A Characterization on Online Influence Events in Social Media

Michael Twardos
Data Science
Klout, Inc.

Abstract

The consumption of media online is driven largely by the influence of members in one's social network. Influence is distributed across multiple channels within a social network and across many social networks. Using a snapshot of data from Klout's data warehouse, we characterize how influence is distributed across users, channels and social networks. We show that influence distribution across users is similar and strongly skewed towards the top users. Influence for a given user across channels within the same social networks is strongly correlated while influence across social networks is weakly positively correlated.

Measuring influence is key to the evolution of the web. It provides businesses with an indication of a user's value which can be used to advertise more efficiently and optimize customer relations. However, influence has traditionally been considered closer to a human emotion that a metric. While influence, like inspiration or happiness, is undoubtably real, the ability to say that one user is more influential than another has remained an exercize in subjectivity. Still, when a person changes their job or the food they eat or the person they vote for and they attribute those actions to the information they acquired from another person, this event is clearly distinguishable as an ”influence event”. In social media, identifying ”influence events” are even easier: comments, likes and reshares occurring on social networks including Facebook, Twitter, Google+, LinkedIn and Foursquare are just a few examples. In this paper, we use this concept of influence to build a better picture of online influence.

Social networks, consisting of the tightly connected communities of friends with shared interests make interaction easy and can be considered the precursor for online influence. Therefore, understanding the structure of social networks and how grow are an important background to studying influence. There is extensive research on the structure and evolution of social networks and other highly connected systems that are believed to behave similarly. The communities and groups of highly connected sets of users has often been referred to in literature as ”small world” networks. These communities are connected together by social bridges: types of users that span different webs of friends and interest and enable information to spread across the entire network. The probability distribution of connections in small world networks results is a power law and the users at the end of the tail correspond to the bridges. Recent work has used real world data to compared structure of social networks on several platforms [1] and verify the small-world properties. Scaling relations in all transportation networks that want to have been observed in everything from circulatory networks, networks of rivers [2] and energy transport in food webs [3].

Although there has been significant attention directed toward the structure of social networks, studies have shown that the social network is not the same at the network upon which influence occurs. In other words, attention is a prerequisite but not a guarantee for influence. One study has shown that users with many Twitter followers do not necessarily drive the most influential actions such as retweets and mentions [4] while another study has shown that even though a user has many followers, most are passive and do not engage with the content that is shared [5]. Still other work has conjectured that influence is related to reciprocity: a high correlation between inbound and outbound links is the key driver for influence [6].

With a lack influence data, models have been proposed that help to understand how influence occurs on a social network. Early work has considered how the social network evolves over time due to influence and constraints in attention and information propgation [8]. Other studies have suggested that a large amount of measurable influence occurs from a user's direct social network [7] and under time delayed correlations [9]. Other models attempted to capture the balance of influence from nearest neighbors (social network) and the larger public (mass media) [10].

A large amount of this work concerning the propagation of influence on social networks has not used real world data. This is partly because influence data is scattered across many channels within a social network as well as across many social networks. This leads to the question of relationships between channels and between social networks: Given that a user is influential on one, what does this mean about how influential they are on another? This is precisely the questions we investigate in this paper. For a given type of influential event we look at the distribution and compare distributions across social networks and channels of influence. Next we look at the relationship between pairs of influence events within a social network to understand the complimentary nature of intra social network influence. Finally, we perform the same analysis on influence events that occur on different social networks. This characterization provides insight into the places where influence occurs and can be used to predict total influence given limited data.

Data Overview

Klout is an online service that tracks influence events using the social network's data APIs and aggregates that data to compute a "Klout Score" . While we cannot uncover the details of the Klout score algorithm, we can use this data to characterize how these events are generated by influential users active in social media. Influence, broadly defined, means the ability to change the behavior of others. In the context of this discussion, influence has a more narrow definition. Depending what social network we are considering, content shared within can be "liked", "retweeted", "commented", "reshared" and so on. These actions are directly dependent on the initial act of posting and can be referred to as "influence events". This data was taken from a snapshot of Klout's users in March of 2012 or their last 90 day activity.

The Distribution of Influential Events

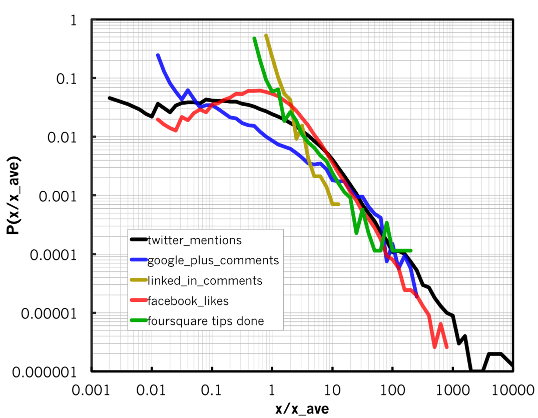

Figure 1: Normalized probability distributions for influential users on major social networks: Twitter, FB, G+, 4S and LI. The amount of influence a user generates is normazed by the average user's amount of influence generated. Overall, the distributions share strong similarities characterized by the majority of users at the lower values of influence with a tail of strong influencers decaying as a power law with an exponent of around -2.

