Network Science **4** (3): 293–313, 2016.
doi:10.1017/nws.2016.13

Evolution through bursts: Network structure develops through localized bursts in time and space

HILLA BROT
Center for Polymer Studies and Department of Physics, Boston University, Boston, MA 02215, USA
and
Kellogg School of Management, Northwestern University, Evanston, IL 60208, USA
(*e-mail:* hillay@gmail.com)

LEV MUCHNIK
School of Business Administration, The Hebrew University of Jerusalem, Jerusalem 91905, Israel
(*e-mail:* levmuchnik@gmail.com)

JACOB GOLDENBERG
School of Business Administration, IDC Herzliya, Herzliya 46150, Israel
(*e-mail:* jacob.goldenberg@mail.huji.ac.il)

YORAM LOUZOUN
Department of Mathematics, Bar-Ilan University, Ramat Gan 52900, Israel
and
Gonda Brain Research Center, Bar-Ilan University, Ramat Gan 52900, Israel
(*e-mail:* louzouy@math.biu.ac.il)

Abstract

Models of network evolution are based on the implicit assumption that network growth is continuous, uniform, and steady. Using the data collected from a large online-blogging platform, we show that the addition and removal of network ties by users do not occur sporadically at isolated nodes spread all over the network, as assumed by the vast majority of stochastic network models, but rather occur in brief bursts of intense local activity.

These bursts of network growth and attrition (addition and removal of network ties) are highly localized around focal nodes. Such network changes coincide with nearly instantaneous densification of the ties between the affected nodes, resulting in an increase of local clustering. Furthermore, we find that these network changes are tightly coupled to the dynamics of individual attributes, particularly the increase in homology between neighboring nodes (homophily) within the scope of the burst. Coincidence of the localized network change with the increase in homophily suggests a strong coupling between the selection and influence processes that lead to simultaneous elevation of assortativity and clustering.

Keywords: *homophily, simultaneous events, individual traits, coupling, bursts, network dynamics, LiveJournal, triadic closure, feedback*

1 Introduction

During the last decades, considerable research has been devoted to understanding network formation processes. This research resulted in a number of insightful models of network formation (Albert et al., 1999; Barabási & Albert, 1999; Kleinberg et al., 1999) that strive to explain and reproduce the most prominent network properties (degree distribution, clustering, etc.). To reproduce these (static) properties, most models assume that the link addition and removal processes are monotonic in time, composed of independent events (e.g. link addition/removal or edge duplication of events) and not affected by the history of events in neighboring nodes. A few previous models suggested different mechanisms of growth such as super linear preferential attachment (Gabel & Redner, 2013; Golosovsky & Solomon, 2012) or a node's attachment based on clustering coefficient, which can in principle lead to bursts in the triangles creation dynamics (Holme & Kim, 2002; Vázquez, 2003). Other models were based on preferential attachment with limited memory (Klemm & Eguiluz, 2002). Some works show empirically that social and cellular networks become denser over time and that their diameter shrinks (Chaintreau et al., 2007; Leskovec et al., 2008; Leskovec et al., 2005). Such results suggest parallel triangle closure in neighboring nodes. However, none of these studies assumes that the links dynamic is a non-monotonic process.

In this paper, we show that a network's dynamic processes are highly local and temporally non-uniform. Using a sequence of snapshots representing the evolution of a large online social network, we show that there is a non-intuitive organization of the dynamics of tie (link) addition and deletion dynamics: Addition and deletion of links occur in localized bursts, with seeming coordination resulting in formation of dense neighborhoods. In particular, areal densification bursts result in a nearly instantaneous increase in clustering among the nodes in the network neighborhood. The supplementary animations connected with this paper (http://peptibase.cs.biu.ac.il/homepage/Evo.avi) demonstrate the addition of new links in the region surrounding a pair of connected and a pair of randomly selected nodes.

In this study, we simultaneously track the co-evolution of the network structure and the local similarity in "lists of interests" of approximately 8 million users of the largest connected component of an online social network. The detailed information about added and removed links and the changing attributes of individual users is used here to uncover details about the dynamics of the link-formation process. We find that the evolution of the social network structure through the addition and removal of network ties (links) occurs through densification bursts. These bursts affect the nodes located within two hops of the focal node and coincide with a nearly instantaneous increase in similarity and clustering among the nodes in that network neighborhood. Such eruptions of local activity lead to the emergence of comparatively dense local groups of users (designated as communities or clusters in different disciplines). Our findings provide insight into the mechanisms underlying the evolution of communities in social networks and contribute to our understanding of the full network structure formation process and its relationship to local link dynamics. These findings have implications for network modelers as well as policy planners.

The non-homogeneous network growth patterns and the bursts we report are in line with the known temporal and heterogeneous patterns of individual and collective human activity in several domains, including personal communication (Barabasi, 2005; Rybski et al., 2009), activity of page views and revisions of individual

Wikipedia pages (Kämpf et al., 2012), Internet (Anderson & Srinivasan, 2003; Paxson & Floyd, 1995), web use patterns (Dezsö et al., 2006), financial time series (Yamasaki et al., 2005), and road traffic (Jagerman & Melamed, 1994). A number of explanations for the nature of these patterns have been proposed over the years (Adamic & Huberman, 2000; Barabasi, 2005; Rybski et al., 2009), and a detailed survey was recently published by Min and Goh (Min & Goh, 2013). For example, (Barabasi, 2005) showed that priority-based queuing of tasks leads to appropriately skewed distribution of inter-activity gaps. (Rybski et al., 2009) studied a collection of millions of messages exchanged over time between members of two online communities and found similarly skewed distributions of communication patterns between social network peers. The complementary model, reproducing empirically observed communication patterns, hinges on the existence of long-term correlations in communication activity. These correlations could arise from either internal or external stimuli (i.e. activity of other network peers) and may even diffuse within the network. In addition to patterns of individual activity bursts, (Garber et al., 2004) reported localization of network activity in diffusion of innovation. They reported that spatial clustering precedes takeoff in sales.

In the present research, we show that links in social networks tend to be added in brief batches of local densification. In contradiction to what is usually assumed, link addition and removal should not be seen as a set of continuous independent events dispersed throughout the network.

An interesting aspect of these densification bursts is that they occur simultaneously at the network and content levels. Most network formation models either focused on the link addition and removal process, assuming a given node content (e.g. homophily-induced link addition or selection) (Kimura & Hayakawa, 2008; Sayama et al., 2013), or they studied node content evolution in a fixed network (e.g. dynamic models of neural networks or contact processes) (Cowley et al., 2001; Diesmann et al., 1999). Other works studied co-evolving social and semantic networks (Diesner, 2012; Roth & Cointet, 2010) as well as the development of social networks and their influence on users' behavior (Crandall et al., 2008; Khosla et al., 2014) or a node's popularity (Figueiredo et al., 2014; Papadopoulos et al., 2012; Perc, 2014) and the emergence of similarity in social systems (Aiello et al., 2012; Centola & van de Rijt, 2015; Fu et al., 2012; Grund, 2014). Here, we show that attributes of adjacent nodes become more similar in parallel with the processes leading to densification and greater clustering. In the following section, we show that all aspects of a network co-evolve through concurrent densification bursts.

2 Data and methods

We collected panel data representing the evolution of a large online social network, *LiveJournal* (LJ) (http://www.livejournal.com/), which is one of the earliest and still popular (Alexa rank of 138 as of February 2014) online blog systems used to publish posts on issues ranging from personal experiences to political views, news events, gossip, and sharing of web links. Each LJ user maintains an individual list of blogs and a set of "friendship" relationships that simplifies navigation, grants access to privately published content, and fuels the automatically generated news feed. Taken together, these bookmarks comprise a collaborative social network in which ties are motivated by a wide spectrum of factors, ranging from personal acquaintance characteristics for Facebook, to content acquisition typical for Twitter.

We crawled the LJ social network, acquiring a sequence of periodic snapshots that represent the evolution of the network over a period of about 1 year. The dataset records addition of approximately 13 million and removal of approximately 7 million links (10% and 5% per snapshot between each pair of snapshots, which were taken at approximately 6-week intervals). Table 1 provides descriptive statistics of this dataset.

