Evolution of Retweet Rates in Twitter User Careers: Analysis and Model

Kiran Garimella,[1] **Robert West** [2]

[1]MIT
[2]EPFL
garimell@mit.edu, robert.west@epfl.ch

Abstract

We study the evolution of the number of retweets received by Twitter users over the course of their "careers" on the platform. We find that on average the number of retweets received by users tends to increase over time. This is partly expected because users tend to gradually accumulate followers. Normalizing by the number of followers, however, reveals that the relative, per-follower retweet rate tends to be non-monotonic, maximized at a "peak age" after which it does not increase, or even decreases. We develop a simple mathematical model of the process behind this phenomenon, which assumes a constantly growing number of followers, each of whom loses interest over time. We show that this model is sufficient to explain the non-monotonic nature of per-follower retweet rates, without any assumptions about the quality of content posted at different times.

1 Introduction

This paper studies the evolution of individual impact in the context of social media. We focus on Twitter users' *retweet rates*—the number of retweets received by their posts ("tweets")—as a way to measure the impact and reach of their content. We aim to understand how this measure of impact changes over the course of users' "careers" (the time since they joined the platform). Specifically, we ask: Do users' retweet rates follow certain patterns? How does a user's retweet rate depend on the size of their audience? And when in a user's career is their content retweeted most?

Individual impact has been previously studied in contexts other than social media, including science (Radicchi and Vespignani 2009; Sinatra et al. 2016), work (DeNisi and Stevens 1981; Barrick and Mount 1991), sports (Yucesoy and Barabási 2016), and art (Dennis 1966; Liu et al. 2018). Measuring impact on social media is, however, inherently different, due to the highly dynamic nature of online social networks, which may grow (and sometimes shrink) by orders of magnitude within very short time frames, accompanied by numerous confounding factors, including temporal changes in audience size (number of followers), productivity (number of posts written), experience (age on the platform), interests (content of posts), and context (external events such elections). Studying the effects of these factors, and separating

them from true impact, is not trivial. For instance, we observe that on average the impact of individual users, as measured by the number of retweets their content receives, increases over the course of their careers. Is this due to a real change in content quality, or might it simply be explained by the fact that users accumulate more followers over time? or due to their increasing expertise in using the platform? or due to external events?

Because of the challenging nature of the problem, research into long-term trends of individual impact on social media, and on the factors that influence these trends, has been limited to date. Most research has attempted to understand the impact of isolated *pieces of content,* e.g., by predicting retweet counts on Twitter (Suh et al. 2010; Figueiredo and Almeida 2011; Kupavskii et al. 2012; Martin et al. 2016), video view counts on YouTube (Figueiredo and Almeida 2011), or image like counts on Instagram (Zhang et al. 2018). On the contrary, the literature is much thinner regarding the nature and role of the impact of individual *users*.

In order to make progress in this direction, we use an extensive dataset built from Twitter, which spans nearly a decade and contains complete careers of users with a wide range of audience size and experience. We also obtain historical follower-count estimates of the users in our dataset from the Internet Archive. Building on this novel dataset, we reconstruct the careers of users—from their very first posts until data collection time—and characterize the ebbing and flowing patterns of individual impact.

Our exposition proceeds in three parts. In Sec. 2, an empirical analysis of absolute and relative (per-follower) retweet rates over the course of Twitter user careers reveals an unexpected non-monotonic progression. In Sec. 3, we propose a mathematical model that captures the crucial non-monotonic aspect of the empirical data while relying on two simple, intuitive assumptions only. In Sec. 4, we verify that these assumptions are supported by the empirical data. Finally, in Sec. 5, we discuss our findings and sketch future directions.

2 Analysis of Retweet Rates Over Time

Data. We start from a large dataset spanning almost 10 years (2009–2018) and consisting of all 2.6 billion tweets posted by the more than 600K users who retweeted a U.S. presidential or vice-presidential candidate at least five times during

(a) Retweet count
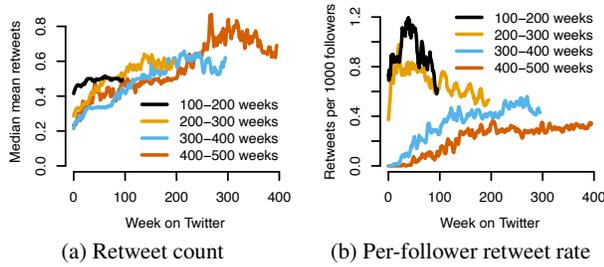
(b) Per-follower retweet rate

Figure 1: Evolution of (a) retweet counts, (b) per-follower retweet rates in user careers (running averages of width 5).