Large amounts of influence events occur every day within the largest social networks. In this section we consider how these influence events are distributed across users. In Fig. 1 the probability distribution of influence events for some of the most common types of influence on the most common social networks is shown. The events are normalized by the average number of influence events generated per user: x/xave. This normalization provides an easier way compare a wide variety of influence events.

There are a few important observations about these distributions. First, they are 'long tailed' and most extend several orders of magnitude past the mean value, xave. These long tails are indicative of a small subset of users who are generating large amounts of influence events over the period of time considered. When looking at distributions that exhibit power law behavior it is common to determine the exponent that describes the slope on a log log plot. While we do not attempt to fit the power law in this discussion, it can be noted that each distribution has an exponent of close to -2 for the largest influencers. A power law of -2 in indicative of a highly connected, "small world" network that is being dominated by the social bridges connecting tightly connected communities. Furthermore, comparing the exponents of each type of influence events provides insight into the inequality of influence within that channel. The flatter the slope, the more influence is dominated by the top influencers. The steeper the slope for large values, the more evenly distributed influence is across more users at moderate values.

These distributions can also provide a better understanding of the breadth or engagement in influential actions that have occurred. Twitter mentions extends through many order of magnitude of influencers, from users 100x less influential than average within this channel to users that are over 1000x more influential. The breadth is followed by facebook likes and google plus comments and then foursquare tips done. Linked in comments shows the least breadth, extending from just below the average to about 10x more than the average.

A final interesting feature of these distributions is the existence of a finite peak for facebook likes and to a lesser extent twitter mentions. These peaks, in contrast to a singularly decaying function, is an indication of a characteristic guarantee of influence. In other words, the facebook ecosystem may provide more opportunity for users to generate some amount of influence. This is in contrast to the events that do not exhibit peaks and represents a steadily decreasing likelihood of generating more influence.

Influence Events Within a Social Network

In this section we look at how influential users generate influence through multiple channels within the same network. This work can be important for several reason. First, there is speculation that users can be categorized by their ability to be influential within one channel and the more they generate influence though that channel, the less they generate through other channels. In contrast, there is other speculation that influence is observable through each influential channel and the more influential events a user generates in one channel the more they can represent in another channel. In other words, in this section we explore the correlation between influential events within the same social network

Figure 2 shows the two dimensional density distribution for observable influential events within 3 social networks: facebook, twitter and google+. The axes are logarithmic meaning that increases in orders of magnitude are separated

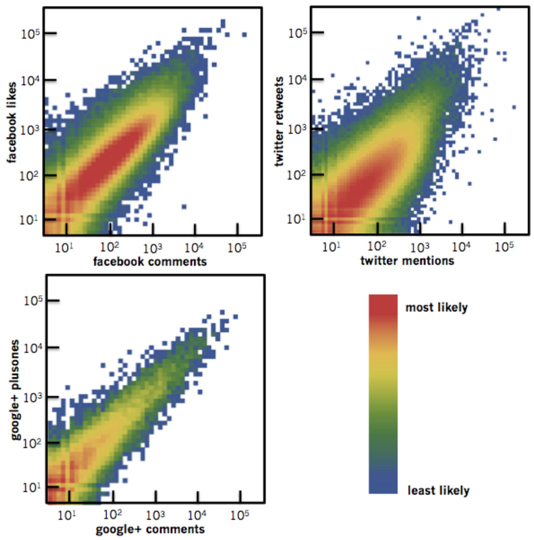

Figure 2: Density distributions for influential channels occurring on the same social network. The correlation between channels is easily seen.

by equal distances. The coloration represents the density or number of users that generated that amount of influence through each channel and it is also logarithmic. Therefore, the red regions represent the areas with the most users.

It is clear to the observer that these events are strongly correlated. As a user generates more influence through one channel on a given network, it appears highly likely that they will generate influence across all possible channels across that social network. The table below represents the correlation between these influence channels. This result should not be taken as conclusive evidence about the correlation of influence within social networks. There may be cases in which influence is not correlated or where it may even be anti correlated. However within these networks where we have access to the data and represents a large portion of social media, the correlation is unmistakeable.

Influence Across Social Neworks

In the last section of our characterization of influence events, we look at how users generate different influential actions across different social networks. Similar to our discussion in the previous section, it is indicative of the state of online influence wether these events are positively or negatively correlated. It characterizes how users spend their time and provides the ability for prediction when one know influence within one social network. Also, keep in mind that this data exists for users who do generate influence events across multiple social networks. In this way, we can also use this information to understand the capacity to influence on a another social network that they currently do not engage with.

