Modeling Temporal Networks Using Random Itineraries

Alain Barrat,[1,2] Bastien Fernandez,[1] Kevin K. Lin,[3] and Lai-Sang Young[4]

[1]*Centre de Physique Théorique, Aix-Marseille Université, CNRS UMR 7332, Université de Toulon, 13288 Marseille cedex 9, France*
[2]*Data Science Lab, Institute for Scientific Interchange (ISI) Foundation, Torino 10126, Italy*
[3]*Department of Mathematics and Program in Applied Mathematics, the University of Arizona, Tucson, Arizona 85721, USA*
[4]*Courant Institute of Mathematical Sciences, New York University, New York, New York 10012, USA*
(Received 28 January 2013; published 8 April 2013)

We propose a procedure to generate dynamical networks with bursty, possibly repetitive and correlated temporal behaviors. Regarding any weighted directed graph as being composed of the accumulation of paths between its nodes, our construction uses random walks of variable length to produce time-extended structures with adjustable features. The procedure is first described in a general framework. It is then illustrated in a case study inspired by a transportation system for which the resulting synthetic network is shown to accurately mimic the empirical phenomenology.

DOI: 10.1103/PhysRevLett.110.158702 PACS numbers: 89.75.Hc, 02.50.Ey, 05.40.Fb

Many systems in nature or related to human activities are conveniently represented as networks of interacting units that can exchange material or information. This approach, combined with techniques from graph theory, statistical physics and data analysis, has led to countless interesting studies and insights in numerous fields [1–6].

Until recently, network structures were often regarded as being stationary, both for simplicity and because of scarceness of data sets on the time variability of connectivity patterns. Thanks to advanced acquisition technologies and large scale production of time-resolved data, temporal information has become more accessible in numerous contexts, from communication networks [7–11] to proximity patterns [12,13] and infrastructure networks [14,15]. This has led to the recent surge of activity in the field of "temporal networks" [16]. Data analysis revealed the coexistence of statistically stationary properties and local variations, interaction burstiness [7–20] and the occurrence of nonlocal repetitive patterns [15,21]. These structural properties affect the dynamical processes taking place on networks [22–31]. Therefore, they have to be accounted for in modeling approaches.

Despite the proliferation of temporally resolved data sets, strong limitations perdure. Indeed, data are often only accessible in restricted forms, such as single samples of limited sizes and statistical relevance. Comparison of connection patterns at different times is not always possible. Some data sets consist only of aggregated information and do not provide access to the temporal course of events. In such circumstances, it is necessary to be able to generate synthetic time-extended structures whose aggregation would reproduce the data at hand. This would enable one to go beyond the approximation of static networks and to incorporate dynamical components in network structures.

Several models of temporal networks have been proposed in the literature [14,32–36]. Their dynamics mostly consist of link updates and show that a global complex space-time organization can emerge as the result of simple pairwise cooperation or competition rules between individual units. Data set randomization is also employed to create null models against which the temporal pattern complexity of empirical data can be evaluated [16,22].

The modeling approach developed in this Letter is to some extent more direct and pragmatic. It is intended to provide a versatile procedure for the construction of time-dependent networks with adjustable characteristics that mimic bursty and possibly repetitive behaviors, as well as extended temporal correlations, as they are observed in real world systems. The goal is to obtain realistic temporal structures independently of any access to data sets or assumptions about basic interaction mechanisms.

The proposed procedure originates from the observation that in many graphs representing systems such as transportation, trade, or communication networks, connection patterns are the result of activity spreading along noisy itineraries within a complex structure. The weighted directed graph used to represent this activity is then given by the accumulation of such itinerary traces. Our construction proceeds in the opposite manner. Starting from a given weighted directed graph $\mathcal{G}$, it assumes that this graph is the superposition of random walks in a given time window $\mathcal{W}$, and seeks to "unfold" these walks, as schematically illustrated in Fig. 1.

