Hot Streaks on Social Media

Kiran Garimella[*]
MIT
garimell@mit.edu

Robert West
EPFL
robert.west@epfl.ch

Abstract

Measuring the impact and success of human performance is common in various disciplines, including art, science, and sports. Quantifying impact also plays a key role on social media, where impact is usually defined as the reach of a user's content as captured by metrics such as the number of views, likes, retweets, or shares. In this paper, we study entire careers of Twitter users to understand properties of impact. We show that user impact tends to have certain characteristics: First, impact is clustered in time, such that the most impactful tweets of a user appear close to each other. Second, users commonly have "hot streaks" of impact, i.e., extended periods of high-impact tweets. Third, impact tends to gradually build up before, and fall off after, a user's most impactful tweet. We attempt to explain these characteristics using various properties measured on social media, including the user's network, content, activity, and experience, and find that changes in impact are associated with significant changes in these properties. Our findings open interesting avenues for future research on virality and influence on social media.

1 Introduction

Evaluating the impact of individual performance is common practice in many realms of life, including art, science, and sports. Given the importance that society attaches to evaluations of impact—resulting in promotions, awards, and pay raises—, studying how impact evolves over the course of lives and careers is critical for making sense of human group behavior. Notably, trends in impact over the course of careers have previously been studied in order to understand the life cycle of creativity (DeNisi and Stevens 1981; Barrick and Mount 1991), as well as in the field of "science of science" (Sinatra et al. 2016).

In this paper, we study the impact of users on social media, where impact is usually defined in terms of the reach of a user's content by metrics such as the number of retweets, likes, views, shares, or comments. More specifically, we focus on *individual impact*—which we define as the reach of an individual user's content over the course of their "career" (the time since they joined the platform). Particularly, we are

---

motivated by recent studies on impact by Sinatra et al. (2016), who show that scientists' most impactful works are randomly distributed over their careers, and by Liu et al. (2018), who show the existence of periods of increased creativity in scientific, artistic, and cultural careers. These works lead us to ask the question whether there are similar trends in impact on social media.

Impact is, however, difficult to study, and particularly so in social media, due to the subtle interplay between individual and social aspects of impact. On the one hand, impact is inherently social: the impact created by a person's content depends on how others in their network perceive the content. On the other hand, impact is also individual, via factors such activity (number of posts written), experience (age on the platform), and interests (content of posts). Due to the challenging nature of the problem, research into long-term trends of individual impact on social media, and the factors that influence these trends, has been limited to date. Most research has investigated the impact of *isolated pieces of content,* e.g., by predicting retweet counts on Twitter (Suh et al. 2010; Martin et al. 2016) or exploring factors that lead to content being shared (Gao, Ma, and Chen 2015). On the contrary, the literature is much thinner regarding the nature and role of the impact of users over the course of their *entire careers.*

To make progress in this direction, we built four datasets from Twitter, which span over a decade and contain complete careers of users with a wide range of followers, experience, and topical interests. We choose retweet counts as the measure of impact and reconstruct the careers of users—from their very first post until today—as a time series of retweet counts obtained for their tweets. We aim to discover characteristic patterns in user impact over time and to understand the role played by factors such as the user's network, the content they post, and their levels of activity.

We start by looking at the position of the most impactful tweets during a user's career and find that, for a large fraction of users, the most impactful tweets are clustered together in time (Section 3).

Moreover, this trend is not just limited to a few most successful tweets in a user's career, but generalizes to entire periods of high impact ("hot streaks") during which a user's posts are retweeted significantly more. We analyze various

factors behind this phenomenon, including tweet content, changes in follower counts, retweeting behavior, and activity of the user. We find that hot streaks are driven by new retweeters who suddenly start following and retweeting a user, thus leading to a sudden gain in followers during a hot streak. Even though these new retweeters retweet highly, their interest is quite limited in time, and they tend to quickly decrease their retweeting activity. Users also become more active during hot streaks, tweeting longer content containing media and a more diverse range of topics (Section 4).

Finally, we zoom into the activity around the most retweeted tweets for a user and show that there is a pattern of build-up before, and drop-off after, the most retweeted tweet (Section 5).

Our study is one of the first to show the presence of such patterns of individual impact on social media. These findings could help us in revisiting existing problems in social network analysis such as understanding virality or influence maximization from an individual perspective. For instance, the clustering of high-impact tweets and the presence of hot streaks shows when a user is the most impactful in her career. This could be useful in finding influential users who are currently on a hot streak to maximize the spread of a marketing campaign. Another example could be to use the presence of the build-up towards a user's most retweeted tweet to explain how tweets become viral for certain users.

2 Datasets

2.1 Dataset description

We use four large datasets with varying characteristics from Twitter. For each user in our dataset, we obtain all their tweets and construct their *career* as the time series of retweet counts for original tweets posted by the user (retweets by the user are not included in the career). Each post is associated with a timestamp when it was posted and a measure of impact (retweet count), as measured when the data was crawled (June 2018). Below, we briefly describe how each dataset was sampled.

VERIFIED. Starting from a list of all verified users on Twitter from June 2018,[1] we filter out users according to various criteria: they should have between 2,000 and 3,200 tweets, and between 50,000 and 2 million followers.[2]

POLITICAL. This dataset contains over 600,000 politically active users spanning a period of almost 10 years (2009–2018). The dataset has been used in previous work (Garimella and Weber 2017).

DESPOINA (RANDOM) and DESPOINA. We randomly sampled two sets of 1 million users with fewer than 3,200 tweets

---

[1] 297,000 users, from http://redd.it/8s6nqz

[2] The Twitter API gives access to no more than the 3,200 most recent tweets for each user. Since we want to make sure that we capture the complete career of a user, we restrict our analysis to users with at most 3,200 tweets. We require at least 2,000 tweets per user because we want to study users with a certain minimum level of activity. Finally, we remove users with over 2 million followers due to the strict rate limiting constraints on the Twitter API endpoint for getting followers.

Table 1: Statistics of our datasets: Number of users, median number of followers, tweets, median of mean retweets and total number of tweets.

Dataset	#users	med. #foll	median #tweets	median #retweets	total #tweets
VERIFIED	2,710	105,007	2,595.5	37.0	5.3M
POLITICAL	10,563	360	2,353	1.52	22.3M
DESPOINA	76,777	606	2,566	2.06	130M
DESPOINA (RANDOM)	5,465	515	1,926	0.48	7.9M

from a large sample of 93 million Twitter users collected by Antonakaki, Ioannidis, and Fragopoulou (2018). For DESPOINA (RANDOM), we sampled completely at random, whereas for DESPOINA we first constrained on certain criteria to ensure high-quality data. In particular, we only selected users who (i) have between 2,000 and 3,200 tweets, (ii) created their account at least 6 months before April 2016 (time of the crawl by Antonakaki, Ioannidis, and Fragopoulou (2018)), and (iii) have at least 100 followers. This gave us 6 million users. From this, we randomly sampled 1 million users.

