

# The Road to Popularity: The Dilution of Growing Audience on Twitter

**Przemyslaw A. Grabowicz**  
MPI-SWS, Germany

**Mahmoudreza Babaei**  
MPI-SWS, Germany

**Juhi Kulshrestha**  
MPI-SWS, Germany

**Ingmar Weber**

Qatar Computing Research Institute, HBKU, Qatar

## Abstract

On social media platforms, like Twitter, users are often interested in gaining more influence and popularity by growing their set of followers, aka their *audience*. Several studies have described the properties of users on Twitter based on static snapshots of their follower network. Other studies have analyzed the general process of link formation. Here, rather than investigating the dynamics of this process itself, we study how the characteristics of the audience and follower links change as the audience of a user grows in size on the road to user's popularity.

To begin with, we find that the early followers tend to be more elite users than the late followers, i.e., they are more likely to have verified and expert accounts. Moreover, the early followers are significantly more similar to the person that they follow than the late followers. Namely, they are more likely to share time zone, language, and topics of interests with the followed user. To some extent, these phenomena are related with the growth of Twitter itself, wherein the early followers tend to be the early adopters of Twitter, while the late followers are late adopters. We isolate, however, the effect of the growth of audiences consisting of followers from the growth of Twitter's user base itself. Finally, we measure the engagement of such audiences with the content of the followed user, by measuring the probability that an early or late follower becomes a retweeter.

## 1 Introduction

Politicians are sometimes accused of abandoning principles from their earlier days as they rise to power. However, one might argue that they simply adapt their message to be in line with their growing – and evolving – target audience. Similarly, as a once unknown actor reaches stardom, their audience changes from a few “niche” fans and personal friends, to millions worldwide. In this paper, we study how an average Twitter user's audience, and her engagement with this audience, changes as she accumulates followers.

We start from an observation by Chen et al. that “the pope is not a scaled up bishop” (Chen, Weber, and Okulicz-Kozaryn 2014). They observed that same-religion linkage preference on Twitter is lower for users with lots of followers, such as the pope, than it is for less visible users. This

led us to hypothesize that as a user grows her audience, the level of topical interest and engagement of the audience decays – or as we will say “dilutes”. Understanding if such an effect exists has important consequences, because general alignment between audience and message is fundamental for successful communication (Grabowicz, Ganguly, and Gummadi 2016), not only on Twitter.

For instance, changing properties of a growing audience may have an influence on information diffusion. Previous studies show that pieces of topical information spread more readily among users interested in the topic of such memes (Grabowicz, Ganguly, and Gummadi 2016) and that the retweeting behavior is influenced by the posting time of information (Yang et al. 2010). These results suggest that information diffusion is less efficient if a growing audience becomes less topically aligned with the messages or less synchronized in time.

Studying the change in audience composition and its dilution for a given Twitter user is challenging, as many confounding factors have to be taken into account. First of all, a random new Twitter user today will be different from a random new Twitter user eight years ago. In fact, Liu et al. (Liu, Kliman-Silver, and Mislove 2014) show that the distribution of location and language of new Twitter users changes over time. Second, existing users also continue to evolve. Garcia et al. (Garca-Gavilanes et al. 2014) observe that “mature users evolve to adopt Twitter as a news media rather than a social network”. Third, the links in the social network evolve as the network grows (Leskovec et al. 2008). Though an evolving network will surely effect information diffusion, it is less clear how information diffusion relates to the changes in audience composition.

Here, we distinguish between two different hypotheses. Namely, we hypothesize that an audience, as it grows, it dilutes its properties, such as: elitness or the coherence of time zone, language, and topics of expertise. We define as an *audience*: (i) Twitter's internal audiences, where each audience is defined as all followers of an expert, and (ii) the user base of Twitter itself. We would like to learn if any of these audiences or both of them dilute their properties as they grow.

In our analysis, we isolate the effects induced by the dilution of a growing audience of a particular expert user and the dilution of Twitter's growing user base itself. To this end, as a control variable we use the time when a given follower

joined Twitter, i.e., the *age* of that user in the system. This allows us to measure both effects of the growing number of followers and Twitter’s user base on the composition of such audiences. We find that:

- As the number of followers grows for a given expert user, the composition of these followers dilutes. We find examples of this rule in the eliteness of followers and their topical coherence.
- The composition of Twitter’s user base also dilutes. Namely, the eliteness of users joining Twitter drops and so does their time zone and language coherence.

In what follows, we first describe the dataset that illustrates our findings. Then, we introduce the random shuffling method that is used to distinguish between the growth of the number of followers and Twitter’s user base (Section 3.1). Next, we measure the properties of these growing audiences (Section 3) and their engagement estimated as the fraction of followers diffusing information (Section 3.3).

