Decentralized Diffusion: Decomposing Information Diffusion in Federated Social Networks

Li Zeng[1], **Sijia Ma**[1], **Seeun Kim**[1], **Giulia Sturlese**[2,3]
[1]Tilburg University, Netherlands
[2]IUSS Pavia School of Advanced Studies, Italy
[3]Leiden University, Netherlands

Abstract

This work examines information diffusion in decentralized social media platforms, specifically on Mastodon due to its federated architecture. Unlike centralized platforms like Twitter, Mastodon distributes content across independently operated yet interconnected servers, known as *"instances"*. In this decentralized environment, inter-instance diffusion plays a critical role in shaping communication patterns. Using a dataset of Mastodon posts, or *"toots"*, related to the Gaza conflict, we analyzed how user-, toot-, and instance-level characteristics drive cross-instance diffusion. Our findings demonstrate that these features collectively shape diffusion, with instance-level characteristics playing a particularly crucial role. This work contributes to the theories of decentralized communication by emphasizing the importance of federated architectures and inter-instance interactions. Practical implications include strategies for platform governance, crisis communication, and misinformation control, with a focus on the potential of decentralized platforms to balance localized interactions with global reach.

Introduction

In recent years, the rise of decentralized social media platforms has been reshaping how information is shared and consumed online (La Cava, Greco, and Tagarelli 2022; Goel, Bakshi, and Agrawal 2022). Unlike traditional, centralized platforms such as Twitter and Facebook, decentralized platforms are built on federated architectures that emphasize distributed control and user autonomy. Mastodon, a leading example of decentralized platforms, operates within a broader ecosystem of the Fediverse, a network of interoperable social media instances that communicate using standardized protocols such as ActivityPub. Each server, or *"instance"*, in the Fediverse allows communities to establish localized norms, moderation policies, and technical configurations while remaining interconnected with other instances. This decentralization fosters diversity in community values and governance but also presents unique challenges in understanding how information propagates across a network of instances without central authority (Raman et al. 2019).

The federated design of platforms like Mastodon enables users to post, share, and interact across instances, creating a rich and dynamic environment for information diffusion. However, this design also introduces complexities absent in centralized systems. For example, differences in moderation practices and user engagement patterns between instances can influence how information spreads within and across the network. Additionally, the absence of central authority raises important questions about the reliability, efficiency, and polarization of information flows. These attributes distinguish decentralized platforms as critical yet under-explored spaces for studying the dynamics of information diffusion, particularly during high-stakes events such as crises.

Understanding information diffusion in the context of decentralized platforms is critical for addressing several societal challenges. During crisis events, timely and accurate information can significantly influence public awareness, collective action, and decision-making (Zeng, Starbird, and Spiro 2016). However, the decentralized nature of platforms like Mastodon raises pressing questions: How do messages traverse the boundaries between instances? Which types of instances act as hubs of activity, amplifying the reach of information? What factors determine whether a message propagates widely or remains confined within its originating instance? Despite the growing popularity of decentralized platforms, these questions remain largely unanswered, underscoring the need for systematic research.

Existing studies on information diffusion have predominantly focused on centralized platforms. Research has explored the dynamics of information cascades (Bakshy et al. 2012; Lerman and Ghosh 2010), the influence of network structure on dissemination (Centola 2010), and the role of individual user characteristics in amplifying content (Cha et al. 2010). On centralized platforms, algorithms play an important role in shaping diffusion by curating content and optimizing user engagement (Gausen, Luk, and Guo 2022). However, decentralized platforms lack such centralized curation mechanisms, which likely alters the patterns and determinants of information spread. Studies that have examined the Fediverse remain sparse, often limited to descriptive analyses of user behavior (Zignani, Gaito, and Rossi 2018). While these works provide valuable insights into user interactions and network structure, they do not systematically evaluate the importance and effectiveness of various antecedents of information diffusion in federated systems.

Several research gaps persist. First, while the topology of decentralized networks has been described (La Cava, Greco, and Tagarelli 2021), the role of individual instances in the information diffusion process remains unclear. Instances may act as bridges that facilitate cross-instance communication, as amplifiers that enhance message visibility, or as barriers that isolate information within localized communities. Second, there is limited understanding of how user-level, post-level, and instance-level characteristics interact to shape diffusion patterns. While prior research has examined factors such as message sentiment (Stieglitz and Dang-Xuan 2013) and user influence (Weng et al. 2012) in centralized settings, these relationships may manifest differently in decentralized environments.

To address these gaps, this study examines information diffusion on Mastodon during a global crisis. Specifically, we focus on the following research questions: (1) *How does Mastodon's federated architecture shape the flow of information across instances?* (2) *What are the key structural and engagement characteristics that facilitate cross-instance information diffusion?* (3) *How do instances vary in their roles in information diffusion?* By addressing these questions, this work aims to advance our understanding of information diffusion in federated networks. Our findings contribute to the growing body of research on decentralized social media, offering theoretical and practical insights into the unique dynamics of platforms like Mastodon.

Related Work

Decentralized Social Media and the Fediverse

Decentralized social media, particularly the Fediverse, represents a paradigm shift from centralized networks to federated architectures. These platforms, such as Mastodon, operate through independently hosted instances interconnected via open protocols like ActivityPub. Recent years have seen an increasing number of studies examining decentralized social media and how it differs from centralized counterparts (Jeong et al. 2024; He et al. 2023).

One of the key aspects of decentralized social media is the ability for anyone to create new instances in a federated network. This allows for a more even distribution of users and traffic across the social web on a large scale. For example, Zignani, Gaito, and Rossi (2018) examined the topology and evolution of Mastodon, finding that its decentralized nature facilitates diverse community interactions but also presents challenges for achieving global reach. Another dimension of decentralization is the ability to foster communities tailored to the interests of individuals and groups (Garompolo, Molinaro, and Iera 2022). For example, La Cava, Mandaglio, and Tagarelli (2024) examined polarization in decentralized networks, where instances are divided into groups holding opposing views on controversial topics.

While existing studies have provided valuable insights, they often emphasize descriptive analyses or community-level phenomena, overlooking the mechanisms driving information diffusion in decentralized networks. Key factors that potentially influence information diffusion such as hashtags, mentions, and media content, have been shown to drive engagement on traditional centralized platforms like Twitter (Zeng, Starbird, and Spiro 2016). Boosting and favoriting behaviors, though conceptually similar to retweeting and liking on centralized platforms, may operate differently in the decentralized context, which amplifies content within and across instances without relying on global recommendation algorithms. La Cava, Greco, and Tagarelli (2022) explored user interactions in decentralized networks and pointed out that boundary-spanning users who are actively engaged in multiple instances play critical roles in cross-instance information propagation.

Overall, most studies on the Fediverse remain descriptive, and current research lacks a systematic focus on the underlying mechanisms of information diffusion in decentralized networks. These gaps motivate our work on understanding the dynamics and patterns of information diffusion, particularly on Mastodon as a leading example of decentralized social media.

Information Diffusion in Social Media

Information diffusion has been extensively studied on centralized social media platforms such as Twitter, Facebook, and Instagram. These studies have primarily focused on understanding how information propagates through networks via retweets, shares, and likes. For instance, Bakshy et al. (2012) investigated the role of social networks in shaping exposure to information and its subsequent spread. Lerman and Ghosh (2010) analyzed the dynamics of information contagion, emphasizing the role of user interactions in amplifying content visibility. Centralized platforms also leverage algorithms to curate content and optimize user engagement, which significantly influences diffusion patterns (Gausen, Luk, and Guo 2022). Despite these insights, the mechanisms observed on centralized platforms may not directly apply to decentralized systems, where centralized curation is replaced by a federated network of instances.

