An activity window model for social interaction structure on Twitter

1 Introduction

The dynamics of information diffusion and interaction in online social networks have attracted increasing attention in recent years [1]. Various studies have shown that the characteristics of human activity have a great effect on information diffusion [2]. In particular, time is a critical factor that deeply influences the dynamics on and of social media because it is decided by the users’ multi-levels activities which are composed of browsing, using and controlling information [3]. All the timestamped activities reflect how users spend time in interaction in online environments [4,5]. However, the temporal process which emerges from the interaction of individuals on websites such as Twitter, Facebook and YouTube needs to be explored in more depth [6].

In the last decade, the access to high resolution datasets from social media platforms has provided opportunities to uncover the structure and dynamics of the social interaction network at different levels, from the small-scale individual’s perspective to the large-scale, collective behavior of the masses [7]. Twitter is in the process of being appropriated for conversational interaction and is starting to be used for collaboration, as well [8]. Thus, the social interactions on Twitter, which is defined as an adaptive network and models the feedback from network structure [9]. Zi Yang et al., explore the behaviors’ timestamp for each user and found that the values of retweet intervals may be influenced by daily routines, which is important to retweeting prediction [10]. K Lerman et al., studied the chain of the URL recommendation for each user and found that the values of re-tweet intervals may be influenced by daily routines, which is important to retweeting prediction [10]. K Lerman et al., studied the chain of the URL recommendation for each user and found that the values of re-tweet intervals may be influenced by daily routines, which is important to retweeting prediction [10].

K Lerman et al., studied the chain of the URL recommendation for each user and found that the values of re-tweet intervals may be influenced by daily routines, which is important to retweeting prediction [10]. K Lerman et al., studied the chain of the URL recommendation for each user and found that the values of re-tweet intervals may be influenced by daily routines, which is important to retweeting prediction [10]. K Lerman et al., studied the chain of the URL recommendation for each user and found that the values of re-tweet intervals may be influenced by daily routines, which is important to retweeting prediction [10]. K Lerman et al., studied the chain of the URL recommendation for each user and found that the values of re-tweet intervals may be influenced by daily routines, which is important to retweeting prediction [10]. K Lerman et al., studied the chain of the URL recommendation for each user and found that the values of re-tweet intervals may be influenced by daily routines, which is important to retweeting prediction [10]. K Lerman et al., studied the chain of the URL recommendation for each user and found that the values of re-tweet intervals may be influenced by daily routines, which is important to retweeting prediction [10]. K Lerman et al., studied the chain of the URL recommendation for each user and found that the values of re-tweet intervals may be influenced by daily routines, which is important to retweeting prediction [10]. K Lerman et al., studied the chain of the URL recommendation for each user and found that the values of re-tweet intervals may be influenced by daily routines, which is important to retweeting prediction [10]. K Lerman et al., studied the chain of the URL recommendation for each user and found that the values of re-tweet intervals may be influenced by daily routines, which is important to retweeting prediction [10]. K Lerman et al., studied the chain of the URL recommendation for each user and found that the values of re-tweet intervals may be influenced by daily routines, which is important to retweeting prediction [10]. K Lerman et al., studied the chain of the URL recommendation for each user and found that the values of re-tweet intervals may be influenced by daily routines, which is important to retweeting prediction [10]. K Lerman et al., studied the chain of the URL recommendation for each user and found that the values of re-tweet intervals may be influenced by daily routines, which is important to retweeting prediction [10]. K Lerman et al., studied the chain of the URL recommendation for each user and found that the values of re-tweet intervals may be influenced by daily routines, which is important to retweeting prediction [10]. K Lerman et al., studied the chain of the URL recommendation for each user and found that the values of re-tweet intervals may be influenced by daily routines, which is important to retweeting prediction [10]. K Lerman et al., studied the chain of the URL recommendation for each user and found that the values of re-tweet intervals may be influenced by daily routines, which is important to retweeting prediction [10]. K Lerman et al., studied the chain of the URL recommendation for each user and found that the values of re-tweet intervals may be influenced by daily routines, which is important to retweeting prediction [10].
interactions and telephone communication [13], there is still a lack of models of human behavior for social interaction on Twitter to understand the communication patterns and how they correlated with the time-varying topological structure of social interaction network.