## 2 Description of the dataset

In this work, we analyze followers of *expert* users. We define an expert as a user that was included in multiple Twitter lists (Ghosh et al. 2012).<sup>1</sup> For each of the expert users, we know the topics of their expertise as they have been explicitly “tagged” by the list curators. Also, we represent expertise of each user as a normalized topical distribution over her topics of expertise (Kulshrestha et al. 2015). Expert users tend to produce more high-quality content and less spam and they are more engaged than random Twitter users (Ghosh et al. 2013). We choose to focus on random expert users instead of all users, due to these qualities.

In total, we identified around 1.6M experts in Twitter as of November 2015. Note that users with many followers are likely to be included in this set.<sup>2</sup> For the purpose of this study, we sample a set of 1K random expert users. More specifically, we select random experts from five different bins of the number of followers. Namely, we select 200 experts that each had 100 to 1K, 1K-10K, 10K-100K, 100K-1M, or 1M-10M followers in the end of November 2015.

Then, using Twitter’s REST API, we retrieve all followers of our expert users and the profile information of these followers. In total, we gather information about 241M users. Furthermore, we collect all tweets of our expert users that were posted during three months between September and October 2015 and all retweets of these tweets, amounting to 1M tweets and 23M retweets.

## 3 The dilution of growing audience

In this section, for each user from our expert set, we look at the properties of their followers as a function of the recency of these followers. This way, we measure how the early audiences differ from the later ones.

<sup>1</sup>A list is a set of users tagged by the creator of the list. See <https://blog.twitter.com/2009/theres-a-list-for-that> for details.

<sup>2</sup>For comparison, Followerwonk listed 824k Twitter users with at least 10,000 followers as of 5th of January 2016, see <https://moz.com/followerwonk/bio/?fmin=10000>.

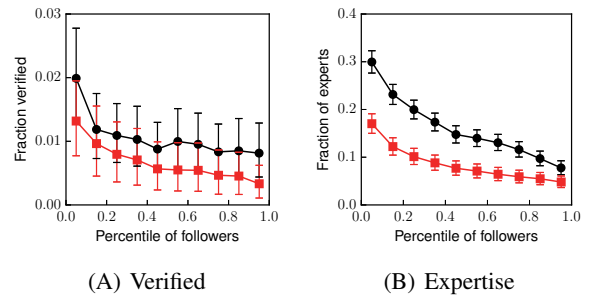


Figure 1: The eliteness of sub-groups of followers of equal size ordered by the time of link formation from the earliest to the latest for real (black circles) and reshuffled (red squares) followers. The plot is prepared for each user separately and then averaged over all users from our set of experts. The error bars are 95% confidence intervals computed on the set of followers and averaged over experts; such estimate is an upper bound of the error bars. The remaining plots are prepared in the same way.

### 3.1 The eliteness of followers

First, we quantify the eliteness of followers of our expert users. To this end, we use two different metrics of eliteness by checking whether (i) the accounts of followers are *verified* by Twitter<sup>3</sup> and (ii) whether they are collectively tagged as expert accounts (Ghosh et al. 2012). We aim to measure how the eliteness of followers changes as the number of followers grows over time. To this end, we split the followers of each expert user into ten chronological bins: the first bin contains the 10% earliest followers, whereas the last bin contains the 10% latest followers. We name these chronological bins as the *recency* of followers. Next, we average the measurements over followers in each bin, and then over our set of expert users.

We find that the earlier followers tend to be more elite than the later followers (Figure 1), both in the terms of verified and expert accounts. Later followers are around half as likely to have verified accounts and over three times less likely to be expert users (compare the leftmost and the rightmost black circles in Figure 1A and 1B).

We note here, that these results can be explained by at least two processes: either (i) the dilution of growing audiences of followers or (ii) the dilution of Twitter’s user base itself, because later followers are users who joined Twitter more recently and soon after joining started forming their follower links (Figure 2). To separate the impact of these two related processes on our results, we randomly reshuffle all follower links of expert users while fixing the age of follower users in the system. Namely, for each expert, every follower is exchanged with a random follower of another expert, under the crucial constraint that an actual follower who joined Twitter in year Y can only get replaced by a random follower who also joined Twitter in year Y. Therefore, for

<sup>3</sup>A verification badge is granted to accounts that correspond to authentic identities of individuals or brands. More at <https://support.twitter.com/articles/119135>.

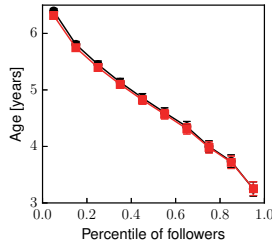


Figure 2: The account’s age for real (black circles) and reshuffled (red squares) followers versus their recency.