Finally, all datasets were further filtered to retain only those users whose most retweeted tweet received at least 50 retweets. This was done to ensure that our notion of the most retweeted tweet is meaningful. For the users who remained after applying these filters, we obtained all tweets, followers, and retweeters of all tweets using the Twitter REST API.

Table 1 shows a high-level summary of the datasets, including the number of users, median number of followers, number of tweets, and median of mean retweet count per user (we first compute mean retweet count per user and then take the median over all users). The datasets were selected in such a way that they span a wide range of characteristics in terms of content (e.g., politics), user behavior (celebrities vs. normal users), number of followers, and impact. The POLITICAL dataset contains tweets and users related to a specific topic (politics) and might have different behavior due to the 2016 U.S. elections and increased political activity; the DESPOINA and DESPOINA (RANDOM) datasets capture a random sample of "normal" users on Twitter; and the VERIFIED dataset captures "celebrity" behavior on Twitter, as evident from the large median follower count in Table 1.

In the rest of the paper, due to space constraints, whenever the trends are similar on all the datasets, we only show results for the VERIFIED and DESPOINA datasets.

Code and datasets used in this paper are publicly available at https://github.com/gvrkiran/hot-streaks-social-media.

2.2 Historical follower-count estimation

To study the dependence of the follower network on impact, we need historical information on the number of followers a user has at any point in history. However, the official Twitter API does not provide this information. We therefore used the Internet Archive Wayback Machine,[3] which archives web pages to obtain historical snapshots of Twitter profiles of

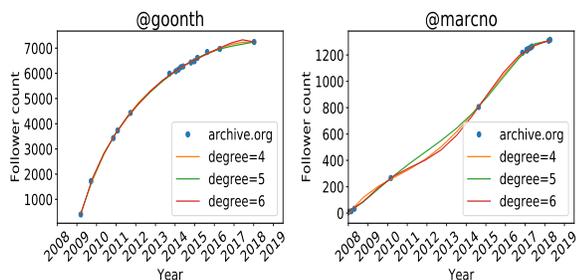---

[3] https://archive.org/web/web.php

Figure 1: Examples of two random userss number of followers over time, obtained from the Internet Archive Wayback Machine (blue dots). The orange, green, and red lines are polynomial fits of degree 4, 5, and 6, respectively.

users.[4] We downloaded and parsed the raw HTML from the Internet Archive to obtain follower counts of users at different time stamps.

Not all users have snapshots at all points in time. We therefore restricted the data to only those users for whom we have at least 10 snapshots, and we fit polynomial curves to estimate the number of followers for each user at any given point in time. For each user, we fit polynomials of degrees 4, 5, and 6 and picked the fit with the best R^2. We only consider users for whom the best fit attained $R^2 > 0.9$. Examples of such fits for two random users from our dataset are shown in Figure 1. We see that the fits capture follower curves of different shapes and provide accurate interpolations for follower counts. Now, for each user, using the best polynomial fit, we can estimate the follower count at points in time for which we do not have data from the Internet Archive. For instance, for user @marcno in Figure 1, we can estimate the number of followers in 2012 as 420. Using this technique, we obtained the imputed historical number of followers for each user in our dataset. All the users included in Table 1 have at least 10 snapshots on the Internet Archive and attain a good fit.

3 Clustered impact

We start with some notation. All our analysis will be at a user level, and we aggregate over all users. Assume a user u has N tweets, $t_1, \ldots, t_N$. We call the life time of a social media user, from their first tweet to their last tweet (at data collection time, June 2018), the *career* of that user. The position of a tweet t_i in a user's career is measured in terms of (i) time in weeks since the first tweet by the user, $w(t_i)$ and (ii) tweet index i, given by $i = P(t_i)$. Though time and tweet index might be correlated, there are users with long periods of inactivity, for whom using absolute weeks to measure position would be misleading. Hence, we perform all our experiments using both measures. Unless explicitly specified, we use tweet index to denote the position of a tweet. The k most retweeted tweets in a user's career are denoted by

$T_1, T_2, T_3, \ldots, T_k$.

We begin by investigating the timing of the five most retweeted tweets in a user's career, T_i for $i \in [1, 5]$. Scatter plots of $P(T_1)$ vs. $P(T_i)$ for $i \in [2, 5]$ are shown in Figure 2. We observe that the positions of T_1 and T_i are highly correlated. We also computed all pairwise correlations between $P(T_i)$ and $P(T_{i+1})$, and between $w(T_i)$ and $w(T_{i+1})$ (for $i \in [1, 4]$), and obtained similar results. To check the robustness of these results, we shuffled the order of the tweets in each career, and the correlation drops from around 0.6 to below 0.10 for all datasets, indicating that this phenomenon holds strongly for real careers.

A user's retweet counts grow over their career, in part due to the growing number of followers a user attracts over her career (Suh et al. 2010; Jenders, Kasneci, and Naumann 2013). Hence, the likelihood of a user having their most retweeted tweets later in their career is higher *a priori*. To mitigate such effects, we normalize retweet counts at a point in time by the number of estimated followers at that time (cf. Section 2.2), computing retweets per follower for every tweet in a user's career. We denote the k tweets with the most retweets per follower by $T_1^*, T_2^*, T_3^*, \ldots, T_k^*$. In order to check how well the clustering in Figure 2 goes beyond the the top 5 most retweeted tweets, we computed the correlation between the position of the tweet with the most retweets per follower, $P(T_1^*)$, with $P(T_2^*)$, $\ldots$, $P(T_{200}^*)$. The results for the VERIFIED and DESPOINA datasets are shown in Figure 3, where each point represents the Pearson correlation between T_1^* and T_i^* for $i \in [2, 200]$. We see that the correlation remains above 0.5 for up to 25 tweets. In Figure 3, we also show the correlation for shuffled careers (green and black lines) and we can see that real careers have a much higher correlation than shuffled careers.

Next, we look at the difference in the position of the top two most retweeted tweets and compute the normalized difference $(P(T_1) - P(T_2))/N$ for each user. Figure 4 shows the distribution of the difference. We see that the distribution peaks and is centered at zero, providing further evidence that T_1 and T_2 are close to each other. The fact that the distribution is symmetric around zero indicates that $P(T_1)$ can be before or after $P(T_2)$. We compared this with a distribution of the difference for shuffled careers (orange bars in Figure 4) and observe that the strong peak close to zero disappears, as expected. Similar patterns hold for the other datasets and when using $(P(T_i) - P(T_{i+1}))/N$, for $i \in [2, 4]$, instead of the above-defined difference.

