The Emergence of Threads: The Birth of a New Social Network

Peixian Zhang, Yupeng He, Ehsan-Ul Haq, Jiahui He, and Gareth Tyson

The Hong Kong University of Science and Technology (GZ)

Abstract. Threads, a new microblogging platform from Meta, was launched in July 2023. In contrast to prior new platforms, Threads was born from an existing parent platform, Instagram, for which all users must already possess an account. This offers a unique opportunity to study platform evolution, to understand how one existing platform can support the "birth" of another. With this in mind, this paper provides an initial exploration of Threads, contrasting it with its parent, Instagram. Our findings reveal that Threads engages more with political and AI-related topics, compared to Instagram which focuses more on lifestyle and fashion topics. Our analysis also shows that user activities align more closely on weekends across both platforms. Engagement analysis suggests that users prefer to post about topics that garner more likes and that topic consistency is maintained when users transition from Instagram to Threads. Our research provides insights into user behaviour and offers a basis for future studies on Threads.

Keywords: Instagram · Threads · Topic analysis

1 Introduction

Meta's new microblogging service, Threads, was launched on July 5th, 2023. It became the fastest-growing 100 million consumer product within five days of its launch. In contrast to prior new social networks, Threads requires an Instagram account, thus enabling rapid social bootstrapping from a user's prior social network. Threads' design allows users to follow their Instagram social connections who already on Threads. The relationship between Threads and Instagram produces a distinct form of attention migration attributed to this unique "parent-child" setup.

The mass sign-up on Threads offers a unique opportunity to characterize the discourse of early adopters on Threads. More importantly, it is possible to track these early adopters on the parent platform and offer a comparative analysis of users' behaviors on the new platform. This is motivated by prior work that has revealed that users tend to exhibit varying norms on different platforms. For example, users share more negative sentiment and work-related content on Twitter than on Instagram [1]. Similarly, Instagram users tend to utilize profile images with lower smile scores compared to Twitter users [2], and they share more content on weekends than on weekdays [3]. In contrast, users migrating from Twitter to Mastodon did not change topical preferences [4]. We argue that the identification of such communication patterns is important for subsequent research on user modelling and discourse analysis [5].

Given the lack of research on Threads, we present the first analysis of early adopters of Threads, and compare their discourse with the same users on Instagram. This analysis provides a first exploration of Threads communication norms. We explore two research questions:

RQ1: What are topics are discussed on Instagram and Threads, and how do they differ? (Section 4).

RQ2: What factors result in a user changing their topic of discussion on Instagram and Threads? Do these factors impact users' selection of topics when they migrate from Instagram to Threads? (Section 5).

We find a difference in topical focus, with topics like `threads`, `twitter`, and `trump` being more prevalent on Threads as compared to Instagram. Then, we show that users who consecutively write on similar topics have lower feedback on their posts than those who write on different topics in their consecutive posts.

2 Related Work

We first discuss prior work on multi-social platform comparisons and the platform migration of users.

Multi-social network analysis. In order to conduct a comparative analysis of the same users' activities across many platforms, prior studies employ algorithms to establish connections between users' accounts on other platforms [6]. Additionally, they utilize biography websites [2] and examine self-mentions within the material across several platforms [7]. These methods might bring different selection biases among users who utilize biography websites or self-report their personal information. In contrast, the direct relationship between Threads and Instagram obviates the need for intermediary services, simplifying the process of cross-platform analysis. We posit this offers a powerful ground truth dataset. There have been prior studies that focus on analyzing cross-platform behaviours. Studies have found that users exhibit differences in self-description [3] and image usage [2] in profiles, alongside topic preferences [1] and activity time [3]. The engagement for similar content also varies across different platforms [1, 8]. Based on these previous findings, we aim to determine whether Threads exhibits comparable topical preferences to its parent, Instagram. To the best of our knowledge, this is the first work to measure user activity on Threads.

Platform Migration. Several works have studied the migration of users to different social media platforms. There are two types of migration: permanent migration and attention migration [10, 11]. [12] explore the push and pull factors that trigger migration on Reddit. There has also been prior work studying attention migrations from Twitter to Mastodon [11], and comparing the activities of the same users after migration [4]. A key contribution of our work is the identification of a third type of migration, which we term *parent-child migration*. This is where an existing platform spawns a new service, as is the case with Threads, which was born out of Instagram (e.g., only Instagram account owners can access Threads, and Threads posts appear on Instagram timelines as a promotional tool).