Below, we first describe the construction in a general setting. We then apply it to a case study inspired by a transportation system, and use this real-world system to benchmark the synthetic network. Finally, we put forward a characterization of the topological and temporal correlations that takes into account the heterogeneity of the activity in the network.

Temporal network construction.—The construction takes as input a graph $\mathcal{G} = (\mathcal{S}, \mathcal{E})$ and a time window $\mathcal{W} = \{0, 1, \cdots, T-1\}$. Here, $\mathcal{S}$ is a set of nodes and $\mathcal{E}$ is a set of weighted directed edges, either obtained from data or built

 © 2013 American Physical Society

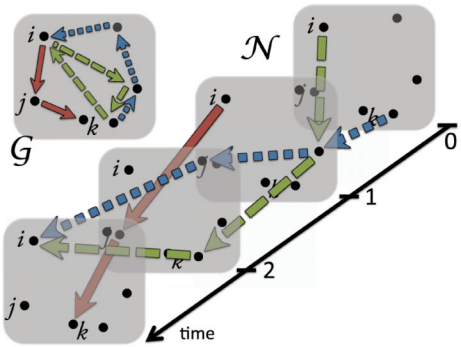


FIG. 1 (color online). Schematic representation of the construction procedure: a weighted directed graph $\mathcal{G}$ can be regarded as a superposition of paths. Unfolding these paths in time results in itineraries (of variable characteristics) that generate a temporal network $\mathcal{N}$.

with prescribed properties such as in and out degree distributions and/or distributions of link weights (for definitions of degree, weight, and strength, see, e.g., Ref. [6]). The temporal network is built as follows. (i) First, the following random walk characteristics are chosen: distributions of starting times in $\mathcal{W}$, of starting locations in $\mathcal{S}$, of walk lengths, and of residence times at each node. (ii) Then, the random walks are generated independently, one after another, using the characteristics specified above. Each walk defines an itinerary that consists of a list of events $\{(i_0, i_1, t_1), (i_1, i_2, t_2), \cdots, (i_{\ell-1}, i_\ell, t_\ell)\}$ with $t_1 < \cdots < t_\ell$ (Fig. 1). At each step, the node i_p is uniformly chosen among the out neighbors of i_{p-1}, and the residence time $t_{p+1} - t_p$ is drawn from the residence time distribution associated with i_p. Any walk reaching a node with no out link terminates there, even if it has not yet reached its prescribed length. (iii) As the construction proceeds, the graph $\mathcal{G}$ is continuously altered as follows: if a walk passes through the link $i \to j$, the weight w_{ij} is decreased by 1; whenever w_{ij} reaches 0, the edge $i \to j$ is discarded from the graph and cannot be used anymore. (iv) Walks are generated until all edges have been discarded from $\mathcal{G}$. Then, the process terminates.

The temporal network $\mathcal{N}$ is defined as the union of all the constructed itineraries. By construction, the set of weighted directed edges resulting from the collection of all events in $\mathcal{N}$ coincides with the original set $\mathcal{E}$. Itineraries are interpreted differently depending upon context. For instance, in transportation systems, the event (i, j, t) is regarded as a material displacement from i to j at time t. In communication networks, events correspond to transmission of information.

Note also that the construction can be used to model routine or seasonal processes occurring in consecutive time windows, using a "noisy deterministic" rule: for each n, a tunable fraction of walks is redrawn in $\mathcal{W}_n = \{nT, nT + 1, \ldots, (n+1)T - 1\}$, while the remaining itineraries are repeated identically from $\mathcal{W}_{n-1}$ to $\mathcal{W}_n$.

We claim that, under suitable choices of input parameters, the network $\mathcal{N}$ can mimic real-world features, in particular bursty temporal patterns and extended correlations. To support these assertions, we benchmark the synthesized network in a case study inspired by a cattle trade system analysis [15]; note, this example is aimed to illustrate the construction algorithm, rather than to accurately fit specific data sets. The robustness of the emerging phenomenology is evaluated in Ref. [37], where statistics obtained with other choices of parameters (e.g., residence time distributions, mean path lengths, and weight distributions) are reported.