that period (Garimella et al. 2018). For each user, the dataset contains all tweets they posted. We focus on each user's original content and discard retweets of others' content. For each tweet, we also obtained the set of users who retweeted it and the number of times it was retweeted (as of June 2018). Furthermore, we obtained the set of followers for each user (as of June 2018), as well as time series of historical follower counts estimated via the Internet Archive's Wayback Machine, using Garimella and West's (2019) method. Reliable historical estimates of follower counts could be obtained for 23K users, who posted a total of 111M tweets. Our analyses focus on these 23K users.[1]

Retweet counts over time. We start with an empirical analysis of the evolution of retweet counts over the course of users' careers. Fig. 1a tracks aggregate retweet counts at weekly granularity (where $x = 1$ corresponds to a user's first week on Twitter). To obtain these curves, we first computed, for each user and week, the mean number of retweets obtained by the user for tweets posted that week, and then computed, for each week, the median over all users. We chose to consider medians of means because retweet-count distributions are heavy-tailed, with a small number of "viral" tweets being orders of magnitude more popular than most tweets (Goel et al. 2016). Such outliers can matter a lot for individual users, so to capture them, we consider per-user *means*. We do not, however, want weekly aggregates to be skewed by outlier tweets, so we consider per-week *medians* of per-user means.

Note that Fig. 1a contains four curves, each corresponding to a different career-length range (users who have been on Twitter for 100–200 weeks, 200–300 weeks, etc.). Studying users with different career lengths separately allows us to tease apart the evolution of users from the evolution of the platform and to avoid instances of Simpson's paradox, which might ensue by mixing user groups with systematically different characteristics.

Fig. 1a shows that the number of retweets obtained per week tends to grow over the course of user careers (with the exception of a final drop for the group of users who have been on Twitter for the longest time). Moreover, the curves have an overall concave shape, with growth rates slowing down as time progresses.

---

[1] Data available at https://github.com/epfl-dlab/retweet-evolution

Per-follower retweet rates. The growing number of retweets per week may be expected, considering that users tend to accrue followers over time. To account for this effect, we next consider a normalized version of the retweet rate, where each user's weekly mean number of retweets per tweet is additionally divided by the number of followers the user had at the time. Aggregating again over all users per week via the median gives rise to Fig. 1b. We observe that for some user groups the time series of per-follower retweet rates have a non-monotonic shape with a single peak. This is particularly the case for users who have been members of Twitter for shorter periods of time (100–300 weeks), whereas for longer-term members (300–500 weeks), the curves grow up to some point, from where on they start to plateau.

Evidence for a "peak age" has been found in a wide variety of fields such as poetry, mathematics, and theoretical physics (Adams 1946; Dennis 1966; Lehman 1953). In line with this literature, one potential narrative for explaining the non-monotonicity could be that users start as "newbies", then learn to be good social media users, and finally become old and lose touch with the rest of the community (Danescu-Niculescu-Mizil et al. 2013). In what follows, we will see, however, that such a complex narrative is not necessary to explain the data. A much simpler model suffices to generate the crucial qualitative shape of the retweets-per-follower curves.

3 Model of Retweet Rates Over Time

Our simple mathematical model makes two assumptions, shown visually in Fig. 2a and to be verified empirically later (Sec. 4). **Assumption I** states that a user's audience grows at a constant rate, e.g., by one follower per day. **Assumption II** states that each follower's retweet rate $f(t)$ (the expected number of retweets per time unit) decays as a power law:

$$f(t) = ct^{-\alpha}. \tag{1}$$

The number $F(t)$ of retweets at time t is then the sum of the retweets received from all followers accumulated by that time: $F(t) = \sum_{\tau=1}^{t} f(\tau)$. An example is shown in Fig. 2a: the magenta and green vertical lines mark specific points in the focal user's career; the intersections of the vertical lines with the followers' retweet-rate curves are indicated by magenta and green dots. By summing the values of the magenta or the green dots, we obtain the values indicated by the respective dots shown in Fig. 2b. By repeating this procedure for each time step, we obtain the full curve of Fig. 2b. The function F is increasing and concave (Fig. 2b), since f is decreasing and positive. The retweet rate per follower at time t is obtained via $G(t) = F(t)/t$ (without loss of generality, assuming one new follower per time step). As shown in Fig. 2c, G is a monotonically decreasing function.