In addition, individual based models of collective social behavior traditionally include two basic ingredients: the mechanism of interaction and the network of interactions [14]. Thus, we introduced a model of temporal networks that allows the human dynamics and social interactions obtained from extensive Twitter data to be modeled simultaneously. The temporal network is significantly different from static networks in many ways, including methods of representation, models, and spreading dynamics [15,16]. Here we concerned at the temporal graph with nodes having functions, where the emergency of edges are depended on the behavior of nodes [17,18]. It captures the activity dynamics of the nodes and shapes the attention and time constrains in human communication. In general, models of temporal networks can be used to interpret generating social interaction from the collective level. Specially, the power-law degree distribution is founded in some temporal networks which are used to represent contact networks or social interaction networks [19-21]. In these studies, the null model is used to randomly shuffle the event sequences, the intervals of the edges and the number of tweets backwards in his input set.

\[
\text{St}_j, \text{user A writes a new tweet(black)}. \text{User B chooses the red tweet and retweeted it, by looking backwards in his timeline}. \text{Similarly, at time St}_k, \text{user B retweeted the red tweet posted by A}. \text{At time St}_j, \text{user A wrote a new tweet(black)}. \text{User B neglected it but user C retweeted it at time St}_k
\]

Motivated by the above analysis, in this paper we study the users’ communication patterns on a class of temporal network with nodes having an activity window which represents the dynamics of social interaction on Twitter. Then, based on the Stanford dataset, we extracted the statistics of the activity window, analyzed the structure of the social interaction network and found the users’ communication patterns under the activity window model. In addition, we propose a generation model for an interaction network on Twitter. Combined with the null model to the dataset, the results showed that the distribution of the activity window influences the degrees’ distribution of the interaction network.

2 Description of the active-window-based interaction network

2.1 The social interaction network

Twitter is a typical message transfer network. On this kind of communication platform, the relationships and the interaction among users build new channels for information diffusion, as shown in Figure 1, each user A has an input set \( \text{Input}_A = \{ (m_j, t_1), (m_k, t_2), \ldots \} \), inside which the tweets \( m_j \) or comments \( m_k \) are posted by their followees j,k. These tweets or comments arrive in turn and are marked by their arrival time \( t_n \). When A is online, he will browse a limited number of tweets backwards in his input set from \( t_k \) to \( t_{k-l} \), choose \( m_l \) and retweet it to his followers. Thus, the interaction network \( T = \{ (V(A(w)), E(a(A, B, p \rightarrow w_A)), T) \} \) is composed of the set of nodes \( V \), the set of edges \( E \), and the length of time edges \( E \), and the duration of the interaction network \( T = [t_0, t_k] \). For \( \forall A \in V \), \( w \) is named the activity window and means the possible length of his backtracking which cascades from his followees. The activity window is actually a local prediction of the behavior of the user when messages posted by his friends have already appeared on his timeline [10]. For \( \forall e(A, B, p \rightarrow w_A) \in E, A \) is the followee of B. If \( p \) is 1, it represents that the edge is active, which depends on the length of \( w_A \) and whether there are tweets posted by his followees that have fallen into this activity window.

2.2 Extracting the activity window

However, with the limited crawl technology, it is too difficult to get the complete scenario of how users use the Twitter App. Therefore we used a machine-learning algorithm to estimate the activity window. Twitter user A has a sequence of activities of tweeting and retweeting. We define it as \( \text{Seq}_A = \{ (t_{wA}, St_1), (RT_j, t_2, St_2), (CM_j, t_3, St_3), \ldots \} \), which makes it explicit that he is online. For example,
(tw_4, St_4) means that A wrote a new tweet at time St_4.
(RT_j, t_2, St_2) means that at time St_2, A retweeted a message of j which arrived at time t_2. CM_j, t_3, St_3 means that at time St_3, A commented on a message of j which arrived at time t_3. Thus, intervals like [St_4, t_4] mark the time difference between the time of j’s message arrival at his input A and the time of A retweeting or commenting. With the EM-GMM algorithm, we can find out two types of time difference sets and their probability. One is short-intervals SI_A with prior probability p^short, which means regular online behaviour (e.g., browse, RT or tweet). The other is long-intervals LI_A with prior probability p^long, which are the result of daily routines, such as meeting time or sleeping. According to [10], both the long-intervals and short-intervals contribute to the discovered patterns of regularity so we mix them together to estimate W_A and therefore W_A is defined as W_A = SI_A * p^short + LI_A * p^long. Thus, the activity window represents a user’s ability or habit of using social media.

3 Results

We analyzed the records of 550,000 users in one month, consisting of 207 million tweets and 124 million followee-follower relationships [29]. The topology of the followee-follower network has small-world properties with L=4.36, <C>=0.01 and M=0.12. In addition, we built the interaction network according to the “RT” or “@” tagged in the dataset and measure the degree distribution of it. We collected the users’ activity sequence mentioned in Section 2.2, estimated the activity windows by EM-GMM algorithm and measure the communication patterns based on the activity window. Combined with the activity-driven model, we examined the effect of the activity window on the interaction network.

