Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media

Link and Triadic Closure Delay: Temporal Metrics for Social Network Dynamics

Matteo Zignani*, Sabrina Gaito*, Gian Paolo Rossi*, Xiaohan Zhao†, Haitao Zheng†, Ben Y. Zhao†
*Department of Computer Science, Universitá degli Studi di Milano, Italy
†Department of Computer Science, UC Santa Barbara, USA

Abstract

Today, numerous models and metrics are available to capture and characterize static properties of online social networks. When it comes to understanding their dynamics and evolution, however, research offers little in terms of metrics or models. Current metrics are limited to logical time clocks, and unable to capture interactions with external factors that rely on physical time clocks. In this paper, our goal is to take initial steps towards building a set of metrics for characterizing social network dynamics based on physical time. We focus our attention on two metrics that capture the "eagerness" of users in building social structure. More specifically, we propose metrics of *link delay* and *triadic closure delay*, two metrics that capture the time delay between when a link or triadic closure is possible, and when they actually instantiate in the trace. Considered over time or across traces, the value of these metrics can provide insight on the speed at which users act in building and extending their social neighborhoods. We apply these metrics to two real traces of social network dynamics from the Renren and Facebook networks. We show that these metrics are generally consistent across networks, but their differences reveal interesting properties of each system. We argue that they can be attributed to factors such as network maturity, environmental and social contexts, and services offered by network provider, all factors independent of the network topology and captured by our proposed metrics. Finally, we find that triadic closure delays capture the ease of neighbor discovery in social networks, and can be strongly influenced by friend recommendation systems.

1 Introduction

Online social networks (OSNs) have been extensively studied in the last decade, with most efforts focusing on static properties computed on single snapshots of the network. Gradually, attention of the community has shifted towards temporal properties of OSNs, with the goals of understanding patterns and mechanisms underlying their growth. However, most temporal studies are limited to analyses of dynamics using logical clocks (Leskovec et al. 2008), despite the growing recognition that a full understanding of graph dynamics requires analyses using physical clocks (Kossinets, Kleinberg, and Watts 2008).

Recent studies of network dynamics on the Renren network provided a macroscopic view of network evolution, by measuring the trend of some measures in each temporal snapshot (Wilson et al. 2009; 2012; Zhao et al. 2012). Later work (Gaito et al. 2012) took it a step further, by adopting a microscopic approach using physical time. Its node-centric analysis showed that users create links in a highly bursty temporal pattern.

In this work, we focus on the dynamic formation of two fundamental network-building components: *dyads* and *triads*. We propose two new metrics to aid the temporal analysis on physical clocks: link creation delay and triangle closure delay. *Link delay* is the time period required before two users of a network create a link between them, *i.e.* the delay between when a friendship is possible and when a friendship link actually forms. On the other hand, triangle closings capture the transitivity of friendships, which has proven to be effective in modeling network evolution and link prediction. *Triadic closure delay* captures the time of formation of all the links of triads by considering the temporal process of triad formation.

These two metrics enable us to study the dynamic creation of dyads and triads, and to highlight network behavior that would otherwise remain hidden. The main contribution of this paper concerns the temporal features characterizing the links and the triangle formation measured on two temporal annotated datasets from Facebook and Renren (Section 3). In our analysis, we find that link delays are generally very low in absolute time, meaning if two people want to become OSN friends, they do so very shortly after both have joined the network. Link establishment is especially fast in early stages of social networks. In addition, link delay results are largely independent of the dates people join the network. To highlight the social nature of this metric, we introduce the term *synchrony* to quantify how well linked users overlap in lifetimes. Finally, we study if link delays correlate to distance in the social network, and find that links spanning more distant nodes generally form faster.

Our *triadic closure delay* takes into account how long a temporal triangle takes to form. We first introduce an algorithm to extract temporal triangles, which enable us to monitor the triangle formation process, and to detect sudden changes in the triangle formation behavior, possibly related to external events. In particular, we show that the introduction of Facebook's 'People You May Know' (PYMK) functionality had a disruptive impact on the triangle creation process in the network, while in Renren we highlight the impact of real life changes on the network structure.

From a microscopic perspective, we shed a light into the physical time of the triadic closure process by introducing a formal definition of the delay of the triangle formation. The triangle delay represents a normalization that accounts for the node and link arrival processes. By analyzing the above quantity we find that triad formation is very fast, accounting for the fact that if two persons have a common friend and want to be online friends, they instantiate the relationship very quickly. Yet this new metric shows slightly different behaviors in our two datasets. In fact the bootstrap phase captured by the Renren dataset is also faster in the triangle formation dynamics compared to Facebook growth. Finally, triangle closure delay allows us to identify when latent triangles have been triggered by external events, thus allowing us to evaluate the impact of these types of events.

2 Related Work

With the availability of datasets on online social networks that evolve over time, researchers have begun to study mechanisms by which nodes arrive and links form or disappear. One promising approach has been the microscopic perspective, *i.e.* the growth and the properties of a network result from choices made by nodes based on local information from their neighborhood or other close-by nodes. However, microscopic mechanisms have been studied only using logical time, where the number of nodes/links represent the 'clock' of the system. This has led to the study of link formation in terms of likelihood or in term of age of nodes (Leskovec et al. 2008). Only (Sun et al. 2012) discussed the issue of predicting when a link will be built. Recently, online social networks have been studied from a physical time perspective, with particular focus on bursty behavior in the neighborhood expansion mechanism (Kikas, Dumas, and Karsai 2013; Gaito et al. 2012). However, the temporal properties of each link have never been faced in the physical time framework.

Among other proposed link creation mechanisms, triadic closure is the most basic and powerful principle to model the evolution of social networks. Specifically, this principle states that individuals with a common friend have a higher chance to become friends themselves at some point in the future (Rapoport 1953). Due to the strong relationship between triadic closure and clustering coefficient, a few works have faced the dynamics of the triadic closure process, by analyzing the temporal trends of the average clustering coefficient. By exploring the social network dynamics of a portion of the arXiv repository, (Amblard et al. 2011) showed that the average clustering coefficient is quite constant (≈ 0.5 since 1992). A more detailed analysis of the average clustering coefficient in a Chinese social network has been presented by (Zhao et al. 2012). Variations in the clustering coefficient have also been observed by (Gonzalez et al. 2013) and (Gong et al. 2012). They show that Google+ structure has become less clustered as new users joined the largest connected component, and that the average clustering coefficient seems to follow a three-phase evolution pattern: first decreasing, then increasing slowly and finally decreasing again. In general, the previous approaches all suffer from the averaging effect, *i.e.* wide fluctuations affecting the clustering coefficient of single nodes are lost in the average. This produces a measure of the whole network that poorly captures the triadic closure process.