Figure 3 shows the density of users that generate influence events on two different social networks. In contrast to the discussion of intra social network behavior, users behavior in these cases is clearly less correlated. For users that demonstrated the ability to drive influential actions on a given network is not a strong indicator of their influence on an additional social network.

In the following table, we show the correlation between the influence events for a selection of pairs.

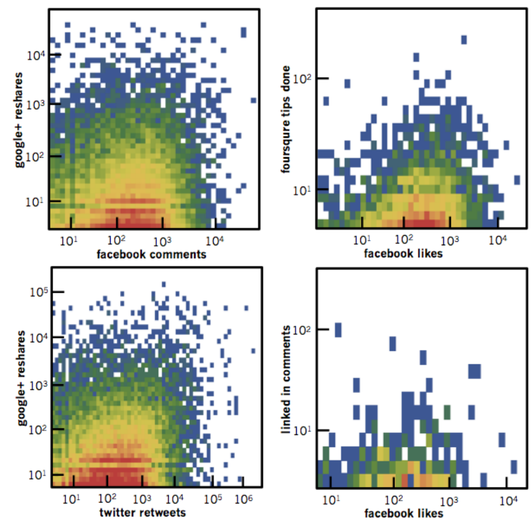

Figure 3: A logarithmic density distribution for influential events occurring for the same user on different social networks. No strong correlation is observable but there is more positive correlation than negative implying similar behavior across social networks by the same user.

Conclusion

Measuring influence is becoming a reality. As more time is spent online and the internet transitions to a continually more social experience, understanding and identifying influence and where it occurs is critical This paper provides a first step in this understanding. In this paper we characterized how engagement is initiated and how influence is achieved across many social media outlets. First we noted that all the influential events considered here have a long tail with the leaders contributing a large fraction of the overall activity. This distribution also implies small world networks with strongly connected clusters strongly connected by bridges. Next we looked at the relationship of different influential channels within the same social network. In this case, the correlation between these behaviors was very high. Users who were more successful driving one type of action within their network were likely to generate action across the other channels. Finally, even when we looks at behavior across social networks we still see a non zero correlation. This implies that measuring the influence of one channel on one network is an indicator of their general influence across the social web. Finally, we explored influence across social networks and noted that although it is not high as inter social network, it is still net positive.

This work serves as baseline for further studies of online influence measures. While the breadth of opportunity for influence to occur is challengin, the relationship between very different types of influence is notable. It implies that models that attempt to predict a user's influence do not need to measure every piece of influence, but can extrapolate from subset of the user's data. As users begin to rely more heavily on navigating social media and their ability to influence it, monitoring these behaviors and being able to predict a user's potential impact will become invaluable.

References

[1] A. Mislove, M. Marcon, K.P. Gummadi, P. Druschel, B. Bhattacharjee, *Measurement and Analysis of Online Social Networks*, In Proceedings of the 7th ACM SIGCOMM conference on Internet measurement (2007), pp. 29-42

[2] Banavar, J.R., A. Maritan, A. Rinaldo, Size and form in efficient transportation networks, Nature, 399, 130-132, 1999

[3] L.A. Barbosa, A. Castro Silva, J. Kamphorst Leal da Silva, *Scaling Relations in Food Webs*, Physical Review E, V 73, 2006

[4] C. Meeyoung, H. Haddadi, F. Benevenuto, K. Gummadi, *Measuring User Influence in Twitter: The Million Follower Fallacy*, International AAAI Conference on Weblogs and Social Media (ICWSM), May 2010

[5] Romero, D. M., Galuba, W., Asur, S., Huberman, B. A.*Influence and Passivity in Social Media*, Proceedings of the 20th International Conference on World Wide Web (WWW) (2011)

[6] D. Saez-Trumper, D. Netleton, R. Baeza-Yates *High Correlation between Incoming and Outgoing Activity as a Distinctive Property of Onine Social Networks* International AAAI Conference on Weblogs and Social Media (ICWSM), May 2011

[7] Z. Zhou, R. Bandari, J. Kong, H. Qian, V. Roychowdhury *Information Resonance on Twitter: Watching Iran*, In Proceedings of the First Workshop on Social Media Analytics (2010), pp. 123-131

[8] R. Kumar, J. Novak, A. Tomkins *Structure and Evolution of Online Social Networks*, Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, 2006

[9] A. Anagnostopoulos, R Kumar, M. Mahdian, *Influence and Correlation in Social Networks*, Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining , 2008

[10] J.C. Gonzallez-Avella, M.G. Cosenza, K.Tucci, *Nonequilibrium transition induced by mass media in a model for social influence*, Physical Review E, V 73, 2006