University of Konstanz

Department of Politics and Public Administration

POL-30110: Network Science of Socio-Economic Systems

The Cinematic Universe of Media Polarization: Structural Analysis of Movie Reviewer Community

Name: Evgeniia Tcoi

Matriculation Number: 1357222

Study Program: MSc. Social and Economic Data Science

Email: evgeniia.tcoi@uni-konstanz.de

April 9, 2025

Abstract

This study investigates the structural dynamics of online movie reviewer communities using network science. By modeling user-movie interactions as a bipartite network and applying a statistically validated projection, we construct a unipartite reviewer-reviewer network to reveal patterns of influence and community formation. A stratified sampling approach preserves the skewed activity distribution typical of user-generated platforms. The analysis uncovers a highly interconnected network characterized by disassortative mixing and dense clustering, with a small group of prolific reviewers acting as central hubs.

1 Introduction

Online communities built around film and media have become influential agents in shaping both public discourse and commercial outcomes. These reviewer communities often organize around shared interests, such as genre preferences or sociocultural perspectives, forming distinct clusters that influence movie ratings, trends, and visibility. As these communities grow in visibility and influence, they have also become vulnerable to coordinated manipulation. One prominent example is review bombing — a practice in which organized groups attempt to artificially inflate or deflate a film's ratings, often in response to political or social controversies.

These dynamics position reviewer communities not merely as consumers but as active participants in cultural debates. Their activity can amplify polarization, reinforce group identity, and shape public perception of creative content. Understanding the structure and behavior of these communities is critical to assessing how opinion forms and spreads in digital spaces.

Network science offers a robust framework for analyzing these dynamics. By modeling reviewer interactions as networks, it becomes possible to detect patterns of clustering, identify central or influential users, and evaluate the extent of polarization or coordination. In particular, projection techniques allow us to study relationships between reviewers based on shared activity, enabling the identification of tightly-knit subgroups and potential echo chambers.

This study focuses on the structural analysis of the movie reviewer community and addresses the research question:

1. What is the structure of the movie reviewer community as a network?

The central hypothesis is that the reviewer community includes highly active users who function as opinion leaders or influencers.

The dataset used for the analysis and the Jupyter notebooks are available at the GitHub repository.

2 Research Design

In network science (e.g. [Beguerisse Díaz, 2008]) movie reviewer communities are commonly analyzed as bipartite networks.

A bipartite network consists of two disjoint sets of nodes (here, reviewers and movies) where edges exist only between nodes of different types. This structure is widely applicable in affiliation networks, such as user-product interactions or movie-actor relationships.

Bipartite networks are characterized by two distinct degree distributions, one for each set of nodes. For instance, reviewer nodes may vary widely in the number of movies reviewed, while movie nodes vary in how many users reviewed them. An important property of bipartite networks is the absence of closed triangles, as links only connect nodes across sets.

To study community-level interactions, a bipartite network can be projected onto one of its layers. In the context of this study, projecting onto the reviewer layer results in a network where two reviewers are connected if they have reviewed at least one movie in common. This projection allows us to analyze reviewer communities, detect influential users, and identify clusters of similar behavior. However, direct projection from large bipartite networks can result in overly dense graphs with high computational cost and limited interpretability.

2.1 Data Collection and Preprocessing

The Dataset

The dataset used in this study originates from [Rossi and Ahmed, 2015] and contains user-generated movie reviews represented as a bipartite graph. Nodes consist of 69,878 unique reviewers and 10,677 movies. Each edge represents a single rating event, resulting in over 10 million total reviewer-movie connections.

The degree distribution for reviewers follows a heavy-tailed pattern typical of user-generated content systems. A small number of reviewers are extremely active, with the most prolific user contributing 7,359 reviews, while the majority of users reviewed fewer than 50 films. This skew is reflected in the distribution of reviewer degrees and poses challenges for unbiased sampling.

