Universiteit Leiden
The Netherlands

Opleiding Informatica

Topology of Decentralized Social Networks: A Case Study of Bluesky

Christos Georghiou

Supervisors:
Dr. A. Saxena & F. Corriera

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)
www.liacs.leidenuniv.nl
01/07/2025

Abstract

This thesis investigates the structural properties of the Bluesky follower-following network. Bluesky is a decentralized social media platform designed to operate without centralized content curation. This research investigates how users connect and organize on Bluesky by analyzing a large-scale follower-following dataset collected via the platform's official API. Through a network science approach, the analysis focuses on key topological properties, including degree distributions, clustering patterns, centrality measures, community structure, and assortative mixing, to assess the underlying dynamics of user connectivity and influence. The analysis shows a scale-free network with a densely connected core of highly followed users and a long tail of sparsely connected accounts. Despite the platform's decentralized architecture, the network demonstrates considerable structural inequality, with influence concentrated in a small subset of nodes. Community detection and homophily analysis further demonstrate that users tend to form tightly knit clusters, with strong preferences for within-community connections. These results suggest that social hierarchies and centralization of influence can emerge organically in decentralized systems, sparking discussions about the extent to which decentralization alone can address structural asymmetries in online networks.

Contents

1 Introduction

Bluesky is a decentralized microblogging platform that was introduced in 2019 as an internal project by Twitter and became an independent company in 2021. While it offers familiar features such as posting short messages, following users, and engaging through replies or likes, its technical foundation sets it apart. Bluesky is built on the Authenticated Transfer Protocol (AT Protocol), which enables a federated network of independently operated services. This architecture gives users more flexibility in how they interact with the platform, including how their data is stored and how content is filtered and moderated.

Online social networks now play a central role in how people communicate, form communities, and access information. Platforms like Twitter and Facebook influence what news users see, how public conversations unfold, and which voices are amplified. These platforms operate using ranking and recommendation algorithms that are typically hidden from users [Gil18]. These systems shape content visibility based on predicted engagement, which often results in a focus on popular or emotionally triggering content rather than balanced or diverse viewpoints.

This lack of transparency has raised concerns about fairness, manipulation, and the broader societal role of large platforms. When algorithms determine what users see, they also influence what conversations take place, which perspectives are elevated, and how communities are formed. The fact that these systems are developed and controlled by private companies creates an imbalance between the users who rely on these platforms and the entities that shape their experiences. It is plausible to suppose that such dynamics may reduce diversity in public discourse, reinforce existing inequalities, or create filter bubbles where people are only exposed to views that match their own.

In response to these issues, alternative models of online interaction have seen a rise in public attention. One of the most prominent approaches is decentralization, where control is distributed rather than concentrated in a single organization. Platforms like Bluesky aim to shift power back to users by letting them choose how their content is moderated, how recommendations work, and where their data is stored. This model has the potential to increase transparency and autonomy, but it also raises new questions. Without centralized systems guiding content visibility, how do communities evolve? Does user engagement become more evenly distributed, or do some users still dominate the network? Despite the growing interest in decentralized platforms, many important questions about how they function remain unanswered. Most existing studies continue to focus on centralized networks, where user data is more readily accessible and platform dynamics are better understood. As a result, there is still limited empirical insight into how decentralized networks form, how users engage with one another, and whether these systems lead to more equitable or democratic outcomes.

This thesis seeks to address this gap by examining the structural properties of the Bluesky follower-following network. The goal is to understand how users connect and organize themselves in a decentralized environment, and whether the resulting network structure differs meaningfully from that of traditional, centralized platforms. Using methods from network science, the research analyzes key features such as degree distribution, clustering, connectivity, and component sizes. Network analysis has helped in understanding various kind of networks, including transaction networks [SPV+21], online social networks [TPFG18], communication networks [MS24], co-authorship networks [SI16], citation networks, and so on. Network science has also been used to model dynamic processes taking place on these networks, such as influence propagation [SIG15], dynamics of influential leaders [BCG+19], opinion modeling [DGM14], influence blocking [SSR22], and so on. By focusing on the follower-following graph, this thesis contributes to a better understanding of whether decentralization of social networks meaningfully impacts the structural dynamics of social media networks.

1.1 Research Question

This thesis investigates how decentralized social networks, and more specifically Bluesky's follower-following network, function in the absence of centralized control and algorithmic curation. The central research question guiding this study is:

How does user interaction and network structure manifest on a decentralized platform like Bluesky, and how does this compare to traditional, centralized social networks?

To address this, the thesis is guided by the following main question:

1. What are the topological characteristics of the Bluesky follower-following network, and how do these compare to those found in centralized social networks (e.g., Twitter), particularly in terms of degree distribution, clustering, and connectivity?

1.2 Contributions

The contributions of this research are mentioned below.

- A topological analysis of the network, examining properties such as degree distributions, clustering coefficients, and the size of the giant component to understand user connectivity.
- A qualitative comparison of these findings with existing literature on centralized platforms like Twitter, to reflect on whether decentralization leads to more distributed engagement or the persistence of dominant user clusters.

1.3 Thesis overview

The remainder of this bachelor thesis is structured as follows. Section 2 outlines the theoretical foundation, introduces the core network analysis concepts used throughout the study, and reviews relevant literature on social network structures. Section 3 presents the dataset and sampling methodology. Section 4 reports the experimental results, while Section 5 interprets the findings and compares them to Twitter as a representative centralized platform. Finally, 6 concludes the thesis and reflects on potential directions for future research.

2 Related Work

2.1 Social Network Analysis

Social Network Analysis (SNA) focuses on examining the underlying structures that emerge from interactions between actors within a network. A major part of SNA is the study of topological metrics, including degree distribution, clustering coefficient, average path length, and centrality indices such as betweenness and PageRank [SI20]. These metrics offer insights into phenomena such as information diffusion, community formation and the emergence of influential users. Studies of online social networks, particularly Twitter, have identified skewed degree distributions typically following a power-law, where a few highly connected hubs exist among many sparsely connected nodes. Grandjean's analysis of Twitter demonstrated the presence of small-world properties, with short average path lengths and high clustering[Gra16], facilitating both rapid information diffusion and the formation of tightly knit communities. However, the majority of existing research focuses on centralized platforms, where governance and algorithms influence network formation. Recent work by Quelle and Bovet on Bluesky users[QB25] shows that small-world structures persist even in decentralized settings.

2.2 Centralized Social Networks

Centralized social networks, such as Twitter and Facebook, are characterized by the governance of the platform in shaping the formation of the network. In such environments, opaque algorithms shape content recommendations and trending topics, directly influencing which users gain visibility and how connections are formed. Thus, centralized platforms often reinforce preferential attachment mechanisms, whereby already prominent users attract a disproportionate number of new followers. Over time, this dynamic promotes the emergence if social hierarchies within the network, concentrating influence among a small subset of highly connected users. Additionally, the platforms' ability to moderate content and user behavior allows them to further shape the boundaries and internal structures of their social graphs.

