Boston University

OpenBU **http://open.bu.edu**

BU Open Access Articles BU Open Access Articles

2014-07-30

Discovering social events through online attention

Dror Y Kenett, Fred Morstatter, H Eugene Stanley, Huan Liu. 2014. "Discovering Social Events through Online Attention." PLOS ONE, Volume 9, Issue 7, 7 pp. https://doi.org/10.1371/journal.pone.0102001
https://hdl.handle.net/2144/39947
Downloaded from DSpace Repository, DSpace Institution's institutional repository

OPEN ACCESS Freely available online

PLOS ONE

Discovering Social Events through Online Attention

Dror Y. Kenett[1]*, Fred Morstatter[2], H. Eugene Stanley[1], Huan Liu[2]

1 Center for Polymer Studies and Department of Physics, Boston University, Boston, Massachusetts, United States of America, **2** School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, Tempe, Arizona, United States of America

Abstract

Twitter is a major social media platform in which users send and read messages (''tweets'') of up to 140 characters. In recent years this communication medium has been used by those affected by crises to organize demonstrations or find relief. Because traffic on this media platform is extremely heavy, with hundreds of millions of tweets sent every day, it is difficult to differentiate between times of turmoil and times of typical discussion. In this work we present a new approach to addressing this problem. We first assess several possible ''thermostats'' of activity on social media for their effectiveness in finding important time periods. We compare methods commonly found in the literature with a method from economics. By combining methods from computational social science with methods from economics, we introduce an approach that can effectively locate crisis events in the mountains of data generated on Twitter. We demonstrate the strength of this method by using it to locate the social events relating to the Occupy Wall Street movement protests at the end of 2011.

Citation: Kenett DY, Morstatter F, Stanley HE, Liu H (2014) Discovering Social Events through Online Attention. PLoS ONE 9(7): e102001. doi:10.1371/journal.pone.0102001

Editor: Matjaz Perc, University of Maribor, Slovenia

Received May 14, 2014; **Accepted** June 13, 2014; **Published** July 30, 2014

Copyright: © 2014 Kenett et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability: The authors confirm that, for approved reasons, some access restrictions apply to the data underlying the findings. Data is collected from twitter API study whose authors may be contacted at drorkenett@gmail.com

Funding: HES and DYK thank the Office of Naval Research (ONR, Grant N00014-09-1-0380, Grant N00014-12-1- 0548), Keck Foundation, and the National Science Foundation for support. FM and HL thank the support of the Office of Naval Research (ONR, Grant N000141010091 and N000141110527). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: drorkenett@gmail.com

Introduction

Over the past several years various Internet social media platforms have enabled people to communicate, locate resources, and disseminate information during times of turmoil, e.g., natural disasters, health epidemics, or social unrest. Twitter, one major social media platform, has emerged as a leading social media outlet. With 200 million users sharing 140-character text messages (''tweets'') over 400 million times each day [1], Twitter's popularity and influence on world events have made it a hot topic for social media research [2]. Research on Twitter began in 2010 when researchers saw its potential for rapid communication and information diffusion. The field of computational social science has been rapidly expanding in response to the influence of Twitter and other online social platforms [3,4], and new insights into social structure and social dynamics are emerging [5–15]. Twitter has also been a focus in studies of humanitarian assistance/disaster relief (HA/DR) efforts [16–18] and in the tracking of disease epidemics [19]. Because Twitter enables the real-time propagation of information to large groups of users, it is an ideal environment for the dissemination of breaking news from news gatherers and from on-site locations where events are taking place.

Twitter has several features of interest to the research community. Twitter's ''retweet'' feature, which allows users to push content through the network by forwarding it to their followers, has elicited much research on how information propagates in social media [20,21], how retweets facilitate online conversation [22], and how retweets factor in times of crisis [23]. Twitter uses a special text feature (a ''hashtag'') in which transmitted words are prefixed with a ''#'' sign. Every hashtag has a page showing the history of all the tweets containing that hashtag in the text, and this creates a community of users discussing that particular hashtag [24]. This encourages users interested in the topic to use the associated hashtag in their tweets to increase the audience of their tweet, and the study of this tagging behavior in Twitter has become an extremely active area of research [25–27]. In addition to text, users can also annotate their tweet with their current location, adding what is called a ''geotag.'' Only about one percent of all tweets are geotagged, yet they still provide background information about an event. Recent work has focused on combining location with textual content to detect topics more relevant to specific regions [28–30]. Because geotags are so sparse, recent work has also focused on associating non-geotagged tweets with a location to better understand the context of the tweet [31,32].

