Simplicial Activity Driven Model

Giovanni Petri[1,*] and Alain Barrat[2,1]
[1]ISI Foundation, 10126 Turin, Italy
[2]Aix Marseille Univ, Université de Toulon, CNRS, CPT, 13009 Marseille, France

(Received 17 May 2018; revised manuscript received 10 October 2018; published 29 November 2018)

Many complex systems find a convenient representation in terms of networks: structures made by pairwise interactions (links) of elements (nodes). For many biological and social systems, elementary interactions involve, however, more than two elements, and simplicial complexes are more adequate to describe such phenomena. Moreover, these interactions often change over time. Here, we propose a framework to model such an evolution: the simplicial activity driven model, in which the building block is a simplex of nodes representing a multiagent interaction. We show analytically and numerically that the use of simplicial structures leads to crucial structural differences with respect to the activity driven model, a paradigmatic temporal network model involving only binary interactions. It also impacts the outcome of paradigmatic processes modeling disease propagation or social contagion. In particular, fluctuations in the number of nodes involved in the interactions can affect the outcome of models of simple contagion processes, contrarily to what happens in the activity driven model.

DOI: 10.1103/PhysRevLett.121.228301

The use of a network representation has become commonplace for describing and studying many complex systems: nodes represent the elements of the systems and links represent pairwise interactions [1,2]. However, in many contexts, representing interactions as pairwise does not tell the whole story. Examples include collaborations among groups of actors in movies [3], spiking neuron populations [4,5], and coauthorships in scientific publications [6].

Let us consider the latter for illustration purposes: In a network representation, a paper coauthored by n scientists yields a clique of $n(n-1)/2$ links, which is, however, treated in the same way as $n(n-1)/2$ papers authored by pairs of scientists (or any combination of subgroups among these scientists leading to the same number of links). While this is equivalent for $n=2$, the number of coauthors of a paper is often larger than 2. For instance, the average number of authors of APS papers has steadily increased from 2 to 6 between the 1940s and now [see Supplemental Material (SM), Sec. I [7]]. In such cases ($n>2$), simplicial representations are more apt to preserve the information observed in data. To take this into account, simplicial descriptions have recently been adopted in models of emerging geometry [8,9], null models for higher order interactions [10,11], network inference [12], brain structure, and dynamics [4,13,14].

We recall that formally a $(d-1)$-dimensional simplex σ is defined as the set of d vertices $\sigma=[x_0,x_1,\ldots,x_{d-1}]$. A collection of simplices is a simplicial complex K if for each simplex σ all its possible subfaces (defined as subsets of σ) are themselves contained in K (see SM, Sec. II [7], and Ref. [15]). In the case of group interactions, this requirement can be considered trivially satisfied, as each group interaction implies all the possible subinteractions. Finally, the 1-skeleton of a simplicial complex is the collection of all its edges, i.e., the underlying network.

Networks and interactions moreover evolve in time, and the field of temporal networks is indeed very active [16,17]. In particular, several models of time-evolving networks have been put forward, using microscopic rules for the establishment and end of interactions between pairs of nodes [18–20]. Among these, the activity driven (AD) temporal network model [19] has attracted a lot of attention. In this model, each agent (node) is assigned an activity potential that determines at each time its probability to create pairwise interactions with other agents selected at random. The AD model and its extensions [21–24] have become a paradigm of temporal networks and have been used to study the impact of the network's temporal evolution on dynamical processes occurring on top of it [19,25].

Models of temporally evolving simplicial interactions are, however, still missing. Here, we bridge this gap by proposing a modeling framework for temporal group activation data: the simplicial activity driven (SAD) model. Our aim is to provide a simple framework that can serve as a basis for richer temporal models including the simplicial nature of interactions, and on which dynamical processes can be studied analytically and numerically to shed light on the impact of both simplicial and temporally evolving interactions.