We represent a timeline of each network potential link (i.e. a pair of nodes) by an 8-bit binary number in which each position represents the existence of the link in the corresponding snapshot between these nodes. In other words, for a sequence of eight snapshots, the potential link between a pair of nodes is flagged with 1 in the snapshots in which it exists and 0 in the ones where it does not exist. For example, the sequence 00111100 represents a pair of nodes that established a link sometime between the second and the third snapshots and maintained it until the link was deleted between snapshots six and seven. In the following analysis, we specifically focus on the subset of links denoted by 00001111, the links created in the middle of the inspected period, between snapshots four and five. This selection offered a sufficiently long observation of the network and user attribute evolution, covering the period preceding and following the link addition.

To enhance the analysis of the network evolution dynamics, we collected a complementary, high-frequency sample of the LJ network. We monitored local networks of 359 randomly selected nodes by performing a two-step snowball sampling at 12-hour intervals. Although the resulting sample does not cover the entire network, we use a sequence of 83 consecutive snapshots represented by a time series of bits, as described above, to trace network evolution dynamics at a finer resolution. We collected this additional dataset from the same social network to ensure that the measurements conducted in both cases represent the same social processes, and to avoid possible discrepancies between the results obtained in each dataset. The higher temporal resolution is instrumental in establishing the tight, intraday coupling between the processes (Table 1).

2.1 Homophily calculation

Social networks exhibit a high degree of assortative mixing (McPherson et al., 2001) that may evolve as a result of selection when subjects choose similar ones to connect to; or, as a result of peer influence when already connected, peers become similar due to interaction[1]. To study these dynamics, we introduce a similarity measure that leverages detailed data contained in user profiles to compute similarity of any two members of the LJ social network. More specifically, we use the evolving lists of interests declared by the users to compute the change in similarity over time and relate it to the changes in the connectivity of the underlying social network. We thereby define the similarity between pairs of (self-declared) domains of interest (DOIs) and use it to compute the similarity between pairs of nodes.

[1] The third possible cause of assortative mixing, correlated exogenous factors simultaneously influencing connected nodes, lie outside the scope of this work.

https://doi.org/10.1017/nws.2016.13 Published online by Cambridge University Press

Table 1. *Summary details of the first and second datasets of lower and higher temporal resolution.*

	Number of snapshots	Number of nodes	Time between snapshots	Number of links at the first snapshot	Number of links at the last snapshot	Average percent of link increase in each snapshot
First dataset	8	9066862	6 weeks	126810165	139289908	1.1%
Second dataset	83	2411693	12 hours	66145254	67041309	0.02%

Note: The first column is the number of snapshots used. The second one is the total number of nodes used for the entire analysis (even if they appeared in a single snapshot). The third column is the time interval between snapshots. The following columns are the initial and final number of links that are very similar. The link numbers are very similar, with a total increase of 10% over the 8 snapshots, and less than 2% along the 83 snapshots of the second dataset, as detailed in the last column.

Simple measures such as the Jaccard index (Hamers et al., 1989) cannot be used, since the overlap between any two random persons is minimal, even if their DOIs are similar. For example, the DOIs of "electric guitar" and "guitar" are different but may represent common interests. Instead, we define a more complex approach is required, based on the similarity between DOIs. First, we generate a list of DOIs C_i for each user i. Then, we develop a similarity matrix B representing the number of co-occurrences of each possible pair of DOIs in the same user profile. For example, if 20 users listed both "Music" and "Physics" as their DOIs, then their corresponding value in the matrix B would be 20. Once constructed, each element b_{jk} of B is normalized by the number of occurrences of the DOI j, d_j, to produce the asymmetric matrix:

$$\tilde{b}_{jk} = \frac{b_{jk}}{d_j}. \tag{1}$$

We artificially set $b_{jj} = 0$. This matrix represents the common usage of DOI and can be used as a proxy for similarity.

Considering the large number of potential DOI pairs ($\sim 10^{14}$) , we chose to consider only the ones that co-occur four or more times. This yields a sparse and manageable, yet meaningful, matrix B.

We then define $S_{i,l}$ —a similarity measure between any two nodes i and l using the similarity matrix above. Given a set of DOIs in node i—C_i, the similarity is defined based on the fraction of shared DOIs and the similarity between non-shared DOIs:

$$S_{ij} = \frac{|C_i \cap C_l|}{\max(|C_i|, |C_l|)} + \frac{\sum_{j \in C_i, k \in C_l} \left(b_{jk} + b_{kj}\right)}{\max(|C_i|, |C_l|)}. \tag{2}$$

Note that other similar measures could be used. However, they all must include the indirect similarity measure based on the similarity between different DOIs—b_{jk}.

3 Results

3.1 Bursting in individual node degree evolution

Using the observed LJ snapshots, we define the in-degree and the out-degree timelines for each node by counting the number of its incoming and outgoing links observed in each network snapshot. We can then describe any pair of nodes in terms of four numbers: the in- and out-degree of the first node in the pair, and the in- and out-degree of the second node in the pair. Consider, for instance, the effect of a new tie connecting a pair of nodes V1 and V2 on the in- and out-degrees of each node (Figure 1a). Addition of a new link between any two nodes may be related to the number of common first and second undirected neighbors, or the clustering coefficient of the nodes, before or after the addition event. Here, we define the common first neighbors to be the nodes directly connected to each of the nodes in the pair, whereas common second neighbors are the nodes located at a distance of up to two hops from any node, but are not first neighbors (Figure 1(b)). All node pairs can be classified as belonging to one of three groups: those sharing at least one common neighbor, those having only a common second neighbor, and those having no first or second common neighbor.

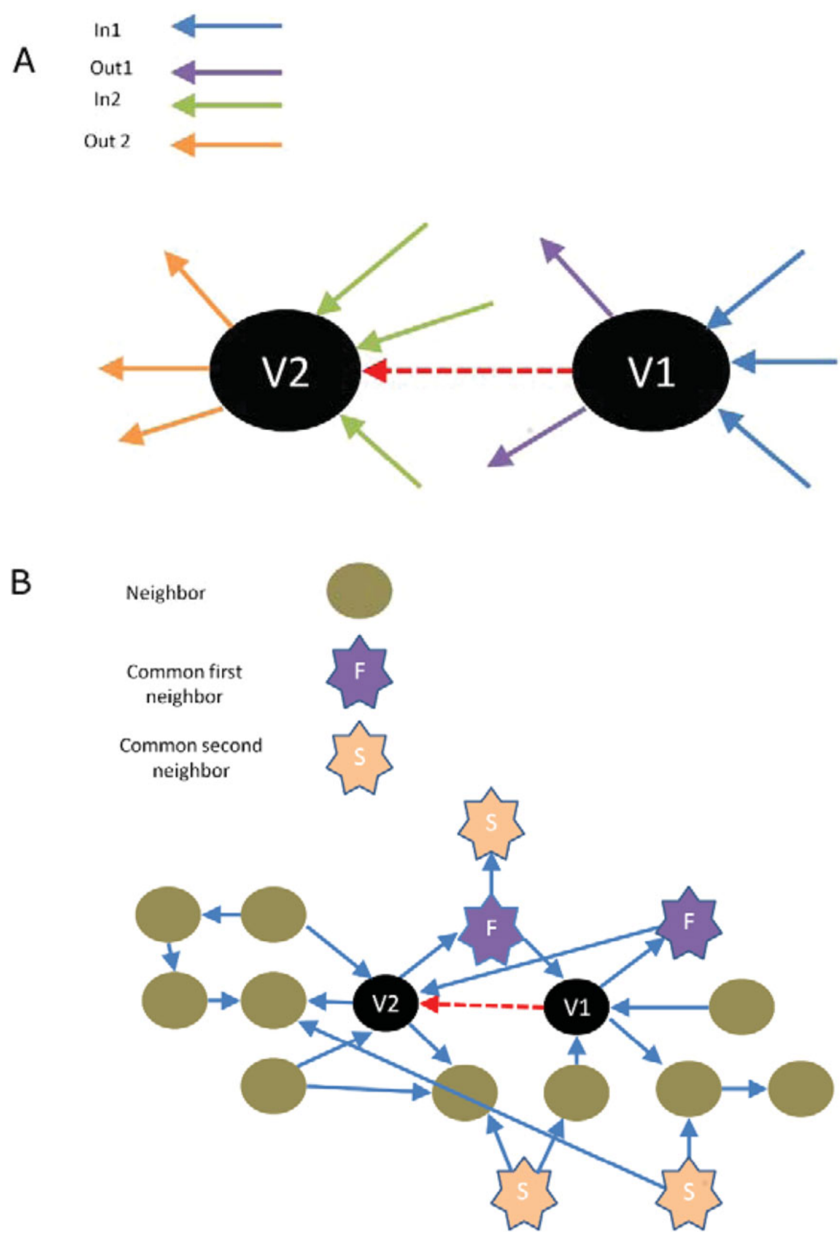

Fig. 1. Schematic representation of directional degree, neighbors, and second neighbors. The figure illustrates the variety of measures that capture local network attribute dynamics over time. Panel A shows the link target node's in-degree and the source node's out-degree for a pair of the newly connected nodes. We use the In1, In2, Out1, and Out2 in all following figures. Panel B shows the in- and out-degrees of the nodes surrounding the newly added link, and the definition of first and second neighbors. (Color online)