Social media platforms now strongly factor in the spreading of ideas and the organization of social movements. Over the past few years, social media has played a key role in such significant events as the Arab Spring uprisings and the violent demonstrations organized in London. Twitter is popular with users seeking to spread information about a cause. Because each message can be no longer than 140 characters, communication spreading information concerning protest gatherings, earthquake relief, or the location of aid stations is extremely rapid [33,34]. Participants in the Arab Spring used Twitter to quickly coordinate protests [35,36]. Occupy Wall Street, a movement protesting the wealth disparity in the United States, was largely organized on Twitter under the hashtag ''#OccupyWallStreet.'' As the movement spread and authorities began to retaliate, protesters used Twitter

to report abuses by police, thus bringing more attention to their cause. Social media became so central during the Arab Spring protests that the regimes in such countries as Egypt and Syria cut the protesters' access to the Internet. During Hurricane Sandy, authorities used Twitter to spread news of power outages and the locations of resources for those affected by the storm.

Because Twitter provides rapid communication and information diffusion, millions of people use it to keep up with current events and create their own discussion threads. Because activity on the Twitter site is huge, it is difficult to differentiate periods of focused discussion from periods of casual chatter. How do we identify the key periods of discussion? How do we filter out the noise and locate the main issues of discussion people are discussing at any given time?

We will first attempt to locate the periods where tweets reflect actual events on the ground. To harness the abundance of data produced by Twitter, we need a highly-scalable method to find key time periods of big events in social media. We focus on the Twitter activity surrounding Occupy Wall Street–the vast Twitter discussion of that event worldwide–and compare several methods of quantifying social communication.

Occupy Wall Street Movement

The Occupy Wall Street movement began on 17 September 2011 in New York City. The movement was largely promoted on social media, and many hashtags were used to discuss the event. The chief driving force behind this movement was the growing wealth disparity between rich and poor in the United States [37]. As the movement gained attention, other Occupy movements emerged in cities across the US. As citizens in other countries identified with the core concerns of the movement, similar actitivies spread across the globe. By 15 October 2011, 951 similar protests had occurred in 82 countries [38]. As the movement continued to grow it was officially endorsed by a number of city governments and labor unions [39].

In this study we collected tweet data from 14 September 2011 through 3 April 2012 using the parameters shown in Table 1 and encompassing 15,736,835 tweets with 402,758 unique hashtags and 6,967,392 retweets. We used Twitter's free, publicly-available data source, the Streaming API (see https://dev.twitter.com/docs/streaming-apis) to collect the data, in which three parameters are supported: keywords (which can be supplied in the form of words, phrases, or hashtags), locations (supplied as a geographic bounding box), and users. Every parameter is treated as an "OR" condition. That is, a tweet will be returned from the Streaming API if it contains at least one of the keywords, if it is produced from within the bounding box using a "geotag", or if it is authored by one of the users specified in the parameters. When a user geotags their tweet, their location is provided as part of the metadata using the GPS sensor on their device (for more information see http://support.twitter.com/articles/78525-faqs-about-the-tweet-location-feature). All parameters supplied to (and tweets returned by) the Streaming API were managed using TweetTracker [40].

Many of the tweets collected were geotagged, with a large number of the geotagged tweets coming from New York City. Figure 1 shows a heatmap of the tweets produced on different days and we can see the extreme cases of geotagged tweets. Figure 1(a) shows the tweets for 15 November 2011, when the New York Police Department attempted to remove protesters from Zuccotti Park. Figure 1(b) shows the tweets for 26 December 2011, when protesting had dwindled. In between these two extremes of activity, is a more general pattern of discussion centered around the protests in Zuccotti Park.

Measures of Social Attention

The Herfindahl-Hirschman index (also known as the Herfindahl index, or HHI) is a measure of the size of firms in relation to an industry and indicates the degree of competition among them. Named after economists Orris C. Herfindahl and Albert O. Hirschman, it is an economic concept widely applied in competition law, antitrust law, and technology management. The measure is also used by the United States Department of Justice when evaluating mergers (see http://www.justice.gov/atr/public/guidelines/hhi.html). The result is proportional to the average market share, weighted by market share. As such, it can range from 0 to 1, moving from a huge number of very small firms (with a value reaching zero) to a single monopolistic producer (with a value reaching 1). Increases in HHI generally indicate a decrease in competition and an increase of market power, whereas decreases indicate the opposite.

We use a normalized HHI [42], H^*, which is defined as

$$H^* \equiv \frac{H - 1/N}{1 - 1/N} \tag{1}$$

where

$$H \equiv \sum_{i=1}^{N} s_i^2 \tag{2}$$

N is the number of hashtags, and s is the percentage of the aggregate measure ($\sum_1^N s_i = 1$).

We utilize the HHI as a "thermostat" of social attention. Each hashtag represents a "firm" and the number of users tweeting this hashtag relative to the total number of users in a given time period represents the hashtag's "market cap." This enables us to examine the HHI value of different hashtags for a given time period. High HHI values indicate a strong focus on a specific topic, and low HHI values indicate a diffused focus among a wide variety of topics.

We use HHI analysis to study the OWS dataset and calculate the HHI value for a time horizon of a single day, using the number of users and hashtags. One concern of the HHI is that it is dependent on the number of tweets produced in a given time interval. Figure 2 shows the time evolution of the HHI. Figure 3 compares the HHI with its underlying parameters: the number of users and the number of hashtags. Here the diagonal figures represent the histogram of values for each of these three parameters, whereas the off-diagonal panels represent a comparison of the values of two different parameters. Studying this figure, it is clear that the HHI is not merely a function of either of these two parameters.