While these studies focus on the temporal trend of the clustering coefficient to quantify the magnitude of the closure process, others combine the snapshot paradigm with the likelihood of a link given the number of common friends. One of the seminal work has been carried on by (Kossinets and Watts 2006). They found that two users are more likely to close a triangle if they share many friends. While the method does not explicitly focus on the process timing, it heavily depends on the choice of the temporal gap between two snapshots. A more detailed and wider study of the triadic closure process has been conducted by (Leskovec et al. 2008). They model how a source node decides to add an edge to some other node two hops away. They have found the most likely mechanism is given by a combination of the number of common friends between the nodes and the presence of recent activity at the candidate neighbor. While this work is the most related to ours, it focuses on the choice mechanism and not the temporal trend of triangles or the time it takes for them to form.

(Mislove et al. 2008), (Viswanath et al. 2009), (Garg et al. 2009) have studied triadic closure in the evolution of online social networks. In particular they analyzed the proximity bias, *i.e.* the tendency of nodes to link with those nearby in the network graph. The results on Flickr and FriendFeed have shown that proximity bias influences the formation of new links, making two nodes which are two hops distant more likely to form a link. These studies only confirm the predominant role of the triadic closure, but do not explore the closure process. Finally, (Romero and Kleinberg 2010) studied the triadic closure process on the Twitter network, finding heterogeneities in the process when they consider high degree nodes.

In contrast, we investigate the temporal dynamic properties of the triadic closure process and link formation. To the best of our knowledge, this is the first study that explores how long the triangle formation lasts in triadic closures. Given our use of physical time, we are able to correlate changes in the triadic closure process with social and environmental contexts, as well as the introduction of new services by the social network provider.

3 Measurement Methodology

The main goal of our work is to introduce metrics that can quantify the microscopic dynamics in the growth of online social networks, using physical time as reference system. One advantage that physical time offers is the possibility of relating the global and local changes in the growth process to events external to the network system, since physical time connects events inside the social network structure and any external events.

To introduce the new metrics, we start by presenting the theoretical framework we adopted to describe the growth processes in physical time. Then we introduce the datasets on which we apply the dynamic metrics. In particular, we need data from online social networks with fine-grained temporal information, so as to stress the microscopic dynamics of their evolution. Using artifacts in our datasets, *i.e.* the bootstrap phase in Renren and the Facebook transition from a university service to a business company, we highlight and measure how different types of external events impact on the network topology and its dynamics.

Notation

In the following, we introduce the notation we adopt to describe the growth of the whole network and its constitutive elements. Formally, we represent the social network of our datasets as an evolving undirected graph. Usually the network growth is represented by a sequence $< G_1, \ldots, G_T >$ where each $G_t = (V_t, E_t)$ is an undirected graph denoting the state of the network at time t having $|V_t|$ nodes and $|E_t|$ links. As we have no information about node and edge removals, the number of nodes and edges always increases in time up to the end of the measurement process indicated by T. At last, graph $G_T = (V_T, E_T)$ will contain the whole set of nodes and edges appearing during the growth.

To analyze the microscopic structural properties and to simplify definitions, we project the sequence $< G_1, \ldots, G_T >$ into an undirected graph $G = (V, E)$ where links are time-stamped by the *time function* $\tau : E \to \mathbb{R}$ that assigns to each edge $e = (u, w)$ its creation time. Given the increasing monotonicity of the sequence, $V = V_T$ and $E = E_T$. In accordance with this framework, we use $\Gamma_t(u) = \{w|\tau(u, w) \leq t)\}$ to denote the set of neighbors of u at time t and consequently the degree of node u at time t, as $k_t(u) = |\Gamma_t(u)|$, while its final degree is $k_T(u)$. Since we only have temporal information about edges, we define the time of the first appearance of node u into the network $b(u)$. We call this its *birth date* assuming that $b(u) = \min(\tau(u, w)|w \in \Gamma_T(u))$, *i.e.* the time of the first link incident to u.

Finally, since the main subject of Section 5 is the dynamics of transitivity, we denote a triangle composed by the vertices u,w and z as $u\hat{w}z$. Exploiting the time function we can assume that for each triangle, $\tau(u, w) < \tau(w, z) < \tau(z, u)$ holds. Consequently, we have an ordering of the edges in a triangle and we lose the triangle isomorphism typical of static undirected graphs, *i.e.* $u\hat{w}z$ is not equivalent to $w\hat{z}u$.

Dataset

The main obstacle to study dynamics is the challenge of obtaining detailed data describing OSN dynamics. In this study, we study two online social networks, Renren and Facebook. First, we received from Renren access to an anonymized dataset that contains the timestamped creation of all users and edges in the Renren network (Jiang et al. 2010). Renren is the largest online social network in China. Like Facebook, Renren was originally designed for college students. Renren was original named Xiaonei (*i.e.*, in school), and changed its name to Renren (i.e. everyone) when it expanded beyond universities. The anonymized dataset describes the growth of each node as a timestamped edge in the network. The first edge created in the entire Renren history dates back to November 22, 2005, and the dataset includes the complete evolution of the first 600,000 nodes for a total of 8 million timed edges events. In total, our dataset covers the first year of Renren, from November 2005 to December 2006.

The second dataset is publicly available[1] (Viswanath et al. 2009). It includes the growth of the New Orleans Facebook network with about 60,000 nodes and 800,000 links from September 2006 to January 2009. While we have evidence of the time creation of all edges in Renren because this dataset covers the network dynamics from the kick-off date, the Facebook dataset contains the timestamped creation of all users and edges, except for 4.2% of vertices and 6.0% of links. These were not considered in our analysis. As this phenomenon is very limited, we believe the results we obtained are applicable to the entire New Orleans network.