2.3 Decentralized Networks and Bluesky

Decentralized social networks aim to distribute control over content moderation, recommendation algorithms, and network governance among third parties, rather than concentrating it within a single platform provider. This model is intended to increase transparency and user autonomy. Two recent examples of decentralized social networks are Mastodon and Bluesky, both microblogging platforms.

An analysis of Mastodon's follower-following network by Zignani, Gaito, and Rossi revealed heavy-tailed degree distributions, high clustering coefficients, and a tendency for users to form tightly knit communities within server instances[ZGR18]. Despite Mastodon's federated structure and the absence of global recommendation algorithms, its network topology mirrors the small-world properties typically found in centralized platforms.

Similarly, a recent study by Quelle and Bovet investigated the interaction network of approximately five million Bluesky users, analyzing follows, replies, likes, and reposts [QB25]. Their findings reveal that Bluesky also demonstrates heavy-tailed degree distributions, pronounced clustering, and short path lengths. These results suggest that the emergence of small-world structures and social hierarchies is a robust feature of online social networks, largely independent of platform governance models. Thus, while decentralization transforms governance and moderation mechanisms, it does not necessarily produce fundamentally different network topologies.

2.4 Relevance and Research Gap

While several studies have begun exploring decentralized platforms, analyses focusing on the follower-following networks of those platforms remain rare. As already mentioned, Zignani et al. and Quelle and Bovet offered insights into decentralized network topology. However, their work remains limited, particularly in relation to large-scale follower-following structures [ZGR18, QB25]. Moreover, prior studies often combine multiple interaction layers or focus on early-stage network snapshots, leaving open questions regarding the evolving topological characteristics of decentralized follower-following graphs at scale. Addressing the follower-following relationships within decentralized environments requires deeper and more targeted analyses.

This thesis builds on existing work by offering a targeted, empirical analysis of Bluesky's follower-following network. The aim is to gain a better understanding of how decentralization influences the network's topological dynamics.

3 Data Collection

3.1 Data Acquisition

The follower-following network data was collected using the official Bluesky API. Two primary endpoints were used: app.bsky.graph.getFollowers, which retrieves the list of followers for a given user, and app.bsky.graph.getFollows, which retrieves the list of accounts that a user follows.

Data collection began by accessing the "What's Hot" feed, which lists the top trending posts on Bluesky. The authors of the top ten posts were extracted to form the initial seed set of users. Using the API, the followers and followings of each of these user were collected and incorporated into the dataset. The dataset was then expanded iteratively: at each step, a user who had been discovered but whose follower and following lists had not yet been explored was selected at random. By successively retrieving and incorporating the follower and following lists of unexplored users, the dataset gradually expanded into a large and interconnected subset of the network.

Data collection was performed between February and June 2025, and included only publicly accessible profiles. No private or restricted information was accessed.

3.2 Dataset Description

The complete dataset collected consists of 4,410,329 users and 25,056,375 directed edges, representing follower-following relationships across Bluesky. However, due to computational constraints, the analysis presented in this thesis is performed on a sampled subgraph of 20,000 users, extracted via random walk traversal.

This subgraph preserves key structural properties of the full network and serves as the base for the experimental results discussed in subsequent chapters. The sampled subgraph is visualized in Figure 1. Nodes are sized and colored based on their in-degree scores, with larger and warmer-colored nodes (e.g., red or orange) representing users with high follower counts. The visualization revealed a densely connected core of high in-degree nodes, surrounded by a sparse periphery of low-degree users. Several small, weakly connected clusters at the margins indicate structural fragmentation. This pattern reflects the network's scale-free topology, where a minority of users accumulate disproportionate visibility despite the absence of centralized recommendation systems.

The data was stored as a two-column CSV file, with each row recording a directed connection: the first column identifies the source node (the follower), and the second column identifies the target node (the user being followed). This format, known as an edge list, reflects the directed nature of the social graph and facilitates the application of network analysis methods, including degree distribution studies, clustering coefficient measurements, and connectivity evaluations.

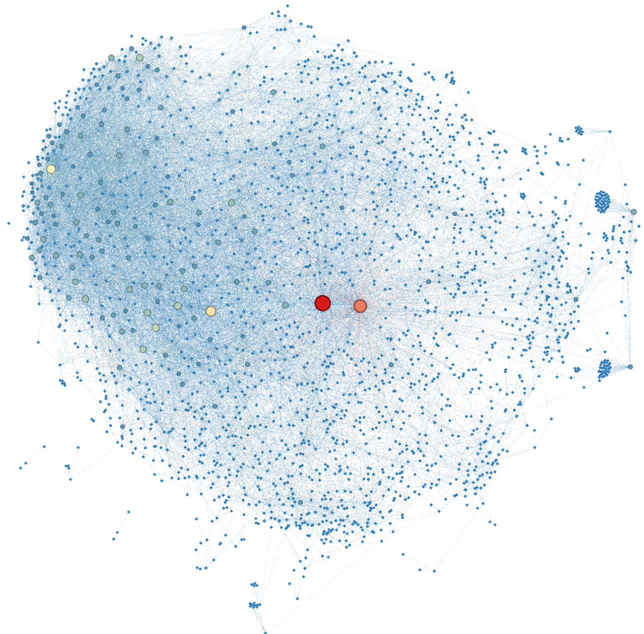

Figure 1: Visualization of a 20,000-node subgraph obtained via random walk sampling. Node in-degree is encoded by size and color (blue = low, yellow = medium, red = high).

4 Network Topology Analysis

This section presents a comprehensive analysis of the structural properties of the Bluesky follower-following network. Various network science metrics are employed to characterize user connectivity, identify influential nodes, detect communities, and assess the overall structure of the graph.

4.1 Centrality Measures

Centrality metrics quantify the structural importance of nodes within a network. This thesis uses the following centrality measures: degree [SGI18], betweenness, PageRank, and closeness. These measures are used to assess the influence, reachability, and network positioning within the Bluesky follower-following graph.

4.1.1 Degree

Degree refers to the number of connections a node has. In directed graphs, in-degree measures how many followers a user has, while out-degree measures how many users a given account follows. These metrics reveal visibility and user activity within the network [New03]

In-degree $k_{\text{in}}(v)$ measures how many followers a user has. Formally, it is defined as:

$$k_{\text{in}}(v) = |\{u \in V : (u, v) \in E\}|$$

This captures the number of incoming edges to node v, representing how often the user is followed.