In its simplest version, the model considers N nodes, whose interactions change over time as follows. (i) Each node i is endowed with an activity rate a_i taken from a predefined distribution F. (ii) At each time step Δt, each node i fires with probability $a_i\Delta t$; when it fires, it creates an

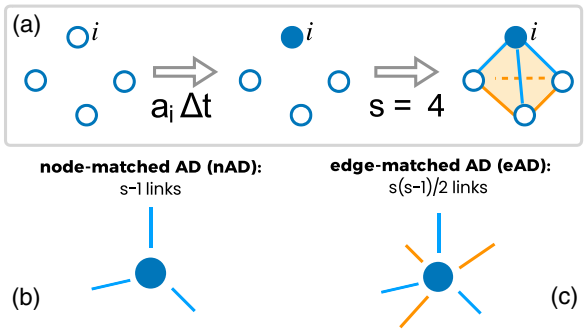

FIG. 1. SAD model. (a) At each time step, a node i activates with probability $a_i \Delta t$. Upon activation it creates a coherent unit of s nodes [an $(s-1)$-simplex], with links between all pairs. (b) In contrast, in the standard AD model (nAD), only the $s-1$ edges stemming from the activated node are added. (c) In the eAD model, $\binom{s}{2}$ links stem from the activated node, conserving at each interaction the number of links of the SAD model.

$(s-1)$ simplex (in networks' terms, a clique of size s) with $s-1$ other nodes chosen uniformly at random. Each activation hence yields $s(s-1)/2$ interactions. (iii) At the following time step, the existing simplices are erased and the process starts anew.

In the framework of collaborations, nodes represent scientists and the activity a_i their propensity to create collaborations: step (ii) corresponds to the creation of a collaboration of s scientists who coauthor a paper. We underline the main difference with the AD model in Fig. 1: in the AD model, each active node creates *binary* interactions with the chosen nodes (in the language of collaborations, $s-1$ papers each with 2 authors) while, in the SAD model, inactive nodes targeted by an active node obtain links to all the other nodes in the simplex, creating a coherent unit. The parameter s defines the size of the collaborations and can be either fixed or a random variable extracted at each activation from a distribution $p(s)$. As for the AD model, the SAD model is Markovian: agents have no memory of previous time steps, and it can be refined by adding memory or community effects [21,24].

In the following, we first study this model from a structural point of view, highlighting the differences between considering the obtained system as a network (given by its 1-skeleton) and taking into account its simplicial nature. We also provide an analysis of a paradigmatic dynamical process occurring on top of the SAD model. Since the AD model has been widely studied as a paradigm for temporal networks, we underline in each case how the introduction of coherent units of s nodes as building blocks yields radically different structural properties and impacts the properties of dynamical processes.

The comparison with the activity driven model can be done in two ways: we can consider AD models involving either the same number of nodes [node-matched AD model (nAD)] or the same number of interactions [edge-matched AD model (eAD)] as the SAD model at each time step (see Fig. 1). In the former, for each activation with group size s in the SAD model, we consider an AD activation with size $m = s-1$; i.e., the activated node creates interactions with $s-1$ other nodes chosen at random: this leads to the same total number of contacted nodes per activation in the nAD and in the SAD model. In the eAD model, for each activation with group size s, we consider instead an AD activation with $m = \binom{s}{2}$, hence preserving the total number of interactions of each activation.

Structure.—Let us focus on the structural properties of the SAD model aggregated over T time steps. We first consider the 1-skeleton of the SAD model, to compare its properties to AD networks, and then consider pure simplicial properties—not reducible to a network approach.

In the aggregated SAD model, each node i is linked by an edge to all the nodes with which it has interacted at least once during T. The degree of i in the corresponding aggregated network corresponds to the number of distinct nodes with which i has interacted (in the interpretation of a scientific collaboration network, the total number of distinct collaborators of a scientist).