For the sake of simplicity, we perform the computations of the common first and second neighbors in the undirected network underlying the original directed network (Figure 1(b)).

3.2 Correlated densification of local network structure

Observation of the in-degree, out-degree, and the fraction of the pairs sharing first- or second-order neighbors over time reveals bursts of link addition and their effect on network evolution (Figure 2(a) and (b)). The changes in these elements are not random, as one would expect from a common stochastic process, but coincide with the addition of the new tie connecting the pair. If link addition in the network were independent of recent history and uniformly distributed in time, we would expect no increase in any of these features upon addition of a single link, since the probability to add or remove links from neighboring nodes would be equal in equilibrium. However, although the network is approximately in equilibrium (a net increase of less than 5% in the average degree per sample), this is not the case: Both incoming and outgoing degrees of the nodes comprising the newly connected pair increase insignificantly before and after formation of the new link. Moreover, we observe a very sharp increase of these properties (an increase of more than 25% in the average degree between two consecutive snapshots) occurring around the time when this link is formed. Note, that we scale all plots by their value at the time of link creation ($t = 0$) and that the added link is not included in the analysis (Figure 2(a)). These measurements suggest that links tend to be added in batches around some nodes. Note that the link studied is not included in the analysis.

Furthermore, note that although increase of the out-degree of the source node (k_{out_1}) can be explained by bursts in activity of a single user (users are likely to start following several new blogs once they update their profile), the rise of the in-degree of either the source or the target nodes results from actions taken by other nodes, and it can be explained only by a burst of activity around the source or target nodes. Interestingly, these scattered actions are performed by nodes residing in the same network area and lead to rapid densification and increased clustering. This phenomenon results in the surge of the probability that a pair of nodes will have common first and second neighbors following the addition of a link between them. These neighbors translate to short paths between nodes, resulting in a sharp decrease in the distances between the nodes in the affected area (the nodes within a short distance of the node pair; see Figure 2(b)).

Given the large divergence in the degree distribution, the effect presented here could mainly represent the nodes with a very low degree or oppositely—the result of a sub-group of nodes with a very high degree and simultaneous drastic changes in their degrees. To dismiss the possibility that the results presented in this work depend on nodes' connectivity, we have validated that they are maintained for low, intermediate and high degree nodes (Appendix A Figure A1).

To confirm that our results could not merely be an artifact of a stochastic edges' addition and deletion process, we compared them to the ones produced by a regular preferential attachment model. More specifically, we conducted the same measurements on a simulated dynamic network with $N = 5000$ nodes and average degree of $K = 100$. In each iteration, edges were created according to a linear preferential attachment rule, and a similar number of edges were randomly deleted. There were no significant changes in the directional degree before and after an edge creation event in this system. The simulation results are show in Appendix A Figure A2 where they are presented on the same scale as in Figure 2(a).

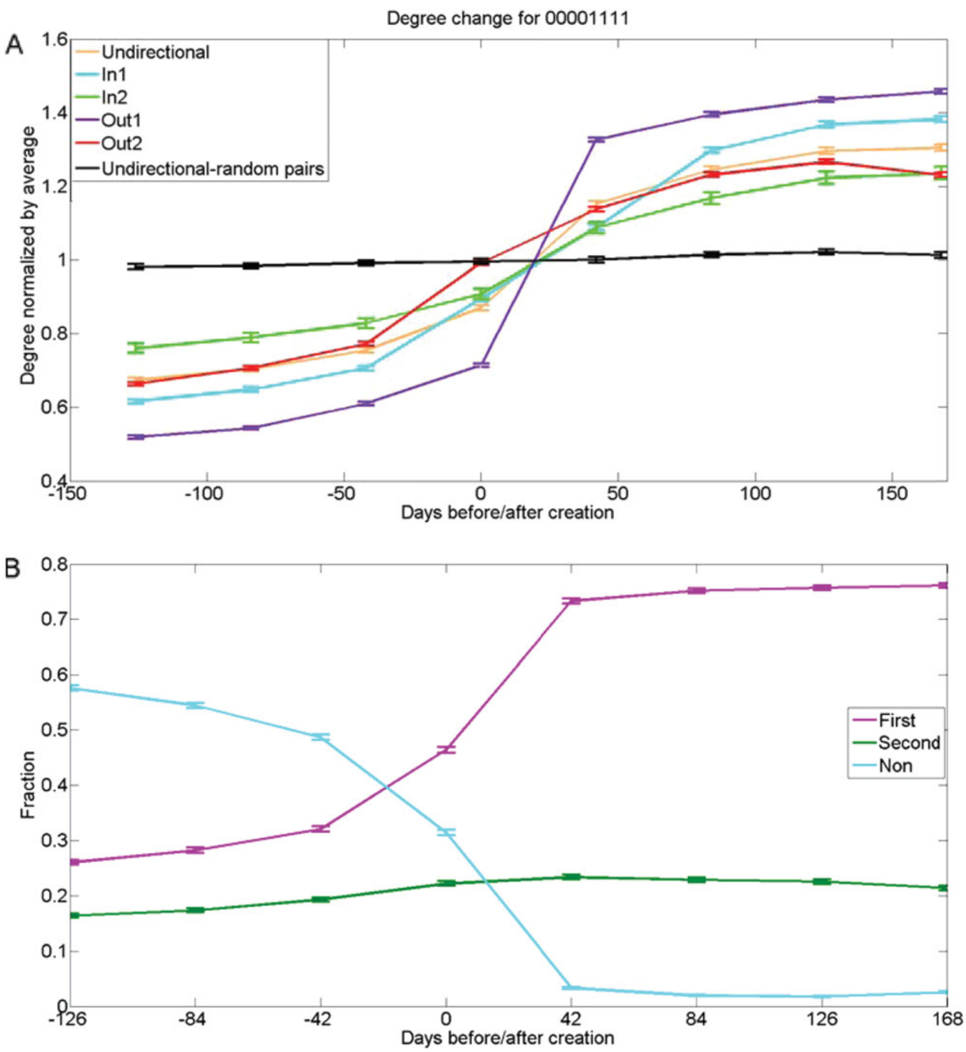

Fig. 2. Evolution of node attributes before, during, and after they connect. The figure shows the change in the in- and out-degrees of the newly connected nodes (00001111) in the first dataset (snapshots, taken at 6-week intervals). We computed the in- and out-degree of each node surrounding the newly added link (panel A), as well as the number of their common neighbors (panel B) (Unidirectional, orange; In1, light blue; In2, green; Out1, purple; Out2, red). For simplicity, we scale all plots by their value at the time of link creation ($t = 0$). All-time series show slow change in the months preceding and following the connection, but rapid dramatic changes coinciding with addition of the new link (time 0). In other words, nodes in newly established pairs acquire many new links (both incoming and outgoing) over a period surrounding the event. There is no change in the properties of random pairs of nodes (black line). This conclusion is supported by the increase in the fraction of the node pairs having at least one first (purple) or second (green) neighbor around the time of link creation (panel B). Acquisition of common neighbors by newly connected nodes leads to a sharp drop (from nearly 60% to virtually 0% in the population) in the fraction of such pairs having neither first nor second-order common neighbors (light blue). These measurements show that link addition cannot be seen as a sequence of isolated events, but should be considered synchronized densification leading to formation of triangles in the area. In A, the error bars represent standard deviation of the directional degree of the observed nodes. In B, the error bars represent standard deviation of the existing/non-existing first or second neighbors for the observed nodes. (Color online)

https://doi.org/10.1017/nws.2016.13 Published online by Cambridge University Press