4 Link delay

In this section we introduce *link delay*, a novel indirect measure of the eagerness of a tie, measuring the elapsed time between the potential establishment of a link and its real creation. A link is possible when all the enabling conditions are set but the link has not yet been created. Below, we first define the metric capturing the time spent to establish a link between two nodes, *i.e.* the delay of that link. Then we evaluate the link delay properties on Renren and Facebook. We find that link delay is very low, meaning if two users wish to establish a friendship relation, they create the link very quickly once both users join. Finally, we correlate link delay with the "ages" of the connected nodes (how long they have been in the network) and with their topological proximity. The results confirm that the delay is independent of age difference of the nodes, and point out that few fast links have been established between topologically distant users.

Definition

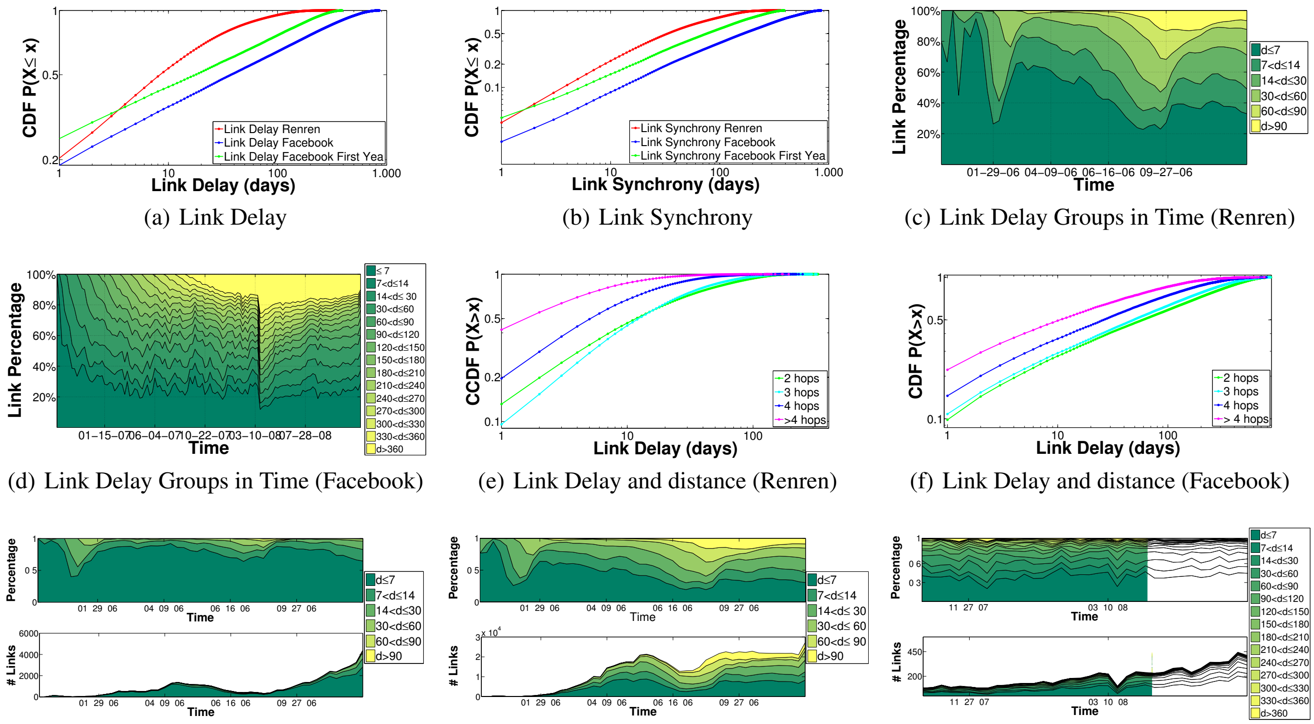
We assume that nodes are free to enter the network anytime during the network lifetime. To properly mirror this assumption in the link delay definition we apply a simple normalization on the values returned from the time function τ including the birth date b of a node. This leads us to define *link delay* $d(u, w)$ as follows:

Given $G = (V, E)$ and its time function τ, the delay $d : E \to \mathbb{R}$ of the link (u, w) is defined as

$$d(u, w) = \tau(u, w) - \max(b(u), b(w)) \quad (1)$$

where the max function on the birth date implies that both nodes need to be created in the graph. $d(u, w)$ measures the elapsed time between the potential link creation time (when all conditions hold) and the actual link creation time. The lower the delay, the faster the two nodes actualize the potential link.

[1] http://socialnetworks.mpi-sws.org



(a) Link Delay (b) Link Synchrony (c) Link Delay Groups in Time (Renren)

(d) Link Delay Groups in Time (Facebook) (e) Link Delay and distance (Renren) (f) Link Delay and distance (Facebook)

(g) Link Delay and distance $>$ 4 hops (Renren) (h) Link Delay and distance 2 hops (Renren) (i) Link Delay and distance $>$ 4 hops (Facebook)

Figure 1: 1(a) and 1(b): CDFs of link delay and link synchrony measured on Renren, Facebook First Year and Facebook. Figures have different y-axis scales: 1(a) starts from 0.2 while 1(b) starts from 0.1. 1(c) and 1(d): The trends of links organized by link delay during the growth of Renren and Facebook. Color intensity is inversely proportional to link delay (upper and lower bound of the groups are in days). 1(e) and 1(f): CDF of the link delay for links grouped by hop distance. 1(g), 1(h) and 1(i): relative and absolute volume of links grouped by link delay, considering links spanning more than 4 hops (g) and 2 hops (h) in Renren, and links spanning more than 4 hops (i) in Facebook New Orleans.

Link delay analysis

Link delay analysis can shed light on a few properties of OSN friendship: i) how much time people take to become friends once they both join the OSN; ii) how link delay differs between different networks as a function of their different user populations and the stage of their evolution captured by our datasets. We analyze dynamics of a network at its infancy (Renren), and the first year of Facebook dataset where Facebook becomes an open service, and its consolidation period (overall Facebook dataset).

First, we compute and analyze the distribution functions of link delay in both datasets. Figure 1(a) shows the Cumulative Distribution Function (CDF) of the link delay extracted from the Renren dataset (red line), from the first year of Facebook dataset (green line) and from the entire Facebook dataset (blue). The link delay distribution always shows a very quick shift from potential to actual link state. In fact, for all three distributions, links were created within a day in $20-27\%$ of the cases, and within a week in $32-50\%$ of the instances. These results highlight that generally speaking, a pair of nodes create friendship links soon after both users have joined the network.