4 Existence of hot streaks

In this section, we try to generalize the finding from the previous section to see if such highly retweeted tweets occur within extended periods of higher-than-usual-impact tweets. We call such periods "hot streaks" following its usage in recent work (Liu et al. 2018).

4.1 Identifying hot streaks

We propose a simple method to identify hot streaks. We represent the retweets a user gets throughout her career as a time series and identify periods of consistently high activity in
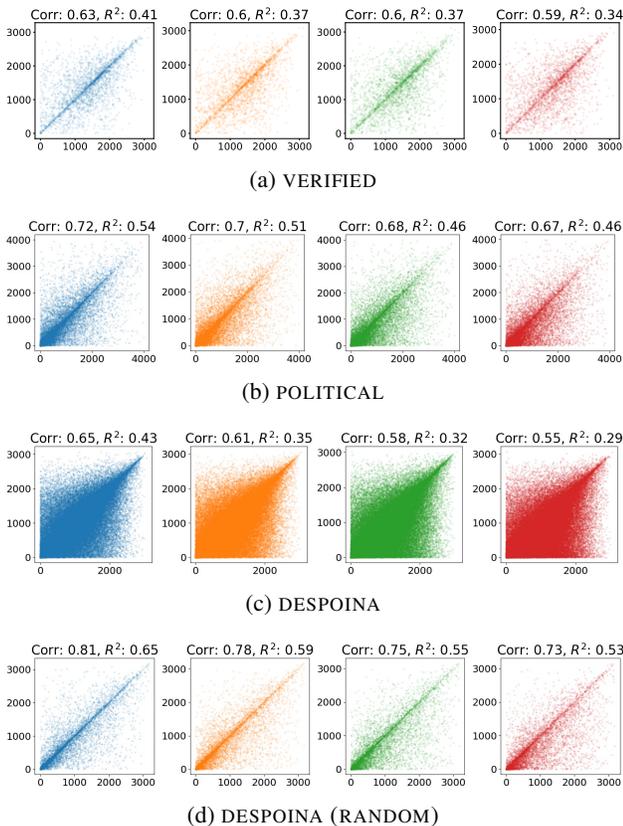
---

[4]e.g., Barack Obama's Twitter profile has over 7,000 snapshots spanning over 10 years. A snapshot of Barack Obama's Twitter page in 2014 (https://bit.ly/2ssJrkP) shows that Obama had 46.6M followers at that time.

(a) VERIFIED

Corr: 0.63, R^2: 0.41 | Corr: 0.6, R^2: 0.37 | Corr: 0.6, R^2: 0.37 | Corr: 0.59, R^2: 0.34

(b) POLITICAL

Corr: 0.72, R^2: 0.54 | Corr: 0.7, R^2: 0.51 | Corr: 0.68, R^2: 0.46 | Corr: 0.67, R^2: 0.46

(c) DESPOINA

Corr: 0.65, R^2: 0.43 | Corr: 0.61, R^2: 0.35 | Corr: 0.58, R^2: 0.32 | Corr: 0.55, R^2: 0.29

(d) DESPOINA (RANDOM)

Corr: 0.81, R^2: 0.65 | Corr: 0.78, R^2: 0.59 | Corr: 0.75, R^2: 0.55 | Corr: 0.73, R^2: 0.53

Figure 2: Scatter plots showing the correlation of the tweet position of the five most retweeted tweets in a user's career. Each row shows the correlation between the position of the most retweeted tweet ($P(T_1)$) and that of the i-th most retweeted tweet ($P(T_i)$), for $i \in [2,5]$ (left to right). Each dot represents a user. We see that most users lie close to the diagonal, indicating that the most retweeted tweets occur closely in time. "Corr." denotes Pearson's correlation coefficient. Note that in VERIFIED and DESPOINA, all users have between 2,000 and 3,200 tweets.

this time series. We identify such periods in a user's career by solving an optimization problem. The objective is to minimize the least-squares fit of a piecewise-constant function to the retweet counts time series and observe if there are extended periods with a high constant. To avoid overfitting, we add a regularization term on the number of constant functions used.

Specifically, the problem can be defined as such: given a user u whose career consists of N tweets with retweet counts $r_1, \ldots, r_N$, find N constants $c_1, \ldots, c_N$ that

$$\text{minimize} \quad \sum_{i=1}^{N}(c_i - r_i)^2 + \alpha \sum_{i=1}^{N-1} \delta(c_i, c_{i+1}), \qquad (1)$$

where $\delta(a,b) = 0$ if $a = b$, and $\delta(a,b) = 1$ otherwise.

The objective function in Equation 1 tries to balance between too many constants (first term) and too few (second term). The penalty parameter α controls the number of con-
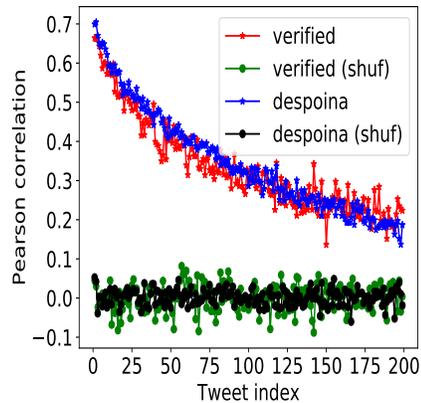

Figure 3: Pearson correlation for T_1^* and T_i^* for $i \in [2, 200]$.
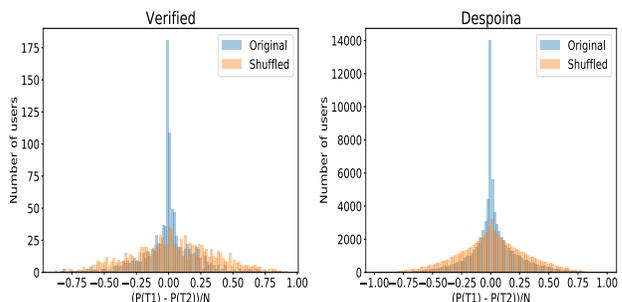

Figure 4: Normalized difference between the position of the first and second most retweeted tweets.

stants. There are exponentially many functions (sets of constants) that would need to be evaluated when solving Equation 1 by brute force. We can, however, also solve it efficiently using a dynamic programming solution with time complexity $O(n^2)$ (Wang and Song 2011). Examples of two users' careers with fits obtained using Equation 1 are shown in Figure 5. The blue line shows the raw retweet-count time series; the black lines show the piecewise-constant fit.

