Athens University of Economics and Business

Social Network Analysis, 2024 - 2025

Report On

Analysis of a Bluesky account network.

Author: Vlasios Zacharis (3220263)
Course Instructor: Dr. Demitrios Pournarakis

Contents

List of Figures

1 Introduction

1.1 Context & Motivation

In the digital age, social networks have become an essential medium for communication and information exchange. Understanding how users connect, influence one another, and form communities provides valuable insights into group behavior, information flow, and the emergence of influential voices.Network analysis, a core methodology in data science, allows us to model these interactions as graphs, revealing patterns and structures that are not immediately visible from raw data alone.

Bluesky is a relatively new social platform focused on decentralized principles. Backed by the AT Protocol (Authenticated Transfer Protocol), it aims to give users more control over their data and social interactions. Given its unique positioning in the social media ecosystem and growing user base, Bluesky offers a compelling domain for exploring real-world network properties.

1.2 Objective of Analysis

This report presents a network analysis of the @atproto.com account and its immediate connection graph on Bluesky. The goal is to apply fundamental network analysis concepts—such as measuring degrees, identifying connected components, assessing centralities, examining community structures, and exploring clustering—to better understand how this account's network is organized and where influence or fragmentation might reside.

Key objectives include:

- Quantifying the basic topological properties (number of nodes, edges, connectivity).
- Identifying key players (high-degree or high-centrality nodes) within the network.
- Detecting community structure (clusters of closely linked users) to see if the network naturally segments into subgroups.
- Examining bridges, density, and clustering behavior to see how users cluster around shared interests or influences..

1.3 Choice of Network (Why @atproto.com)

The @atproto.com handle on Bluesky is closely associated with the underlying AT Protocol team, making it a central reference point for official updates and community discussions. Consequently, it garners a sizable and diverse follower base, providing a rich dataset for network analysis. By focusing on a well-connected account:

- We can capture substantial user interaction data (over 100k nodes before filtering).
- We observe varied relationships, as the account's followers and followings likely span developers, early adopters, enthusiasts, and curious onlookers.
- We ensure the final dataset still represents a large, interesting sub-network even after cleaning and filtering—ideal for illustrating key social graph metrics and phenomena.

In summary, @atproto.com is a strategic choice for showcasing the breadth of network analysis concepts required by this coursework. The following sections detail the data collection, cleaning steps, visualization strategies, and metrics computations that lead to our final insights.

2 Data Collection & Preparation

2.1 Tools Used

- **Gephi (Version 0.10.1):** A popular open-source software for network visualization and analysis.
- **BlueSky Gephi Plugin:** Used to interface directly with the Bluesky (AT Protocol) API for retrieving followers and following data of specific user accounts.
- **ForceAtlas2 Layout (in Gephi):** Utilized initially for a quick visual check of the raw imported network and later again after filtering. This force-directed algorithm treats edges like springs, drawing connected nodes closer and pushing unconnected nodes apart, thereby creating an "organic" map of the network.

2.2 Data Import Process

- **Target Account:** @atproto.com (Bluesky handle).
- **Plugin Configuration:**
  - **Followers and Following selected:** This ensures we capture both inbound and outbound connections, yielding a more complete view of the user's immediate network.
  - **Crawl Limit:** Set to 1,000 to cap the breadth/depth of the fetch, preventing an unmanageably large dataset.
  - **N+1 Depth:** Enabled to include the network of those who follow / are followed by the target account (i.e., followers-of-followers), broadening the scope to second-degree connections.

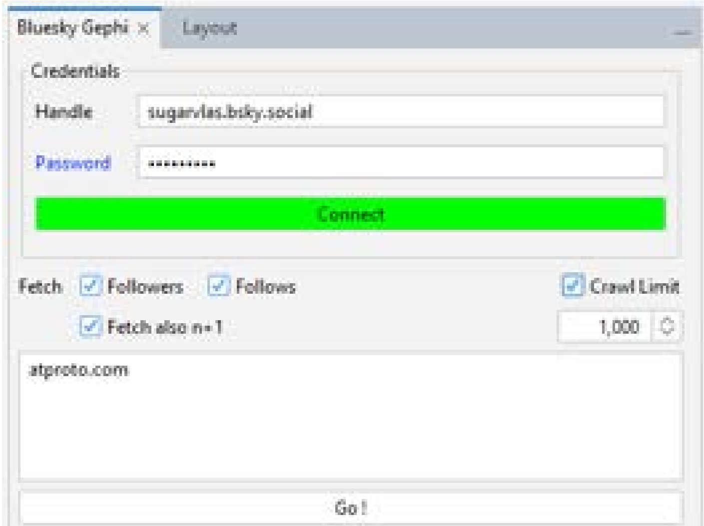


Figure 1: Plugin Configuration

- **Initial Dataset Size:**
  - **Nodes:** 115,025
  - **Edges:** 128,852

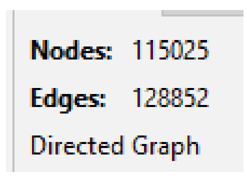


Figure 2: Initial Dataset Size

  These counts reflect raw data before any cleaning or filtering.

- **Preliminary Visual Check:** Applied a ForceAtlas2 layout to verify that data was imported correctly and to get a rough overview. The network was very large, so further refinement steps were necessary.

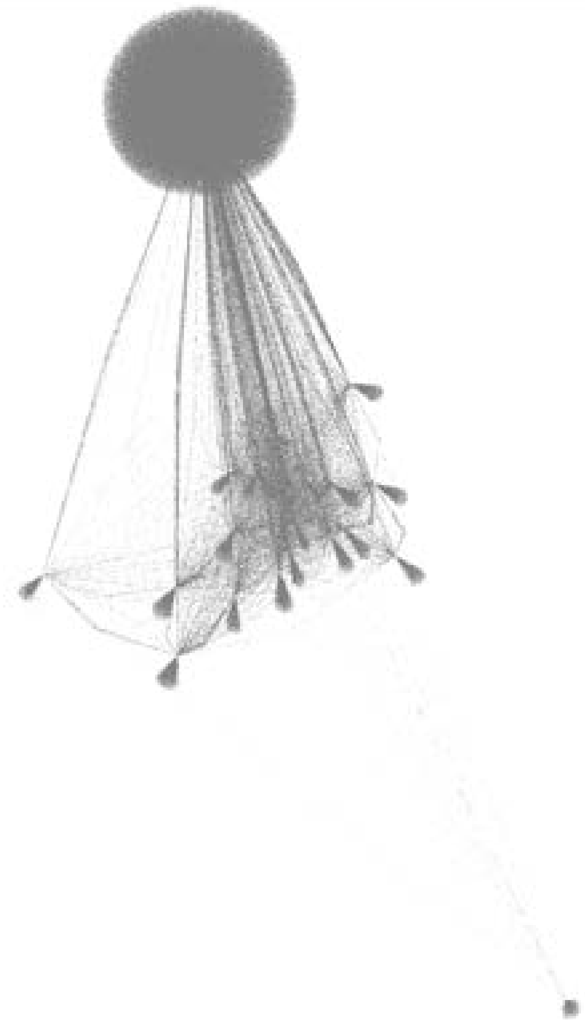


Figure 3: Preliminary Visual Check

2.3 Data Cleaning

- **Duplicate Nodes:**
  - In the Data Laboratory (Nodes table), a check for duplicates revealed 1 duplicated node.
  - Merged the duplicate to maintain unique node entries.

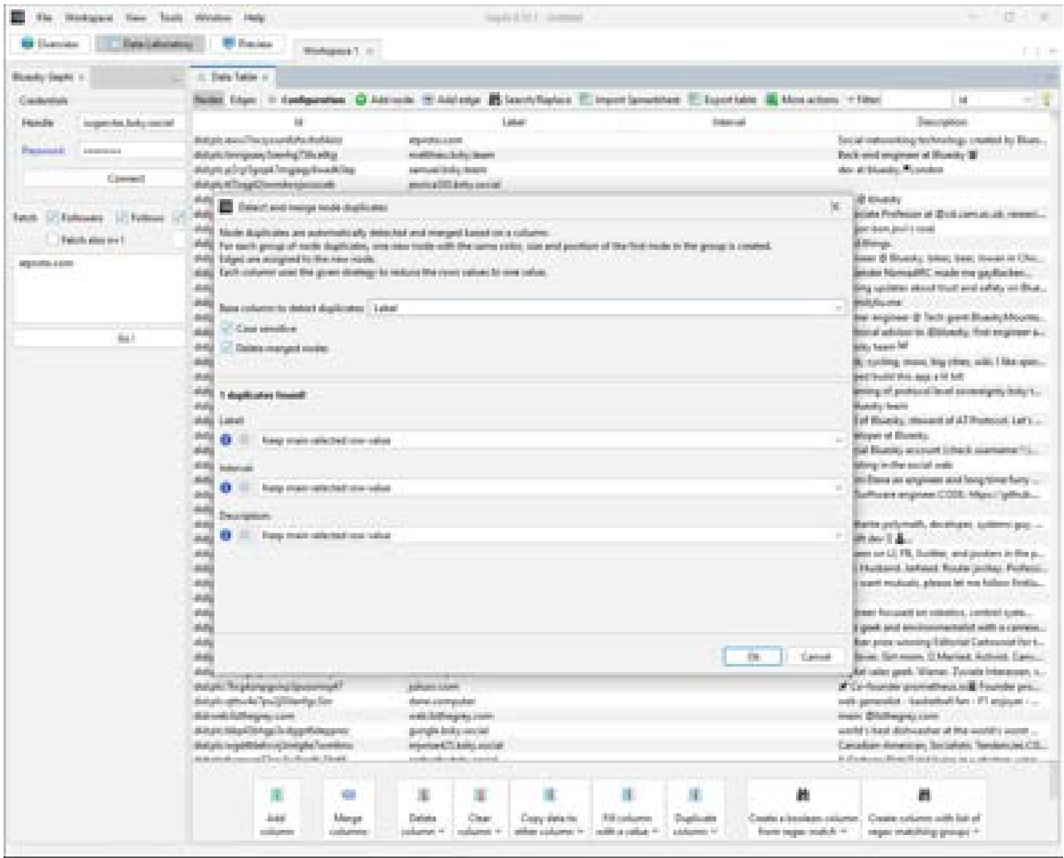


Figure 4: Duplicate Nodes Check

- **Invalid Nodes:**
  - Identified nodes with an id = 0 and a label = "handle.invalid", which typically occur due to incomplete or erroneous API responses.
  - Removed these invalid entries to ensure all nodes refer to real, valid user accounts.

