Similarity Forces and Recurrent Components in Human Face-to-Face Interaction Networks

Marco Antonio Rodríguez Flores* and Fragkiskos Papadopoulos†
Department of Electrical Engineering, Computer Engineering and Informatics, Cyprus University of Technology, 33 Saripolou Street, 3036 Limassol, Cyprus and Social Computing Research Centre (SCRC), Cyprus University of Technology, 3036 Limassol, Cyprus

(Received 3 August 2018; revised manuscript received 1 November 2018; published 18 December 2018)

We show that the social dynamics responsible for the formation of connected components that appear recurrently in face-to-face interaction networks find a natural explanation in the assumption that the agents of the temporal network reside in a hidden similarity space. Distances between the agents in this space act as similarity forces directing their motion towards other agents in the physical space and determining the duration of their interactions. By contrast, if such forces are ignored in the motion of the agents recurrent components do not form, although other main properties of such networks can still be reproduced.

DOI: 10.1103/PhysRevLett.121.258301

Understanding the mechanisms that drive the dynamics of face-to-face interaction networks is crucial for better analyses of spreading phenomena. In particular, phenomena that evolve as fast as real-time face-to-face interactions, such as respiratory transmitted diseases, word-of-mouth information transfer, and viruses in mobile networks [1–3]. Furthermore, deriving efficient epidemic control strategies requires an accurate description of fast-evolving contagions [1,4–7]. However, a complete understanding of the processes responsible for the structural and dynamical properties of face-to-face interaction networks has been an elusive task [3,8,9].

Face-to-face interaction networks portray social interactions in closed settings such as schools, hospitals, offices, etc. A typical representation consists of a series of network snapshots. Each snapshot corresponds to an observation interval, which can span from a few seconds to several minutes depending on the devices used to collect the data [10,11]. The agents (nodes) in each snapshot are individuals and an edge between any two agents represents a direct face-to-face interaction.

Analyses of such networks have uncovered universal properties, such as the heavy-tailed distributions of the interaction duration and time between consecutive interactions, cf. [12]. Previous results point to the idea of social attractiveness as a mechanism responsible for these universal properties and for other structural characteristics of the time-aggregated network of contacts, like its degree, weight, and strength distributions [11,13,14]. Specifically, in the attractiveness model [13,14] agents have an activation probability r_i and a *global attractiveness* value a_i that are sampled uniformly at random from [0, 1]. Time is slotted and in each slot each noninteracting agent i is active with probability r_i. Active agents perform *random walks* in a closed Euclidean space moving towards a random direction every slot with a constant velocity (displacement) v. Agents stop moving to interact whenever they encounter another agent within a threshold distance d. The activation probability represents the activeness of each agent in the social event. The global attractiveness of the agents defines an *escaping probability* from the interactions. For instance, an agent i that has stopped moving in order to interact with other agents within distance d, can resume mobility with probability $1 - \max_{j\in\mathcal{N}_i}\{a_j\}$, where $\mathcal{N}_i$ is the set of agents interacting with i [13]. Therefore, longer interactions occur when an individual with a high global attractiveness a_j is involved.

However, it has been recently revealed that face-to-face interaction networks exhibit structural and dynamical properties such as community formation, which originate from motion patterns that are far from random [8]. In a temporal setting, communities are dynamic, meaning that their structure and size change over time. A common strategy to track dynamic communities is to construct their evolution timelines by aggregating connected components of at least three nodes in different time slots, according to some similarity measure [8,15]. In other words, the building blocks of dynamic communities are connected components that appear recurrently. If we extract the connected components in each time slot of a real face-to-face interaction network, we can see that many of the exact same components appear several times throughout the observation period. Indeed, in Figs. 1(a)–1(c) we have extracted and assigned identities (IDs), in order of appearance, to the unique components found in three real-world data sets from SocioPatterns [10]: a hospital, a primary school, and a high school [16–18] (see Table I and Supplemental Material [19], Secs. I, II, where we also consider a fourth data set from a conference [12]). The blue lines in Figs. 1(a)–1(c) represent *recurrent components*, i.e., components that appeared at least once in a previous