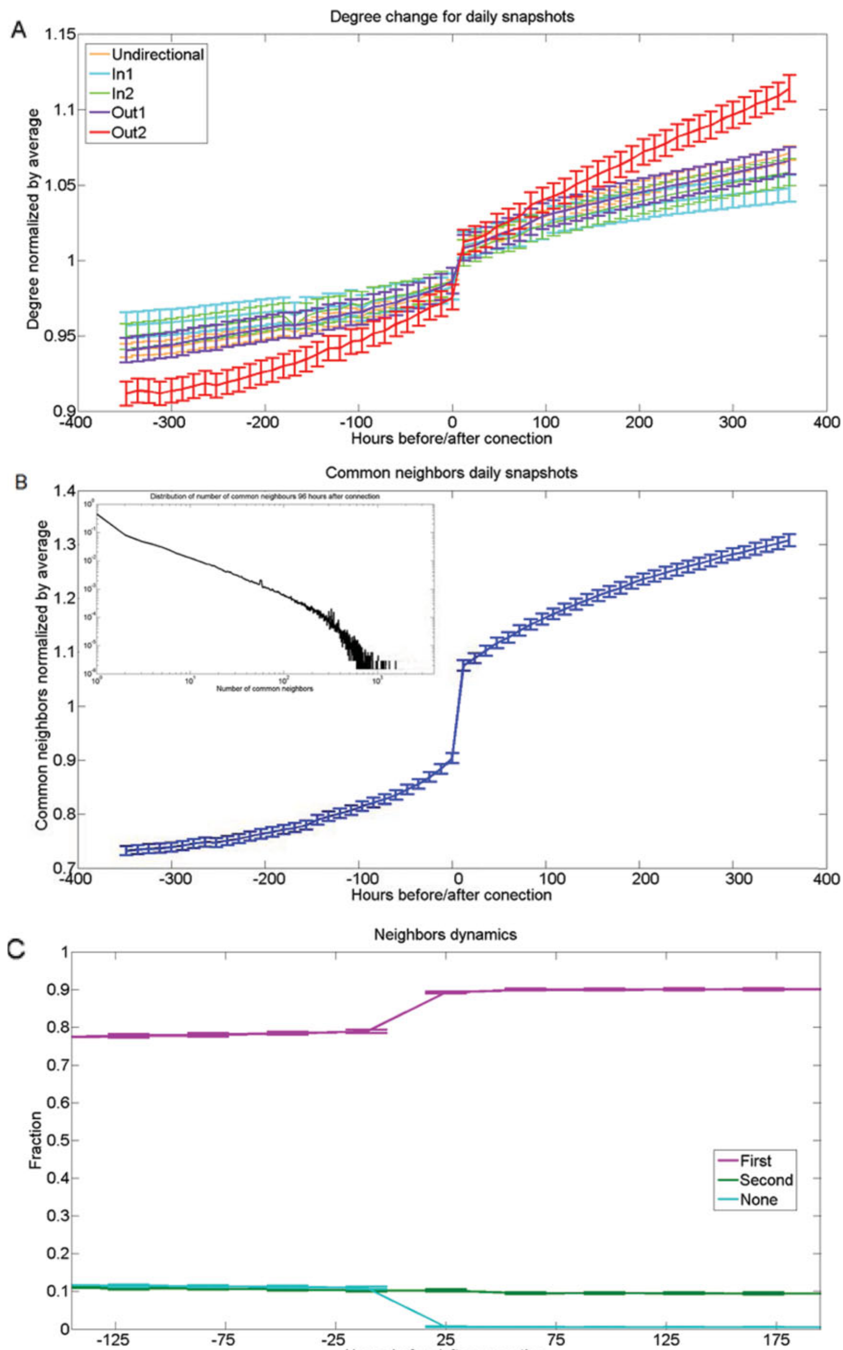

Fig. 3. Evolution of degrees and common neighbors of newly created pairs using the second (high-frequency) dataset. The change of the number of in- and out-degrees of each of the nodes in the pair (panel A) is analogous to that observed in Figure 2A, although the data, collected at higher sampling frequency, covers a significantly shorter time interval (Unidirectional, orange; In1. light blue; In2, green; Out1, purple; Out2, red). We can still observe a gradual monotonic increase in the average in- and out-degrees of each node, a sharp surge around the time of their connection, and an immediate reversion to the original gradual slope. The inset at panel A shows the degree distribution of 1 represented snapshot (96 hours after connection).

https://doi.org/10.1017/nws.2016.13 Published online by Cambridge University Press

Not surprisingly, due to triadic closure, about one-third of the newly connected links fall between nodes with common neighbors prior to the connection. However, the fraction of the node pairs sharing at least one common neighbor increases to 70% immediately after the new link is established. This phenomenon is even more pronounced when we consider longer trajectories involving second neighbors. Practically, all newly connected pairs acquired first or second common neighbors around the time of the new link formation (Figure 2(b)). Thus, link addition is correlated in both time and space, as links tend to be added in batches over a short period within small regions. While the production of dense clusters in networks has been previously shown (Gallos et al., 2012), the level of clustering presented here is surprisingly high.

To confirm that network change events indeed occur in bursts, and to estimate the appropriate time scale determining the changes in degree and clustering, we replicated the analysis using the second dataset. This dataset was collected through a series of 2-step snowball sampling procedures initiated at 359 random seed nodes, resulting in more than 2,000,000 nodes. This process was repeated at 12-hour intervals (Table 1). Application of the analysis described above to this high-frequency dataset confirms that social network users tend to add links in temporal bursts confined to 12 hours, with results equivalent to those obtained with the longer time scale (i.e. a slow increase before and after the addition of the new link, and a sharp increase in all measures above concurrent with the link addition; Figure 3(a)–(c)). Note that the analysis was repeated for different focal snapshots (the period when links were added) to dismiss possible effects of a specific interval or exogenous event. We scale all plots by their value at the time of link creation ($t = 0$), as in Figure 2(a). The analysis of the high-frequency dataset confirms that new ties tend to be created in batches, in line with previous research on bursts of general human activity (e.g. Muchnik et al., 2013). We will further show that these batches represent local densification events.

3.3 Formation of clusters (communities) through localized network densification

A community is typically defined as a set of nodes with a comparatively high density of links between its members and significantly weaker connections to the rest of the network (Clauset et al., 2004; Eriksen et al., 2003; Holme et al., 2003). In this section, we show that the correlated bursts of link addition described above contribute to community formation through a sharp increase of local density of links among the nodes affected by the activity burst. Assuming that all nodes in the affected region

Fig. 3. *Continued*. A similar pattern is exhibited by the change in the number of the nodes' common neighbors (panel B). In both A and B panels, we scale all plots by their value at the time of link creation ($t = 0$). The final set of plots (panel C) represents the change of the fraction of pairs having at least one first (purple) or second (green) common neighbor and their complementary (neither first nor second common neighbor). A: Error bars represent the standard deviation of the directional degree of the observed nodes. B: Error bars represent the standard deviation of the number of common neighbors of the observed nodes. C: Error bars represent the standard deviation of the binary flag of existence of first or second neighbors for the observed nodes. (Color online)