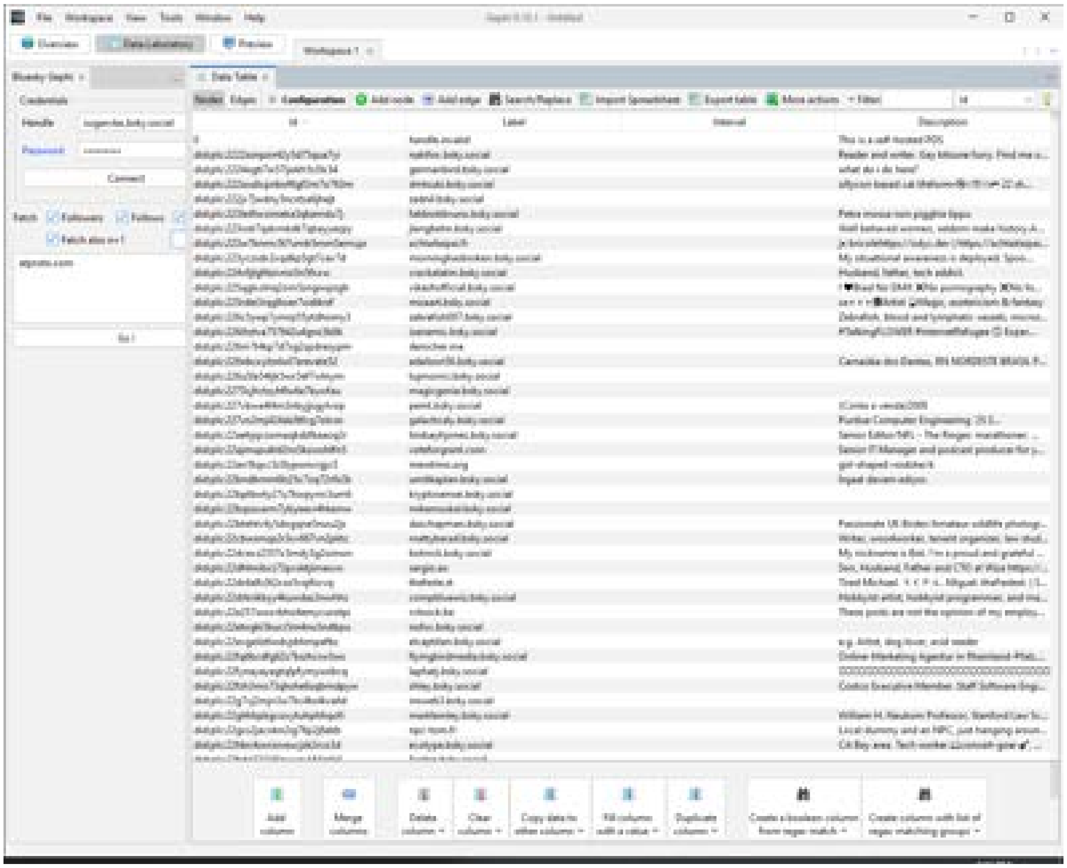


Figure 5: Invalid Nodes Check

- **Post-Cleaning Counts:** After merging the duplicate and deleting invalid nodes, the network contained 114,522 nodes and 128,312 edges.

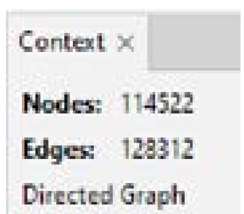


Figure 6: Post-Cleaning Counts

- **Filtering:**
  - **Giant Component Filter:** Isolated the largest connected sub-network to focus on the portion of the graph where most users can reach each other.
  - **Degree Range Filter:** Kept only nodes with degree ≥ 2, removing peripheral or isolated nodes that do not significantly contribute to the analysis (e.g., those with degree 0 or 1).

- **Resulting Dataset:**
  - 6,522 nodes ($\approx 5.69\%$ of the original)
  - 20,312 edges ($\approx 15.83\%$ of the original)

- **Final Dataset:**
  - By focusing on the giant component and higher-degree nodes, we arrive at a more manageable subgraph.
  - This refined dataset remains sufficiently large to reveal meaningful patterns and communities, while limiting noise and purely isolated users.

3 Graphical Representation of the Network

3.1 ForceAtlas2 Configuration

After filtering the dataset down to a more cohesive subgraph (6,522 nodes and 20,312 edges), a second ForceAtlas2 layout was applied in Gephi with adjusted parameters to enhance clarity and reduce node overlap:

- **Scaling:** Adjusted from a value of 10.0 to a value of 40.0, to prevent excessive overlap of high-degree nodes. This stretches the layout, pushing clusters further apart, making it easier to distinguish different communities.

- **Prevent Overlap:** Enabled. to ensure that large, high-degree nodes do not physically overlap smaller ones, improving label and node visibility.

Despite these changes, other ForceAtlas2 settings remained as set by default (e.g., Threads = 15, Gravity = 1.0, etc.).

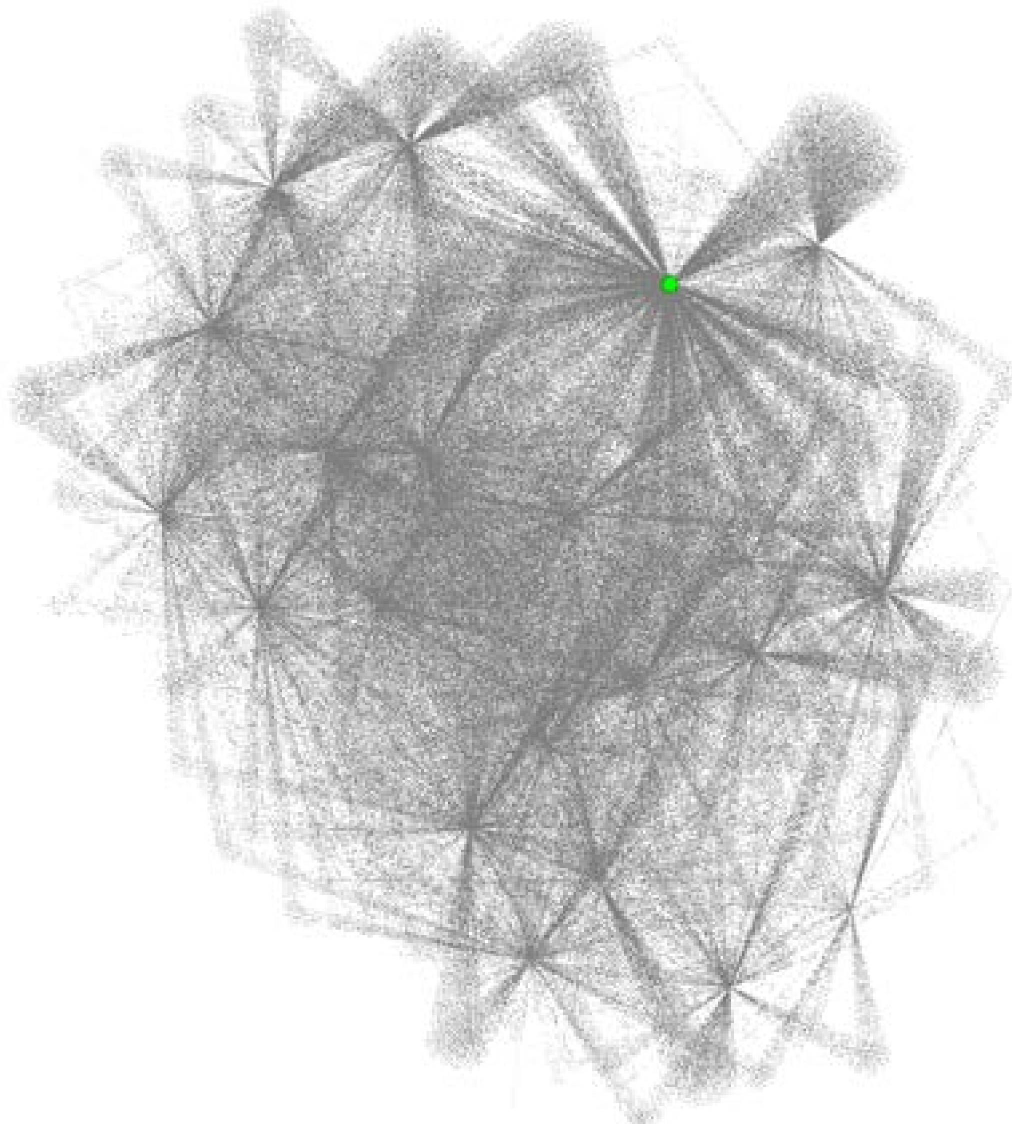


Figure 7: Graph Post Second ForceAtlas2 Layout Application

3.2 Modularity Partition & Color Coding

A Modularity run (Louvain algorithm) identified 7 modularity classes, labeled from 0 through 6.

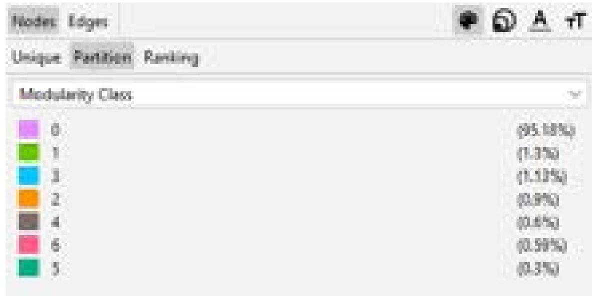


Figure 8: Color Partitioning by Modularity Class

- Class 0 (magenta) – 95.18% of nodes
- Class 1 (orange) – 1.30%
- Class 3 (turquoise) – 1.13%
- Class 2 (light blue) – 0.90%
- Class 4 (green) – 0.60%
- Class 6 (light pink) – 0.59%
- Class 5 (fuchsia) – 0.30%

Hence, the vast majority of users fall within Class 0, while the other six classes combined account for about 5% of the network.

3.3 Node Sizing & Labels

- **Degree-Based Sizing**
  - Nodes are scaled according to their Degree (number of connections).

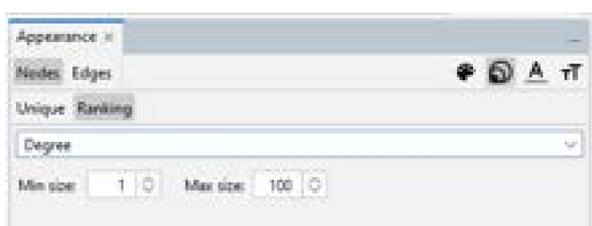


Figure 9: Degree-Based Node Sizing

  - As expected, @atproto.com is among the largest nodes (the largest one, to be precise) due to its extensive follow/follower network.

- **Labels**

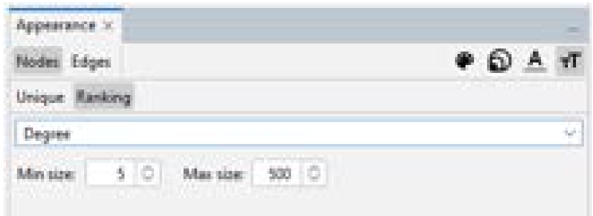


Figure 10: Degree-Based Label Sizing

  - Turned on in the Preview settings and sized proportionally to node degree.
  - Only the most connected nodes (in each class) have labels large enough to read clearly, preventing clutter in a dense graph.