Empirical analogues of F and G were displayed in Fig. 1. While the shape of the theoretical F (Fig. 2b) is approximately mirrored by empirical data (Fig. 1a), there is a discrepancy for G, which is monotonically decreasing in the model (Fig. 2c), but not so in empirical data (Fig. 1b). The discrepancy is resolved by letting users start not with zero but with $s > 0$ followers. As we will see later (Fig. 3a), this is in line with empirical data. An initial follower count $s > 0$ could potentially be caused by the platform recommending
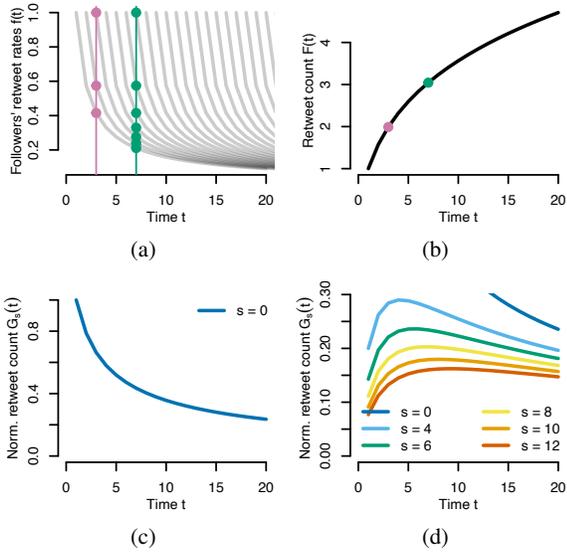
Figure 2: Model of retweet rates. (a) Retweet rates $f(t)$ of individual followers (one curve per follower, shifted along t-axis) decay as a power law. (b) A user's retweet count $F(t)$ at time t is obtained by summing the retweets from all followers at time t. (c–d) The relative retweet rate per follower, $G_s(t)$, is obtained by dividing the absolute number of followers by $s+t$, where s is the initial number of followers.

newcomers to other users more aggressively right after joining, so they would quickly accumulate a certain number of followers immediately in the beginning. With $s \geq 0$ initial followers, the relative number of retweets per follower is $G_s(t) = F(t)/(s+t)$, which has an internal maximum for $s > 0$ (Fig. 2d) and captures the essence of the empirical curves of Fig. 1b.

The curves shown in Fig. 2a–d were obtained empirically for a specific set of parameters ($\alpha = 0.8, c = 1$). In the general case, the model is more easily analyzed when assuming continuous, rather than discrete, time:

$$F(t) = \int_0^t f(\tau + \epsilon)\, \mathrm{d}\tau \qquad (2)$$

$$= \int_0^t c(\tau + \epsilon)^{-\alpha}\, \mathrm{d}\tau \qquad (3)$$

$$= \frac{c}{\alpha - 1}\left(\epsilon^{-(\alpha-1)} - (t+\epsilon)^{-(\alpha-1)}\right), \qquad (4)$$

for $\alpha \neq 1$.[2] The offset $\epsilon > 0$ is necessary to be able to handle the case $\alpha \geq 1$, where $\int_0^t f(\tau)\, \mathrm{d}\tau = \int_0^t c\tau^{-\alpha}\, \mathrm{d}\tau = \infty$ would diverge.

In the continuous case, F has the same increasing concave shape as in the discrete case (Fig. 2b). The shape of G_s (the retweets per follower over time) is analyzed in detail in Ap-

---

[2] The case $\alpha = 1$ needs to be handled separately: here, $F(t) = c\left(\log(1 + t/\epsilon)\right)$.
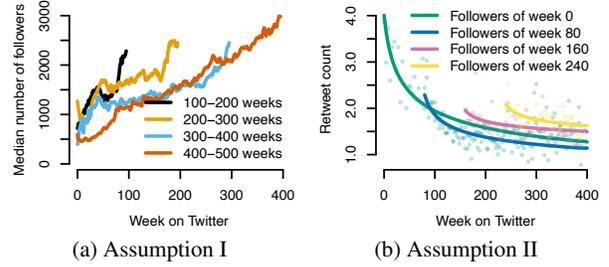

(a) Assumption I (b) Assumption II

Figure 3: Empirical validation of model assumptions.

pendix A,[3] where we show that G_s attains a unique internal maximum at a time $t^* > 0$ if and only if $s > 0$ (as in the discrete case; Fig. 2d).

To summarize, the non-monotonic curves indicating a "peak age" on Twitter (Fig. 1b) do not require a complex narrative as sketched earlier (inexperienced youth, age of maximal alignment with community norms, losing touch with newer members in old age). Rather, they can be explained with a much simpler model containing just two ingredients: (1) a constant influx of new followers, (2) each of whom loses interest over time according to a power law.

To be clear, we do not claim that our model fully captures all the dynamics of retweet rates. Rather, we have shown that the seemingly complex, non-monotonic evolution of retweets per follower directly follows from two intuitive assumptions, which are shown to approximately hold in empirical data in the following section.