To the best of our knowledge, the closest study to our own is [13]. This interesting work explored the characteristics differentiating Twitter users who transition to Threads from those who stay behind. However, in our study, we utilize a larger dataset comprising millions of users on Threads. Furthermore, all of these previous investigations [11, 13, 4] focus on functionally similar platforms. In our research, Threads and Instagram represent distinct platforms that face a special migration [10]. To the best of our knowledge, this represents the first study to amalgamate data from both Instagram and Threads for a comparative analysis of user engagement.

3 Dataset

We now describe our dataset collection, pre-processing, and statistics in the following sections.

Threads. We use the threads-net API[1] to gather Threads data. We use the fact that each Threads account is associated with an auto-generated numeric ID. Starting from August 14th to September 4th (2023), we iterate over all integers from 0 to 12 million to identify any account with IDs in this range. This approach helps us overcome the limitation of searching the Threads platform for specific keywords and provides an extensive search range for early Threads adopters. For each Threads account, we collect the 25 most recent threads posts with their details from that account (25 is the limit imposed by the API). From September 4th to September 13th, we further extend this data by snowballing the identification of additional users, and collecting their posts, based on the reposted threads (*i.e.* users sharing other users' content, similar to retweets on Twitter). In total, there are 1,253,438 Threads accounts with 4,716,626 posts. There are 582,459 (46.47%) users without any threads, 543,522 (43.36%) users with fewer than 25 threads, and 127,457 (10.16%) users with 25 threads. We also find 212 threads posted before the Threads launched (which we exclude from the dataset). A possible explanation could be the presence of test accounts on Threads that Threads developers used to test the application's functions before its launch.

Instagram Threads and Instagram users share the same username, thus we search for all the usernames from the Threads dataset on Instagram. We use the CrowdTangle API[2] to retrieve all corresponding Instagram accounts and their posts from May 5 to September 13, 2023. In total, we collect 683,168 accounts with 10,862,421 posts on Instagram based on the user list on Threads. Note, we

[1] https://github.com/dmytrostriletskyi/threads-net
[2] `https://crowdtangle.com`

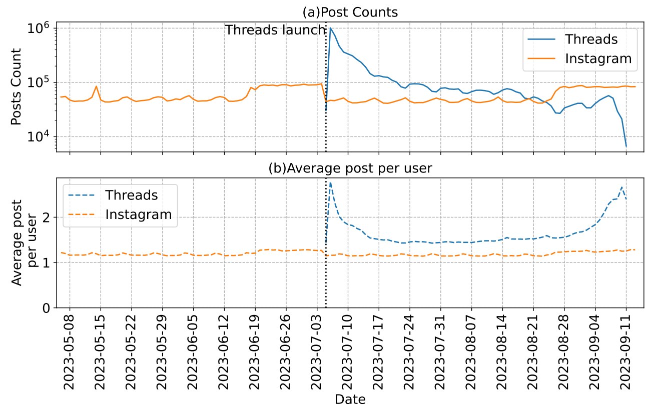


Fig. 1: (a) Daily volume of posts on Instagram and Threads (b) Daily average of posts per users

do not collect any Instagram accounts that have their profiles set as private, or any accounts that have changed their usernames. For better comparative analysis with Threads, we split the Instagram data into two parts: ***(i)*** Before Threads launch from May 4th to July 4th with 5,215,434 posts from 575,204 users; and ***(ii)*** After Threads launch from July 5th to Sept 9th with 5,646,987 posts from 587,941 users. This segmentation helps compare users on two platforms after they start posting on Threads. There are 479,877 common users in the two datasets.