Beyond this property, some differences can be noted among the two networks and different time periods. First, a comparison between Renren and Facebook shows that the two distributions diverge right after the first week, when Renren users begin forming friendships more quickly than their Facebook counterparts. The same trend remains true if we consider the first year of Facebook. Secondly, the comparison between Facebook and its first year strengthens the existence of different behavior in link delay as the network grows, *i.e.* links are formed faster in early stages of the network.

A possible explanation of the slight differences between Renren and Facebook lies in the different phases captured by the datasets. In the bootstrap of the network (Renren) the information about the presence of possible friends' nodes spreads faster than in a later period (Facebook) because of the social environment where Renren initially operated: namely, a university campus where Renren was adopted as a social service. Information about a friend's network presence might have spread not only on the "network," but also via other forms of offline networks (classroom, dormitory, etc.), accelerating the adoption of the service. This effect, although smoother because its coverage of users over a larger geographical area, remains visible in the first year of Facebook. The same phenomenon may underlie the different behaviors during the different Facebook macroscopic snapshots; in a network's early phase the information about the establishment of new friendship relation reaches nodes faster, while it has slowed down significantly for a network in its consolidation phase.

We can obtain more information by taking into account when links are created. This way we obtain the temporal trend of the link delay groups as shown in Figures 1(c) and 1(d). We group all links into "delay groups" based on their link delays. We compute, for each week, the percentage of new links belonging to the different groups. Obviously the number of groups depends on the dataset, because they cover different time periods. Renren is characterized by three different stages. From its birthdate to the end of January, *i.e.* the first three months, most links are established suddenly (with low link delay). At the beginning of February, a drastic change in link delay occurs, because users tend to establish unexpressed links, and the volume of quick links increases abruptly. During the summer the delayed links increase in number and quick links decrease, before a reversal after the summer. The resulting phenomenon is characterized by a certain degree of fluctuation in the ratio of link delay.

In order to quantify the instability, we measure the standard deviation of the group's time-series. The further the standard deviation is from 0, the more the time-series is dispersed and fluctuates in time. The computed values indicate that $[7-14]$ and the $[14-30]$ days delay groups are the most stable ($\sigma = 3\%$) and contain on average 28% of the overall created links, while the quickest link and most delayed groups are the most unstable. In addition, the relative time-series seem to be inversely correlated: when the portion of quick link increases, the most delayed diminishes and vice versa.

We observe different results in our Facebook dataset. The component of links having a delay within a week represents 32% of overall links, and is characterized by a quite constant trend. In fact, the dark green group in Figure 1(d) is always between 20% and 40%. A more evident phenomenon happens on March 26, 2008, where we observe a drastic increase in the volume of delayed links. This date corresponds to the introduction of Facebook's "People You May Know" (PYMK) functionality. By analyzing the delay, we can highlight *i)* how this features acts and *ii*) how long its effect lasts.

The friend recommendation system highly amplifies the tendency of establishing "old potential" links that could have been created a long time ago. 60% of links created in the week of the PYMK introduction have a delay greater than 6 months, and 20% had a delay greater than a year. Observing the group trends the weeks after PYMK introduction, we note that the initial behavior in preferring delayed links disappears, and after the summer of 2008 reaches pre-PYMK percentages. Although the link delay reveals interesting characteristic in edge creation process, it is not able to capture the reason behind it, *i.e.* which process causes the observed effects or which algorithms were active in the early rollout of the PYMK feature. In Section 5 we explore these effects correlating them to the triadic closure process.

Finally we apply the stability time-series analysis as we did for Renren, excluding the 10 weeks after March 26 to reduce the PYMK impact. The analysis of the standard deviation shows that fluctuated and stable components simultaneously act during the Facebook New Orleans growth: 1) unstable quick links and 2) stable more delayed links.

Link speed and link synchrony

The link delay observations provided so far are independent of the reciprocal network age of the nodes involved in links. However, the birth date b allows us to verify whether or not link creation favors pairs of nodes with similar network ages. To quantify this type of behavior, we introduce the notion of *synchrony* of a link:

Given $G = (V, E)$ and the birth date function b, synchrony $s : E \to \mathbb{R}$ of a link (u, w) is defined as

$$s(u, w) = |b(u) - b(w)| \tag{2}$$

In Figure 1(b), we show the metric of synchrony obtained from our datasets (Renren: red and Facebook: blue). In both cases we observe that the probability of having a link between peer nodes is low, and nearly 50% of links are established between nodes with different creation times: one month apart in Renren and 5 months in Facebook. We argue that this result is the direct consequence of previously observed bursty behaviors in the edge creation process (Kikas, Dumas, and Karsai 2013; Gaito et al. 2012). In fact, old nodes continue to generate or receive (as we consider an undirected graph) links even if they are aging because 20% of links are created 9 months or one year after a node's birthday respectively.

By comparing link delay and synchrony distribution, as shown in Figure 1(b), we can derive that low link delay has an higher influence on the link population than low link synchrony. In fact, the impact of low synchrony ($\leq$ 7 days) is at most 10%, and lower than the contribution given by low delay links (at most 50%). This proposes the link delay as a relevance feature in the future growth of the networks and on the underlying processes.

Link delay and edge locality

We relate two kinds of locality of the edge: topological and temporal. Topological locality of a link (u, w) is measured by the number of hops h it spans, i.e. the length of the shortest path from u to w removing the edge (u, w). Temporal proximity is given by the link delay. In Figure 1(e) and Figure 1(f) we study the distribution of the link delay in four groups of geodesic distance (2 hops, 3 hops, 4 hops and >4 hops[2]).

For Renren we observe that the distribution for the $>$ 4-hops group lies above the other groups, indicating that edges that connect nodes more than 4 hops distant establish earlier than closer nodes. The same behavior, to a lesser extent, involves also the 4-hops group. This fact characterizes not

[2]For computational constraints distances have been computed by a truncated version of the shortest path Dijkstra's algorithm that terminates 4 hops far the node, so >4 hops group could contain edges connecting different connected components.

only the early bootstrap phase of the network, where the service has a small group of subscribers but remain constant during the network growth, as shown in Figure 1(g). It reports both the absolute number and the percentage of links spanning more than 4 hops divided in the same delay groups adopted in the temporal analysis. We note that more than 80% of the far links keep actualizing in less than 2 weeks. Figure 1(h) shows a different phenomenon occurs in the 2-hops group. Here we observe that some friendships were created with an delay greater than 3 months, even though the two users were already connected by at least one common friend.

