

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

Temporal dynamics of posts and user engagement of influencers on Facebook and Instagram

This is a pre print version of the following article:

Original Citation:

Availability:

This version is available <http://hdl.handle.net/2318/1843099> since 2022-02-23T14:56:23Z

Publisher:

Association for Computing Machinery

Published version:

DOI:10.1145/3487351.3488340

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

Temporal Dynamics of Posts and User Engagement of Influencers on Facebook and Instagram

Luca Vassio*, Michele Garetto[†], Carla Chiasserini[‡] and Emilio Leonardi[§]

* [‡] [§] Politecnico di Torino, Italy

[†] University of Turin, Italy

Email: *luca.vassio@polito.it, [†]michele.garetto@unito.it, [‡]carla.chiasserini@polito.it, [§]emilio.leonardi@polito.it

Abstract—A relevant fraction of human interactions occurs on online social networks. In this context, freshness of content plays an important role, with content popularity rapidly vanishing over time. We therefore investigate how influencers' generated content (i.e., posts) attracts interactions, measured by number of likes or reactions. We analyse the activity of Italian influencers and followers over more than 5 years, focusing on two popular social networks: Facebook and Instagram, including more than 13 billion interactions and about 4 million posts. We characterise the influencers' and followers' behaviour over time, show that influencers' posts are short-lived with an exponential temporal decay, and characterise the time evolution of the interactions from their initial peak till the end of a post lifetime. Finally, leveraging our findings, we develop an analytical model of the interactions temporal dynamics and validate it against experimental data.

Index Terms—Online Social Networks; Temporal dynamics; Popularity evolution; User engagement; Facebook; Instagram.

I. INTRODUCTION

Billions of people use online social media applications such as Facebook (FB) and Instagram (IG) as part of their daily online activities. Social media applications represent indeed a place to exchange opinions, getting news and maintaining social interactions through posts, comments, and likes. In particular, FB has been for quite a long time the most popular social media application, while IG has experienced a surge in popularity in the last few years. In both FB and IG, *influencers* (i.e., popular users, groups, newspapers, or companies) post content (i.e., the so-called *posts*) in the form of photos, videos or texts. All users of these social networks can *follow* influencers and *like/react* to such posts.

In spite of the significant attention paid to user-generated content in online social networks (OSN), most of the existing studies have analyzed content popularity as a function of the total number of interactions (views, likes, etc.), measured at the time data was crawled. Instead, little attention has been paid on how content popularity evolves over time.

A body of work that aims at predicting content popularity, considering its intrinsic characteristics and social interactions

features, can be found in [1], [2]. The main factors that impact FB post popularity are identified in [3], through an empirical analysis using multiple linear regressions. Similarly, [4] highlights the characteristics related to the dynamics of content production and consumption in IG, while [5], [6] predict the popularity of a future post on IG by combining user and post features.

Few works, instead, focus on understanding the temporal dynamics of content generated in OSNs. The decay in popularity over time, i.e., the rate of new interactions, of Internet memes [7] is shown to be well modeled by a negative exponential function. The work in [8] measures the time evolution of popularity of images in Flickr, finding that heavy-tailed distributions can represent the decay in rate of new interactions over time. Instead, [9] finds that the most popular Flickr pictures exhibit a close to constant interaction rate. The study in [10] models the popularity evolution of posts with Hawkes point processes, using Twitter data to fit the involved parameters. Finally, [4] shows that the distribution of likes to posts on IG is best fit by a power law, and it suggests that popularity of media measured by the number of likes might grow by a preferential attachment mechanism. However, [4] provides no evidence of this kind of evolution.

Notice that a large-scale characterization of the temporal evolution of the popularity of posts in OSNs is still missing. In this work, we aim at filling this gap by providing an experimental analysis of the time evolution of interactions with user-generated content, both on a per-post and per-influencer basis. To this end, we focus on the two popular social networks, namely, FB and IG, and investigate social network interactions through posts. We analyze, model, and compare user engagement and interactions, by leveraging a dataset of more than 13 billion interactions over approximately 4 million posts of 651 Italian influencers on these two OSNs. The collected dataset covers a period of more than 5 years, from January 1, 2016 to June 1, 2021.