Another attention-based measure of social attention is the entropy [43] of the hashtags over a given time period. We here consider the hashtag probability to be the number of times the hashtag is used over the number of times all hashtags are used in a given time interval. The hashtag entropy is calculated by first assigning the probability of a given hashtag, p_i, using the fraction of users who tweeted this hashtag in the given time horizon, summing over all hashtags such that:

Table 1. Parameters supplied to the Streaming API for each of the data sources.

Data Set	Keywords	Geoboxes	User Timelines
Occupy Wall Street	#occupywallstreet, #ows, #occupyboston, #p2, #occupywallst, #occupy, #tcot, #occupytogether, #teaparty, #99percent, #nypd, #takewallstreet, #occupydc, #occupyla, #usdor, #occupysf, #solidarity, #15o, #anonymous, #citizenradio, #gop, #sep17, #occupychicago, #occupyphoenix, #occupyoakland	None	None

Coordinates below the boundary box indicate the Southwest and Northeast corner, respectively. No users were tracked during the course of data collection.
doi:10.1371/journal.pone.0102001.t001

$$S_{\text{Hashtag}} = -\sum_{i}^{N} p_i log(p_i), \quad (3)$$

where N is the number of hashtags in the given time horizon. In evaluating the effectiveness of our HHI-based approach, we compare its performance as a classifier of the ground truth relative to that of the other three models.

Indicators of Activity in Social Media

To search for periods of focused discussion, we locate time periods with a large number of tweets or time periods with a large number of unique hashtags and test whether these two simple measures can enable us to identify the focused discussion periods in the dataset. We quantitatively test the two simple measures by performing a receiver operating characteristic (ROC) curve analysis. The ROC curve plots the fraction of true positives out of the positives and the fraction of false positives out of the negatives for a binary classifier system. ROC curve analysis is a standard method in signal detection theory as well as in psychology, medicine, and biometrics [41]. One key measure from the ROC curve is the area-under-curve (AUC) score, the measure of the area under the ROC curve. The ROC AUC varies from 0.50 (totally random classification) to 1.0 (perfect classification).

We vary the measurement threshold to identify important days, and compare the results with the ground truth. The true positive rate is defined as the fraction of the actual significant days, as listed by the ground truth, that are also identified by the measure. The false positive rate is the fraction of days that are not identified in the ground truth, but are identified as significant by the measure. Each point in the ROC curve corresponds to one selection threshold. A random classifier yields a diagonal line (AUC = 0.50) from the bottom-left to the top-right corner. The greater the curve's distance above the diagonal line, the stronger the model's predictive power. To obtain ground truth, we extract dates from the Wikipedia timeline of the OWS protests (see http://en.wikipedia.org/wiki/Timeline_of_Occupy_Wall_Street). Next, by varying the threshold that indicates "important" days, we find the ROC curve, shown in Figure 4(a). The ROC AUC of the top hashtags is 0.36 and the ROC AUC of the top tweets is 0.42, both scoring worse than a perfectly random classifier.

Although we can mitigate the poor results obtained in the experiment by inverting the class labels–giving the inverted hashtag and tweet indicators ROC AUCs of 0.64 and 0.58,

Figure 1. Heatmap of geotagged Twitter activity. Twitter activity related to the Occupy Wall-Street (OWS) Movement, collected for hashtags, or topics, used by protests or members of the movement. The "redder" areas indicate regions with more tweets. Here we see two extremes of geotagging behavior. Panel (a) shows the tweets for 15 November 2011, when the New York Police Department attempted to remove protesters from Zuccotti Park. Panel (b) shows the tweets for 26 December 2011, when protesting had dwindled. In between these two extremes of activity, is a more general pattern of discussion centered around the protests in Zuccotti Park.
doi:10.1371/journal.pone.0102001.g001