Researchers have developed a variety of models to understand the mechanisms driving information diffusion. Epidemic models, such as the Susceptible-Infectious-Recovered (SIR) framework, conceptualize information spread analogously to infectious diseases, where users transition between states of susceptibility, infection, and recovery (Kumar and Sinha 2021; Govindankutty and Gopalan 2024). These models provide a simplified yet powerful framework for analyzing information cascades. Stochastic models, like the independent cascade model, incorporate randomness by assigning activation probabilities to network edges, effectively simulating the unpredictability of user behavior (Saito, Nakano, and Kimura 2008). Additionally, machine learning approaches, including node-level and graph-level methods, have been employed to predict diffusion. Node-level methods use features like user attributes, post content, and interaction history, while graph-level methods focus on network structures to understand propagation patterns (Lagnier et al. 2013; Cao, Han, and Zhu 2021). These methods deepen our understanding of diffusion dynamics in centralized social media as a whole.

In contrast, the research on information diffusion in decentralized social media remains in its infancy. Platforms in the Fediverse operate without centralized curation or control, relying on decentralized architectures where individual instances maintain autonomy. Zignani, Gaito, and Rossi (2018) examined the structure and evolution of Mastodon, finding that its decentralized nature fosters diverse interactions but complicates global information reach. Similarly, Struett et al. (2024) pointed out the governance challenges within the Fediverse in terms of effective moderation and balance between openness and competition. Unlike centralized platforms, where algorithms mediate and amplify content, diffusion in decentralized platforms depends on user-driven interactions, instance policies, and federated network structures.

Despite these initial explorations, significant gaps exist. First, there is a limited understanding of how information propagates naturally across instances without centralized algorithms. Second, the heterogeneity of instances introduces unique challenges, as structural and behavioral differences may either facilitate or inhibit the flow of information. Third, existing diffusion models, often developed for centralized systems, may not adequately capture the complexities of federated architectures. By integrating user-level, post-level, and instance-level characteristics, this work contributes to the growing body of literature on decentralized networks and information diffusion theories.

Methods

Context: the Spread of Gaza Conflict Information on Mastodon

This work selects Mastodon as the research context due to its decentralized nature and unique structural features, suitable for studying information diffusion in distributed networks. This decentralized architecture enables diverse interactions across instances, providing rich data to analyze the platform structure's role in shaping diffusion patterns. Moreover, Mastodon's open-source framework provides researchers with access to detailed metadata, allowing for an in-depth analysis of how content elements, user engagement, and an instance's network position shape information spread.

Crisis events, such as the Gaza conflict, are ideal for information diffusion studies due to their high public engagement, rapid speed, and diverse narratives. The Gaza conflict, in particular, rooted in Israeli-Palestinian tensions since 1948, underwent key events, such as the First and Second Intifadas in 1987 and 2000, and the tragic escalation in 2023 (Diez, Albert, and Stetter 2008). The 2023 conflict, closely connected to public discourse, both reflects and shapes conversations increasingly mediated through social media. Its complex geopolitical, social, and humanitarian dimensions provide a rich context to explore how user-generated content, platform dynamics, and the viral spread of information shape public understanding during crises.

Data Collection

In this study, we collected Mastodon posts, or *"toots"* related to the Gaza conflict from nine instances: *mastodon.social*, *masto.nyc*, *seattle.wa.us*, *sfba.social*, *gardenstate.social*, *better.boston*, *mastodon.nz*, *mastodon.au*, and *mastodon.ie*. These instances were identified through a pre-data collection process. Candidate instances were initially drawn from the public Mastodon server list[1] and evaluated based on activity metrics (e.g., number of active users and posting frequency), geographic diversity, public accessibility, and federation policies.

This instance selection reflects a trade-off between data accessibility and instance diversity. We prioritized instances with high activity levels and open federation policies to ensure sufficient coverage and consistent API access. In contrast, many regional servers outside English-speaking contexts have lower activity or impose restrictions that limited data collection. Additionally, focusing on English content allows for effective manual validation of event relevance and qualitative content analysis, which is critical given the reliance on keyword-based retrieval. While this selection strategy may limit the demographic diversity of originating instances, the federated architecture of Mastodon ensures that the dataset still contains replies and reshares from a much broader network of over 1,200 instances, mitigating the sampling constraints.

To collect Mastodon data, we developed Python scripts to interact with the Mastodon API. Using keyword-based queries[2], we retrieved toots from the federated timelines of the nine selected instances. Since Mastodon treats both seed posts and replies as toots, the data collection includes both. For replies in the dataset, we traced back to their corresponding seed posts to ensure all seed posts were captured. This process resulted in a dataset containing only seed toots. Finally, we utilized the "Context" API calls to collect the complete threads for each seed toot, including all associated replies. This approach collected an initial dataset of 200,096 toots from September 1 to December 19, 2023. Each toot includes content, user metadata, threading metadata, visibility settings, timestamp, and engagement metrics (e.g., replies, boosts, favorites).

Data Processing

Due to Mastodon's federated nature, a single toot can appear across multiple instances if the originating instance federates with others. To ensure reliable de-duplication, we used each toot's globally unique URI to identify and remove 97,390 duplicates. Beyond the URI-based de-duplication, we also flagged self-repeated seed toots, where the same user posted identical content multiple times, including both text and embedded media or links. Among 831 such toots, we identified 287 unique seed toots. For each group, we retained the version with the most replies; if tied, we kept the earliest toot. This approach limits cascade inflation, preserves independence among cascades, and retains the version with the greatest engagement. In total, this step removed 612 toots, including 544 redundant seed toots and 68 associated replies.

[1]https://joinmastodon.org/servers

[2]The ten keywords — *gaza*, *israel*, *palestine*, *hamas*, *genocide*, *freepalestine*, *ceasefire*, *idf*, *warcrimes*, and *palestinians*— were determined through empirical analysis in the pre-collection phase and through external sources, such as news and media.

Next, we identified and removed irrelevant toots. To guide this process, we first performed a preliminary analysis using topic modeling, which reveals that some toots were captured by general keywords (e.g., “genocide”) but were not contextually related to the Gaza conflict. Based on this insight, we implemented a manual validation stage in which two researchers independently reviewed a substantial subset of toots. Posts were excluded if they were not specifically related to the event. Disagreements were resolved through discussion to ensure consistency in inclusion criteria. As a result, 4,407 irrelevant toots were removed.

To account for potential spamming behavior and its impact on cascade structure, we identified and flagged all self-replied toots, defined as replies authored by the same user who posted the original toot being replied to. This behavior, often referred to as self-amplification, may artificially inflate cascade size and depth without involving new participants. Among all replies in our dataset, 26.23% are self-replies. At the thread level, 71.19% of threads contains no self-replies, indicating that while self-replying is not dominant, it is still a substantial behavior that warrants consideration in diffusion analysis. Notably, we observed that cascades dominated by self-replies typically remain within the originating instance; hence, their influence on cross-instance spread, also the main outcome of interest in our study, is limited.

As a result, we collected a total of 97,687 toots with complete diffusion information for seed toots and their associated replies. These toots involve 1,289 unique instances and 18,442 unique users. Of the total, 43,354 toots (44.38%) are replies, whereas 54,333 are seed toots. The seed toots alone involve 746 unique instances and 7,307 unique users. Notably, 41,829 seed toots are single toots, meaning they did not receive any replies.

