

## Betweenness Preference: Quantifying Correlations in the Topological Dynamics of Temporal Networks

René Pfitzner,<sup>\*</sup> Ingo Scholtes,<sup>†</sup> Antonios Garas,<sup>‡</sup> Claudio J. Tessone,<sup>§</sup> and Frank Schweitzer<sup>||</sup>

*ETH Zurich, Chair of Systems Design, Weinbergstrasse 56/58, 8092 Zurich, Switzerland*

(Received 8 August 2012; revised manuscript received 19 March 2013; published 10 May 2013)

We study correlations in temporal networks and introduce the notion of betweenness preference. It allows us to quantify to what extent paths, existing in time-aggregated representations of temporal networks, are actually realizable based on the sequence of interactions. We show that betweenness preference is present in empirical temporal network data and that it influences the length of the shortest time-respecting paths. Using four different data sets, we further argue that neglecting betweenness preference leads to wrong conclusions about dynamical processes on temporal networks.

DOI: [10.1103/PhysRevLett.110.198701](https://doi.org/10.1103/PhysRevLett.110.198701)

PACS numbers: 89.75.Hc, 05.40.Fb, 89.65.-s

Recent works have argued that properties of dynamical processes evolving on complex networks change profoundly when the dynamics of the network topology is taken into account. For a number of empirical temporal networks obtained from time-stamped contact data, simulations have shown that their topological dynamics can both slow down [1–3] or speed up spreading processes [4]. At the same time, it has been observed that, compared to time-aggregated topologies, the exploration dynamics of random walks in temporal networks is significantly slower [5]. Furthermore, it has been shown that network dynamics alone can give rise to collective phenomena like synchronization [6]. These observations have generated significant interest in the mechanism underlying these phenomena. A series of recent works focused on the influence of interevent time distributions and temporal correlations in the time series of interactions [3,5,7–9]. Bursty activity patterns of nodes have been identified as one possible source that slows down spreading [3] and random walk processes [10]. Similarly, bursty node activities have been suggested to slow down information diffusion, particularly when the diffusion process is initiated in phases of low activity [11]. Furthermore, for a number of social contact networks, it has been shown that heterogeneous interevent times increase the length of time-respecting paths [12]. Apart from interevent time distributions, it has been argued that link appearance frequencies and their correlation with community structures are another characteristic of temporal networks that can slow down spreading dynamics [3]. Another line of research is concerned with the study of temporal motifs [13,14], i.e., whether there are classes of frequently occurring temporal contact patterns. It was shown that the presence of certain temporal motifs (like, e.g., “chains” of consecutive edges continuing time-respecting paths) can *decrease* the length of time-respecting paths, thus speeding up spreading processes [12].

Going beyond previous works, in this Letter, we study betweenness preference, which captures the over- or underrepresentation of particular time-respecting paths

passing through nodes. This temporal-topological feature is neither visible in the weighted time-aggregated network nor can it be attributed to interevent time distributions, bursty node dynamics, or the statistics of temporal motifs. Our study is motivated by the idea that in many real-world networks nodes contact other nodes based on previous contacts [9]. As a simple example, consider the influence of context in information dissemination: Emails received by work-related contacts are more likely to be forwarded to a work-related subset of social contacts. Here we study the influence of such special classes of dynamical contact patterns. We argue that betweenness preference, i.e., the tendency of nodes to preferentially connect—in a temporal sense—particular pairs of neighbors, (i) is not captured in the time-aggregated network, (ii) is present in empirical temporal network data, (iii) changes the topology of time-respecting paths, and (iv) critically influences dynamical processes evolving on temporal networks.

A temporal network is defined as a tuple consisting of a set of nodes  $v \in \mathcal{V}$  as well as a set of events:  $e(v, w, t, l\Delta t) \in \mathcal{E}$ . An event is an interaction between two nodes  $v$  and  $w$ , starting at time  $t$  and with a duration  $l\Delta t$  relative to some smallest unit of discrete time  $\Delta t$  (for simplicity, we assume  $\Delta t = 1$ ). Based on time-stamped edges and a discrete notion of time, we construct a flow-preserving static representation of temporal networks by unfolding time into an additional topological dimension. This construction serves as the basis for our models, and we call it a time-unfolded network. Time-unfolded networks of two different temporal networks are illustrated in Fig. 1. In the resulting temporal unfolding, we indicate the presence of a possible flow event by an edge  $(v_t, w_{t+1})$  while replacing the original node set  $\mathcal{V}$  by a set  $\mathcal{V}'$  of temporal copies of nodes  $v_t$ , where  $v \in \mathcal{V}$  and  $t \in \{0, 1, \dots, L\}$  for an observation period of duration  $L$ , similar to [15–17]. As shown in Fig. 1 (left), the two different temporal networks are the same in the time-aggregated representation  $G_{\text{Agg}}$  (right). In  $G_{\text{Agg}}$ , edge weights indicate the number  $l$  of time steps in which an edge was active