Case study.—In this application, we consider $|\mathcal{S}| = 10^4$ nodes representing farms. The graph $\mathcal{G}$ is constructed using a variant of the uncorrelated configuration model [38]. Following statistics reported in Ref. [15], we use as in and out degree distributions the power-law $P(k_{\text{in,out}}) \propto k_{\text{in,out}}^{-2.5}$ (with cutoff at $\sqrt{|\mathcal{S}|}$ to avoid degree correlations [38]). Weights are also power-law distributed according to $P(w) \propto w^{-2.5}$. Interpreting events in $\mathcal{N}$ as material displacements, we impose flux conservation at every node so that in and out strengths balance; i.e., $\sum_j w_{ji} = \sum_j w_{ij}$ holds for almost all $i \in \mathcal{S}$ (in practice, more than 99% of $\mathcal{S}$ [37]).

We consider a temporal window $\mathcal{W}$ of length $T = 10^3$ units ("days"), with periodic boundary conditions, to generate $\mathcal{N}$ from $\mathcal{G}$. To this aim, we use random walks with uniformly distributed starting times and locations. We assign each node i a residence time τ_i, drawn from $P(\tau_i) \propto \tau_i^{-3}$ for $1 \leq \tau_i \leq 60$; a walk visiting node i stays for time τ_i, plus an additional random delay drawn from a Poisson distribution with mean $\tau_i/5$. (The power-law $P(\tau_i)$ is inspired by Ref. [15]; similar results are obtained for an exponential $P(\tau_i)$ [37].) Walk lengths are generated using a Poisson distribution with average 10. Notice that since the number of graph edges decreases as the construction proceeds, long walks can seldom be achieved when few edges remain and the realized distribution accordingly exhibits an excess of short paths, see Fig. 2.

Three main types of analysis have been proposed in order to study structural and temporal properties of a time-resolved transportation network [14,15,21]: (i) statistics of aggregated networks on various time scales, (ii) distributions of activity and inactivity periods, and (iii) repetition

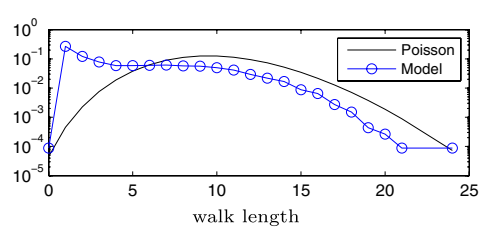


FIG. 2 (color online). Prescribed (solid line) and realized (circles) distributions of random walk lengths in the temporal network $\mathcal{N}$. The discrepancy is due to the decay of the number of remaining edges in $\mathcal{G}$ as the construction proceeds.

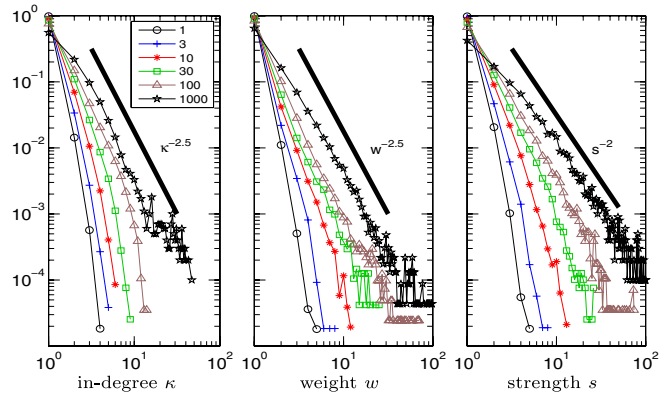
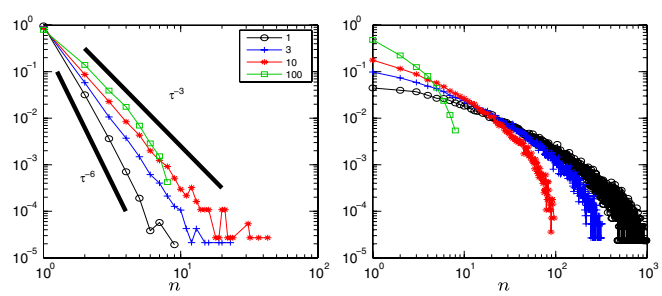