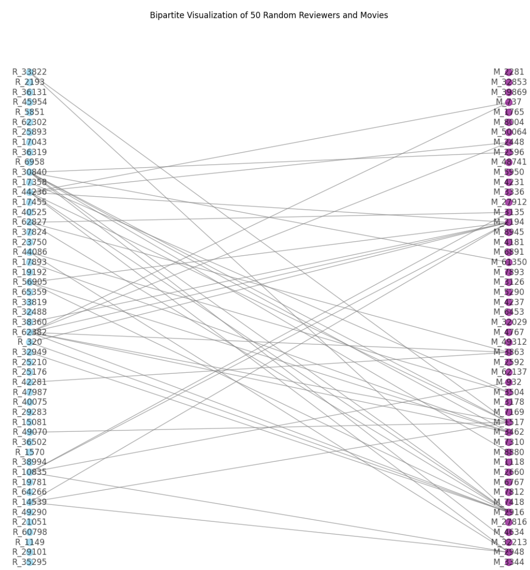

Figure 1: Subgraph visualization of 50 randomly selected reviewers and the movies they rated. Links represent review activity.

Without filtering or simplification, projecting such a dense bipartite network into a unipartite reviewer-reviewer graph would be computationally infeasible. In the worst-case scenario, if all reviewers rated a single shared movie, the resulting projection would contain more than 4.8 billion edges (69878×69878).

Sampling Strategy

To manage complexity and preserve the underlying structure of the network, a stratified sampling approach was employed. Rather than using simple random sampling (which would over-represent high-activity reviewers) the sample was drawn in proportion to reviewer activity, preserving the skewed distribution found in the original data.

The process involved:

- Dividing reviewers into 7 strata using quantile-based binning based on the number of movies reviewed.
- Sampling proportionally from each stratum to retain the original distribution of user activity.
- Targeting a total sample size of 10% of the original reviewer population (6,987 reviewers).

This method ensured that both highly active and low-activity reviewers were represented in the sample, enabling more accurate reconstruction of community patterns in the projected network.

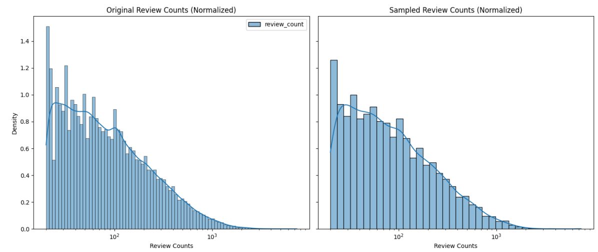

Figure 2: Review Count Distribution: Comparison between the full dataset and the stratified sample.

Resulting Bipartite Network

The final sampled bipartite graph retained the core structural characteristics of the full dataset while reducing computational load. This downscaled version was used as the basis for all subsequent analyses, including statistically validated reviewer projection and community detection.

Table 1: Basic Statistics of the Sampled Bipartite Network

Metric	**Value**
Number of Reviewers	6,987
Number of Movies	10,117
Total Edges (Reviews)	994,563
Average Reviewer Degree	142.45
Average Movie Degree	98.31

Degree Distributions (Log-Log)

The plots in Figure 3 illustrate the degree distribution of reviewers using both a Probability Density Function (PDF) and Cumulative Distribution Function (CDF), each on a log-log scale.