In Facebook we obtain a trend similar to Renren, but smoother. As shown in Figure 1(f), edges spanning four or more hops exhibit a lower delay compared to those that span closer nodes. Also in this case, the lowest components (≤ 7 and $(7-14]$) score the $40-50\%$ of links along the period indicating a stable behavior, as shown in Figure 1(i).

In general these results suggests a quite surprising behavior involving link delay and edge locality. While some works (Leskovec et al. 2008), (Easley and Kleinberg 2010) report that closer nodes are more likely to establish a new link, our results suggest that a high link likelihood does not always corresponds to fast link creation. Links that span farther nodes generally instantiate faster. In such cases, the network connectivity is less important than offline friendships. To go deeper, we explore the degree and the age of the nodes forming far and close links. We find that on average, nodes further away in the network are the youngest and are characterized by low degrees. That strengths the observation that nodes just entering into the network exploit some external relationships (offline or other social networks) during link creation.

5 The Triadic Closure Process

The availability of temporal annotated networks has allowed the study of the network's evolution over time, and has led to a deep understanding of the mechanisms governing node and link arrival and creation. In this section, we study the evolution mechanisms of online social networks, mainly focusing on the basic growth principle underpinning these networks: *triadic closure*.

Observed as one of the most frequent processes of link formation, triadic closure has been widely adopted in different disciplines. For instance, the sociological principle for triadic closure is the transitivity of friendship, which says that two individuals have a high likelihood of establishing a friendship if they share a common friend. The transitivity of friendship has been proved to be effective in modeling network evolution and predicting future link formation. Despite its commonly accepted value, the foundational principles governing triadic closure have not yet been analyzed in depth.

Here we characterize the triadic closure process by delving into its temporal aspects. We consider two perspectives. First we analyze the triangle formation growth from the network point of view, by counting new formed triangles on the overall dataset. To reach this goal we adopt an algorithm able to extract temporal annotated triangles. Temporal information are used to monitor the number of triangles day-by-day. By analyzing the triangle time-series we are able to map sudden changes of the triangle formation onto events external to the network, such as the introduction of a new feature in the service or seasonal events involving most of nodes.

We move apart from the network perspective to embrace a microscopic point of view focusing on the formation of the single triangles. In particular we study the speed of the formation of triangles. Our goal is to shed a light on the dynamical properties of the triadic closure process by introducing the formalism to capture the time a triangle takes to be established. This way we define the triadic closure delay and we show how this new metric captures different behaviors in the datasets under investigation.

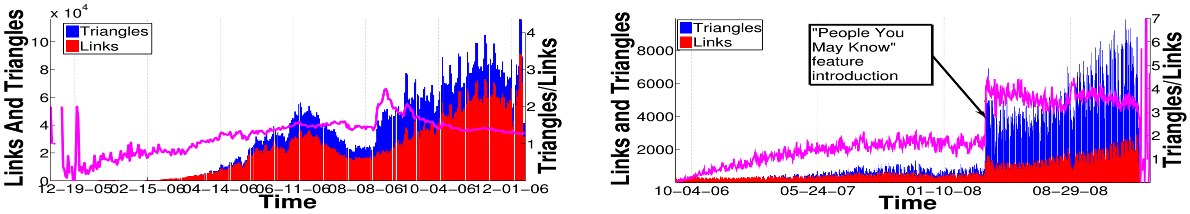
Temporal triadic closure

We believe that the dynamical analysis of network evolution cannot disregard the transitivity closure process, for the literature has shown its importance in the formation of social networks - despite the fact that a temporal analysis of the triadic closure poses both algorithmic and methodology issues. The first concerns the extraction and counting of temporal annotated triads. While many approaches have been proposed in literature, most are suitable for static networks and so cannot be adopted in our microscopic view. Our approach in the study of triadic closure dynamics advocates, rather, an extension of the triangle enumeration methods in order to swallow the temporal information. [3] Our starting point is the observation that time annotation impacts the number of isomorphic triangles. In a simple undirected graph the number triangles isomorphic to $u\hat{w}z$ is 6 (3!), while in the temporal case the ordering induced by time makes the isomorphism disappear.

Once the triangles have been extracted, we have all the information we need to study the triangle creation process during network evolution. In Figure 2, we show the volume of triangles that are created daily in Renren and in Facebook. On December 13, 2006 the Renren network was composed of more than 12 million triangles, two thirds of which resulting from an activity of triangle creation started in August. In Facebook, instead, during the three years of observation, we count more than 1.7 million temporal triads. By observing the triangle trend in the two datasets over the overall periods, we note a general skew in the triadic closure process.

Obviously, triangle and link volumes are strongly related, as the increasing number of triangles could impact the overall number of new links. This is true even though not all links derive from the triadic closure process. For instance, they could be the consequence of new node arrival or some other effects, namely a preferential attachment process or search for new friends by graph exploration.

[3] Methods for frequency-based pattern and temporal graph matching (Berlingerio et al. 2009) are not suitable for our purposes because of combinatorial arguments based on integer partition. For each integer i we should extract all the triangles such that the relative times of their links sum to i. In fact we are not interested in temporal pattern shifted in time.



(a) Renren

(b) Facebook New Orleans

Figure 2: Number of new links (red) and triangles (blue) formed during the growth of Renren and Facebook New Orleans, sampled each day. The magenta line represents the ratio between the triangle and the links created in a day (y-scale on the right).

A first hint in the validation of the impact of triadic closure on the creation of new links can be given by the comparison between the arrival of the new edge and the formation of the new triangle. In fact by comparing the link and the triangle time-series in Figure 2, we observe the same trend; an increase/decrease in the new link volume corresponds to an increase/decrease in the number of triads. In order to quantify this relation, in the same figure we plot the ratio between the triangle and the link time-series. Thus we can quantify, on average, the number of triangles closed by a link. Obviously, the average value does not account for the per-link fluctuation, although it gives an idea of the role played by the triadic closure process. Despite the above limitations, by analyzing the link/triangle ratio we find two interesting results - both of which are related to events more or less external to the network.