Given these piecewise-constant fits, we define a period as a hot streak if its constants lie consistently above a certain threshold, say, the k-th percentile of retweets in a users career. We also experimented with other methods,[5] but found that these algorithms were not robust enough to handle the spiky nature of retweets. For instance, we see in Figure 5 (right) that there are many tweets during the hot streak (marked in red) below the threshold (dashed green line). Using these simple methods fails to identify such cases. Note that the exact method used to identify hot streaks is orthogonal to the analysis in the rest of the paper. Hence, any other method can also be used.

Parameters. We have two parameters in the optimization

---

[5]We tried (i) identifying contiguous portions of a career above a certain threshold, (ii) using time-series change-point detection (Liu et al. 2013; Ross 2015), and (3) Fourier analysis (Moskvina and Zhigljavsky 2003).
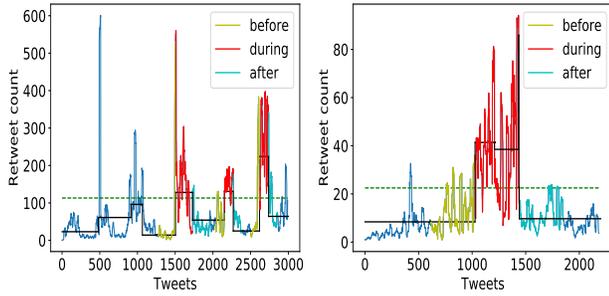
Figure 5: Example careers of two users. The blue line shows the retweet counts over the career of a user. The black line shows the piecewise-constant fit obtained using Equation 1. The dashed green horizontal line shows the 90th percentile. The periods of increased activity indicating hot streaks can be seen in red. Periods before and after the hot streak are labeled in yellow and cyan, respectively.

Figure 6: Distribution of number of hot streaks.

Figure 7: Distribution of length of hot streaks compared to shuffled careers.

problem for identifying hot streaks: (i) α in Equation 1: we tested for multiple values of α, from 1 to 10 and fixed the value of 1 because it gave the best fits (observed via manual inspection); (ii) the percentile threshold k to identify a hot streak: we tested a range of values from 70 to 95 in increments of 5 and fixed the value of $k = 90$ (i.e., 90th percentile); higher/lower values lead to fewer/more hot streaks.

4.2 Characteristics of hot streaks

The real value of identifying hot streaks is in revealing several fundamental patterns that govern individual careers. In this section, we look at various characteristics of hot streaks. Similar to the previous section, for every analysis we perform, we compared the results with shuffled careers, thus ensuring the robustness of our results.

Number of hot streaks. We first look at the fraction of users for whom a hot streak greater in length than 10 tweets exists. We find that 86% of users in VERIFIED, 59% in POLITICAL, 63% in DESPOINA, and 65% in DESPOINA (RANDOM) have at least one such hot streak. We compared these with the fraction of users having at least one hot streak for randomly shuffled careers, and the fractions drop to less than 30%. Given that each of these datasets was sampled in a different way and that a nontrivial fraction of users have hot streaks, we can conclude that hot streaks are a prevalent phenomenon. Figure 6, shows the distribution of the number of hot streaks per use. We see that a majority of users have fewer than 4 hot streaks, with some users having more than 10.

Length of hot streaks. Next, we compute the distribution of the length of the longest hot streak for all users. We can see from Figure 7 that for a majority of users, hot streaks are mostly short in length—up to 20 tweets—, but some users have much longer hot streaks of over 100 tweets. We compare the lengths of hot streaks to shuffled careers and see in Figure 7 that real careers are characterized by a longer tail indicating that real users have longer streaks of high retweet activity. We tested the robustness of these results by
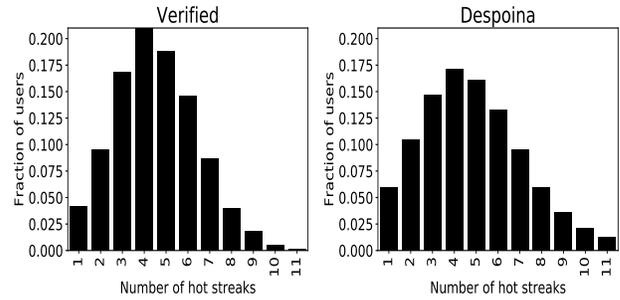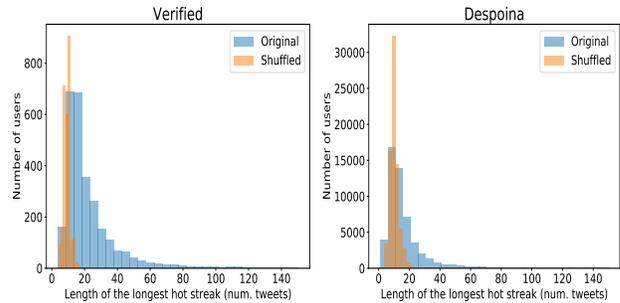
controlling for individual career length and arrived at the same conclusions.

Position of hot streaks. What is the position of hot streaks in a user's career? Are they randomly distributed, as found by Liu et al. (2018) in the context of art, science, and movies? Defining the position of a hot streak as the position of its first tweet (within the respective user's career), we find that the position of hot streaks in a user's career differs depending on how long the user has been on Twitter (which we refer to as the "age on Twitter"). We divided users into 4 categories based on their age on Twitter: 100–200, 200–300, 300–400, and 400–500 weeks.[6] Figure 8 shows the distribution of the position of hot streaks in careers. We see that it is positioned at different places depending on the user's age on Twitter. For older users, it is more likely to be at the end of their career. As users spend more time on the platform, they learn more, gain more followers, and hence their chance of having a hot streak increases. For younger users, the position is more evenly distributed, indicating other factors at play (content, activity, etc.).

User properties. Who are the users who have hot streaks? How are they different from the others who do not have hot streaks? To check if users who have hot streaks differ in any

---

[6](i) Since Twitter introduced the retweet button only in 2009, we only consider users after that. Users with a career over 500 weeks would have their first tweet prior to 2009 and hence were not considered. (ii) The results remain similar for different groupings, e.g., when bucketing into groups spanning 50, rather than 100, weeks.
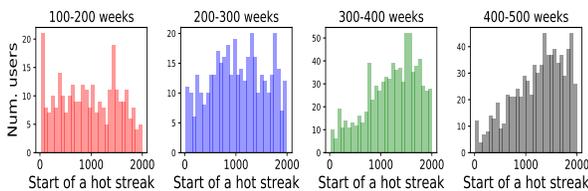
Figure 8: Distribution of the position of hot streaks in user careers (VERIFIED dataset), for various user ages on Twitter (one plot per age). Only the first hot streak of each user is considered.

way from those who do not, we compared the two user groups in terms of profile features such as the number of followers, friends, tweets, and age on Twitter. We find that users with hot streaks have a significantly higher number of followers, but that there is no difference in the other feature values.