As a side note, we remark that, for $\alpha > 1$, $F(t)$, the number of retweets per time unit, is upper-bounded by a constant $\frac{c\epsilon^{-(\alpha-1)}}{\alpha-1}$. That is, for $\alpha > 1$, the model predicts an inherent barrier on the number of retweets a user can get in a single time unit, even if her audience grows at a constant rate: the drop-off of interest would be too steep to be compensated by the influx of new followers. Later, we will observe empirically (Fig. 4) that the empirical exponents α are well below 1, which theoretically implies a potentially unbounded number of retweets per time unit.

4 Validation of Model Assumptions

We now verify that the model's assumptions hold empirically.

Assumption I: Constant audience growth. Fig. 3a shows that the median number of followers indeed grows approximately at a constant rate. We also see that, empirically, users tend to start off with a nonzero number $s > 0$ of followers, which is the condition for which the per-follower retweet rate $G_s(t)$ is non-monotonic in our model (Fig. 2d). Non-monotonic empirical curves (Fig. 1b) are thus in line with the shape predicted by the model.

Assumption II: Decay of followers' interest. To verify the assumption that each follower's retweet rate $f(t)$ decays as a power law, we proceed as follows. For each week t_0 in a user u's career, consider all followers who retweeted u for the

---

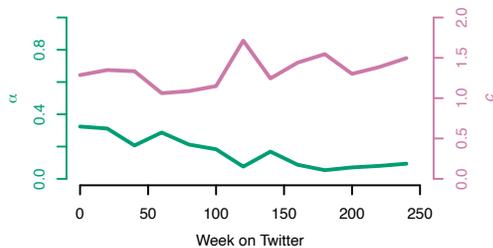[3] Appendices available online at http://arxiv.org/abs/2103.10754

Figure 4: Empirically estimated parameters (power-law exponent α and multiplicative factor c) of followers' retweet rates $f(t) = ct^{-\alpha}$, as functions of the week in the focal user's career when the followers began to follow the focal user.

first time that week and track them for the rest of u's career, computing these followers' mean numbers of retweets of u's tweets for each subsequent week $t' \geq t_0$, and aggregating weekly over all users u via the median. The result is shown in Fig. 3b (for users who have been on Twitter for at least 400 weeks; for clarity's sake, curves are shown for only four starting weeks t_0, but in practice we computed curves for all t_0). We observe that decreasing power laws $ct^{-\alpha}$ (solid lines in Fig. 3b) fit followers' interest well. Irrespective of when a follower starts retweeting, their interest in the followed user drops off with time. Fig. 4, which tracks α and c over time, shows that the drop-off in interest becomes less steep over time (as indicated by the decreasing values of α).

5   Discussion

Summary. Our empirical analysis (Sec. 2) of the "careers" of thousands of Twitter users, starting with each user's very first tweet, revealed that the number of retweets a user receives per week grows steadily, at a decreasing speed (Fig. 1a). The growth is partly due users' tendency to accrue followers over time (Fig. 3a); the decreasing speed, due to followers' tendency to lose interest over time (Fig. 3b).

Accounting for the growing number of followers by dividing weekly retweet counts by the number of followers at the given time reveals partly non-monotonic curves of retweets per follower, especially for more recent users (Fig. 1b). One might be tempted to explain this shape by positing a prototypical user life cycle, where users start off unacquainted with the norms of the platform, subsequently produce content that is gradually becoming more attractive to others, until they reach a peak age after which they cease to be "in tune" with the rest of the community, resulting in decreasing retweet rates per follower from there on.

By formulating a mathematical model (Sec. 3), we show that such a complex narrative is not required in order to explain the non-monotonic shape of retweets-per-follower curves. Rather, we can reproduce the qualitative shape of the curves by making only two simple, intuitive assumptions, both of which were verified to hold empirically in the data (Sec. 4). The assumptions state that (1) users gain followers at a steady rate, and (2) the followers lose interest in the followed user's content according to a power law. We emphasize

that it is not our goal to prove that the model captures all the dynamics that drive retweet rates—in fact, there are probably many factors and facets that remain unmodeled. Rather, we show that the simple model is sufficient to explain key aspects of the data.