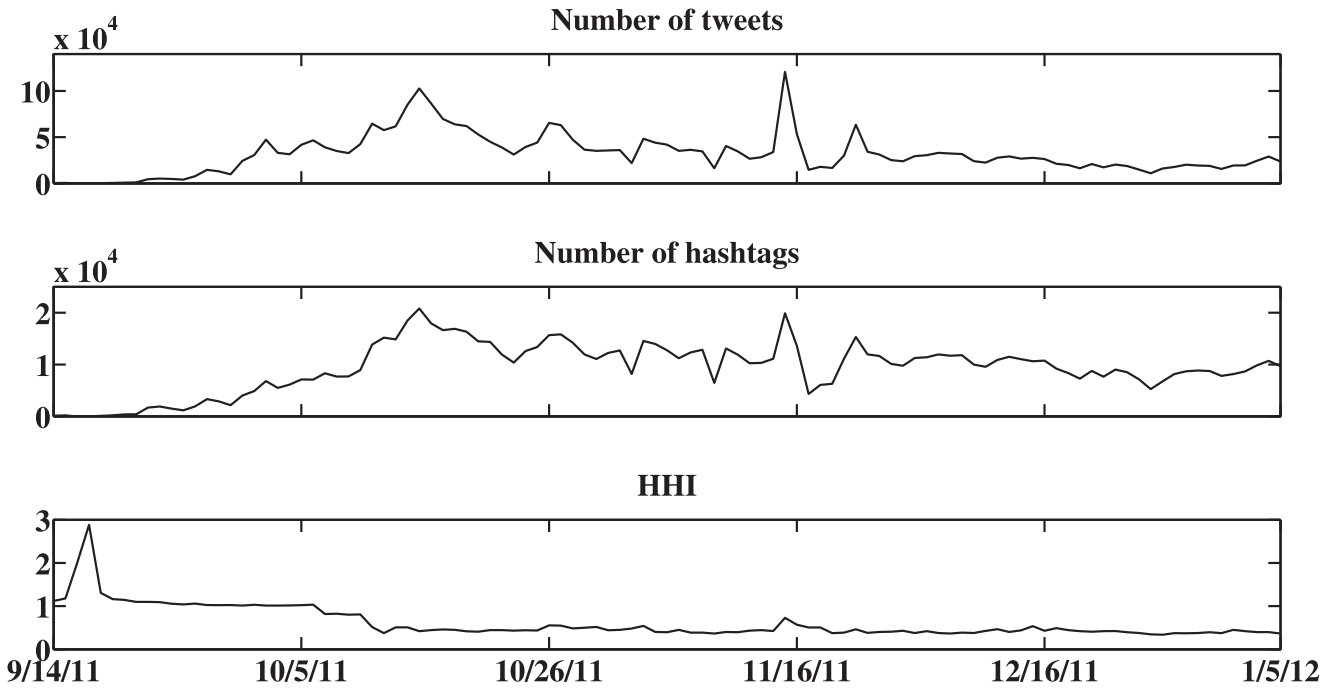

Figure 2. Time evolution of the number of tweets (top), number of hashtags (middle), and Herfindahl-Hirsch Index (HHI) parameter (bottom) for the OWS dataset, on a daily time horizon. The HHI calculates how diverse the discussion is on Twitter, by calculating how many messages are associated with a given hashtag, and ranges from a value of 0, for highly diverse discussion, to 1, when all messages are focused on only one hashtag.
doi:10.1371/journal.pone.0102001.g002