Out-degree $k_{\text{out}}(v)$ measures how many users a given account follows:

$$k_{\text{out}}(v) = |\{u \in V : (v, u) \in E\}|$$

This corresponds to the number of outgoing edges from node v, indicating user activity.

Analyzing degree distributions is fundamental to understanding how connectivity, influence, and structural inequality emerge within a social network.

These distributions are known to reveal important topological signatures of social graphs, such as scale-free structure and preferential attachment, both of which have been extensively observed in large-scale online social networks like Twitter and Facebook[KLPM10, MMG+07]. To investigate whether similar patterns are present in a decentralized setting, the in-degree and out-degree distributions of the Bluesky graph were analyzed on a log-log scale.

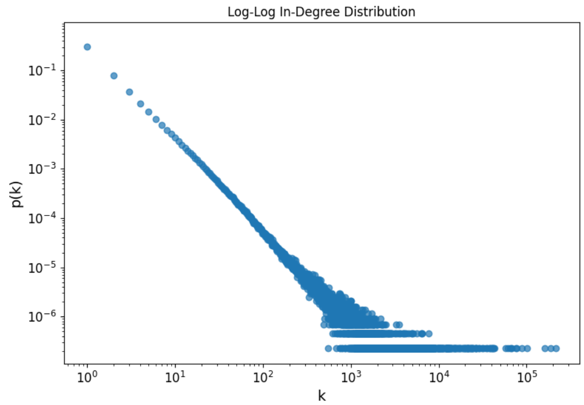

Figure 2: Log-log plot of the in-degree distribution in the Bluesky follower-following network. The x-axis shows the in-degree k and the y-axis shows the probability $P(k)$.

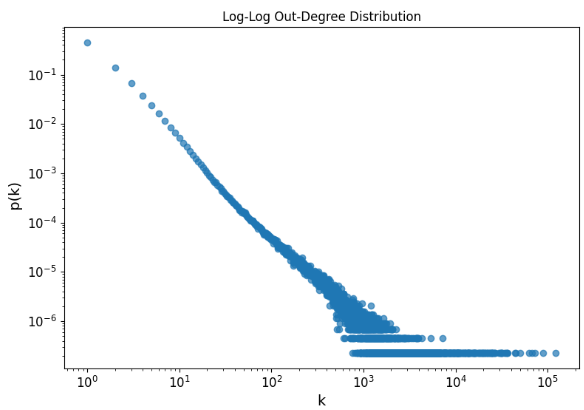

Figure 3: Log-log plot of the out-degree distribution in the Bluesky follower-following network. The x-axis shows the out-degree k and the y-axis shows the probability $P(k)$.

4.1.2 In-Degree Distribution

Figure 2 shows that the in-degree distribution is heavy-tailed. The majority of users receive very few followers, while a small subset accumulates a disproportionately large number. Some accounts in the sample exceed 175,000 followers, acting as highly influential hubs within the platform.
This shape is characteristic of scale-free networks, where a small number of nodes dominate connectivity[BA99]. The near-linear tail in the log-log plot supports the presence of a power-law distribution. This pattern suggests that Bluesky exhibits preferential attachment, where users are more likely to follow already-popular accounts[New05, KLPM10]. Despite its decentralized design, the platform reproduces centralizing dynamics that are commonly found in mainstream social media.

4.1.3 Out-Degree Distribution

Figure 3 shows that the out-degree distribution also follows a heavy-tailed pattern. Most users follow only a small number of accounts, while a few follow tens of thousands. In some cases, users follow over 70,000 accounts. These outliers likely represent automated or bot-like behavior, potentially aimed at inflating visibility or engagement.
This distribution aligns with scale-free properties and reflects strong heterogeneity in user activity. The concentration of highly active users indicates that connection patterns may be shaped by activity-based biases or automated behaviors, which could distort the organic growth of the network. Combined with the in-degree analysis, this reinforces that decentralization does not inherently produce a more balanced or egalitarian structure.

4.1.4 Betweenness Centrality

Betweenness centrality measures how frequently a node appears on the shortest paths between other nodes. This measure captures a user's ability to control or mediate information flow between otherwise unconnected parts of the graph. In online social networks, users with high betweenness centrality often play structurally influential roles by connecting clusters or communities that would otherwise remain isolated[New05].
For a node v, betweenness centrality is defined as:

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

where σ_{st} denotes the total number of shortest paths between nodes s and t, and $\sigma_{st}(v)$ refers to the number of those paths that traverse node v.

The distribution of betweenness centrality scores is shown in Figure 4. The majority of users have low or near-zero scores, while a small fraction exhibit significantly elevated values.
This pattern reflects a common finding in social networks: most users do not facilitate connections between communities, while a small subset acts as critical intermediaries. These high-betweenness users may not be the most followed or visible in the network, but their structural position grants them potential influence over how information spreads across clusters. The long-tailed nature of the distribution also highlights a persistent structural inequality in connectivity, a feature previously documented in studies of centralized platforms such as Twitter[KLPM10].

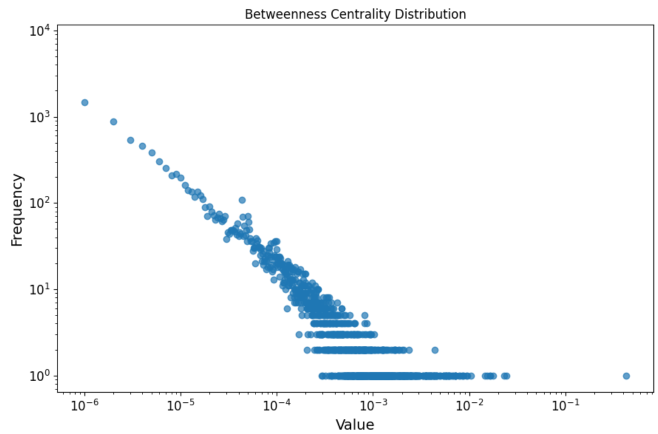

Figure 4: Distribution of betweenness centrality scores in the Bluesky follower-following network.

Despite Bluesky's decentralized design, the emergence of users with disproportionately high betweenness scores suggests that decentralized governance alone does not prevent the formation of hierarchical structures based on positional influence.

4.1.5 PageRank Centrality

PageRank is a centrality measure that assesses a node's influence in a directed network by accounting for both the number and the importance of its incoming links. Originally introduced by Brin and Page in the context of web graphs[PBMW99], it has since been broadly adopted in social network analysis to capture recursive importance. A node is considered influential not only if it receives many links, but especially if those links come from influential nodes.
The PageRank score of a node v is defined recursively as:

$$PR(v) = \frac{1-d}{N} + d \sum_{u \in \mathcal{N}^-(v)} \frac{PR(u)}{k_{\text{out}}(u)}$$

where d is the damping factor (d = 0.85), N is the total number of nodes in the graph, $\mathcal{N}^-(v)$ denotes the set of nodes linking to v, and $k_{\text{out}}(u)$ is the out-degree of node u. The damping factor models the probability that a user will jump to a random node, preventing rank sink in disconnected components.