3.4 Observations from the Visual Graph

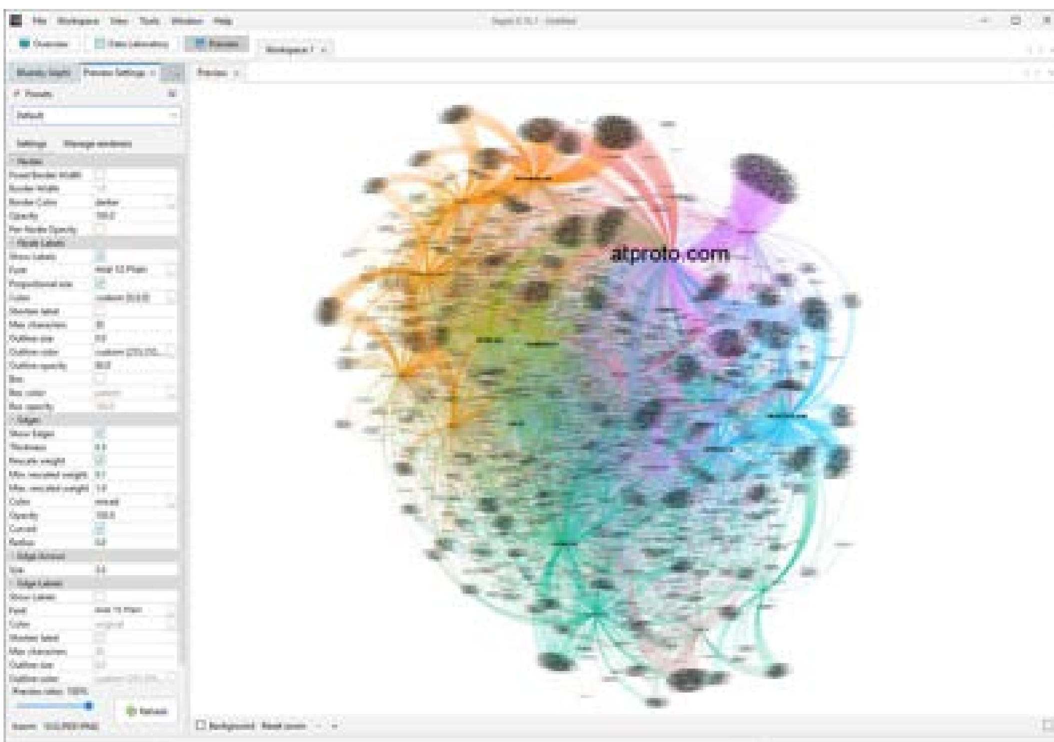


Figure 11: Graph Post Vizualization Adjustments

- **One Predominant Cluster (Class 0)**
  With about 95% of the total nodes, the network effectively forms a single "mega-community." The algorithm likely detects only minimal structural differences between most users, grouping them en masse.

- **Minor Sub-Communities**
  The orange, turquoise, light blue, green, light pink, and fuchsia clusters each represent less than 2% of the network. These smaller pockets may be niche interest groups or specialized follower sets branching off from the main network.

- **High-Degree Hubs**
  Within the dominant magenta mass (Class 0), certain large-degree nodes stand out—these may serve as local bridges or influential accounts besides @atproto.com. Further centrality measures (betweenness, eigenvector, PageRank) can reveal if they truly act as brokers or key influencers.

Overall, this latest analysis underscores a highly unified network around @atproto.com, with only a few small outlying communities. Although the minor clusters are relatively small in node count, each could have its own internal structure, warranting closer investigation for specialized interests or unique interaction patterns.

4 Basic Topological Properties

After consolidating the data for the @atproto.com network—removing duplicates, invalid nodes, and low-degree isolates—this analysis focuses on a subgraph of 6,522 nodes and 20,312 edges. The network is treated as directed, reflecting the follow/following relationships on Bluesky.

4.1 Node and Edge Counts

After filtering the dataset down to a more cohesive subgraph (6,522 nodes and 20,312 edges), a second ForceAtlas2 layout was applied in Gephi with adjusted parameters to enhance clarity and reduce node overlap:

- **Nodes:** 6,522
  Each node represents an individual Bluesky user that either follows or is followed (directly or in second-degree connections) by @atproto.com.

- **Edges:** 20,312
  A directed edge from User A to User B indicates that A follows B. No weighting attributes were provided, so all edges have equal weight.

These counts demonstrate a moderate-sized social sub-network—large enough to observe interesting patterns (like community structure and influential hubs), yet still manageable for visualization and analysis in Gephi.

4.2 Network Diameter

- **Diameter (Directed):** 4
  Computed via Gephi's Network Diameter algorithm, the diameter is the longest shortest path within the directed graph. A value of 4 implies that, in the worst case, you can reach any node from any other node by traversing at most four follow-steps. This relatively small diameter is typical of many social networks, suggesting a "small-world" effect: most users are only a few hops away from each other.

Figure 12: Network Diameter

4.3 Average Path Length

- **Average Path Length:** 2.8374
  Computed via Gephi's Average Path Length (directed) algorithm, this metric

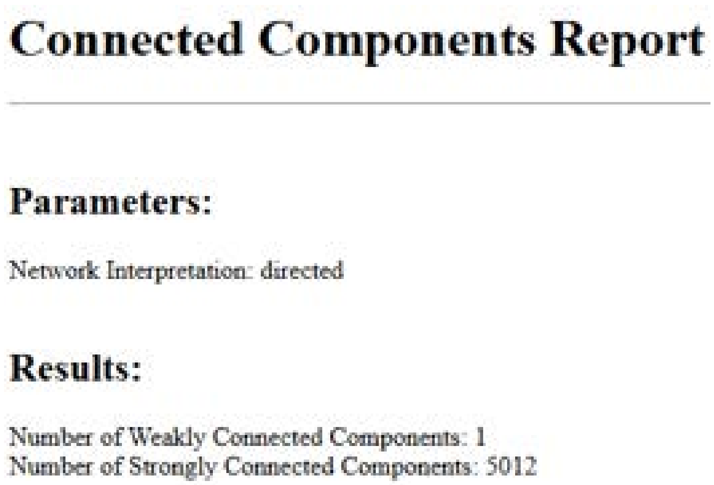
indicates that on average it takes just under three steps (2.84) to travel from one user to another in the network. Together with the small diameter, this finding further underscores the high interconnectedness among these users—information can potentially reach distant parts of the network with only a few link traversals.

Figure 13: Average Path Length

5 Component Measures

Because this analysis treats the @atproto.com network as a directed graph, Gephi (or any network analysis tool) can distinguish between weakly and strongly connected components:

- **Weakly Connected Component (WCC):** A set of nodes that remain connected if we ignore edge direction.
- **Strongly Connected Component (SCC):** A set of nodes where every node can reach every other node following edge directions.


Connected Components Report

Parameters:

Network Interpretation: directed

Results:

Number of Weakly Connected Components: 1
Number of Strongly Connected Components: 5012

Figure 14: Connected Components Report

5.1 Weakly Connected Components

- **Number of WCCs:** 1

This indicates that, if we disregard edge directions (treating the graph as undirected), all 6,522 nodes form a **single "giant" component**. Practically, it means there is a path—ignoring direction—between any two users in the network.

Size Distribution

- Since there is only one weakly connected component, its size is effectively 6,522 (the entire subgraph).
- The attached "Size Distribution" chart reflects this with a single data point at 6,522 nodes.

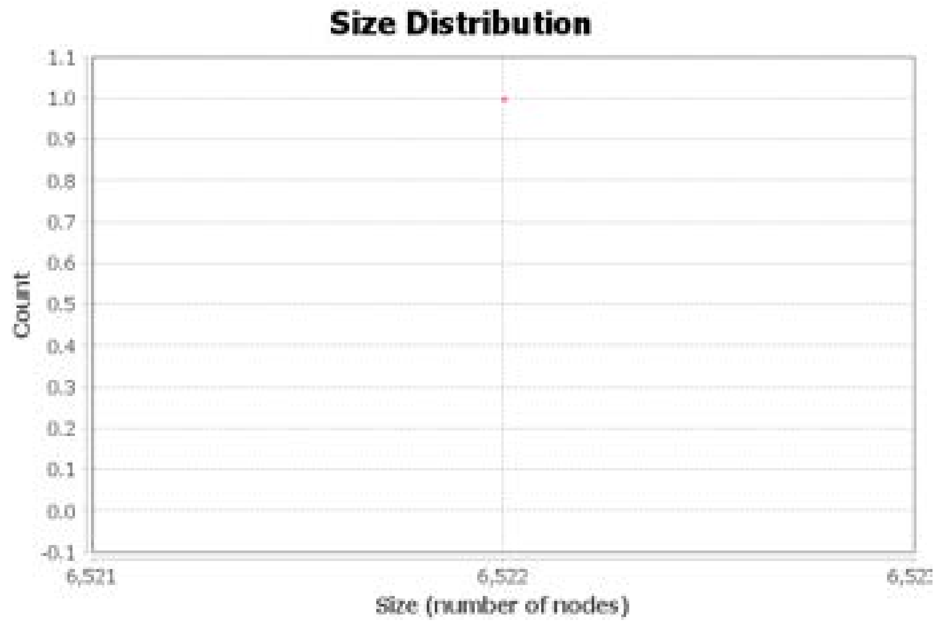


Figure 15: Connected Components Size Distribution

5.2 Strongly Connected Components

- **Number of SCCs:** 5,012

When direction is considered, the network breaks into 5,012 smaller strongly connected components. Many are likely single nodes or small clusters where each user can reach the others via directed paths. This large number suggests that most users do not form mutual loops with one another, which is common in "broadcast-style" social follow graphs (e.g., many people following a prominent account, but fewer reciprocal connections).

Giant Strongly Connected Component?
While there is a single giant WCC, the strongly connected view is more fragmented. In typical large social networks, the largest SCC may still be modest in size compared to the entire network—reflecting that user relationships are often asymmetric. Further analysis (e.g., SCC size distribution) could identify whether one SCC dominates or most are small "islands."

Overall, these component measures illustrate a high-level network structure:

- Ignoring direction, nearly everyone is in a unified cluster around @atproto.com.
- Considering direction, the network fractures into many smaller SCCs, highlighting the predominantly one-way nature of follows in Bluesky.

6 Degree Measures

In a directed network, each node has two main degree values:

- **In-Degree:** Number of incoming edges (followers).
- **Out-Degree** Number of outgoing edges (followings).

When referencing a node's "Degree" without specification in Gephi, it means Total Degree = In-Degree + Out-Degree. Below are the key observations from the @at-proto.com sub-network's degree data.

6.1 Average Degree

- **Average Degree:** 3.114

This indicates that, on average, each node connects to about three others (either following them, being followed by them, or both). While most social networks have relatively low average degrees (mostly 2), a few high-degree "hub" nodes tend to skew many of the related metrics.

Figure 16: Average Degree

6.2 Degree Distributions

The degree distributions in a social network typically show a long-tail shape, where most users have relatively few connections, and a small fraction of nodes exhibit extremely high connectivity. In the attached charts:

- **Degree Distribution** (all-degree view):
  - Most nodes cluster at the low-degree end (under 10).
  - A few nodes extend the tail past 500, 1,000, and one beyond 3,000, indicating hubs or influencers.

