Analyzing the dynamics of user influence in Threads

(Discussion Paper)

Gianluca Bonifazi[1,†], Christopher Buratti[1,†], Enrico Corradini[1,†], Michele Marchetti[1,†], Federica Parlapiano[1,†], Davide Traini[1,†], Domenico Ursino[1,*,†] and Luca Virgili[1,†]

[1]*DII, Polytechnic University of Marche*

Abstract

One of the most common analyses in social networks concerns power users (also called influencers, lead users, influential users, etc.), i.e. users who play a crucial role in the dissemination of information in a social platform. In this paper, we want to make a double contribution to this line of research by proposing a new definition of power users that takes into account the four main centralities of Social Network Analysis and then applying it to Threads, a social platform that is still little studied by social network analysts because of its young age.

Keywords

Threads, Power Users, Influencers, Influential Users, Lead Users, Social Network Analysis

1. Introduction

The study of power users [1, 2, 3] (also known as lead users [4], influential users or influencers [5, 6, 7, 8, 9, 10]) is a central topic in Social Network Analysis (SNA) [11]. Indeed, they play a central role in catalyzing the dissemination of content, accelerating its speed of spread, and amplifying its visibility. They also shape the dynamics of user interactions and are critical to network cohesion, often connecting communities that would otherwise remain isolated.

Threads[1] was launched by Meta in July 2023 as a direct alternative to X to encourage content sharing and people's participation in public discussions. It allows users to share short texts, images and videos. Threads is tightly integrated with Instagram. Its text-based nature and Instagram-based growth model make it a unique case study for analyzing the dynamics of social platforms. Given the young age of Threads, the study of the dynamics of user influence in it is still in its early stages.

In this paper, we want to contribute to both issues (power user investigation and study of the dynamics of user influence in Threads). In particular, we propose a definition of the

SEBD 2025: 33rd Symposium on Advanced Database Systems, June 16-19, 2025, Ischia, Italy

*Corresponding author.

†These authors contributed equally.

g.bonifazi@univpm.it (G. Bonifazi); c.buratti@pm.univpm.it (C. Buratti); e.corradini@univpm.it (E. Corradini); michele.marchetti@univpm.it (M. Marchetti); f.parlapiano@pm.univpm.it (F. Parlapiano); d.traini@pm.univpm.it (D. Traini); d.ursino@univpm.it (D. Ursino); luca.virgili@univpm.it (L. Virgili)

0000-0002-1947-8667 (G. Bonifazi); 0009-0007-1675-0155 (C. Buratti); 0000-0002-1140-4209 (E. Corradini); 0000-0003-3692-3600 (M. Marchetti); 0009-0009-5475-8101 (F. Parlapiano); 0009-0007-3098-9349 (D. Traini); 0000-0003-1360-8499 (D. Ursino); 0000-0003-1509-783X (L. Virgili)

[1]https://www.threads.net

concept of power user and then we use this definition to propose an approach for detecting and characterizing power users in Threads.

Our definition of power user is intended to be highly selective and based on well-known concepts in SNA in order to take advantage of the knowledge that social network analysts have discovered in the past. In particular, it is based on the idea that to be a power user, it is not enough for a user to have many connections, but she/he must be close to as many users as possible, act as a bridge between communities of users who would otherwise not communicate, and have connections with other power users. In SNA, these properties are identified with the four classical forms of centrality (degree, closeness, betweenness, and eigenvector) [12]; therefore, in our conception, a power user must simultaneously have very high values of all the four centralities (and thus be among the top users for each of them).

There are so many approaches to power user detection in the literature that it would be impossible to cover all of them in this paper. For example, the approaches described in [13, 14, 15, 16, 17, 18, 19, 20] are only recent approaches that search for power users by considering information other than centralities. In contrast, the approaches of [5, 6, 4, 7] are close to ours in that they are based on centralities. However, none of them considers closeness centrality, which instead plays a very important role and is orthogonal to other forms of centrality such as degree centrality [12]. Clearly, by imposing the need for high values of all four centralities simultaneously, our approach is extremely selective, implying that the power users it finds (if any) are very strong.

The rest of this paper is organized as follows: in Section 2, we illustrate the Threads dataset and the model used to represent Threads. In Section 3, we present the concept of power users and formulate an approach for power user detection in Threads. In Section 4, we characterize the detected power users. Finally, in Section 5, we draw our conclusions and look at possible future developments.