Diffusion Cascade Construction

To analyze the spread of information on Mastodon, we constructed a diffusion tree for each thread that begins with a seed toot and has at least one reply. A cascade in this context refers to the sequence of interactions where the initial seed toot generates replies, which may elicit further replies, creating a branching structure. In a diffusion tree, a seed toot serves as the root node, with each direct reply forming a child node. This structure extends iteratively, with replies to replies forming successive levels in the tree. Formally, let T_t denote the cascade for seed toot t, represented as a tree structure $T_t = (V_t, E_t)$, where V_t is the set of nodes (toots) and E_t is the set of directed edges representing reply relationships.

To characterize the cascades, we calculated the following metrics:

- **Cascade Size** (**S_t**): the total number of toots in the cascade for seed toot t.
- **Cascade Depth** (**D_t**): the length of the longest path from the root node (seed toot) to any leaf node, where a leaf node is defined as a reply with no further responses.
- **Cascade Breadth** (**B_t**): the maximum number of replies at any single level l in the cascade, given by:

$$B_t = \max_l \{|V_{t,l}|\}$$

where $V_{t,l}$ is the set of nodes at level l.

- **Cross-Instance Spread** (**C_t**): the number of replies in the cascade from instances outside the seed toot’s origin. This metric reflects contribution diversity across the federated network.
- **Cross-Instance Spread Ratio** (**R_t**): the proportion of replies in the cascade that originate from other instances, computed as:

$$R_t = \frac{C_t}{S_t}$$

- **Spread Account Diversity** (**A_t**): the number of unique user accounts involved in the cascade.

Reply-based Network Construction

Using the full diffusion cascades, we constructed a network to analyze the flow of replying interactions between instances. In this network, nodes represent individual instances, and directed edges capture reply relationships between them. Let $G = (N, E)$ denote the directed reply-based network, where N is the set of nodes and E is the set of directed edges.

The size of each node N_j is the total number of toots (both seed toots and replies) from instance j, expressed as:

$$N_j = S_j^{\text{seed}} + S_j^{\text{reply}}$$

where S_j^{seed} and S_j^{reply} are the number of seed toots (roots of cascades) and replies, respectively, from instance j.

For a directed edge e_{ij} from instance i to instance j, the weight w_{ij} represents the volume of replies sent from users in instance i to toots in instance j, which measures the intensity of interaction and the flow of information between instances.

After building the reply-based network, we calculated the following node properties to analyze the structural roles of instances within the network:

- **In-Degree Centrality**: a measure of how central a node is based on the number of incoming connections, indicating the prominence of an instance as a target of replies;
- **Out-Degree Centrality**: a measure of how central a node is based on the number of outgoing connections, indicating an instance’s activity in engaging with others;
- **Betweenness Centrality**: a measure of a node’s importance based on the number of shortest paths between other nodes that pass through it, indicating its role in connecting different parts of a network.
- **Closeness Centrality**: the inverse of the average shortest path distance from a node to all other nodes, representing how quickly an instance can reach or be reached by others in the network;
- **Eigenvector Centrality**: a measure of the influence of a node within the network, based on its connections to other highly connected nodes, highlighting instances that interact with influential hubs;
- **Influence Score**: the proportion of nodes in the network that can be reached from a given node (directly or indirectly).

Cascade Regression Modeling

Next, we used linear regressions with various fixed effects and control variables to evaluate how an instance's network features influence cascade outcome, including size, depth, breadth, cross-instance spread, spread ratio, and spread account diversity. The unit of analysis is the cascade originating from seed toots, including both those that generate cascades and those that do not. The regression model is specified as follows:

$$\text{Cascade Outcome}_{tui} = \text{Instance Network Features}_i + \text{User Features}_u + \text{Toot Features}_t + \text{User}_u + \text{Year}_y + \text{Month}_m + \epsilon_{tui}$$

Variables in the estimation model are explained below:

- Cascade Outcome$_{tui}$ represents various outcomes of cascades resulting from toot t sent by user u located in instance i. Cascade outcomes include the cascade size, depth, breadth, cross-instance spread, spread ratio, and spread account diversity.
- Instance network features include in-degree centrality, out-degree centrality, betweenness centrality, closeness centrality, eigenvector centrality, and influence score. Note that network features were included one by one because of their high correlations with each other.
- User features include the number of followers, followings, bot (whether a user is designated as a bot account), tenure (number of days a user has been using the Mastodon service), and total statuses (the total number of posts made by a user). User features may update and vary as users post more toots. The user-level fixed effect is included to account for unobserved user-specific features that may affect the cascade outcomes.
- Toot features include the number of within-instance mentions (mentions of users from the same instance as the toot), cross-instance mentions (mentions of users from instances different from the toot's origin), tag count, URL count, word count, language, media content (images, videos, gifs, audios), sensitive (indicators of whether the toot is marked sensitive), self-reply ratio (the ratio of replies from the posting user), and toot type (whether the toot belongs to within-instance diffusion or cross-instance diffusion).

The summary statistics of all numerical variables are shown in Table 1. Instance network centrality metrics were normalized during network construction. All other numerical features, including user features, toot features, and cascade outcomes, have been processed with logarithm transformation during modeling to mitigate the impact of extremely large values. Variables that are highly correlated with each other were removed. Multiple fixed effects (FEs), including year, month, and language FEs, were included to account for overall time trends and cultural differences in social media discourse.

Within- and Cross-Instance Spread Classification

Building on the regression analysis, we took the next step to predict whether a seed toot can spread beyond its originating instance. This task is framed as a binary classification. A seed toot is classified as "cross-instance spread" if it receives any reply from a user in an instance other than its origin. Otherwise, if the seed toot does not receive any replies or if all its replies stay within the originating instance, it is classified as "within-instance spread". This classification framework allows us to identify the predictive characteristics concerning the broader propagation of information and the dynamics of localized versus cross-instance interactions within decentralized networks.

In this classification task, we utilized features from three key categories: user-level, toot-level, and instance-level properties, as seen in Table 1. These features capture the multi-layered attributes that may relate to cascade spread, including characteristics of the initial toot, user engagement patterns, and the structural properties of the originating and participating instances.

We chose tree-based machine learning models for this task because of their ability to capture non-linear relationships and interactions among features, which are common in complex diffusion dynamics. During a preliminary analysis, we evaluated decision trees, random forests, and boosted tree models. Among these, XGBoost (eXtreme Gradient Boosting) outperform others in both accuracy and computational efficiency. Therefore, our focus is on the implementation and evaluation of XGBoost models for this task.

The machine learning setup began with assigning a random seed. To account for variability in random processes, we used 10 different random seeds randomly selected from the range [0, 1000]. The entire machine learning pipeline was repeated for each random seed. The pipeline was structured as follows: the dataset was randomly split into 80% training data and 20% test data. For model optimization, we applied 10-fold cross-validation with grid search to tune four key hyperparameters: learning rate (controls the step size for updates), max depth (limits tree complexity to prevent overfitting), subsample ratio (fraction of data used for training each tree), and number of estimators (total trees in the ensemble). The evaluation metric for parameter tuning is logarithmic loss (logloss). Once the optimal hyperparameter combination was determined, the model was evaluated on the intact test dataset. The final model performance, including precision, recall, accuracy, and F_1 score, is reported as the aggregate performance with error ranges across the 10 random seeds. Additionally, feature importance analysis identifies the most predictive features and feature categories, offering insights into the key drivers of cross-instance spread.