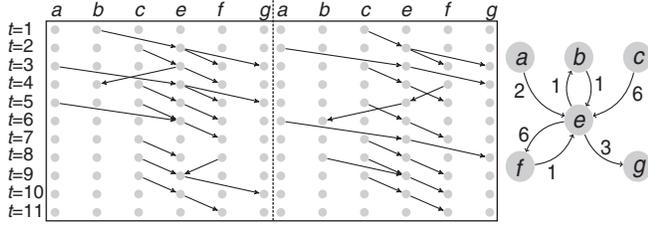


FIG. 1. Time-aggregated weighted network  $G_{\text{Agg}}$  (right) and time-unfolded network of two different temporal networks  $G_{\text{Dyn}_1}$  and  $G_{\text{Dyn}_2}$  (left), both of which are consistent with  $G_{\text{Agg}}$ .

throughout the observation period. In analogy to statistical mechanics, one might think of such a time-aggregated network as a macrostate which is compatible with different temporal networks, i.e., microstates.

*Betweenness preference.*—An important aspect when studying dynamical processes like diffusion or synchronization on static, time-aggregated networks is that one assumes transitive paths. However, this transitivity does not necessarily hold in a temporal network that gives rise to the respective time-aggregated network [18]. To illustrate this fact, consider the time-aggregated network  $G_{\text{Agg}}$  depicted in Fig. 1. In a static system, information could spread in a transitive way from  $b$  via  $e$  to  $g$ . In a temporal network, the order in which edges appear imposes an additional constraint: Information can flow only along time-respecting paths [18]. Hence, in a temporal network underlying  $G_{\text{Agg}}$ , information can flow from  $b$  to  $g$  only if the connection  $(b, e)$  is followed by a connection  $(e, g)$ . Thus, even though the links  $(b, e)$  and  $(e, g)$  are present in both temporal networks (Fig. 1, left), a time-respecting path between  $b$  and  $g$  exists only in the left example. To quantify this transitivity-limiting property of temporal networks, we study whether certain time-respecting paths are preferentially realized as compared to the time-aggregated perspective. Focusing on elementary building blocks of time-respecting paths, we particularly study two-paths, i.e., a path of length two, representing two consecutive edge activations interconnecting three nodes. The statistics of the ensemble of realized two-paths will reveal to what extent path transitivity holds in the temporal network. Based on the time-unfolded representation of temporal networks, we define the elements of a per-node betweenness preference matrix  $\mathbf{B}^v(t)$  as follows:

$$B_{sd}^v(t) := \begin{cases} 1, & \text{if } (s_{t-1}, v_t) \in \mathcal{E} \text{ and } (v_t, d_{t+1}) \in \mathcal{E}, \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

Each matrix element  $B_{sd}^v(t)$  captures whether node  $v_t$  in a time-unfolded network was in between a source  $s_{t-1}$  and a destination  $d_{t+1}$  on a two-path  $(s_{t-1}, v_t) \rightarrow (v_t, d_{t+1})$ . This definition builds on a notion of time-respecting paths comprised of edge activations following each other *immediately*, which can be relaxed though by including some notion of memory. Based on definition (1), we define the

elements of a time-aggregated betweenness preference matrix  $\mathbf{B}^v$ :

$$B_{sd}^v := \sum_t B_{sd}^v(t) \left[ \sum_{s'd'} B_{s'd'}^v(t) \right]^{-1}, \quad (2)$$

and a normalized betweenness preference matrix  $\mathbf{P}^v$ :

$$P_{sd}^v := B_{sd}^v \left[ \sum_{s'd'} B_{s'd'}^v \right]^{-1}. \quad (3)$$

Essentially,  $P_{sd}^v$  is the probability distribution of the two-paths  $(s_{t-1}, v_t) \rightarrow (v_t, d_{t+1})$  over all  $t$ . We use this to quantify to what extent  $v$  exhibits a preference to interconnect particular pairs of source and target nodes. Based on the concept of mutual information, we define a betweenness preference measure as