3.1 Structure of the social interaction network

The social interaction network is constructed by users’ activities, such as “RT” and “@”. Here we represented these temporal data as a static graph and found that it was much sparser and more highly-clustered than the corresponding “followee-follower” network. Although there are only 70,994 nodes and 216,582 edges in it, the cluster coefficient and the modularity respectively increased to 0.13 and 0.621.

We divided the interaction network into four aggregated networks by weeks and observed their degree distributions. Figure 2(a) represents four fragment interaction networks, in which the length of time is one week. Obviously, they have the same degree distribution, which complies with power-law distribution and the slope is about -1.45 ± 0.02 (R^2=0.95). Panel (b) The power-law distribution of degree k_i of aggregated interaction networks, while the slope increases as time goes by. The valuations of the slope are -1.45, -1.31, -1.26, -1.19 (R^2=0.95). For TG_1 → _a, <K>=6.10, <C>=0.13, M=0.621, <L>=10.32.

3.2 Features of the activity window

Based on the Stanford Twitter dataset, we analyzed the statistical characteristics of the activity windows. From Figure 2(a), we can see a power-law distribution with exponential cut-off for W. This distribution is different from other online human activities, such as email activity, online games, or web page browsing, which have exponent α < 2 [16, 21]. The reason is that the distribution of W complies with power-law distribution with slope -0.86, when it is smaller than 45 minutes. Meanwhile there is exponential distribution with exponent -0.02, when W is big.
大于45。它表明大多数用户的活动窗口范围从5到30分钟。这意味着目的时，人是接受消息的可能。在有限的时间内。活动模式将突发特性可能由刺激相的交互导致。其他人将浏览45分钟后，他们的浏览行为是无记忆的，可能被长度影响的输入集。

3.3 Patterns of communication

除了用户的活动，增加的关注是集中在关系社交的网络度和网络的密度。例如，Miritello等人已经研究了这一点：在与Dunbar的理论[12]一致。我们观察到互动网络由用户的行为，包括RT和@。在这里，我们观察到一个更复杂的活动模式。例如，当k_i的值较大时，我们测量到的互动稳定性将下降。因此，我们观察到的互动稳定性不仅增加了窗口的大小，而且取决于两个用户参与的社会关系的度。

3.4 Effect of the activity window on the interaction network

如上所述，活动窗口形状的通信模式的用户在Twitter上的暂态在线活动和结果的时间动态的互动网络。这可能是因为不同的活动窗口大小将带入时间序列，延迟，和在异步大规模联系的Twitter。因此，我们观察到的互动稳定度将下降与活动窗口的大小。

这表明更高的稳定互动将发生在那些有更有效的对象和更小的后退长度的窗口中。
An activity window model for social interaction structure on Twitter

Figure 4: The interaction burstiness (denoted as $c_{ij}$) of each edge $e_{ij}$ in the interaction network, where $i$ and $j$ are the source and sink vertexes of the directed edge respectively. Thus we plot all the edges in a 3-dimensional parameter space. $k_i$ is the degree of vertex $i$, $k_j$ is the degree of vertex $j$, and $w_i$ denotes the size of $i$’s active windows. The value of interaction burstiness fluctuates from 0 to 2.3. It indicates that the interaction burstiness $c_{ij}$ increases with the degrees of vertex. In addition the larger the activity window the larger the range of the interaction burstiness.

Figure 5: The interaction stability (denoted as $\Delta_{ij}$) of each edge $e_{ij}$ in the interaction network, where $i$ and $j$ are the source and sink vertexes of the directed edge respectively. Thus we plot all the edges in a 3-dimensional parameter space. $k_i$ is the degree of vertex $i$, $k_j$ is the degree of vertex $j$, and $w_i$ denotes the size of $i$’s active windows. When the degrees of vertexes in an edge are more than 10, most of the interaction stability $\Delta_{ij}$ increase with the size of the activity window, fluctuating from 1 to 6 hours. Otherwise, the $\Delta_{ij}$ is smaller than 1.5 hours.

count, the dynamics of information spreading will have significant difference with the SI, SIR, or SIS models [25]. In this study, we explored the effect of the activity window on the generation of an interaction network with the sampling dataset which retains the relationships and the timelines for users. Based on the dataset we combine the activity-driven model [26] with the activity window to simulate the dynamics of the interaction network with time-varying features. The model is named ADAW (Activity-driven model with activity window) and is as follows:

- Load the follower-followee network, the Input sets and the activity window $p(w)$ complying to power-law distribution with exponential cut off. The slope of the power-law distribution is -0.86 and the rate of the exponential distribution is -0.02. Choose 514 nodes as seeds (0.1% of all nodes) to create new tweets and push them to their friends.

- At each discrete time step $t$ the instant interaction network $G_t$ starts with $N$ nodes.