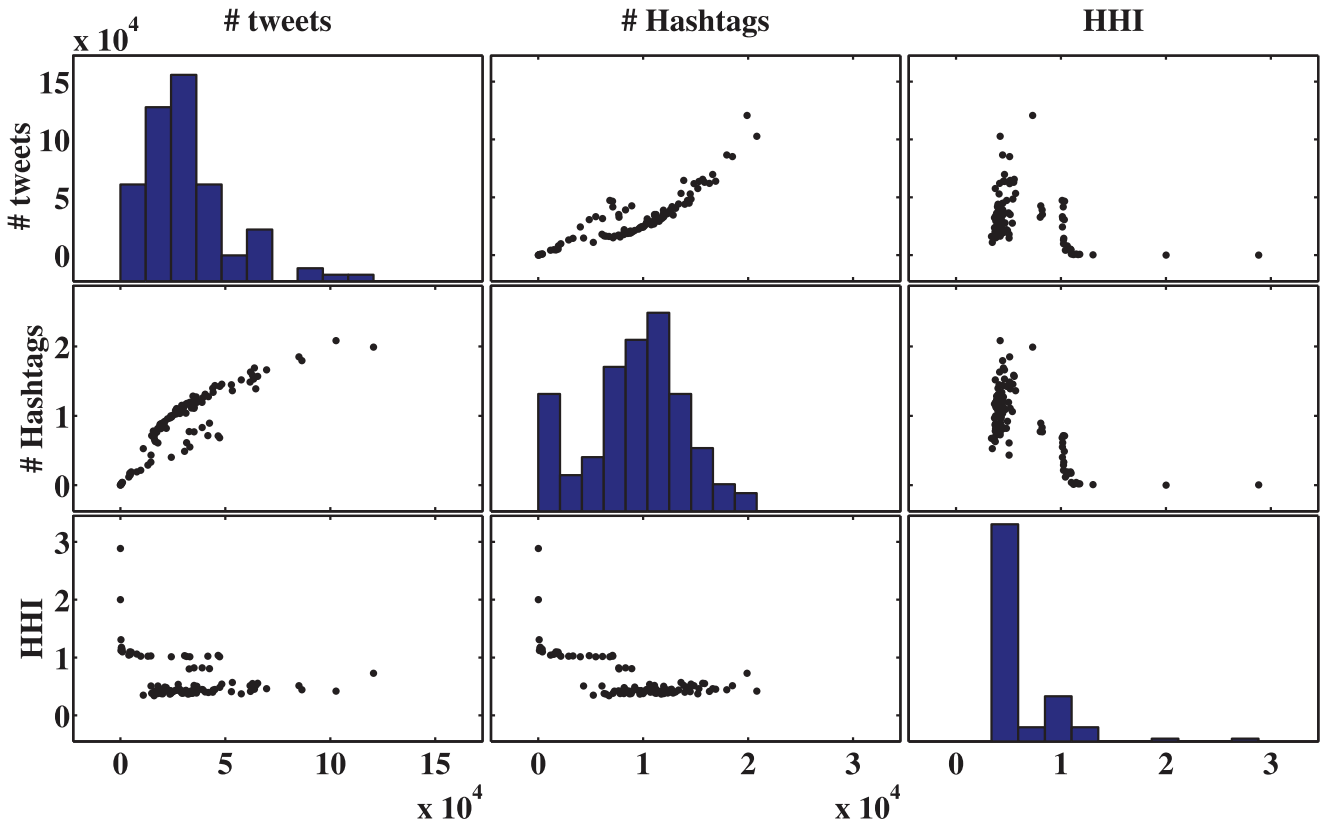

Figure 3. Comparison of the HHI to its underlying parameters: the number of tweets, and number of hashtags. Here, the diagonal figures represent the histogram of values for each of these three parameters, whereas the off diagonal panels represent a comparison of the values of two different parameters. It is clear by studying these figures that the HHI is not merely a function of either the number of tweets or number of users.
doi:10.1371/journal.pone.0102001.g003