Denoting by $k_T(i)$ the expected aggregated degree at time T of node i with activity a_i, we compute it by separating it into two contributions. The first comes from node i's own activation events, which occur at each time step Δt with probability $a_i \Delta t$: after T time steps, i will have activated $\sim T a_i$ times; for fixed simplex size s, this means it will have made $T a_i \bar{m}$ interactions ($\bar{m} = s-1$). The second contribution comes from the activations of other nodes: every node $j \neq i$ will have activated $T a_j$ times; in each activation of $j \neq i$, i was selected with probability $\bar{m}/(N-1)$ and, if selected, provided with $\bar{m}$ interactions. Hence, at time T node i will have accumulated $\kappa_T(i)$ interactions with

$$\kappa_T(i) = \bar{m} a_i T + \sum_{j \neq i} \frac{\bar{m}^2 T a_j}{N-1} \simeq \bar{m} T (a_i + \bar{m} \langle a \rangle), \quad (1)$$

where the approximation holds for $N \gg 1$. For any node distinct from i, the probability not to have been involved in any of these interactions is $[1 - 1/(N-1)]^{\kappa_T(i)}$, and hence, finally the number of distinct nodes having interacted with i is

$$k_T^{\mathrm{SAD}}(i) = (N-1)\left[1 - \left(1 - \frac{1}{N-1}\right)^{\kappa_T(i)}\right] \quad (2)$$

$$\simeq N[1 - e^{-[T\bar{m}(a_i + \langle a \rangle \bar{m})]/N}], \quad (3)$$

where the approximation holds for large N and small T/N (in the SM we also give the derivation for the aggregated degree distribution [7]). In Fig. 2(a), we show the excellent agreement between the prediction for the aggregated degree averaged over all nodes, $\langle k_T^{\mathrm{SAD}}(i) \rangle$, at fixed s, and numerical

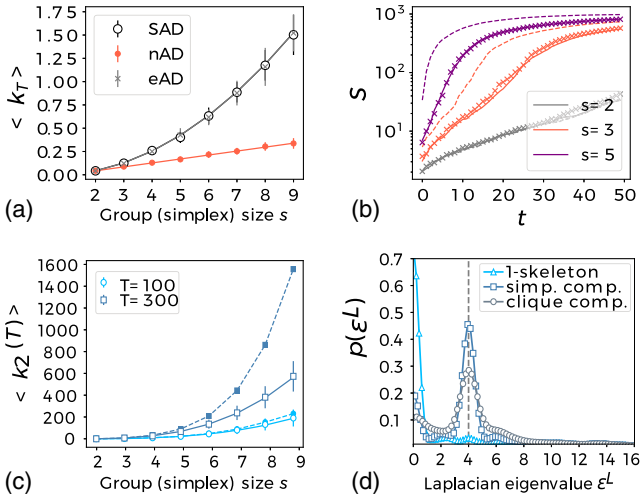

FIG. 2. Structural properties of SAD model. In all plots we use $N = 2000$ nodes and activities sampled from $F(a) = (a/a_0)^{-\alpha}$, where $\alpha = 2.1$ and $a_0 = 5 \times 10^{-3}$. (a) Average aggregated degree $\langle k_T \rangle$ for the SAD model and corresponding nAD and eAD models versus simplex size s, with $T = 10$. Symbols, numerical values; lines, theoretical predictions [for the SAD model, Eq. (3) averaged over all nodes]. (b) Temporal growth of the aggregated GCC size S for the SAD (solid lines), nAD (crosses), and eAD (dashed lines) models for various fixed simplex size s. (c) Empty symbols are the average $\langle k_2(i,T) \rangle$ over all nodes i of the number of 2-simplices to which i belongs, in the SAD model aggregated until time T, for various values of s and of T. Continuous lines are the prediction [Eq. (4)] averaged over nodes. The filled symbols give instead the average number of triangles to which a node i belongs in the 1-skeleton of the aggregated SAD model. (d) Eigenspectrum of the simplicial Laplacian $\mathcal{L}_1$ computed on the aggregated SAD 1-skeleton (here with $s = 4$), on the actual aggregated SAD simplicial complex, and on the clique complex of the 1-skeleton of the aggregated SAD simplicial complex [in which each $(k+1)$ clique of the 1-skeleton is promoted to a k-simplex]. Aggregation time $T = 100$.

simulations. We also compare it with the nAD and eAD models, for which $k_T^{\text{nAD or eAD}}(i) = N(1 - e^{-Tm(a_i + \langle a \rangle)/N})$ [19], with $m = \bar{m}$ for nAD model and $m = \binom{s}{2}$ for eAD model.