Results

Exploratory Data Analysis

We began with an exploratory analysis to examine how user characteristics, toot-level elements, and reply interactions related to the structure of information diffusion on Mastodon. Specifically, we investigated who initiated interactions, what kinds of toots attracted interactions, and how different reply interactions led to distinct cascade structures. Our analysis focuses on the seed toots. We categorized these seed toots into three groups based on the type of replies their

Statistic	N	Mean	St. Dev.	Min	Max
User-level Features					
Follower Count	54333	3571.673	20085.610	0	3360000
Following Count	54333	691.622	1446.538	0	33670
Bot	54333	0.061	0.239	0	1
Statuses Count	54333	15832.960	31599.550	0	1147495
Tenure	54333	919.035	523.061	390	15381
Toot-level Features					
Within-Mention Count	54333	0.015	0.145	0	10
Cross-Mention Count	54333	0.112	0.444	0	33
Tag Count	54333	5.108	4.425	0	82
URL Count	54333	0.814	0.989	0	42
Word Count	54333	39.178	40.207	0	1399
Audio Count	54333	0.001	0.025	0	1
Video Count	54333	0.090	0.286	0	1
Gif Count	54333	0.0018	0.043	0	1
Image Count	54333	0.283	0.595	0	4
Sensitive	54333	0.042	0.200	0	1
Self Reply Ratio	54333	0.055	0.220	0	1
Instance-level Features					
Total Users	53847	655734.800	986844.600	1	2297913
In-Degree Centrality	53981	0.178	0.190	0	0.480
Out-Degree Centrality	53981	0.125	0.126	0	0.323
Betweenness Centrality	53981	0.069	0.085	0	0.208
Closeness Centrality	53981	0.464	0.099	0	0.601
Eigenvector Centrality	53981	0.279	0.423	2.59e-23	0.984
Influence Score	53981	0.639	1.406	0	0.672
Cascade Outcomes					
Cascade Size	54333	1.798	3.642	1	332
Cascade Depth	54333	0.549	2.040	0	147
Cascade Breadth	54333	0.383	1.665	0	243
Cross-Instance Spread	54333	0.392	2.001	0	121
Cross-Spread Ratio	54333	0.075	0.190	0	0.978
Spread Account Diversity	54333	1.375	1.472	1	140

Table 1: Descriptive statistics of user-level, toot-level, instance-level features and cascade statistics

resulting cascades received: (1) **non-replied**, which received no replies; (2) **self-replied**, which received mostly replies from the original author (over 80%); and (3) **externally-replied**, which received replies primarily from other users. Notably, 71.2% of cascades contains no self-replies.

Who Initiates Reply Interactions? We compared user-level characteristics of seed toots that led to three types of interaction outcomes: non-replied, self-replied, and externally-replied. Specifically, we examined follower count, following count, and posting rate, using log-transformed values to address skewness. Posting rate was calculated as the total number of statuses posted by a user divided by the number of days since their account was created on Mastodon.

As shown in Figure 1, users whose seed toots received replies from others (externally-replied) consistently have higher follower counts, follow more users, and post at a higher rate than users whose seed toots received no replies or mostly self-replies. In contrast, non-replied and self-replied toots are associated with users who have smaller audiences and lower activity levels. Kruskal-Wallis tests reveal significant differences across all three groups for each metric (all $p < 0.001$). These findings suggest that reply patterns of seed toots are associated with the author's social visibility and activity level.

Which Seed Toots Receive Replies? Table 2 shows statistically significant differences in toot-level features of the seed posts across groups. These features capture aspects of content richness (word count and tag count), element richness (presence of images or URLs), and mentions (whether the seed toot mentioned another user from the same instance or a different one).

Externally-replied seed toots tend to be moderately long (median = 41.0), slightly shorter than self-replied ones (47.0) but longer than non-replied seed toots (30.0). Tag count is also lower for externally-replied toots (median =

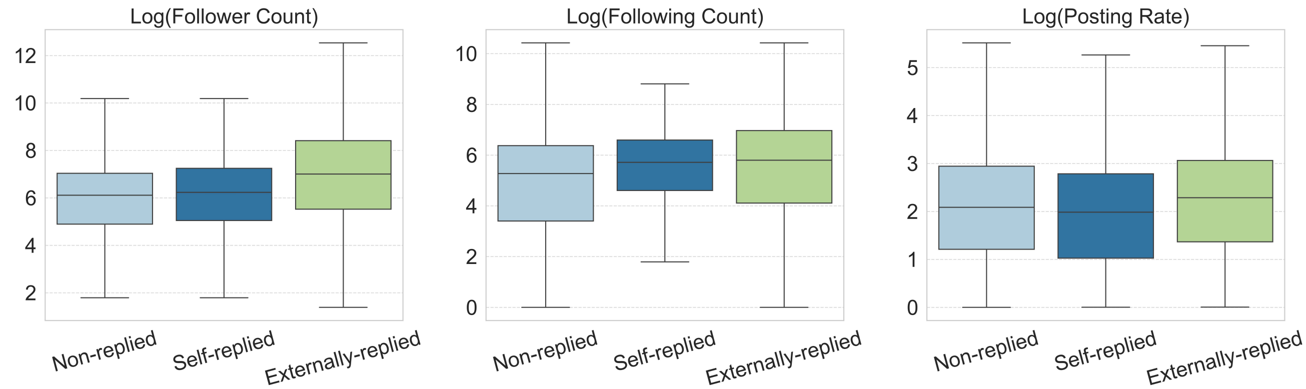

Figure 1: Distribution of user features (log-transformed) by seed toot's reply interaction outcomes.

Toot Feature	Non-replied	Self-replied	Externally-replied	Test	p-value
Word Count	30.0 [16.0–49.0]	47.0 [25.0–67.0]	41.0 [25.0–61.0]	H = 1842.6	$< 0.001^{***}$
Tag Count	4.0 [2.0–6.0]	4.0 [2.0–7.0]	3.0 [2.0–5.0]	H = 428.3	$< 0.001^{***}$
Image Present	23.1%	32.2%	23.1%	$\chi^2 = 115.2$	$< 0.001^{***}$
URL Present	71.6%	59.6%	64.5%	$\chi^2 = 328.9$	$< 0.001^{***}$
Within- @ Present	1.4%	0.7%	1.2%	$\chi^2 = 10.0$	0.007^{**}
Cross- @ Present	8.4%	8.0%	7.2%	$\chi^2 = 17.5$	$< 0.001^{***}$

Table 2: Toot-level features by interaction type. Medians with IQRs are shown for numeric features (Kruskal-Wallis test); percentages are shown for binary features (chi-square test).

3.0), suggesting that excessive tagging may not encourage replies from others. Self-replied toots are more likely to include images (32.2%) compared to both externally-replied and non-replied toots (23.1%), while URLs are most common in non-replied toots (71.6%). Mentions are infrequent overall but slightly more common in non-replied toots. These patterns suggest that externally-replied toots are less media-heavy and less tag-saturated, whereas self-replied toots often include additional media elements, possibly to support self-amplification.

How Replies Relate to Cascades? To examine how reply types relate to cascade outcomes, we compared the cascades initiated by self-replied and externally-replied seed toots. Non-replied seed toots were excluded here because they have no resultant cascades.

As shown in Figure 2, externally-replied cascades tend to be slightly larger (median = 3.0) than self-replied ones (2.0), but more importantly, they show greater breadth, cross-instance spread, and account diversity. For example, self-replied cascades show virtually no cross-instance replies (median = 0.0) and minimal diversity in participating accounts, while externally-replied cascades involve more distinct users and replies across different instances. In contrast, cascade depth is slightly higher for self-replied cascades, likely reflecting sequential self-posting behavior. All five cascade outcomes show statistically significant group differences ($p < 0.001$). These results suggest that while self-replies contribute to deeper but more isolated cascades, externally-replied toots are associated with broader, more distributed diffusion involving diverse participants.