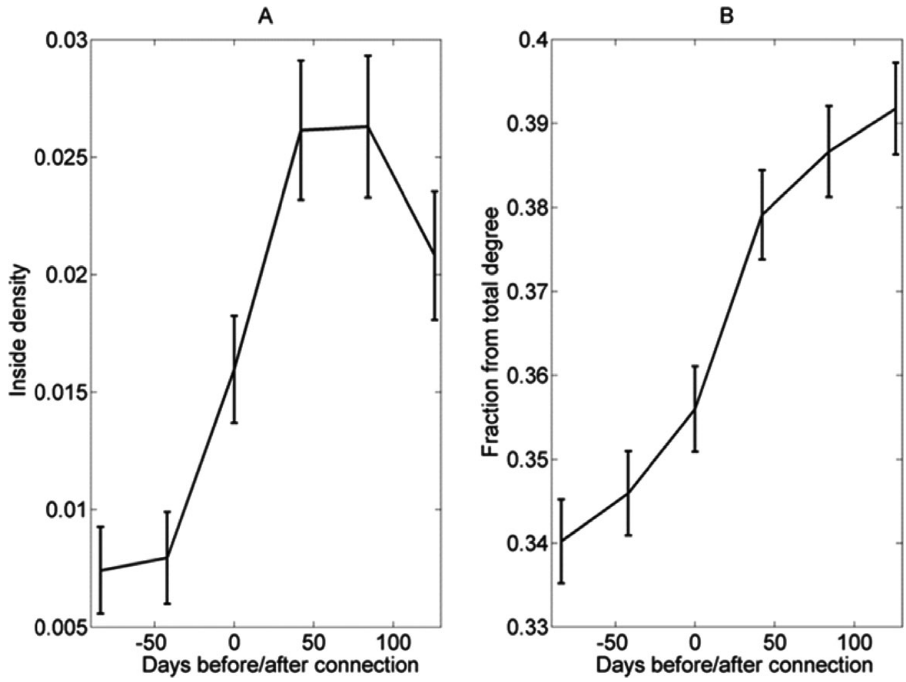

Fig. 4. Evolution of density of the local environment surrounding newly created links (panel A) and fraction of the intra-cluster links out of all links arriving to or originating from the local environment (panel B). This analysis based on sample of 100,000 newly created edges. Error bars represent standard deviation of the fraction of edges fraction of the observed pairs.

become first or second neighbors of the newly connected nodes (i.e. the radius of the cluster is limited to three), we traced the temporal evolution of the fraction of links within the neighborhood of the newly formed link, as defined below, and the fraction of links between this region and the outside.

We start by defining a *local environment* as a set of nodes that are either first or second neighbors of either of the two newly connected nodes. We designate links fully residing within the local environment as *intra-cluster links,* and links connecting nodes within the local environment to the rest of the network as *external links*. The local environment link *density* is the fraction of the existing intra-cluster links out of all possible links between the local environment members ($N_1(N_1 - 1)/2$, with N_1 being the number of nodes in the local environment). The local environment, its density, and the number of intra-cluster links and external links obviously change from one snapshot to another. Newly added links may introduce new nodes into the cluster, while removed existing links may detach old local environment members.

As expected, both the number and, what is more important, the density of the intra-cluster links increase over time (Figure 4(a)). While the total number of links in the local environment increases, as discussed, so does the total number of nodes in this region, since new nodes become first and second neighbors of the pair formed by the newly added link. Still, the density of links within the local environment surrounding the newly formed pair increases significantly (Figure 4(a)).

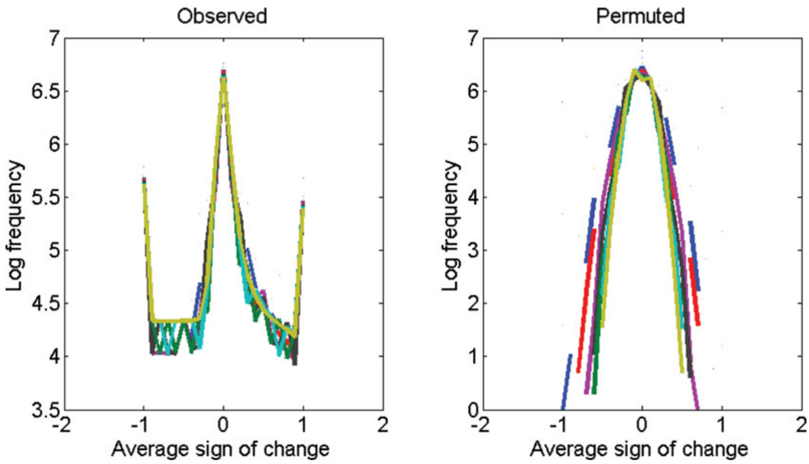

Fig. 5. Histogram of sign change in all observations in the second (fast) dataset. The lines represent different smoothing windows. We computed for each node the sign of its degree change in each time point (82 time points) and obtained series of −1,0,1 values. We then smoothed these series with square windows of width 3 to 20 and computed the histogram of values in the smoothed series (left panel). The same analysis was performed when the time series were permuted (right panel). The curves present the 10 base log of the histogram. While in the real time series, values of 1 and −1 are grouped, leading to extreme values in the histogram (−1 and 1), in the permuted series, the histogram converges to a Gaussian. (Color online)

Even more interestingly, the fraction of the intra-cluster links out of all links in the local environment grows significantly during this period (Figure 4(b)). Thus, densification is accompanied by increasing isolation of the active area from the remainder of the network. In other words, bursts and localization of new link creation in social networks lead to formation of well-pronounced clusters.

3.4 Additional measures of activity burst

A clear indication of activity bursts would be a correlation between changes in the in-degree and out-degree of nodes. While the absolute in- and out-degrees have previously been reported to be correlated (Brot et al., 2015; Fagiolo, 2007; Foster et al., 2010), one would not expect a correlation in the changes in the two types of degrees, unless link addition occurs in bursts. We tested the changes in degree over 82 snapshots of more than 700,000 nodes in the second dataset, and the Pearson correlation of the changes in the in- and out-degrees is more than 0.5 ($p < 1.e\text{-}100$). To further check whether peaks of increase in the in-degree coincide with peaks of increase in the out-degree, we aligned the time series of all nodes based on the peak of the change in their in-degree and computed the average out-degree change. We then performed the reverse analysis (Figure 5). Both plots show a complete alignment of the peaks in the change in the in- and out-degrees. Note that this is not the result of drastic changes in the number of links added to the network between any pair of snapshots.

Finally, we computed for each node and for each time point whether its degree increased, decreased or stayed the same over 82 time points (81 observations—each observation is the difference between two of the 82 snapshots). If link addition and removal events were clustered in time, then one would expect continuous stretches of degree increase or decrease. To test for those, we assigned each increase event a value of 1, and each decrease event a value of −1 for both the in- or out-degree of each node. We then computed the frequency of distribution of these values averaged over multiple time points. As expected from the aggregation of events, even when averaged over 20 time points, there were still many cases of continuous growth. In order to confirm statistical significance of these findings, we compare them to a null model based on the same analysis performed on permutations of the same time series. We directly studied the same time series for burstiness. Each set of bursts was assigned a value of +1, 0, or −1, representing the sign of the change. While in the real data there is a large fraction of windows with all +1 or −1 values even for large smoothing radii. However, the permuted data are used, the data converge to a bell-shaped distribution around 0 (Figure 6).

3.5 Coupling between network evolution and homophily

We have shown that densification occurs in localized bursts, inducing regions with high internal density and a lower density of links to more distant nodes. This purely network perspective of the process is complemented in this section by the study of how content evolution relates to the localized activity bursts and network densification discussed earlier. In addition to a set of structural properties (e.g., degree distribution and clustering) shared by many social networks, such networks are also known to exhibit homophily, or the tendency of similar people to be connected or located in relative proximity. Such irregular distributions (content similarity) might arise from peer *influence*-when people change their interests due to interactions with their peers-or *selection*-strategic choices of peers having similar background, exhibiting similar behavior, or declaring overlapping interests (Aral et al., 2009; 2013; Manski, 1993; McPherson et al., 2001).

In this section, we show that in addition to local densification, activity bursts coincide with a sharp increase of similarity between the nodes in the affected area. To test this conjecture, we collected detailed information about user profiles, specifically their self-reported interests. These observations facilitate examination of the interplay between network changes and the increased similarity of the affected users' interests. More specifically, we tested whether network densification events are induced (and preceded) by changes in content similarity or actually induce (and precede) them. Note that time precedence is obviously not evidence of causality and cannot reveal which of the processes, selection or peer influence, dominates network evolution dynamics. However, the opposite—that causality requires precedence—is often true.