$$I^v(S; D) := \sum_{\substack{d \in D \\ s \in S}} P_{sd}^v \log_2 \left( \frac{P_{sd}^v}{P^v(s)P^v(d)} \right), \quad (4)$$

where  $P^v(s) = \sum_d P_{sd}^v$  and  $P^v(d) = \sum_s P_{sd}^v$ . In general,  $I^v(S; D)$  captures to what extent the knowledge of the source  $s$  of a time-respecting path through  $v$  determines the next step  $d$ . Or, within a context of information flow, it measures how selective  $v$  is in mediating information preferentially between certain pairs of nodes  $s$  and  $d$ . We note that this measure is minimal if the random variables  $S$  and  $D$  are independent. This allows us to calculate the matrix elements  $P_{sd}^v$  resulting in  $I^v(S; D) = 0$  solely based on the underlying static, time-aggregated network with edge weights  $w_{ij}$ :

$$\hat{P}_{sd}^v := p^v(s) \cdot p^v(d), \quad (5)$$

where  $p^v(s) = w_{sv} [\sum_{s'} w_{s'v}]^{-1}$  and  $p^v(d) = w_{vd} [\sum_{d'} w_{vd'}]^{-1}$ .

We now introduce a configuration model to generate temporal networks that are members of an ensemble defined by a given set of betweenness preference matrices. We limit ourselves to a subset of all possible realizations, in which exactly one edge is active per time step and which consist of *only* two-paths. The model creates a temporal network with given betweenness preference matrices as follows: First, we define the number of two-paths  $N_2$  to be realized. Second, we draw a random two-path  $(s, v) \rightarrow (v, d)$  according to  $p(s, v, d) := B_{sd}^v / \sum_{v, s, d} B_{sd}^v$ . Third, we create temporal network edges  $(s_t, v_{t+1})$  and  $(v_{t+1}, d_{t+2})$ . We increment  $t = t + 3$  and  $n_2 = n_2 + 1$  (number of realized two-paths) and repeat this procedure until  $n_2 = N_2$ . Having empirical temporal network data available, we use this model to create microstates that (i) preserve the betweenness preference distribution (in an infinitely long sequence), (ii) have the same macrostate, and (iii) destroy all other correlations (such as bursty activity patterns). We call the so-constructed temporal network the betweenness preference preserving case. Using this configuration model, we also construct microstates with low betweenness preference based on a given macrostate, by using the

probability  $\hat{P}_{sd}^v$  defined in Eq. (5). We call a temporal network created in such a way the uncorrelated case, since it preserves only the macrostate (the weighted aggregate network) but destroys betweenness preference. Precisely, the uncorrelated case has betweenness preference that is expected from a random microstate of finite duration [in a temporal network of infinite duration, the model reproduces the limiting case of  $I^v(S; D) = 0$ ].

*Empirical results.*—To demonstrate that betweenness preference is an important property in real-world data sets, we use a one-week subset of the empirical contact sequence (September 8–15, 2004) from the RealityMining Project (RM) [19], featuring 64 individuals with 20 000 recorded interactions and a granularity of 5 min. Figure 2 shows the distribution of betweenness preference in (a) the uncorrelated case, (b) the original data sample, and (c) the betweenness preference preserving case. The uncorrelated (a) and betweenness preference preserving (c) cases were created by using the betweenness preference matrices of the RM data, utilizing the configuration model described above. As expected, the uncorrelated case (a) shows a spike around  $I(S; T) = 0$ , indicating small betweenness preference for most nodes. The theoretical expectation of  $I(S; T) = 0$  for all nodes in the uncorrelated case is not realized due to finite duration of the temporal sequence. Analyzing the betweenness preference distribution of the empirical temporal network (b), one realizes that it is very different from the one in (a). The distribution in (b) is rather broad, with an average  $\langle I(S; T) \rangle = 1.9$  and a median  $Q_{0.5}[I(S; T)] = 1.99$ , as compared to  $\langle I(S; T) \rangle = 0.58$  and  $Q_{0.5}[I(S; T)] = 0.21$  in (a). Performing a two-sided Kolmogorov-Smirnov test, we can reject the hypothesis that distribution (b) is identical to distribution (a) ( $p < 10^{-9}$ ). This shows that there is a pronounced amount of betweenness preference in the empirical contact sequence, whereas it is mostly absent in the uncorrelated case. Panel (c) shows the betweenness preference distribution of the betweenness preference preserving case. With  $\langle I(S; T) \rangle = 1.91$  and median  $Q_{0.5}[I(S; T)] = 2.02$ , distributions (b) and (c) are very similar. Since we create the model in a