Data pre-processing. We undertake several pre-processing steps. On Instagram. The distribution of posts per user is ($\mu = 9.6, median = 4, min = 1, max = 4,942$), ($\mu = 9, median = 4, min = 1, max = 4,016$) for before and after launch data segment, respectively. Thus, we exclude any users with posts more than the 99th percentile of number of posts per user distribution. This covers 83 accounts for Instagram before Threads launched and 89 after Threads launched. We also exclude the same users on Threads to ensure that the comparison is conducted on the same set of users. Figure 1 shows the (a) total post counts (solid line) and (b) the average post per user (dashed line) on two platforms daily from May 5 to September 13. For the topical analysis, we exclude reposts to avoid repeat counting topics on Threads. This covers 20.83% of Threads' posts. Note, Instagram does not have a repost feature. Hence, we do not apply this to Instagram. On Threads, 2.88% posts do not contain any text (*i.e.* they only contain images), preventing us from performing topical analysis. Thus, we also exclude these. Moreover, in the Instagram dataset, 9.38% of images have embedded text in the images, and these parts of the text have a median length of 18 words. For these posts, we combine the text in the images (CrowdTanlge provides this text) with the users' added caption. We do so because Instagram is an image-centric platform, and if users have to share a text-based message (such as a quote), they sometimes share it in the image [1]. Finally, we perform the following text pre-processing on the remaining Threads

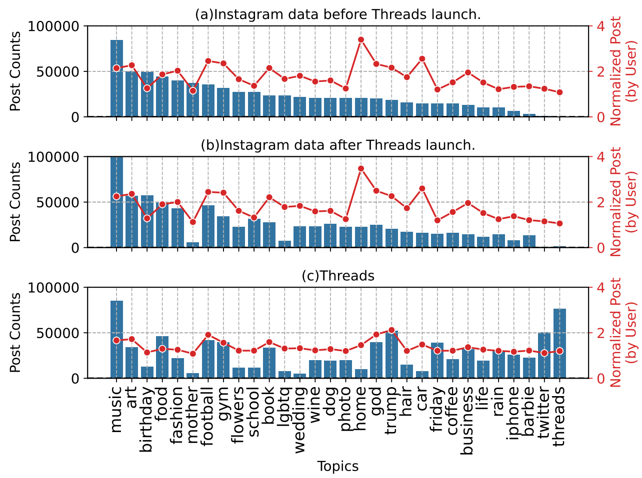


Fig. 2: Posts count for the top-30 topics (left-y) and normalised posts per user(right-y). (a) Instagram before Threads launch. (b) Instagram after Threads launch. (c) Threads. The x-axis is ranked based on the number of posts on Instagram before Threads launch.

and Instagram posts' text: ***(i)*** We only include English language posts, filtered by langdetect library.[3] 41.47% and 41.82% of total posts are in non-English languages on Threads and Instagram, respectively. ***(ii)*** We remove all hyperlinks. ***(iii)*** We remove any mentions of users (@username). ***(iv)*** We remove all non-alphanumeric characters, including emojis, punctuation, and special symbols (e.g., #, @, $). This leaves 3,533,095 posts from 556,730 users on Threads, 2,471,012 posts from 421,426 users on Instagram before Threads launch, and 2,673,157 number of posts from 436,223 users on Instagram after the Threads launch.

4 RQ1: Content-level analysis

We now analyse the topics on both platforms.

Topical Extraction We first look at the most common topics on each platform, and their weekly trends. We use BERTopic [14], which employs sentence transformers and c-TF-IDF to generate topics. We combine the posts from both platforms to create one corpus to ensure consistency and coherence in topic comparisons [4]. Using a threshold of at least 3,500 posts per topic, we identify 206 topics across the combined corpus. These topics cover 50.27% of posts in the data.

Topic Volume. We first look at the most common topics within each platform. We see that the prominence of topics varies across the two platforms. Figure 2

[3] `https://pypi.org/project/langdetect/`

presents the absolute number of posts (bars) and the normalized number of posts per user (red line) for the top 30 topics. The x-axis shows the topics ranked in order of their post count in the Instagram dataset before the Threads launch. Normalization is performed by calculating the number of users engaging in a given topic. Figure 2(a) and (b) have similar trends in terms of both the total number of topics and the normalized result. Some topics vary on Instagram before and after the Threads launch in the total number of posts. These topics are related to specific events occurring during that period, such as `mother`, `barbie (film)` and `lgbtq`, which refer to Mother's Day, film-release and LGBT pride, respectively.