As was the case for the other features, we find not only that densification of the network and increasing similarity coincide, but the two occur virtually simultaneously, so that neither precedes the other. To refine the resolution of our examination and rule out issues related to insufficient sampling frequency, we confirmed these results using the high sampling rate dataset. Both sets of measurements indicate tight coupling between network densification and increased similarity (see Figure 7). In our opinion, the simplest and most straightforward explanation for close co-evolution is that it is driven by strong feedback between content and structure. Prior research

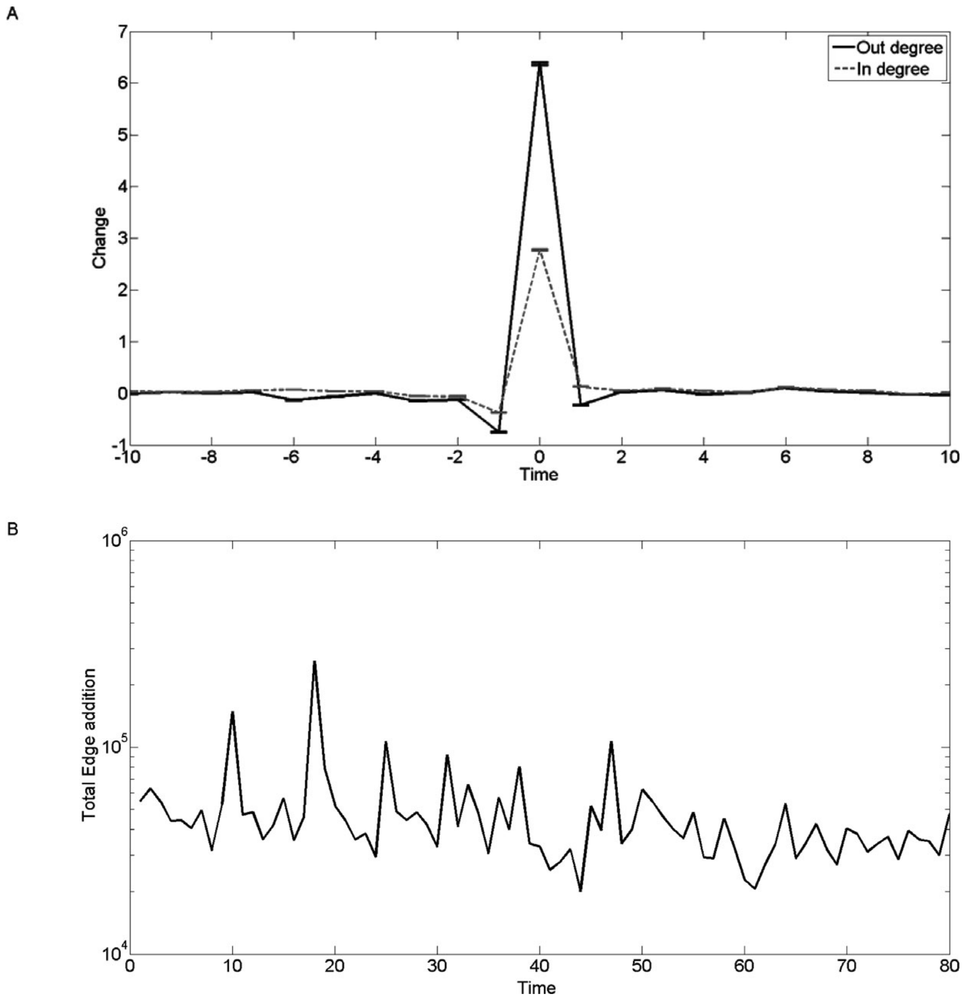

Fig. 6. Correlation between in- and out-degree change. Following the results in Figure 5, we computed for each node the peak of the in-degree change and set it as the 0 point for each node (this time differs for each node). We then averaged the change of the out-degree over all nodes aligned to this 0 point. One can see a clear peak in the out-degree change when the in-degree rises the most (panel A). We then repeated the opposite analysis with similar results (dashed lines in panel A). In order to check that the results are not an artifact of a large number of links added in a given time point, we computed the total number of links added between each two snapshot (panel B), and the results are quite uniform. Thus, a parallel addition of incoming and outgoing edges occurs around active nodes.

(Brot et al., 2012) showed that such feedback can explain many network properties, including all fundamental social network attributes (power-law degree distribution, high clustering, and short path lengths), which is also the case here. An alternative explanation suggests that densification of structure and increased homophily are both induced by external events simultaneously affecting both processes. Being unable to observe exposure of individual network users to all the external events, we cannot dismiss the exogenous source of the observed network-content coupling.

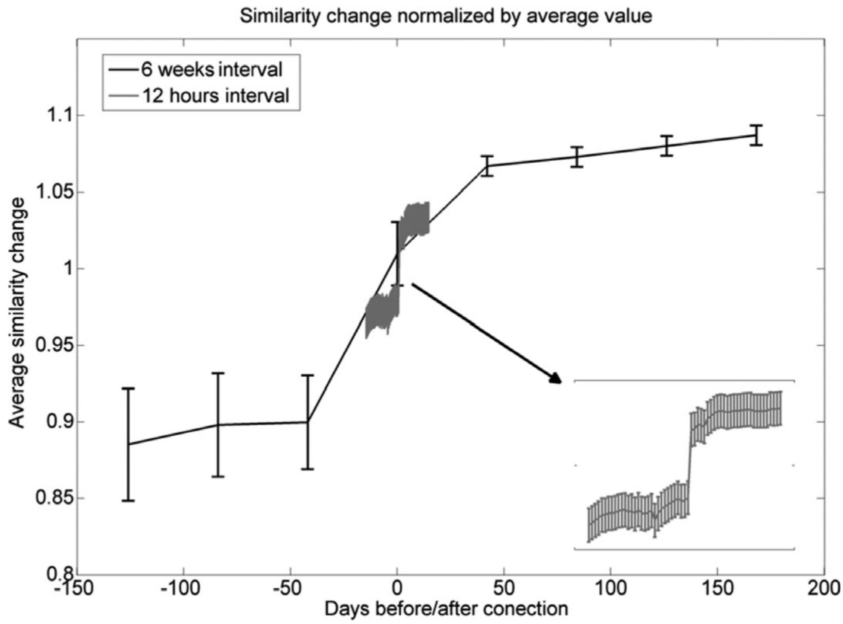

Fig. 7. Normalized change in similarity over time. The plot shows the change in similarity of the connected nodes for the complete dataset (main plot) and the high-frequency dataset (inset). Average degree and similarity are normalized by their value at the time of link creation. Error bars represent standard deviation computed for the multiple instances observed in each dataset. Analogous to network densification, the gradual increase in similarity spikes around the time of link creation and changes to slow growth once the burst of activity is over.

4 Discussion

We have shown that the activity of adding or removing links has a highly pronounced local and temporal nature. Newly emerging network links are not sporadically spread over the network, as (in many cases, implicitly) assumed by most current network models; instead they are localized in both (network) "space" and time in the form of short bursts occurring in network neighborhoods confined to a three-hop radius. Such densification bursts result in relative isolation of the affected network areas-within brief periods of time they increase in size and become significantly denser and relatively more isolated from the remainder of the network. At the aggregate level, these bursts are difficult to spot in large networks because they occur continuously in different regions of the network. This mechanism has implications for network stability, information dissemination, and forecast reliability.

We further investigated the relationship between network evolution and changes in the underlying content. The two processes are found to be tightly coupled and co-evolve so closely that neither can be considered to precede the other. These observations shed light on the evolution of homophily in social networks and suggest that both selection and influence processes are in play. It is also important to note that similarity between peers does not evolve gradually as one might naively expect, but increases in short bursts coinciding with the evolution of the surrounding network. These findings have implications for our understanding of the rate and manner in which social networks evolve and adjust to change.

The research is naturally subject to several limitations. The first is the use of snapshots instead of continuous tracking. We addressed this limitation by collecting a high-frequency sample, but unlike the main dataset, this high-frequency sample is incomplete. The second limitation relates to the lack of data on the events exogenous to the network. As we noted, in principle, such events could drive the network and the content. Although we cannot rule out this option empirically, the large number of coordinated bursts traced in the data in different periods casts doubt that they all result from such exogenous events. Access to such data and the absence of their correlation with the network regions affected by bursts might rule out this explanation entirely.