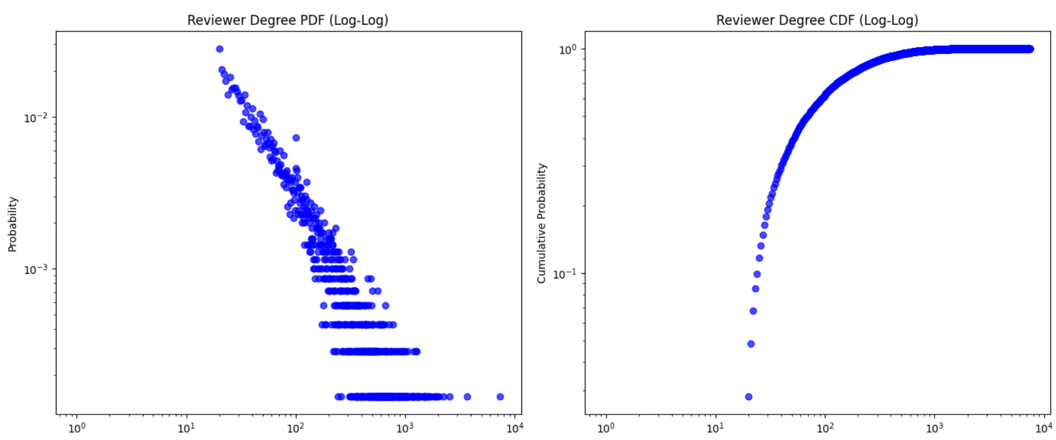

Figure 3: Sample Bipartite Network Degree Distribution: Probability Density Function (left) and Cumulative Distribution Function (right), plotted on log-log scales.

The left plot (PDF) shows a clear heavy-tailed distribution, where the frequency of reviewers sharply declines with increasing degree. The approximately linear trend in log-log space suggests a power-law distribution, a charachteristic of scale-free networks.

The right plot (CDF) further confirms this pattern by showing the cumulative probability of having a given degree or less. The steep early rise and flattening tail indicates that most reviewers have low degrees, while a small subset possesses extremely high degrees, likely corresponding to "super-reviewers" or potential influencers. These users could play outsized roles in shaping opinion dynamics within the network and are of particular interest for later sections on centrality and polarization.

This structure has significant implications for both network robustness and the propagation of influence: while most nodes are sparsely connected, highly active users form critical hubs that may drive community behavior.

Unipartite Projection

In order to analyze the social structure and potential communities among movie reviewers, we projected the bipartite network of reviewers and movies into a unipartite graph of reviewers. In this projection, each reviewer is a node, and an edge between two reviewers indicates that they have reviewed at least one movie in common. However, a direct projection without filtering results in an extremely dense graph, where even weak overlaps (such as sharing a single movie) generate connections. This undermines the interpretability of the resulting network.

To avoid it, the bipartite graph is first represented using a sparse matrix $\mathbf{B} \in R^{m \times n}$, where m is the number of reviewers and n is the number of movies. Each element $B_{ij} = 1$ if reviewer i has reviewed movie j, and 0 otherwise. This binary matrix captures all reviewer-movie interactions.

To construct the reviewer-reviewer projection, we compute the co-occurrence matrix:

$$\mathbf{C} = \mathbf{B} \cdot \mathbf{B}^\top$$

Here, C_{ij} indicates the number of movies both reviewers i and j have rated. This operation captures all pairwise intersections of movie reviews among the reviewers.

After the matrix multiplication, the resulting matrix $\mathbf{C}$ is symmetric and square, with each non-zero entry indicating a potential edge in the projected graph.

To filter spurious edges that arise due to random overlaps, we apply a statistical test based on the hypergeometric distribution. For each pair of reviewers i and j, we define:

- N: Total number of movies in the dataset
- $K = k_i$: Number of movies reviewed by reviewer i
- $n = k_j$: Number of movies reviewed by reviewer j
- k: Number of shared movies between reviewers i and j (i.e., C_{ij})

Under the null hypothesis that both reviewers select movies randomly and independently, the probability that they share at least k movies is given by the survival function of the hypergeometric distribution:

$$P(X \geq k) = \sum_{x=k}^{\min(K,n)} \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}} = \text{Hypergeom.sf}(k-1; N, K, n)$$

If the p-value computed from this test is less than the significance threshold $\alpha = 0.05$, the edge between reviewers i and j is considered statistically significant and retained in the network. This method ensures that only meaningful connections, based on non-random overlap in viewing behavior, are included.

Edges passing the statistical filter are used to construct a reviewer-reviewer unipartite graph $G = (V, E)$, where each node represents a reviewer and each edge signifies a statistically significant shared review pattern. The weight of each edge corresponds to the number of shared reviews.