Reply-Based Instance Network Analysis

To start with, we constructed cascades by gathering all replies to each seed toot in our dataset. Using the Mastodon "Context" API call, as detailed in the Methods section, we collected the full thread for each seed toot. The metadata for each descendant toot in the thread includes the ID of the toot it replies to, allowing us to reconstruct the hierarchical structure of replies. Importantly, the context of a seed toot captures not only its direct replies but also replies at all subsequent levels of the thread. This enabled us to build 12,504 diffusion trees (i.e., cascades), encompassing a total of 55,858 toots: 12,504 seed toots and 43,354 replies. Summary statistics for the cascade properties are provided in Table 1.

Next, we constructed an instance-level reply network. In this network, nodes represent instances participating in the cascades, and directed edges indicate replying behavior between instances. The resulting graph comprises 1,156 nodes and 6,549 directed edges, reflecting the replying dynamics across the federated network. Figure 3 visualizes the central part of this network, with key nodes and interactions.

Cluster analysis using the Louvain method (modularity = 0.296) indicates the hierarchical structure of the network. *Mastodon.world* is the hub of the largest cluster (30.36% nodes), while *mastodon.social*, *mastodon.online*, *mas.to*,

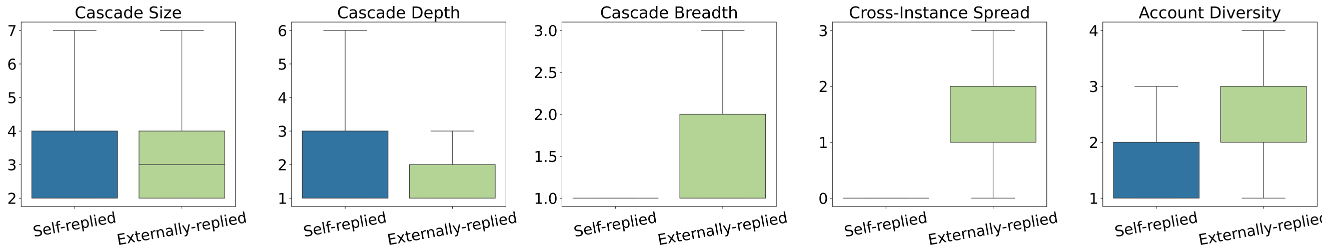

Figure 2: Distribution of cascade outcomes by seed toot's reply types.

and *mstdn.social* lead other major clusters with varying densities and cohesion levels. These hubs generally exhibit lower clustering coefficients, suggesting that they act as bridges between instance nodes and enable efficient information flow across the network.

Instance rankings by centrality metrics in Table 3 further shows that these largest instances are the most active and influential nodes, with dominating degree and centrality metrics. Instances in smaller clusters, such as *mastodon.ar.al*, play a unique role by connecting localized interactions to the broader network. For example, *mastodon.ar.al*'s relatively low clustering coefficient (0.09) and high betweenness centrality highlight its importance as a bridge between otherwise distant communities.

In short, these patterns suggest a dual structure in the reply-based network: large instances facilitate global-scale diffusion, while instances in smaller clusters act as intermediaries, linking regional clusters to the broader community.

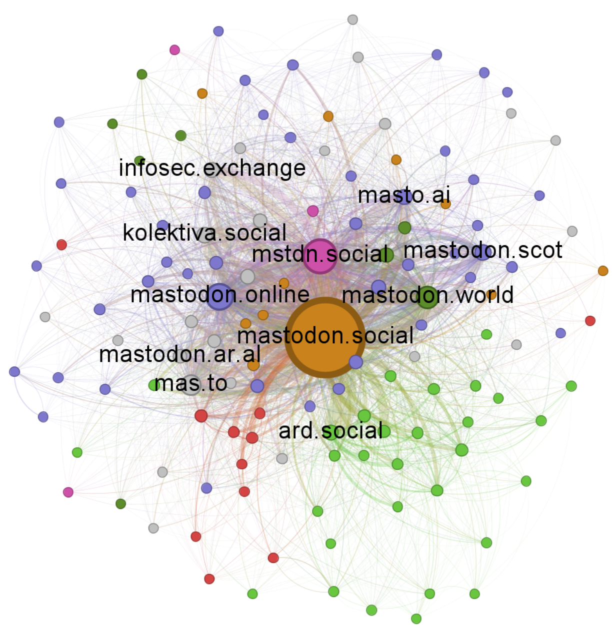

Figure 3: Reply-based network: Louvain clustering (*modularity* = 0.296). Largest six clusters with 132 nodes (11.42%) and 3,011 edges (45.98%) filtered by weighted degree $\geq$ 20.

Impacts of Instance Features on Cascade Outcomes

We used linear regressions to evaluate the impact of instance-level network features on cascade outcomes. We analyzed the impact of instance network features one by one. The estimation results are shown in Table 4. As we can see, the majority of the coefficients are positive and significant, indicating that the instance network features generally facilitate information cascade. There are several key findings that can be drawn from the regression results above:

- Among various instance network features, betweenness centrality has the largest impact on cascade outcomes in terms of the impact magnitudes. Thus, increasing the betweenness centrality of an instance is the most effective way to enhance the cascade outcomes of toots from that instance.
- Among various cascade outcomes, cascade size, breadth, and cross-instance spread receive the largest impact from instance network features in terms of impact magnitudes. In other words, instance network features are more effective in increasing cascade size, breadth, and cross-instance spread. The magnitude of the impact on the spread ratio is smaller compared to other cascade outcomes, potentially due to its percentage-based measuring scale.
- Eigenvector centrality and closeness centrality are more effective in increasing cascade size, cross-instance spread, and spread account diversity than impacting cascade breadth and depth.
- Influence score has the weakest impact on cascade outcomes, with only limited positive effect on cascade size, cross instance spread, and spread account diversity.

Cross-Instance Spread Classification

Model Performance and Global Feature Importance

The XGBoost classifier effectively classifies the within-spread and cross-spread groups using user, toot, and instance-level features. Table 5 presents model average performance metrics $\pm$ standard deviation over 10 times with different random seeds. The model achieves high training performance: precision $\approx$ 0.850, recall $\approx$ 0.930, F_1 score $\approx$ 0.888, and accuracy $\approx$ 0.966. The test performance is slightly lower, but still remains strong: precision $\approx$ 0.813, recall $\approx$ 0.888, F_1 score $\approx$ 0.849, and accuracy $\approx$ 0.953. These results show a robust generalization, with a small performance gap between training and test datasets across all

Metrics	#1	#2	#3	#4	#5
Degree	mastodon.social	mstdn.social	mastodon.online	mas.to	mastodon.world
In-degree Centrality	mastodon.social	mstdn.social	mastodon.online	mas.to	mastodon.world
Out-degree Centrality	mastodon.social	mstdn.social	mastodon.online	mastodon.world	mas.to
Betweenness Centrality	mastodon.social	mstdn.social	mastodon.online	mas.to	mastodon.ar.al
Closeness Centrality	mastodon.social	mstdn.social	mastodon.online	mas.to	mastodon.world
Eigenvector Centrality	mastodon.social	mstdn.social	mastodon.online	mastodon.world	mas.to