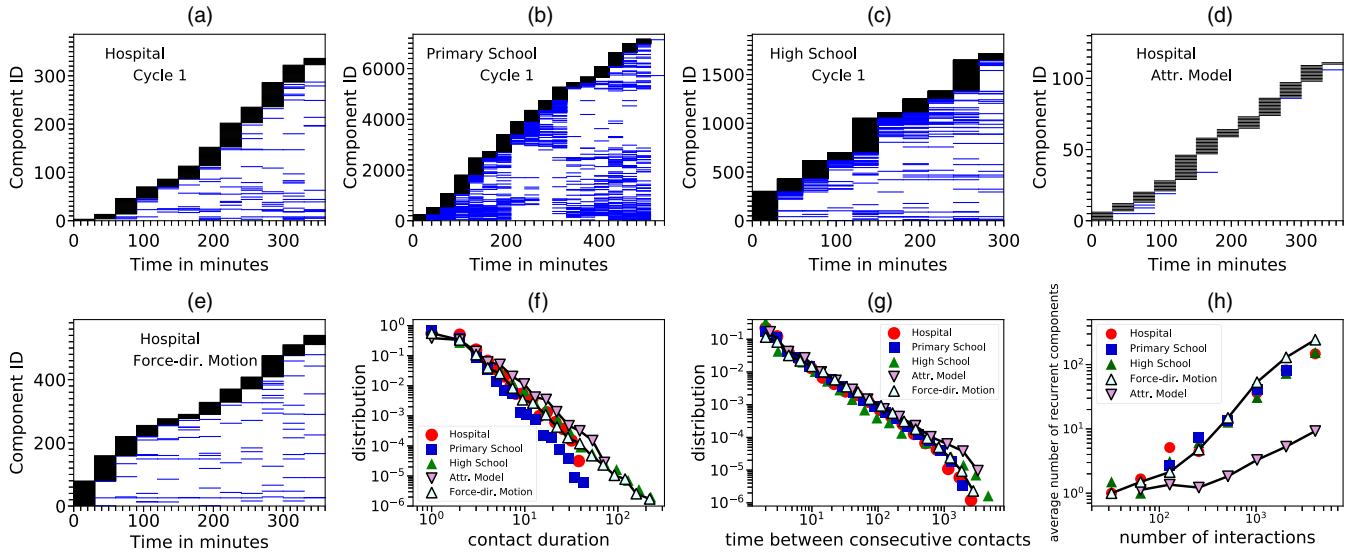

FIG. 1. Recurrent component patterns and distributions of contact durations and of times between consecutive contacts in three real-world data sets and in simulated networks. (a)–(c) Components found in the first activity cycle of the hospital, primary school, and high school (6, 8.6, and 5 h, respectively). (d) Components found in a simulation of the attractiveness model with the same duration as in (a). (e) Same as (d) but with the FDM (force-directed motion) model. (f),(g) Distribution of contact duration and of time between consecutive contacts in real and simulated networks. (h) Average number of recurrent components where an agent participates as a function of its total number of interactions in real and simulated networks. The blue lines in (a)–(e) correspond to *recurrent* components while the black lines to components appearing for the first time, i.e., to the *unique* components. The x axis is binned into 30 min intervals, while the y axis shows the component IDs observed in each bin; all components consist of at least three nodes. The simulations with the models use the parameters of the hospital (Table I and Supplemental Material [19], Sec. IV). In (f)–(h) the results with the models are averages over 10 simulation runs. Results for all activity cycles, the conference data set, and for the simulated counterparts of the rest of the real networks are found in Supplemental Material [19], Secs. II, III.

time interval. By contrast, in the attractiveness model we observe very few recurrent components [Fig. 1(d) and Supplemental Material [19], Sec. II], even though the model accurately reproduces the broad distributions of contact durations and of times between consecutive contacts [Figs. 1(f) and 1(g)]. This is because in the model nodes drift according to their own random trajectories and the probability for a group of at least three nodes to meet again is vanishing. In other words, components form in this model purely based on chance.