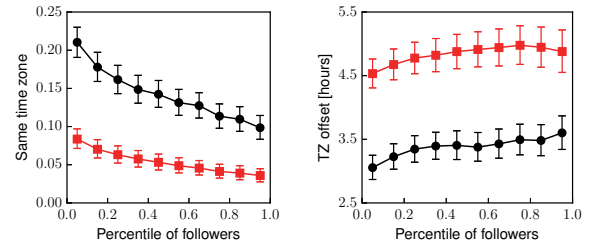
each given expert we preserve the distribution of ages of her followers, up to variations within year  $Y$ . Furthermore, we assume that these reshuffled links for each expert were created in the same chronological sequence as the real follower links that got replaced. Thus, this reshuffling purposely preserves the correlation between age and recency of followers (compare red and black lines in Figure 2), while destroying the correlations between recency and other properties that are not implied by their correlation with age. In such a reshuffled network, existence of a correlation between recency and a given property signifies that this correlation is due to the age of followers in the whole Twitter system, instead of their recency among the given set of followers.

We plot the metrics of eliteness for the reshuffled followers as red squares in Figure 1. Noticeably, the values of eliteness drop with the recency of reshuffled followers in a similar way to real followers, so the dilution of eliteness of the growing user base of Twitter is present. In other words, it is likely that “tech-savvy” experts joined Twitter early on and since that time the user base of Twitter is diluting to a more casual population. In the case of verified accounts, we do not find evidence for the dilution among followers of a user (curves in Figure 1A are nearly parallel). However, in the case of expertise, we find that the fraction of expert users drops down faster with the recency among real followers than among reshuffled followers (the two curves meet at the rightmost point in Figure 1B, but the black curve is steeper than the red curve). This finding means, that the fraction of expert users among followers drops down with the recency of these followers, providing an evidence for the dilution of expertise among growing audience at the level of followers of individual expert users.

### 3.2 Time zone, language, and topical congruence

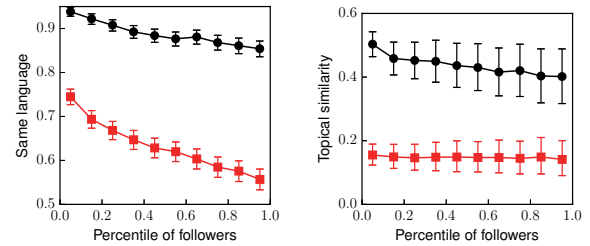
Next, we measure audience characteristics and compare them to the expert they follow. To this end, we extract the time zone and the language of each user from the self-reported fields specified by each user in their profile.

Having this information, we measure if the expert and their audience are in the same time zone (Figure 3) and if they report the same language (Figure 4). We find that the earlier followers are more likely to be in the same time zone as the person they follow (Figure 3). The average absolute difference between the time zone of the expert and her followers is 3 hours for the earliest followers and above 3.5



(A) TZ match (B) Difference in TZ offset

Figure 3: Time zone match between a user and her followers ordered from the earliest to the latest.



(A) Language (B) Topic

Figure 4: The language match and the topical similarity between a user and her followers ordered from the earliest to the latest.

for the latest followers; that is a nearly 20% increase in the difference between the time zones. An increased desynchronization of time zones may lead to a coverage bias, if the experts produce certain type of content in the evening, when some of their followers are already asleep.

Also, we note that the earlier followers are more likely than the later followers to use the same language as the person they follow (Figure 4A). This result is followed by a related finding showing that the earlier followers are more likely to share topics of expertise with the person they follow (Figure 4B). We compute the topical similarity of expertise between users as the cosine similarity of their expertise distributions (Kulshrestha et al. 2015). Note that this calculation can be done only for followers who are experts themselves, for whom expertise distributions are known.

Again, we compare these results against the results for the randomly shuffled followers. Interestingly, the dilution of language is more pronounced for the reshuffled followers. This is an intriguing result that can be explained by the fact that initially almost all tweets were in English, while nowadays there exist large non-English communities on Twitter (Liu, Kliman-Silver, and Mislove 2014). On the other hand, we note that the topical similarity between an expert and her followers is twice larger than between the user and her reshuffled followers (compare the black and red lines in Figure 4B). Most importantly, the topical alignment of Twitter’s user base does not decrease, while it visibly decreases for the growing audiences of followers (compare the slopes

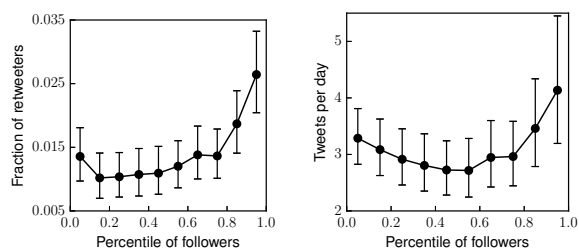


Figure 5: (A) The fraction of retweeters among followers ordered from the earliest to the latest. (B) The average number of tweets produced per day by a follower versus the sequence of followers ordered from the earliest to the latest.

of red and black lines in Figure 4B, respectively), suggesting that the dilution is present at the level of followers of individual users, but it is not at the level of whole Twitter.

