

Echo Chambers, Bridging, and Structural Bias in BlueSky: A Multi-Method Network Analysis of Starter Packs

Melih Dilbaz (31169), Emre Tuygan (30547), Eda Ecesu Çınar (32467), Boray Saydam (29094)

10 June 2025

Abstract

This study investigates echo chamber formation and polarization dynamics within decentralized social media platform Bluesky, focusing on its starter pack feature. We analyze 45 starter packs spanning Natural Science and Political Science domains by constructing directed, weighted interaction networks based on replies and reposts. Using hybrid models that integrate behavioral and topological signals, we simulate diffusion processes incorporating reinforcement and edge decay. We further examine cross-pack interaction behavior through Cross Pack Interaction Rate (CPIR) and Normalized Bridge Scores, and apply community detection via Leiden modularity and Infomap. Regression Discontinuity Design (RDD) is employed to detect gender based structural thresholds. Our results show that while Natural Science packs exhibit high internal clustering with limited external reach, Political Science packs demonstrate broader engagement and bridge dynamics. Although homophily levels were moderate (0.4–0.6) across all starter packs the pack structure still encouraged information to circulate mostly within the same group, without causing extreme polarization. Results support the idea that starter packs help build group cohesion and echo chamber like behavior. Results also show that users in Political Science are more intellectually open and connected to diverse communities compared to users in the natural sciences.

1 Introduction

The rapid growth of decentralized social networks has introduced new challenges in understanding how information flows within online communities. Bluesky, a newly emerging decentralized social media platform, has offered a unique experience for users to bootstrap their social graph through its starter packs. Starter packs are curated lists of users typically centered around a specific topic or community. Based on the previous research of Pamfil et al.[8] it is known that these starter packs offer an opportunity to examine how people interact, form content, and more significantly, how they influence one another within and across different interest groups. This paper aims to explore the interaction patterns among users in these starter packs, focusing on the impact of Blusky starter packs on polarization across the communities of natural science and political science on the basis of following hypotheses:

1. **H1.** Starter packs as seed structures amplify in group cohesion but also reinforce social fragmentation by concentrating interactions within defined user clusters.
2. **H2.** Interaction networks within specific starter packs exhibit echo chamber characteristics, with high reciprocity and topical homogeneity reinforcing internal consensus.
3. **H3.** Structural bias exists within starter packs, reinforcing hierarchical dominance of certain identity groups through uneven centrality and assortative connectivity.

4. **H4.** Political Science and Natural Scienve users exhibit different levels of intellectual openness, measured by cross-starter-pack interaction rates.

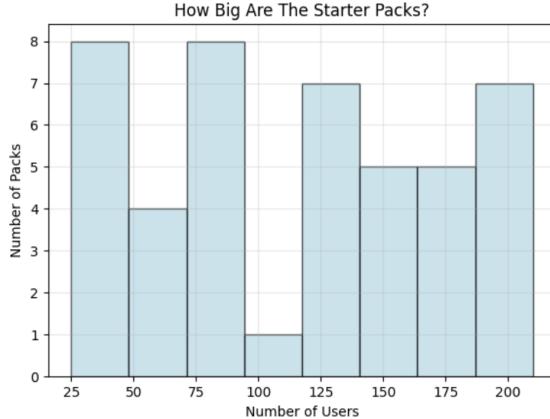
To evaluate these hypotheses, we manually selected 25 natural science and 20 political science starter packs, where those numbers are considered based on pack quality and topical relevance.. We constructed directed, weighted interaction networks (replies and reposts) using Bluesky's API and a Selenium scraper. Our analysis encompasses diffusion simulations, structural metrics, and complementary community detection methods (Leiden modularity and Infomap). By integrating behavioral and topological signals, we investigate how starter packs shape the polarization dynamics and echo chamber phenomena on a decentralized platform.

This study was conducted by Melih, Eda Ece, Emre, and Boray. Melih was responsible for constructing and visualizing the interaction networks, implementing and visualizing the community detection algorithms, and developing the echo chamber analysis. Eda Ece finalized the data collection process and carried out cross-analyses between starter packs and the detected communities. Emre handled data pre-processing, network construction, the implementation of diffusion modeling and social reinforcement mechanisms, and the analysis of network statistics. Boray focused on designing and implementing the gender analysis as well as the regression discontinuity design.

2 Dataset and Data Collection

The dataset focuses on user accounts in BlueSky's starter packs, which are lists of users created by users and mostly centered around specific topics or interests. In order to collect the necessary data, the data collection process involved three stages:

1. **Selecting Starter Packs:** 25 packs were chosen based on relevance, activity, content quality, and coherence.
 - Whether the listed accounts were still active or not
 - Whether the feed content was relevant rather than being spam
 - Overall coherence and quality of the starter pack
2. **Extracting User Handles:** A Selenium-based web scraper was used to visit each starter pack URL and extract user handles. This step was necessary because Bluesky does not provide an API to access the starter pack data.
3. **Fetching User Data:** Using Bluesky's official API, we retrieved each user's followers, followings, and up to 1000 of their latest posts, replies, and reposts. To ensure robustness, the script included a retry mechanism and periodic delays to avoid rate limiting. All data were saved as individual .json files. The entire data collection process took over 72 hours.



(a) Histogram of starter pack sizes (range: 25–210 users).

Figure 1: **Starter pack size distribution.** shows that pack sizes span the full range from 25 to 210 users (mean = 114.1, median = 122, SD = 58.3).

A total of 45 starter packs were included in the analysis, each containing between 25 and 210 users. These packs were manually selected for relevance to STEM or Political Science . Pacs are grouped into four size categories—Small (25–75 users), Medium (76–125), Large (126–175), and Very Large (176+)—and found, as shown in Figure 1, that Small and Large packs are the most prevalent, ensuring no single category dominates the dataset.

3 Network Construction and Analysis

The goal was to build a directed network using the data collected from the starter packs that reflect the interaction dynamics between the accounts inside the starter packs. Therefore, nodes were chosen to represent individual users (influencers), while edges represented interactions such as how many times a user replied to or reposted another user. For the rest of the paper, the users will be called as influencers.

3.1 Edge Representation

Logan et al. [5] identify four defining characteristics of modern social network analysis: (1) structural intuition based on ties linking social actors, (2) grounding in systemic empirical data, (3) use of graphic imagery, and (4) quantification through rigorous mathematical modeling. Together, these features provide a comprehensive framework for understanding online information propagation.

Within this framework, the most effective diffusion of messages occurs through highly influential users commonly referred to as *influencers*, which are also the one of the primary focuses of the starter-pack analysis. These accounts typically maintain large followings and act as compelling, trusted sources of information.

However, as Erlandsson et al. [3] argue, raw follower counts alone are insufficient to quantify true influence. Instead, one must consider the nuanced patterns of user interactions. Logan et al. [5] outline three such interaction types: mentions, retweets, and replies. While these do not offer a direct metric of tie strength, they can be used to approximate it.

To emphasize the deliberate nature of rarer interactions, Logan et al. propose using the inverse of interaction frequencies. Under a uniform interaction distribution, each type would contribute

roughly 3 to an arc weight (since $1/0.33 \approx 3$). However, in their observed data—comprising 14.3% replies, 28.6% mentions, and 57.2% retweets—replies contribute a weight of 7, mentions 3.5, and retweets 1.75.

In the implementation, this inverse-frequency weighting strategy is mirrored.

3.2 Edge Construction Workflow

1. **Assembling interaction data:** For each starter pack, all influencers are iterated over to collect reply and repost counts, along with topological features such as common incoming and outgoing neighbors. These are aggregated into a list of edge data for further processing.
2. **Computing dynamic weights:** For each pack, the total number of replies and reposts are computed, their relative frequencies are derived, and the inverse-frequency weights are calculated:

$$\text{weight}_{\text{reply}} = \frac{1}{\text{freq}_{\text{reply}}}, \quad \text{weight}_{\text{repost}} = \frac{1}{\text{freq}_{\text{repost}}}$$

These weights assign greater significance to less frequent interactions, they become more deliberate.

3. **Applying weights to the network:** For each directed edge (u, v) in the base graph G_{base} , an interaction score is computed:

$$\text{interaction_score}_{u,v} = (\text{replies} \times \text{weight}_{\text{reply}}) + (\text{reposts} \times \text{weight}_{\text{repost}})$$

This score is further enriched with relational assortativity metrics, separately focusing on common outgoing, common incoming, and both of those scenarios. This separation measures the degree of shared neighbors and potential for information reinforcement.

This workflow operationalizes rigorous mathematical modeling and ensures a more accurate measurement of an influencer’s impact beyond superficial follower counts.

3.3 Directed Network

Following Logan et al. [5], a directed network representation is adopted to capture interaction asymmetries. This modeling choice allows to differentiate between users such as celebrities who may rarely post but attract many interactions and bots or spammers, who may exhibit high activity but lack meaningful inbound engagement. The direction of each edge thus encodes the flow of interaction (who is influencing whom), which is critical for understanding influence propagation dynamics.

3.4 Hybrid Model: Topology and Interaction

The paper builds a hybrid link model by integrating both behavioral and topological signals between starter-pack influencers in the network, drawing on Valverde-Rebaza and de Andrade Lopes [13] community-aware framework and Logan et al. [5] inverse-frequency weighting of Twitter interactions. First of all, the set of reply and repost counts for each influencer alongside topological affinities (common in- and out-neighbors) into a base directed graph. Then three variants of a Resource Allocation (RA) score are computed, the formulas for the neighbors linked with incoming and outgoing edges as follows:

Specifically:

- It's built a base directed graph where edges are formed by reply and repost counts between influencers and weighted accordingly.
- Compute *topological affinity* using shared neighbors. For each edge (u, v) , it's measured three variants of the Resource Allocation (RA) index:

$$RA_{\text{incoming}}(u, v) = \sum_{z \in \Gamma^-(u) \cap \Gamma^-(v)} \frac{1}{|\Gamma^-(z)|}$$

$$RA_{\text{outgoing}}(u, v) = \sum_{z \in \Gamma^+(u) \cap \Gamma^+(v)} \frac{1}{|\Gamma^+(z)|}$$

$$RA_{\text{both}}(u, v) = RA_{\text{incoming}}(u, v) + RA_{\text{outgoing}}(u, v)$$

where $\Gamma^-(u)$ and $\Gamma^+(u)$ denote the set of users who have incoming and outgoing edges with user u , respectively.

This modeling choice penalizes high-degree shared neighbors, emphasizing rare shared connections as stronger indicators of similarity or influence. After normalizing these RA scores alongside our inverse-frequency interaction score, simple combination is obtained:

$$l_{\text{combo}}(u, v) = a \times l_{\text{int}}(u, v) + (1 - a) \times s_{u,v} RA$$

This formula is applied to each of three weighted digraphs (incoming, outgoing, both). [11] This hybrid edge property merges community-based RA and activity-based interaction metrics such that the hybrid network reflects both engagement behavior and structural proximity, offering a nuanced view of influence potential within and across starter packs.

3.5 Network Representation

To construct the interaction graph, a preprocessed CSV dataset is prepared. Each row in the dataset contained the following fields: source user, target user, reply count, repost count, number of shared followers, and number of shared followings. For example, a tuple like ('userA', 'userB', 3, 1, 5, 2) indicates that **userA** replied to **userB** three times and reposted them once, and that the two users share five followers and follow two of the same accounts.