FIG. 3 (color online). Distributions of node in degrees, link weights, and node strengths for networks aggregated over intervals of various lengths. For integration lengths $<10^3$, the figures display distributions obtained by averaging over all corresponding intervals in $\mathcal{W}$. Power laws (thick black lines) with expected exponents are plotted for reference.

of connection patterns such as temporal paths. Here, we show that when applied to the synthetic network $\mathcal{N}$, these diagnostics proffer characteristics very similar to those in empirical data.

Real world data exhibit robust heterogeneous behaviors at all temporal scales, as asserted by power-law distributions of in degrees, strengths, and link weights [12,14,15]. These distributions remain almost stationary when integrated over distinct windows of equal length. As shown in Fig. 3, the same properties are observed in $\mathcal{N}$. These plots also validate the construction algorithm: statistics aggregated over the whole $\mathcal{W}$ reveal power-law distributions with slopes in close agreement with the corresponding ones in $\mathcal{G}$.

Robust heterogeneous features in transportation networks have also been observed in activity and inactivity period statistics, i.e., in the distributions of the periods during which a node (or a link) is continuously active or inactive. Such observables typically exhibit broad distributions [14]. Figure 4 reports these distributions for nodes,



(a) Node activity time distrib. (b) Node inactivity time distrib.

FIG. 4 (color online). Distributions of activity (a) and inactivity (b) periods aggregated over intervals of various lengths $T_{\rm agg} = 1$, 3, 10, 100, i.e., fractions of nodes that are active at least once in each of (respectively, inactive during) n consecutive intervals of duration $T_{\rm agg}$ (the maximal value of n is $T/T_{\rm agg}$).

obtained from $\mathcal{N}$ using various aggregation scales $T_{\rm agg}$, from the most detailed ("daily", $T_{\rm agg} = 1$) resolution to coarser time scales mimicking weekly or monthly aggregations (for similar results on links, see Ref. [37]). Activity distributions reveal power-law behaviors in quantitative agreement with empirical data [15], with decay exponents in the same range of values (increasing from -6 to -3 as the aggregation interval $T_{\rm agg}$ increases). Moreover, distributions of inactivity periods are broader, more concave, and extend across all time scales, as in Ref. [15].

In addition to burstiness, the activity in $\mathcal{N}$ shows positive temporal correlations, as can be anticipated from the nature of the construction. In particular, global activity correlations can be appreciated by using

$$C(t) = \frac{\langle A(\cdot)A(\cdot + t)\rangle_{ti} - \langle A(\cdot)\rangle_{ti}^2}{\mathrm{Var}[A(\cdot)]},$$

where $A(t)$ denotes the total number of events (i, j, t) on day t and $\langle\cdot\rangle_{ti}$ means time averaging over $\mathcal{W}$. As Fig. 5 (left) shows, $C(t)$ is positive for all t in an extended interval, whereas its analog in a null model, obtained by reshuffling all links, is 0 for any $t \neq 0$. Note that such a null model is equivalent to a temporal network construction in which we impose all the random walks to have length 1.

Beside global correlations, local structures can be highlighted by exhibiting chronological link sequences that occur more often than by chance [15,21]. A typical example is given by τ-tolerant temporal paths, i.e., paths composed of consecutive active links within τ days of each other [37]. The number of such paths is much larger in empirical temporal networks than in reshuffled data [15]. (Note that only the case $\tau = 1$ was considered in Ref. [15].) Figure 5 (right), which shows the number of τ-tolerant temporal paths that occur at least twice in $\mathcal{N}$, possesses similar features, providing further evidence of the presence of strong correlations in the synthetic network.