Table 3: Top 5 ranked instances for reply-based network properties

Instance Network Features	**Size**	**Depth**	**Breadth**	**Cross Instance Spread**	**Spread Ratio**	**Spread Account Diversity**
Out-degree Centrality	16.032***	22.123**	23.271**	19.155***	1.692***	11.625***
	(1.478)	(10.732)	(10.600)	(1.328)	(0.269)	(1.001)
In-degree Centrality	9.314***	12.660**	13.797**	11.295***	1.066***	7.021***
	(0.822)	(5.971)	(5.897)	(0.739)	(0.150)	(0.557)
Betweenness Centrality	27.231***	40.103**	44.364***	33.172***	3.221***	21.136***
	(2.421)	(17.575)	(17.358)	(2.174)	(0.440)	(1.639)
Closeness Centrality	1.001***	1.147	1.296	1.361***	0.109***	0.587***
	(0.205)	(1.483)	(1.465)	(0.184)	(0.037)	(0.139)
Eigenvector Centrality	26.322***	35.714	41.530*	32.579***	3.439***	21.174***
	(3.198)	(23.200)	(22.914)	(2.873)	(0.581)	(2.166)
Influence Score	0.325**	-0.210	-0.129	0.512***	0.034	0.178*
	(0.135)	(0.979)	(0.967)	(0.121)	(0.025)	(0.091)

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Results were estimated with fixed effect OLS regressions, which include user features, post features, year fixed effect, month fixed effect, and user fixed effect.

Table 4: Instance Network Features and Cascade Outcomes

metrics. This model also demonstrates a strong consistency, with low standard deviations for all metrics, suggesting reliable performance with different random seeds.

	Train	Test
Precision	0.850 ± .002	0.813 ± .009
Recall	0.930 ± .003	0.888 ± .010
Accuracy	0.966 ± .001	0.953 ± .001
F_1 Score	0.888 ± .001	0.849 ± .002

Table 5: Training and test performance (mean ± std)

The implications of precision and recall are particularly important given the class imbalance (85.4% within-spread vs. 14.6% cross-spread). Despite this imbalance, the model maintains a strong recall on the minority cross-spread group, indicating its effectiveness in identifying inter-instance diffusion. This provides a solid foundation for the further analysis of model misclassifications.

To understand what the model has learned, we examined global feature importance using SHAP (SHapley Additive exPlanations) values. Figure 4 presents the top six features contributing to the model's predictions. These include user-level, toot-level, and instance-level features. Features are ranked by their absolute SHAP value, with higher-ranking features contributing more to the predictions. Each dot represents a test record, and its position indicates whether the feature increases or decreases the SHAP value (impact on model output), while the colors (red for high values, blue for low values) represent feature magnitudes.

Among the top six features, mention-related features, particularly within-mention count and cross-mention count, show the strongest impact on the prediction outcome. The model is more likely to predict cross-instance spread when either mention count is high. This indicates that frequent mentions, even within a single instance, may signal a broader diffusion potential. This reflects the model's sensitivity to interaction-heavy content as a signal of cross-instance reach.

Content-related features also play an important role. URL count emerges as one of the most influential predictors, suggesting that posts containing links are more likely to be referenced, shared, or discussed across instances, thereby increasing their likelihood of cross-instance diffusion.

Network-based features such as in-degree centrality and betweenness centrality further contribute to the model predictions. A high in-degree centrality suggests that the originating instance receives a large number of replies, which may indicate greater visibility or engagement potential.

Betweenness centrality captures the instance's role as a bridge in the network, indicating its importance in connecting otherwise distant communities. However, compared to mention- and content-related features, its SHAP contributions are more concentrated around zero, suggesting that it serves as a contextual structural signal rather than a dominant driver of cross-instance spread.

User-level features also remain important. Follower count is among the top predictors, likely because users with larger audiences have greater visibility and are more prone to cross-instance diffusion.

While these results show that the model utilizes a diverse set of meaningful features, it is equally important to examine where the model fails. In the following section, we analyzed false positives and false negatives to better understand the limitations of the classifier and the contextual patterns behind its misclassifications.

Error Analysis: What Drives Misclassifications? We conducted a SHAP-based error analysis comparing feature contributions across four prediction types: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). In this context, TP refers to correctly identified cross-spread toots, FP to toots incorrectly predicted as cross-spread, TN to correctly identified within-spread toots, and FN to toots incorrectly predicted as within-spread. Figure 5 presents the average SHAP value of the top six features by prediction type. This breakdown demonstrates how different features influence correct and incorrect predictions, and reveals systematic differences in model behavior across error types.

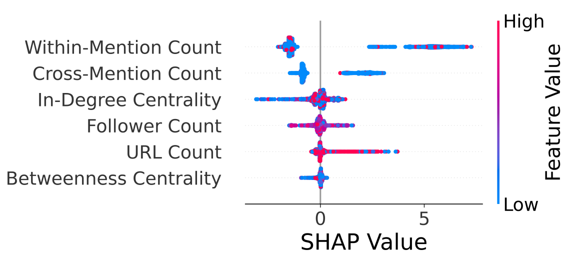

Figure 4: SHAP summary plot.

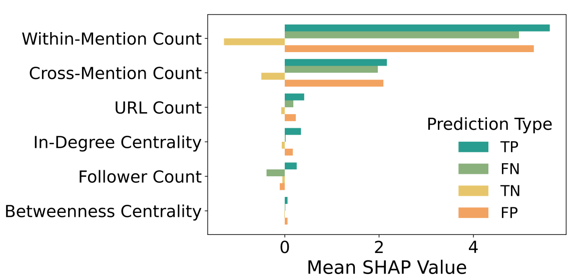

Figure 5: Directional SHAP by prediction type.

First, both TPs and TNs show consistent SHAP directions for within-mention and cross-mention counts. These features have strong positive SHAP values for TPs and strong negative values for TNs, indicating the model's confidence when the features align with the true class. However, these features have especially strong positive SHAP values for FPs, suggesting that the model may over-interpret high mention activity as a signal for cross-instance diffusion, even when the toot remains confined to a single instance. For FNs, mention-related features still have a positive influence on the prediction, but the magnitude is weaker than in TPs or FPs. This suggests that while the model detects signals of interaction-heavy content, it does not weigh them strongly enough to shift the prediction from within- to cross-instance spread.

Content-related features, such as URL count, display a similar pattern, consistently pushing predictions toward cross-instance spread, particularly in FP cases. This suggests that the model may overestimate the diffusion potential of externally referenced content, while also failing to fully leverage this signal in some FN cases.

Network-based features also contribute to prediction patterns. In-degree centrality shows a similar directional effect as mention-related features, reflecting the visibility and engagement level of the originating instance. In contrast, betweenness centrality exhibits more moderate SHAP values concentrated around zero, indicating that while bridge-like structural positions provide useful contextual information, their direct contribution to misclassification is comparatively limited.

Finally, follower count displays a distinct SHAP pattern across prediction types. It has positive SHAP values for TPs and negative values for TNs, showing that the model generally associates high follower counts with cross-instance spread. However, both FPs and FNs also have negative SHAP values, with the magnitude being strongest for FNs. It indicates that in many FN cases, follower count plays a major role in pushing the model toward an incorrect within-spread prediction. In contrast, in FP cases, follower count tends to have a weaker influence, as its effect is often overridden by stronger signals such as mention activity and URL presence.

Overall, the SHAP-based error analysis shows that misclassifications follow consistent patterns rather than occurring randomly. FPs are mainly driven by the model's overestimation of mention activity, URL presence, and in-degree centrality. FNs, in contrast, are more likely when the model over-relies on follower count or underweights mention activity.