Further Twitter datasets. The above analyses were performed on a Twitter dataset that is particularly well suited for studying entire user careers, as it contains all tweets posted by a predetermined set of users. Collecting this dataset was demanding, as it required custom-made crawling tools, due to the fact that Twitter's API only provides access to each user's 3,200 most recent tweets (Garimella and West 2019). Although our user sample is not random, but biased towards politically interested users (Garimella et al. 2018), we consider this topical bias to be less severe than the bias that would be induced by studying only users with fewer than 3,200 tweets, the limit imposed by Twitter's API. Nevertheless, to further corroborate our findings, we ran our analyses on three other Twitter datasets (Garimella and West 2019), which are not topically biased, but are subjected to the 3,200-tweet limit. The results, reported in Appendix B (cf. footnote 3), are qualitatively coherent to those reported here.

Beyond Twitter. Going even further, we collected and analyzed full user careers from two other platforms, Instagram and YouTube. Although these platforms are rather different from Twitter, we found patterns for absolute and per-follower like counts on Instagram and view counts on YouTube that mirror the patterns for retweet counts found on Twitter. The results are presented in Appendix B (cf. footnote 3).

Future work. To the best of our knowledge, this is the first longitudinal study of complete user careers on social media. Our findings, public dataset, and models can serve as a stepping stone for further research in this area. For instance, whereas here we only considered the *number* of followers, one could also use heuristics (Meeder et al. 2011) to approximate the actual *set* of followers. This could allow us to study how the follower network evolves and how it influences, and is influenced by, retweeting behavior (Myers and Leskovec 2014). Moreover, it would be interesting to define scores for measuring the impact of social media users, similar to the scores used to measure the impact of scientists (Sinatra et al. 2016). From a practical perspective, it would be useful to infer long-term behavioral patterns associated with successful user careers. For instance, can the peak age be delayed, or the drop-off be slowed down, by acting in certain ways? Beyond science alone, the answers to such questions would be of value for users, marketers, and social media platforms striving for better user retention.

Acknowledgments

References

Adams, C. W. 1946. The age at which scientists do their best work. *Isis* 36(3/4): 166–169.

Barrick, M. R.; and Mount, M. K. 1991. The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology* 44(1): 1–26.

Danescu-Niculescu-Mizil, C.; West, R.; Jurafsky, D.; Leskovec, J.; and Potts, C. 2013. No country for old members: User lifecycle and linguistic change in online communities. In *WWW*.

DeNisi, A. S.; and Stevens, G. E. 1981. Profiles of performance, performance evaluations, and personnel decisions. *Academy of Management Journal* 24(3): 592–602.

Dennis, W. 1966. Creative productivity between the ages of 20 and 80 years. *Journal of Gerontology* 21(1): 1–8.

Figueiredo, F.; and Almeida, J. 2011. The tube over time: Characterizing popularity growth of YouTube videos. In *WSDM*.

Garimella, K.; Morales, G. D. F.; Gionis, A.; and Mathioudakis, M. 2018. Political discourse on social media: Echo chambers, gatekeepers, and the price of bipartisanship. In *WWW*.

Garimella, K.; and West, R. 2019. Hot streaks on social media. In *ICWSM*.

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

Kupavskii, A.; Ostroumova, L.; Umnov, A.; Usachev, S.; Serdyukov, P.; Gusev, G.; and Kustarev, A. 2012. Prediction of retweet cascade size over time. In *CIKM*.

Lehman, H. C. 1953. *Age and Achievement*. Princeton University Press.

Liu, L.; Wang, Y.; Sinatra, R.; Giles, C. L.; Song, C.; and Wang, D. 2018. Hot streaks in artistic, cultural, and scientific careers. *Nature* 559(7714): 396.

Martin, T.; Hofman, J. M.; Sharma, A.; Anderson, A.; and Watts, D. J. 2016. Exploring limits to prediction in complex social systems. In *WWW*.

Meeder, B.; et al. 2011. We know who you followed last summer: Inferring social link creation times in Twitter. In *WWW*.

Myers, S. A.; and Leskovec, J. 2014. The bursty dynamics of the Twitter information network. In *WWW*.

Radicchi, F.; and Vespignani, A. 2009. Diffusion of scientific credits and the ranking of scientists. *Physical Review E* 80: 056103.

Sinatra, R.; Wang, D.; Deville, P.; Song, C.; and Barabási, A.-L. 2016. Quantifying the evolution of individual scientific impact. *Science* 354(6312).

Suh, B.; Hong, L.; Pirolli, P.; and Chi, E. H. 2010. Want to be retweeted? Large scale analytics on factors impacting retweet in Twitter network. In *SocialCom*.

Yucesoy, B.; and Barabási, A.-L. 2016. Untangling performance from success. *EPJ Data Science* 5(1): 17.

Zhang, Z.; Chen, T.; Zhou, Z.; Li, J.; and Luo, J. 2018. How to become Instagram famous: Post popularity prediction with dual-attention. In *BigData*.