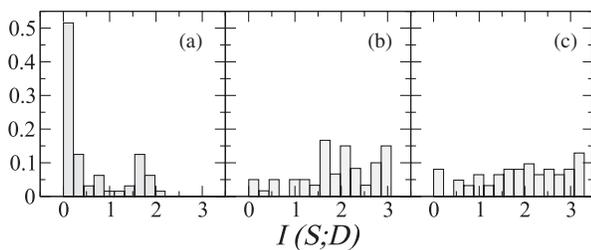


FIG. 2. Betweenness preference distributions for nodes of (a) 200 temporal networks of the uncorrelated case, (b) the empirical RealityMining data, and (c) 200 temporal networks of the  $\mathbf{P}^v$  preserving case. All temporal networks have the same duration of  $N_2 = 5 \times 10^4$  two-paths. (a) and (c) are based on the same empirical data as (b).

statistical fashion based on the normalized betweenness preference matrix  $\mathbf{P}^v$ , the two distributions are not completely identical due to finite  $N_2$ . Performing the two-sided Kolmogorov-Smirnov test, we cannot reject the hypothesis that the two distributions are identical ( $p = 0.68$ ). Hence, in this case the model preserves the betweenness preference of the real network, whereas all other correlations (e.g., bursty node activities) are destroyed by construction.

An important aspect of betweenness preference is that it influences transitivity and, thus, changes the topology of time-respecting paths in temporal networks. Of particular interest in the study of temporal networks is the notion of fastest time-respecting paths. A fastest time-respecting path between nodes  $s$  and  $v$  is defined as the path by which information from node  $s$  first reaches node  $v$  [18]. To quantify the transitivity-limiting effect of betweenness preference, i.e., the under- or overrepresentation of particular two-paths, we study fastest time-respecting paths in temporal networks, where two-paths are the only means of information flow, as follows. We measured the length of the fastest time-respecting path between all pairs of nodes for uncorrelated  $[L_u(s, v)]$  and betweenness preference preserving realizations  $[L_p(s, v)]$ , generated by our configuration model. For  $s = v$ , we set zero path length. We compute the relative length difference  $\delta(s, v) = [L_p(s, v) - L_u(s, v)] / L_u(s, v)$  and average over several realizations. Results for the RM data set are shown in Fig. 3. The left panel shows the distribution of  $\delta(s, v)$  which exhibits a clear shift to positive values. Moreover, the null hypothesis that the distribution of length of fastest time-respecting paths in the betweenness preference preserving and uncorrelated cases are the same can be rejected (two-sided Kolmogorov-Smirnov test yields  $p < 10^{-15}$ ). Thus we conclude that betweenness preference has a profound effect on

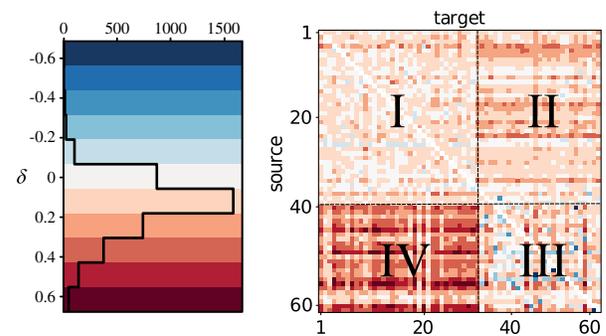


FIG. 3 (color online). Relative difference  $\delta$  between the lengths of fastest time-respecting paths in betweenness preference preserving and uncorrelated temporal networks constructed from the RealityMining data set. Each cell of the matrix represents the length difference for one pair of nodes, averaged over 200 realizations of betweenness preference preserving and uncorrelated temporal networks. Rows and columns are sorted in the same way. Notation I–IV is included to highlight the block structure of the matrix. A histogram of relative differences is overlaid in the color legend (left panel).