In our previous work, we already showed how these interactions and their temporal dynamics are fundamental to understand phenomena occurring in OSNs, such as political discussion during the elections [11], and the impact of Covid-19 [12]. Here we analyze such data aiming at answering the following research questions. *What are the behavior over time and the popularity level of the posts published by the influencers? How do followers interact with such posts? In particular, what is the time evolution of the reactions to these*

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ASONAM '21, November 8–11, 2021, Virtual Event, Netherlands

© 2021 Association for Computing Machinery.

ACM ISBN 978-1-4503-9128-3/21/11/\$15.00

<https://doi.org/10.1145/3487351.3488340>

TABLE I: Features of the Italian influencers dataset we built, as recorded in 2021

	Influencers	Posts	Interactions	Avg followers
<i>Instagram</i>	244	0.31 mil.	9.36 bil.	1.19 mil.
<i>Facebook</i>	407	3.57 mil.	4.02 bil.	1.55 mil.

posts, and how can we model it?

Our main findings can be summarized as follows:

- Both influencer activity and user activity exhibit a characteristic daily pattern, but with a different shape;
- The inter-arrival time of posts has a long-tail distribution, reasonably fit by a log normal;
- On average 50% of user interactions occur within the first 4 hours after content creation on FB, and after 2 hours on IG; interactions arrival rate has an exponential temporal decay;
- Most of the posts are short-lived, with a lifetime between 20 and 50 hours, after which they no longer attract interactions;
- Content popularity can be well predicted by the interactions received within the first hour or even the first 15 minutes;
- The fraction of total interactions obtained within a given time interval is correlated with the number of newly published posts in the same interval.

We then discuss the above results, some of which have appeared in our conference paper [13], and develop an analytical model that captures the main aspects emerged from our experimental study. Importantly, we validate our model against experimental data and show that it can accurately represent user interactions on IG or FB, thus representing a useful tool for predicting user behavior on OSNs.

II. DATA COLLECTION

We focus on monitoring the activities triggered by top Italian influencers on FB and IG. To this end, we have built lists of the most popular Italian influencers, including different categories. Notice that, by the term influencer, we refer not only to individuals, but also to groups, football teams, newspaper, or companies. Those marked as Italian are the ones that communicate on the online social platform mainly using the Italian language. To get popular profiles, we have exploited the online analytics platform hypeauditor.com for IG, and www.socialbakers.com and www.pubblicodelirio.it for FB. The analysis has been restricted to the influencers that had at least 10,000 followers on June 1, 2021. The lists of influencers we used have been made publicly available.¹

For each monitored profile, we have downloaded the corresponding metadata, i.e., the profile information, and all the generated posts, using the CrowdTangle tool and its API². CrowdTangle is a content discovery and social analytics tool owned by Facebook, which is open to researchers and analysts

worldwide to support research, upon having a partnership agreement. Furthermore, for each post, we have downloaded the number of associated interactions, along with their timestamp. Monitored posts are sampled by CrowdTangle within the first 20 days (480 hours), with a higher sampling rate (up to few minutes) closer to the publication time of the post. Notice that, on IG, users can *like* posts, whereas on FB, they can *react* to posts with a thumbs up or other five pre-defined emojis. Thus, for each post, we have collected the *number of likes/reactions* the post received, hereinafter referred to as *interactions*, which are provided by CrowdTangle in an anonymized manner. Finally, we have stored the data, which takes around 110GB of disk space, on a Hadoop-based cluster and we have used PySpark for scalable processing.