2. Dataset and Threads modeling

For our experiments, we constructed a dataset containing all posts and comments published in Threads from December 14, 2023 to February 21, 2024. It can be downloaded from the following GitHub repository: https://github.com/ecorradini/Threads_Dataset. It is anonymized to protect the privacy of Thread users. It is important to highlight that Threads has a feature that distinguishes it from other content-based social platforms in that each comment is itself a post. Therefore, for each post/comment, we stored the possible "parent post" so that we could reconstruct discussions conducted by multiple users through chains of posts/comments.

Once the dataset was constructed, it was necessary to define a model to represent Threads. To do this, we use a network $\mathcal{T} = \langle N, A \rangle$. N is the set of nodes in $\mathcal{T}$. There is a node $n_i \in N$ for each user who posted on Threads. Since there is a biunivocal correspondence between a node n_i and its corresponding user u_i, we will employ these two terms interchangeably in the following. A is the set of arcs in $\mathcal{T}$. An arc $a_{ij} = (n_i, n_j) \in A$ indicates that u_i posted a comment in response to a post made by u_j and, by implication, that u_j piqued u_i's interest.

3. Defining and detecting power users in Threads

In Table 1, we report some basic measures of $\mathcal{T}$. The examination of the values of these measures reveals a scenario typical of a new social network, in which interactions are still limited, users know each other little, and tend to interact on the basis of their content of interest rather than on the basis of their indegree or outdegree, as evidenced by the almost null value of indegree and outdegree assortativity.

Property	Value
Number of nodes	45,349
Number of arcs	72,333
Density	0.000035
Average clustering coefficient	0.000743
Diameter	13
Average shortest path	4.540
Maximum connected component's size	45,349
Average indegree	1.595
Average outdegree	1.595
Indegree assortativity	-0.042
Outdegree assortativity	-0.003

Table 1
Values of some basic measures of $\mathcal{T}$

After this initial analysis, since our definition of power users is based on the four centralities, we calculated the corresponding distributions. They are shown in Figure 1. We considered only the indegree centrality and not the outdegree centrality because it is precisely the indegree centrality that indicates whether a user in $\mathcal{T}$ has attracted the interest of other users.

From the figure, we can see that the distribution of indegree centrality follows a very steep power law, the distribution of closeness centrality resembles the superposition of two bell-shaped curves with different heights and a "half-bell-shaped" curve, and the distributions of betweenness centrality and eigenvector centrality follow a very steep power law. Basically, the four distributions respect what is predicted for them by social network theory [12]. Looking at them, we notice the presence of a small number of nodes that have extremely high centrality values. This is not entirely surprising, except for the fact that this is also the case for closeness centrality, which generally does not show this characteristic [12]. At this point, we can ask whether the nodes with high centrality in the four distributions are always the same or whether they are different. SNA tells us that they are generally different in the different centralities [12]. Therefore, if they were the same, we would be in the presence of very strong users.

As a first test of this hypothesis, we computed the Spearman's correlation coefficient [21] between the different centrality measures and saw that there is a strong correlation (equal to 0.46 on a scale between -1 and 1) between indegree centrality and closeness centrality, which should be uncorrelated for social network theory. This reinforces the idea that there may be some nodes in $\mathcal{T}$ having high values for all four centralities. To test this idea, we calculated the top 20% of nodes for each centrality, resulting in four sets of nodes. The 20% value is empirical and was chosen based on the fact that three of the four distributions follow a power law, as well as the desire to focus only on the most important nodes while not losing strong nodes, and thus potential power users. The 20% threshold represents a reasonable tradeoff between the latter two requirements.

At this point, we calculated the intersection of the four sets thus constructed and saw that