Interestingly, k_T^{SAD} depends on $\bar{m}^2$; thus, if the simplex size s is allowed to fluctuate, the size of such fluctuations will affect the aggregated degree (see Fig. S2 of the SM [7]). This is in contrast with the nAD model, which has no dependence on the second moment of s, while the eAD model inherits it from the matching of the number of edges created at each activation [since each activation creates $m = s(s-1)/2$ edges, the total number of interactions and the integrated degree depend on the fluctuations of s].

For the aggregated network degree, we thus observe a similar behavior for the SAD and eAD models. Figure 2(b), however, highlights that the SAD model building mechanism also leads to an important structural difference with the eAD model: as each activation creates $\binom{s}{2}$ interactions that involve only $s-1$ nodes, the size of the largest connected component (GCC) in the SAD model integrated until T grows with T much more slowly than in the eAD model, for which each activation creates a star reaching $\binom{s}{2}$ nodes; in fact, it grows in the same way as in the nAD model, despite creating more interactions at each step (for $s > 2$, as for $s = 2$ the three models are the same). Overall, the structural properties of the SAD model, from the point of view of its 1-skeleton, present thus both similarities and important differences with AD models with the same number of events.

Let us now investigate purely simplicial structural properties of the SAD model. First, we compute the average number $k_2(i,T)$ of 2-simplices to which a node i belongs in the SAD aggregated until T, in a way similar to the computation of $k_T(i)$. We obtain (see SM, Sec. IV, for details [7])

$$k_2(i,T) = \binom{N-1}{2}\left(1 - e^{-\{[(s-1)(s-2)]/[(N-1)(N-2)]\}T(a_i + \bar{m}\langle a \rangle)}\right). \tag{4}$$

Note that $k_2(i,T)$ corresponds to the number of distinct cliques of three nodes to which i has participated from 0 to T, which is different from the number of triangles to which i belongs in the 1-skeleton of the aggregated SAD model: indeed, a triangle (i,j,k) can be obtained even if links (i,j), (i,k), and (j,k) are never present in the same $(s-1)$-simplex. We show in Fig. 2(c) that the average of $k_2(i,T)$ over all nodes is correctly predicted by Eq. (4) (see SM, Sec. IV, for a more extensive validation [7]). The figure also shows that the average number of triangles to which a node belongs grows faster with both s and T than $\langle k_2(i,T) \rangle$, highlighting the differences between simplices and triangles and thus the importance of the simplicial nature of the SAD model.

We moreover present in Fig. 2(d) the eigenspectrum of the simplical Laplacian (see SM, Sec. V [7], and Refs. [26,27]) of the aggregated SAD model, a cornerstone of studies of how dynamical processes are affected by the underlying structure [28–30]. The figure highlights how the eigenspectrum differs, depending on whether we compute the Laplacian on the 1-skeleton, on the aggregated SAD simplicial complex, or on the clique complex obtained by considering each clique of the 1-skeleton as a simplex. These differences illustrate further how the SAD model contains information not reducible to its 1-skeleton, i.e., to a network.

To further our analysis, we now explore how dynamical processes are impacted by the SAD model.

Dynamical processes.—We consider the paradigmatic susceptible-infected-susceptible (SIS) model for disease spreading [2]. In this model, nodes can be either susceptible (S) or infectious (I). Infectious propagate the disease to susceptibles at rate β whenever they are interacting, and recover spontaneously at rate μ, becoming again

susceptible. In a homogeneous population, the epidemic threshold λ_c separates an epidemic-free state at low values of the parameter $\lambda = \beta/\mu$ from an endemic state at high values of λ.