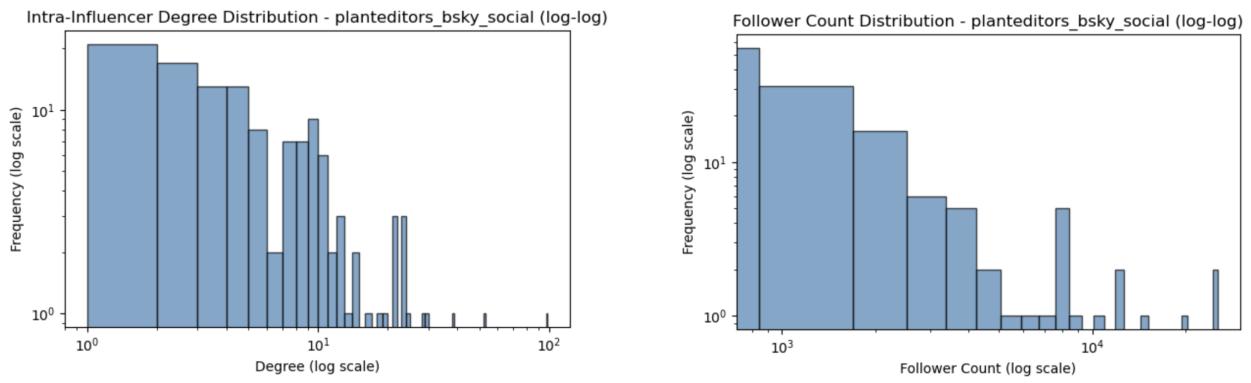
These attributes were parsed into dictionary structures to streamline iteration. For each user, the algorithm iterated over all reply and repost targets to compute total reply and repost counts. Simultaneously, mutual followers and mutual followings were calculated while explicitly excluding self-loops to avoid artificial inflation of influence.

After constructing the initial directed graph, we reversed the edge directions to better simulate the flow of influence. For example, if **User A** reposted **User B**, the raw edge would originally point from A to B. After reversing, the edge points from B to A, reflecting the idea that **User B**'s content influenced **User A**'s behavior.

Incoming-weighted Influencer Graph for Starter Pack: `planteditors_bsky_social`



Figure 2: An example directed interaction network of one starter pack `planteditors_bsky`. Node size reflects in-degree and edge thickness encodes interaction weight.



(a) Intra-influencer degree distribution (log-log)

(b) Follower count distribution (log-log)

Figure 3: Distribution plots for the starter pack `planteditors_bsky`. The left shows internal degree variation; the right shows external reach via follower counts.



Figure 4: Social Network of Natural Science starter packs. The different colors correspond to distinct starter packs. In total 25 starter pack is included. Node size reflects total degree and edge thickness encodes interaction weights.

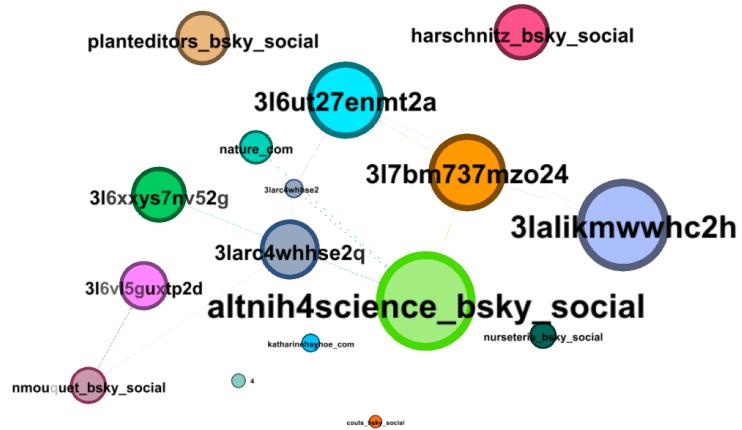


Figure 5: Major starter packs for the same network, representative nodes are selected on the total degree basis.

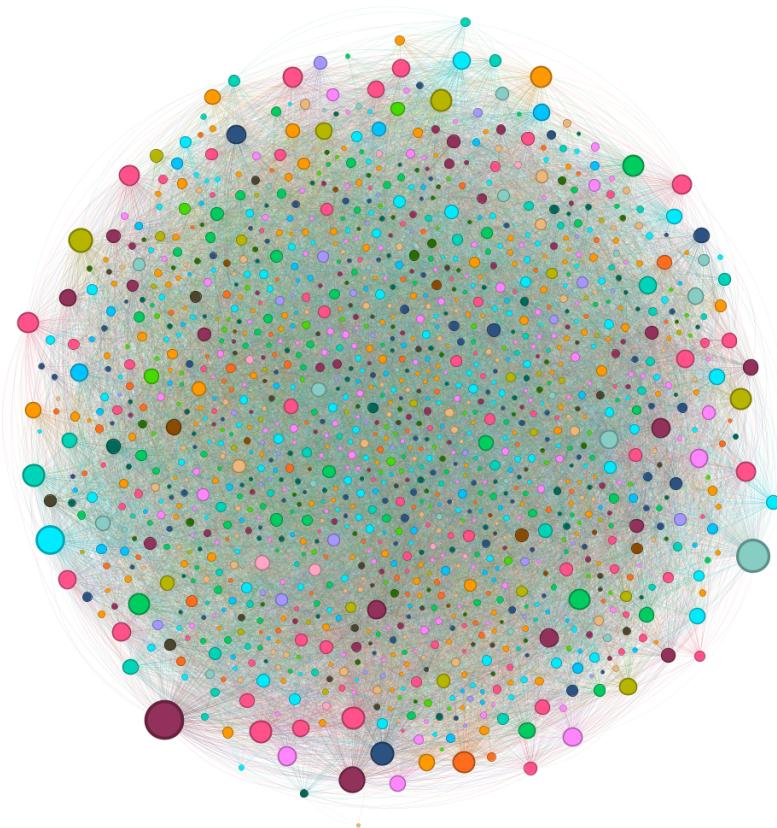


Figure 6: Social Network of Political Science starter packs. The different colors correspond to distinct starter packs. In total 20 starter pack is included. Node size reflects total degree and edge thickness encodes interaction weights.

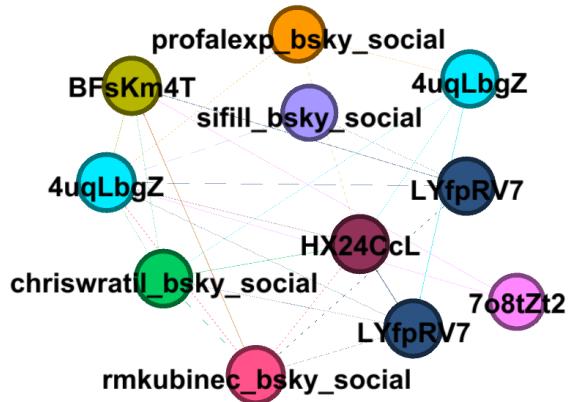


Figure 7: Major starter packs for the same network, representative nodes are selected on the total degree basis but node size is equal due to homogeneity of the node degrees.

4 Diffusion Modeling and Social Reinforcement

4.1 Simulation

One of the core insights of this study, is driven by the work of Centola and Macy [2] on *social reinforcement*, suggesting that individuals rarely adopt new behaviors or beliefs after a single exposure. Instead, multiple independent confirmations from peers are typically required to drive meaningful adoption.

Translating this principle to an online setting with millions of users is computationally intractable. Therefore, the problem is isolated to a more tractable subset: *starter packs*, consisting of domain-specific influencers connected by mutual repost and reply behaviors.

By focusing on this elite circle, it is assumed that if a piece of content receives repeated exposure within such a group, it is more likely to achieve persuasive momentum than through a single broadcast.

The SI-style is modified and the produced simulation is run across a range of interaction weight thresholds and baseline ε values. Since diffusion behavior in this domain is not well studied, these parameters are varied systematically to uncover consistent patterns and common diffusion trajectories.

The reinforcement mechanism is implemented through a sigmoid function, following Centola and Macy [2], which increases infection probability as the number of infected neighbors (k) grows. This captures the idea that individuals require repeated, independent exposures to adopt behaviors or ideas. The infection probability from each infected neighbor is calculated as the product of the normalized edge weight (representing interaction strength) and the Centola-style sigmoid output:

$$\text{prob}_{i \rightarrow j} = \text{edge_weight}_{i,j} \cdot \frac{1}{1 + e^{-\alpha(k-\beta)}}$$

where α controls the steepness, and β the midpoint of the reinforcement curve.

Edge weight normalization plays a central role in ensuring comparability and interpretability of interaction strengths across the network. The raw interaction counts are computed as the sum of replies and reposts between user pairs and they vary widely in scale and are heavily skewed due to power-law behavior in social networks. To address this, it is first applied a $\log(1 + x)$ transformation to all raw interaction counts. This dampens extreme values while preserving the rank order, reducing the influence of outliers without discarding interaction intensity.

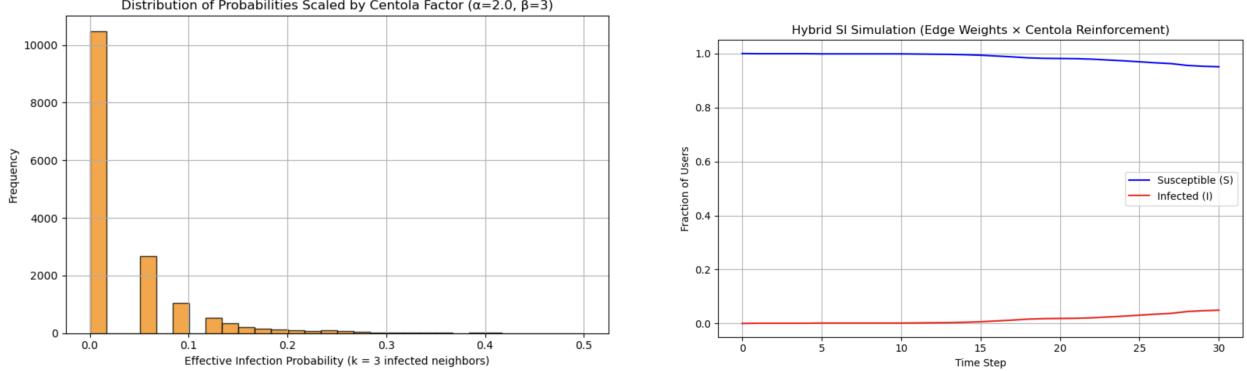
Next, the transformed weights are scaled to the $[0, 1]$ interval using min-max normalization. In degenerate cases where all transformed values are equal, a default weight of zero is assigned. To prevent very weak ties from being entirely excluded from the simulation, a small positive floor is applied (e.g., $\varepsilon = 10^{-3}$), ensuring every edge retains some minimal probability of influence.

Edge weight decay is introduced to simulate the weakening of influence over time. In practice, repeated failed infection attempts from an infected node to a susceptible neighbor may lead to reduced influence, reflecting phenomena such as attention decay, content fatigue, or diminished trust. Additionally, as the original content ages, it is less likely to cause engagement.

This behavioral fading can be modeled by multiplying the edge weight by a decay factor $\delta \in (0, 1)$ after each time step or following failed infection attempts:

$$\text{edge_weight}_{i,j}^{(t+1)} = \delta \cdot \text{edge_weight}_{i,j}^{(t)}$$

This extension helps to avoid overestimating the impact of static, high weight connections and offers a more realistic representation of saturation effects in online influence.



(a) Distribution of edge-level infection probabilities, computed as normalized edge weights scaled by a Centola-style [2] sigmoid factor with $k = 3$, $\alpha = 2.0$, and $\beta = 3$. This reflects the effect of repeated exposures.

(b) Fraction of susceptible and infected users over time from the hybrid SI simulation for a given patient zero. Infection likelihood depends on edge weights and number of infected neighbors via Centola reinforcement.

Figure 8: Diffusion modeling results using a Centola reinforced SI process on the interaction network. Panel (a) shows the initial probability distribution used in infection attempts; panel (b) visualizes the resulting cascade dynamics.

4.2 Cross Pack Interaction Rate

4.2.1 Cross-Pack Interaction Rate (CPIR)

To quantify openness, we defined the **Cross-Pack Interaction Rate (CPIR)** as:

$$\text{CPIR} = \frac{\text{Number of cross-pack interactions}}{\text{Total number of interactions}}$$

CPIR captures the proportion of a user’s replies and reposts that are directed toward users *outside* their own starter pack. Cross-pack interactions were identified by comparing the starter pack label of each interaction target with the user’s own. A higher CPIR suggests greater engagement with external communities, whereas a lower CPIR indicates more closed behavior.

4.2.2 Normalized Bridge Score and Classification

In addition to CPIR, we examined the number of *unique external starter packs* a user engaged with. Users who interacted with members of different starter packs were considered **bridges**, facilitating information flow across community boundaries. This aligns with prior work on cross group diffusion in social systems [12].