Qualitative Insights in Misclassified Posts Finally, we conducted an exploratory review on the content of the ten most misleading FPs and FNs, based on model confidence: toots where the prediction probability is highly confident yet incorrect (e.g., predicted probability near 0 for true positives, or near 1 for true negatives). Each toot was coded using two simple schemes:

- Tone (third-person or personal): whether the toot describes external events or expressed subjective reactions.
- Fact support (supported, unsupported, or inapplicable): whether the toot includes any evidence support (e.g., links, quoted sources).

While exploratory, this review reveals clear contrasts. Most FPs (8/10) are in a third-person tone, often resembling news reports, and six include external links or media as factual support. In contrast, half of the FNs (5/10) are from a personal perspective and lack any factual claims. Two others promote Gaza-related mobile apps or events, while the remaining three make unsupported claims without sources. Additionally, within each group, we identified two misclassified toots posted by a specific user. This suggests that certain user-level characteristics may systematically contribute to highly confident misclassifications. This observation aligns with our earlier findings of the SHAP-based error analysis.

This preliminary content analysis highlights how textual perspective and the presence of factual support may relate to model misclassifications. While limited in scale, our findings suggest that posts resembling well-sourced news may be overestimated in their spread potential, whereas personal or promotional content may be overlooked. These observations open opportunities for more in-depth analyses of how content style, user behavior, and audience context jointly contribute to information diffusion across federated platforms.

Discussion

Decentralized social media platforms present a unique context for understanding information cascades, especially through the structure of interconnected instances. This study leverages this architecture to examine how instance-level features shape various cascade outcomes during a global crisis. We show that user characteristics (e.g., follower count) and toot-level features (e.g., mentions) contribute to information diffusion, but that instance-level network properties play a crucial role in facilitating or constraining cross-instance spread.

Our results also align with established findings from centralized platforms like Twitter, where user influence and content features (e.g., hashtags, mentions) strongly predict engagement and virality. For example, users with larger follower bases and toots that include mentions are more likely to trigger replies on centralized diffusion (Bakshy et al. 2011; Kwak 2010). However, in federated environments, these features do more than enhance visibility: they help bridge structural gaps between instances, making them especially critical for enabling cross-instance diffusion.

While prior work has largely focused on centralized platforms, directly applying those insights to federated systems may obscure fundamental differences. Decentralized platforms like Mastodon operate without centralized curation or a unified user network. Instead, diffusion is shaped by the distinct policies, federation relationships, and moderation norms of independently run instances. Our work addresses this gap by analyzing diffusion from an instance-centric perspective suited to the decentralized architecture.

Rather than tracing individual users, we focus on instance networks as the fundamental unit of analysis in federated systems. Instance-level dynamics reflect how communities, shaped by shared moderation policies and federation choices, influence the visibility and spread of information. While user- and instance-level networks may share structural properties (e.g., ego-centric diffusion patterns) (Arnaboldi et al. 2016, 2017), our findings highlight the distinct role of instance attributes on cascade outcomes like size, cross-instance spread, and spread account diversity.

These findings extend prior diffusion research (e.g., Goel et al. (2016)) by showing how decentralized architectures reshape the conditions under which information spreads. They also complement recent studies on the Fediverse (e.g., Zignani, Gaito, and Rossi (2018)) by integrating multi-level features, user, toot, and instance, into a unified modeling framework. This layered approach reflects the complexity of diffusion in federated systems and offers a foundation for future research on decentralized communication.

Practically, these findings offer important implications for platform governance and crisis communication strategies. First, federated networks in information cascades not only depend on user and toot features, but also on instance features, emphasizing the structural and administrative roles of instances. Highly connected instances, like *mastodon.social* and *mastodon.world*, take on significant responsibilities for accurate and responsible information dissemination as digital gatekeepers (Shoemaker and Vos 2014). While the design of federated social networks may inherently limit global diffusion to some extent, they offer potential for controlling misinformation and polarization. Additionally, predictive models for cascade behaviors could support real-time content moderation, helping platform administrators prioritize critical information and mitigate misinformation.

Conclusion

This study examines information diffusion within federated social media, focusing on Mastodon during the Gaza conflict. By analyzing user-, toot-, and instance-level features, we studied how behaviors, content, and structure shape cascades through exploratory analysis, network modeling, and predictive modeling.

Our findings offer several theoretical and practical contributions. The study advances the literature on decentralized social media by analyzing cross-instance information diffusion patterns and identifying the key role of instance features in shaping cascade dynamics. It highlights how federated architectures mediate diffusion processes, emphasizing the influence of well-connected instances in enabling cross-instance cascades. By incorporating instance, user, and toot features in predictive models, this work deepens the understanding of antecedents to information cascades. Practically, these insights can inform platform governance, from moderating information flows to designing tools for crisis communication and misinformation control.

Despite its contributions, this study has limitations. First, it focuses on a single crisis event - the Gaza conflict, and uses data from a subset of Mastdon instances. While the federated nature of the platform enabled the collection of toots from a wide range of instances, our initial selection of nine primarily English-speaking instances may still introduce demographic bias. Future work should expand to more diverse linguistic, regional, and topical contexts to improve generalizability. Second, we constructed the instance network using reply-based interactions, a representative but partial method to capture inter-instance relationships. Future studies could also incorporate user following networks, which reflect long-term social ties, and toot mentioning network, which facilitate content visibility across instances. Finally, our analysis primarily captures static snapshots of diffusion patterns, overlooking temporal diffusion dynamics. Future work could incorporate longitudinal methods to fully uncover how user and instance roles evolve and affect cascades over time.

References

Arnaboldi, V.; Conti, M.; La Gala, M.; Passarella, A.; and Pezzoni, F. 2016. Ego network structure in online social networks and its impact on information diffusion. *Computer Communications*, 76: 26–41.

Arnaboldi, V.; Conti, M.; Passarella, A.; and Dunbar, R. I. 2017. Online social networks and information diffusion: The role of ego networks. *Online Social Networks and Media*, 1: 44–55.

Bakshy, E.; Hofman, J. M.; Mason, W. A.; and Watts, D. J. 2011. Everyone's an influencer: quantifying influence on twitter. In *Proc. of the* 4^{th} *WSDM*, 65–74.

Bakshy, E.; Rosenn, I.; Marlow, C.; and Adamic, L. 2012. The role of social networks in information diffusion. In *Proc. of the 21st international conf. on WWW*, 519–528.

Cao, Z.; Han, K.; and Zhu, J. 2021. Information diffusion prediction via dynamic graph neural networks. In *2021 IEEE 24th international conference on computer supported cooperative work in design (CSCWD)*, 1099–1104. IEEE.

Centola, D. 2010. The spread of behavior in an online social network experiment. *science*, 329(5996): 1194–1197.

Cha, M.; Haddadi, H.; Benevenuto, F.; and Gummadi, K. 2010. Measuring user influence in twitter: The million follower fallacy. In *Proc. of ICWSM*, volume 4, 10–17.

Diez, T.; Albert, M.; and Stetter, S. 2008. *The European Union and border conflicts: The power of integration and association*. Cambridge University Press.

Garompolo, D.; Molinaro, A.; and Iera, A. 2022. Bridging separate communities with common interest in distributed social networks through the use of social objects. *Future Generation Computer Systems*, 129: 440–452.

Gausen, A.; Luk, W.; and Guo, C. 2022. Using agent-based modelling to evaluate the impact of algorithmic curation on social media. *ACM Journal of Data and Information Quality*, 15(1): 1–24.

Goel, A. K.; Bakshi, R.; and Agrawal, K. K. 2022. Web 3.0 and decentralized applications. *Materials Proc.*, 10(1): 8.