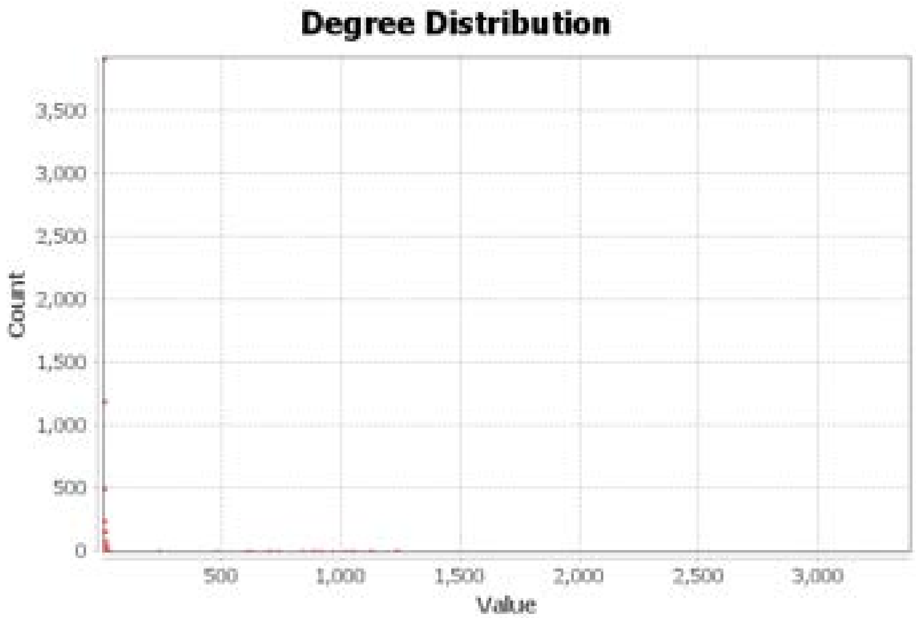


Figure 17: Degree Distribution

- **Out-Degree Distribution:**
  - Typically smaller than in-degree for large hubs, as fewer users follow thousands of accounts.
  - Most nodes cluster at the low-degree end (under 4).
  - A handful of nodes have out-degrees in the hundreds.

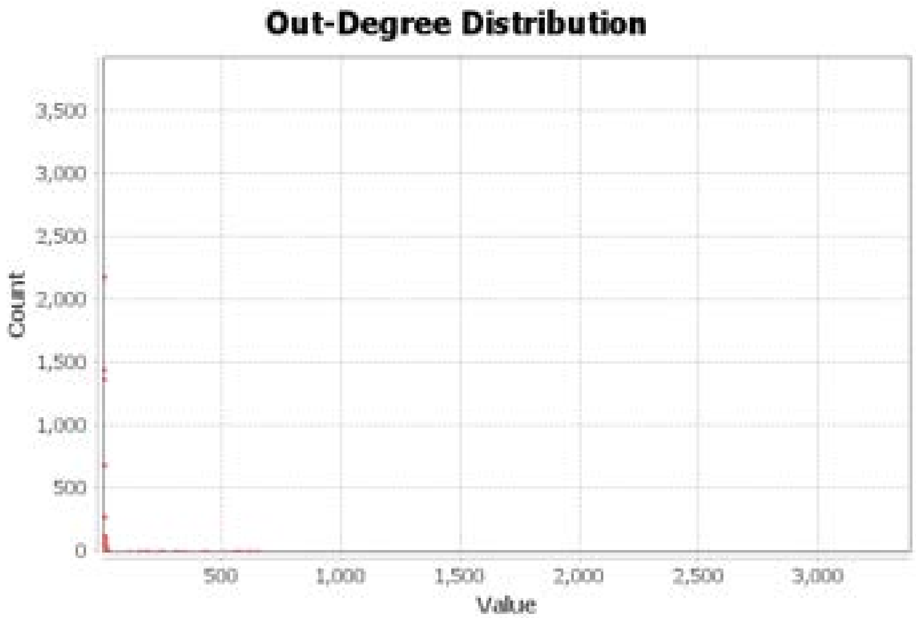


Figure 18: Out-Degree Distribution

- **In-Degree Distribution:**
  Nodes like @atproto.com exhibit a very high in-degree, reflecting their popularity. The chart shows a strong spike at the low end and a thin tail stretching to over 3,000, implying a handful of highly followed accounts.

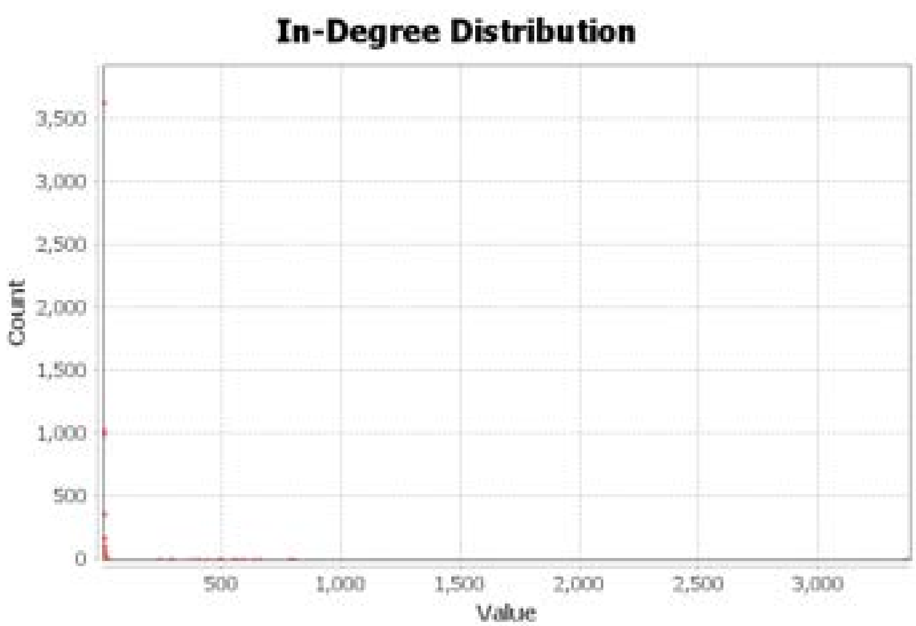


Figure 19: In-Degree Distribution

6.3 Maximum Degrees

- **Max Degree:** 3,383 for @atproto.com
  This user stands out as having the highest total (in + out) connections.

Data Table | Nodes | Edges | Configuration | Add node | Add edge | Search/Rep

Id	Label	Interval	Description	Degree
did:plc:ewvi7n...	atproto.com		Social networking t...	3383
did:plc:44ybar...	bnewbold.net		dweb, cycling, sno...	1233
did:plc:yk4dd2...	dholms.xyz		dreaming of protoc...	1226
did:plc:3jpt2m...	esb.lol		Engineer @ Bluesky...	1124
did:plc:vjug55k...	emilyliu.me		emilyliu.me	1114
did:plc:fpruhu...	danabra.mov		like jon bon jovi's r...	1045
did:plc:p2cp5g...	samuel.bsky.te...		dev at bluesky Lo...	1035
did:plc:fgsn4gf...	anshnanda.com		Helped build this a...	1011
did:plc:linrigsa...	matthieu.bsky....		Back-end enginee...	958
did:plc:vpkhqo...	why.bsky.team		Technical advisor t...	908
did:plc:upo6iq...	foysal.it		Build things.	905
did:plc:l3rouwl...	divy.zone		bluesky team	882
did:plc:tpg43q...	jacob.gold		Former engineer @...	876
did:plc:oisofpd...	hailey.at		eng @ bluesky	834
did:plc:tl7zqgil...	jessica200.bsky...			732
did:plc:ragtjsm...	pfrazee.com		Developer at Bluesky.	699
did:plc:qjeavhl...	rose.bsky.team		Bluesky team	689
did:plc:q6gjna...	jaz.bsky.social		JazGender Nomad...	619
did:plc:oky5cz...	jay.bsky.team		CEO of Bluesky, st...	600
did:plc:vzmlifz...	martin.kleppm...		Associate Professor...	473
did:plc:eon2iu...	safety.bsky.app		Sharing updates ab...	234
did:plc:tykz5c7...	5iii.bsky.social		to be or not to be.	20
did:plc:z72i7hd...	bsky.app		official Bluesky acc...	19
did:plc:h7ta4xs...	paleobyleo.bsk...		Dutch Food writ...	18
did:plc:tggluyd...	mdtalam.bsky....		PhD student @De...	18
did:plc:a3yds6...	kikidollms.bsk...		#BlueCrew #True...	17
did:plc:admfo...	jmgstudio.bsk...		En Antequera hay u...	17
did:plc:np5kajl...	kimmysterr.bs...		Wife to one of New ...	17
did:plc:2ibfj4kf...	amazonstyleq...		" Welcome to @...	17
did:plc:xl5yc4v...	silvrrolls.bsky.s...		Retired Automotive ...	17
did:plc:x6fdazb...	lv-dragons.bsk...		LV Dragons is an ap...	16

Figure 20: Max Degree

- **Max Out-Degree:** 647 for @danabrav.mov.

This indicates the highest number of accounts followed by any single user.