Detailed activity correlations.—The nature of the temporal network construction, which relies on random walks, suggests that nontrivial local activity correlations should

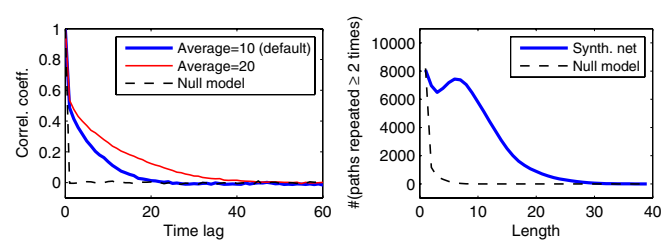


FIG. 5 (color online). Left: Autocorrelation function $C(t)$ vs t, averaged over 40 realizations of $\mathcal{N}$, for two distinct walk length averages and compared to a null model where all links have been temporally randomized, equivalent to imposing that all walks in the construction have length 1. Right: Number of τ-tolerant temporal paths ($\tau = 5$) occurring at least twice vs their length, both in $\mathcal{N}$ and in the null model.

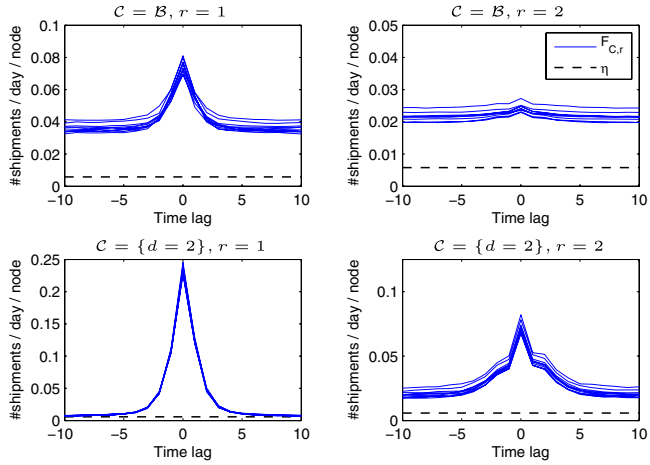


FIG. 6 (color online). $F_{\mathcal{C},r}(\tau)$ vs τ for 2 categories of nodes $\mathcal{C} = \mathcal{B}$ (busy nodes) and $\{d = 2\}$, and two neighboring ball radii $r = 1$ and $r = 2$. Each panel displays the curves corresponding to 10 realizations of $\mathcal{N}$. Dashed lines refer to the average activity $\eta = |\mathcal{N}|/T/|\mathcal{S}|$ per day per node.

also simultaneously emerge at both topological and temporal levels. Indeed, if a node is active at a certain time, then an elevation of activity should be detected in its topological and temporal neighborhood. (By contrast, such correlations are nonexistent in a model where all links have been reshuffled.) Here, we provide quantitative estimates of these correlations.

In heterogeneous networks such as the current synthetic $\mathcal{N}$, local activity patterns are extremely diverse and are hardly captured by global averages such as $C(t)$. Therefore, prior to averaging, we first need to divide the nodes of $\mathcal{S}$ into categories according to their total activity (defined as the total number of events in $\mathcal{W}$, starting from the considered node). The activity at a node in $\mathcal{S}$ ranges from 1 to over 200, with average $|\mathcal{N}|/|\mathcal{S}| \sim 6$. Accordingly, we define a node as "busy" if its integrated activity exceeds a given threshold, say, e.g., 40 (results are qualitatively insensitive to this choice). Furthermore, nonbusy nodes are categorized into level sets of the distance d to the set of busy nodes $\mathcal{B}$, where d is measured following the edges in $\mathcal{E}$. Altogether, the categories $\mathcal{B}=\{d=0\}, \{d=1\}, \ldots, \{d=4\}$ account for more than 99% of $\mathcal{S}$.