Figure 2(c) presents the topics on Threads. When compared with Figure 2(b), we notice that topics about art (`music` and `art`), food (`food`, `coffee` and `wine`), and sports (`football` and `gym`) are popular on both Instagram and Threads. However, Threads also has distinct topics from Instagram. Topics on social media applications (`Threads` and `Twitter`) are more discussed on Threads. Politics (`trump`) is another topic discussed more on Threads than on Instagram. However, discussion patterns on the two platforms display a different trend. A total of 24,659 (4.43%) users wrote about topic `trump` on Threads. Yet there are only 9,113 (2.09%) users writing about `trump` on Instagram. Interestingly, each user on Instagram is more active though, with 2.26 posts per user about `trump` vs. 1.43 posts per user on Threads. Furthermore, Threads has fewer posts on daily lifestyle topics like fashion (`fashion` and `hair`) and pets (`dog`) as compared to Instagram, where these topics are popular [15]. Moreover, users prefer to share festivals and anniversary topics such as `birthday` and `wedding` on Instagram. Instagram has a higher total number of posts on these topics than Threads. 3,850 users post `wedding` relevant content on Threads, and 12,855 users post wedding on Instagram. 925 (7.20% of 12,855) users post this topic on two platforms. 11,111 users post `birthday` on Threads, and 44,677 users on Instagram. There are 1,914 (4.28% of 44,677) common users who post this topic on two platforms. Finally, we also notice a topic `AI` outside Figure 2. 10,509 (1.89%) users engaging in discussion on `ai` with an average of 2.12 posts per user on Threads. Only 3,282 (0.75%) users discuss `ai` on Instagram. This indicates a potential interest in AI-relevant topics on Threads that might not be as prominent on Instagram.

Overall, we note that while there are topical similarities on the two platforms, users on Threads post less on fashion-related topics and more on topics related to politics and technology than Instagram.

5 RQ2: Topic-consistency exploration

We next analyze the impact of the feedback received by posts in individual users' topical choices. We explore whether users' topical choices change as they move from Instagram to Threads. We first define the term topic consistency and then explore the changes in users' topic consistency within and across the platform.

5.1 Methodology

We define topic consistency as continuing the same topic in two consecutive posts. We further split this into two levels to examine topic consistency at the intra- and inter- platforms scale.

Intra-platform topic-consistency. If two consecutive posts by a user on a single platform share the same topic, the preceding post in time is labeled as having intra-platform topic consistency.

Inter-platform topic-consistency. Inter-platform topic consistency measures a user's continuation of the same topic from Instagram to Threads. To explore the topic consistency of users when migrating from Instagram to Threads, we inspect posts for each user. We first extract the topics of n recent posts on Instagram before the launch of Threads. We then extract the topics of n posts on Threads. If there is an overlap between these two sets, the user will be regarded as having inter-platform topic consistency. However, it is difficult to select the optimal threshold (n) for the number of posts to calculate this overlap. Hence, we experiment with three thresholds (*i.e.*, $n = 1, 2, 3$) and compare the results of these three alternatives.

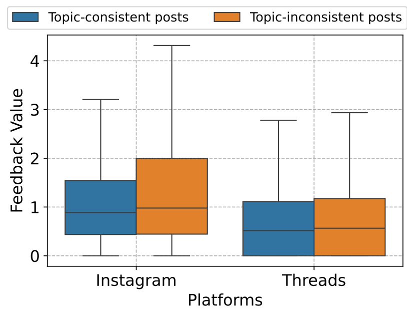
Feedback. We also measure feedback on posts, using the number of likes for each posts on both platforms. To calculate the *feedback* of a post, we first calculate the mean value of the likes of all posts published before this post. Taking the mean value of the previous likes and comparing it with the likes received by the current post helps in measuring whether the new post received better or worse engagement. We use the obtained mean value to divide the likes count by the current post, and use Laplace smoothing to avoid a denominator of zero.

5.2 Intra-platform exploration

We now estimate the feedback measure's impact on topic consistency at the post (intra-platform) and user (inter-platform) levels. We hypothesize that topic-consistent posts tend to have worse feedback since users with unpopular posts tend to focus more on their interests (narrow interests) compared with more popular users [16].

For this, we first divide all the posts into topic-consistent posts and topic-inconsistent posts. From all the posts that have been assigned a valid topic from BERTopic model (same model from RQ1), there are 709,272 (28.25%) topic-consistent posts and 1,801,559 (71.75%) topic-inconsistent posts on Instagram, while 115482 (8.93%) topic-consistent posts and 1178260 (91.07%) topic-inconsistent posts on Threads.