In the Renren network the ratio in the last eight months remains quite constant and stabilizes at 1.5 triangle/link, suggesting a steady-state. Yet interestingly, we can also observe the peak in the triangle/link ratio in late August, when a spike in triadic closures occurs. This peak deserves further study. We conjecture that external events are intervening to speed up the network growth processes. In this case we could speak of 'summer effect' due to the combination of two facts: i) in the period we are analyzing Renren got targeted as a service for college students and ii) in those days Chinese students had a break from courses, so had extra time to pursue other interests and meet new people. As a consequence, offline encounters were also mirrored in the online network.

A more extensive and substantial result emerges from the triadic closure counting on the Facebook network. As evident in Figure 2(b), the network shows an abrupt transition after March 26, 2008. This date corresponds to the introduction of Facebook "People You May Know" (PYMK) functionality, which promptly impacts both network and triangles. In fact, prior to the launch of PYMK, the triangles/links ratio is quite similar to Renren's, then rapidly increases and stabilizes at the greater value of 3-4 triangles/link. We note how the PYMK mechanism highly impacts the microscopic characteristic of the network structure. In particular, it influences the link creation process to highly favor triadic closures. This strong effect cannot be captured by analyzing only the link creation over time. In fact we observe only a medium increase in the new link volume as shown in Figure 2(b). This observation stresses the importance of adopting different indexes in describing the network evolution; in fact, the number of new links alone would not be enough to let the phase transition emerge in the triadic closure process.

As in the Renren case, we see how events external to the network topology can highly influence its dynamical properties. But Renren and Facebook show totally different triggering events. In Renren the event is seasonal and behavioral, absolutely external to the network. In Facebook the event is external to the network but internal to the service. These observations have two main implications: 1) new features of the service could rapidly and massively modify the structure it manages, in a sort of feedback effect; 2) truly external events trigger changes but have a limited temporal impact on the network topology and, as a consequence, are harder to detect.

We have shown how the triadic closure process is a fundamental mechanism in the growth of online social networks and how it impacts their evolution. Nevertheless, we only consider the result of the process, *i.e* the triangle which has already been formed, and make no mention of how it got there. The question which comes to mind is how long we have to wait before observing the triadic closure effects, *i.e.* how long a triangle takes to be established.

Triadic Closure Delay

The triadic closure process has never been analyzed in temporal networks evolution. The total amount of time a triangle takes to be closed and the temporal relation among the constitutive links still await further study.

Definition The definition of the time taken by a triangle to establish is tricker since we are considering a dynamic process, where the components could appear at different times. The usual definition of triangle closure is the conditional probability that a link (u, z) is formed given that links (u, w) and (w, z) exist. In physical time, its delay is captured simply by $\tau(z, u) - \max(\tau(u, w), \tau(w, z))$. It accounts for the time to close the last link of a triangle. While this metric is very useful to understand triangle features within a static context, it cannot capture the timing creation of the links of a triangle, and is thus unable to shed a light on the temporal formation process of triads. We introduce a new definition of triadic closure delay which embeds the time of formation

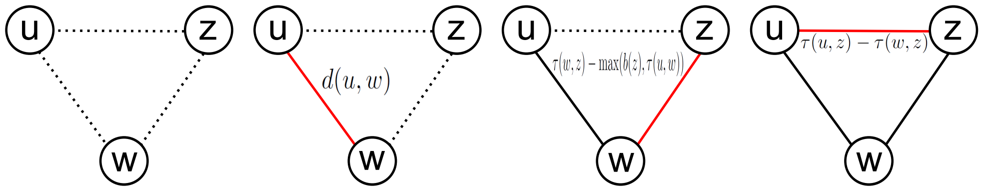


Figure 3: Steps in the formation of the triangle and the definition of its constitutive elements. In 3(a) the potential triangle (dotted links) which will form at the end of triadic closure. In 3(b) the link (u, w) (red) establishes in $t_\Delta(u, w)$ unit time. In 3(c) the second link (w, z) forms in $t_\Delta(w, z)$ time and finally in 3(d) the last link takes $t_\Delta(z, u)$ to be created and the process ends.

of all the links of triads, by considering the temporal process of triad formation as given by both the nodes and the links appearing in the network.

For example, in a triangle $u\hat{w}z$ the creation of the edge (u, w) depends on the presence of nodes u and w and the remaining links. To take into account these arguments, we employ the *birth date* $b(u)$, which denotes the time of the first appearance of node u into the network.

Once the birth date has been defined, we can normalize the triangle creation time swallowing the temporal gap of node appearance, thereby capturing the real feasibility of a triangle. To attain a global definition of triadic closure time, we focus on the definition of its constitutive elements. We indicate the normalized time of the link (u, w) in a triangle $\Delta = u\hat{w}z$ as $t_\Delta(u, w)$.

To give a general definition of the triangle delay, we follow the steps that characterize the triadic closure process shown in Figure 3. In 3(a) we show the potential triangle with no links among nodes. The first element to be created is the red link (u, w) in Figure 3(b) and we have to measure how long it takes to be established. It corresponds to the delay of the link (u, w), so

$$t_\Delta(u, w) = d(u, w)$$

The triadic process is still in node w as the link (w, z) has not been created yet. Two possible situations could arise just before the creation of the red link in Figure 3(c): 1) node z is already in the network, so $\tau(u, w) > b(z)$ and $t_\Delta(w, z) = \tau(w, z) - \tau(u, w)$; 2) node z is absent, so the closure has to wait for its appearance. In the latter case we have $b(z) > \tau(u, w)$, so we discount the waiting time of the process in the node w, $\tau(u, w) - b(z)$, obtaining $t_\Delta(w, z) = \tau(w, z) - b(z)$. Putting together the conditions we obtain the general definition for $t_\Delta(w, z)$:

$$t_\Delta(w, z) = \tau(w, z) - \max(b(z), \tau(u, w))$$

The last step involves the creation of the link z, u as depicted in Figure 3(d). By definition of b and the ordering of the time values of the links, at the creation of the link (z, u), nodes w and z are already participating, so

$$t_\Delta(z, u) = \tau(z, u) - \tau(w, z)$$

.

Once the delay of each constitutive element has been defined, we can define the triadic closure delay.