To account for domain size differences, we defined a **Normalized Bridge Score (NBS)** as:

$$\text{Normalized Bridge Score} = \frac{\text{Number of unique external packs interacted with}}{\text{Total domain packs} - 1}$$

This score lies in the range $[0, 1]$ and reflects the breadth of a user’s external interactions. Based on NBS thresholds, users were classified into four categories:

- **Isolate:** 0 external packs
- **Specialist:** $\leq 10\%$ of possible external packs
- **Connector:** 10–30%
- **Bridge:** $> 30\%$

5 Community Detection : Modularity and Infomap

5.1 Introduction

1. **Why Community Detection Matters:** Community detection in networks is useful to identify the way how agents interact, form belief, and create a flow of information. Social networks could correspond to tightly connected clusters, which can work as *echo chambers* where reinforcement of similar ideologies can dominate over the platform. Thus, uncovering those structures could help detect behavioral or ideological polarization.

Community detection goes beyond individual interactions, and showcases whether observed groupings are of random noise or logically formed. This phenomenon is of use for echo chamber analysis, where the focus is drawn into persistent communication loops.

2. **Overview of Modularity and Infomap Approaches:** To perform detection, methodologically different but complementary two algorithms are taken up for this paper’s purpose: **Leiden modularity optimization** and **Infomap**.

This implementation is built upon Leiden algorithm, and Louvain modularity to improve stability as maximizing intra-community density [14]. The observed graph has a modularity score favoring some partitions that create dense clusters. This approach suggests comparing the observed score to a null model preserving the degree distributions. The Leiden method has returned 11 coarse-grained communities.

In contrast, Infomap technique models a flow-based view of community detection via probability calculation at each step of walk across the network. Then, it compresses the total information through partitioned nodes, this helps describe such flows[9]. Another matter differing it from modularity is its sensitivity to edge direction and weight such that it could capture the behavioral persistence that is typical case for echo chambers. This has led to detection of 150 communities in Natural Science social network.

Recent empirical research supports using both methods. Blekanov et al. [1] argue combining Infomap and Leiden for detecting communities on social media would do better, specifically for weighted and large-scale graphs. Their findings show that while Infomap captures nuanced flow dynamics, Leiden is efficient at revealing topological structure—making their hybrid use suited for identifying echo chambers.

How Echo Chambers Appear in the Network

- Leiden identifies where nodes are densely connected. Infomap identifies where information flow tends to remain trapped. This behavior can mirror how echo chambers act in real social systems.
- If Leiden and Infomap produce similar communities, it shows robustness across paradigms.

5.2 Modularity-Based Implementation

1. **Application of Modularity and Maximization:** In this study, echo chambers are considered as a phenomenon that is formed by tightly-knit clusters of starter-pack influencers. Due to those topological characteristics of the network, it is aimed to apply modularity-based approaches *maximizing the modularity-score*—a quality function introduced by Newman and Girvan [7]. As the next step, the degree of the modularity is assessed by comparing it to a null model score.

The modularity score measures the difference between the score observed from the studied network and the one from the null model, which is an average score obtained as a result of some number of runs. A higher modularity than the null model, suggests formation of meaningful partitions across the network.

Newman–Girvan Spectral Algorithm Newman and Girvan [7] initially proposed a spectral algorithm where the modularity matrix is used to recursively maximize the modularity Q_{NG} . Although it could work effectively, it suffers from failing to detect small yet significant communities.

Louvain Method The Louvain method and its improved version, the Leiden algorithm, iteratively optimize modularity by greedy local merging and aggregation. In the implementation, the Leiden method with *directed modularity* is adopted, as it is shown to be robust in real-world large-scale graphs [8].

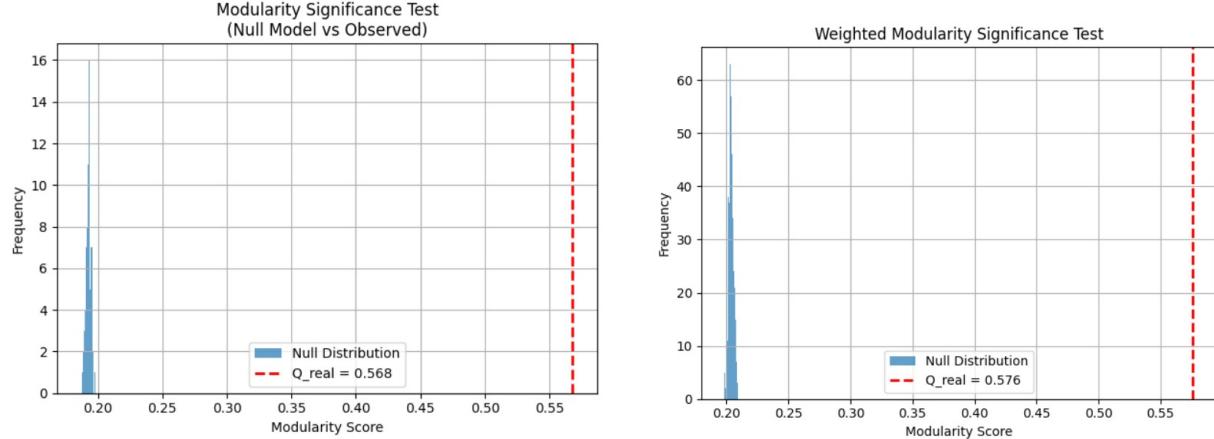
2. **Preprocessing the Graph and Implementation:** For community detection, first the studied network is binarized through edge weights (whether there is an arc or not), then the graph is converted into Graph object with enumerated nodes.

The Leiden algorithm is applied on the binarized graph using

```
leidenalg.ModularityVertexPartition
```

from the library `leidenalg`. To measure the significance of the modularity score, p-value is calculated and comparison is made to randomly generated 100 null graphs (aimed to spot a mean value), which preserves the in- and out-degree distribution.

The null model is consistent with the configuration model proposed by Zhang and Chen [15], which provides a well-defined foundation for assessing statistical significance.



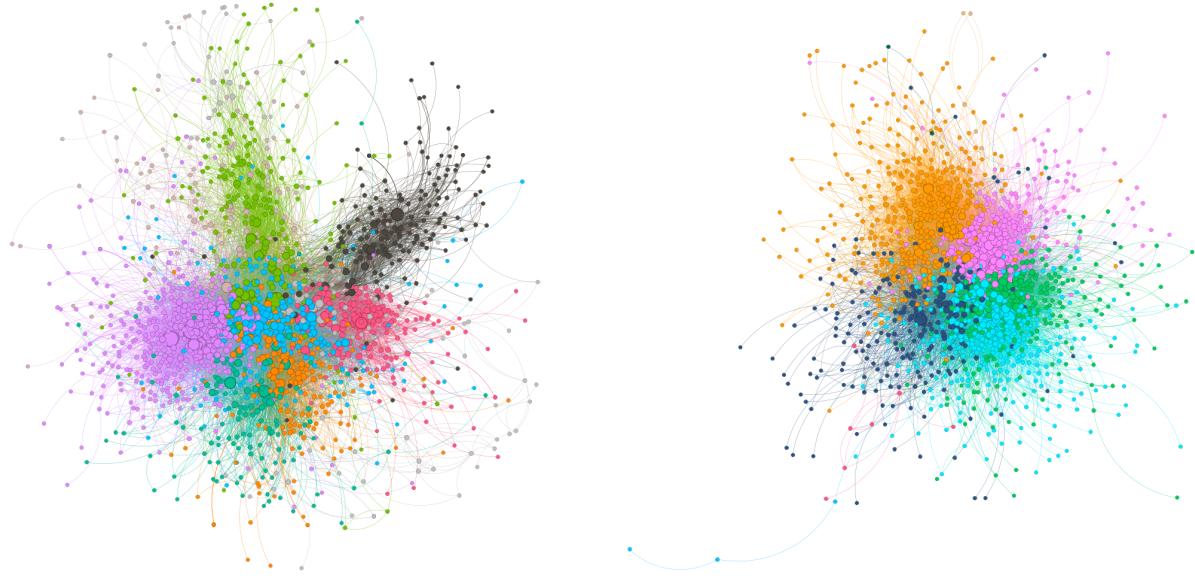
(a) This chart shows the default application of Leiden algorithm on Natural Sciences Network with binarized edges whose weights are reduced to 0 or 1 values. The number of generation for null network is 100.

(b) To experiment the difference, this time edge weights are not binarized, but passed into the Leiden method, and the Natural Science Network is kept the same. This run has generated null models 500 many times.

Figure 9: The red dashed lines indicate to the modularity score of the observed network, but the cumulated blue histograms are degree-preserving null network scenarios with the modularity score around 0.20 at both. Those findings will give rise to the null hypothesis testing in the chapter 6.

Community Detection by Modularity: Visuals

Modularity by its core algorithm focuses on the topological characteristics of the network and the general use suggests the application of it in a binarized manner. However, in this paper it is aimed to see the effects of the interaction rates between nodes; thus, modularity is implemented in both ways, binarized and weighted. Figure 9 suggests that those two approaches do not differ significantly. Yet, considering weighted edges leads to increase in the number of communities detected.



(a) Natural Science starter-pack network with 8 communities. (b) Political Science starter-pack network with 9 communities.

Figure 10: Community detection by `modularity` is applied ignoring edge weights, producing coarse-grained communities.



(a) Natural Science starter-pack network with 12 communities. (b) Political Science starter-pack network with 16 communities.

Figure 11: Community detection by `modularity` is applied considering edge weights, producing coarse-grained communities.

5.3 Infomap: Flow-Based Detection

1. **Intuition: Random Walks:** Infomap differs from modularity techniques by grounding on flow dynamics such that communities are the regions where a random walker remains for a relatively long time. This formalizes community detection as a minimization problem through iterative walks, which aims to describe the path of the walker using the fewest number of bits.

The Infomap approach is integrated to echo-chamber analysis. It is apart from modularity, which searches for topological partitions while Infomap detects circulations in which information flow is trapped and isolated mirroring effects of echo chamber phenomenon [14].

The Map Equation is the minimum number of bits to represent a random walker's movement, which reflects the underlying community structure. Infomap seeks the partitions with the minimized number of nodes. This has shown strong performance obtaining fine-grained modules in weighted and directed networks, complementing modularity-based approaches [14].

2. **Preprocessing the Graph and the Infomap Algorithm Steps:** IGraph library and `community_infomap()` method are used, plus edge weights are passed in. Some number of runs with random seeds are performed, producing a partition of the network into communities.

Robustness and Consistency are evaluated through partitions by computing pairwise Normalized Mutual Information (NMI) and a mean score [15]. A high NMI score indicates similar structure detection, supporting the robustness of the discovered echo chambers.

Using Infomap along with modularity-based methods is recommended by Leiden. Blekanov et al. [1] arguing that combining the two produce complementary insights in social networks. Whereas Leiden excels in explores large, topologically cohesive groups, Infomap uncovers smaller modules that may demonstrate tighter echo chambers.

Community Detection by Infomap: Visuals

Infomap suggests an algorithm with a behavioral focus across a network that captures the modules in which information flows are trapped. In addition, this paper utilizes the the infomap technique as a complementary method to capture finer-grained communities.



(a) Natural Science starter-pack network.

(b) Political Science starter-pack network.

Figure 12: Community detection by `infomap` is applied, preserving directed and weighted network properties. A Fine-grained community detection is provided. Unlike the result of modularity, it produced a lot of communities as many as 150 for both cases above.

5.4 Gender Analysis through Regression Discontinuity Design

Social media does not host only political or ideological polarization across its users. Moreover, there are occurrences where gender-based biases could shape the behavioral and topological structures in the social media. Thus, this study has also taken the gender factor into account to investigate potential structural biases and threshold effects in gender composition within Bluesky starter packs. Therefore, a comprehensive gender analysis has been conducted using Regression Discontinuity Design (RDD) across the decentralized social network studied in this paper.

1. Methodological Framework and Data Processing

Gender Classification Architecture A sophisticated three-stage gender classification pipeline was implemented to process the studied social networks. Prior to the implementation, dataset is collected via temporary email generation through the **1secmail.com** API. It is aimed to manage rate limiting across multiple **GenderAPI.io** keys, with robust data collection being ensured despite API constraints.