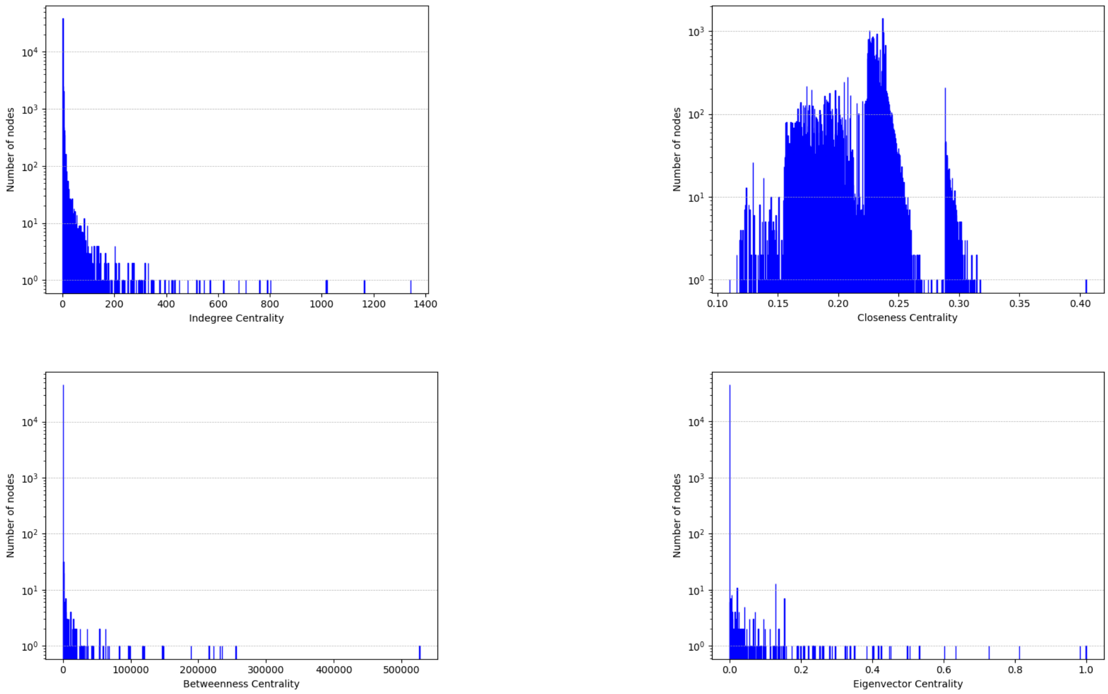

Figure 1: Distribution, in semi-log scale, of the nodes of $\mathcal{T}$ with respect to indegree centrality (top-left), closeness centrality (top-right), betweenness centrality (bottom-left) and eigenvector centrality (bottom-right)

it returned 1,176 users corresponding to the 2.59% of total users. Thus, we found that in our Threads dataset there is indeed a set of power users according to our definition. In the next section, we will analyze the main characteristics of these power users.

4. Characterizing power users in Threads

Considering the semantics of the four centrality measures, we can already define some characteristics of Threads power users. In fact: *(i)* they are important reference points for other users; *(ii)* the information they transmit can reach other users very quickly; *(iii)* they are able to transmit information between different Threads communities; and *(iv)* they are connected to other equally central users, which would lead us to hypothesize the presence of a backbone between them.

We calculated and compared the indegree of users and power users, and then the mean and median of these values. We saw that the mean (resp., median) indegree of power users is 11.96 (resp., 5) times greater than that of users. We expected this for the mean, given our definition of power users, but it was not obvious for the median. These median values tell us one important thing, i.e., that the overall indegree distribution is shifted upward for power users.

At this point, we checked whether there is a backbone connecting power users in Threads, i.e., whether they tend to prefer contacts with other power users rather than with other users. To do this, we considered the subnet $\mathcal{P}$ of $\mathcal{T}$ consisting only of power users and their connections, and measured the following parameters of $\mathcal{P}$ and $\mathcal{T}$: number of nodes, number of arcs, density, average clustering coefficient, diameter, average shortest path, average indegree, and normalized average indegree. The latter parameter was introduced by us and is defined as the ratio of the average indegree to the number of nodes in the network. It takes into account the fact that the same value of average indegree on a very large network or on a very small network has different implications. Table 2 shows the value obtained for these parameters.

Parameter	**Value in $\mathcal{T}$**	**Value in $\mathcal{P}$**
Number of nodes	45,349	1,176
Number of arcs	72,333	2,724
Density	0.000035	0.001971
Average clustering coefficient	0.000743	0.027748
Diameter	13	12
Average shortest path	4.540146	3.914330
Average indegree	1.595	2.316
Normalized average indegree	0.00003517	0.001969

Table 2
Values of some basic parameters in $\mathcal{T}$ and $\mathcal{P}$

From the analysis of this table, we can see that: *(i)* the density, average clustering coefficient, and normalized average indegree are much higher in $\mathcal{P}$ than in $\mathcal{T}$, meaning that power users tend to interact and be connected to each other much more than other users; *(ii)* the average shortest path and diameter are smaller in $\mathcal{P}$ than in $\mathcal{T}$. These results all point in the same direction, that is, they lead us to conclude that there is indeed a backbone among power users in Threads. This is an extremely significant result, because it suggests that there is a structured organization among these users that allows them to strongly influence the behavior of other users, despite the fact that they are very few in number.