To calculate the SIS epidemic threshold for the SAD model, we use a temporal heterogeneous mean-field approach similar to the one used for the AD model [19]: nodes are classified according to their activity, and we denote by N_a the number of nodes with activity a; I_a^t and S_a^t denote, respectively, the numbers of infectious and susceptible nodes with activity a at time t. We thus have $N_a = S_a^t + I_a^t$, $N = \int da N_a$ is the total population, and $I^t = \int da I_a^t$ the total number of infectious at time t.

Let us consider the case of the SAD model with fixed clique size s. The variation during a time step Δt of the number of infectious is given by the following equation taking into account the evolution of both interactions and spreading process:

$$I_a^{t+\Delta t} - I_a^t = -\mu \Delta t I_a^t + \beta \Delta t S_a a (s-1) \int da' \frac{I_{a'}^t}{N} + \beta \Delta t S_a \int da' a' \frac{I_{a'}^t}{N}(s-1) + \beta \Delta t S_a \int da' a' \frac{S_{a'}}{N}(s-1) \int da'' \frac{I_{a''}}{N}(s-2). \tag{5}$$

The first term corresponds to the recovery of nodes with activity a. The second term corresponds to susceptible nodes with activity a that become active at time t (with probability $a\Delta t$) and create a simplex of size s (with $s-1$ other nodes) that includes infectious nodes with any activity (hence, the integration over a'). The third term stems from the fact that susceptible nodes with activity a can be chosen as clique partners by infectious nodes with activity a' that become active (with probability $a'\Delta t$). While these three terms also appear in the case of a spreading process on an AD network, the last term is specific to the SAD model: it describes the cases in which susceptible nodes with activity a are chosen by a susceptible with activity a', which becomes active (with probability $a'\Delta t$) and creates a simplex that also includes an infectious node with activity a''.

Straightforward computations detailed in the SM (Sec. VI) yield then the epidemic threshold condition [7]

$$\frac{\beta}{\mu} > \frac{2}{s(s-1)\langle a\rangle + (s-1)\sqrt{s^2\langle a\rangle^2 + 4(\langle a^2\rangle - \langle a\rangle^2)}}, \tag{6}$$

to compare with the result $(\beta/\mu) > 1/(m\langle a\rangle + m\sqrt{\langle a^2\rangle})$ for an AD model with parameter m.

If the sizes of the cliques formed in the SAD model are extracted at random at each activation from a distribution $p(s)$, the rhs of Eq. (5) needs simply to be integrated as $\int ds p(s)$, if the size s is independent of the activity a. The epidemic threshold becomes

$$\frac{\beta}{\mu} > \frac{2}{\langle s(s-1)\rangle\langle a\rangle + \sqrt{\Delta}}, \tag{7}$$

with $\Delta = \langle (s-1)(s-2)\rangle\langle (s-1)(s+2)\rangle\langle a\rangle^2 + 4\langle s-1\rangle^2\langle a^2\rangle$ (see details in the SM, Sec. VI [7]). Notably, it depends not only on the average clique size but also on the second moment of $p(s)$, and vanishes as $1/\langle s^2\rangle$ if the clique size fluctuations diverge (see Fig. S6 in the SM [7]). This is in contrast with the case of the SIS model on the AD model, in which fluctuations of the numbers of links created at each time step would not change the epidemic threshold (m just being replaced by its average).