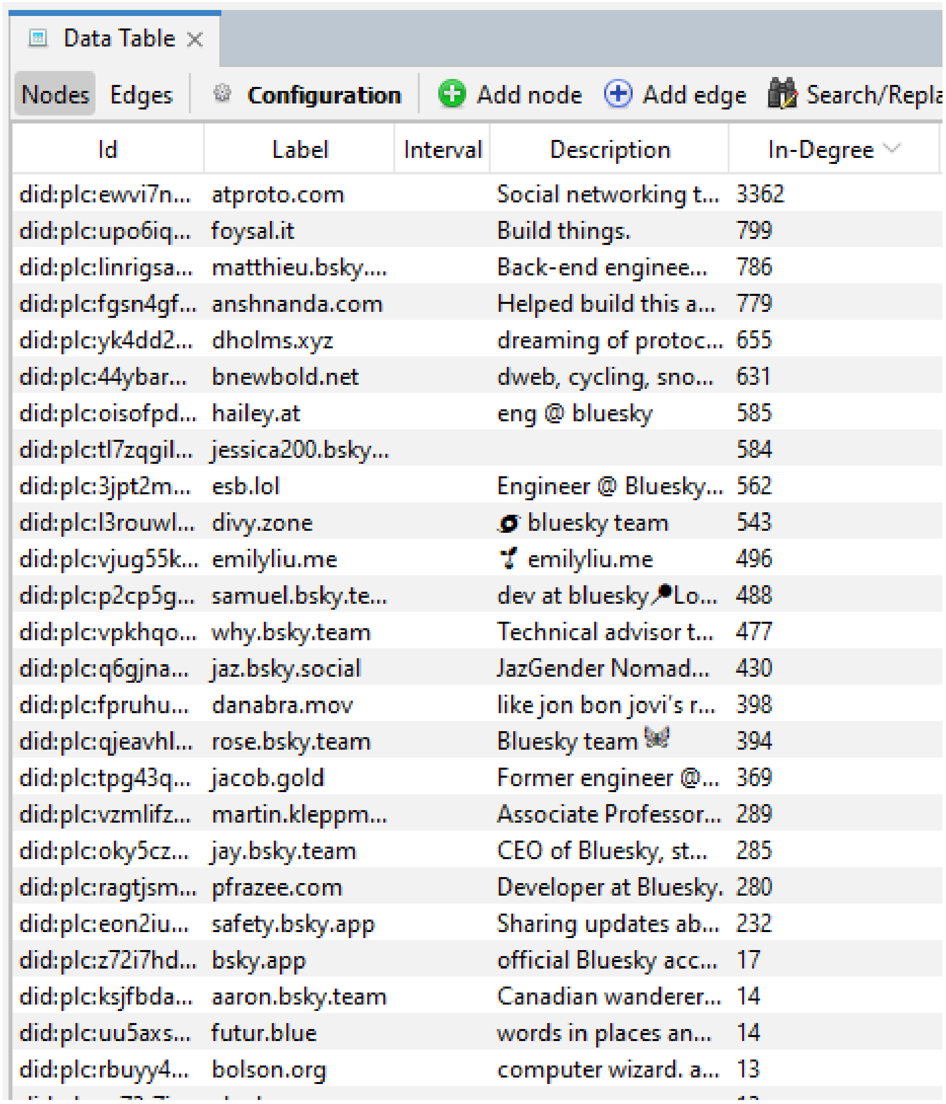
Data Table ×

Nodes Edges Configuration Add node Add edge Search/Replace

Id	Label	Interval	Description	Out-Degree
did:plc:fpruhu...	danabra.mov		like jon bon jovi's r...	647
did:plc:vjug55k...	emilyliu.me		emilyliu.me	618
did:plc:44ybar...	bnewbold.net		dweb, cycling, sno...	602
did:plc:yk4dd2...	dholms.xyz		dreaming of protoc...	571
did:plc:3jpt2m...	esb.lol		Engineer @ Bluesky...	562
did:plc:p2cp5g...	samuel.bsky.te...		dev at bluesky Lo...	547
did:plc:tpg43q...	jacob.gold		Former engineer @...	507
did:plc:vpkhqo...	why.bsky.team		Technical advisor t...	431
did:plc:ragtjsm...	pfrazee.com		Developer at Bluesky.	419
did:plc:l3rouwl...	divy.zone		bluesky team	339
did:plc:oky5cz...	jay.bsky.team		CEO of Bluesky, st...	315
did:plc:qjeavhl...	rose.bsky.team		Bluesky team	295
did:plc:oisofpd...	hailey.at		eng @ bluesky	249
did:plc:fgsn4gf...	anshnanda.com		Helped build this a...	232
did:plc:q6gjna...	jaz.bsky.social		JazGender Nomad...	189
did:plc:vzmlifz...	martin.kleppm...		Associate Professor...	184
did:plc:linrigsa...	matthieu.bsky....		Back-end enginee...	172
did:plc:tl7zqgil...	jessica200.bsky...			148
did:plc:upo6iq...	foysal.it		Build things.	106
did:plc:ewvi7n...	atproto.com		Social networking t...	21
did:plc:tykz5c7...	5iii.bsky.social		to be or not to be.	20
did:plc:h7ta4xs...	paleobyleo.bsk...		Dutch Food writ...	18
did:plc:tggluyd...	mdtalam.bsky....		PhD student @De...	18
did:plc:a3yds6...	kikidollms.bsk...		#BlueCrew #True...	17
did:plc:admfo...	jmgstudio.bsk...		En Antequera hay u...	17
did:plc:np5kajl...	kimmysterr.bs...		Wife to one of New ...	17
did:plc:2ibfj4kf...	amazonstyleq...		" Welcome to @...	17
did:plc:xl5yc4v...	silvrrolls.bsky.s...		Retired Automotive ...	17
did:plc:x6fdazb...	lv-dragons.bsk...		LV Dragons is an ap...	16

Figure 21: Max Out-Degree

- **Max In-Degree:** 3,362 for @atproto.com.
Reflects that @atproto.com is the most-followed node in the dataset (the primary "hub" of this network).



Data Table

Nodes | Edges | Configuration | Add node | Add edge | Search/Repla

Id	Label	Interval	Description	In-Degree
did:plc:ewvi7n...	atproto.com		Social networking t...	3362
did:plc:upo6iq...	foysal.it		Build things.	799
did:plc:linrigsa...	matthieu.bsky....		Back-end enginee...	786
did:plc:fgsn4gf...	anshnanda.com		Helped build this a...	779
did:plc:yk4dd2...	dholms.xyz		dreaming of protoc...	655
did:plc:44ybar...	bnewbold.net		dweb, cycling, sno...	631
did:plc:oisofpd...	hailey.at		eng @ bluesky	585
did:plc:tl7zqgil...	jessica200.bsky...			584
did:plc:3jpt2m...	esb.lol		Engineer @ Bluesky...	562
did:plc:l3rouwl...	divy.zone		bluesky team	543
did:plc:vjug55k...	emilyliu.me		emilyliu.me	496
did:plc:p2cp5g...	samuel.bsky.te...		dev at bluesky Lo...	488
did:plc:vpkhqo...	why.bsky.team		Technical advisor t...	477
did:plc:q6gjna...	jaz.bsky.social		JazGender Nomad...	430
did:plc:fpruhu...	danabra.mov		like jon bon jovi's r...	398
did:plc:qjeavhl...	rose.bsky.team		Bluesky team	394
did:plc:tpg43q...	jacob.gold		Former engineer @...	369
did:plc:vzmlifz...	martin.kleppm...		Associate Professor...	289
did:plc:oky5cz...	jay.bsky.team		CEO of Bluesky, st...	285
did:plc:ragtjsm...	pfrazee.com		Developer at Bluesky.	280
did:plc:eon2iu...	safety.bsky.app		Sharing updates ab...	232
did:plc:z72i7hd...	bsky.app		official Bluesky acc...	17
did:plc:ksjfbda...	aaron.bsky.team		Canadian wanderer...	14
did:plc:uu5axs...	futur.blue		words in places an...	14
did:plc:rbuyy4...	bolson.org		computer wizard. a...	13

Figure 22: Max In-Degree

These extreme values underscore the uneven or skewed nature of social follow networks: while most participants have only a few connections, a tiny subset accumulates a very large following.

Overall, these degree measures and distributions confirm that @atproto.com's follower graph aligns with classic "hub-and-spoke" dynamics found in many online networks: a majority of nodes have low degree, and a small minority of highly connected accounts (hubs) profoundly influence the network's structure and information flow.

7 Centrality Measures

Centrality metrics identify influential nodes or those with strategic positions in the network. In a directed social graph, these measures help reveal which users have broad reach, which act as bridges, and which are closely connected to other key nodes.

7.1 Degree Centrality

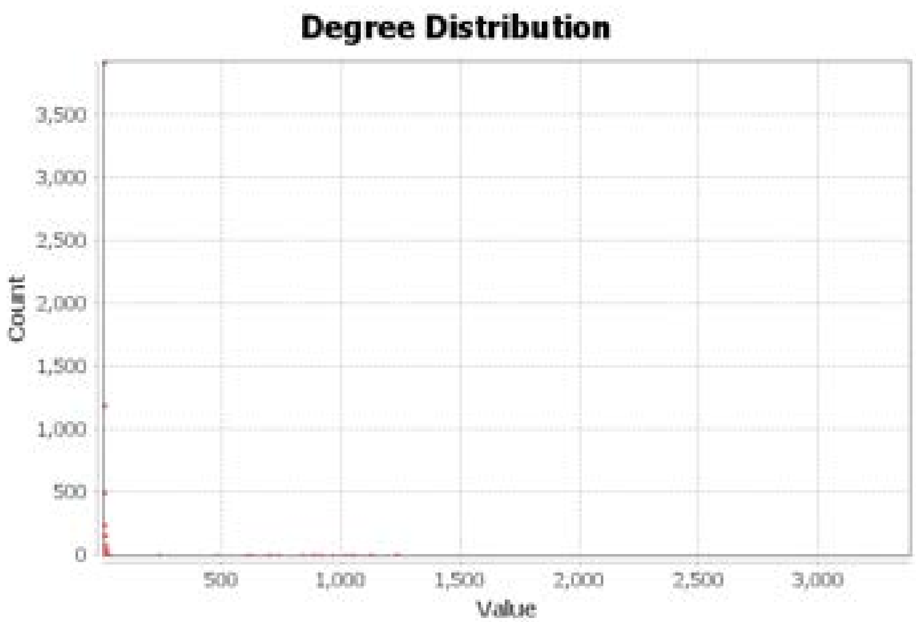


Figure 23: Degree Distribution

- **Interpretation:** Degree centrality measures the number of direct connections (edges) a node has in the network.
- **Distribution:** The attached Degree Distribution histogram exhibits a skewed (long-tail) shape, characteristic of many social networks. Most nodes have relatively few connections, whereas a small fraction exhibits extremely high degree values—signifying "hub" or "influencer" nodes. Overall, Degree Centrality underscores the hub-and-spoke dynamics in the @atproto.com network, with several high-degree accounts dominating the follow landscape and acting as major entry points for information flow.

7.2 Betweenness Centrality

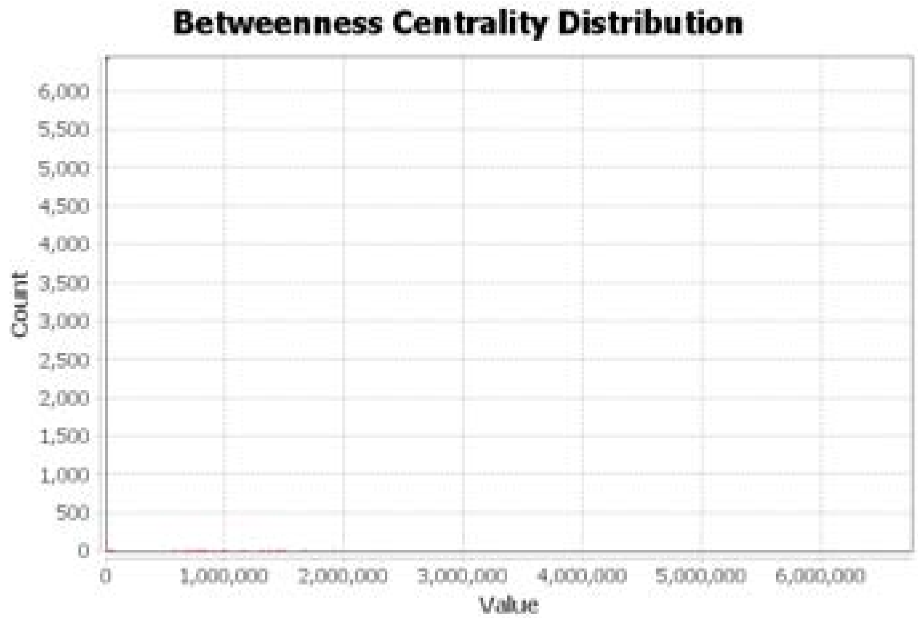


Figure 24: Betweenness Centrality Distribution

- **Interpretation:** Betweenness measures how often a node lies on shortest paths between other pairs of nodes. High betweenness nodes are effective "bridges" for information flow.