Here, we present a model of mobile agents where their motion is not totally random, but instead it is also directed by pairwise similarity forces. We show that this model can capture the most distinctive features of face-to-face interaction networks including their observed recurrent component patterns. In addition to the two-dimensional Euclidean space where agents move and interact (an $L \times L$ square), agents in the model also reside in a hidden similarity space, where coordinates abstract their similarity attributes. Distances between the agents in this space act as similarity forces directing their motion towards other agents in the physical space and determining the duration of their interactions. We consider the simplest metric space as the similarity space, which is a circle of radius $R = N/2\pi$ where each agent $i = 1, 2, \ldots, N$ is assigned a random angular coordinate $\theta_i \in [0, 2\pi]$. Therefore, the similarity distance between two agents i, j is $s_{ij} = R\Delta\theta_{ij}$, where $\Delta\theta_{ij} = \pi - |\pi - |\theta_i - \theta_j||$ is the angular distance between the agents. (We also consider nonuniformly distributed coordinates in Supplemental Material [19], Sec. VIII, obtaining similar results.)

Time in the model is slotted and at the beginning of each slot agents can be in one of two states: *inactive* or *interacting*. Inactive agents move in the slot only if they become active, while interacting agents move only if they escape their interactions. At the beginning of each slot t, each inactive agent i is activated with a preassigned probability r_i. Furthermore, each interacting agent i escapes its interactions with probability

TABLE I. Analyzed data sets. N is the total number of agents; T is the total duration of the data set in slots of 20 sec; $\bar{n}$, $\bar{l}$ are the average numbers of interacting agents and links (interactions) per slot. The activity cycles correspond to observation periods in different days (see Supplemental Material [19], Sec. I). μ_1, F_0, μ_2 are the FDM parameters used in the simulated counterpart of each real network (see text).

Data set	N	T	$\bar{n}$	$\bar{l}$	Cycles	μ_1	F_0	μ_2
Hospital	70	4400	7.09	4.7	4	0.8	0.12	0.9
Primary school	242	3100	56.38	40.57	2	0.35	0.2	0.78
High school	327	7375	41.89	25.56	5	1.2	0.11	0.86
Conference	113	7030	4.98	2.96	3	2.65	0.02	3.6

$$P_i^e(t) = 1 - \frac{1}{|\mathcal{N}_i(t)|} \sum_{j \in \mathcal{N}_i(t)} e^{-s_{ij}/\mu_1}, \quad (1)$$

where $\mathcal{N}_i(t)$ is the set of agents that i is currently interacting with and s_{ij} is the similarity distance between agents i and j. The summands in Eq. (1) can be seen as *bonding forces* that decrease exponentially with the similarity distance, while parameter $\mu_1 > 0$ is the decay constant controlling the importance of these forces as the similarity distance increases and allowing us to tune the average contact duration (Supplemental Material [19], Sec. IV). The model assumes that the contact duration in number of slots between two agents i, j is exponentially distributed with rate s_{ij}/μ_1. The discrete analog of this distribution is the geometric distribution with success probability $p_{ij} = 1 - e^{-s_{ij}/\mu_1}$. Therefore, Eq. (1) is the average of $p_{ij}, j \in \mathcal{N}_i(t)$.

Each moving agent i in the slot updates its position (x_i^t, y_i^t) according to the following motion equations:

$$x_i^{t+1} = x_i^t + \sum_{j \in \mathcal{S}(t)} F_{ij} \frac{(x_j^t - x_i^t)}{\sqrt{(x_j^t - x_i^t)^2 + (y_j^t - y_i^t)^2}} + R_i^x, \quad (2)$$

$$y_i^{t+1} = y_i^t + \sum_{j \in \mathcal{S}(t)} F_{ij} \frac{(y_j^t - y_i^t)}{\sqrt{(x_j^t - x_i^t)^2 + (y_j^t - y_i^t)^2}} + R_i^y, \quad (3)$$