### 3.3 The engagement of followers

Finally, we would like to understand how the diluted properties of a growing audience influence engagement metrics. Among such metrics we distinguish different types of interactions between a follower and the expert that they follow, for instance: retweets, favorites, replies. Here, we focus on the results for retweets.

We measure the engagement of followers with respect to the user that they follow, by measuring the fraction of retweeters among the followers of given recency (Figure 5A). This fraction initially drops by around 30% for early followers and then it grows 3-fold for late followers.

On the one hand, we expect that the fraction of retweeters among early followers should be higher, because they are more synchronized with the person that they follow in terms of time zone, language, and topics of expertise. On the other hand, the tweeting rate of the followers increases with their recency (Figure 5B), because some users get bored and cease using Twitter, while other users are spam bots. This may be the reason why the fraction of retweeters increases with the recency of the followers.

## 4 Discussion

Our results show that the composition of growing audiences tends to get diluted. First, we find that the eliteness of followers of an expert and their topical coherence decreases with their recency. Second, we show that the composition of Twitter’s user base also dilutes. For example, the eliteness and the time zone and language congruence of the users joining Twitter drops with their recency. We hypothesize that these phenomena are the effect of a common dilution process that happens for growing audiences.

Future studies may want to understand how global evolving trends are affected and interact with the rule of dilution of growing audiences. Furthermore, it is unclear what is the impact of the growing and diluted audiences for information diffusion. We show initial measurements in this respect, but

further rigorous studies of the dynamics of the latent properties of users, such as the language or topics of expertise, shall explore this question.

Finally, we note that our measurements are platform-specific. Thus, is the law of dilution of organically growing audiences present in other systems, also offline social collectives? Furthermore, what is the importance of this law with respect to recommender systems? Oftentimes, recommender systems suggest content that is topically aligned to our interests. What minimalist recommender system leads to the dilution of growing audiences? Addressing these questions will shed more light on this intriguing phenomenon.

## References

- Chen, L.; Weber, I.; and Okulicz-Kozaryn, A. 2014. U.s. religious landscape on twitter. In *SocInfo*, 544–560.
- García-Gavilanes, R.; Kaltenbrunner, A.; Sez-Trumper, D.; Baeza-Yates, R.; Aragn, P.; and Laniado, D. 2014. Who are my audiences? a study of the evolution of target audiences in microblogs. In *SocInfo*, 561–572.
- Ghosh, S.; Sharma, N.; Benevenuto, F.; Ganguly, N.; and Gummadi, K. 2012. Cognos: Crowdsourcing Search for Topic Experts in Microblogs. In *SIGIR*, 575–590.
- Ghosh, S.; Zafar, M. B.; Bhattacharya, P.; Sharma, N.; Ganguly, N.; and Gummadi, K. 2013. On sampling the wisdom of crowds. In *Proc. 22nd ACM Int. Conf. Conf. Inf. Knowl. Manag. - CIKM '13*, 1739–1744. New York, New York, USA: ACM Press.
- Grabowicz, P.; Ganguly, N.; and Gummadi, K. P. 2016. Distinguishing between Topical and Non-topical Information Diffusion Mechanisms in Social Media. *ICWSM '16: Proceedings of the 10th International AAAI Conference on Web and Social Media*.
- Kulshrestha, J.; Zafar, M. B.; Noboa, L. E.; Gummadi, K. P.; and Ghosh, S. 2015. Characterizing Information Diets of Social Media Users. In *ICWSM*.
- Leskovec, J.; Backstrom, L.; Kumar, R.; and Tomkins, A. 2008. Microscopic evolution of social networks. In *Proceeding 14th ACM SIGKDD Int. Conf. Knowl. Discov. data Min. - KDD '08*, 462–470. Las Vegas, Nevada, USA: ACM.
- Liu, Y.; Kliman-Silver, C.; and Mislove, A. 2014. The Tweets They are a-Changin: Evolution of Twitter Users and Behavior. *Proc. Eighth Int. AAAI Conf. Weblogs Soc. Media* 305–314.
- Yang, Z.; Guo, J.; Cai, K.; Tang, J.; Li, J.; Zhang, L.; and Su, Z. 2010. Understanding Retweeting Behaviors in Social Networks. *Idea* 1633–1636.