For each influencer, we have downloaded all the data related to the posts published between January 1, 2016 and June 1, 2021. Table I reports the main features of our dataset, separately for each OSN. The 651 influencers show a large variability in the distribution of number of posts: some influencers published few tens of posts, while others, such as newspapers pages, up to 10^5 posts. Also, in the period under study, influencers on FB published more than those on IG. The main reasons are twofold: i) on FB more influencers are actually pages or organizations, rather than single individuals, and ii) many popular IG influencers did not exist at the beginning of the considered time period (i.e., in 2016), or have become active much later. The number of followers per influencer (as recorded on June 1, 2021) varies between 10k and tens of millions.

III. TEMPORAL USER ENGAGEMENT WITH POSTS

In this section, first we characterize the patterns of the influencers' and followers' activity (Sec. III-A), and then we analyze the time evolution of interactions (Sec. III-B). Finally, we investigate the correlation between the interactions a post attracts and the number of newly published posts (Sec. III-C).

A. Activity of influencers and followers

We first characterize the daily patterns. Fig. 1 presents the Probability Density Function (PDF) of the influencers' hourly activity, obtained considering the time instants at which the posts were published. The PDF accounts for all the analyzed 4 million posts. A 24-hour local-time clock is used, reported according to the ISO 8601 standard. Likewise, Fig. 2 shows the daily activity distribution of the followers (all 13 billions reactions have been considered).

We observe that followers' and influencers' activities exhibit similar patterns over the two considered social networks: followers' and influencers' activities significantly decrease during the night, with two peaks during the day. However, it is interesting to notice that followers tend to be more active later in the evening with respect to influencers. Also, the observed PDF of the influencers' activity suggests a strong non-stationary behavior.

We now investigate the distribution of the inter-arrival time between different posts. In particular, we focus on the

¹<https://mplanestore.polito.it:5001/sharing/KhoYSXAGR>

²<https://github.com/CrowdTangle/API>

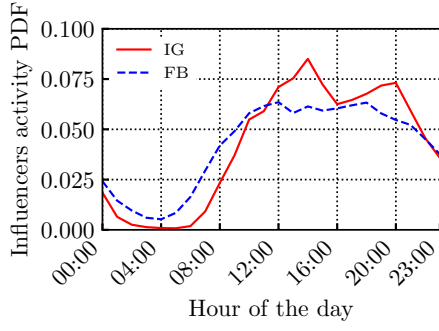


Fig. 1: Daily activity (i.e., creation of posts) of the sampled Italian influencers.

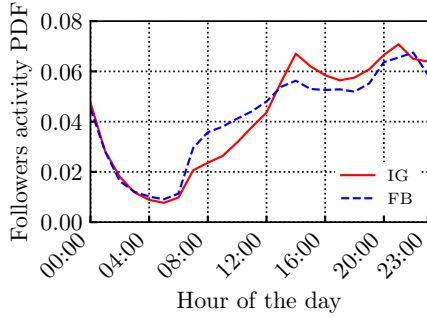


Fig. 2: Daily followers' activity (i.e., their interactions) with the influencers.

tail of the distribution, considering a time-scale of several hours. Fig.3 depicts the tail of the intertime of all posts generated by the influencers, including the best fitting log-normal distribution. The log scales in the plot suggest that the log-normal distribution provides a substantially better fit than what would be obtained by an exponential distribution (i.e., under a Poisson process assumption). This is due to the fact that influencers sometimes remain silent for (quite) long periods. We also analyzed single influencers, and found that, for the median influencer, the average posts interarrival time is equal to 19 hours on FB, and 57 hours on IG. Then, fitting separately each influencer with a lognormal distribution, on average we obtained a mean value of 2.0 and a standard deviation of 1.4 on FB, and a mean value of 3.1 and a standard deviation of 1.3 on IG.