We now analyze the feedback differences on topic-consistent and inconsistent posts. Figure 3(a) shows the distribution of feedback for both Threads and Instagram. We find that the feedback value on Instagram (median = 0.95, std = 10853.97) is higher than that on Threads (median = 0.55, std = 180.74). In general, topic-inconsistent posts (median = 0.88, std = 9637.13) have relatively higher feedback than topic-consistent posts (median = 0.79, std = 7800.72).



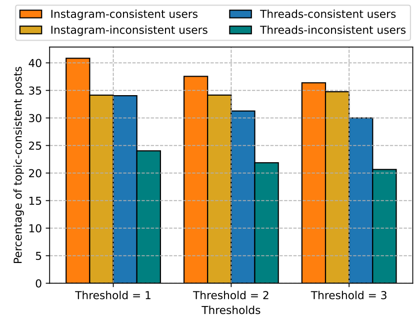


Fig. 3: (a) Box plot of feedback value for both topic-consistent posts and inconsistent posts on Instagram and Threads. (b) Percentage of topic consistency posts for consistent/inconsistent users on two platforms (with three different thresholds).

Unsurprisingly, these results indicate that topic-inconsistent posts are more likely to have better feedback, *i.e.* users who talk about a topic that receives a high number of likes are more likely to change the topic of their next post. This phenomenon is more common on Instagram than Threads. A potential explanation for this phenomenon is that users who prefer to publish posts on a wide variety of topics tend to be popular with audiences. Naturally, receiving positive feedback reinforces this message for publishers. But users who focus on same topics care little about the feedback.

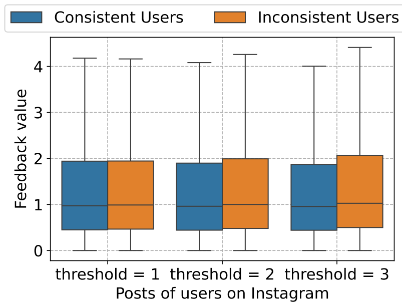
5.3 Inter-platform exploration

Next, we utilize the inter-platform consistency definition (users writing similar topics on Instagram and Threads in their consecutive posts) and divide users into consistent and non-consistent groups. We test three different thresholds on number of consecutive posts (from 1 to 3) to define the inter-platform topic consistency. Table 1 shows the number of consistent and inconsistent users with all thresholds.

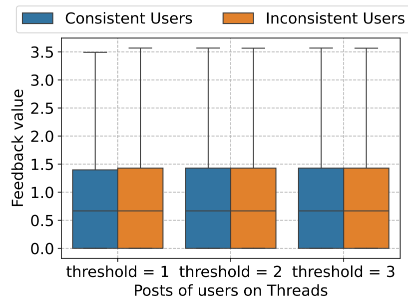
	Threshold =1	**Threshold =2**	**Threshold =3**
Consistent users	46,051	84,631	102,807
Inconsistent users	131,293	92,713	74,537

Table 1: The number of consistent/inconsistent users for different threshold.

We first analyze the percentage of topic-consistent posts by topic-consistent and -inconsistent users. Figure 3(b) shows these percentages for both Instagram and Threads with all thresholds. This figure shows that the percentage of topic-consistent posts (with threshold=1), on Instagram, from consistent users



(a) Feedback (Ins)



(b) Feedback (Threads)

Fig. 4: (a) Box plot of feedback for consistent/inconsistent users on Instagram. (b) Box plot of feedback for consistent/inconsistent users on Threads.

(40.83%) is higher than inconsistent users (34.14%). The corresponding percentage on Instagram is also higher than Threads. These results indicate that users who maintain topic consistency when they migrate from the "parent" platform to the "child" are also more likely to publish posts on the same topics consecutively. Further, the same group of users is more likely to maintain topic consistency on Instagram but change topics more frequently on Threads.