Advanced preprocessing techniques were employed by the name extraction algorithm, including CamelCase pattern recognition (e.g., "JohnSmith" → "John Smith"), special character normalization, common word filtering using a curated lexicon of 70+ stop words, and minimum length validation (≥ 2 characters). Gender classification was performed using **Gender-API.io** with multiple API keys being managed through an intelligent rotation system, with exponential backoff retry logic and chunked processing (20 names per batch) being included.

Data Scope and Processing The analysis was conducted across 1,677 users in 25 Natural Science starter packs and 1,816 users in 20 Political Science starter packs being represented. Gender classification was performed on the follower networks rather than following networks to capture audience demographics, with a three-category classification being yielded: Female (0), Male (1), and Unknown (2).

2. **Regression Discontinuity Design Implementation:** The RDD analysis was grounded in the hypothesis that gender composition operates as a structural organizing principle in network formation, with sharp discontinuities being potentially created at critical threshold values. Three theoretically motivated thresholds were selected: Gender Parity (50%) based on demographic balance theory, Female Minority (33%) reflecting critical mass theory in organizational behavior, and Female Majority (67%) capturing reverse minority dynamics.

The primary RDD specification was estimated as:

$$Y_i = \alpha + \beta_1(X_i - c) + \beta_2 T_i + \beta_3(X_i - c)T_i + \varepsilon_i$$

where Y_i represents outcome variables (homophily, network size, density, total followers), X_i is the running variable (female ratio), c denotes the threshold value, T_i is the treatment indicator, and β_2 captures the treatment effect of interest. Multiple bandwidths (0.05, 0.10, 0.15, 0.20, 0.25) were systematically tested to ensure robustness, with optimal bandwidth selection being conducted using MSE minimization criteria.

6 Results

6.1 Results of Individual Starter Pack Network Analysis: Natural vs. Political Science

As given in the figure 2, every starter pack is analyzed and plotted. Key structural metrics are computed for 25 natural science and 20 political science starter pack subgraphs. The following summarizes the key findings and comparative network properties of these two domains.

Density. Natural science packs exhibited a higher average density ($\mu = 0.130$) than political science packs ($\mu = 0.078$), indicating that the network in the natural sciences domain tend to be more densely interlinked. The maximum densities were 0.482 and 0.192 respectively, suggesting denser structures in some science packs.

Connectivity. The average size of the largest weakly connected component was 66.6 for science packs and 81.95 for political packs. Natural science packs had more variation ($\sigma \approx 64.76$) compared to politics ($\sigma \approx 48.24$). Strongly connected components averaged 40.56 nodes in science and 51.6 in politics, indicating similar but slightly lower mutual engagement across the natural science network.

Radius. Political science graphs had a higher average radius ($\mu = 3.4$ vs. $\mu = 2.5$), implying slightly longer shortest paths and more elongated influence flows within the strongly connected core.

Reciprocity. Reciprocity was marginally higher in political science packs ($\mu = 0.300$) than in natural science packs ($\mu = 0.335$), no significant difference exists but it might suggest that mutual interaction is more common among political science influencers.

Assortativity. Both domains showed negative assortativity (science: -0.218 , politics: -0.159), reflecting the tendency of high-degree nodes to connect with low-degree ones. This is consistent with

broadcast-style dynamics, where a few central figures have influence over many other peripheral users.

Degree Statistics. Political science starter packs have slightly higher average degrees (mean: 9.09, median: 6.7) than science packs (mean: 7.55, median: 5.38), though both show high variance. The standard deviation of the total degree in natural science network was 7.47, again indicating a small subset of highly connected influencers in both domains.

Clustering. The mean clustering coefficients were very similar: 0.409 (natural science) and 0.371 (political science), implying slightly stronger local connectivity among influencers within the natural science communities..

While overall structures appear similar, political science packs showcase slightly more reciprocity, longer influence chains, and marginally higher degree centrality. These patterns may reflect domain-specific engagement styles: political discourse emphasizing extended dialogue and mutual visibility, whereas scientific discourse may prioritize hierarchical amplification within a locally cohesive structure. But such causalities require further research.

These insights support the diffusion modeling suggested in the following chapters by highlighting which domains have stronger potential for the echo chamber formation and reinforcement loops.

(a) Natural Science Starter Packs					(b) Political Science Starter Packs				
Metric	Mean	Std	Min	Max	Metric	Mean	Std	Min	Max
Density	0.130	0.113	0.024	0.482	Density	0.078	0.058	0.023	0.192
Radius	2.50	1.18	1.00	5.00	Radius	3.40	0.88	2.00	6.00
Weakly CC Size	66.60	64.76	7.00	180.00	Weakly CC Size	81.95	48.24	18.00	162.00
Strongly CC Size	40.56	47.69	1.00	144.00	Strongly CC Size	51.60	38.62	11.00	119.00
Reciprocity	0.335	0.197	0.000	0.769	Reciprocity	0.300	0.111	0.130	0.544
Assortativity	-0.218	0.217	-0.758	0.511	Assortativity	-0.159	0.084	-0.342	-0.015
Mean Degree	7.55	4.51	2.20	19.95	Mean Degree	9.09	4.54	3.21	19.63
Median Degree	5.38	3.34	2.00	15.00	Median Degree	6.70	4.13	2.00	18.00
Std Degree	7.47	5.16	1.14	19.69	Std Degree	8.59	3.54	2.35	15.68
Clustering	0.409	0.205	0.000	0.824	Clustering	0.371	0.118	0.111	0.572

Table 1: Summary of network structural metrics for 25 natural science and 20 political science starter packs.

6.2 Results of Complete Starter Pack Network Analysis: Natural vs. Political Science

In this chapter, starter packs are analyzed within their relations across the studied social networks.

Density. The cumulative graph of starter packs for political science was denser (0.0103) than that of natural science (0.0060), suggesting a slightly higher frequency of interactions among political influencers overall.

Connected Components. The largest weakly connected component in the natural science network contained 1,639 nodes, compared to 1,419 in the political network showing slightly broader reachability in natural science. However, the largest strongly connected component was larger in political science (1,124 nodes) than in natural science (1,109 nodes), there is no significant difference

to derive a conclusion but sizes of strongly connected components in both of domains being around half of the total network science can give a teaser of echo chamber.

Graph Radius. The radius of the strongly connected core was higher in natural sciences being 6 compared to that of political science network being 5.

Reciprocity. Reciprocity was comparable between domains: 0.198 in natural science and 0.193 in political science. This indicates that approximately 20% of connections are mutual in both cases, reflecting a mix of broadcast and reciprocal behavior for both domains.

Assortativity. Degree assortativity differs slightly between the two domains. Natural science network shows weak disassortativity ($r = -0.050$), consistent with broadcast-style diffusion in which central influencers get linked to less-connected users. In contrast, political science network exhibits a mildly positive assortativity ($r = 0.029$), hinting a modest tendency for influencers to engage with similarly connected peers. Although both values are near zero, the difference in direction may reflect subtle structural distinctions in engagement style.

Clustering The natural science network has clustering coefficients as high as 0.52 and the political science network, 0.6. These values show that both networks have prominent hubs and moderately clustered subgraphs.

Distributions. Figures 13 and 14 display the degree distributions and the follower count distributions (both on log-log scales) for each cumulative network. These plots further illustrate scale-free-like patterns and heterogeneous reach among influencers in both domains.

Overall, the cumulative network analysis reveals that while both domains share structural similarities such as sparse connectivity and moderate reciprocity the observed differences are relatively minor and do not support strong domain specific conclusions. This highlights the need for further analysis using dynamic simulations and community detection algorithms.

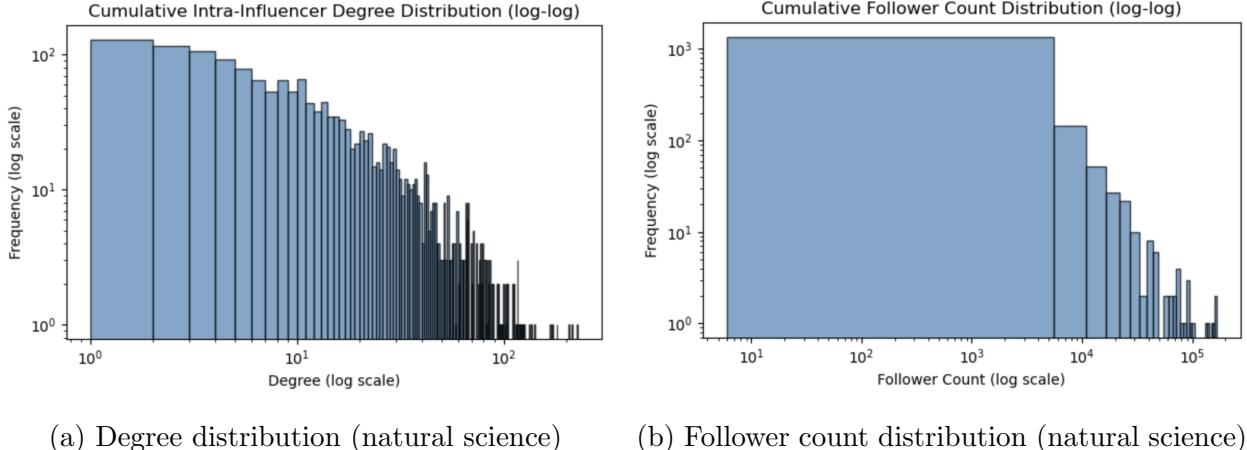


Figure 13: Cumulative network distributions for natural science influencers. The left plot shows internal connectivity variation; the right reflects external audience size.

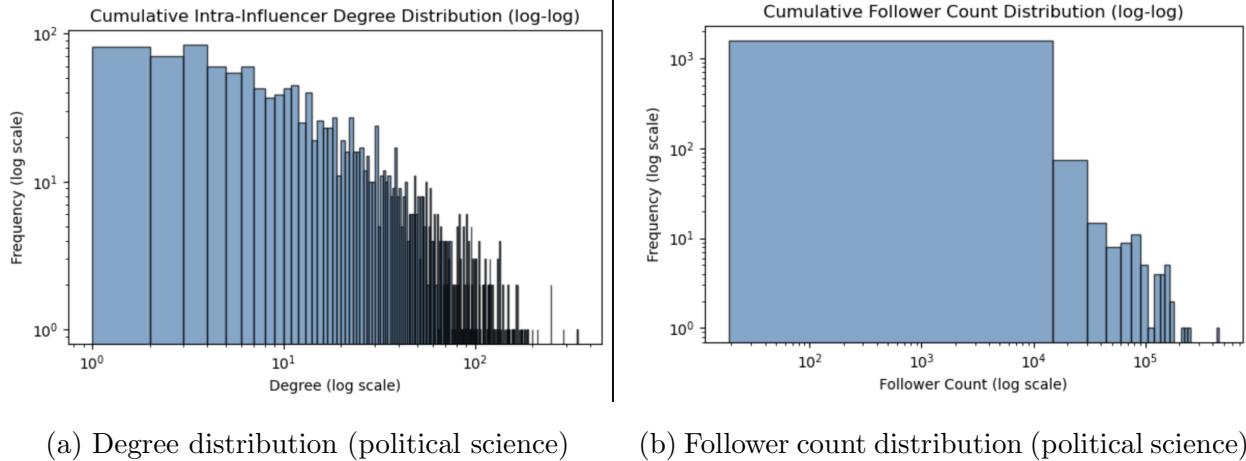


Figure 14: Cumulative network distributions for political science influencers. The left plot shows internal connectivity; the right illustrates external reach.

6.3 Results of Diffusion Model

In this section, it is aimed to show the influence of diffusion parameters and identify the most effective influencers in terms of their reach across all simulations.

Parameter Sensitivity. Fixing $\beta = 3$, we varied α and ε to study the sensitivity of the cascade size. The resulting heatmap (Figure 15) reveals that both stronger reinforcement (higher α) and higher probability floors lead to larger average cascades, highlighting a nonlinear interaction between internal motivation and baseline influence.