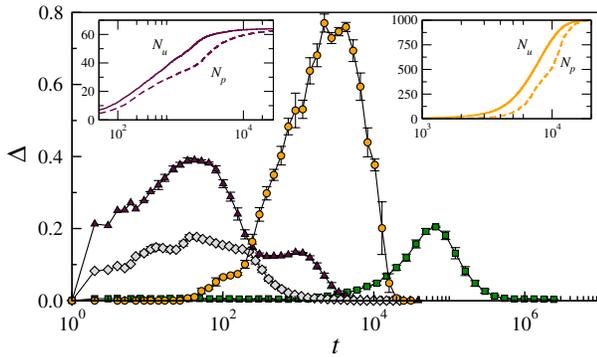


FIG. 4 (color online). SI-spreading dynamics for temporal networks of RealityMining data (RM, purple triangles), the ant network (AN, gray diamonds), the coauthorship network (CN, green squares), and the 1000 node synthetic network (SN, orange circles). The main figure shows time evolution of average relative differences  $\Delta$  ( $10^4$  realizations for CN, 200 for the others) in SI spreading dynamics between the uncorrelated and the betweenness preference preserving cases. Insets (left, RM; right, SN) show the time evolution of the number of infected nodes for the uncorrelated ( $N_u$ ) and betweenness preference preserving cases ( $N_p$ ). Error bars represent standard deviations (when not shown, error bars are smaller than symbol size).

the topology of fastest time-respecting paths in temporal networks ( $\langle\langle\delta(s, v)\rangle\rangle = 0.08$ ,  $Q_{0.5}[\delta(s, v)] = 0.06$ ). In Fig. 3, right panel, we present the matrix  $\delta(s, v)$ , exhibiting a pronounced block structure. In particular, high values of  $\delta(s, v)$  occur in the off-diagonal regions IV and II (albeit less pronounced in the latter). This result implies the existence of a temporal community structure induced by the existence of betweenness preference in the RM data set: Fastest time-respecting paths are up to 60% longer between communities (blocks II and IV), whereas the prolonging effect almost vanishes within the communities (blocks I and III). For some pairs of nodes even the opposite effect is apparent in region III: Betweenness preference selectively shortens fastest time-respecting paths.

*Spreading dynamics in temporal networks.*—Given this result on the effect of betweenness preference on time-respecting paths, we now study the paradigmatic SI (susceptible-infected) epidemic model to exemplify the effect of betweenness preference on dynamic processes evolving on temporal networks. To exclusively quantify the impact of betweenness preference correlations, we compare spreading dynamics on temporal networks created with the uncorrelated case (a) and the betweenness preference preserving case (c) of our configuration model. More specifically, we study four different data sets: (AN) social interactions in ant colonies [20]; (CN) a large temporal network of scientific coauthorships [21]; (RM) the RealityMining data set introduced earlier; and (SN) a medium-sized synthetic network [22]. To start the contagion process, in all simulations the first appearing node has been infected and the infection probability set to  $p = 1$ . The insets in Fig. 4 show the evolution of the number of infected

individuals for the uncorrelated and the betweenness preference preserving cases for the RM and SN data sets. The number of infected individuals clearly follows a typical S-shaped curve (however, notice the semilogarithmic scaling) in the uncorrelated  $N_u$  as well as the betweenness preference preserving  $N_p$  cases. The slopes in the middle part of the infection dynamics are, however, clearly different, indicating slower spreading in the temporal network with nonvanishing betweenness preference  $N_p$ . This slowdown is indicated by the time to saturation, which, in the RM data (Fig. 4, left inset), is 1 order of magnitude larger for  $N_p$ . To substantiate this important effect of large betweenness preference, the main panel shows the time evolution of the relative difference of infected individuals in the uncorrelated and the betweenness preference preserving cases, i.e.,  $\Delta = [N_u(t) - N_p(t)]/N_u(t)$ . Our results clearly show that the uncorrelated model significantly overestimates the average number of infected individuals—for RM at times up to  $\approx 40\%$  and for AN and CN up to  $\approx 20\%$ . Results for the synthetic network with artificially high betweenness preference (Fig. 4, circles and right inset) confirm these findings and show evidence that in large systems high betweenness preference can have an even more pronounced effect on spreading processes: The slowdown of spreading processes can be as large as  $\approx 80\%$ . Interpreting  $\Delta$  as the error made when not accounting for betweenness preference correlations, it becomes obvious that taking a time-aggregated perspective on temporal networks, and hence neglecting betweenness preference, can lead to significantly misleading statements about the properties of dynamical processes evolving on networks with dynamic topology. Betweenness preference, captured in terms of the measure introduced in Eq. (4), quantifies this potential pitfall and helps to decide whether time-respecting paths are statistically distributed as expected from a weighted time-aggregated perspective. Additionally, betweenness preference matrices [Eqs. (2) and (3)] allow us (i) to study the over- or underrepresentation of particular time-respecting paths passing through nodes and (ii) to define proxy models that reproduce the temporal-topological dynamics of empirical temporal networks.