The distribution of PageRank scores across the Bluesky network is presented in Figure 5. The resulting scores were binned and plotted on a log-log scale, using a total of 400 logarithmic bins. This distribution is characteristic of scale-free networks, where influence becomes highly concentrated among a small subset of nodes. Although the majority of accounts exhibit low PageRank scores, a small fraction of accounts possess disproportionately high centrality, resulting in a long right tail in the plot. This is not solely a function of high in-degree, but also of the fact that these users are followed by others who themselves hold significant structural importance.

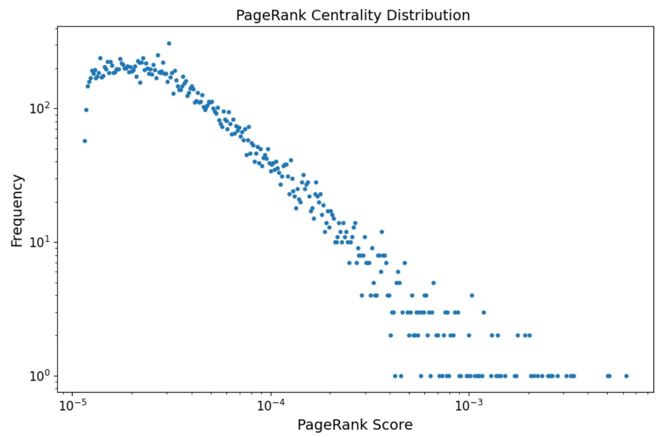

Figure 5: Distribution of PageRank scores in the Bluesky follower-following network.

The findings align with prior observations from centralized social media platforms, where PageRank distributions similarly reveal pronounced inequalities in user influence. Notably, the emergence of such concentration within Bluesky, despite its decentralized architecture, suggests that hierarchical patterns of influence may arise within social networks, independent of centralized control mechanisms such as algorithmic curation or content moderation. This raises important questions about whether decentralization alone is sufficient to counteract structural inequalities in user visibility and influence.

4.1.6 Closeness Centrality

Closeness centrality measures how near a given node is to all other nodes in a network, based on the shortest path distances between them. It reflects the potential efficiency with which a node can spread information through the network, or alternatively, how quickly it can access information from other users. [F+02] In the context of online social platforms, users with high closeness centrality tend to be structurally well-positioned to reach broad parts of the network with minimal intermediary steps.
Formally, the closeness centrality $C_C(v)$ of a node v is defined as:

$$C_C(v) = \frac{1}{\sum_{t \in V \setminus \{v\}} d(v,t)}$$

where $d(v,t)$ denotes the length of the shortest path from node v to node t, and V is the set of all nodes in the graph. This formula assumes that all nodes are reachable; when applied to disconnected graphs or samples, it is typically computed only over the largest connected component or adjusted to exclude unreachable nodes.

In this study, closeness centrality was computed using the reversed sampled graph of 20,000 nodes, obtained via random walk traversal. The sampled graph was reversed in order to measure how efficiently a user can be reached by others. Due to the nature of random walk sampling, the resulting subgraph may not be fully connected, and thus contains nodes with undefined closeness, which are assigned a score of zero.

The resulting distribution of closeness scores is displayed in Figure 6. The distribution of closeness centrality values is distinctly bimodal, with two sharply separated clusters. The primary mass of users has closeness centrality values concentrated tightly around 10^{-1}, while a separate group is spread across extremely low values ranging from approximately $10^{-4.5}$ to $10^{-3.5}$. Notably, there is a clear gap between these two regions, with virtually no users occupying the intermediate range between $10^{-3.5}$ and 10^{-2}. This discontinuity suggests the presence of structurally distinct subpopulations in the network.

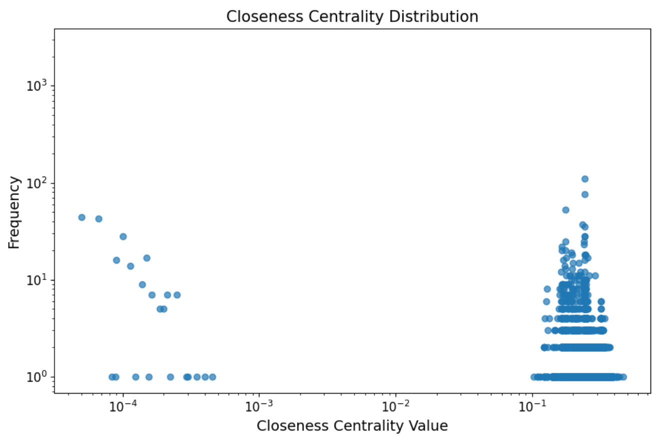

Figure 6: Distribution of closeness centrality values in the Bluesky follower-following network.

The dominant cluster near 10^{-1} represents users embedded within the largest connected component, where average path lengths to others are short. These users occupy positions that allow them to access much of the network with minimal intermediary steps. Conversely, the tail of nodes with very low closeness centrality likely reflects users in disconnected or weakly connected components. This is expected, given the limitations of random walk sampling in capturing all reachable components from the perspective of inward paths. These users may represent isolated accounts, fringe communities, or remnants of partial sampling.

These results highlight that structural centrality in Bluesky is unevenly distributed, with a few nodes having topological proximity to most others. Despite the platform's decentralized architecture, the presence of such core nodes suggests that centrality of reach can still emerge without algorithmic mediation, reinforcing the broader finding that social hierarchies are not solely the product of centralized design.

4.2 Clustering Coefficient

The clustering coefficient measures the likelihood that a user's neighbors are also connected to one another, forming closed triads.[WS98] Although the Bluesky follower-following network is directed, the clustering coefficient is computed on its undirected form to capture local cohesion. The local clustering coefficient of a node v in an undirected graph is formally defined as:

$$C(V) = \frac{2T(v)}{k_v(k_v - 1)}$$

where $T(v)$ is the number of triangles that are passing through node v, and k_v is the degree of v. This measure quantifies the extent to which a node's neighbors are also connected to each other, capturing the presence of tightly knit groups in the network. High clustering coefficients indicate locally cohesive structures, while low values suggest sparse or hub-like connectivity.

To examine how local clustering varies with node connectivity, the average clustering coefficient was evaluated across all nodes with the same degree k, resulting in the conditional expectation $\langle C \mid k \rangle$. As shown in Figure 7, the average shows a clear negative correlation with degree. Low-degree nodes tend to be embedded within highly clustered local structures, suggesting the presence of tightly-knit communities. Conversely, high-degree nodes often serve as bridges or hubs, linking distinct parts of the network while remaining structurally isolated from cohesive local groups.