Let $\mathbb{G}$ be a temporal undirected graph and $u\hat{w}z$ a temporal admissible triangle, i.e. $\tau(u, w) < \tau(w, z) < \tau(z, u)$, the *triadic closure delay* of $u\hat{w}z$, $d(u\hat{w}z)$ is defined as the sum of the normalized times of its links

$$d(u\hat{w}z) = t_\Delta(u, w) + t_\Delta(w, z) + t_\Delta(z, u)$$

which corresponds to

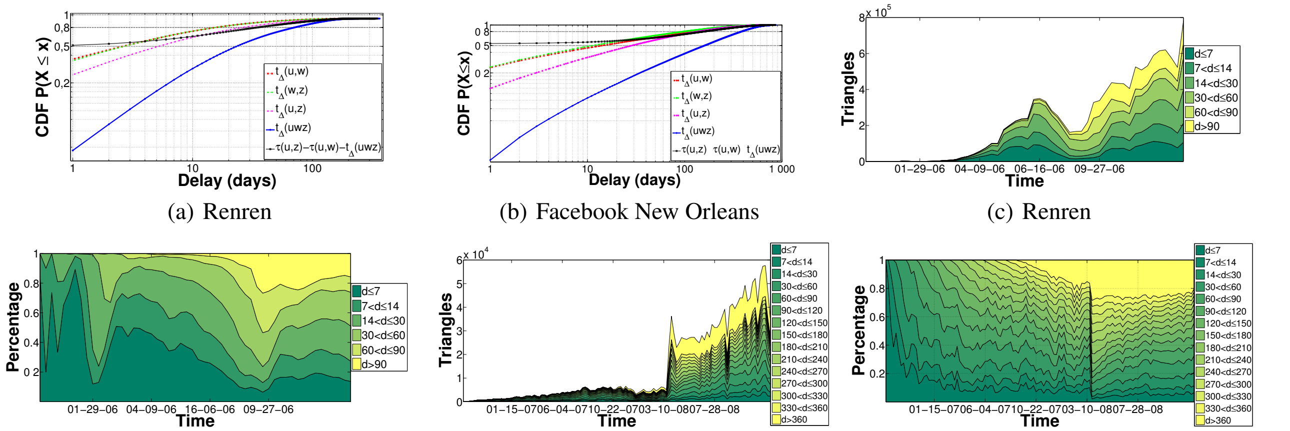
$$d(u\hat{w}z) = d(u, w) - \max(b(z), \tau(u, w)) + \tau(z, u)$$

From the above definition we must observe that the triadic closure delay does not depend on the creation time of the middle link (w, z). The normalization given by the birth date is quite important since it covers all nodes. In Figures 4(a) and 4(b) we quantify the effects of the normalization through the distribution of $\tau(z, u) - \tau(u, w) - d(u\hat{w}z)$, that represents the difference between the triadic delay and the delay not normalized, *i.e.* $\tau(z, u) - \tau(u, w)$. We can observe that the normalization impacts on 50% of triangles. In particular, 40% of triads are involved in a delay normalization of more than a month.

Triadic closure delay properties We analyze the triadic closure delay and the t_Δ of each link in a triangle and then compare the different evolutions of the two networks we are studying. In fact, the temporal information not only allows us to measure the triangle delay but also to temporally place it. This enabled us to verify whether or not the fast triangle trend is stable during the network growth and to see if external mechanisms, e.g. Facebook's PYMK, modified this trend.

In Figure 4 we report the delay CDF for Renren and Facebook. Considering the Renren distribution in Figure 4(a) we observe that most triangles have a high speed formation, given that half of the triangles close in less than 25 days. This fact stresses the importance of the study of triangle formation dynamics. In effect, triadic closure impacts the network structure both significantly and quickly. As shown in Figure 4(b), we observe a different behavior in Facebook. The measured delay is much greater than in the Renren case, as half of the triangles get established in five months at most.

We question if all the elements are necessary to predict the triadic closure delay. This corresponds to verifying whether or not certain relationships occur among the different elements. For example, if $t_\Delta(u, w)$ and $t_\Delta(w, z)$ are low, what can we say about the delay of $t_\Delta(u, z)$. Will it be low



(a) Renren (b) Facebook New Orleans (c) Renren

(d) Renren (e) Facebook New Orleans (f) Facebook New Orleans

Figure 4: CCDF of the triadic closure delay and its constitutive elements. The black CDF in both figures represents the effect of the normalization of the triadic delay definition w.r.t. the simple definition that does not consider the node arrival process. 4(c) and 4(e): The number of triangle created in each week and divided in different delay groups. 4(d) and 4(f): The volume trends (percentage) of the triangle delay groups during the growth of Renren and Facebook. Group green intensity is inversely proportional to triangle delay (upper and lower bound of the groups are in days).

too? To stress possible relationships among the elements we adopt two approaches. First, we randomize the $t_\triangle$s to delete any relations between the triangle elements. Then we compare the resulting delay distribution with the real one. The randomization is obtained by shuffling the elements in each column of the matrix $T_\triangle = [t_\triangle(u,w), t_\triangle(u,w), t_\triangle(u,w)]$. We find that the delay and the shuffled delay distributions are quite similar. This observation suggests a lack of a particular relation between the delay elements of a triangle. Furthermore we confirm the above result by computing the correlation matrix among $t_\triangle(u,w)$,$t_\triangle(z,w)$ and $t_\triangle(z,u)$. More specifically we find correlation coefficients close to 0 (from 0.03 to 0.05) for each pair of variables.

Generally, these observations suggest that the single delay of the constitutive elements is not sufficient to explain the total delay of triangles. In other words, the triadic closure delay cannot be predicted by simply observing a single element.

Delay dynamics in the network growth As shown in Figure 2, in both datasets the triangle formation trend is not regular. Now through triadic closure delay we are to able to capture what kind of triangles (low or high delay) contribute to the observed irregularity. In Figure 4 we show the impact of the triadic closure delay groups during network growth. In Figure 4(c) and 4(e), we divide the new triangles created in each week into groups according to their delays. In Figure 4(d) and 4(f), we maintain the same groups but we normalize the contribution of each group with respect to the total number of triangles formed during the week. This way we quantify the absolute and relative contributions of each group to the triangle formation dynamics. For example, a group could undergo a rapid increase (absolute volume) but have an overall low impact (relative) simply because a general boost of formation activity of the triads.