Influence. How influential are hot streaks in defining success in a user's career? To answer this question, we compute the fraction of total retweets obtained during all hot streaks in a user's career. Figure 9 shows the distribution of the retweet count fraction for two datasets. We see that for a majority of users (above 80%), only 10–20% of retweets are obtained during a hot streak, indicating that hot streaks are not the main driver of retweet activity in a user's career.
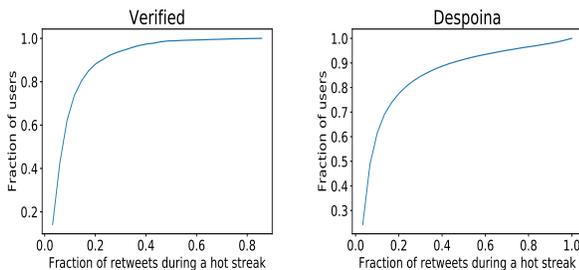

Figure 9: CDF of the fraction of retweets obtained during a hot streak.

Finally, we investigate whether hot streaks are connected to a recently introduced feature on Twitter— threads[7]—by looking at hot streaks for a random sample of 50 users manually. In only one of the 50 cases did the hot streak occur in a thread, indicating that hot streaks are typically not threaded tweets.

4.3 What happens during a hot streak?

In the analysis below, for each user, we only consider the longest hot streak and obtain tweets from periods of equal length (in terms of number of tweets) before and after the hot streak. That is, if a user has a hot streak of length k tweets, we consider a period of k tweets before and after the hot streak (examples are shown in Figure 5, colored in yellow and cyan). All the results reported are statistically significant

($p < 0.01$) after applying Bonferroni correction for multiple tests, wherever applicable. All the results presented below are consistent across datasets, but we focus on the VERIFIED dataset due to space constraints.

Retweet count. In Figure 10(a), we see that the mean retweet count increases significantly[8] during a hot streak when compared to the period before and after. This is expected by the definition of hot streaks. However, we also observe that the retweet counts after the hot streak are significantly higher than before, indicating some spillover effects.

Follower gain. We measure the growth in number of followers before, during, and after a hot streak. Figure 10(b) shows the follower gain. We observe that the follower gain during a hot streak is significantly higher than before or after the hot streak. This might be because more retweets lead to more followers (Myers and Leskovec 2014). Similar to above, we also observe that the follower gain after a hot streak is significantly higher than the follower gain before.

Retweets per follower. To understand if the increased retweet count after the hot streak is due to the increased number of followers, we look at the retweets per follower by dividing the average number of retweets before, during and after by the average number of followers during the same period. We can see from Figure 10(c) that the increase after the hot streak vanishes in this case. Retweets per follower before and after the hot streak are significantly less frequent than during the hot streak, but there is no significant difference between before and after.

Activity. We look at the activity (all the tweets by the user) before, during, and after a hot streak. Note that this also includes retweets (unlike in Section 4.1), because we want to know how a user's overall tweeting behavior is associated with the impact created by their own tweets. To measure this, we compute the number of tweets per hour during, before and after a hot streak. We find that, during hot streaks, the activity is higher (Figure 10(d)).

4.4 Factors affecting hot streaks

In the previous section, we saw that hot streaks are characterized by a higher retweet count, an increased follower gain, and an increase in activity. In this section, we try to dig deeper into each of those factors and try to understand the role they play in the hot streak phenomenon. We particularly look at three factors: network (behavior of followers and retweeters), content (tweets posted during a hot streak), and activity (number and type of tweets). We want to measure how much of the hot streak characteristics we observed in the previous section are accounted for by these three factors.

Network. In the previous section, we found that during a hot streak, users gain many followers, indicating changes to the network. To further analyze the role of the network during a hot streak, we look at individual retweeters who retweeted a user before, during, and after the hot streak.

We fix the set of retweeters who retweeted the user before a hot streak and tracked their activity during and after the hot streak. Surprisingly, we find that these retweeters retweet
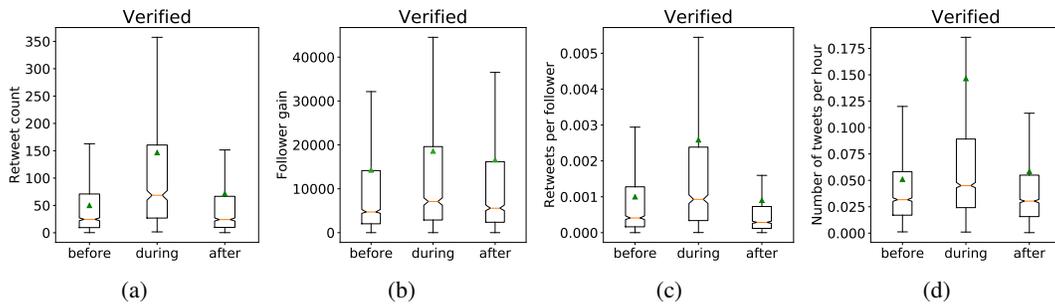

Figure 10: Notched box plots showing the distributions before, during, and after a hot streak. The orange line shows the median, green triangle the mean. The notches indicate 95% confidence intervals around the median obtained via bootstrapping. (a) Retweet count, (b) follower gain, (c) retweet count per follower, (d) number of tweets per hour.

the user significantly *less* during a hot streak, compared to before. Figure 11 shows a box plot of retweeter behavior. We also analyze the retweet count by only retweeters who started retweeting during the hot streak and compared that to retweeters who started before or after the hot streak. We find that retweeters who start during the hot streak have a significantly higher retweet count (Figure 12). We can also see from Figure 12 that the retweeters who started retweeting during a hot streak lose interest more quickly than those who started retweeting before or after. These findings indicate that hot streaks are driven mostly by new retweeters who start retweeting the user suddenly, but lose interest quickly.
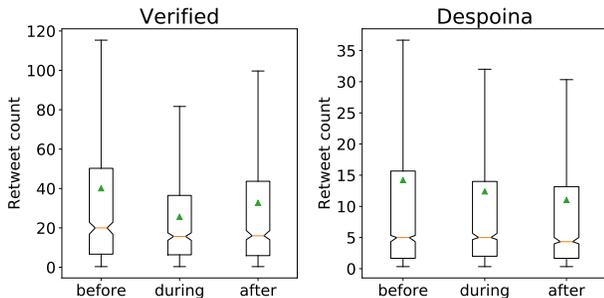

Figure 11: Retweet counts when only considering retweeters who retweeted before a hot streak.