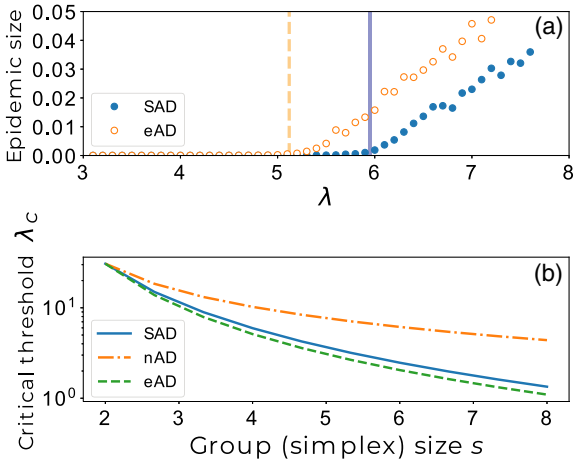

FIG. 3. SIS epidemic threshold for AD and SAD models. (a) Epidemic prevalence versus $\lambda = \beta/\mu$: the epidemic transition in the SAD model is delayed as compared with an SIS model on the corresponding eAD model (here, $N = 1000$, $s = 4$, $T = 20\,000$). Vertical lines correspond to the theoretical values of the epidemic thresholds [Eq. (6) for the SAD model]. (b) Increasing the average connectivity of the underlying network lowers the epidemic threshold in all models; for the s-regular SAD model λ_c [Eq. (6)] is always larger than for the corresponding eAD model. In both panels, node activities were sampled from $F(a) = (a/a_0)^{-\alpha}$, where $\alpha = 2.1$ and $a_0 = 5 \times 10^{-3}$.

Figure 3(a) displays the result of numerical simulations of a SIS model on temporal eAD and SAD networks, showing agreement with the theoretical values of the epidemic threshold. We, moreover, compare in Fig. 3(b) the epidemic threshold obtained in a SAD model with fixed clique size s with the one obtained in the nAD and eAD models. In the former case, the epidemic threshold is smaller in the SAD model, which can be related to the fact that the SAD model has more interactions than the nAD network. In the latter case instead, the fact that the $s(s-1)/2$ interactions are created as cliques hampers the spread on the SAD model with respect to the eAD

network, leading to a higher epidemic threshold for the SAD case (see SM, Table S1, for examples [7]).

To conclude, we have presented a new model for temporal networks, based on the fact that the fundamental building blocks of many social networks are coherent units of individuals interacting as groups, rather than dyadic interactions. Our simplicial activity driven model considers indeed agents who create simplices with other agents, yielding a simplicial complex once aggregated. We have shown how this mechanism leads to fundamental differences with respect to a well-known model in which active agents create sets of dyadic interactions, and how the structural properties of the SAD model differ from those of its 1-skeleton, showing the necessity to take into account its simplical nature and not to reduce it to a network interpretation. These differences appear at the structural level and also have strong consequences on how dynamical processes unfold on these networks, as we have illustrated on a paradigmatic epidemic model (we also show in Sec. VII of the SM that a social contagion process [31,32] on the SAD model displays a rich phenomenology, very different with respect to an AD model [7]).

As noted in its definition, the SAD model is Markovian, as the AD model: it does not yield non-Poissonian nor bursty temporal patterns. However, thanks to this simplicity, our model lends itself to analytical investigations of its structural properties and of contagion processes, which has allowed us to highlight the need to correctly take into account the simplicial nature of interactions and the fluctuations of the numbers of nodes involved in these interactions. Moreover, it can serve as a starting point for many refinements, such as adding memory effects, node categories, and interacting probabilities depending on these categories, or correlations between activity of an agent and size of the simplicial complex it creates. Moreover, it would be interesting to study further dynamical processes on the SAD model and its variations. Finally, the SAD model constitutes a first null model for the homology of temporal complex systems with high-order interactions. We hope that our work will stimulate research in such directions.