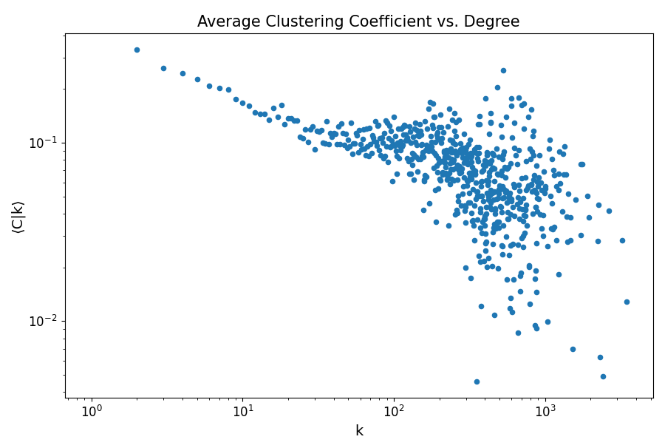

Figure 7: Average clustering coefficient $\langle C \mid k \rangle$ as a function of node degree k in the Bluesky follower-following network.

This inverse relationship reflects a hierarchical organization of the network, where low-degree nodes are part of tightly knit clusters, while high-degree nodes function as inter-community connectors. Such structural patterns are typical of complex real-world networks, particularly online social networks, where highly connected hubs facilitate integration by linking disconnected local communities. [RB03]

4.3 Connectivity

This section examines the structural cohesion and navigability of the Bluesky follower-following network by analyzing its connected components, local neighborhood properties, and reachability dynamics. Together, these measures offer insights into how information may propagate and how tightly users are embedded within the social graph.

4.3.1 Connected Components

In network analysis, a connected component is a group of nodes within which each node is reachable from any other. In directed graphs, two types of components are typically distinguished. A strongly connected component (SCC) requires that every node be reachable from every other via directed paths, whereas a weakly connected component (WCC) allows reachability when edge direction is ignored. A large WCC suggests that most users are part of a globally navigable structure, while the distribution of SCCs reveals the extent to which users form tightly integrated subgroups or remain isolated.

To assess the global connectivity of the Bluesky follower-following network, we analyze both SCCs and WCCs in the sampled graph.
The network contains 2 weakly connected components, with the largest WCC including 99.99% of all nodes (19,997 out of 20,000). This indicates that, structurally , the graph is nearly entirely connected when directionality is disregarded. However, the picture becomes more fragmented when considering direction. The network contains 2,828 strongly connected components, with the largest SCC containing 85.38% of all nodes (17,076 out of 20,000). This implies that while the majority of users form a mutually reachable core, a large number of users exist in more peripheral positions.

This structure is typical in directed online social networks, where popular accounts are disproportionately followed, creating an asymmetric topology. The existence of a large SCC reinforces the view that the network is cohesively structured around a highly interconnected core, while large WCC suggests the potential for wide information diffusion even among loosely connected peripheral users. This aligns with the findings of Guille et al. [GHFZ13], who highlight that online social networks often exhibit a core-periphery structure that facilitates efficient information diffusion across the network.

4.3.2 Average Neighbor Degree

To explore the local structural patterns of the network, the average neighbor degree $k_{nn}(v)$ is computed, which captures whether users tend to connect with others who are more or less connected than themselves. The average neighbor degree measures the average connectivity of a node's immediate neighbors.[PSVV01] It is defined as the mean degree of all nodes directly connected to a given node. For a node v, it is defined as:

$$k_{nn}(v) = \frac{1}{k_v} \sum_{u \in \mathcal{N}(v)} k_u$$

where k_v is the degree of node v, $N(v)$ is its set of neighbors, and k_u is the degree of neighbor u. By computing the average neighbor degree for nodes of degree k, we capture the network's assortativity profile. If average neighbor degree increases with node degree, the network exhibits assortative mixing. A decreasing trend indicates disassortative structure, which is often found in hierarchical social graphs.

The results of the analysis of the average neighbor degree are shown in Figure 8. The average neighbor degree initially increases with node degree, peaks at intermediate values of k, and then declines for the nodes of higher degree. This bell-shaped distribution indicates that low and mid-degree users tend to connect to highly connected hubs, whereas high-degree users mainly connect to lower-degree nodes. This pattern reflects a disassortative mixing, in which a small set of influential users form the core of the network while maintaining asymmetric links with a broader, less connected periphery.

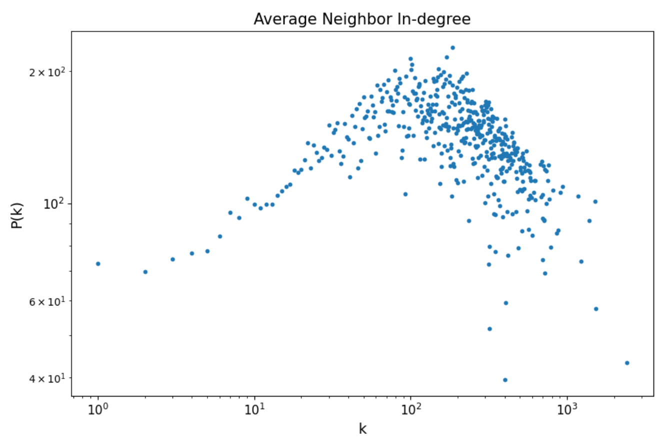

Figure 8: Average neighbor degree k_nn as a function of node degree k in the Bluesky follower-following network.

The findings from the average neighbor analysis provide a more localized perspective on connectivity patterns and align with the global assortativity analysis discussed in Section 4.4.2. While the latter quantified overall degree correlations using Pearson's correlation coefficient, the average neighbor degree captures how these patterns emerge at the level of immediate connections. Together, both analyses reveal consistent dissassortative tendencies and reinforce interpretation of the Bluesky network as hierarchically structured around central, high-degree hubs.

4.3.3 Reachability via Snowball Sampling

Snowball sampling is a graph traversal method used to simulate the expansion of influence or information through a network. Starting from one or more seed nodes, it iteratively collects all direct neighbors at each layer, expanding outward in breadth-first order.[Goo61] This process is typically repeated for a fixed number of layers or until growth saturates. The method is useful for assessing reachability, identifying the size of connected regions, and understanding how quickly a network can be traversed from arbitrary entry points.

To evaluate the network's global expansion dynamics, a snowball sampling was conducted using Breadth-First Search (BFS) from six randomly selected seed nodes, extending up to seven layers from each origin. This approach simulates how influence or information would spread outward from a user through successive layers of connections, revealing how quickly the network becomes saturated as more users are reached.