Content and activity. Above, we looked at *who* is driving the increase in retweet count during a hot streak. Next, we dig deeper into content-related factors, looking at *what* drives such increase in impact. To understand this, we looked at various aspects in terms of content posted and types of activities by users during a hot streak and compared it with periods before and after the hot streak. Specifically, we looked at attributes from tweet patterns to measure activity features such as these: fraction of tweets that are retweets/replies/mentions; fraction of tweets containing media/URLs/hashtags; tweet length; and content related features such as entropy of the tweet in terms of the word distribution, entropy of the topic distribution, and average sentiment of tweets. Topics were extracted using latent Dirchlet allocation (LDA) as implemented in Python's scikit-learn package. Sentiment was ex-
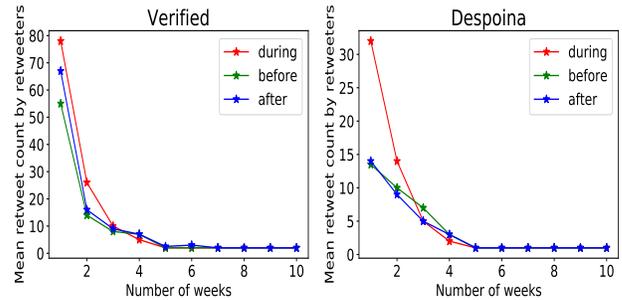

Figure 12: Retweet counts by retweeters starting before, during, and after hot streaks. The *x*-axis shows the number of weeks since the beginning of the period before, during, or after the respective hot streak. Retweeters starting during a hot streak have a higher retweet count early on, but it drops off quickly.

tracted using SentiStrength (Thelwall 2017).

Table 2 shows the results comparing the period during the hot streak to periods of the same number of tweets before and after the hot streak. Each column compares two values: before vs. during, during vs. after, and before vs. after. A cross (✗) indicates no statistically significant differences ($p < 0.01$). A ">" ("<") indicates that the element on the left (right) in a column is significantly greater than the element on the right (left) for all the datasets. For instance, the ">" in Table 2, the first row, second column ("retweet, during vs. after") indicates that the fraction of tweets that are retweets is significantly higher during the hot streak when compared to the period after the hot streak. Significance is tested using a *t*-test of means, and a Bonferroni correction is applied to the *p*-values. A value is marked ">" or "<" only if it holds across all four datasets. We can observe the following results from Table 2. During a hot streak, users have (i) a higher fraction of tweets that are retweets or contain media, (ii) a lower fraction of tweets that are replies/mentions, (iii) longer tweets, with higher entropy both in terms of words and topics, and (iv) no change in sentiment.

Content and activity vs. network. We saw that content, activity, and network features become significantly different

Table 2: Content and activity features before, during, and after a hot streak. Rows: fraction of tweets that were retweets/replies/mentions; fraction of containing hashtags/URLs/media; tweet length; entropy of word distribution, entropy of topic distribution; average sentiment. Each column compares pairs of values before/during/after a hot streak. A cross (✗) indicates no statistically significant differences ($p < 0.01$). A ">" ("<") indicates that the element on the left (right) in a column is significantly greater than the element on the right (left) for all the datasets.

	before vs. during	during vs. after	before vs. after
retweet	<	>	✗
reply	>	<	✗
mention	>	<	✗
hashtag	✗	✗	✗
URLs	✗	<	<
media	<	>	✗
tweet length	<	>	✗
entropy tweets	<	>	✗
entropy topics	<	>	✗
sentiment	✗	✗	✗

during a hot streak. To understand the role of each of these factors, we look for users for whom there was no significant change in the network, i.e., users for whom there was no significant increase in follower count during a hot streak, and investigate their content and activity. The following trends from Table 2 still persist for this subset of users: (i) lower fraction of replies/mentions, (ii) higher fraction of media, and (iii) higher entropy of topics. The effects we find here show that irrespective of changes in the network, users exhibit a strong change in how they act and how they produce content during a hot streak.

Predicting hot streaks. We now build a simple binary classifier to predict for a user whether a week belongs to a hot streak or not, based on the three sets of features described above: content, activity, and network. Concretely, we use the following features: (a) content and activity features: fraction of tweets that are retweets/replies/mentions, fraction of tweets containing hashtags/URLs/media, tweet length, entropy of the word distribution, entropy of the topic distribution (obtained using LDA), sentiment of the tweets; (b) network features: number of followers gained, average activity of the retweeters, number of retweeters tweeting for the first time, number of existing retweeters active during that week.

Table 3 shows the accuracy of a random forest classifier for different sets of features on a class-balanced dataset. The numbers in brackets show standard deviation over a 10-fold cross-validation. We also evaluate the classifier using just content/activity and just network features. We see that both network and content/activity features individually do better than a random prediction (accuracy of 50%). However, their combination improves the accuracy significantly. We computed the features that play a role in the prediction using the decrease in impurity in the random forest classifier. Entropy of topics, number of followers, and number of tweets come up as the most important features in all the datasets. It is interesting to see that the most important features span all three categories of features—content, network, and activity. Our

Table 3: Accuracy predicting whether a particular week is in a hot streak or not. "Content" and "Network" refer to just content/activity- and network-related features in isolation. "Combined" indicates all features combined.

	Content	Network	Combined
VERIFIED	0.67 (0.03)	0.62 (0.02)	0.72 (0.04)
POLITICAL	0.72 (0.01)	0.63 (0.01)	0.78 (0.01)
DESPOINA	0.72 (0.0)	0.63 (0.0)	0.77 (0.0)
DESPOINA (RANDOM)	0.70 (0.02)	0.63 (0.01)	0.75 (0.01)

main goal with the prediction task was to determine which factors are most associated with hot streaks, by looking at the features that would best predict hot streaks. However, the answer is not clear. A mix of all the three feature sets is involved, along with some external noise which we cannot account for (constituting the remaining 30% in accuracy in our model), pointing to the limits of predicting real-world phenomena (Martin et al. 2016).

5 Impact around the most retweeted tweet

In the previous section, we looked at trends in impact over the entire career of a user. In this section, we zoom into the career and understand the dynamics of retweet activity surrounding the *most impactful* tweets of a user.

Looking at the period of 10 tweets before and after the most retweeted tweet of a user, we found that the most retweeted tweet does not occur in isolation, but that there is a build-up to it. The top row in Figure 13 shows the mean retweet count across all users for the five most retweeted tweets ($T_1, \ldots, T_5$) and the 10 tweets preceding/succeeding them. The bottom row just shows the 10 tweets before and after the most retweeted tweet (T_1), excluding T_1. We can clearly see that there is a build-up towards, and drop-off after, the most retweeted tweet. This pattern holds for various stratifications of the users, along experience, retweet counts, and datasets. Comparing this pattern with 10 tweets before and after the most retweeted tweet in a shuffled career, we see from Figure 13 (bottom row, black line) that the trend disappears.