- **Distribution:** The attached histogram reveals a long-tail distribution, with most nodes near zero and a few nodes exhibiting very high betweenness values (up to the million range). These high-betweenness accounts act as brokers connecting different parts of the graph.

7.3 Closeness Centrality

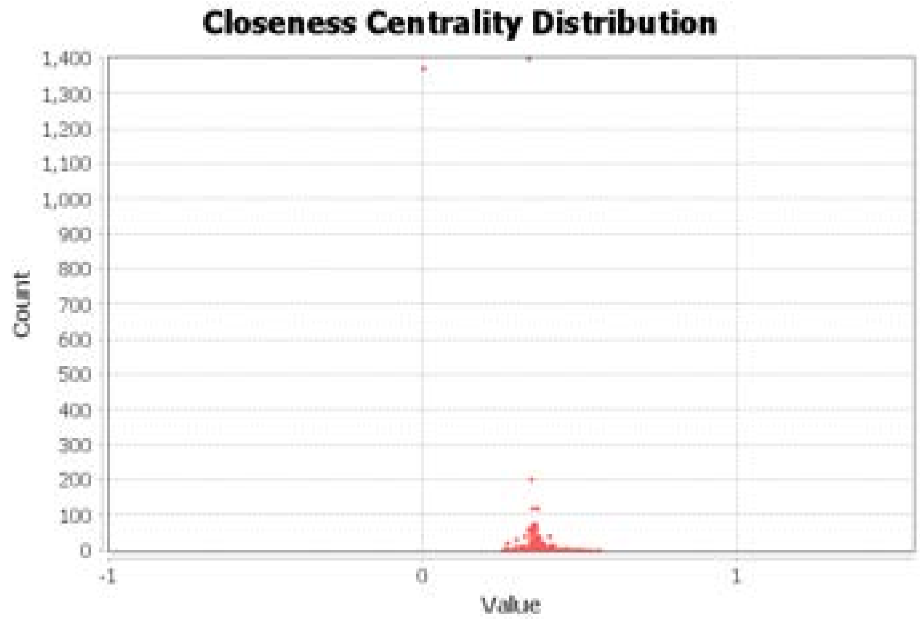


Figure 25: Closeness Centrality Distribution

- **Interpretation:** Closeness indicates how quickly a node can reach others via directed paths—higher closeness means shorter average distance to all other nodes.
- **Distribution:** As the chart shows, closeness centralities cluster toward low values, with a few outliers near zero or slightly higher. The graph's diameter of 4 and average path length of 2.84 likely keep many closeness scores clustered in a narrow range, as most nodes can be reached in just a few steps.

7.4 Eigenvector Centrality

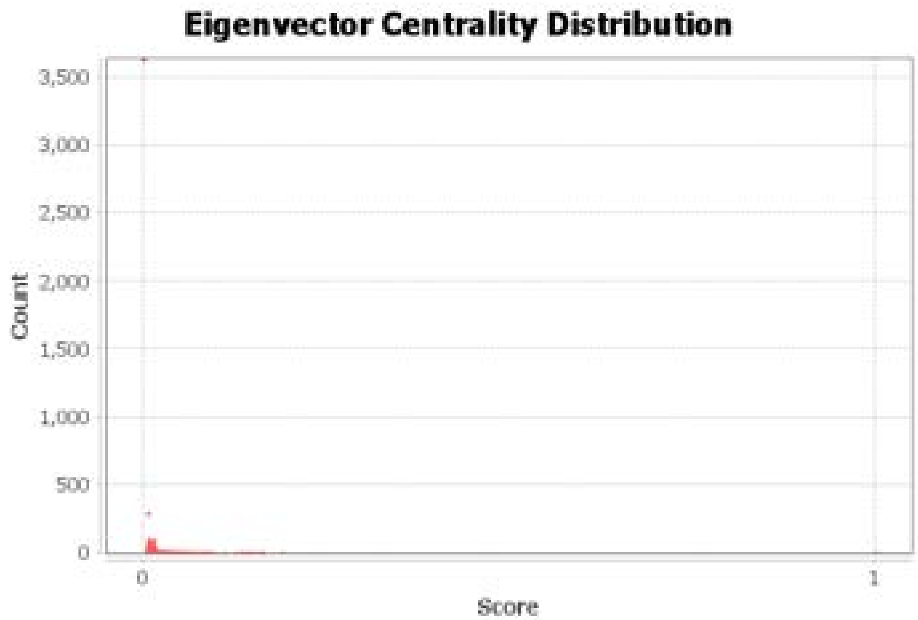


Figure 26: Eigenvector Centrality Distribution

- **Parameters:**
  - **Network Interpretation:** Directed
  - **Iterations:** 1,000
  - **Sum Change:** 0.0021582671788027237 (convergence criterion)
- **Interpretation:** Eigenvector centrality elevates nodes that are connected to other highly connected nodes. In a social network, it highlights "influential" hubs who themselves link to other powerful hubs.
- **Distribution:** The histogram shows the majority of nodes having near-zero scores, with a small fraction taking moderate to higher values. High eigenvector centrality often overlaps with high in-degree, but it more specifically emphasizes connections to other key nodes.

Overall, these centrality measures reveal a star-like structure centered on @atproto.com and a handful of other influential accounts. High betweenness nodes facilitate inter-cluster connections, high-degree nodes form "broadcast" hubs, and strong eigenvector scores reflect alignment with other powerful nodes. This combination highlights a mixed network topology with both highly centralized hubs and smaller bridging accounts that knit the network's sub-communities together

8 Clustering Effects in the Network

In social networks, clustering refers to the tendency of a node's neighbors (followers/-following) to also connect among themselves, forming small triangles (triads). A high clustering coefficient suggests users within the same neighborhood frequently interlink, indicative of tightly knit subgroups.

8.1 Average Clustering Coefficient

- **Value (Directed in Gephi):** 0.754 This metric, computed within Gephi, indicates that on average, 75.4% of a node's immediate neighborhood connections are themselves interconnected. This is a relatively high average clustering coefficient (ACC), especially for a directed network. In general, an ACC above 0.5 implies that many local neighborhoods exhibit a dense interconnection.

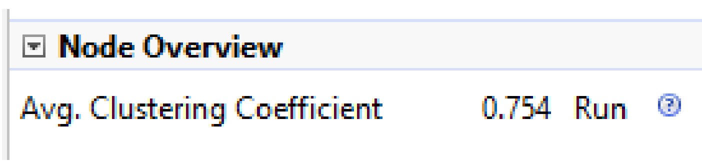


Figure 27: Average Clustering Coefficient

8.2 Number of Directed Triangles

- **Exact Directed 3-Cycles:** 4,527
  - Although Gephi's default directed clustering coefficient statistic doesn't always report the total triangle count explicitly, a high ACC implies numerous triads within local neighborhoods.
  - To extract the exact number of directed triangles, the edges table was extracted from Gephi's Data Laboratory. Then using the following Python script and NetworkX on the exported edges.csv, 4,527 fully directed triangles were identified. That is, triplets (u,v,w) where edges u→v, v→w, and w→u all exist.
  - This direct count reaffirms that, even in a directed setting, the @atproto.com network has thousands of small, closed loops—further evidence of tight local clusters.

Listing 1: Python script for detecting directed triangles

```
import pandas as pd
import networkx as nx

```

```
def count_directed_triangles(G):
count = 0

# Iterate over each node u
for u in G.nodes():
    # Get all successors (out-neighbors) of u
    successors_u = set(G.successors(u))

    # For each successor v of u
    for v in successors_u:
        # Get successors of v
        successors_v = set(G.successors(v))

        # Find common successors w (i.e., u->w and v->w exist)
        common_successors = successors_u.intersection(successors_v)

        # Check if w->u exists to complete u->v->w->u
        for w in common_successors:
            if w != u and w != v:
                if G.has_edge(w, u):
                    count += 1

return count

if __name__ == "__main__":
    # 1. Load the edges from a CSV file exported by Gephi
    #    Make sure your CSV has at least "Source" and
    #    "Target" columns.
    csv_file = "edges.csv"
    df = pd.read_csv(csv_file)

    # 2. Create a directed graph and add edges
    G = nx.DiGraph()
    for _, row in df.iterrows():
        G.add_edge(row["Source"], row["Target"])

    # 3. Count directed 3-cycles (triangles)
    num_triangles = count_directed_triangles(G)
    print(f"Number of directed triangles (3-cycles) in the graph: {num_triangles}")
```

8.3 Clustering Coefficient Distribution & Triadic Closure

As the attached chart suggests (with no visible data points plotted), Gephi may not have produced a detailed histogram for each node's coefficient under the directed interpretation, or the values are heavily concentrated in a certain range. Regardless, the combined findings (high ACC + large number of 3-cycles) indicate that nodes often share neighbors who also link back to each other. In other words, the triadic closure phenomenon is robust: if two users follow the same influencer, there's a strong chance they form (or are part of) a reciprocal loop, facilitating quick information flow within that sub-neighborhood.

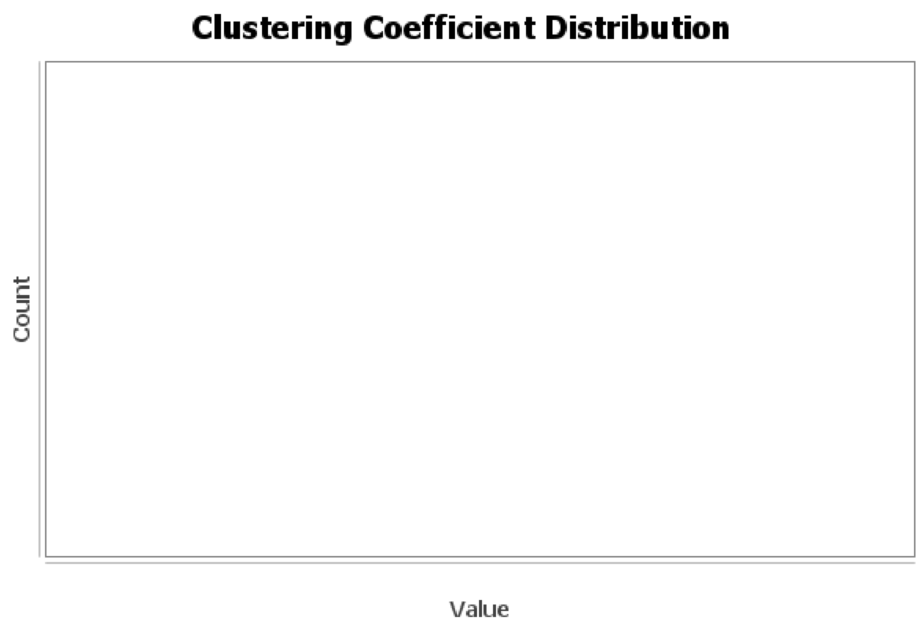


Figure 28: Clustering Coefficient Distribution

In summary, the high average clustering coefficient and thousands of directed 3-cycles reveal that, although the network is dominated by a few star-like hubs, those local sub-networks also exhibit significant internal linkage, illustrating a strong triadic closure effect around the @atproto.com account.