B. Temporal dynamics of interactions

We now analyze the temporal evolution of the interactions on a post, considering up to 20 days (480 hours) after the creation of the post. We compute, for all the 4 million posts and for every sample-time (rounded to the closest integer hour), the fraction of received interactions with respect to the total number of interactions obtained by a post after 20 days. We consider fractions in order to compare different posts, and

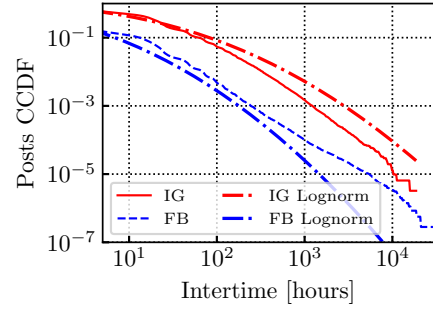


Fig. 3: CCDF of the posts intertime for the influencers (loglog scale).

different influencers. Finally, we compute the average over all posts.

The results, representing the dynamics of the average fraction of interactions over the first 3 days, are shown in Fig.4, using a log x-scale. One can notice that the majority of the interactions occurs within the first few hours. On average, the first hour accounts for 31% of all the interactions on FB (40% on IG), reaching more than 80% after 1 day. Moreover, on average, 50% of user interactions occur within the first 4 hours after content creation on FB, and after 2 hours on IG. Thus, the freshness of the post has a big impact on the attractiveness of a post. Interestingly, the growth of the number of user interactions is faster on IG than on FB. However, after around 30 hours, the two curves converge. Considering each influencer separately and considering all the times of the interactions, we obtain that posts on FB are indeed characterized by a higher interaction-time: on average, the expected value of the interaction-time is 15 hours for FB, and 11 hours for IG. Studying the evolution of the rate of new interactions, we found that we can approximate it almost perfectly in the first 24 hours with a negative exponential decay function (with mean equal to 5.4 for FB, and 8.7 for IG).

As expected, individual posts can have wildly different patterns of interactions accumulation over time. For example, we show results related to two specific posts on FB published by a well-known influencer (namely, Giuseppe Conte, former Italian Prime Minister). The temporal dynamics over the first 3 days since the posts creation, depicted with blue and red markers in Fig.5, shows a non-monotonic behavior of the rate of interactions, which first decreases, and, then, after some hours, increases again. We verified that this behavior is essentially due to the non-stationary behavior of users' activity during the day (see Fig. 2), i.e., quasi-flat portions of the curves correspond to night hours.

We now turn to the interesting question whether the total number of interactions collected by a post can be forecast by observing just the interactions received during a first initial period after publication. This is confirmed by the results in Fig.6, presenting the total number of interactions versus the interactions attracted after 1 hour, for the same influencer

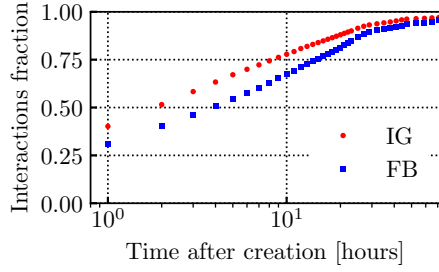


Fig. 4: Average evolution of a post in terms of interactions over 3 days.

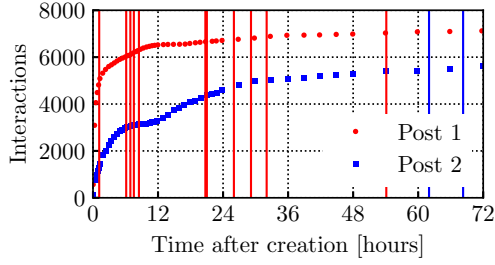


Fig. 5: Two examples of the evolution over 3 days of the interactions to a post.

(Giuseppe Conte). The corresponding value of the Pearson correlation coefficient is 0.85. This strong correlation holds even when considering the interactions received in the first 15 minutes (for which case, the Pearson correlation coefficient is 0.83). Further, note the high variance of the number of interactions received by a post, which can be attributed to the large diversity in the posts' intrinsic attractiveness. Similar considerations apply to other influencers on FB, as well as on IG.