As illustrated in Figure 9, the sampling curves show a two-phase pattern. In the early stages (depth 1-4), the number of newly discovered users increases rapidly, approximating exponential growth. This behavior reflects the presence of high-degree hubs that serve as central channels, dramatically increasing reachability in just a few steps. These hubs act as bridges between otherwise distant parts of the network, facilitating fast expansion from local origins.

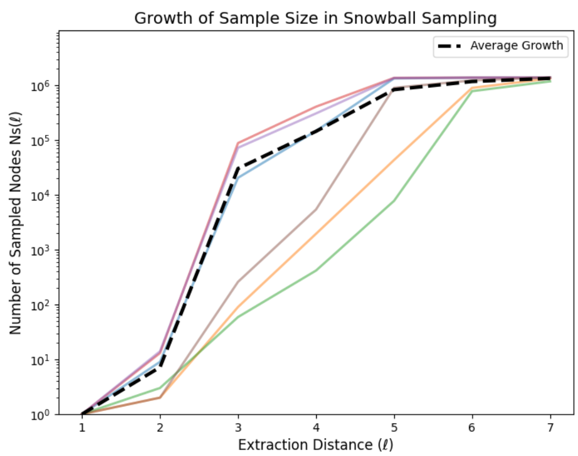

Figure 9: Growth of sample size in snowball sampling.

In later stages (depth 5-7), the curves plateau across all trials, indicating that most reachable users have already been discovered by this point. This pattern suggests the presence of a giant connected component, in which most nodes are structurally embedded and accessible within a limited number of steps from any given origin. Such a pattern is characteristic of small-world networks, which are marked by short average path lengths and high navigability, even at scale.

These structural features suggest that the Bluesky network supports efficient information diffusion and exhibits substantial global cohesion. Together with the connected component analysis, the reachability results provide further evidence that the Bluesky possesses a small-world topology, which is a common property of online social systems [WS98].

4.4 Community Structure

A community represents a densely connected group of nodes and sparsely connected with the rest of the network, and there exists many different types of methods for identifying communities [APS22]. This section investigates the community of the Bluesky follower-following network by analyzing three interrelated metrics: modular community detection, assortativity, and homophily. These aspects provide insight into the degree to which users are embedded in tightly knit groups how such groups are organized at the network level, and whether users tend to connect to structurally or categorically similar users.

4.4.1 Community Detection

Community detection identifies clusters or groups of nodes that are more densely connected internally than with the rest of the network. These structures help reveal latent groupings and social segmentation in complex networks.

This thesis applies the Louvain algorithm, a widely adopted heuristic for detecting communities by maximizing modularity. Modularity measures the difference between the observed density of edges within communities and the expected density in a randomized network with the same degree distribution [BGLL08]. This method optimizes modularity Q, defined as:

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

where A_{ij} represents the adjacency matrix, k_i and k_j are the degrees of nodes i and j, respectively, m is the total number of edges in the network, and $\delta(c_i, c_j) = 1$ if nodes i and j belong to the same community, and 0 otherwise. A higher value of Q indicates a stronger community structure, signifying a greater concentration of intracommunity edges than expected in a comparable random network.

While clustering characterizes local cohesiveness, community detection focuses on identifying larger-scale structures in which nodes are more densely connected internally than externally. To detect such meso-scale groupings, the Louvain algorithm was applied to the undirected version of the sampled Bluesky network.

The algorithm partitioned the network into 14 communities, with a modularity score of 0.3716, indicating a moderate level of community structure. The detected communities vary widely in size, from 2 to 5,754 nodes, as summarized in Table 1
The modular structure of the network is further illustrated in Figure 10, which presents a gephi-rendered visualization with nodes colored by their community affiliation. In the visualization, the largest communities occupy dense central regions, while smaller groups tend to be closer to the edge. Moreover, despite color coordination between the different communities, it is difficult to identify all 14 communities. Notably, the gephi visualization 10 revealed overlapping boundaries between communities, which suggests limited modular separation and inter-community blending. This observation aligns with the relatively modest modularity value and is further examined in the next part, which investigates assortativity and homophily patterns.

Community ID	Size (nodes)	Percentage of total
2	5,754	28.77%
4	4,263	21.31%
13	2,779	13.90%
0	2,397	11.99%
6	1,855	9.28%
1	1,242	6.21%
9	788	3.94%
8	574	2.87%
7	194	0.97%
12	98	0.49%
5	29	0.15%
10	20	0.10%
3	4	0.02%
11	2	0.01%

Table 1: Community sizes identified by the Louvain algorithm.

(a) (b)

Figure 10: Community Structure visualization using Gephi.

4.4.2 Degree assortativity

Assortativity describes the tendency of nodes in a network to connect with others that are similar in some respect. In directed networks, assortativity can be measured along different dimensions, most notably degree-based and attribute-based assortativity.[New02]

Degree assortativity measures the statistical correlation between the degrees of nodes on either side of a directed edge. Typically, in directed networks, the assortativity is computed separately for in-degree and out-degree using the Pearson correlation coefficient. Positive assortativity values indicate a tendency for nodes to connect with others that have similar degree characteristics. Conversely, negative values suggest dissasortative mixing, where high-degree nodes tend to connect to low-degree nodes, often demonstrating hierarchical-like structures.

In the sampled Bluesky network, the in-degree assortativity was calculated as -0.0424, and the out-degree assortativity as -0.0812. Both values are slightly negative, indicating a weak but consistent dissasortative trend. This pattern reflects the tendency of high-degree nodes (hubs) to attract connections from peripheral, low-degree users, rather than forming reciprocal links with other highly connected nodes.

4.4.3 Community-Based Assortativity (Homophily)

Attribute-based assortativity, or homophily, captures the extent to which nodes connect to others with similar attributes. In this analysis, homophily was assessed based on community labels produced by the Louvain algorithm. A high level of homophily indicates that users tend to form connections within their own community, reflecting cohesive and internally clustered groupings.

The resulting coefficient of 0.4732 indicates a moderate-to-strong preference for intracommunity connections. This means that users are significantly more likely to follow users within their own community than outside of it. This result reinforces the earlier observation of partially overlapping but internally cohesive communities. While the modularity score of 0.3716 indicates only a moderate separation between communities, the homophily coefficient of 0.4732 suggests that users have a clear preference for intracommunity connections. Thus, users predominantly engage with others within their assigned clusters, although intercommunity interactions remain present.

Metric	**Value**	**Interpretation**
In-degree assortativity	–0.0424	Weak disassortative mixing
Out-degree assortativity	–0.0812	Weak disassortative mixing
Homophily	0.4732	Moderate-to-strong intra-community preference

Table 2: Summary of assortativity and homophily metrics.