To understand if this is a prevalent phenomenon across users, we perform linear fits for the retweet counts of 10 tweets before and after the most retweeted tweet T_1 (excluding T_1 itself) and plot the average slope of these lines. Figure 14 shows the average slope of the increase (in red) and decrease (in black). We also show the individual fits for all users (in light shades). We can see from Figure 14 that for most users (over 60%) there is an average positive (negative) trend before (after) T_1.

We further look at the role of content in this phenomenon, by computing the entropy of the topic distribution of the 10 tweets before/after T_1. Figure 15 shows that the entropy of the topic distribution decreases specifically at T_1. To understand if this is just an artifact of tweets being shorter, we also plotted the average length of the tweets before and after T_1. We find that even though the tweets are longer at T_1, the entropy of topics is lower. We looked at whether these tweets (10 before/after) are part of a thread on Twitter, but could not
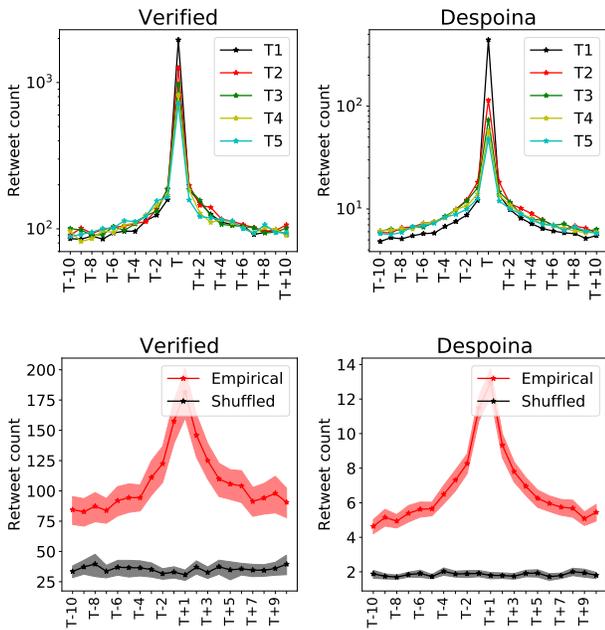
Figure 13: Top row: Mean retweet counts of 10 tweets before $(T-10,\ldots,T-1)$ and after $(T+1,\ldots,T+10)$ each of the top 5 most retweeted tweets (T) per user. Bottom row: The 10 tweets before and after, without the focal tweet (T) itself, compared to a shuffled timeline (in black), with 95% confidence intervals.

Figure 14: Linear fits for retweet counts before and after T_1 (the user's most retweeted tweet), for two datasets. We see that the average slope (dark) is positive (negative) before (after) T_1, indicating a clear build-up (drop-off) around T_1.

Figure 15: Entropy of the topic distributions before, during, and after T_1 (the user's most retweeted tweet) along with 95% confidence intervals (black line). The green line shows the average length of the tweet (on the y-axis on the right).

find any evidence for that. Future research should analyze the content of most retweeted tweets more closely in order to explain the marked decrease in entropy.

6 Related work

Hot hand. The "hot hand" hypothesis originates from basketball, where a shooter is allegedly more likely to score if their previous attempts were successful, while having "hot hands". Gilovich, Vallone, and Tversky (1985) were the first to show that there is no evidence of hot hands in basketball, arguing that short random sequences occur by chance and are not representative of their generating process. However, these claims have been refuted in numerous studies later, pointing to flaws in the statistical analysis in the original study and showing the existence of hot hands on different datasets (Miller and Sanjurjo 2018). Similar phenomena exist in other domains, too, such as gambling (Ayton and Fischer 2004). See the supplementary material of Liu et al. (2018) for a thorough review of studies on hot hands and the proposed potential reasons for this phenomenon. Perhaps most relevant to our work is recent research by Liu et al. (2018), who show that in many disciplines such as science, art, and movies, careers exhibit hot streaks of increased creative and impactful output.

Our work is different from the above studies in many ways. Firstly, we study entire careers of social media users looking at impact, rather than activity. Our contribution is a combination of identifying trends in impact of a user in her career and showing for the first time the presence of the hot hand
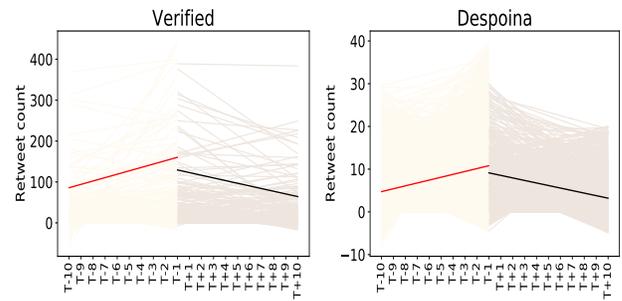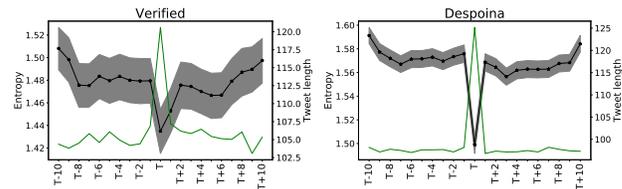
phenomenon on social media. Secondly, an important aspect that is missing from all the above analysis is social influence, where the performance/impact of a person depends not only on her personal abilities, but also on who is judging them. Our analysis on social media impact takes that into account. We try to untangle the relationship between content, activity, and network and show that each of these factors plays their own role in hot streaks. Thirdly, social media is inherently different than domains such as art and science. For instance, activity on social media, unlike in science, sports, arts, or any other skill-based occupation, has a much lower barrier. Being productive in terms of the number of tweets produced is easier than producing scientific papers or art work. This might explain why some of our results are different from Liu et al. (2018).

Bursts of activity on social media. Sudden bursts of activity on Twitter have been studied previously. A sudden surge in the number of tweets/retweets were used to identify events (Chierichetti et al. 2014). Zhao et al. (2015) looked at the distribution of such bursts for retweets from a tweet. They conclude that the interest in the tweet follows a power-law decay. Myers and Leskovec (2014) show the impact of a burst of retweets on follower growth. They study how the follower network changes following a burst of retweet activity. Barabasi (2005) proposes that human activity in general is characterized by burstiness, with periods of high activity followed by rest periods. However, the grounds for such phenomena are not yet clearly understood. Gandica et al.