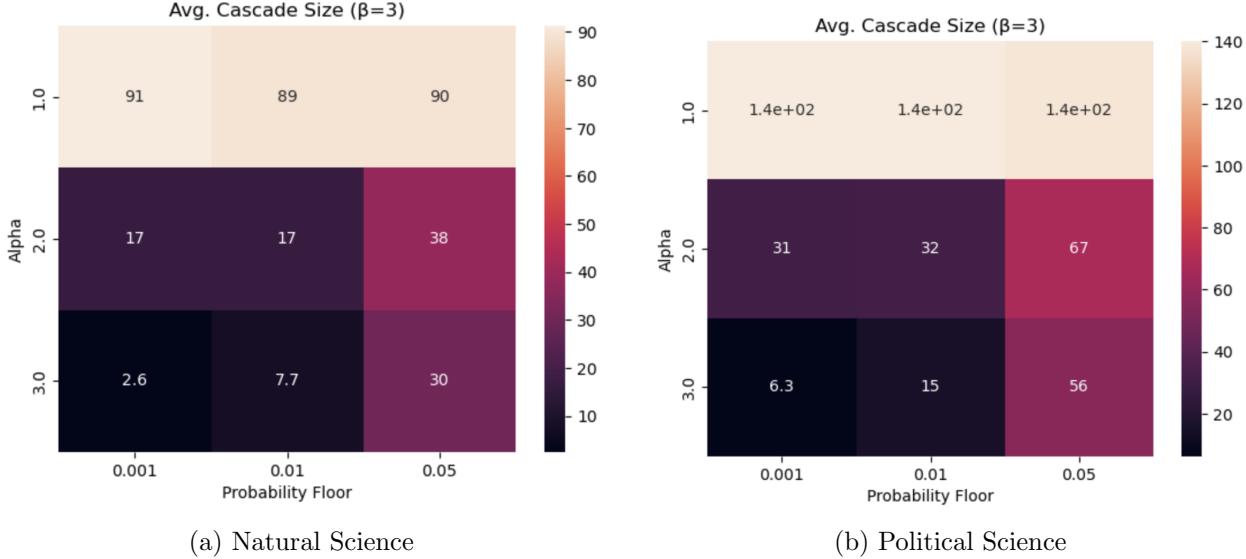


Figure 15: Heatmap of average cascade sizes across α and probability floor ε ($\beta = 3$ fixed).

Spread Potential across Parameters. Figure 16 shows a horizontal bar plot of average spread size for each parameter combination. Some parameter sets yield significantly greater reach, suggesting that real world diffusion may depend heavily on both internal network structure and thresholds for action. Political science packs tend to result in larger spread in all settings. This suggests the need to search for users that cause larger diffusion.

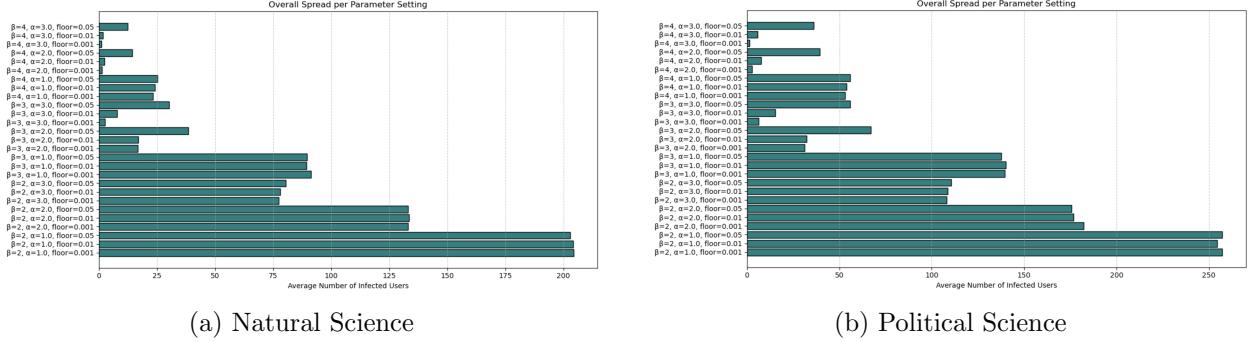


Figure 16: Average number of infected users under each parameter configuration.

Global Influencer Impact. Aggregating all simulation runs, we computed each influencer’s average spread. Figure 17 shows the distribution of average infections per influencer. The majority caused limited spread, but a few consistently trigger large cascades, acting as key spreaders. As we can locate a subset of users causing larger spread regardless of the diffusion properties we can consider them as the main actors of an echo chamber.

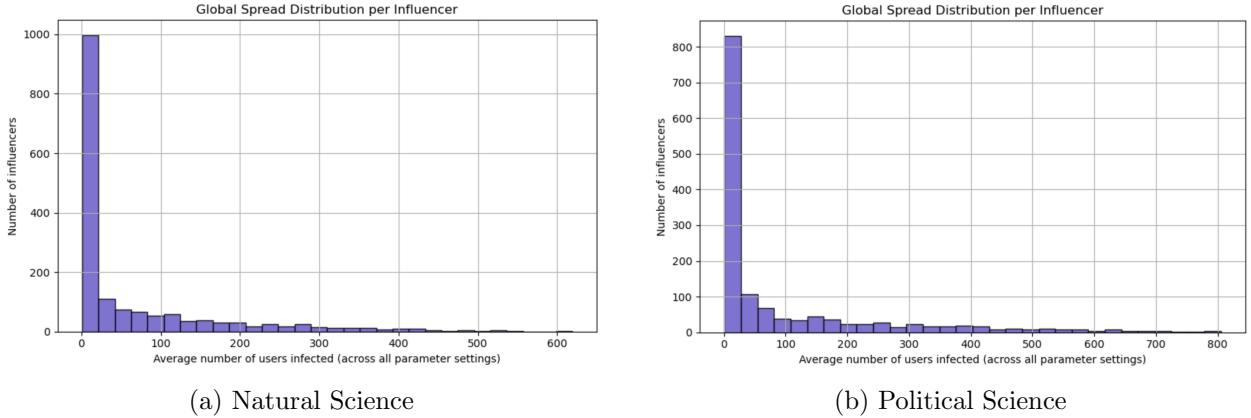


Figure 17: Histogram of average spread size per influencer across all parameter runs.

Echo Chamber vs. Bridge Dynamics. How in pack and out of pack degrees relate to average diffusion spread is examined. As shown in Figure 18, there is a strong positive correlation between intra-pack degree and spread in both natural science ($\rho = 0.786, p < 10^{-10}$) and political science ($\rho = 0.787, p < 10^{-298}$), indicating that tightly knit clusters can amplify spread within their own communities.

However, out of pack degree shows an even stronger correlation in the political science network ($\rho = 0.830, p \approx 0$), compared to natural science ($\rho = 0.665, p \ll 0.001$), suggesting that bridging connections play a greater role in cross community influence in political domains. These find-

ings support the hypothesis that echo chamber reinforcement and bridging dynamics jointly shape diffusion, but the bridging dynamics may be more critical for broad reach in politically diverse environments. These findings support our second hypothesis (H2), demonstrating that interaction networks within starter packs exhibit clustering (a possible playground for echo chamber dynamics). However, they also highlight the critical role of out of pack connections in enabling broader diffusion, particularly in political domains.

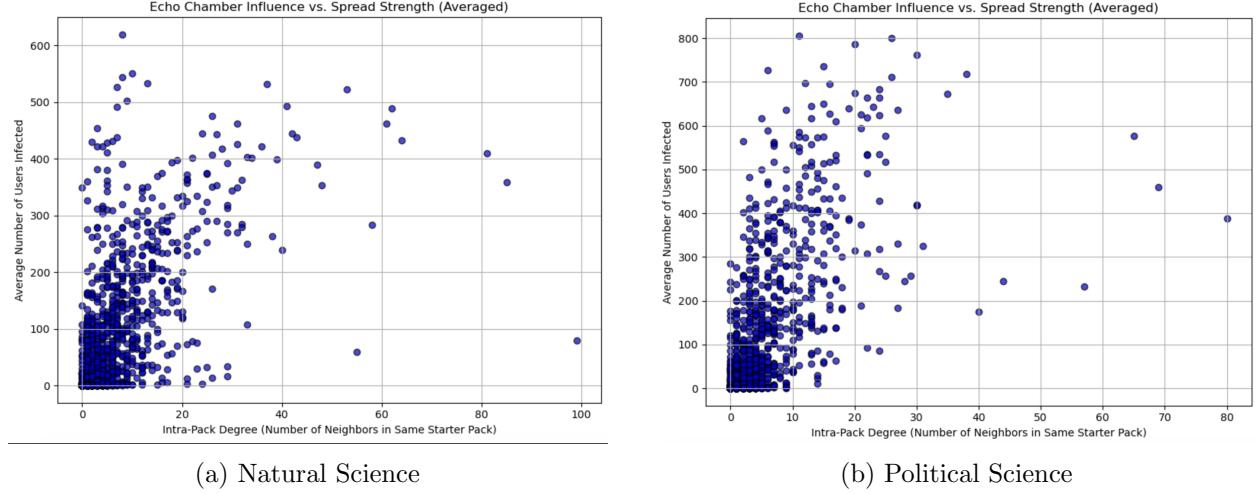


Figure 18: Relationship between in pack connectivity and diffusion.

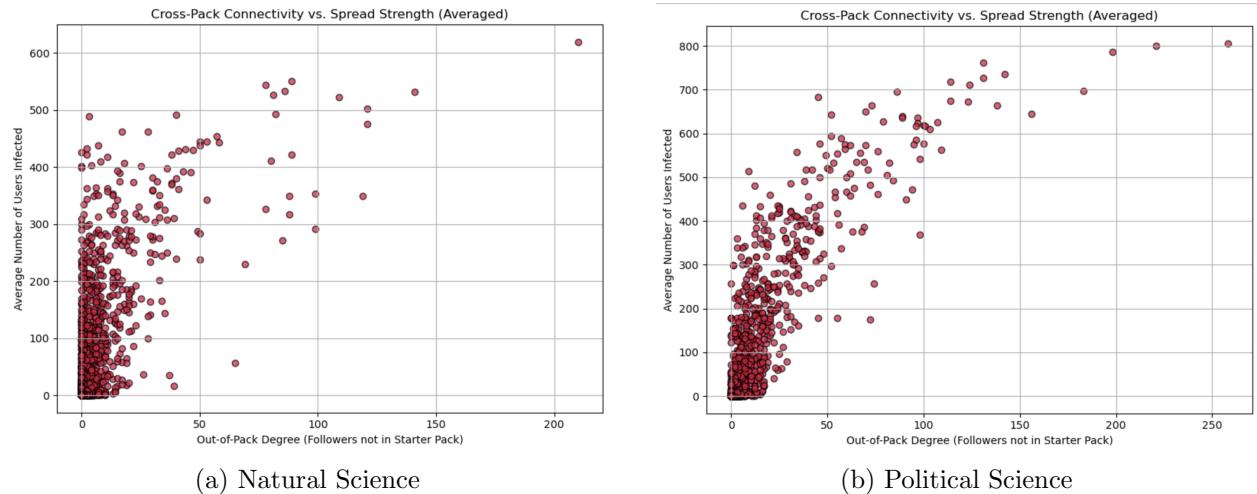


Figure 19: Relationship between out of pack connectivity and diffusion.

Cross Pack Influence. Beyond raw spread, we measured how many distinct starter packs each influencer reached on average. As shown in Figure 20, most influencers tend to be connected within their own communities, but few cause high cross pack diffusion, indicating bridge roles in the network. This correlation is intuitively expected as influencers with diverse audiences can reach broader. Political science community has larger cross pack reach in comparison to natural science community.