Future research may help us understand whether, in view of these dramatic local changes, different centrality measures adequately capture the status of individual nodes. We believe that such burst dynamics may play a significant role in community formation in a network. However, this question is beyond the scope of the present study.

References

Adamic, L. A., & Huberman, B. A. (2000). Power-law distribution of the World Wide Web. *Science*, **287**(5461), 2115–2115.

Aiello, L. M., Barrat, A., Schifanella, R., Cattuto, C., Markines, B., & Menczer, F. (2012). Friendship prediction and homophily in social media. *ACM Transactions on the Web (TWEB)*, **6**(2), 9.

Albert, R., Jeong, H., & Barabási, A.-L. (1999). Internet: Diameter of the world-wide web. *Nature*, **401**(6749), 130–131.

Anderson, R. E., & Srinivasan, S. S. (2003). E-satisfaction and e-loyalty: A contingency framework. *Psychology & Marketing*, **20**(2), 123–138.

Aral, S., Muchnik, L., & Sundararajan, A. (2009). Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences*, **106**(51), 21544–21549.

Aral, S., Muchnik, L., & Sundararajan, A. (2013). Engineering social contagions: Optimal network seeding in the presence of homophily. *Network Science*, **1**(2), 125–153.

Barabasi, A.-L. (2005). The origin of bursts and heavy tails in human dynamics. *Nature*, **435**(7039), 207–211.

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, **286**(5439), 509–512.

Brot, H., Muchnik, L., Goldenberg, J., & Louzoun, Y. (2012). Feedback between node and network dynamics can produce real world network properties. *Physica A: Statistical Mechanics and its Applications*, **391**(24), 6645–6654.

Brot, H., Muchnik, L., & Louzoun, Y. (2015). Directed triadic closure and edge deletion mechanism induce asymmetry in directed edge properties. *The European Physical Journal B*, **88**(1), 1–11.

Centola, D., & van de Rijt, A. (2015). Choosing your network: Social preferences in an online health community. *Social Science & Medicine*, **125**, 19–31.

Chaintreau, A., Mtibaa, A., Massoulie, L., & Diot, C. (2007). The diameter of opportunistic mobile networks. Paper presented at the Proceedings of the 2007 ACM CoNEXT conference. New York, NY: ACM.

Clauset, A., Newman, M. E. J., & Moore, C. (2004). Finding community structure in very large networks. *Physical Review E*, **70**(6), 066111.

Cowley, M. A., Smart, J. L, Rubinstein, M., Cerdán, M. G., Diano, S., Horvath, T. L., . . . Low, M. J. (2001). Leptin activates anorexigenic POMC neurons through a neural network in the arcuate nucleus. *Nature*, **411**(6836), 480–484.

Crandall, D., Cosley, D., Huttenlocher, D., Kleinberg, J., & Suri, S. (2008). Feedback effects between similarity and social influence in online communities. Paper presented at the Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining. New York, NY: ACM.

Dezsö, Z., Almaas, E., Lukács, A., Rácz, B., Szakadát, I., & Barabási, A.-L. (2006). Dynamics of information access on the web. *Physical Review E*, **73**(6), 066132.

Diesmann, M., Gewaltig, M.-O., & Aertsen, A. (1999). Stable propagation of synchronous spiking in cortical neural networks. *Nature*, **402**(6761), 529–533.

Diesner, J. (2012). Uncovering and managing the impact of methodological choices for the computational construction of socio-technical networks from texts, DTIC Document, Dissertations. Paper 194.

Eriksen, K. A., Simonsen, I., Maslov, S., & Sneppen, K. (2003). Modularity and extreme edges of the Internet. *Physical Review Letters*, **90**(14), 148701.

Fagiolo, G. (2007). Clustering in complex directed networks. *Physical Review E*, **76**(2), 026107.

Figueiredo, F., Almeida, J. M., Gonçalves, M. A., & Benevenuto, F. (2014). On the dynamics of social media popularity: A YouTube case study. *ACM Transactions on Internet Technology (TOIT)*, **14**(4), 24.

Foster, J. G, Foster, D. V., Grassberger, P., & Paczuski, M. (2010). Edge direction and the structure of networks. *Proceedings of the National Academy of Sciences*, **107**(24), 10815–10820.

Fu, F., Nowak, M. A., Christakis, N. A., & Fowler, J. H. (2012). The evolution of homophily. *Scientific Reports 2*, doi:10.1038/srep00845.

Gabel, A., & Redner, S. (2013). Sublinear but never superlinear preferential attachment by local network growth. *Journal of Statistical Mechanics: Theory and Experiment*, **2013**(2), P02043.

Gallos, L. K., Rybski, D., Liljeros, F., Havlin, S., & Makse, H. A. (2012). How people interact in evolving online affiliation networks. *Physical Review X*, **2**(3), 031014.

Garber, T., Goldenberg, J., Libai, B., & Muller, E. (2004). From density to destiny: Using spatial dimension of sales data for early prediction of new product success. *Marketing Science*, **23**(3), 419–428.

Golosovsky, M., & Solomon, S. (2012). Stochastic dynamical model of a growing citation network based on a self-exciting point process. *Physical Review Letters*, **109**(9), 098701.

Grund, T. (2014). Network size and network homophily: Same-sex friendships in 595 Scandinavian schools. *Analytical Sociology*, 297–314.

Hamers, L., Hemeryck, Y., Herweyers, G., Janssen, M., Keters, H., Rousseau, R., & Vanhoutte, A. (1989). Similarity measures in scientometric research: The Jaccard index versus Salton's cosine formula. *Information Processing & Management*, **25**(3), 315–318.

Holme, P., Huss, M., & Jeong, H. (2003). Subnetwork hierarchies of biochemical pathways. *Bioinformatics*, **19**(4), 532–538.

Holme, P., & Kim, B. J. (2002). Growing scale-free networks with tunable clustering. *Physical Review E*, **65**(2), 026107.

Jagerman, D. L., & Melamed, B. (1994). Burstiness descriptors of traffic streams: Indices of dispersion and peakedness. *Proceedings of the 1994 Conference on Information Sciences and Systems*, **1**, 24–28.

Kämpf, M., Tismer, S., Kantelhardt, J. W., & Muchnik, L. (2012). Fluctuations in Wikipedia access-rate and edit-event data. *Physica A: Statistical Mechanics and its Applications*, 391(23), 6101-6111. doi: http://dx.doi.org/10.1016/j.physa.2012.07.004.

Khosla, A., Das Sarma, A., & Hamid, R. (2014). What makes an image popular? Paper presented at the Proceedings of the 23rd international conference on World wide web. New York, NY: ACM.

Kimura, D., & Hayakawa, Y. (2008). Coevolutionary networks with homophily and heterophily. *Physical Review E*, **78**(1), 016103.

Kleinberg, J. M., Kumar, R., Raghavan, P., Rajagopalan, S., & Tomkins, A. S. (1999). The web as a graph: Measurements, models, and methods. In: T. Asano, H. Imai, D. T. Lee, S. Nakano, & T. Tokuyama (Eds.), *Computing and Combinatorics Volume 1627 of the series Lecture Notes in Computer Science* (pp. 1–17). Berlin, Heidelberg: Springer.

Klemm, K., & Eguiluz, V. M. (2002). Highly clustered scale-free networks. *Physical Review E*, **65**(3), 036123.

Leskovec, J., Backstrom, L., Kumar, R., & Tomkins, A. (2008). Microscopic evolution of social networks. Paper presented at the Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining. New York, NY: ACM.

Leskovec, J., Kleinberg, J., & Faloutsos, C. (2005). Graphs over time: Densification laws, shrinking diameters and possible explanations. Paper presented at the Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining. New York, NY: ACM.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, **60**(3), 531–542.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, **27**, 415–444.

Min, B., & Goh, K.-I. (2013). Burstiness: Measures, models, and dynamic consequences. In: P. Holme & J. Saramäki (Eds.), *Temporal networks* (pp. 41–64). Berlin, Heidelberg: Springer.

Muchnik, L., Pei, S., Parra, L. C., Reis, S. D. S., Andrade, J. S. Jr, Havlin, S., & Makse, H. A. (2013). Origins of power-law degree distribution in the heterogeneity of human activity in social networks. *Scientific Reports*, 3, doi:10.1038/srep01783.