where $\mathcal{S}(t)$ is the set of all moving and interacting agents in the slot, while F_{ij} is the magnitude of the *attractive force* between agents i and j, which also decreases exponentially with their similarity distance,

$$F_{ij} = F_0 e^{-s_{ij}/\mu_2}. \quad (4)$$

Parameter $F_0 \geq 0$ is the force magnitude at the minimum similarity distance, $s_{ij} = 0$, while $\mu_2 > 0$ is the decay constant controlling the importance of the force magnitude as the similarity distance increases. Therefore, the sums in Eqs. (2) and (3) are the total attractive forces exerted to agent i by the agents $j \in \mathcal{S}(t)$ along the x and y directions of the motion. The random motion components are $R_i^x = v \cos\phi_i$, $R_i^y = v \sin\phi_i$, where ϕ_i is sampled uniformly at random from $[0, 2\pi]$ and $v \geq 0$ is the magnitude of the random displacement. We can think of R_i^x, R_i^y as accounting for omitted degrees of freedom, akin to Langevin dynamics [25]. At $v = 0$ the motion becomes deterministic, while at $F_0 = 0$ it degenerates to random walks. Once the moving agents update their positions they either transition to the interacting state if they are within interaction range d from other noninactive agents, or to the inactive state. We call the described model *force-directed motion* (FDM) model. We make its implementation available at [26].

To understand how the formation of components depends on F_0, μ_2, v, we first consider deterministic motion. In this case, the magnitude of the expected agent displacement is controlled by F_0 and μ_2. This magnitude can be kept fixed if, when F_0 decreases, μ_2 increases accordingly. As μ_2 increases, larger components form that involve agents at larger similarity distances, until the agents eventually collapse into a giant component. At the same time, the number of components initially increases and then decreases, see Fig. 2(a). The motion in Eqs. (2) and (3) is deterministic motion with random noise. This noise decreases the chances for similar—close in the similarity space—agents to meet, which reduces the size of components. At the same time, it can either increase (if its magnitude v is sufficiently small) or decrease (if v is sufficiently large) the number of components [Fig. 2(b)].

To tune FDM's parameters in simulations of real networks we follow the procedure in the Supplemental Material [19], Sec. IV. In a nutshell, we fix $v = d = 1$. The number of agents N and time slots T are the same as in the real networks (Table I). The activation probability r_i is either $r_i = 0.5$ for every agent i (primary school and high school), or sampled uniformly at random from [0, 1]. Parameters μ_1, F_0, μ_2 (Table I) and the size of the Euclidean space L (Supplemental Material [19], Table I) are adjusted in order to approximately match the following quantities between simulated and real networks: (i) the average contact duration (using μ_1); (ii) the average number of recurrent components per interval of 10 min, while ensuring a similar size of the largest component formed (using F_0, μ_2); and (iii) the average agent degree in the time-aggregated network (using L).

In Fig. 1(e) we see that the FDM can reproduce a similar pattern of unique and recurrent components as in the hospital [Fig. 1(a)], in stark contrast to the attractiveness

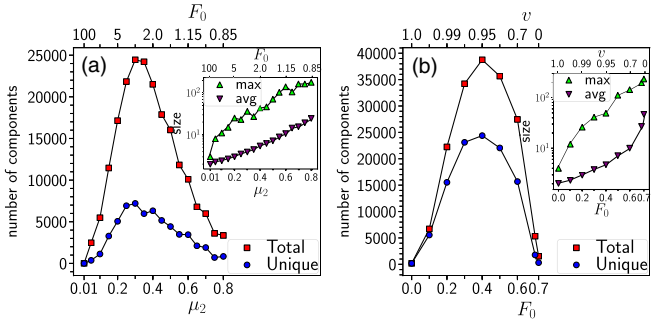

FIG. 2. Formation of components in the FDM. (a) Number of components formed (total and unique) in deterministic motion ($v = 0$) for pairs of parameters μ_2 (bottom x axis) and F_0 (top x axis). (b) Same as (a) but for pairs of F_0 and $v \geq 0$. In both (a),(b) as one parameter increases the other decreases so that the expected agent displacement per slot is always $\approx d = 1$. The insets show the maximum and average size across all components. In both plots $N = 242$, in (b) $\mu_2 = 1$. See also Supplemental Material [19], Sec. IV.