Goel, S.; Anderson, A.; Hofman, J.; and Watts, D. J. 2016. The structural virality of online diffusion. *Management science*, 62(1): 180–196.

Govindankutty, S.; and Gopalan, S. P. 2024. Epidemic modeling for misinformation spread in digital networks through a social intelligence approach. *Sci. Rep.*, 14(1): 19100.

He, J.; Zia, H. B.; Castro, I.; Raman, A.; Sastry, N.; and Tyson, G. 2023. Flocking to mastodon: Tracking the great twitter migration. In *Proc. of the 2023 ACM on Internet Measurement Conference*, 111–123.

Jeong, U.; Sheth, P.; Tahir, A.; Alatawi, F.; Bernard, H. R.; and Liu, H. 2024. Exploring platform migration patterns between twitter and mastodon: A user behavior study. In *Proc. of ICWSM*, volume 18, 738–750.

Kumar, P.; and Sinha, A. 2021. Information diffusion modeling and analysis for socially interacting networks. *Social Network Analysis and Mining*, 11(1): 11.

Kwak, H. 2010. What is Twitter, a social network or a news media? *Department of Computer Science, KAIST*.

La Cava, L.; Greco, S.; and Tagarelli, A. 2021. Understanding the growth of the Fediverse through the lens of Mastodon. *Applied network science*, 6: 1–35.

La Cava, L.; Greco, S.; and Tagarelli, A. 2022. Information consumption and boundary spanning in decentralized online social networks: the case of mastodon users. *Online Social Networks and Media*, 30: 100220.

La Cava, L.; Mandaglio, D.; and Tagarelli, A. 2024. Polarization in Decentralized Online Social Networks. In *Proceedings of the 16th ACM Web Science Conference*, 48–52.

Lagnier, C.; Denoyer, L.; Gaussier, E.; and Gallinari, P. 2013. Predicting information diffusion in social networks using content and user's profiles. In *Advances in Information Retrieval: 35th European Conf. on IR Research, 2013, Moscow, Russia, Proc. 35*, 74–85. Springer.

Lerman, K.; and Ghosh, R. 2010. Information contagion: An empirical study of the spread of news on digg and twitter social networks. In *Proc. of ICWSM*, volume 4, 90–97.

Raman, A.; Joglekar, S.; Cristofaro, E. D.; Sastry, N.; and Tyson, G. 2019. Challenges in the decentralised web: The mastodon case. In *Proc. of the internet measurement conf.*, 217–229.

Saito, K.; Nakano, R.; and Kimura, M. 2008. Prediction of information diffusion probabilities for independent cascade model. In *International conf. on knowledge-based and intelligent information and eng. systems*, 67–75. Springer.

Shoemaker, P. J.; and Vos, T. P. 2014. Media gatekeeping. In *An integrated approach to communication theory and research*, 89–103. Routledge.

Stieglitz, S.; and Dang-Xuan, L. 2013. Social media and political communication: a social media analytics framework. *Social network analysis and mining*, 3: 1277–1291.

Struett, T.; Sinnreich, A.; Aufderheide, P.; and Gehl, R. 2024. Can this platform survive? Governance challenges for the fediverse. *Governance Challenges for the Fediverse*.

Weng, L.; Flammini, A.; Vespignani, A.; and Menczer, F. 2012. Competition among memes in a world with limited attention. *Scientific reports*, 2(1): 335.

Zeng, L.; Starbird, K.; and Spiro, E. S. 2016. Rumors at the speed of light? Modeling the rate of rumor transmission during crisis. In *2016 49th Hawaii International Conference on System Sciences (HICSS)*, 1969–1978. IEEE.

Zignani, M.; Gaito, S.; and Rossi, G. P. 2018. Follow the "mastodon": Structure and evolution of a decentralized online social network. In *Proc. of AAAI on Web and Social Media*, volume 12, 541–550.

Reproducibility Checklist

1. General Paper Structure

1.1. Includes a conceptual outline and/or pseudocode description of AI methods introduced (yes/partial/no/NA) NA

1.2. Clearly delineates statements that are opinions, hypothesis, and speculation from objective facts and results (yes/no) yes

1.3. Provides well-marked pedagogical references for less-familiar readers to gain background necessary to replicate the paper (yes/no) yes

2. Theoretical Contributions

2.1. Does this paper make theoretical contributions? (yes/no) yes

If yes, please address the following points:

2.2. All assumptions and restrictions are stated clearly and formally (yes/partial/no) partial

2.3. All novel claims are stated formally (e.g., in theorem statements) (yes/partial/no) yes

2.4. Proofs of all novel claims are included (yes/partial/no) yes

2.5. Proof sketches or intuitions are given for complex and/or novel results (yes/partial/no) yes

2.6. Appropriate citations to theoretical tools used are given (yes/partial/no) yes

2.7. All theoretical claims are demonstrated empirically to hold (yes/partial/no/NA) yes

2.8. All experimental code used to eliminate or disprove claims is included (yes/no/NA) NA

3. Dataset Usage

3.1. Does this paper rely on one or more datasets? (yes/no) yes

If yes, please address the following points:

3.2. A motivation is given for why the experiments are conducted on the selected datasets (yes/partial/no/NA) yes

3.3. All novel datasets introduced in this paper are included in a data appendix (yes/partial/no/NA) no

3.4. All novel datasets introduced in this paper will be made publicly available upon publication of the paper with a license that allows free usage for research purposes (yes/partial/no/NA) partial

3.5. All datasets drawn from the existing literature (potentially including authors' own previously published work) are accompanied by appropriate citations (yes/no/NA) NA

3.6. All datasets drawn from the existing literature (potentially including authors' own previously published work) are publicly available (yes/partial/no/NA) NA

3.7. All datasets that are not publicly available are described in detail, with explanation why publicly available alternatives are not scientifically satisficing (yes/partial/no/NA) NA

4. Computational Experiments

4.1. Does this paper include computational experiments? (yes/no) yes

If yes, please address the following points:

4.2. This paper states the number and range of values tried per (hyper-) parameter during development of the paper, along with the criterion used for selecting the final parameter setting (yes/partial/no/NA) yes

4.3. Any code required for pre-processing data is included in the appendix (yes/partial/no) no

4.4. All source code required for conducting and analyzing the experiments is included in a code appendix (yes/partial/no) no

4.5. All source code required for conducting and analyzing the experiments will be made publicly available upon publication of the paper with a license that allows free usage for research purposes (yes/partial/no) yes

4.6. All source code implementing new methods have comments detailing the implementation, with references to the paper where each step comes from (yes/partial/no) yes

4.7. If an algorithm depends on randomness, then the method used for setting seeds is described in a way sufficient to allow replication of results (yes/partial/no/NA) yes

4.8. This paper specifies the computing infrastructure used for running experiments (hardware and software), including GPU/CPU models; amount of memory; operating system; names and versions of relevant software libraries and frameworks (yes/partial/no) no

4.9. This paper formally describes evaluation metrics used and explains the motivation for choosing these metrics (yes/partial/no) yes

4.10. This paper states the number of algorithm runs used to compute each reported result (yes/no) yes

4.11. Analysis of experiments goes beyond single-dimensional summaries of performance (e.g., average; median) to include measures of variation, confidence, or other distributional information (yes/no) yes

4.12. The significance of any improvement or decrease in performance is judged using appropriate statistical tests (e.g., Wilcoxon signed-rank) (yes/partial/no) yes

4.13. This paper lists all final (hyper-)parameters used for each model/algorithm in the paper's experiments (yes/partial/no/NA) partial