This filtered graph represents a much sparser and interpretable structure than the naive projection.

3 Analysis and Results

After constructing the filtered reviewer-reviewer network using hypergeometric projection, we performed an in-depth analysis of its structural characteristics. The final graph consists of:

- **6,982 nodes**, representing individual movie reviewers

- **19,118,275 edges**, representing statistically significant co-review relationships

This results in a highly dense network. The average degree of a node (i.e., the average number of meaningful reviewer connections) is computed as:

$$\bar{k} = \frac{2 \times |E|}{|V|} = \frac{2 \times 19,118,275}{6,982} \approx 5,479$$

Such a high average degree suggests that most reviewers share significant overlap in reviewed movies with a very large portion of the reviewer base, indicating the presence of a highly interconnected community.

The degree assortativity coefficient was found to be:

$$r = -0.15$$

This negative value indicates a disassortative mixing pattern, meaning that high-degree nodes (i.e., highly connected = influential reviewers) tend to form connections with low-degree nodes, rather than with other high-degree nodes. This pattern is typical in networks where hubs serve as bridges or broadcasters to less connected individuals, reinforcing their role as potential opinion leaders or information spreaders across subgroups.

Community detection was performed using the Louvain algorithm, a modularity-based method that partitions the network into internally cohesive groups. The algorithm identified 3 communitites. Given that the data initially has no additional node attributes, we can only assume that the clustering may reflect the movies' genres which would indicate that reviewers' preferences in that regard are established in the community.

Degree distribution of the projected graph differs significantly from the original graph.

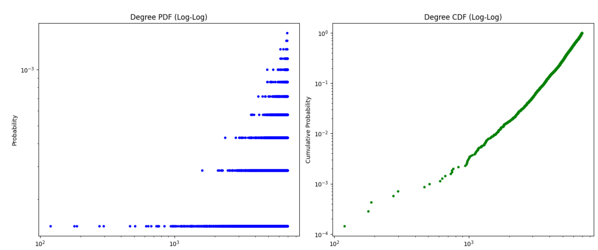

Figure 4: Projected Reviewer Network Degree Distribution: Probability Density Function (left) and Cumulative Distribution Function (right), plotted on log-log scales.

The distribution is heavily right-skewed, with a large concentration of nodes possessing high degrees. This pattern emerges as a direct result of the hypergeometric projection, which retains only statistically significant reviewer co-occurrences based on shared movie ratings.

The skew indicates that a majority of retained reviewers share high overlap in reviewed content with many others (average degree centrality = 0.79), forming a densely connected backbone. Low-degree nodes are rare, suggesting that sparse or weakly associated reviewers were pruned during filtering. This behavior deviates from classical power-law or exponential distributions seen in many natural networks and suggests the presence of a highly cohesive core of influential or prolific reviewers.

4 Conclusion

This study employed network science techniques to analyze the structural dynamics of an online movie reviewer community. Using a stratified sampling approach and a statistically validated projection, we constructed a meaningful reviewer-reviewer network that preserved key structural features while filtering out noise. The resulting network revealed a densely connected community with disassortative mixing, suggesting the presence of central opinion leaders who connect less active users.

Community detection uncovered distinct clusters, likely reflecting shared preferences or reviewing behavior. The network's high connectivity and skewed degree distribution point to a cohesive core of influential reviewers, highlighting their potential role in shaping discourse and amplifying trends.

These findings demonstrate the value of network-based approaches in understanding digital communities. Future work could integrate temporal or semantic data to explore how influence, coordination, and polarization evolve over time.

References

[Beguerisse Díaz, 2008] Beguerisse Díaz, M. (2008). Analysis of a bipartite network of movie ratings and catalogue network growth models.

[Rossi and Ahmed, 2015] Rossi, R. A. and Ahmed, N. K. (2015). The network data repository with interactive graph analytics and visualization. In *AAAI*.