9 Bridges and Local Bridges

Bridges—sometimes referred to as "cut edges"—are edges whose removal increases the number of connected components. In a directed network, this concept typically translates into weak or strong bridges, depending on whether we look at the network as undirected (weak connectivity) or strictly respect edge directions (strong connectivity). A local bridge, meanwhile, is an edge (u,v) where u and v share no other common neighbors, meaning it provides a unique local connection between two neighborhoods. For this section, the network was addressed as undirected, because in this context, the direction of the edges doesn't play a big role. Below is attached the Python script utilized to calculate the required values:

Listing 2: Python script for detecting local and weak bridges

```
import pandas as pd
import networkx as nx

def count_local_bridges(UG):
    local_bridges_list = []
    for u, v in UG.edges():
        neighbors_u = set(UG[u]) - {v}
        neighbors_v = set(UG[v]) - {u}
        # If there's no overlap in the remaining neighbors,
        # it's a local bridge
        if neighbors_u.isdisjoint(neighbors_v):
            local_bridges_list.append((u, v))
    return local_bridges_list

def main():
    # 1. Read the edges from a CSV file exported by Gephi
    csv_file = "edges.csv"
    df = pd.read_csv(csv_file)

    # 2. Create a directed graph
    DG = nx.DiGraph()
    for _, row in df.iterrows():
        DG.add_edge(row["Source"], row["Target"])

    # 3. Create an undirected copy of that graph
    UG = DG.to_undirected()

    # 4. Find all bridges in the undirected graph
    #    NetworkX provides an efficient built-in method
    #    to detect bridges
    bridges = list(nx.bridges(UG))
    num_bridges = len(bridges)
```

```

        # 5. Find all local bridges
        local_bridges_list = count_local_bridges(UG)
        num_local_bridges = len(local_bridges_list)

        print(f"Number of weak bridges (undirected): {num_bridges}")
        print("Examples of bridges:",
        bridges[:10], "..." if num_bridges > 10 else "")

        print(f"Number of local bridges: {num_local_bridges}")
        print("Examples of local bridges:",
        local_bridges_list[:10], "..." if num_local_bridges > 10 else "")

    if __name__ == "__main__":
        main()
```

9.1 Bridges (Undirected)

After exporting the @atproto.com follow graph from Gephi and reading the edges into Python, the script created an undirected view (G.to_undirected()) of the directed graph to search for weak bridges. The results:

- **Number of Bridges:** 7
- **Sample Bridges:**

$$('did:plc:vpkhqolt662uhesyj6nxm7ys','did:plc:hvzwk2da7mafuhef6agdujmo'),$$

$$('did:plc:qjeavhlw222ppsre4rscd3n2','did:plc:353gzbhcjbznfxegvayffqek'),$$

$$('did:plc:44ybard66vv44zksje25o7dz','did:plc:loyxs64im5clm7cuomkjmic4'),$$

$$\ldots$$

Removing any of these 7 edges would break the weakly connected network into additional components, indicating these edges are critical "links" in the broader follower graph. Though relatively few in number, their structural importance is high.

9.2 Local Bridges

Local bridges connect two nodes whose immediate neighborhoods do not otherwise overlap. If such an edge disappears, it doesn't necessarily split the entire network, but it locally isolates those neighborhoods:

- **Number of Local Bridges:** 939
- **Sample Local Bridges:**

$$('did:plc:p2cp5gopk7mgjegy6wadk3ep', 'did:plc:yqfmy2p54vqgekrcz5zzykhl'),$$

$$('did:plc:p2cp5gopk7mgjegy6wadk3ep', 'did:plc:3s5wtzqrvrsvbxl6afiko2yw'),$$

$$('did:plc:p2cp5gopk7mgjegy6wadk3ep', 'did:plc:ahcw75tcvif2lf5q6fryz25n'),$$

$$\ldots$$

Notably, the node did:plc:p2cp5gopk7mgjegy6wadk3ep appears in many of these edges, indicating it sits at the crossroads of distinct local neighborhoods. Although removing a single local bridge might not fragment the entire graph, it does disconnect close-knit groups from each other locally, potentially reducing how swiftly information traverses these sub-regions.

Overall, the presence of 7 bridges and 939 local bridges in the @atproto.com follower network suggests an architecture largely dominated by a single giant component, yet peppered with numerous edges that individually connect smaller sub-neighborhoods. This underscores both stability (few global cut edges) and local fragility (many single-edge inter-neighborhood connections), shaping how information might propagate or be siloed in the Bluesky social graph.

10 Gender and Homophily

Homophily refers to the tendency of individuals to form ties with others who are similar to themselves, whether by demographic attributes (e.g., gender, age) or shared interests. In many social networks, homophily manifests as clusters of users sharing a particular characteristic.

10.1 Data Availability

In this project, the Bluesky user information (node attributes) did not include a “gender” field or other demographic markers. Consequently, no direct measure of gender-based homophily could be performed (e.g., no partitioning of the graph by “M,” “F,” “Other,” etc.).

10.2 Potential Approaches (If Data Were Available)

10.2.1 Partition & Visualization

- Using Gephi’s Appearance → Partition to color nodes by gender would allow a quick visual check for gender-specific clusters.
- Computing the modularity of the network once partitioned by gender to see if same-gender nodes cluster more strongly than expected at random.

10.2.2 Homophily Metrics

- **Edge-based measure:** The fraction of same-gender edges (M-M or F-F) versus cross-gender edges (M-F).
- **Assortativity coefficient:** In network analysis, the assortativity by a categorical attribute (like gender) quantifies how strongly nodes of the same category connect to each other vs to others.

10.3 Interpretation and Use

- **High gender homophily** suggests more segregated communities by gender, which can influence discussion topics, support mechanisms, or echo chambers.
- **Low gender homophily** means interactions are more cross-gender, indicating that the network is less prone to demographic clustering.

10.4 Conclusion on Gender Analysis

In the absence of explicit gender data, it is not currently possible to assess the homophily or assortativity within the @atproto.com network. Should such data become available in the future, the methodological steps outlined here would facilitate an examination of whether and how gender-based subgroups form in the Bluesky ecosystem.

11 Graph Density

In network analysis, density measures the ratio of existing edges to the maximum number of possible edges. For a directed graph with n nodes, the maximum number of edges is n × (n - 1). A density close to 1 indicates a fully saturated network where nearly all possible connections exist, whereas a density near 0 indicates a very sparse graph.

11.1 Density Result

- **Network Interpretation** Directed
- **Density: 0.000** (rounded to three decimal places)

Given the @atproto.com network has over 6,000 nodes but only about 20,000 edges, even a moderate edge count pales in comparison to the potentially millions of directed connections (n × (n - 1)). Hence, the density effectively rounds down to 0.000.

Figure 29: Graph Density

11.2 Interpretation

11.2.1 Sparse Connections

A near-zero density confirms that most users do not follow one another directly. Instead, the network is structured around select hubs and sub-communities, with many possible "follow" links remaining unutilized.

11.2.2 Prevalence of Short Paths

Despite the low density, the network still exhibits a small-world property (short average path length, diameter of 4). This combination—low density yet high connectivity—often arises in large-scale social networks where just a few links per person suffice to keep the graph globally reachable.

11.2.3 Implications

A near-zero density confirms that most users do not follow one another directly. Instead, the network is structured around select hubs and sub-communities, with many possible "follow" links remaining unutilized.

- **Information Flow:** A single user is typically linked to a small fraction of the total population, implying reliance on multi-step paths or key hubs for broader dissemination.

- **Growth Potential:** A low baseline density suggests ample room for new connections to form without saturating the network, potentially influencing how new communities or bridging ties evolve over time.

In summary, although the @atproto.com network is large, it is comparatively sparse, aligning with classic patterns in social graphs. Much of the connectivity is concentrated around high-degree nodes, leaving the overall edge-to-node ratio extremely low yet still cohesive enough for efficient information transit.

12 Community Structure (Modularity)

Communities in a network are groups of nodes more densely interconnected with each other than with the rest of the graph. Modularity is a measure (ranging roughly from 0 to 1) of the strength of such community divisions—a higher value indicates more distinct clusters.

12.1 Modularity Report Summary

- **Randomize:** On
- **Use Edge Weights:** Off (all edges treated equally)
- **Resolution:** 1.0
- **Modularity Score: 0.332**
- **Number of Communities: 7**

Figure 30: Modularity Size Distribution

These results stem from Gephi's implementation of the Louvain community detection algorithm. With a modularity of about 0.332, the network displays moderate structural grouping, indicating that while sub-communities exist, they are not extremely isolated from each other.

12.2 Interpretation

12.2.1 Seven Community Partitions

- The algorithm divides the network into 7 modules (clusters). Each cluster represents users more closely interlinked within that group than with users outside it.
- Some modules may be quite large, reflecting the network's hub-oriented structure, while others are smaller niche groups.

12.2.2 Moderate Modularity

A score of 0.332 suggests some cohesive sub-networks, but also significant interconnectivity between these clusters. This is common in social graphs where influencers or bridging nodes link multiple communities.

12.2.3 Visual Representation

In Gephi's Partition tab, coloring nodes by their Modularity Class highlights each cluster, providing a clear view of sub-community "blobs." Combined with ForceAtlas2, these colored groupings become visually distinct. The visual representation was tackled in the *first section* of the report.

12.3 Significance for the @atproto.com Network

- **Information Flow:** Communities can act as semi-independent "local conversations." Understanding these groups helps identify potential echo chambers or specialized interest clusters around certain influencers.
- **Cross-Community Bridges:** Nodes connecting different modules potentially have higher betweenness, playing a key role in disseminating content beyond a single group.

Overall, the modularity results confirm that the @atproto.com network is neither a completely uniform mass nor highly fragmented; instead, it comprises several moderately distinct sub-communities, each interconnected yet partially siloed in their interactions.

13 PageRank

PageRank is a link analysis algorithm famously introduced by Google to rank web pages. In a social network context, PageRank highlights nodes that receive many (or high-quality) incoming links, under the assumption that being followed by influential accounts imparts additional importance.

13.1 Report Parameters

- **Epsilon:** 0.001
  Convergence threshold—iteration stops when PageRank changes are below this value.
- **Probability (Damping Factor):** 0.85
  Probability that a user continues following links at each step (vs. teleporting to a random node). A typical value of 0.85 balances local and global influence.