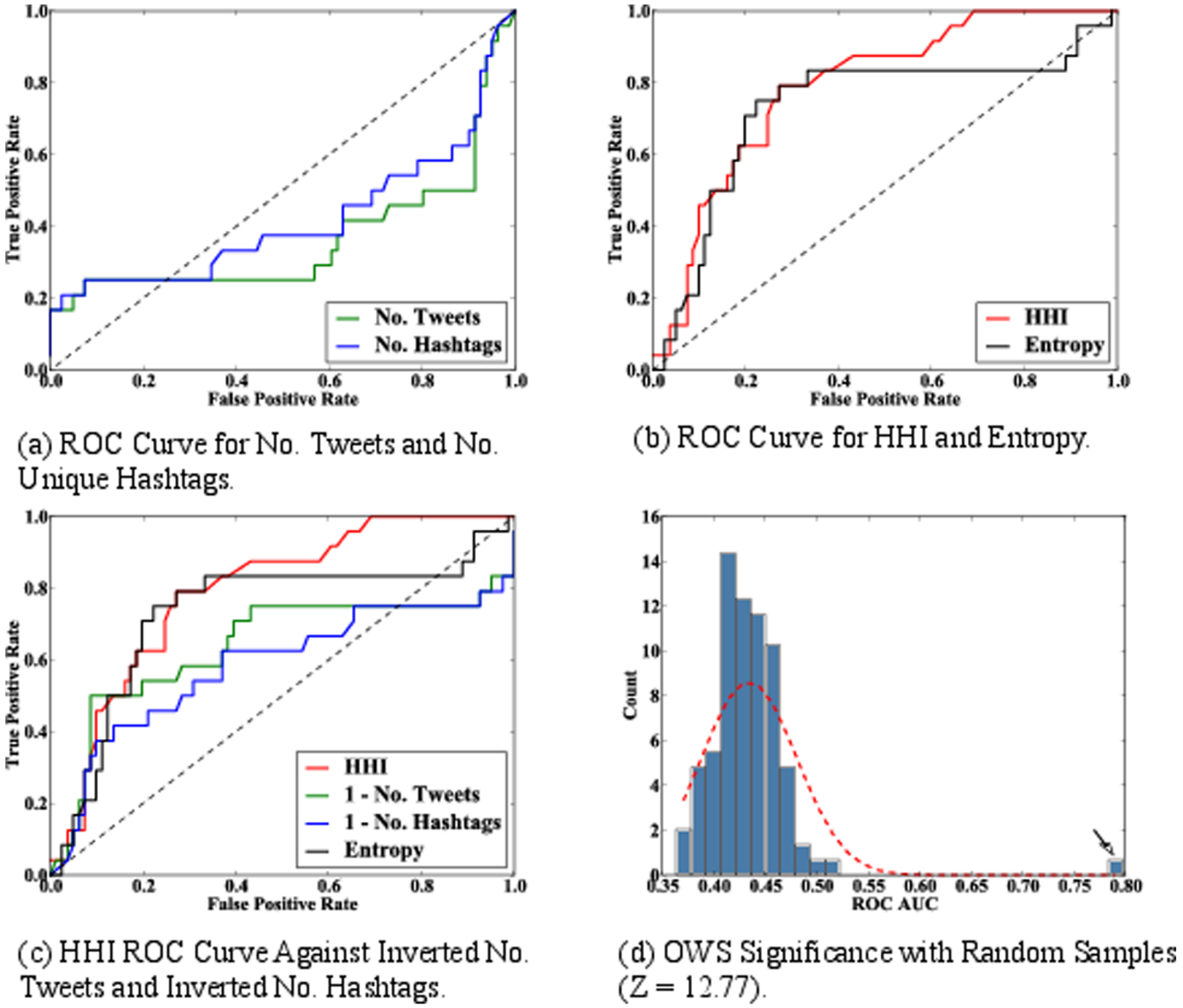

Figure 4. HHI ROC analysis. (a) ROC curve of number tweets and number unique hashtags as classifiers for finding significant dates in the dataset. Number of tweets AUC = 0.42 and number of unique hashtags AUC = 0.36. (b) ROC curve of the HHI and Entropy classifiers. HHI AUC = 0.79, entropy AUC = 0.72. The focus-based classifiers provide the best classification when compared with the other methods, with the HHI being the best predictor. (c) ROC curve of the four classifiers - one minus number of tweets, one minus number of hashtags, and hashtag entropy - and their performance in identifying the ground truth. This is done as a below-random (<0.50) AUC means that the class labels should be inverted. (d) Distribution of the HHI AUC values for prediction of the ground truth for many random samples of the OWS dataset. The arrow in this figure represents the measure of the unshuffled data.
doi:10.1371/journal.pone.0102001.g004

respectively–this approach has intuitive problems. Predicting periods with few unique hashtags and few tweets is not relevant to the problem of finding periods of intense discussion. Therefore, there is a need for a measure of social attention that focuses not only on the number of tweets or unique hashtags, but also on their "attention"–the degree to which users congregate around them.

Social Attention as a Detector of Real-World Events

We next use the HHI as a thermostat for social focus during times of crisis. Alternate approaches would be to use the number of tweets, the number of unique hashtags produced in a given day, or the entropy of the hashtags used in the time period.

Figure 4(b) and Figure 4(c) shows the results of performing all four indicators on the OWS dataset, with HHI and entropy attaining ROC values of 0.79 and 0.72, respectively. The attention-based indicators provide the best classification when compared with the other methods, with the HHI being the best predictor.

To confirm that the classification accuracy of the HHI comes from the hashtag selection made by the users and is not merely an artifact of the volume of tweets, we randomly shuffle the tweets based on the time they were produced. If the effectiveness of the HHI is due to the volume of tweets, then there should be no significant difference between the initial AUC and those from the datasets with the randomly shuffled timestamps.

To this end, we create a unique set, T, of all the timestamps from tweets in the dataset. For each tweet we then randomly choose a timestamp from T and assign it to the tweet, without replacement. Using this shuffled dataset we calculate the ROC AUC score. We repeat this process 100 times to determine the distribution of the shuffled tweets. Finally, we compare the AUC score of the original data with the shuffled data to see if it differs significantly ($\mu \pm 3\sigma$) from the center of the random shuffles. Figure 4(d) shows the distribution of ROC AUC scores of the randomly shuffled data. The Z–score of the original data, calculated as

$$Z_{score} = \frac{AUC - \mu}{\sigma}, \quad (4)$$

is +12.77, significantly outside of the control bounds.

Summary

In this work we investigate the problem of finding real-world events quickly as they unfold in large, noisy social media data. We seek to find a measure of attention in social media. The naive choice for this aim is to investigate the number of tweets and number of unique hashtags, and we find that this approach is unsatisfactory. One possible explanation for the poor performance of these measures could be that extraneous conversation on Twitter leads to spikes in activity not relevant to the event. We investigate two additional methods, HHI and entropy, and find that they are successful detectors of these periods of intense discussion. HHI, a measure borrowed from the economics literature adapted for use in social media, yields the best results for identifying the times of intense discussion.

Our results indicate that significant social events cause the discussion on Twitter to move from many subjects to a few, as demonstrated through the Herfindahl index. In terms of classical information theory, this can be conversely related to a measure of entropy of the discussion topics, where our results show that significant events are related to drops in the entropy (or high HHI). Entropy has been used in the past to study traditional media and online media [44–46]. Our results show that while the two measures are closely related, the HHI outperforms entropy as a detector of significant events. This work presents a first use of the HHI to study social attention on Twitter.

Although discussions in Twitter and in digital social media in general are extremely heterogeneous, when a significant event occurs discussions converge to the event and become extremely homogeneous. The point at which this switching occurs indicates the magnitude of the event. Because of this, the proposed Herfindahl index provides a means of detecting significant events, and provides a simple measure to filter significant events and centers of attention in the social online media. This simple yet sophisticated measure can provide important insights to people of different background and needs, such as scientists, social-media based marketing professionals, policy and decision makers, and a multitude of relief agency workers.

Acknowledgments

We wish to thank Shlomo Havlin for all of his comments and suggestions for this work.

Author Contributions

Conceived and designed the experiments: DYK FM HES HL. Performed the experiments: DYK FM HES HL. Analyzed the data: DYK FM HES HL. Contributed reagents/materials/analysis tools: DYK FM HES HL. Contributed to the writing of the manuscript: DYK FM HES HL.

