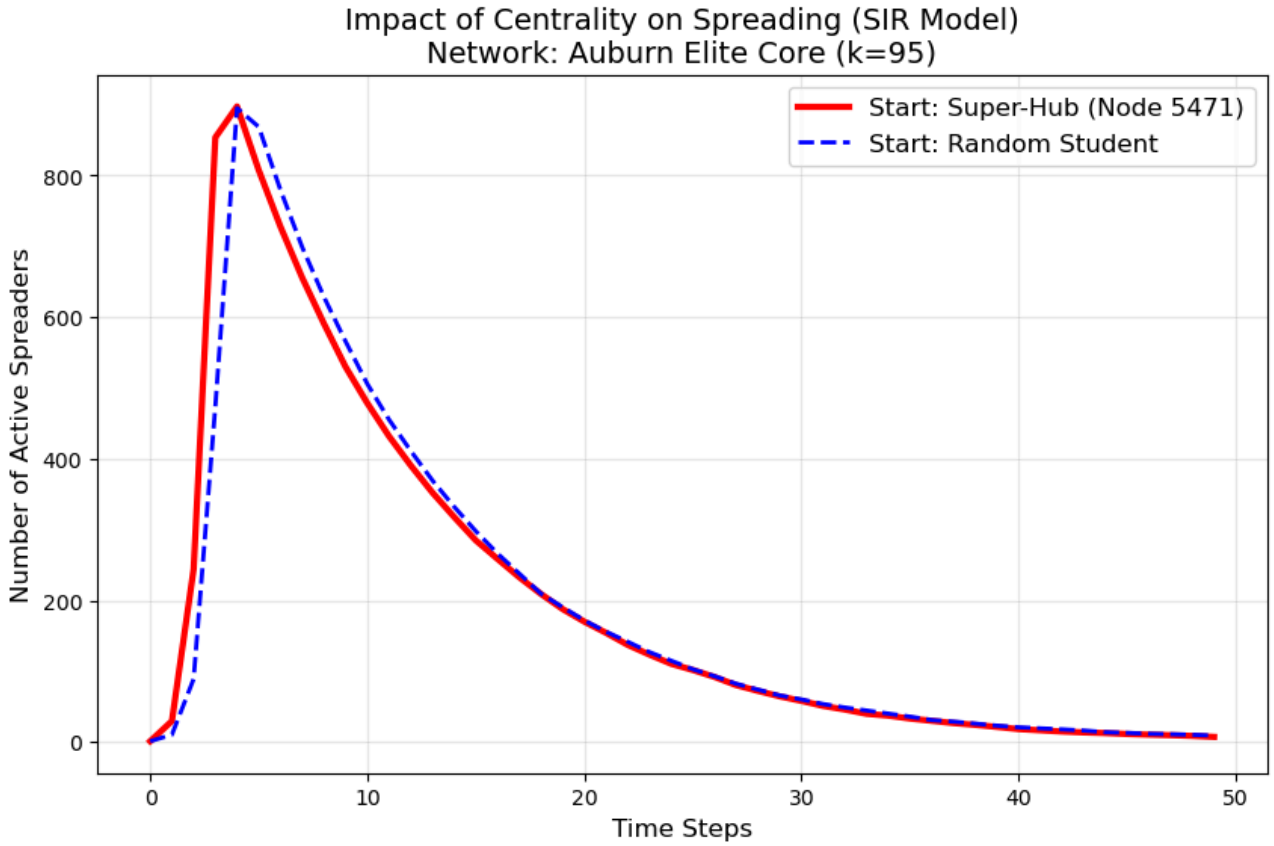

Fig. 6: Simulation of information diffusion (SIR Model)

The simulation demonstrates the critical role of the initial spreader's structural position:

- Explosive Diffusion (Hub Start): The red curve illustrates an immediate, vertical spike in active spreaders, infecting nearly 80% of the network within 5 time steps. Because Node 5471 connects directly to over 56% of the elite core, the rumour bypasses the typical "percolation phase."
- Slow Percolation (Random Start): The blue curve shows a delayed and lower peak (~200 active spreaders). A random student must rely on chance to pass the rumour to a hub; until that happens, the spread is contained locally.

The Auburn Elite Core is highly susceptible to information cascades, but only if initiated by a central actor. A rumour started by the Student Body President becomes campus knowledge instantly; a rumour started by a random student often fizzles out.

Limitations of the Analysis

While this study provides insights into the structural dynamics of the network, several limitations must be acknowledged:

1. Data Age: The dataset is from 2005. Social media dynamics (and the definition of a "Facebook Friend") have evolved significantly, meaning these findings reflect offline campus structures more than modern digital behaviour.
2. Selection Bias (K-Core): By filtering for k=95, we excluded over 90% of the population. Our findings of "explosive spreading" apply strictly to the elite, hyper-connected sub-population and likely overestimate the speed of diffusion for the general student body.
3. Simplification: The model assumes unweighted, undirected edges. In reality, influence is often directed and weighted. Treating all ties as equal oversimplifies the true flow of influence.
4. SIR Idealization: The simulation assumes a homogeneous mixing rate and does not account for complex human behaviours such as resistance to rumours or varying susceptibility based on social context.

Conclusion

The analysis of the socfb-Auburn71 dataset reveals a network characterized by a dense, highly interconnected core. The social structure is centralized around a few dominant "Super-Hubs"; specifically, Node **5471**, who possess both high popularity and strong influence. While distinct communities exist, the high density and lack of gatekeepers suggest that the social environment is porous. Information spreads rapidly and freely across group boundaries, driven by the immense connectivity of the central actors. The findings confirm that in the Auburn elite core, structural centrality is the primary determinant of information diffusion speed.

Reference

1. Traud, A. L., Mucha, P. J., & Porter, M. A. (2012). Social structure of Facebook networks. Physica A: Statistical Mechanics and its Applications, 391(16), 4165-4180.
2. Traud, A. L., Kelsic, E. D., Mucha, P. J., & Porter, M. A. (2011). Comparing community structure to characteristics in online collegiate social networks. SIAM Review, 53(3), 526-543.
3. Bastian, M., Heymann, S., Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks.
4. Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment, 2008(10), P10008.
5. Newman, M. E. J. (2002). Assortative mixing in networks. Physical Review Letters, 89(20), 208701. (The defining paper for assortativity).
6. Pastor-Satorras, R., & Vespignani, A. (2001). Epidemic spreading in scale-free networks. Physical Review Letters, 86(14), 3200.