We finally analyze the differences in feedback on posts from these user groups and show the distribution in Figure 4. The median feedback count on Threads is less than 1 for both consistent and inconsistent users on all thresholds. On Instagram, the median value is close to 1 for both user types. A K-W test shows that there is no significant difference in feedback received by both user-groups, even though users' feedback on Instagram (median = 0.98, std = 15528.53) is relatively higher than that on Threads (median = 0.67, std = 118.56). These results indicate that there is little correlation between whether users maintain topic-consistency at inter-platforms scale and the feedback they receive. Further, generally users receive better feedback on Instagram compared with Threads.

6 Conclusion

Our study serves as the first characterization of Threads. We focus on the inteheritted overlaps between Threads (the "child") and Instagram (its "parent"). We have shown that there is a difference in topical focus across the two platforms (RQ1). Instagram has a higher percentage of topics related to lifestyle and fashion, while Threads has more topics related to politics and technology. Finally, we have explored the topic-consistency across platforms and compare the feedback (RQ2). We find consistency is more aligned with the content itself than about the feedback their posts receive (especially on Instagram). We also notice that users who maintain topic consistency when they migrate from Instagram to Threads are more likely to publish posts consecutively on the same topics.

7 Ethic Statement

We limit our analysis to public posts on both platforms. When gathering data, we witness usernames for both Instagram and Threads users. However, we exclude this from our analysis and make no attempt to de-anonymize users. We further exclude usernames from our content-based analysis.

References

1. L. Manikonda, V. V. Meduri, and S. Kambhampati, "Tweeting the mind and instagramming the heart: Exploring differentiated content sharing on social media," in *AAAI ICWSM*, 2016.
2. C. Zhong, H.-w. Chang, D. Karamshuk, D. Lee, and N. Sastry, "Wearing many (social) hats: How different are your different social network personae?" in *AAAI ICWSM*, 2017.
3. B. H. Lim, D. Lu, T. Chen, and M.-Y. Kan, "# mytweet via instagram: Exploring user behaviour across multiple social networks," in *IEEE/ACM ASONAM*, 2015.
4. J. He, H. B. Zia, I. Castro, A. Raman, N. Sastry, and G. Tyson, "Flocking to mastodon: Tracking the great twitter migration," in *ACM IMC*, 2023.
5. H. Kwak, C. Lee, H. Park, and S. Moon, "What is twitter, a social network or a news media?" in *ACM WWW*, 2010.
6. Y. Li, Z. Zhang, Y. Peng, H. Yin, and Q. Xu, "Matching user accounts based on user generated content across social networks," *Future Generation Computer Systems*, 2018.
7. M. Han Veiga and C. Eickhoff, "A cross-platform collection of social network profiles," in *ACM SIGIR*, 2016.
8. R. Farahbakhsh, A. Cuevas, and N. Crespi, "Characterization of cross-posting activity for professional users across facebook, twitter and google+," *Social Network Analysis and Mining*, vol. 6, pp. 1–14, 2016.
9. S. Kumar, R. Zafarani, and H. Liu, "Understanding user migration patterns in social media," in *AAAI AAAI*, 2011.
10. C. Fiesler and B. Dym, "Moving across lands: Online platform migration in fandom communities," *ACM on HCI*, no. CSCW1, 2020.
11. U. Jeong, P. Sheth, A. Tahir, F. Alatawi, H. R. Bernard, and H. Liu, "Exploring platform migration patterns between twitter and mastodon: A user behavior study," *arXiv preprint arXiv:2305.09196*, 2023.
12. E. Newell, D. Jurgens, H. Saleem, H. Vala, J. Sassine, C. Armstrong, and D. Ruths, "User migration in online social networks: A case study on reddit during a period of community unrest," in *AAAI ICWSM*, 2016.
13. U. Jeong, A. Nirmal, K. Jha, X. Tang, H. R. Bernard, and H. Liu, "User migration across multiple social media platforms," *arXiv preprint arXiv:2309.12613*, 2023.
14. M. Grootendorst, "Bertopic: Neural topic modeling with a class-based tf-idf procedure," *arXiv preprint arXiv:2203.05794*, 2022.
15. Y. Hu, L. Manikonda, and S. Kambhampati, "What we instagram: A first analysis of instagram photo content and user types," in *AAAI ICWSM*, 2014.
16. E. Ferrara, R. Interdonato, and A. Tagarelli, "Online popularity and topical interests through the lens of instagram," in *ACM HT*, 2014.