References

1. Tsukayama H (2013) Twitter turns 7: Users send over 400 million tweets per day. The Washington Post. Available'' http://www.washingtonpost.com/business/technology/twitter-turns-7-users-send-over-400-million-tweets-per-day/2013/03/21/2925ef60-9222-11e2-bdea-e32ad90da239_story.html. Accessed 2014 Jul 4.
2. Kumar S, Morstatter F, Liu H (2014) Twitter Data Analytics. Springer.
3. Lazer D, Pentland AS, Adamic L, Aral S, Barabasi AL, et al. (2009) Life in the network: the coming age of computational social science. Science (New York, NY) 323: 721.
4. Conte R, Gilbert N, Bonelli G, Cioffi-Revilla C, Deffuant G, et al. (2012) Manifesto of computational social science. The European Physical Journal Special Topics 214: 325–346.
5. Rybski D, Buldyrev SV, Havlin S, Liljeros F, Makse HA (2012) Communication activity in a social network: relation between long-term correlations and inter-event clustering. Scientific reports 2.
6. Gallos LK, Rybski D, Liljeros F, Havlin S, Makse HA (2012) How people interact in evolving online affiliation networks. Physical Review X 2: 031014.
7. Ciulla F, Mocanu D, Baronchelli A, Gonçalves B, Perra N, et al. (2012) Beating the news using social media: the case study of american idol. EPJ Data Science 1: 1–11.
8. Gonzalez MC, Hidalgo CA, Barabasi AL (2008) Understanding individual human mobility patterns. Nature 453: 779–782.
9. Eagle N, Pentland AS, Lazer D (2009) Inferring friendship network structure by using mobile phone data. Proceedings of the National Academy of Sciences 106: 15274–15278.
10. Havlin S, Kenett DY, Ben-Jacob E, Bunde A, Cohen R, et al. (2012) Challenges in network science: Applications to infrastructures, climate, social systems and economics. European Physical Journal-Special Topics 214: 273.
11. Gao J, Hu J, Mao X, Perc M (2012) Culturomics meets random fractal theory: insights into long-range correlations of social and natural phenomena over the past two centuries. Journal of The Royal Society Interface 9: 1956–1964.
12. Kenett DY, Portugali J (2012) Population movement under extreme events. Proceedings of the National Academy of Sciences 109: 11472–11473.
13. Moat H, Curme C, Avakian A, Kenett DY, Stanley HE, et al. (2013) Quantifying wikipedia usage patterns quantifying wikipedia usage patterns before stock market moves. Scientific Reports 3: 1801.
14. Preis T, Moat HS, Stanley HE (2013) Quantifying trading behavior in financial markets using google trends. Scientific Reports 3: 1684.
15. Moat HS, Preis T, Olivola CY, Liu C, Chater N (2014) Using big data to predict collective behavior in the real world. Behavioral and Brain Sciences 37: 92–93.
16. De Longueville B, Smith RS, Luraschi G (2009) "OMG, from here, I can see the flames!": A use case of mining location based social networks to acquire spatio-temporal data on forest fires. In: Proceedings of the 2009 International Workshop on Location Based Social Networks. New York, NY, USA: ACM, LBSN'09, pp. 73–80. doi:10.1145/1629890.1629907. URL http://doi.acm.org/10.1145/1629890.1629907.
17. Sakaki T, Okazaki M, Matsuo Y (2010) Earthquake shakes twitter users: real-time event detection by social sensors. In: Proceedings of the 19th international conference on World wide web. New York, NY, USA: ACM, WWW'10, pp. 851–860. doi:10.1145/1772690.1772777. Available: http://doi.acm.org/10.1145/1772690.1772777. Accessed 2014 Jul 4.
18. Morstatter F, Lubold N, Pon-Barry H, Pfeffer J, Liu H (2014) Finding eyewitness tweets during crises. In: Association of Computational Lingustics Workshop on Language Technologies and Association of Computational Lingustics Workshop on Language Technologies and Computational Social Science.
19. Pastor-Satorras R, Vespignani A (2001) Epidemic spreading in scale-free networks. Phys Rev Lett 86: 3200–3203.
20. Nagarajan M, Purohit H, Sheth A (2010) A qualitative examination of topical tweet and retweet practices. In: Fourth International AAAI Conference on Weblogs and Social Media. AAAI.
21. Kwak H, Lee C, Park H, Moon S (2010) What is twitter, a social network or a news media? In: Proceedings of the 19th international conference on World wide web. New York, NY, USA: ACM, WWW'10, pp. 591–600. doi:10.1145/1772690.1772751.
22. Boyd D, Golder S, Lotan G (2010) Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In: System Sciences (HICSS), 2010 43rd Hawaii International Conference on. pp. 1–10. doi:10.1109/HICSS.2010.412.
23. Mendoza M, Poblete B, Castillo C (2010) Twitter under crisis: can we trust what we RT? In: Proceedings of the First Workshop on Social Media Analytics. New York, NY, USA: ACM, SOMA '10, pp. 71–79. doi:10.1145/1964858.1964869. Available: http://doi.acm.org/10.1145/1964858.1964869. Accessed 2014 Jul4.