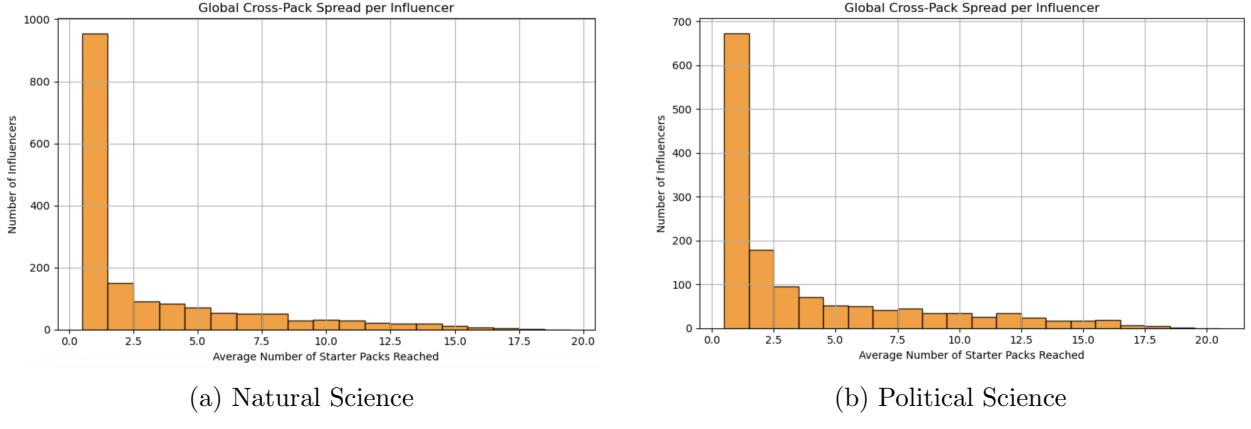


Figure 20: Distribution of average number of starter packs reached per influencer.

Cross pack interaction rates. CPIR and Bridge Scores are computed, users were grouped by domain. Mean CPIR values were compared between the STEM and Political Science groups. To assess the statistical significance of the difference in openness, an independent two-sample **t-test** is conducted. Additionally, **Cohen’s d** [12] was calculated to estimate the effect size of the observed difference. The mean and standard deviation of the Cross-Pack Interaction Rate (CPIR) for each group are reported in Table 2.

Table 2: Mean and standard deviation of CPIR by domain

Domain	Mean CPIR	Std. Dev.
STEM	0.032	0.063
Political Science	0.097	0.114

The average CPIR for Political Science users was three times higher than that of STEM users, suggesting a markedly greater tendency toward cross starter pack engagement in the Political Science domain.

Users were further classified based on their *Normalized Bridge Score*, which quantifies the diversity of starter packs each user engaged with. Table 3 summarizes the distribution of users across key bridge types.

Table 3: Bridge classification distribution across domains

Bridge Type	STEM Users	Political Science Users
Isolate	High (largest group)	Much lower
Bridge	796 (42%)	1,319 (70%)

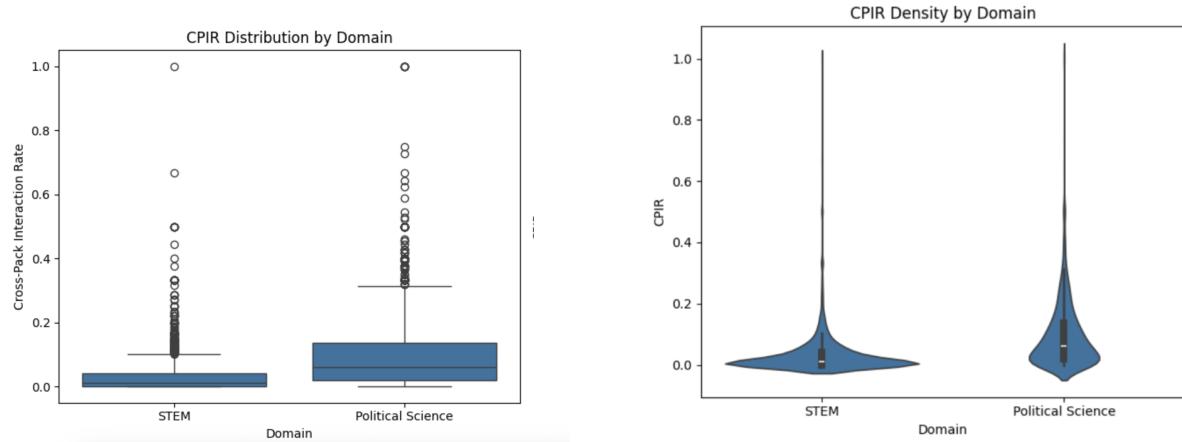
This distribution suggests that Political Science users are significantly more likely to act as *intellectual bridges*, engaging with multiple starter packs, whereas STEM users tend to remain more isolated within their own packs.

To evaluate the significance of the observed difference in CPIR between the two domains, an independent two-sample **t-test** was conducted:

- **T-statistic:** -21.601
- **P-value:** < 0.001
- **Effect Size (Cohen's d):** -0.705

A t-statistic of -21.601 indicates that the observed difference in CPIR is highly unlikely to be due to chance. The negative sign reflects the direction of the effect, with STEM users exhibiting significantly lower CPIR values. The p-value confirms that this result is statistically significant at any conventional threshold.

The Cohen's d value of -0.705 indicates a medium to large effect size, reinforcing the conclusion that the difference in cross-pack interaction behavior is both statistically and practically meaningful. The negative value again highlights the relative insularity of STEM users compared to their Political Science counterparts.

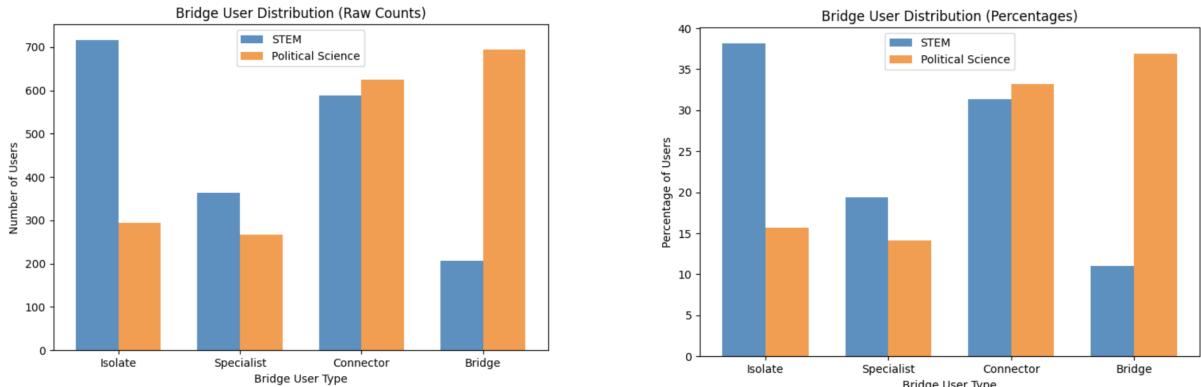


(a) **CPIR Distribution by Domain.** Box plot illustrates a higher and more varied Cross-Pack Interaction Rate (CPIR) values of political science users compared to STEM users, whose CPIR values are lower and more tightly concentrated.

(b) **CPIR Density Comparison.** Density plot shows CPIR values of STEM users being sharply clustered near zero, indicating limited cross-pack engagement, whereas Political Science users demonstrate a broader distribution with more individuals exhibiting high CPIR values.

Figure 21: Relationship between out of pack connectivity and diffusion.

Figure 21 (a) and Figure 21 (b) present the comparison of the CPIRs between the users in STEM and Political Science domains. In Figure (a), it can be observed that Political Science users exhibit a larger median CPIR than STEM users, while in Figure (b), it can be observed that STEM users' CPIR values are sharply concentrated near zero, confirming limited cross-pack engagement, whereas Political Science users show a broader, more gradual distribution. Thus, both of the figures indicate that Political Science users are more intellectually open and outward-facing in their interaction patterns compared to STEM users.



(a) Bridge Category Distribution.

Proportion of users in each bridge classification. STEM users are largely classified as *isolates*, while Political Science users are more commonly *bridges*.

(b) Stacked Bridge Category Comparison.

Comparison of bridge category membership shows more Political Science users in the *bridge* and *connector* categories.

Figure 22: **Bridge User Role Comparison.** Comparison of normalized bridge category distributions across STEM and Political Science domains.

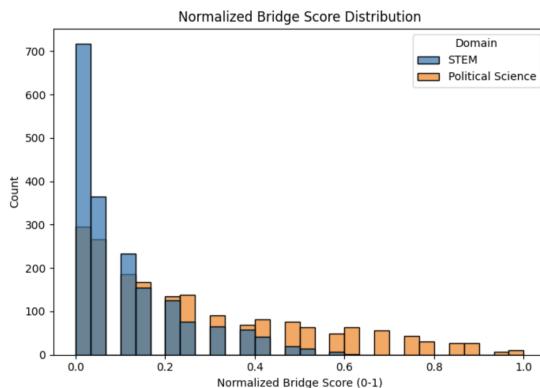


Figure 23: **Normalized Bridge Score Density.** Distribution of normalized bridge scores across domains. STEM users are sharply concentrated near zero, while Political Science users show a more diverse spread.

Figure 22 (a) and Figure 22 (b) illustrate the distribution of users across four bridge categories based on their normalized bridge scores. In Figure 22 (a), STEM users are predominantly classified as *isolates*, indicating limited engagement beyond their own starter packs. In contrast, Political Science users are more frequently found in the *bridge* category, suggesting a greater propensity to engage with multiple distinct communities.

The underlying distribution of normalized bridge scores is further detailed in Figure 23. This figure reveals a sharp concentration of STEM users near a score of 0, reinforcing the notion of limited cross-pack interaction diversity. On the other hand, Political Science users exhibit a more even spread, with many individuals scoring in the mid to high range—highlighting a broader and more inclusive interaction pattern across starter packs.

Given results ($p < 0.001$ and Cohen's $d = 0.705$), **H4** is confirmed: Political Science users are significantly more outward-facing than STEM users.

6.4 Starter-Pack Cohesion vs Fragmentation

Figure 24 shows, for each starter pack, the fraction of interactions that stay inside the pack (in-pack density, blue) versus those that go outside (cross-pack edge ratio, orange).

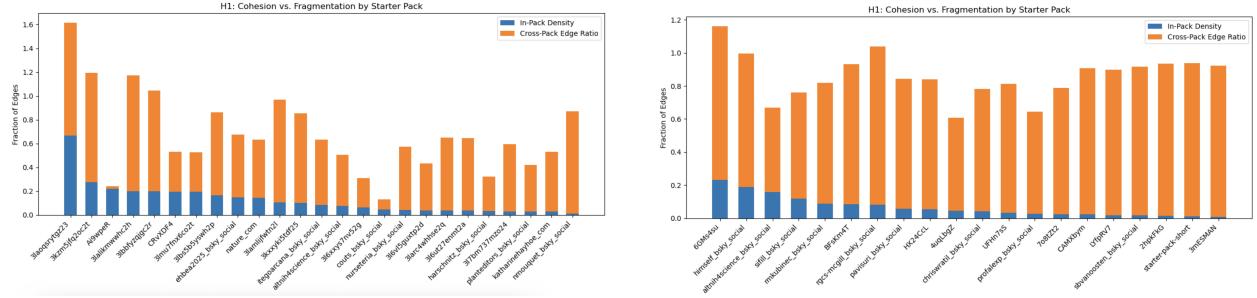


Figure 24: In-pack density vs. cross-pack edge ratio by starter pack, for (a) Natural Science and (b) Political Science domains.

For the Natural Science cumulative graph, a Pearson correlation between in-pack density and cross-pack ratio yields

$$r = 0.36, \quad p = 0.074.$$

This weak positive trend—though not statistically significant at the 0.05 level suggests that, if anything, more densely connected science packs also exhibit slightly higher external connectivity.

In contrast, the Political Science network shows

$$r = -0.17, \quad p = 0.469,$$

indicating a small negative but non-significant relationship between cohesion and fragmentation.

Since neither domain exhibits a strong nor significant inverse correlation, it is not clearly evident that there is a trade-off (high internal cohesion paired with low external connectivity) as suggested by **H1**. Instead, the cohesion and fragmentation patterns appear largely independent in both Natural and Political Science starter pack networks.

6.5 Results of Community Detection

To empirically evaluate the characteristics of echo chambers in studied starter-pack domains, it is conducted both modularity-based and flow-based community detection across two themes: *Natural Sciences* and *Political Science*. Each analysis tests the extent to which starter packs produce meaningful internal cohesion or fragmentation.

Natural Sciences Starter Pack Network

- Modularity Analysis:** Analysis has begun with applying Leiden algorithm on the binarized network (edge weights are reduced 0 or 1), this has resulted in modularity score $Q_{\text{real}} = 0.5558$.