For each node category $\mathcal{C}$, the correlations between the activity in nodes of $\mathcal{C}$ at time t and in their neighborhood at time $t + \tau$ are then measured by

$$F_{\mathcal{C},r}(\tau) = \langle a_{B(i,r)}(t + \tau)\rangle_{\{i\in\mathcal{C}, t\in\mathcal{W}|s_{\mathrm{out}}(i,t)>0\}},$$

where $a_E(t) = \sum_{j\in E} s_{\mathrm{out}}(j, t)/|E|$ is the average activity per node in a subset of nodes E at time t [$s_{\mathrm{out}}(j, t)$ is the number of events starting from j at time t], $B(i, r)$ is the ball of center i and radius r, and the average $\langle\cdot\rangle$ is taken over all $i \in \mathcal{C}$ and $t \in \mathcal{W}$ such that $s_{\mathrm{out}}(i, t) > 0$.

Figure 6 displays "activity spike curves" $F_{\mathcal{C},r}(\tau)$ vs τ for two categories, $\mathcal{B}$ and $\{d = 2\}$, and $r = 1, 2$. In a null model where all links are reshuffled (equivalent to a model built with random walks of length 1), each such curve would be perfectly flat except at $\tau = 0$, where it would take a larger value. In contrast, each curve has a distinctive shape here that is almost insensitive to the network realization. Hence, for each category, these curves provide a characterization of activity correlation patterns in a topological and temporal neighborhood.

The peaked shape of $F_{\mathcal{C},r}(\tau)$ illustrates how the occurrence of activity in the center of a ball is accompanied by an elevated activity in the ball for a number of days before and after. The peaks are substantially sharper for smaller balls (of radius 1), and are instead barely detectable when averaging in the ball of radius 2 around $\mathcal{B}$. At large values of τ, $F_{\mathcal{C},r}(\tau)$ approaches a well-defined baseline level of activity. For $r = 1$, this baseline is higher around busy nodes than around $\{d = 2\}$, pointing to a higher level of mean activity in the neighborhood of $\mathcal{B}$, as it can be anticipated.

Concluding remarks.—In this Letter, a simple framework for the construction of temporal networks displaying bursty, repetitive, and correlated behaviors has been presented, based on random walks on a predefined aggregated graph. The construction is sufficiently versatile to adapt to a wide range of situations, by appropriate tuning of input parameters. In particular, it would be interesting to apply it to initial graphs $\mathcal{G}$ with community structure. As random walks tend to be trapped inside communities, the construction is expected to naturally give rise to the emergence of temporal communities. Another possible extension would be to use, instead of random walks as building blocks, spreading trees of propagation processes on networks.

Our procedure enables one to obtain plausible and realistic instances of temporal networks when only the aggregated structure is known. Hence, it can be employed to compensate for the lack of time-resolved data or to provide alternative scenarios when access to empirical information is limited. Moreover, that the synthetic network has tunable properties is essential to assessing the influence of time-dependent features on the dynamical processes on networks, especially when only aggregated information is available. Finally, structural flexibility also makes it possible to test the relevance of various definitions of centrality measures for nodes and links in temporal networks.

A. B. and B. F. are partly supported by FET Project No. MULTIPLEX 317532 and by CNRS PEPS *Physique Théorique et ses Interfaces*. L. S. Y. is supported in part by NSF Grant No. DMS-1101594, and K. L. by NSF Grant No. DMS-0907927.

[1] Special issue on Complex Networks and Systems, edited by B. Alberts [Science 325, 357 (2009)].

[2] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford University Press, Oxford, 2003).