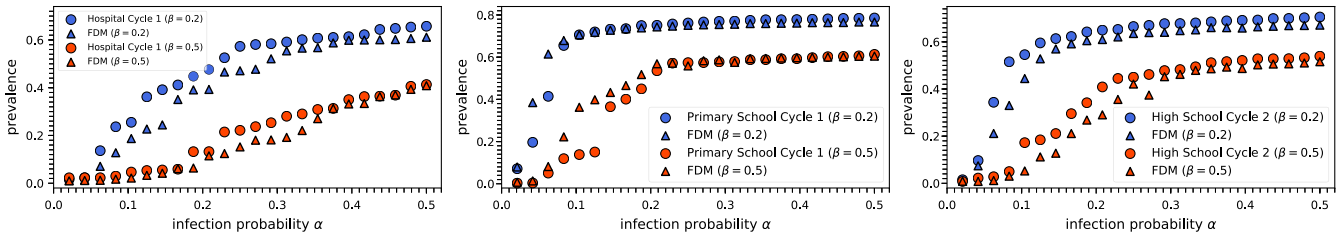

FIG. 3. Average percentage of infected agents per time slot (prevalence) of the SIS model as a function of the infection probability α in real and simulated networks (circles and triangles, respectively), for two recovery probabilities β. In the SIS each agent can be in one of two states, susceptible or infected. At any time slot an infected agent recovers with probability β and becomes susceptible again, whereas infected agents infect the susceptible agents with whom they interact, with probability α. To simulate the SIS process on temporal networks we use the dynamic SIS implementation of the Network Diffusion Library [27]. See Supplemental Material [19], Sec. VII, for further details.

model [Fig. 1(d)]. Similar results hold for all cycles of activity and for all considered data sets (Supplemental Material [19], Sec. II). In Fig. 1(h) we also see that the model can capture the correlations between the average number of recurrent components where an agent participated and the total number of interactions of the agent (see also Supplemental Material [19], Sec. II. C). At the same time, the model reproduces the broad distributions of contact durations and of times between consecutive contacts [Figs. 1(f) and 1(g)]. The model also adequately reproduces a range of other properties of the considered real networks, including weight distributions, distributions of component sizes and of shortest time-respecting paths, and group interaction durations (Supplemental Material [19], Sec. III). It is then not surprising that the susceptible-infected-susceptible (SIS) spreading process [28] behaves similarly in real and simulated networks (Fig. 3). Figure 4 shows that agents close in the similarity space tend to stay closer to each other in the Euclidean space throughout the simulations and interact more often, as expected.

The exponential form of the attractive force in Eq. (4) promotes locality and the formation of small components, as observed in real data. This is also promoted by the metric property of the similarity space, i.e., the triangle inequality,

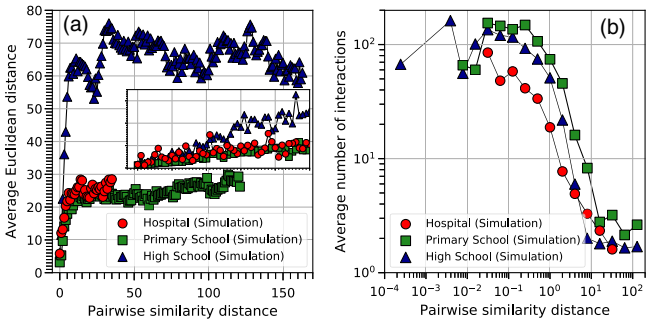

FIG. 4. Average Euclidean distance and number of interactions between two agents as a function of their similarity distance, in simulated counterparts of the Hospital, primary school, and high school. The inset in (a) is an enlarged view of similarity distances up to 5.