All previous results have considered the structure of $\mathcal{T}$; now we want to go further and also examine the content of the posts/comments and see if there are communities of users with the same interests in Threads. From this point of view, Threads can be seen as a network of partially overlapping communities, each interested in a particular topic. In such a scenario, we want to see if power users act as connectors or bridges between different communities. If this were true, the backbone of power users would also act as a "glue" that holds the various Threads communities together.

To perform this analysis, we first had to find a way to analyze the content exchanged in Threads. For this purpose, we thought to look at the topics that users were discussing through their posts/comments. To do this in a simple but effective way, we used OpenAI's GPT-3.5 and asked it to extract, for each post/comment, the topic that best represented it. Doing this over our entire dataset, ChatGPT identified 531 topics. This means that there are 531 partially overlapping communities in our Threads dataset, each comprising all users who published at least one post/comment on the corresponding topic.

In Figure 2, we show the distribution of posts with respect to topics restricted to the top 50 topics. This distribution follows a power law. In particular, the two topics "Entertainment" and "Politics" have a much larger number of associated posts than the other topics. A third topic that still has a significant number of associated posts is "Technology". From the fourth topic on, we see a slow decrease in the number of posts associated with each topic.

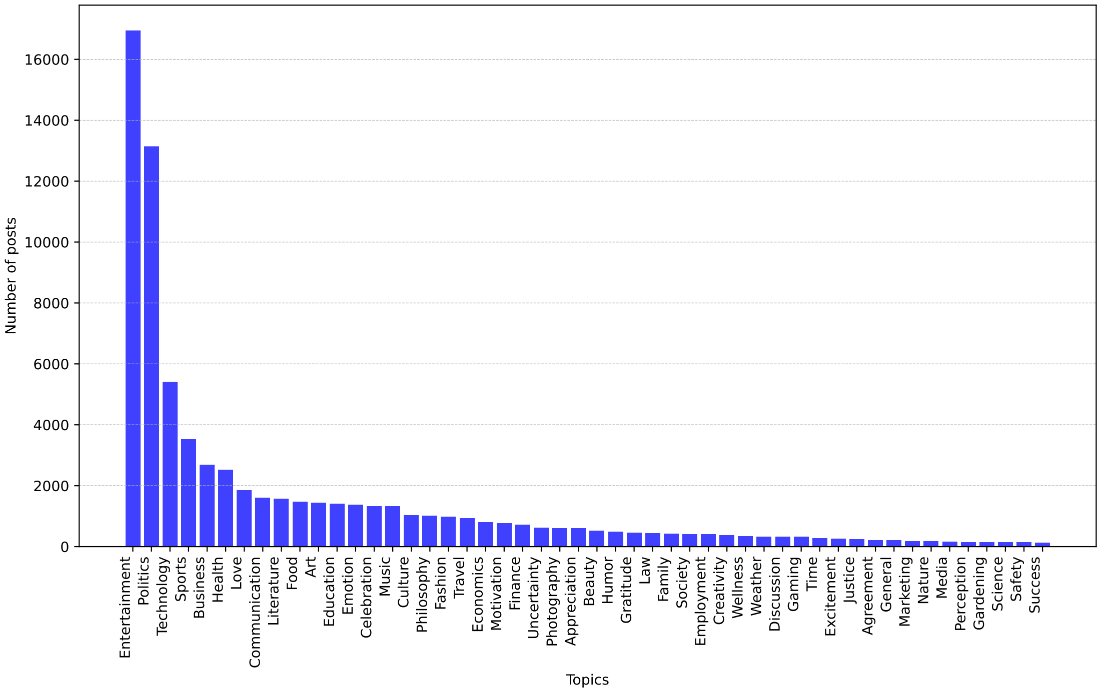

Figure 2: Distribution of posts with respect to topics (50 most frequent topics in $\mathcal{T}$)

In Figure 3, we show the distribution of power users with respect to topics restricted to the top 50 topics. This distribution is much less steep than the previous one. In fact, the two topics "Entertainment" and "Politics" still dominate the others, but this dominance is not as pronounced as in Figure 2. Again, we see a slow decline in the number of power users associated with each topic. Comparing Figure 3 with Figure 2, we can see some interesting differences. For instance, the topic "Technology", which was third in Figure 2, drops to eight in Figure 3, while the topic "Education", which was twelfth in Figure 2, rises to fifth in Figure 3.