Finally, we investigate the lifetime of the posts. We compute the fraction of interactions at the sampled timestamps, and compare them to the total number of interactions after 20 days. Then, for each post, we consider the time at which the post has reached the 95-th percentile of its interactions and define this as the lifetime of the post. To reduce the noise, we considered posts with at least 1,000 interactions. Fig. 7 depicts the distribution of the lifetime in hours, using a log x-scale. Interestingly, the difference between the two social networks is small, even though on average FB attracted a smaller fraction of interactions than IG within the first hours (see Fig. 4). The median value of the lifetime is 33 hours for both FB and IG, while the mean lifetime is 50 hours for FB and 55 for IG.

C. Impact of newly published content

We already observed that the arrival rate of new interactions decay exponentially with time (Sec.III-B). To better understand the nature of the arrival process of interactions generated by a specific post, we asked ourselves whether it is affected

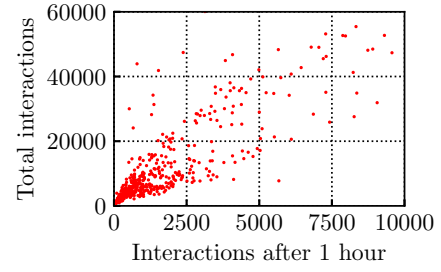


Fig. 6: Total no. of interactions vs. no. of interactions after 1 hour.

by the fact that, meanwhile, new posts are published by the same influencer, reducing the 'novelty' of the post. From Fig. 5 (where newly published posts from the same influencer are highlighted with vertical lines, in red for the first post and in blue for the second one), we observe that in the first post example many new posts are published within the first three days, whereas, in the second post example, no new post is published within the first 62 hours. On the other hand, after 12 hours the first post has already received 91% of its total interactions, while the second post, after the same amount of time, has collected only 70% of its interactions, due to the fact that its interaction rate decays more slowly. Hence, it is not easy to disentangle the impact of newly published content from the variability of the interaction rate.

To better understand the impact of newly published content, we considered a fixed period of 12 hours after the post creation, and computed the number of new posts published within this period. Fig. 8 shows the average fraction of interactions obtained after 12 hours, as a function of the number of newly published posts in the same period, for all posts by the previously considered influencer (Giuseppe Conte). We observe a clear correlation between the two quantities: the higher the number of new posts published within the first 12 hours, the faster the post approaches the end of its lifetime. Indeed, in the absence of newly generated posts, a post on average collects 72% of its total interactions within the first 12 hours; when 7 newer posts are generated in the same period, the average fraction of collected interactions increases to 82%. This clearly shows that the arrival rate of interactions also depends on how many new posts are published meanwhile after the post creation.

IV. DISCUSSION AND CONCLUSIONS

From our measurements and analysis, we have learned some important lessons to be kept in mind while developing a model for the temporal evolution of the number of interactions received by a post: (i) Posts are characterized by an initial attractiveness, which is highly heterogeneous even across the posts published by the same influencer; (ii) The growth rate of interactions naturally decays over time, but the decay rate is itself highly diverse from post to post, beside depending on the

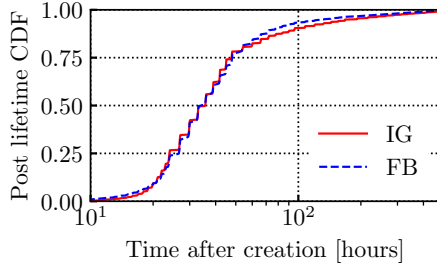


Fig. 7: Lifetime of a post computed as 95-th percentile of interactions.

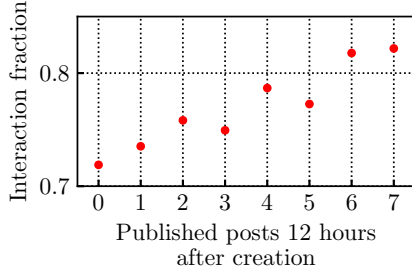


Fig. 8: Average fraction of interactions vs. number of published post, after 12 hours.

considered social network; (iii) The interaction rate should be modulated by the daily pattern of user activity, which appears to be independent of the particular online platform; (iv) The generation of new posts by the same influencer progressively fades away the attractiveness of the original post. This can be interpreted by the fact that users focus their attention at the top of the timeline.