- With probability $a_i$ each node becomes active, where $a_i = \frac{activity_i}{activity}$, and $activity_i$ is the number of activities of node $i$ in one month, which is determined by the empirical data. If node $i$ is active, he can take one of the following behaviors: innovating a new tweet with the probability $p$ or going backward to the timeline with the probability 1-p, the scale of which is equal to $W_i$, and randomly choosing $m$ tweets to retweet or comment on. Then $m$ links of the interaction network are generated.

- At the next time step $t+1$, all edges in $G_t$ are deleted.

To gain insight into the effects activity widows, we employed a null model [15] in the ADAW model. The features in the activity windows in the following null models are separately randomized. The null models are as follows:

W-Random: The whole community structure in the follower-followee network and user timeline are retained. The activity windows distribution is $p(w) = U[5, 231]$. The temporal correlation consisting of online time length between the follower-followee relationships is destroyed.

W-U: The whole community structure in the follower-followee network and user timeline is retained. The activity windows distribution is $p(w) = U[5, 231]$. The temporal correlation consisting of the online time length between the follower-followee relationships is destroyed.

For each simulation we recorded the instant interaction network $G_t$ and accumulated them into aggregated networks when the simulation ended. Figure 6(a) presents the probability distribution of degree for each aggregated network, which are all power-law distributions but with
different slopes. The statistic of degree distribution of the interaction network generated by ADAW is the most similar to that of the Stanford dataset. In 200 repeated simulations of the ADAW models, there are 186 aggregated networks whose degree distributions are tested to be a power-law, and the slope is about -1.45 ± 0.02. Yet in the results of the W-Random and W-U models there are only 134 and 162 aggregated networks that have power-law degree distribution, and the value of the slope is about -1.28 ± 0.02 or -1.32 ± 0.02.

While switching the activity window size will bring more deviation to the aggregated network in the two null models, it has been proven that the activity window is a temporal factor of the dynamic mechanisms for generating the interaction networks on Twitter. Figure 6(b) shows the cumulative proportions of users participating in the interaction in one month. When the correlation between the degree and activity window is shuffled randomly, the cumulative proportions are bigger than those of ADAW and W-U. The result proves that both topological and temporal correlations slow down the spreading. Furthermore, W-U has a higher cumulative number of the nodes than those of ADAW in the first 5 days, and then it reverses. We interpret this phenomenon by a phased analysis of the simulation. At the beginning of the simulation of the ADAW model, most users have a small activity window and bigger $\Delta_{ij}$, which slows the growth of the number of participants. Continuing with the simulation iterator, more and more users appear, who have bigger activity windows, lower $\Delta_{ij}$, and $cv_{ij}$ scattering around 1. The probability of interaction is therefore increased, and the number of participants exceeds that of the W-U model.

### 4 Conclusion

Online social network systems can be extremely dynamic and are generated almost instantaneously. Understanding their dynamical properties is not only of fundamental interest, but also has a broad range of applications [27]. One contribution of this study is to propose a temporal network model with nodes having an activity window, which describes the dynamics of the interaction network on Twitter. The activity window depicts users’ fragmental participation on Twitter, which complies with power-law distribution with an exponential cut off, acquired by way of the EM algorithm. Meanwhile, we have concentrated on how the activity window affects the communication pattern of the online interaction. The results show that the increase in the degrees of the social ties leads to a dramatic drop in the unevenness of the interaction stability between users. However, the interaction burstiness is related to the degrees in the interaction network. The higher the degrees, the higher the burstiness.

The other contribution of this study was developing the ADAW model by combining the activity window with the activity-driven model. It proved that the activity window affects the generation of the interaction network. In addition, we introduced null models into ADAW where the activity window characteristics of the nodes are randomly shuffled [28]. Thus we were able to distinguish the effects of the activity window having different features, including power-law distribution with exponential cut off, random distribution and uniform distribution. The results of the ADAW model are the closest to that of the real dataset because the time characteristics of individuals’ activity will slow down the speed of spreading in a small-world network [2, 23]. Comparing the simulation result of ADAW with the other two distributions, we also found that ADAW reduces the interaction scale but is bigger than the uniform distribution.

Our results also provide new insights for the description of temporal variation of the online social networks. In particular, the ADSW model supplements the way of de-
An activity window model for social interaction structure on Twitter

Picturing information diffusion and can be extended further by including other social processes. Hence our future work will explore how the complex contagion mechanism works on it.

Conflicts of Interest
The authors declare that there is no conflict of interest regarding the publication of this paper.

Funding Statement
This work was supported in part by the National Natural Science Foundation (grant numbers 71401024 and 71801145).

Acknowledgment
JZ is grateful to Professor Ying-Cheng Lai and his students of Arizona State University for the helpful discussions, insights, and the support of the experimental equipment. The authors would also like to thank the anonymous referees for their careful work and thoughtful suggestions that have helped improve this paper substantially.

References