In Figure 4 (resp., 5), we show the distribution of users (resp., power users) with respect to topics restricted to the top 50 topics. For each topic, we show the number of users (resp., power users) who published at least one post on it. In the figure, we consider two classes of users (resp., power users) called "unique" and "shared". Given a topic, the former consists of users (resp., power users) who published posts only on it, while the latter includes users (resp., power users) who published posts on it and at least one other topic.

By comparing the two figures, we can draw some important conclusions. First of all, there are some important differences in the position of topics in the two distributions. For example, "Technology" is ranked third in Figure 4 and eighth in Figure 5, while "Education" is ranked eleventh in Figure 4 and fifth in Figure 5. However, the most important information that can be obtained by comparing the two figures is the difference in the proportion of users and power users belonging to the "unique" and "shared" classes. In fact, for a given topic, the proportion of power users belonging to the "shared" class is generally larger than the corresponding

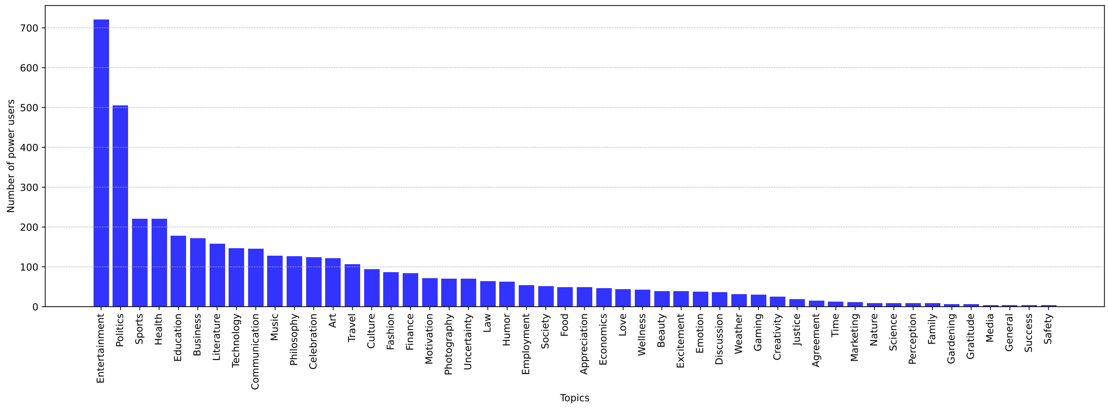

Figure 3: Distribution of power users with respect to topics (50 most frequent topics in $\mathcal{T}$)

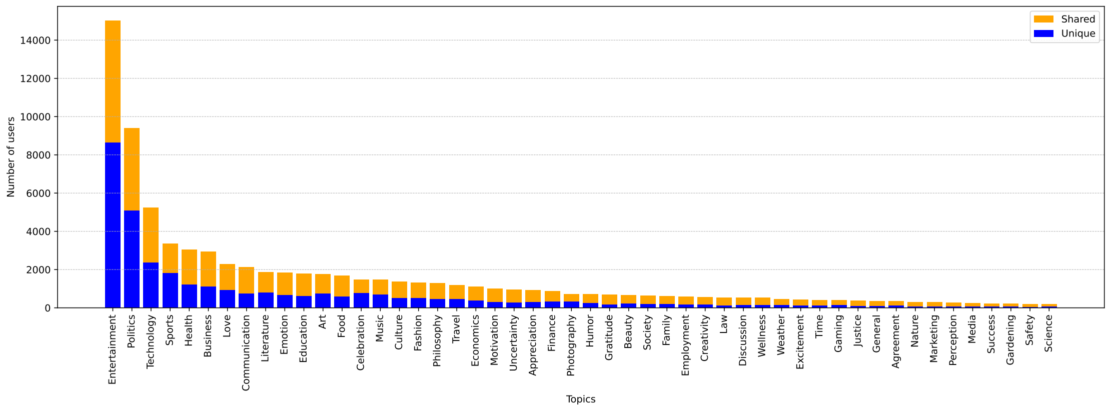

Figure 4: Distribution of users with respect to topics and their division into the classes "unique" and "shared" (50 most frequent topics in $\mathcal{T}$)

proportion of users. This seems to be a first confirmation of our hypothesis that power users can act as “bridges” (and their backbone as a “glue”) to hold the different communities of Threads together.