Acknowledgments

G. P. is supported by Compagnia di San Paolo (Grant No. ADnD).

*giovanni.petri@isi.it

[1] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[2] A. Barrat, M. Barthelemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, Cambridge, England, 2008).
[3] J. J. Ramasco, S. N. Dorogovtsev, and R. Pastor-Satorras, Phys. Rev. E **70**, 036106 (2004).
[4] C. Giusti, E. Pastalkova, C. Curto, and V. Itskov, Proc. Natl. Acad. Sci. U.S.A. **112**, 13455 (2015).
[5] M. W. Reimann, M. Nolte, M. Scolamiero, K. Turner, R. Perin, G. Chindemi, P. Dłotko, R. Levi, K. Hess, and H. Markram, Front. Comput. Neurosci. **11**, 48 (2017).
[6] A. Patania, G. Petri, and F. Vaccarino, EPJ Data Sci. **6**, 18 (2017).
[7] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.121.228301 for details on the computations and additional numerical results.
[8] G. Bianconi and C. Rahmede, Phys. Rev. E **93**, 032315 (2016).
[9] G. Bianconi, C. Rahmede, and Z. Wu, Phys. Rev. E **92**, 022815 (2015).
[10] O. T. Courtney and G. Bianconi, Phys. Rev. E **93**, 062311 (2016).
[11] J.-G. Young, G. Petri, F. Vaccarino, and A. Patania, Phys. Rev. E **96**, 032312 (2017).
[12] A. R. Benson, R. Abebe, M. T. Schaub, A. Jadbabaie, and J. Kleinberg, arXiv:1802.06916.
[13] L.-D. Lord, P. Expert, H. M. Fernandes, G. Petri, T. J. Van Hartevelt, F. Vaccarino, G. Deco, F. Turkheimer, and M. L. Kringelbach, Front. Syst. Neurosci. **10**, 85 (2016).
[14] A. E. Sizemore, C. Giusti, A. Kahn, J. M. Vettel, R. F. Betzel, and D. S. Bassett, J. Comput. Neurosci. **44**, 115 (2018).
[15] A. Hatcher, *Algebraic Topology* (Cambridge University Press, Cambridge, England, 2005).
[16] P. Holme and J. Saramäki, Phys. Rep. **519**, 97 (2012).
[17] P. Holme, Eur. Phys. J. B **88**, 234 (2015).
[18] J. Stehlé, A. Barrat, and G. Bianconi, Phys. Rev. E **81**, 035101 (2010).
[19] N. Perra, B. Gonçalves, R. Pastor-Satorras, and A. Vespignani, Sci. Rep. **2**, 469 (2012).
[20] C. L. Vestergaard, M. Génois, and A. Barrat, Phys. Rev. E **90**, 042805 (2014).
[21] M. Karsai, N. Perra, and A. Vespignani, Sci. Rep. **4**, 4001 (2014).
[22] K. Sun, A. Baronchelli, and N. Perra, Eur. Phys. J. B **88**, 326 (2015).
[23] M. Nadini, K. Sun, E. Ubaldi, M. Starnini, A. Rizzo, and N. Perra, Sci. Rep. **8**, 2352 (2018).
[24] H. Kim, M. Ha, and H. Jeong, Phys. Rev. E **97**, 062148 (2018).
[25] N. Perra, A. Baronchelli, D. Mocanu, B. Gonçalves, R. Pastor-Satorras, and A. Vespignani, Phys. Rev. Lett. **109**, 238701 (2012).
[26] D. Horak and J. Jost, Adv. Math. **244**, 303 (2013).
[27] J. Friedman, Algorithmica **21**, 331 (1998).
[28] S. Mukherjee and J. Steenbergen, Random Struct. Algorithms **49**, 379 (2016).
[29] O. Parzanchevski and R. Rosenthal, Random Struct. Algorithms **50**, 225 (2017).
[30] A. Muhammad and M. Egerstedt, in *Proceedings of the 17th International Symposium on Mathematical Theory of Networks and Systems, Kyoto, Japan* (2006), pp. 1024–1038, http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.716.7479.
[31] D. J. Watts, Proc. Natl. Acad. Sci. U.S.A. **99**, 5766 (2002).
[32] F. Karimi and P. Holme, Physica A (Amsterdam) **392**, 3476 (2013).