13.2 PageRank Distribution

Figure 31: PageRank Distribution

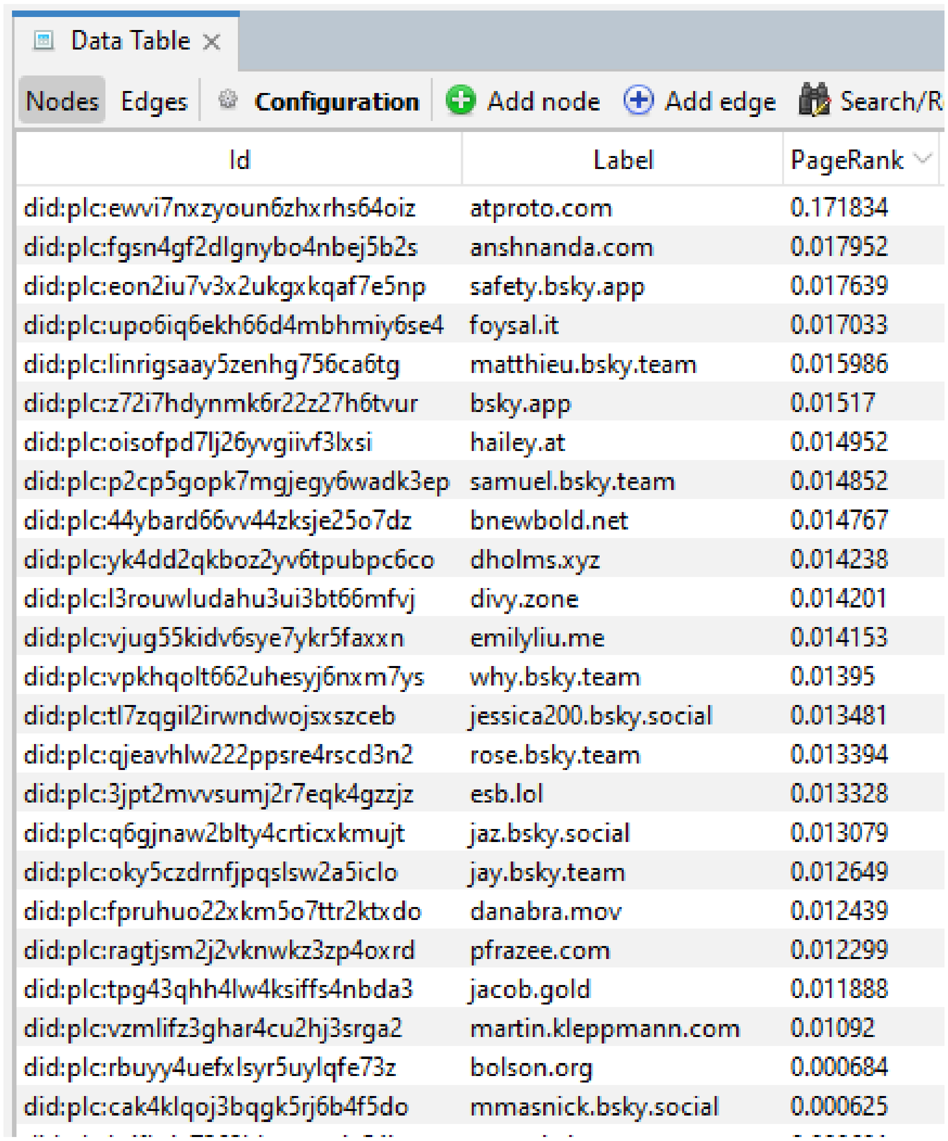
The attached PageRank Distribution chart shows that:

- Most nodes have near-zero PageRank, indicating relatively few (or low-ranked) inbound links.

- A small fraction of nodes hold visibly higher scores, reflecting their status as influential hubs. In many large social graphs, this distribution follows a long-tail pattern with a handful of top accounts dominating visibility.

13.3 Interpretation & Insights

13.3.1 High PageRank Nodes



Id	Label	PageRank
did:plc:ewvi7nxzyoun6zhxrhs64oiz	atproto.com	0.171834
did:plc:fgsn4gf2dlgnybo4nbej5b2s	anshnanda.com	0.017952
did:plc:eon2iu7v3x2ukgxkqaf7e5np	safety.bsky.app	0.017639
did:plc:upo6iq6ekh66d4mbhmiy6se4	foysal.it	0.017033
did:plc:linrigsaay5zenhg756ca6tg	matthieu.bsky.team	0.015986
did:plc:z72i7hdynmk6r22z27h6tvur	bsky.app	0.01517
did:plc:oisofpd7lj26yvgiivf3lxsi	hailey.at	0.014952
did:plc:p2cp5gopk7mgjegy6wadk3ep	samuel.bsky.team	0.014852
did:plc:44ybard66vv44zksje25o7dz	bnewbold.net	0.014767
did:plc:yk4dd2qkboz2yv6tpubpc6co	dholms.xyz	0.014238
did:plc:l3rouwludahu3ui3bt66mfvj	divy.zone	0.014201
did:plc:vjug55kidv6sye7ykr5faxxn	emilyliu.me	0.014153
did:plc:vpkhqolt662uhesyj6nxm7ys	why.bsky.team	0.01395
did:plc:tl7zqgil2irwndwojsxszceb	jessica200.bsky.social	0.013481
did:plc:qjeavhlw222ppsre4rscd3n2	rose.bsky.team	0.013394
did:plc:3jpt2mvvsumj2r7eqk4gzzjz	esb.lol	0.013328
did:plc:q6gjnaw2blty4crticxkmujt	jaz.bsky.social	0.013079
did:plc:oky5czdrnfjpqslsw2a5iclo	jay.bsky.team	0.012649
did:plc:fpruhuo22xkm5o7ttr2ktxdo	danabra.mov	0.012439
did:plc:ragtjsm2j2vknwkz3zp4oxrd	pfrazee.com	0.012299
did:plc:tpg43qhh4lw4ksiffs4nbda3	jacob.gold	0.011888
did:plc:vzmlifz3ghar4cu2hj3srga2	martin.kleppmann.com	0.01092
did:plc:rbuyy4uefxlsyr5uylqfe73z	bolson.org	0.000684
did:plc:cak4klqoj3bqgk5rj6b4f5do	mmasnick.bsky.social	0.000625

Figure 32: Nodes Sorted By PageRank

Accounts like @atproto.com or other large in-degree hubs typically rank near the top. This aligns with the earlier Degree and Eigenvector findings, affirming that certain users not only have many followers but also receive inbound links from other highly connected or well-ranked users.

13.3.2 Network Influence

PageRank highlights global importance rather than just local connectivity. A node with moderate degree can achieve a higher PageRank if it is followed by already high-ranked nodes, indicating a kind of reputation transfer within the network.

13.3.3 Visibility Gap

The disparity between near-zero PageRank for most users and moderate or high values for a select few underscores a typical inequality in social reach. Just as in web page ranking, a small set of "powerful" accounts accumulate a disproportionate share of attention and visibility.

13.3.4 Potential Applications

- **Key Opinion Leaders:** Identifying top PageRank nodes can help locate social "gatekeepers" or influential voices.
- **Content Promotion:** Content shared by high-PageRank users is more likely to spread widely, given their advantage in the network's link structure.

In sum, the PageRank analysis provides another lens on the network's power dynamics: a small group of high-score nodes wield outsized influence, while the majority remain on the periphery with minimal inbound connections from established hubs.

14 Conclusions

This network analysis of the @atproto.com follower graph reveals a large yet sparse structure with a strong central hub and numerous interconnected communities. Below are the key takeaways:

14.1 Global Topology & Connectivity

- With 6,522 nodes and 20,312 edges, the subgraph remains sparse (density 0.000), but maintains a short average path length (2.84) and a diameter of 4, indicative of a small-world effect.
- Almost all nodes form a single weakly connected component, while strong connectivity is more fragmented (5,012 strongly connected components).
- Only 7 weak bridges were found, suggesting that globally, the network is robust in an undirected sense, whereas 939 local bridges underscore numerous single-edge interconnections at the neighborhood level.

14.2 Degree & Centralities

- Degree distributions show a classic long-tail pattern: a few highly connected "hub" nodes (e.g., @atproto.com, @danabrav.mov) dominate in-degree and out-degree, while most users have few connections.
- Betweenness and Closeness confirm a small number of nodes serve as critical bridges or central paths.
- Eigenvector Centrality and PageRank further highlight influential users who not only have many followers but are also followed by other well-connected accounts.

14.3 Clustering & Triadic Closure

An average clustering coefficient of 0.754, combined with 4,527 fully directed 3-cycles, implies that local neighborhoods are unexpectedly dense even in a directed graph. Triadic closure is robust, meaning many pairs of neighbors also link to each other.

14.4 Community Structure

- Modularity of 0.332 indicates a modest but significant presence of sub-communities, with 7 identified clusters. One large class contains the majority of nodes, with several smaller groups on the periphery.
- Bridging edges and high-betweenness nodes facilitate connections across these communities.

14.5 Information Flow & Influence

- Sparse global density yet short paths denote a structure where a few prominent nodes can rapidly disseminate information, while most participants remain lightly connected unless they follow or are followed by these hubs.
- PageRank scores confirm that "hub" accounts enjoy disproportionate visibility and influence within the network.

14.6 Gender & Homophily (Data Limitation)

No gender attribute was available, preventing direct analysis of homophily. Should demographic or interest-group data become accessible, it would be valuable to see if sub-communities align with shared attributes or remain primarily interest-based.

14.7 Bridges & Vulnerabilities

Although globally robust (few cut edges), the abundance of local bridges hints that certain neighborhoods could be easily isolated if these edges were removed or if key users become inactive. Identifying those edges and users—potential "brokers"—could be important for understanding or mitigating fragmentation.

14.8 Future Work

Enriching the data with content-based attributes (topics, hashtags) or demographics would shed light on how shared interests (or identities) shape sub-communities and influence flows.

In summary, the @atproto.com network combines a star-like pattern of major hubs with a web of smaller connections that sustain short path distances and high clustering. This architecture supports both broad reach for key influencers and tightly knit local clusters, underscoring the hallmark complexities of emerging social platforms like Bluesky.

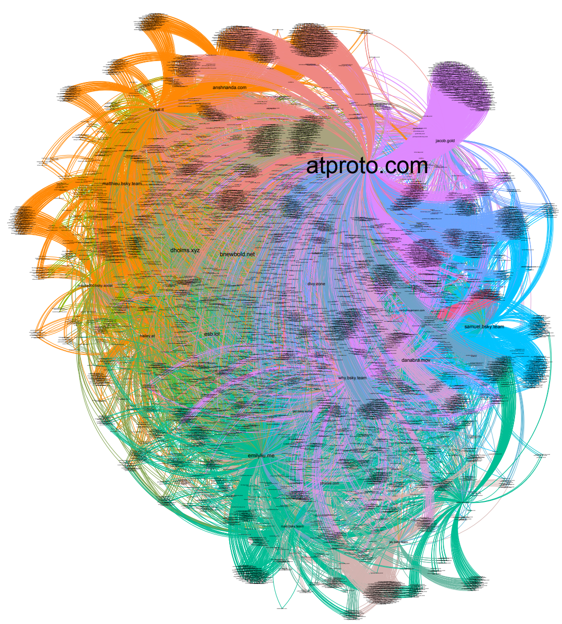


Figure 33: Final Graph (zoom in for details)

References

[1] Bastian, M., Heymann, S., Jacomy, M. (2009). Gephi: An Open Source Software for Exploring and Manipulating Networks. In Third International AAAI Conference on Weblogs and Social Media (ICWSM-09).
Gephi software: `https://gephi.org/`

[2] Bluesky. (2023). Authenticated Transfer Protocol (ATP) documentation.
Official site: `https://atproto.com/`