To test whether our hypothesis was true, we performed the following additional experiment: *(i)* we selected the 50 most frequent topics in terms of the number of users who made at least one post on them; these are the 50 topics that appear in Figure 4; *(ii)* we considered the 2,450 topic pairs that could be obtained from them; *(iii)* for each pair, we calculated the number of users in common between the two topics of the pair; *(iv)* for all pairs with a number of users in common greater than 0, we calculated the ratio of power users in common to users (involving power users) in common; *(v)* we averaged the values thus obtained.

The value of this average is 0.4637, which is much higher (in particular, 17.90 times higher) than the percentage of power users in $\mathcal{T}$ (which, as we have seen, was 0.0259). This result confirms the hypothesis that power users act as “bridges” between different Threads communities, and the backbone of power users acts as the “glue” that holds these communities together.

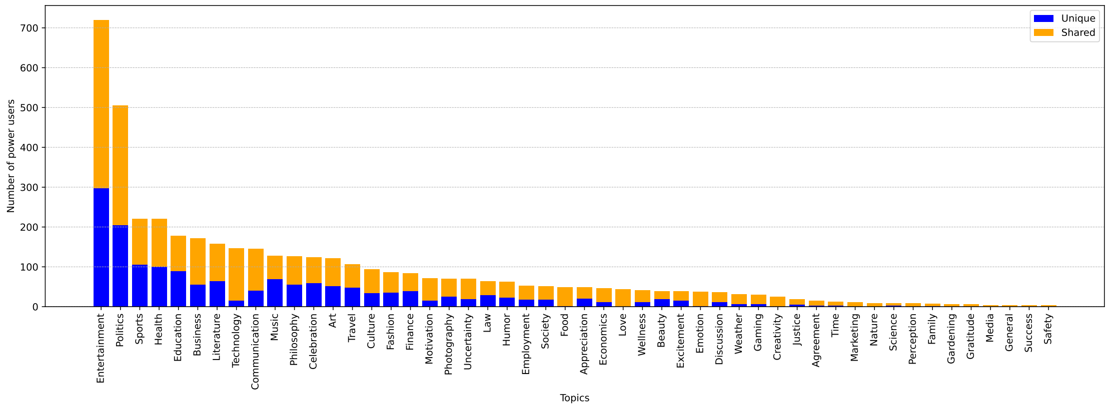

Figure 5: Distribution of power users with respect to topics and their division into the classes "unique" and "shared" (50 most frequent topics in $\mathcal{T}$)

With this analysis, we have completed our characterization of power users in Threads. We have shown that these users are critical to the information diffusion in this social platform. We have also shown that power users are able to: *(i)* influence user behavior, *(ii)* hold together the different communities that make up this social platform, and *(iii)* influence the Threads life and evolution.

5. Conclusions

In this paper, we have first proposed a new definition of power users based on the classic centrality measures of SNA, in order to exploit the knowledge accumulated on this topic over the years. Then, we have defined an approach for power user detection and applied it to Threads, and we have seen that there are indeed power users on this social platform. Finally, we have presented an experimental campaign to characterize power users in Threads. From a structural point of view, we have shown that power users form a backbone capable of rapidly spreading information in Threads and strongly influencing user behavior on this social platform. From a content point of view, we have shown that their backbone acts as a “glue” capable of holding together the different communities that make up Threads, which would otherwise risk remaining isolated.

Threads is a very young network and for this reason it is still little studied. Moreover, it has some peculiarities that distinguish it from all the other existing social networks, first of all its growth model based on Instagram. For this reason, we think it makes sense to conduct research in the future to better understand this platform. For example, we would like to define mechanisms to measure the trust, reputation and reliability of Threads users based on the posts/comments they publish and, more generally, on their behavior. Second, we would like to verify if there are phenomena of assortativity, both of status and of value, within Threads and, if so, we would like to investigate their causes. Last but not least, we would like to study the dynamics by which the different user communities in Threads are born, evolve, and eventually die.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] C. Wilson, B. Boe, A. Sala, K. Puttaswamy, B. Zhao, User interactions in social networks and their implications, in: Proc. of the ACM European Conference on Computer systems (EuroSys'09), Nuremberg, Germany, 2009, pp. 205–218. ACM.