5 Discussion

The structural analysis of Bluesky's follower-following network reveals several characteristics commonly associated with centralized social platforms. Degree distributions are heavily skewed, with a small number of users accumulating the vast majority of followers. Centrality metrics confirm that influence is concentrated in a limited set of structurally dominant nodes, while clustering and community detection indicate the formation of tightly knit local groups. Despite Bluesky's decentralized infrastructure, these patterns reflect the emergence of hierarchical ordering and structural inequality. This chapter interprets the topological implications of the analysis, examines how decentralization interacts with emergent centrality, and positions Bluesky's network structure in comparison with prior analyses of centralized platforms.

5.1 Topological Implications

The structural properties of the Bluesky follower-following network reveal that its decentralized architecture does not eliminate the formation of hierarchy or concentrated influence. Despite the absence of centralized recommendation systems or algorithmic ranking, the network's degree distributions are heavy-tailed. A small fraction of users accumulate over 175,000 followers while most have very few. This reflects a classic power-law structure, with preferential attachment dynamics. Similarly, the out-degree distribution follows a skewed pattern. Some users follow tens of thousands of accounts, a behavior likely driven by strategic engagement practices intended to boost visibility.

Centrality measures reinforce this concentration of influence. Users with the highest PageRank scores are not only those with high in-degree but are also followed by others who themselves hold substantial centrality, reflecting the recursive nature of influence in the network. The PageRank distribution reveals that centrality is highly unequal, with a small number of accounts exhibiting disproportionately high scores. Closeness centrality results show that a small subset of users can reach the rest of the network through relatively short paths, indicating their central positioning within the overall topology. Betweenness centrality also identifies a limited group that functions as critical intermediaries, connecting otherwise distant or disconnected clusters. The convergence of these different centrality dimensions, visibility, reachability, and brokerage, shows that structural dominance in the Bluesky network is not only present but multidimensional. This pattern is further reinforced by the network's slightly negative degree assortativity scores, which indicate that highly connected nodes tend to form links with low-degree users rather than with each other. This disassortative mixing strengthens hub-periphery dynamics and deepens the asymmetry in influence distribution.

The community structure analysis further illustrates how localized cohesion coexists with broader network segmentation. The Louvain algorithm partitioned the network into 14 distinct communities, with a modularity score of 0.37, reflecting a moderate level of internal clustering. Community sizes vary substantially, ranging from isolated clusters of just a few nodes to a dominant group of more than 5,000 users. This distribution suggests that while users form tightly-knit groups, the overall network maintains a loose modular configuration rather than sharp fragmentation [BGLL08]. Clustering coefficients were highest among low-degree users, reinforcing the presence of dense local triads in smaller sub-networks. The homophily coefficient of 0.47 further supports this pattern, indicating a moderate-to-strong preference for intracommunity connections. This tendency strengthens internal cohesion but may reduce exposure to content or interactions originating outside one's immediate social group.

The network also exhibits strong small-world properties, remaining globally navigable. Reachability analysis using snowball sampling confirms this, showing that most users can be reached within five to seven hops from random starting points. This suggests relatively short average path lengths throughout the network. Global connectivity is further supported by a giant weakly connected component that includes 99.99% of sampled users [WS98]. At the same time, a more nuanced core-periphery structure becomes visible when directionality is considered. The largest strongly connected component, comprising 85.38%, forms a mutually reachable core, while the remaining users are positioned at the network's periphery, likely experiencing more limited integration. Ultimately, although the system is technically decentralized, network access and content diffusion appear largely mediated through this core of highly connected nodes.

In sum, the findings demonstrate that decentralization at the protocol level does not guarantee egalitarian network structures. The topology of Bluesky replicates several properties associated with centralized platforms, including concentrated visibility, network-driven hierarchy, and fragmented but cohesive sub-communities. The absence of central control does not prevent the emergence of structural power, which arises through user interaction dynamics and network effects.

5.2 Bluesky vs. Other Social Networks

As previously established, Bluesky's follower-following network exhibits structural patterns commonly observed in centralized platforms such as Twitter. This section compares Bluesky's network structure to findings from Kwak et al.[KLPM10], Grandjean [Gra16], and Myers et al.[MSGL14], in order to assess whether decentralization meaningfully alters the topological outcomes typically observed in large-scale social graphs.

The concentration of influence on Bluesky, as reflected in its heavy-tailed degree distributions and sharply unequal centrality scores, closely mirrors the structure described by Kwak et al. [KLPM10]. Their analysis of Twitter follow graph revealed that a small number of users received most incoming connections, producing a pronounced power-law distribution. Bluesky exhibits a similar pattern, with a few users dominating in-degree and PageRank centrality, while betweenness and closeness metrics highlight their strategic roles as bridges within the network. Myers et al. [MSGL14] found weakly dissassortative degree correlations in Twitter's graph, which is consistent with Bluesky's own negative assortativity values for both in-degree (-0.0424) and out-degree (-0.0812). These values suggest the presence of a hub-periphery structure, where highly connected nodes tend to attract links from peripheral users rather than from each other.

Community formation dynamics show additional similarities. Grandjean [Gra16], in his analysis of Twitter's digital humanities community, observed the emergence of modular subgroups based on shared academic interests. Bluesky's network structure follows a similar trajectory. The Louvain algorithm identified 14 distinct communities, with a modularity score of 0.3716 and a homophily coefficient of 0.4732. These results indicate that users have a tendency to cluster internally, favoring intracommunity connections over external ones. High clustering coefficients among low-degree users also support the presence of dense, localized triads within the network. Such modular structures, existing in both Bluesky and Twitter, seem to be an outcome of user interactions rather than a feature dictated by the platforms' underlying design.

Network connectivity and navigability also reflect similar properties. Kwak et al. [KLPM10] reported that 99.97% of Twitter users were part of a single weakly connected component, while the largest connected component was significantly smaller. Bluesky shows a near-identical structure, with 99.99% of the sampled users belonging to the WCC and 85.38% to the SCC. Snowball sampling on Bluesky revealed that the majority of users could be reached within five to seven hops from random seed nodes. While this figure is slightly higher than Twitter's reported average path length of 4.12 [KLPM10] and Grandjean's community-level estimate of 2.29 [Gra16], the result remains consistent with the small-world property. Although snowball sampling and average path length measure reachability differently, both reflect the network's overall navigability. These findings suggest that Bluesky, like Twitter, supports rapid information spread and efficient connection between users, despite differences in platform architecture.

Overall, these comparisons suggest that decentralization at the protocol level does not inherently reshape the emergent structure of large-scale social graphs. Bluesky mirrors key structural patterns observed in centralized platforms like Twitter, including concentrated influence, modular community formation, and a pronounced core-periphery topology. These findings raise broader questions about the limits of decentralization as a corrective to structural inequality and segmentation in digital social networks. These findings further imply that user-driven interaction dynamics may play a more decisive role in shaping network structure than the underlying platform architecture.