which ensures that if an agent a is similar to an agent b and b is similar to a third agent c, then c is also similar to a. This means that these agents will tend to gather close to each other in the Euclidean space forming triangle abc. On the other hand, if similarity distances do not satisfy the triangle inequality, then agents a and c might be close to some other agents d and e, forming chain $dabce$ in the network. In other words, agents will tend to form larger components. We verify this argument in the Supplemental Material [19], Sec. VI, where we break the triangle inequality by randomly assigning similarity distances to all pairs of agents instead of assigning to the agents similarity coordinates. In this way forces lose their localization effect and we see that a giant connected component, nonexistent when the similarity space satisfies the metric property, forms in the middle of the Euclidean space.

In summary, forces emerging from similarity distances in metric spaces appear to provide a natural explanation for the observed recurrent component dynamics in face-to-face interaction networks. These forces direct the motion of the agents in the physical space and determine the agents' interaction durations. Motion based on these principles can still capture a wide range of other main properties of such networks, in addition to their recurrent component patterns. The interactions do not have to be exactly face-to-face or of few activity cycles. In the Supplemental Material [19], Secs. II and III, we see that similar results hold in a longitudinal data set from an MIT dormitory, where proximity was captured if mobile phones were within 10 m from each other [29].

The modeling approach we consider bears similarities to N-body simulations and Langevin dynamics [25], suggesting that similar techniques and approaches from these well established areas of physics can be applicable to contemporary network science problems. Yet, we note that the similarity forces in our case only direct the motion of the agents in the physical space, and do not depend on the agents' distances in this space akin to gravity.

We also observe that *hyperbolic* spaces appear to underlie the topologies of traditional complex networks, whose degree distributions are heterogeneous [30]. In this case,

the hidden distance between two nodes is not just the angular distance $R\Delta\theta$ but the effective distance $\chi = R\Delta\theta/(\kappa\kappa')$, where κ, κ' are the expected degrees of the nodes [30]. One can replace angular with effective distances in the FDM. However, in all data sets we considered, the distribution of κ s was quite homogeneous to justify the need for this description [31]. Indeed, if we use effective distances in the FDM with the estimated κ s from the real data we obtain very similar results (Supplemental Material [19], Sec. IX).

A natural direction for future work is the *inverse problem* of inferring the similarity coordinates of agents given a sequence of real network snapshots. Another direction is extending the model with the addition of static nodes that exist both in the physical and in the similarity space and represent locations. Finally, it would be interesting to investigate how social influence could also be incorporated into the model, where interacting agents may influence each other and become more similar [32]. This would result in the agents moving both in the Euclidean and similarity spaces. Taken altogether, our results pave the way towards more realistic modeling of face-to-face interaction networks, which is crucial for understanding and predicting social group dynamics and designing efficient epidemic control and navigation [33,34] strategies.

The authors acknowledge support by the EU H2020 NOTRE project (Grant No. 692058).