(2016) show the existence of such burstiness in the behavior of editors on Wikipedia. Goel et al. (2015) study virality on social media and show that modeling what goes viral and how is hard to predict due to the lack of proper structure in the nature of cascades. Our work adds to all these fields. First, we show that there are bursty periods of impact in a user's career on social media, where, unlike Wikipedia (Gandica et al. 2016), the network factor plays an important role. We also try to untangle content and network in this bursty behavior (Myers and Leskovec 2014). Second, most of the above work analyzes single pieces of content (tweets) and the factors that influence their impact. In this paper, we go beyond a single tweet and look at the entire career of a user. The difference is that individual content can have different properties, like becoming viral or having a higher quality, but on average these effects might cancel out when looking at a lot of posts for a single user, hence giving rise to new patterns. Finally, our results show that there is a pattern in the way individual impact is created. These findings could be useful in understanding collective behavior such as event detection, measuring influence, or virality.

7 Discussion

In this work, we analyze individual impact on Twitter using datasets consisting of careers of a large set of users sampled to represent a wide range of properties. We present new insights on the prevalence of structure and patterns in the impact in a user's career. By showing that the results hold on multiple datasets and are robust to shuffling careers, we establish that these results are inherent properties of social media and not artifacts of sampling.

Clustering of impact. We first show that the most impactful tweets of a user occur close together temporally for a large set of users. This result does not hold when user's careers are shuffled, indicating that real careers are characterized by such effects.

Hot streaks. Next, we generalize the above finding and show that there is not just clustering of individual popular tweets, but extended periods of consistently high impact, which we call hot streaks. Our analysis of the length of hot streaks shows that most users have hot streaks of around 10–20 tweets, and some users, of over 100 tweets. We also show that retweets obtained during hot streaks do not constitute a major chunk of the total retweets accrued by a user. Looking at how various factors affect a hot streak, we find significant differences in the network, content, and activity of a user. A classifier to predict whether a week belongs to a hot streak or not performs reasonably well, with an accuracy close to 75%.

Impact around the most retweeted tweet. Finally, we show that there is a build-up and drop-off surrounding a user's most retweeted tweets. We tried to quantify if this is a prevalent phenomenon by looking at the slopes of linear fits for the 10 tweets before and after the most retweeted tweet, finding that this is a common trend among a majority of users. Even though individual tweets may become viral at random (Martin et al. 2016), such underlying structure around the most

retweeted of an individual gives hope for predicting which tweets might become viral.

Limitations. This work has several limitations. First, our datasets are not completely randomly sampled, and hence we cannot be sure about the generalizability of the results. Our datasets are picked in such a way to ensure at least some activity on the users' behalf. So the trends reported here may not hold for a random Twitter user. However, our findings are novel and point towards structural properties in impact for an important subset of users.

Second, our study cannot fully explain the mechanisms by which these effects occur. The underlying process could be a complex mix of the features we considered along with other external, dataset/domain/time-specific factors. It may well be that some of this phenomena cannot be explained with observable features (Martin et al. 2016), and trying to find explanations could be futile (Langer 1975).

Future work. The results presented in this paper are just one step towards understanding how user behavior evolves on social media. The large, rich, long-term dataset we created provides immense opportunities to explore many other aspects further, including downstream activities such as using these findings to study virality, influence, etc. We looked at individual characteristics of impact. It would be interesting to see if such individual patterns propagate in the network. For instance, if a user is in a hot streak, do users in her network also experience hot streaks? In preliminary work, we observed similar phenomena on other social network platforms, such as YouTube and Instagram, though the effects were less pronounced and did not hold for a majority of the users. Hence, future work should investigate how such trends generalize across platforms.

References

Antonakaki, D.; Ioannidis, S.; and Fragopoulou, P. 2018. Utilizing the average node degree to assess the temporal growth rate of Twitter. *Social Network Analysis and Mining* 8(1):12–20.

Ayton, P., and Fischer, I. 2004. The hot hand fallacy and the gambler's fallacy: Two faces of subjective randomness? *Memory & Cognition* 32(8):1369–1378.

Barabasi, A.-L. 2005. The origin of bursts and heavy tails in human dynamics. *Nature* 435(7039):207.

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.

Chierichetti, F.; Kleinberg, J. M.; Kumar, R.; Mahdian, M.; and Pandey, S. 2014. Event detection via communication pattern analysis. In *ICWSM*.

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

Gandica, Y.; Carvalho, J.; Aidos, F. S. D.; Lambiotte, R.; and Carletti, T. 2016. On the origin of burstiness in human behavior: The Wikipedia edits case. *arXiv preprint arXiv:1601.00864*.

Gao, S.; Ma, J.; and Chen, Z. 2015. Modeling and predicting retweeting dynamics on microblogging platforms. In *WSDM*.

Garimella, K., and Weber, I. 2017. A long-term analysis of polarization on twitter. In *Proceedings of the 11th AAAI International Conference on Web and Social Media*, 528–531. AAAI.

Gilovich, T.; Vallone, R.; and Tversky, A. 1985. The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology* 17(3):295–314.

Goel, S.; Anderson, A.; Hofman, J.; and Watts, D. J. 2015. The structural virality of online diffusion. *Management Science* 62(1):180–196.

Jenders, M.; Kasneci, G.; and Naumann, F. 2013. Analyzing and predicting viral tweets. In *WWW*.

Langer, E. J. 1975. The illusion of control. *Journal of Personality and Social Psychology* 32(2):311.

Liu, S.; Yamada, M.; Collier, N.; and Sugiyama, M. 2013. Change-point detection in time-series data by relative density-ratio estimation. *Neural Networks* 43:72–83.

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*.

Miller, J. B., and Sanjurjo, A. 2018. Surprised by the hot hand fallacy? A truth in the law of small numbers. *Econometrica* 86(6):2019–2047.

Moskvina, V., and Zhigljavsky, A. 2003. An algorithm based on singular spectrum analysis for change-point detection. *Communications in Statistics-Simulation and Computation* 32(2):319–352.

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

Ross, G. J. 2015. Parametric and nonparametric sequential change detection in R: The cpm package. *Journal of Statistical Software* 66(3):1–20.

Sinatra, R.; Wang, D.; Deville, P.; Song, C.; and Barabási, A.-L. 2016. Quantifying the evolution of individual scientific impact. *Science* 354(6312):aaf5239.

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*.

Thelwall, M. 2017. The heart and soul of the web? Sentiment strength detection in the social web with SentiStrength. In *Cyberemotions*. Springer. 119–134.

Wang, H., and Song, M. 2011. Ckmeans.1d.dp: Optimal *k*-means clustering in one dimension by dynamic programming. *The R Journal* 3(2):29–33.

Zhao, Q.; Erdogdu, M. A.; He, H. Y.; Rajaraman, A.; and Leskovec, J. 2015. Seismic: A self-exciting point process model for predicting tweet popularity. In *KDD*.