By analyzing the absolute and the relative volume of triangles grouped in group delay in Renren, (Figures 4(c) and 4(d)) we observe three behaviors. First, groups with a delay less than a month run into a continuous decrease from the beginning of June, when they reach the maximum activity, to the beginning of August, *i.e.* during the summer. This fact may be relevant to the relative drop of the link creation. Second, we see a peak of the $[60 - 90]$ days group that spans all August. This fact indicates that during August the potential triangles, begun before the summer, actualize. Finally, after the summer, a component of high delay triangles (yellow group) appears and stabilizes on 30% of the new triangles. In general, seasonal effects on Renren are influential, and the triadic closure delay is able to measure which latent triangles it acts on.

As for the absolute volume in Facebook, in Figure 4(e) we observe that the eight first delay groups keep slowly increasing. That accounts for a component of fast triangles which is independent from the stage of the network and that involves a similar number of triangles. Other groups, characterized by a higher delay, manifest after the PYMK service and stay quite constant until the end of the sampling period. Relative to the volume trends, in Figure 4(f) we observe that the PYMK service primarily acts on the 'old' triangles. This implies that the suggestion mechanism based solely on the common-neighbors friend recommendations, awakens latent links long asleep.

In addition, we can quantity the long period effects of the mechanism. Specifically, we observe that after a brief period from the PYMK introduction, the relative volumes stabilize, with the exception of the [30-90] groups which increase slightly. By comparing the distribution of the triadic closure delay groups before and after the introduction of the friend suggestion system, we observe (Figure 4(f)) that PYMK is likely to promote higher delay triangles. By analyzing the triadic closure delay during the temporal evolution of Facebook, we were able to quantify the effects and the impact of the PYMK feature.

6 Conclusion

This paper takes a first step in the direction of building a set of metrics capable to characterize social network dynamics. Until now, in fact, when it comes to understanding detailed dynamics and evolution inside these networks, current research offers very little in terms of metrics or models. We focus our attention on two metrics: *link delay* and *triadic closure delay*. They can capture the time delay between when a link or triadic closure is *possible*, and when they actually instantiate in the trace. We have applied these metrics to two real traces of social network dynamics from Renren and Facebook, and we have shown that they are generally consistent across networks, but their differences shed light on interesting properties of each system.

References

Amblard, F.; Casteigts, A.; Flocchini, P.; Quattrociocchi, W.; and Santoro, N. 2011. On the temporal analysis of scientific network evolution. In *Proceedings of the International Conference on Computational Aspects of Social Networks*, CASoN '11. IEEE.

Berlingerio, M.; Bonchi, F.; Bringmann, B.; and Gionis, A. 2009. Mining graph evolution rules. In *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*, ECML-PKDD '09.

Easley, D., and Kleinberg, J. 2010. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. New York, NY, USA: Cambridge University Press.

Gaito, S.; Zignani, M.; Rossi, G. P.; Sala, A.; Zhao, X.; Zheng, H.; and Zhao, B. Y. 2012. On the bursty evolution of online social networks. In *Proceedings of the First ACM International Workshop on Hot Topics on Interdisciplinary Social Networks Research*, HotSocial '12. ACM.

Garg, S.; Gupta, T.; Carlsson, N.; and Mahanti, A. 2009. Evolution of an online social aggregation network: an empirical study. In *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement*, IMC '09. ACM.

Gong, N. Z.; Xu, W.; Huang, L.; Mittal, P.; Stefanov, E.; Sekar, V.; and Song, D. 2012. Evolution of social-attribute networks: measurements, modeling, and implications using google+. In *Proceedings of the 2012 ACM Conference on Internet Measurement*, IMC'12. ACM.

Gonzalez, R.; Cuevas, R.; Motamedi, R.; Rejaie, R.; and Cuevas, A. 2013. Google+ or google-?: dissecting the evolution of the new osn in its first year. In *Proceedings of the 22nd International Conference on World Wide Web*, WWW '13. ACM.

Jiang, J.; Wilson, C.; Wang, X.; Huang, P.; Sha, W.; Dai, Y.; and Zhao, B. Y. 2010. Understanding latent interactions in online social networks. In *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement*, IMC '10. ACM.

Kikas, R.; Dumas, M.; and Karsai, M. 2013. Bursty egocentric network evolution in skype. *Social Network Analysis and Mining* 3(4):1393–1401.

Kossinets, G., and Watts, D. J. 2006. Empirical analysis of an evolving social network. *Science* 311(5757):88–90.

Kossinets, G.; Kleinberg, J.; and Watts, D. 2008. The structure of information pathways in a social communication network. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '08, 435–443. New York, NY, USA: ACM.

Leskovec, J.; Backstrom, L.; Kumar, R.; and Tomkins, A. 2008. Microscopic evolution of social networks. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '08. ACM.

Mislove, A.; Koppula, H. S.; Gummadi, K. P.; Druschel, P.; and Bhattacharjee, B. 2008. Growth of the flickr social network. In *Proceedings of the First workshop on Online Social Networks*, WOSN '08. ACM.

Rapoport, A. 1953. Spread of information through a population with socio-structural bias: I. assumption of transitivity. *The Bulletin of Mathematical Biophysics* 15(4):523–533.

Romero, D. M., and Kleinberg, J. M. 2010. The directed closure process in hybrid social-information networks, with an analysis of link formation on twitter. In *Proceedings of the Fourth International Conference on Weblogs and Social Media*, ICWSM'10.

Sun, Y.; Han, J.; Aggarwal, C. C.; and Chawla, N. V. 2012. When will it happen?: relationship prediction in heterogeneous information networks. In Adar, E.; Teevan, J.; Agichtein, E.; and Maarek, Y., eds., *WSDM*, 663–672. ACM.

Viswanath, B.; Mislove, A.; Cha, M.; and Gummadi, K. P. 2009. On the evolution of user interaction in facebook. In *Proceedings of the 2nd ACM Workshop on Online Social Networks*, WOSN '09. ACM.

Wilson, C.; Boe, B.; Sala, A.; Puttaswamy, K. P. N.; and Zhao, B. Y. 2009. User interactions in social networks and their implications. In *Proc. of ACM EuroSys*.

Wilson, C.; Sala, A.; Puttaswamy, K. P. N.; and Zhao, B. Y. 2012. Beyond social graphs: User interactions in online social networks and their implications. *ACM Transactions on the Web* 6(4).

Zhao, X.; Sala, A.; Wilson, C.; Wang, X.; Gaito, S.; Zheng, H.; and Zhao, B. Y. 2012. Multi-scale dynamics in a massive online social network. In *Proceedings of the 2012 ACM Conference on Internet Measurement*, IMC '12. ACM.