6 Conclusions and Further Research

This thesis examined the structure of the Bluesky follower-following network to assess how decentralized platforms evolve at scale and whether their topologies differ meaningfully from those of centralized systems. While Bluesky operates without centralized control, its network architecture exhibits features commonly associated with traditional social media platforms, such as highly visible hubs and cohesive communities.

The results of this analysis point to an important distinction between decentralized design and decentralized outcomes. Protocol-level decentralization does not automatically produce uniformly distributed social structures. Instead, the findings suggest that even in the absence of centralized recommendation systems, interaction dynamics, such as preferential attachment, visibility loops, and localized clustering, continue to drive hierarchical and modular patterns. This challenges the notion that decentralized protocols can exclusively mitigate the concentration of influence, similarly observed on major social media platforms.

Although this study offers a snapshot of Bluesky's early-stage network, it also highlights several limitations and directions for future research. The analysis focused exclusively on the topological features of the follower graph. Future work could incorporate temporal dynamics, interaction layers, and content diffusion, extending beyond static connectivity to actual patterns of influence and communication. One promising direction would be to combine structural analysis with content-based engagement studies. For instance, Quelle & Bovet [QB25] show that the network structure can reinforce exposure biases even without traditional algorithmic curation, highlighting how user-driven sorting and platform design features together influence the dynamics of online discourse.

Looking beyond Bluesky, comparisons with other decentralized platforms such as Mastodon could clarify whether the patterns observed here reflect something unique about Bluesky's design or point to more general trends in decentralized networks. This kind of research could help unpack how platform architecture, governance, and user behavior interact to shape visibility and influence online.

In short, the findings presented here show that decentralization changes the architecture of control, but not necessarily the structure of influence. Understanding how social networks form and stratify in decentralized environments remains a critical challenge that requires further empirical, theoretical, and design-focused research.

References

[APS22] Aikta Arya, Pradumn Kumar Pandey, and Akrati Saxena. Node classification using deep learning in social networks. In *Deep Learning for Social Media Data Analytics*, pages 3–26. Springer, 2022.

[BA99] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[BCG+19] Ivan Bermudez, Daniel Cleven, Ralucca Gera, Erik T Kiser, Timothy Newlin, and Akrati Saxena. Twitter response to munich july 2016 attack: Network analysis of influence. *Frontiers in big Data*, 2:17, 2019.

[BGLL08] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.

[DGM14] Abhimanyu Das, Sreenivas Gollapudi, and Kamesh Munagala. Modeling opinion dynamics in social networks. In *Proceedings of the 7th ACM international conference on Web search and data mining*, pages 403–412, 2014.

[F+02] Linton C Freeman et al. Centrality in social networks: Conceptual clarification. *Social network: critical concepts in sociology. Londres: Routledge*, 1(3):238–263, 2002.

[GHFZ13] Adrien Guille, Hakim Hacid, Cecile Favre, and Djamel A Zighed. Information diffusion in online social networks: A survey. *ACM Sigmod Record*, 42(2):17–28, 2013.

[Gil18] Tarleton Gillespie. *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press, 2018.

[Goo61] Leo A Goodman. Snowball sampling. *The annals of mathematical statistics*, pages 148–170, 1961.

[Gra16] Martin Grandjean. A social network analysis of twitter: Mapping the digital humanities community. *Cogent arts & humanities*, 3(1):1171458, 2016.

[KLPM10] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. What is twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web*, pages 591–600, 2010.

[MMG+07] Alan Mislove, Massimiliano Marcon, Krishna P Gummadi, Peter Druschel, and Bobby Bhattacharjee. Measurement and analysis of online social networks. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, pages 29–42, 2007.

[MS24] Mariana Macedo and Akrati Saxena. Gender differences in online communication: A case study of soccer. *arXiv preprint arXiv:2403.11051*, 2024.

[MSGL14] Seth A Myers, Aneesh Sharma, Pankaj Gupta, and Jimmy Lin. Information network or social network? the structure of the twitter follow graph. In *Proceedings of the 23rd international conference on world wide web*, pages 493–498, 2014.

[New02] Mark EJ Newman. Assortative mixing in networks. *Physical review letters*, 89(20):208701, 2002.

[New03] Mark EJ Newman. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003.

[New05] Mark EJ Newman. Power laws, pareto distributions and zipf's law. *Contemporary physics*, 46(5):323–351, 2005.

[PBMW99] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford infolab, 1999.

[PSVV01] Romualdo Pastor-Satorras, Alexei Vázquez, and Alessandro Vespignani. Dynamical and correlation properties of the internet. *Physical review letters*, 87(25):258701, 2001.

[QB25] Dorian Quelle and Alexandre Bovet. Bluesky: Network topology, polarization, and algorithmic curation. *PloS one*, 20(2):e0318034, 2025.

[RB03] Erzsébet Ravasz and Albert-László Barabási. Hierarchical organization in complex networks. *Physical review E*, 67(2):026112, 2003.

[SGI18] Akrati Saxena, Ralucca Gera, and SRS Iyengar. Estimating degree rank in complex networks. *Social Network Analysis and Mining*, 8(1):42, 2018.

[SI16] Akrati Saxena and SRS Iyengar. Evolving models for meso-scale structures. In *2016 8th international conference on communication systems and networks (COMSNETS)*, pages 1–8. IEEE, 2016.

[SI20] Akrati Saxena and Sudarshan Iyengar. Centrality measures in complex networks: A survey. *arXiv preprint arXiv:2011.07190*, 2020.

[SIG15] Akrati Saxena, SRS Iyengar, and Yayati Gupta. Understanding spreading patterns on social networks based on network topology. In *Proceedings of the 2015 IEEE/ACM international conference on advances in social networks analysis and mining 2015*, pages 1616–1617, 2015.

[SPV+21] Akrati Saxena, Yulong Pei, Jan Veldsink, Werner van Ipenburg, George Fletcher, and Mykola Pechenizkiy. The banking transactions dataset and its comparative analysis with scale-free networks. In *Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 283–296, 2021.

[SSR22] Akrati Saxena, Pratishtha Saxena, and Harita Reddy. Fake news propagation and mitigation techniques: A survey. *Principles of Social Networking: The New Horizon and Emerging Challenges*, pages 355–386, 2022.

[TPFG18] Shazia Tabassum, Fabiola SF Pereira, Sofia Fernandes, and João Gama. Social network analysis: An overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(5):e1256, 2018.

[WS98] Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small-world'networks. *nature*, 393(6684):440–442, 1998.

[ZGR18] Matteo Zignani, Sabrina Gaito, and Gian Paolo Rossi. Follow the "mastodon": Structure and evolution of a decentralized online social network. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 12, pages 541–550, 2018.