This score is compared to the null network distribution obtained out of 100 times of random generation ($\mu = 0.19, \sigma = 0.002$). There is not an instance of a null network having a larger modularity score than the observed one, meaning that $p\text{-value} < 0.0001$. That being said, statistically significant deviation occurs from the mean of null distribution. This confirms the existence of a strong community structure such that topological characteristics has produced communities distinctly identifiable.

The same analysis was applied on the weighted network as well. A slight increase in the mean of the null distribution was observed out of 100 runs, but the significant deviation persisted ($Q_{\text{real}} = 0.5683, \mu = 0.20$).

2. **Infomap Analysis:** Infomap was run over 20 iterations on the same weighted network. Community partitions were evaluated using Normalized Mutual Information (NMI), yielding an average score of $\text{NMI} = 0.9641$, indicating high stability.

Together, these findings support **Hypothesis H2**: the Natural Science starter pack network shows strong internal clustering consistent with echo chamber structures. The strong modularity and high Infomap stability both point to persistent internal consensus.

Political Science Starter Pack Network

1. **Modularity Analysis:** The Leiden algorithm has been applied on both the binarized network and weighted network. This time the modularity score turns out to be $Q_{\text{real}} = 0.4060$, which means that the topological community partition has dampened. Yet, the comparison with the null distribution still points to a significant deviation ($\mu = 0.1465, \sigma = 0.002$). Repeating the application on weighted network has resulted in a slight increase but yet a large deviation from the mean of null distribution. Thus, $p\text{-value} < 0.0001$ and deviation is proven to be statistically significant.
2. **Infomap Analysis:** Similar to the implementation of Natural Science network, Infomap has been applied iteratively 20 times, resulting in an average $\text{NMI} = 0.9250$. This indicates to high stability within communities across the network.

Across both thematic networks, *modularity* and *Infomap* demonstrates a consistent structure in which starter-pack influencers do not interact at random but form statistically robust, self-reinforcing clusters. Natural Sciences shows a well-sealed echo-chamber formation whereas Political Sciences exhibit a lower yet significant partition along with slightly less NMI score.

Overall, the results **support H2** (echo-chamber-like cohesion inside starter packs) and partially bolster **H1** and **H3**: interaction is concentrated within packs and structurally biased toward central figures, while the extent of fragmentation varies by domain.

Community Analysis on Binarized Networks by Modularity

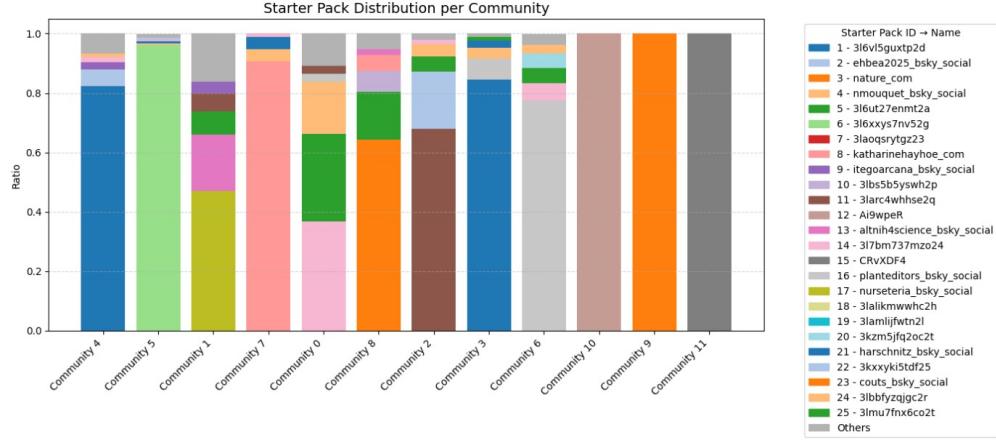


Figure 25: Natural science communities and their user distributions computed considering a binary network.

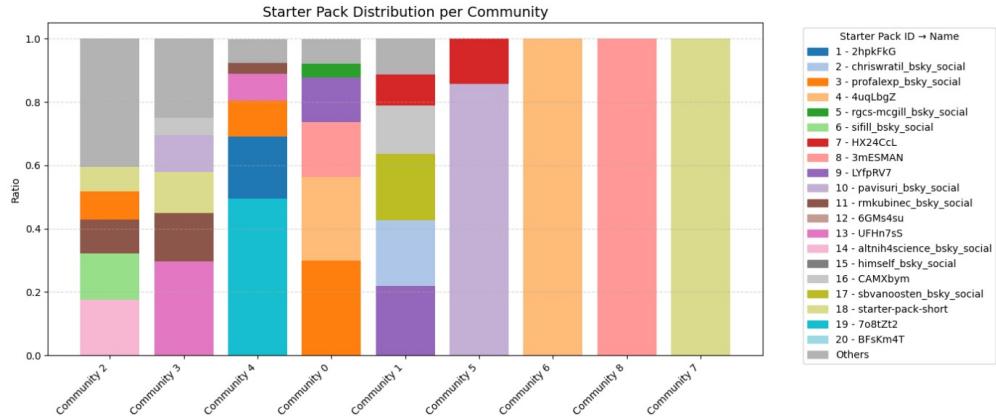


Figure 26: Political science communities and their user distributions computed considering a binary network.

Political science communities are more evenly composed of members from different starter packs. This aligns with our earlier findings, which showed that the political science network exhibits a higher degree of cross-pack spread. These results suggest that not only does the political domain facilitate broader diffusion, but it also features multiple users within the same starter pack who successfully reach distinct communities. This dynamic may contribute to echo chamber tendencies, as influential users have the capacity to spread information both within their starter packs and across different communities, potentially sustaining in group properties while extending their reach beyond the immediate community.

Community Analysis on Weighted Networks by Modularity

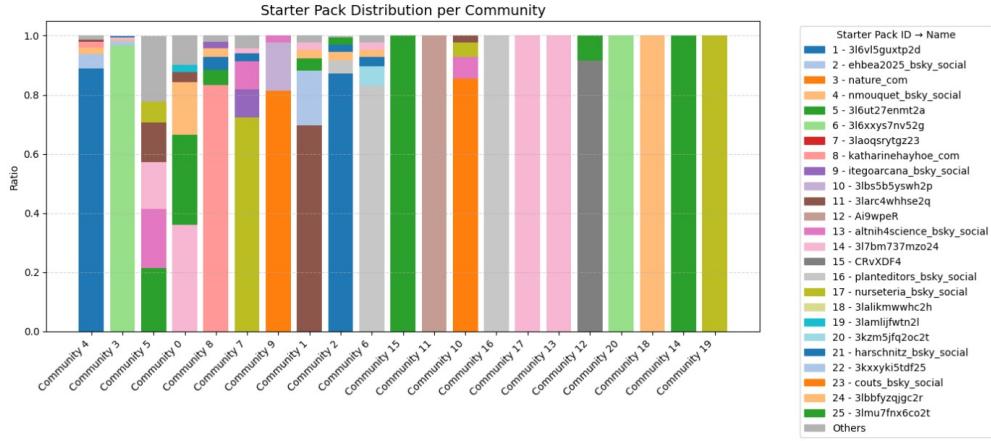


Figure 27: Natural science communities and their user distributions computed considering a weighted network.

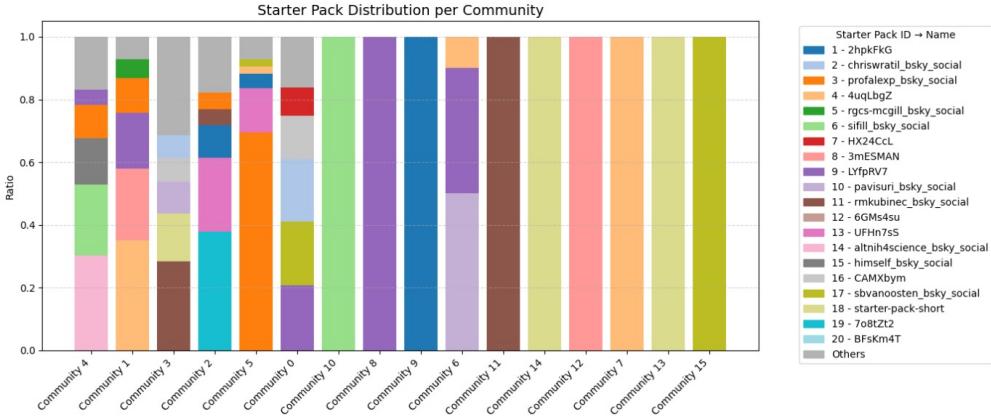


Figure 28: Political science communities and their user distributions computed considering a weighted network.

These weighted distributions further highlight the structural differences between domains. Natural science communities tend to be tightly aligned with individual starter packs with limited cross pack exposure. In contrast, political science communities are more compositionally diverse, with overlapping starter pack representation that allows for broader spread.

Community Analysis on Weighted Networks by Infomap

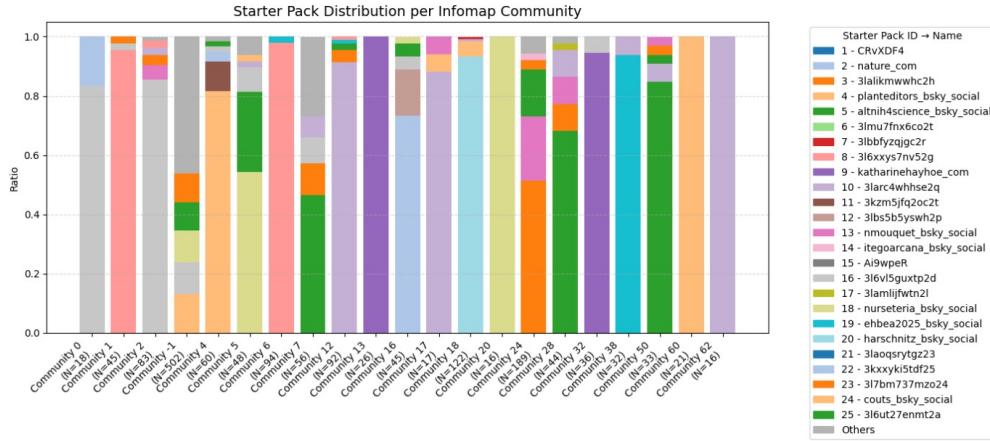
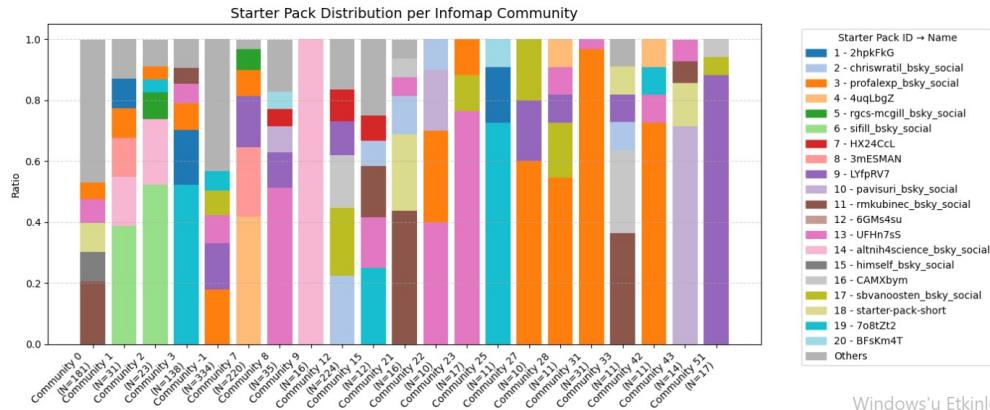


Figure 29: Natural science.



Windows'u Etkinle

Figure 30: Political science.

The communities and their user distributions found by infomap technique are shown above. The results obtained using the Infomap algorithm further support our earlier observations regarding community composition. As seen in Figure 23, natural science communities tend to be more homogeneous, with many communities being dominated by a single starter pack. In contrast, political science communities display greater internal diversity, often composed of users from multiple starter packs. This pattern is consistent with the weighted network results shown earlier and reinforces the interpretation.

The binarized edge network shows 11 distinct communities through modularity optimization, with clear separation between STEM and Political Science domains. Community 0 exhibits the highest variation in starter pack communities, displaying a diverse mix of multiple different starter packs rather than being dominated by a single one. This heterogeneous structure suggests that Community 0 serves as a central hub where different thematic areas converge. STEM starter packs with different focuses: one focuses on invasion biology and invasion ecology outreach, another on general ecology, and another on people who are curious and scientifically inclined, along with those involved in science communication was included in the Community 0.