24. Yang L, Sun T, Zhang M, Mei Q (2012) We know what @you #tag: does the dual role affect hashtag adoption? In: Proceedings of the 21st international conference on World Wide Web. New York, NY, USA: ACM, WWW'12, pp. 261–270. doi:10.1145/2187836.2187872. Available: http://doi.acm.org/10.1145/2187836.2187872. Accessed 2014 Jul 4.
25. Romero DM, Meeder B, Kleinberg J (2011) Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In: Proceedings of the 20th international conference on World wide web. New York, NY, USA: ACM, WWW'11, pp. 695–704. doi:10.1145/1963405.1963503. URL.
26. EfronM(2010) Hashtag retrieval in a microblogging environment. In: Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval. New York, NY, USA: ACM, SIGIR '10, pp. 787–788. doi:10.1145/1835449.1835616.
27. Weng J, Lim EP, He Q, Leung CK (2010) What do people want in microblogs? measuring interestingness of hashtags in twitter. In: Data Mining (ICDM), 2010 IEEE 10th International Conference on. pp. 1121–1126. doi:10.1109/ICDM.2010.34.
28. Yin Z, Cao L, Han J, Zhai C, Huang T (2011) Geographical topic discovery and comparison. In: Proceedings of the 20th international conference on World wide web. New York, NY, USA: ACM, WWW'11, pp. 247–256. doi:10.1145/1963405.1963443.
29. Pozdnoukhov A, Kaiser C (2011) Space-time dynamics of topics in streaming text. In: Proc. of the 3rd ACM SIGSPATIAL Int'l Workshop on Location-Based Social Networks. New York, NY, USA: ACM, LBSN'11, pp. 1–8. doi:10.1145/2063212.2063223.
30. Morstatter F, Pfeffer J, Liu H, Carley KM (2013) Is the sample good enough? comparing data from twitters streaming api with twitters firehose. In: International Conference on Weblogs and Social Media. pp. 400–408.
31. Cheng Z, Caverlee J, Lee K (2010) You Are Where You Tweet: A Content-Based Approach to Geo-locating Twitter Users. In: Proceedings of The 19th ACM International Conference on Information and Knowledge Management. Toronto, Ontario, Canada: International Conference on Information and Knowledge Management, pp. 759–768. doi:10.1145/1871437.1871535.
32. Li R, Wang S, Deng H, Wang R, Chang KCC (2012) Towards social user profiling: unified and discriminative influence model for inferring home locations. In: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. New York, NY, USA: ACM, KDD '12, pp. 1023–1031. doi:10.1145/2339530.2339692.
33. Bennett S (2011). Twitter: Faster than earthquakes. Media Bistro. Available: http://www.mediabistro.com/alltwitter/twitter-earthquake-video_b13147. Accessed 2014 Jul 4.
34. Mourtada R, Salem F (2011) Civil movements: The impact of facebook and twitter. Arab Social Media Report 1.
35. Huang C (2011) Facebook and twitter key to arab spring uprisings: report. The National Abu Dhabi Media 6.
36. Campbell DG (2011) Egypt Unshackled: Using Social Media to @#:) the System. Amherst, NY: Cambria Books.
37. Berrett D (2011) Intellectual roots of wall st. protest lie in academe. The Chronicle of Higher Education. Available: http://chronicle.com/article/Intellectual-Roots-of-Wall/129428/. Accessed 2014 Jul 4.
38. Chappell B (2011). Occupy wall street: From a blog post to a movement. http://www.npr.org/2011/10/20/141530025/occupy-wall-street-from-a-blog-post-to-a-movement. Accessed 2014 Jul 4.
39. (2011) Occupy wall street gets union support. United Press International. Available: http://www.upi.com/Top_News/US/2011/09/30/Occupy-Wall-Street-gets-union-support/UPI-89641317369600/. Accessed 2014 Jul 4.
40. Kumar S, Barbier G, Abbasi MA, Liu H (2011) Tweettracker: An analysis tool for humanitarian and disaster relief. In: Fifth International AAAI Conference on Weblogs and Social Media, ICWSM.
41. Swets JA (1996) Signal detection theory and ROC analysis in psychology and diagnostics: Collected papers. Lawrence Erlbaum Associates Mahwah, NJ.
42. Rhoades SA (1993) The herfindahl-hirschman index. Fed Res Bull 79: 188.
43. Cover TM, Thomas JA (2012) Elements of information theory. Wiley-Interscience.
44. McClelland CA (1961) The acute international crisis. World Politics 14: 182–204.
45. McClelland CA (1968). Access to berlin: the quantity and variety of events, 1948-1963. Available: http://www.econbiz.de/Record/access-to-berlin-the-quantity-and-variety-of-events-1948-1963-mcclelland-charles/10002418818. Accessed 2014 Jul 4.
46. Boydstun AE, Bevan DS, Thomas HF (2014) The importance of attention diversity and how to measure it. Public Policy and Administration 42(2): 173–196.