Despite the intrinsic difficulties to incorporate all above features into a simple and tractable model, we believe that a modeling effort in this direction would be valuable. Indeed, our preliminary analysis suggests that, by observing the very initial phase of the posts lifetime, one can predict their popularity evolution and forecast the total number of interactions collected at the end, which could have interesting applications. Moreover, we plan to apply the derived dynamic popularity model to help predicting the time evolution of

followers variations, possibly including external factors [14].

ACKNOWLEDGEMENTS

This work has been supported by the SmartData@PoliTO center for Big Data and Data Science. We thank Prof. Martino Trevisan from Politecnico di Torino for his help in the data collection and his open source software FastPlot.

REFERENCES

- [1] H. Li, X. Ma, F. Wang, J. Liu, and K. Xu, "On Popularity Prediction of Videos Shared in Online Social Networks," in *Proceedings of the 22nd ACM CIKM*, 2013, p. 169–178.
- [2] G. Rizos, S. Papadopoulos, and Y. Kompatsiaris, "Predicting News Popularity by Mining Online Discussions," in *Proceedings of the 25th International Conference on World Wide Web (WWW)*, 2016, p. 737–742.
- [3] F. Sabate, J. Berbegal-Mirabent, A. Cañabate, and P. R. Leberherz, "Factors influencing popularity of branded content in Facebook fan pages," *European Management Journal*, vol. 32, no. 6, pp. 1001–1011, 2014.
- [4] E. Ferrara, R. Interdonato, and A. Tagarelli, "Online Popularity and Topical Interests through the Lens of Instagram," in *Proceedings of the 25th ACM Conference on Hypertext and Social Media*, 2014, p. 24–34.
- [5] M. Gayberi and S. G. Oguducu, "Popularity Prediction of Posts in Social Networks Based on User, Post and Image Features," in *Proceedings of the 11th ACM MEDES*, 2019, p. 9–15.
- [6] S. Carta, A. S. Podda, D. R. Recupero, R. Saia, and G. Usai, "Popularity Prediction of Instagram Posts," *Information*, vol. 11, no. 9, 2020.
- [7] J. Leskovec, L. Backstrom, and J. Kleinberg, "Meme-Tracking and the Dynamics of the News Cycle," in *Proceedings of the 15th ACM SIGKDD*, 2009, p. 497–506.
- [8] R. van Zwol, "Flickr: Who is Looking?" in *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*, 2007, p. 184–190.
- [9] M. Cha, A. Mislove, and K. P. Gummadi, "A Measurement-Driven Analysis of Information Propagation in the Flickr Social Network," in *Proceedings of the 18th International Conference on World Wide Web (WWW)*, 2009, p. 721–730.
- [10] A. Hassan Zadeh and R. Sharda, "Modeling brand post popularity dynamics in online social networks," *Decision Support Systems*, vol. 65, pp. 59–68, 2014.
- [11] C. H. Ferreira, F. Murai, A. P. Silva, J. M. Almeida, M. Trevisan, L. Vassio, M. Mellia, and I. Drago, "On the dynamics of political discussions on instagram: A network perspective," *Online Social Networks and Media*, vol. 25, p. 100155, 2021.
- [12] M. Trevisan, L. Vassio, and D. Giordano, "Debate on online social networks at the time of COVID-19: An Italian case study," *Online Social Networks and Media*, vol. 23, p. 100136, 2021.
- [13] L. Vassio, M. Garetto, C. Chiasserini, and E. Leonardi, "Temporal dynamics of posts and user engagement of influencers on facebook and instagram," in *Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, ser. ASONAM '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 129–133.
- [14] F. Bertone, L. Vassio, and M. Trevisan, "The Stock Exchange of Influencers: A Financial Approach for Studying Fanbase Variation Trends," in *Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 2021.