A notable observation is that Community 5 in the weighted network contains the same starter packs as Community 0 in the binary network, but with different proportional representations reflected in the color sizes. This suggests that while the core membership of this diverse community remains stable across both approaches, the weighted analysis reveals the varying degrees of user overlap intensity between the starter packs, providing a more nuanced understanding of internal community structure without changing the fundamental grouping.

In political science starter packs Community 0 exhibits the highest variance and includes starter packs focused on political theorists, philosophers, scientists, and sociologists. When examining the weighted edges version of the Political Science communities , the same groups are observed but with altered proportional sizes like the STEM communities.

A very different picture is shown by the Infomap method, with many more communities identified than by the previous approaches. Unlike the modularity methods, in which 11–19 communities were found, the network is broken into much smaller, more specialized groups by Infomap. The most diverse community is identified as Community -1, in which the highest variation in composition is observed, and ecology starter packs are again included among its mix. This increased splitting reveals smaller academic niches and specialized communities that had been grouped together into larger, broader categories by the other methods.

The Infomap analysis of the Political Science network demonstrates patterns consistent with those observed in the STEM network, exhibiting similar fragmentation into numerous specialized communities and revealing fine-grained organizational structures within the disciplinary domain. Together, these findings provide strong support for H2: starter packs not only exhibit reciprocal interactions but also form tightly-clustered, stable, and (where measured) topically homogeneous communities—hallmarks of echo-chamber behavior.

6.6 Results Of Gender Analysis through Regression Discontinuity Design

Domain-Specific Findings and Statistical Results

1. **Natural Science Domain:** Distinctive gender composition patterns were shown by the Natural Science network as well, The analysis is conducted over 1,677 users and 8,075,727 total followers (average: 4,816 followers per user). An average female ratio of 0.359 (SD = 0.032, range: 0.31-0.42) was observed with a gender homophily index of 0.490 (SD = 0.032).

The specialized log gender ratio metric $\log(P(F-Sp))/\log(P(F))$, where $P(F) = 0.38$ represents the Bluesky platform baseline, is calculated to reveal that 1525 starter packs exhibited a fea female representation above the platform with an average log ratio of 1.062. Systematic over-representation of female users in Natural Science communities relative to platform demographics was suggested by this finding.

No statistically significant discontinuities in any outcome variables are produced by RDD analysis across all three thresholds, consistent with previous findings in social network threshold research [6]. All treatment effects approximate to zero with p-values = 1.000, indicating absence of threshold effects in the Natural Science domain. McCrary density tests confirmed no evidence of manipulation around any thresholds (all p-values = 1.000).

2. **Political Science Domain — Significant Threshold Effects:** The Political Science domain is found to encompass 20 starter packs with 1,816 users and 12,719,984 total followers (average: 7,004 followers per user), with different patterns being demonstrated. An average

female ratio of 0.372 ($SD = 0.035$, range: 0.31-0.46) is observed with a gender homophily index of 0.480 ($SD = 0.031$).

It is revealed by log gender ratio analysis that 13 out of 20 starter packs exhibited above-platform representation (average log ratio: 1.025), with lower over-representation compared to Natural Science but still positive bias being indicated.

Statistically a significant threshold effect at the Female Minority threshold (33% female) is demonstrated by Political Science, following the sharp RDD framework established by Thistlethwaite and Campbell [10]. A discontinuity of 1,269,401 followers ($SE = 442,437$, $p = 0.019$) at the 0.05 bandwidth specification is shown by the total follower count. Marginally significant effects with 91.55 nodes ($SE = 67.48$, $p = 0.208$) at 0.05 bandwidth were demonstrated by network size. The first empirically documented gender-based threshold effect in social network formation is represented by this finding, with direct evidence for Hypothesis **H2** regarding echo chamber characteristics being provided.

Network Correlation Analysis

Striking domain differences in how gender composition relates to network clustering were revealed by sophisticated correlation analysis. An extremely strong negative correlation between female ratio and homophily ($r = -0.950$) is demonstrated by Natural Science starter packs, with higher female representation coinciding with lower overall gender homophily being suggested, potentially indicating more integrated gender dynamics. A similarly strong but slightly weaker negative correlation ($r = -0.911$) is shown by Political Science, with domain-specific moderation effects.

Weak correlations between network size and homophily (Natural Science: $r = -0.191$, Political Science: $r = -0.016$) are demonstrated by both domains, with echo chamber dynamics operating independently of scale effects.

Homophily Distribution and Echo Chamber Evidence

Hypothesis H2 is evidenced by a detailed homophily analysis across both domains. It is shown by homophily classification that all starter packs exhibited moderate homophily (0.4-0.6 range), with strong homophily (> 0.6) being found in 0% of both domains, moderate homophily in 100% of both domains, and weak homophily (< 0.4) in 0% of both domains.

That all starter packs exhibit measurable echo chamber characteristics while avoiding extreme segregation was indicated by this universal moderate homophily pattern. Structural constraints that prevent complete polarization while maintaining sufficient clustering for internal consensus formation were suggested by the bounded homophily range (0.4-0.6).

Comparative Domain Analysis

Table 4: Comprehensive gender composition and network characteristics by domain

Metric	Natural Science	Political Science
Demographics		
Starter Packs	19	20
Total Users	1,677	1,816
Total Followers	8.1M	12.7M
Avg Followers/User	4,816	7,004
Gender Composition		
Average Female Ratio	0.359	0.372
Gender Homophily Index	0.490	0.480
Packs with Ratio > 1	78.9%	65.0%
Log Gender Ratio (mean)	1.062	1.025
RDD Results		
Significant Effects	0	1
Maximum Effect Size	N/A	1.27M followers
P-value (minimum)	1.000	0.019
Correlations		
Female Ratio and Homophily	-0.950	-0.911
Network Size and Homophily	-0.191	-0.016

Robustness Checks and Validation

Comprehensive manipulation testing is conducted using McCrary (2008) methodology [6]. All thresholds with discontinuity = 0.0000, test statistic = 0.000, p = 1.000 are shown by Natural Science, while test statistic = 0.000, p = 1.000 (no manipulation) is demonstrated by Political Science Female Minority threshold. Continuity of pre-treatment characteristics is verified through t-tests comparing means above and below thresholds, with no discontinuous baseline characteristics ($p > 0.05$) being shown by network size and total followers.

Robustness of key findings was confirmed by bandwidth sensitivity analysis, following best practices established by Imbens and Kalyanaraman [4]. Significance at 0.05 bandwidth ($p < 0.05$) is maintained by the Political Science follower effect, with effect size decreasing but remaining substantial at larger bandwidths (0.10-0.25). Optimal bandwidth selection procedures recommended by Imbens and Kalyanaraman [4] are employed to ensure robustness across multiple specifications.

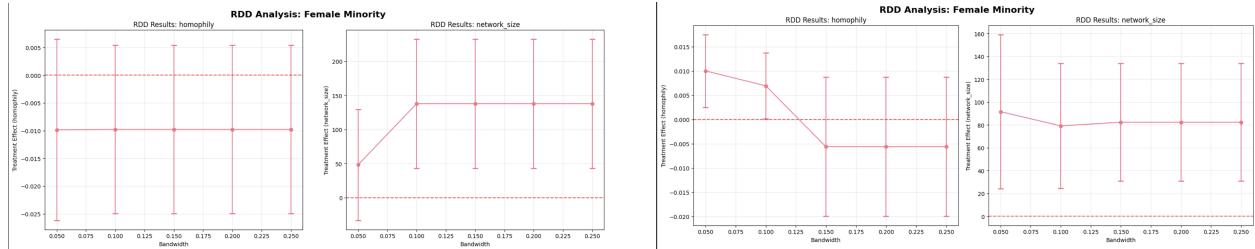
Evidence and Network Characteristics: Visuals



(a) Natural Science comprehensive dashboard showing gender distribution patterns, network size relationships, and correlation matrix revealing strong negative relationships between female ratio and homophily.

(b) Political Science dashboard demonstrating higher gender variability, stronger network connectivity patterns, and more dispersed homophily distribution.

Figure 31: Comprehensive RDD Analysis Dashboard comparing gender distributions and network properties across domains.



(a) Natural Science RDD results for Female Minority threshold showing treatment effects across bandwidth specifications. Non-significant effects ($p > 0.05$) with confidence intervals spanning zero were demonstrated by all outcome variables.

(b) Political Science RDD results revealing significant threshold effects at Female Minority level. Clear discontinuity with large effect size was shown by total followers, with domain-specific gender dynamics being indicated.

Figure 32: Regression Discontinuity Design results for Female Minority threshold (33% female) across domains.

Substantial empirical evidence supporting Hypothesis H2 regarding echo chamber characteristics within starter pack networks was provided by the gender analysis. Systematic gender-based clustering that supports internal consensus formation while avoiding extreme segregation was indicated by universal moderate homophily across all 39 starter packs. That gender composition operates as a structural organizing principle, creating qualitatively different network dynamics below critical threshold values, was demonstrated by the significant threshold effect detected in Political Science at 33% female representation.

Evidence for **Hypothesis H3** regarding structural bias and hierarchical dominance is provided by the detection of significant threshold effects specifically at the Female Minority level. Those networks with less than one-third female representation exhibit fundamentally different influence

patterns, suggested by the 1.27 million follower discontinuity, with critical mass theory being corresponded to where minority representation below 30-35% fails to achieve meaningful influence.

That Bluesky's starter pack mechanism creates structural constraints that prevent complete polarization while maintaining sufficient clustering for echo chamber dynamics was suggested by the bounded homophily pattern (0.4-0.6 range) observed across all starter packs. Contribution to theoretical understanding of how platform design features can moderate extreme polarization while preserving community formation benefits was made by this finding.

References

- [1] I. Blekanov, S. S. Bodrunova, and A. Akhmetov. Detection of hidden communities in twitter discussions of varying volumes. *Future Internet*, 13(11):295, 2021.
- [2] Damon Centola. The spread of behavior in an online social network experiment. *Science*, 329(5996):1194–1197, 2010.
- [3] Fredrik Erlandsson, Piotr Bródka, Anton Borg, Henrik Johnson, and David Camacho. Finding influential users in social media using association rule learning. *Entropy*, 18(5):164, 2016.
- [4] Guido Imbens and Karthik Kalyanaraman. Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies*, 79(3):933–959, 2012.
- [5] A. P. Logan, P. M. LaCasse, and B. J. Lunday. Social network analysis of twitter interactions: A directed multilayer network approach. *Social Network Analysis and Mining*, 13(1):65, 2023.
- [6] Justin McCrary. Manipulation of the running variable in the regression discontinuity design: a density test. *Journal of Econometrics*, 142(2):698–714, 2008.
- [7] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69(2):026113, 2004.
- [8] A. R. Pamfil, A. Saebi, and D. S. Bassett. Null models for community detection in networks. *arXiv preprint arXiv:2501.11605*, 2024.
- [9] Martin Rosvall and Carl T Bergstrom. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105(4):1118–1123, 2008.
- [10] Donald L Thistlethwaite and Donald T Campbell. Regression-discontinuity analysis: an alternative to the ex post facto experiment. *Journal of Educational Psychology*, 51(6):309–317, 1960.
- [11] Sho Tsugawa and Kazutoshi Kito. Retweets as a predictor of relationships among users on social media. *PloS one*, 12(1):e0170279, 2017.
- [12] Thomas W Valente and Kayo Fujimoto. Bridging the gap: Strategic use of group-level ties to build social capital. *Social Networks*, 35(3):370–380, 2013.
- [13] Jorge Valverde-Rebaza and Alneu de Andrade Lopes. Exploiting behaviors of communities of twitter users for link prediction. *Social Network Analysis and Mining*, 3(4):1063–1074, 2013.

- [14] Zhi Yang, René Algesheimer, and Claudio J Tessone. A comparative analysis of community detection algorithms on artificial networks. *Scientific Reports*, 6:30750, 2016.
- [15] X. Zhang and G. Chen. A null model approach to testing community significance in networks. *Statistica Sinica*, 27(3):1633–1656, 2017.