I. S. acknowledges support by the SNF through Grant No. CR12I1\_125298. C. J. T. and A. G. acknowledge support by the SNF through Grant No. 100014\_126865. A. G. and F. S. acknowledge support by the EU-FET Project No. MULTIPLEX 317532.

\*rpfitzner@ethz.ch

†ischoltes@ethz.ch

‡agaras@ethz.ch

§tessonec@ethz.ch

||fschweitzer@ethz.ch

[1] L. Isella, J. Stehlé, A. Barrat, C. Cattuto, J.-F. Pinton, and W. V. den Broeck, *J. Theor. Biol.* **271**, 166 (2011).

- [2] B. Blonder and A. Dornhaus, *PLoS ONE* **6**, e20298 (2011).
- [3] M. Karsai, M. Kivelä, R. Pan, K. Kaski, J. Kertész, A.-L. Barabási, and J. Saramäki, *Phys. Rev. E* **83**, 025102 (2011).
- [4] L. E. C. Rocha, F. Liljeros, and P. Holme, *PLoS Comput. Biol.* **7**, e1001109 (2011).
- [5] M. Starnini, A. Baronchelli, A. Barrat, and R. Pastor-Satorras, *Phys. Rev. E* **85**, 056115 (2012).
- [6] C. J. Tessone and D. H. Zanette, *Europhys. Lett.* **99**, 68006 (2012).
- [7] J. Iribarren and E. Moro, *Phys. Rev. Lett.* **103**, 038702 (2009).
- [8] N. Perra, B. Gonçalves, R. Pastor-Satorras, and A. Vespignani, *Sci. Rep.* **2**, 469 (2012).
- [9] G. Miritello, E. Moro, and R. Lara, *Phys. Rev. E* **83**, 045102 (2011).
- [10] N. Perra, A. Baronchelli, D. Mocanu, B. Gonçalves, R. Pastor-Satorras, and A. Vespignani, *Phys. Rev. Lett.* **109**, 238701 (2012).
- [11] A. Panisson, A. Barrat, C. Cattuto, W. V. den Broeck, G. Ruffo, and R. Schifanella, *Ad Hoc Networks* **10**, 1532 (2012).
- [12] R. Pan and J. Saramäki, *Phys. Rev. E* **84**, 016105 (2011).
- [13] L. Kovanen, M. Karsai, K. Kaski, J. Kertész, and J. Saramäki, *J. Stat. Mech.* (2011) P11005.
- [14] Q. Zhao, Y. Tian, Q. He, N. Oliver, R. Jin, and W.-C. Lee, in *Proceedings of the Nineteenth International Conference on Information and Knowledge Management* (Association for Computing Machinery, New York, 2010), pp. 1645–1648.
- [15] R. Parshani, M. Dickison, R. Cohen, H. E. Stanley, and S. Havlin, *Europhys. Lett.* **90**, 38004 (2010).
- [16] H. Kim and R. Anderson, *Phys. Rev. E* **85**, 026107 (2012).
- [17] B. Mitra, L. Tabourier, and C. Roth, *Comput. Netw.* **56**, 1041 (2012).
- [18] P. Holme and J. Saramäki, *Phys. Rep.* **519**, 97 (2012).
- [19] N. Eagle, A. S. Pentland, and D. Lazer, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 15274 (2009).
- [20] We consider the temporal network for one ant colony, consisting of 81 individuals and 475 interactions [2].
- [21] Collected from a scholarly database, it consists of  $10^4$  authors and  $1.8 \times 10^6$  coauthorship relations. It spans the complete history of papers published within the domain “Condensed Matter Physics” from 1926 to 1999.
- [22] The network has similar modularity, cluster structure, and density as the RM data, but with 1000 nodes, and the nodes additionally have an artificially high betweenness preference (a procedure based on [23]).
- [23] M. E. J. Newman and M. Girvan, *Phys. Rev. E* **69**, 026113 (2004).