[3] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).
[4] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet: A Statistical Physics Approach* (Cambridge University Press, Cambridge, England, 2004).
[5] G. Caldarelli, *Scale-Free Networks* (Oxford University Press, Oxford, 2007).
[6] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, Cambridge, England, 2008).
[7] J.-P. Eckmann, E. Moses, and D. Sergi, Proc. Natl. Acad. Sci. U.S.A. **101**, 14 333 (2004).
[8] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabàsi, Proc. Natl. Acad. Sci. U.S.A. **104**, 7332 (2007).
[9] D. Rybski, S. V. Buldyrev, S. Havlin, F. Liljeros, and H. A. Makse, Proc. Natl. Acad. Sci. U.S.A. **106**, 12 640 (2009).
[10] R. D. Malmgren, D. B. Stouffer, A. S. L. O. Campanharo, and L. A. N. Amaral, Science **325**, 1696 (2009).
[11] M. Karsai, K. Kaski, A.-L. Barabási, and J. Kertész, Sci. Rep. **2**, 397 (2012).
[12] C. Cattuto, W. Van den Broeck, A. Barrat, V. Colizza, J.-F. Pinton, and A. Vespignani, PLoS ONE **5**, e11596 (2010).
[13] M. Salathé, M. Kazandjieva, J. W. Lee, P. Levis, M. W. Feldman, and J. H. Jones, Proc. Natl. Acad. Sci. U.S.A. **107**, 22 020 (2010).
[14] A. Gautreau, A. Barrat, and M. Barthélemy, Proc. Natl. Acad. Sci. U.S.A. **106**, 8847 (2009).
[15] P. Bajardi, A. Barrat, F. Natale, L. Savini, and V. Colizza, PLoS ONE **6**, e19869 (2011).
[16] P. Holme and J. Saramäki, Phys. Rep. **519**, 97 (2012).
[17] P. Holme, Phys. Rev. E **71**, 046119 (2005).
[18] A. L. Barabàsi, Nature (London) **435**, 207 (2005).
[19] A. Vàzquez, J. G. Oliveira, Z. Dezsö, K.-I. Goh, I. Kondor, and A.-L. Barabàsi, Phys. Rev. E **73**, 036127 (2006).
[20] A. L. Barabási. *Bursts: The Hidden Pattern Behind Everything We Do* (Dutton, New York, 2010).
[21] L. Kovanen, M. Karsai, K. Kaski, J. Kertész, and J. Saramäki, J. Stat. Mech. (2011) P11005.
[22] M. Karsai, M. Kivelä, R. K. Pan, K. Kaski, J. Kertész, A.-L. Barabási, and J. Saramäki, Phys. Rev. E **83**, 025102 (R) (2011).
[23] A. Vazquez, B. Rácz, A. Lukács, A.-L. Barabási, Phys. Rev. Lett. **98**, 158702 (2007).
[24] J. L. Iribarren and E. Moro, Phys. Rev. Lett. **103**, 038702 (2009).
[25] G. Miritello, E. Moro, and R. Lara, Phys. Rev. E **83**, 045102(R) (2011).
[26] L. Isella, J. Stehlé, A. Barrat, C. Cattuto, J.-F. Pinton, and W. Van den Broeck, J. Theor. Biol. **271**, 166 (2011).
[27] P. Bajardi, A. Barrat, L. Savini, and V. Colizza, J. R. Soc. Interface **9**, 2814 (2012).
[28] P. Holme, arXiv:1302.0692.
[29] T. Hoffmann, M. A. Porter, and R. Lambiotte, Phys. Rev. E **86**, 046102 (2012).
[30] N. Perra, A. Baronchelli, D. Mocanu, B. Goncalves, R. Pastor-Satorras, and A. Vespignani, Phys. Rev. Lett. **109**, 238701 (2012).
[31] B. Ribeiro, N. Perra, and A. Baronchelli, arXiv:1211.7052.
[32] J. Tang, S. Scellato, M. Musolesi, C. Mascolo, and V. Latora, Phys. Rev. E **81**, 055101(R) (2010)
[33] J. Stehlé, A. Barrat, and G. Bianconi, Phys. Rev. E **81**, 035101(R) (2010).
[34] N. Perra, B. Gonçalves, R. Pastor-Satorras, and A. Vespignani, Sci. Rep. **2**, 469 (2012).
[35] M. Starnini, A. Baronchelli, and R. Pastor-Satorras, arXiv:1301.3698.
[36] T. Gross and B. Blasius, J. R. Soc. Interface **5**, 259 (2008).
[37] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.110.158702 for additional details on the model and case studies with different parameter sets.
[38] M. Catanzaro, M. Boguñá, and R. Pastor-Satorras, Phys. Rev. E **71**, 027103 (2005).