Papadopoulos, F., Kitsak, M., Serrano, M. Á., Boguñá, M., & Krioukov, D. (2012). Popularity versus similarity in growing networks. *Nature*, **489**(7417), 537–540.

Paxson, V., & Floyd, S. (1995). Wide area traffic: The failure of poisson modeling. *IEEE/ACM Transactions on Networking (ToN)*, **3**(3), 226–244.

Perc, M. (2014). The matthew effect in empirical data. *Journal of The Royal Society Interface*, **11**(98), 20140378.

Roth, C., & Cointet, J.-P. (2010). Social and semantic coevolution in knowledge networks. *Social Networks*, **32**(1), 16–29.

Rybski, D., Buldyrev, S. V., Havlin, S., Liljeros, F., & Makse, H. A. (2009). Scaling laws of human interaction activity. *Proceedings of the National Academy of Sciences*, **106**(31), 12640–12645.

Sayama, H., Pestov, I., Schmidt, J., Bush, B. J., Wong, C., Yamanoi, J., & Gross, T. (2013). Modeling complex systems with adaptive networks. *Computers & Mathematics with Applications*, **65**(10), 1645–1664.

Vázquez, A. (2003). Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. *Physical Review E*, **67**(5), 056104.

Yamasaki, K., Muchnik, L., Havlin, S., Bunde, A., & Stanley, H. E. (2005). Scaling and memory in volatility return intervals in financial markets. *Proceedings of the National Academy of Sciences of the United States of America*, **102**(26), 9424–9428.

Appendix A

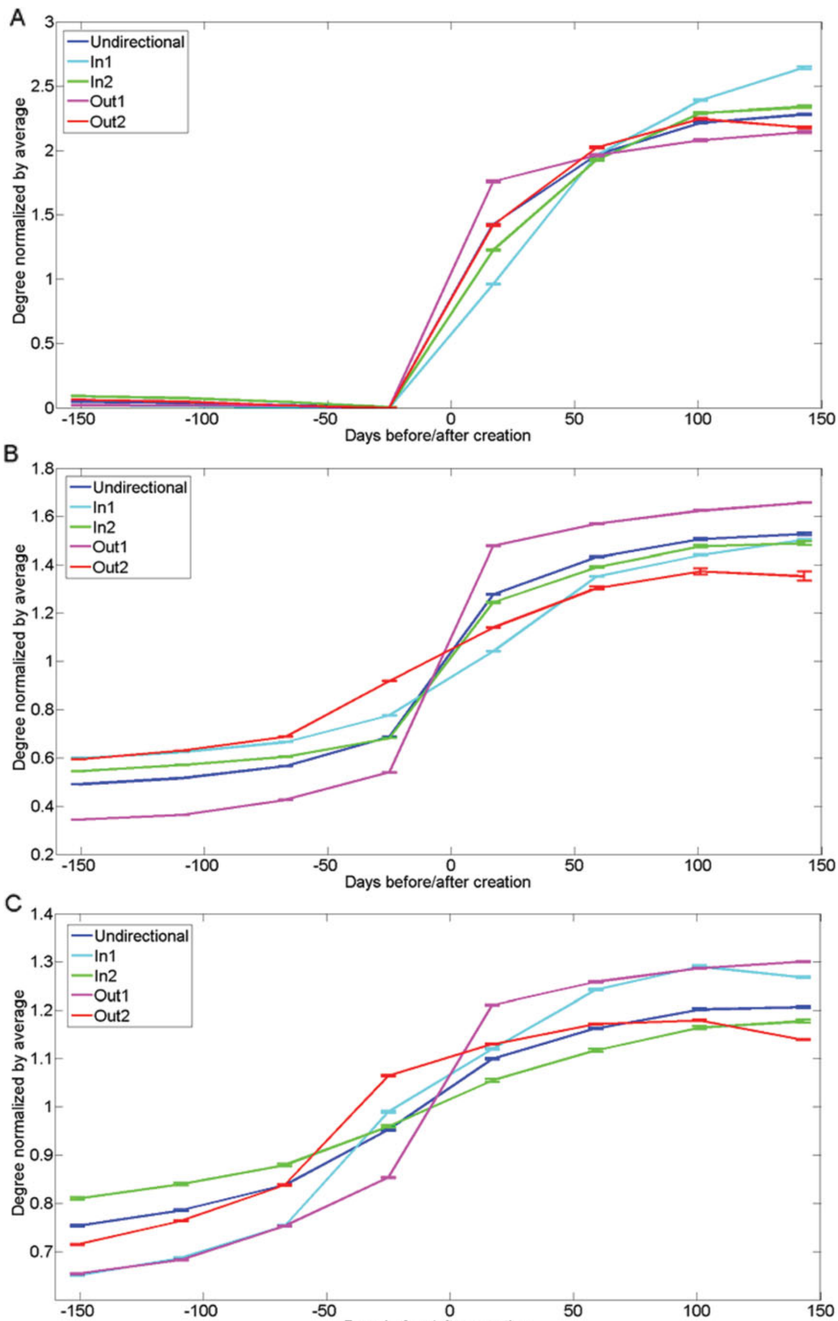

Fig. A1. Evolution of node attributes before, during, and after they connect for three different groups of nodes according to their initial degree. We divided the group of the newly connected nodes (00001111) in the first dataset (snapshots taken at 6-week intervals) into three groups according to the total degree of the combined nodes' pair: high degree, above 900 (panel A); low degree, below 5 (panel B); and intermediate degree, below 900 and above 4 (panel C). We computed the in- and out-degree of each node surrounding the newly added link in the

Fig. A1. *Continued*. same way as in Figure 2(a). We scale all plots by their value at the time of link creation ($t = 0$). These plots show qualitatively similar results for all groups as well as for general groups of newly connected nodes, as presented in Figure 2(a). There is a slow change in the months preceding and following the connection, then we see rapid dramatic changes coinciding with the addition of the new link (time 0). The error bars in all panels represent standard deviation of the directional degree of the observed nodes. (Color online)

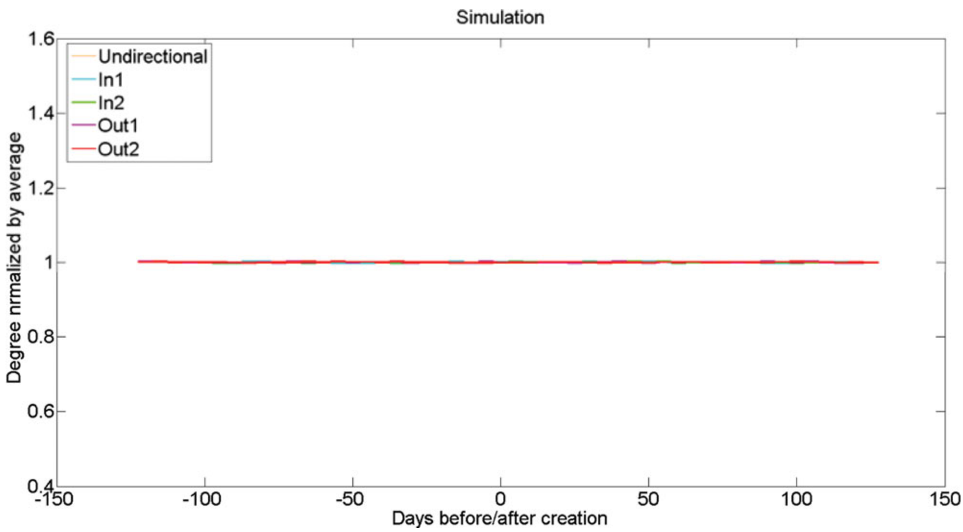

Fig. A2. Evolution of node attributes before, during, and after they connect for an artificial network with simple preferential attachment mechanism. The network contains 5000 nodes with average degree of 100. In each iteration, new edges are created according to a linear preferential attachment rule and a similar number of edges is deleted randomly. We computed the in- and out-degree of each node surrounding the newly added link in this network in the same way as in Figure 2(a). We scale all plots by their value at the time of link creation ($t = 0$). This plot shows that there is no significant change in the different degrees before and after an edge creation event in this system. (Color online)