[2] F. Buccafurri, V. Foti, G. Lax, A. Nocera, D. Ursino, Bridge Analysis in a Social Internetworking Scenario, Information Sciences 224 (2013) 1–18. Elsevier.

[3] F. Buccafurri, G. Lax, S. Nicolazzo, A. Nocera, Comparing Twitter and Facebook user behavior: Privacy and other aspects, Computers in Human Behavior 52 (2015) 87–95. Elsevier.

[4] J. Kratzer, C. Lettl, N. Franke, P. Gloor, The social network position of lead users, Journal of Product Innovation Management 33 (2016) 201–216. Wiley Online Library.

[5] Y. Yustiawan, W. Maharani, A. Gozali, Degree centrality for social network with Opsahl method, Procedia Computer Science 59 (2015) 419–426. Elsevier.

[6] P. Howlader, K. Sudeep, Degree centrality, eigenvector centrality and the relation between them in Twitter, in: Proc. of International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT'16), Bangalore, India, 2016, pp. 678–682. IEEE.

[7] Z. Alp, S. Öğüdücü, Identifying topical influencers on Twitter based on user behavior and network topology, Knowledge-Based Systems 141 (2018) 211–221. Elsevier.

[8] A. Zareie, A. Sheikhahmadi, M. Jalili, Influential node ranking in social networks based on neighborhood diversity, Future Generation Computer Systems 94 (2019) 120–129. Elsevier.

[9] X. Huang, D. Chen, D. Wang, T. Ren, Identifying influencers in social networks, Entropy 22 (2020) 450. MDPI.

[10] N. Subramani, S. V. Easwaramoorthy, P. Mohan, M. Subramanian, V. Sambath, A gradient boosted decision tree-based influencer prediction in social network analysis, Big Data and Cognitive Computing 7 (2023) 6. MDPI.

[11] B. Anastasiei, N. Dospinescu, O. Dospinescu, Word-of-mouth engagement in online social networks: Influence of network centrality and density, Electronics 12 (2023) 2857. MDPI.

[12] M. Tsvetovat, A. Kouznetsov, Social Network Analysis for Startups: Finding connections on the social web, Sebastopol, CA, USA, 2011. O'Reilly Media, Inc.

[13] K. Purba, D. Asirvatham, R. Murugesan, Influence maximization diffusion models based on engagement and activeness on instagram, Journal of King Saud University-Computer and Information Sciences 34 (2022) 2831–2839. Elsevier.

[14] S. Kumar, A. Mallik, A. Khetarpal, B. Panda, Influence maximization in social networks using graph embedding and graph neural network, Information Sciences 607 (2022) 1617–1636. Elsevier.

[15] S. Bartolucci, F. Caccioli, F. Caravelli, P. Vivo, Ranking influential nodes in networks from aggregate local information, Physical Review Research 5 (2023) 033123. APS.

[16] S. Wu, W. Li, H. Shen, Q. Bai, Identifying influential users in unknown social networks for adaptive incentive allocation under budget restriction, Information Sciences 624 (2023) 128–146. Elsevier.
[17] J. Bhadra, A. Khanna, A. Beuno, A Graph Neural Network Approach for Identification of Influencers and Micro-Influencers in a Social Network: Classifying influencers from non-influencers using GNN and GCN, in: Proc. of International Conference on Advances in Electronics, Communication, Computing and Intelligent Information Systems (ICAECIS'23), IEEE, Bangalore, India, 2023, pp. 66–71.
[18] S. Iqbal, R. Khan, R. Khan, F. Alarfaj, A. Alomair, M. Ahmed, Association Rule Analysis-Based Identification of Influential Users in the Social Media., Computers, Materials & Continua 73 (2022). Tech Science Press.
[19] M. Hasan, A. Bakar, M. Yaakub, Measuring user influence in real-time on twitter using behavioural features, Physica A: Statistical Mechanics and its Applications 639 (2024) 129662. Elsevier.
[20] W. Karoui, N. Hafiene, L. B. Romdhane, Machine learning-based method to predict influential nodes in dynamic social networks, Social Network Analysis and Mining 12 (2022) 108. Springer.
[21] J. Zar, Spearman rank correlation: overview, Wiley StatsRef: Statistics Reference Online (2014). Wiley Online Library.