*mj.rodriguezflores@edu.cut.ac.cy
†f.papadopoulos@cut.ac.cy

[1] A. Barrat and C. Cattuto, Face-to-face interactions, in *Social Phenomena: From Data Analysis to Models* (Springer, New York, 2015), pp. 37–57.
[2] N. Masuda and R. Lambiotte, *A Guide to Temporal Networks* (World Scientific, Singapore, 2016).
[3] P. Holme, Modern temporal network theory: A colloquium, Eur. Phys. J. B **88**, 234 (2015).
[4] P. Holme, Temporal network structures controlling disease spreading, Phys. Rev. E **94**, 022305 (2016).
[5] P. Holme and N. Litvak, Cost-efficient vaccination protocols for network epidemiology, PLoS Comput. Biol. **13**, e1005696 (2016).
[6] M. Karsai and N. Perra, Control strategies of contagion processes in time-varying networks, in *Temporal Network Epidemiology* (Springer, Singapore, 2017), pp. 179–197.
[7] M. Ogura and V. M. Preciado, Optimal containment of epidemics in temporal and adaptive networks, in *Temporal Network Epidemiology* (Springer, Singapore, 2017), pp. 241–266.
[8] V. Sekara, A. Stopczynski, and S. Lehmann, Fundamental structures of dynamic social networks, Proc. Natl. Acad. Sci. U.S.A. **113**, 9977 (2016).
[9] H. Barbosa, M. Barthelemy, G. Ghoshal, C. R. James, M. Lenormand, T. Louail, R. Menezes, J. J. Ramasco, F. Simini, and M. Tomasini, Human mobility: Models and applications, Phys. Rep. **734**, 1 (2018).
[10] Sociopatterns, http://www.sociopatterns.org/.
[11] M. Starnini, B. Lepri, A. Baronchelli, A. Barrat, C. Cattuto, and R. Pastor-Satorras, Robust modeling of human contact networks across different scales and proximity-sensing techniques, in *Social Informatics* (Springer International Publishing, Cham, 2017), pp. 536–551.
[12] L. Isella, J. Stehlé, A. Barrat, C. Cattuto, J.-F. Pinton, and W. Van den Broeck, What's in a crowd? Analysis of face-to-face behavioral networks, J. Theor. Biol. **271**, 166 (2011).
[13] M. Starnini, A. Baronchelli, and R. Pastor-Satorras, Modeling Human Dynamics of Face-to-Face Interaction Networks, Phys. Rev. Lett. **110**, 168701 (2013).
[14] M. Starnini, A. Baronchelli, and R. Pastor-Satorras, Model reproduces individual, group and collective dynamics of human contact networks, Soc. Networks **47**, 130 (2016).
[15] D. Greene, D. Doyle, and P. Cunningham, Tracking the evolution of communities in dynamic social networks, in *Proceedings of the 2010 International Conference on Advances in Social Networks Analysis and Mining, ASONAM '10* (IEEE Computer Society, Washington, DC, 2010), pp. 176–183.
[16] P. Vanhems, A. Barrat, C. Cattuto, J.-F. Pinton, N. Khanafer, C. Régis, B. Kim, B. Comte, and N. Voirin, Estimating potential infection transmission routes in hospital wards using wearable proximity sensors, PLoS One **8**, e73970 (2013).
[17] J. Stehlé, N. Voirin, A. Barrat, C. Cattuto, L. Isella, J.-F. Pinton, M. Quaggiotto, W. Van den Broeck, C. Régis, B. Lina, and P. Vanhems, High-resolution measurements of face-to-face contact patterns in a primary school, PLoS One **6**, e23176 (2011).
[18] R. Mastrandrea, J. Fournet, and A. Barrat, Contact patterns in a high school: A comparison between data collected using wearable sensors, contact diaries and friendship surveys, PLoS One **10**, e0136497 (2015).
[19] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.121.258301 for more information and supplementary discussions, which includes Refs. [20–24].
[20] B. A. Galler and M. J. Fisher, An improved equivalence algorithm, Commun. ACM **7**, 301 (1964).
[21] Pathpy, https://github.com/IngoScholtes/pathpy.
[22] A. Muscoloni and C.-V. Cannistraci, A nonuniform popularity-similarity optimization (nPSO) model to efficiently generate realistic complex networks with communities, New J. Phys. **20**, 052002 (2018).
[23] G. J. McLachlan and D. Peel, *Finite Mixture Models*, Wiley Series in Probability and Statistics (Wiley, New York, 2000).
[24] F. Papadopoulos, M. Kitsak, M. Á. Serrano, M. Boguñá, and D. Krioukov, Popularity versus similarity in growing networks, Nature (London) **489**, 537 (2012).
[25] T. Schlick, *Molecular Modeling and Simulation: An Interdisciplinary Guide*, Interdisciplinary Applied Mathematics (Springer, New York, 2010).
[26] FDM simulator and data sets, https://bitbucket.org/mrodrflr/similarity_forces.
[27] G. Rossetti, L. Milli, S. Rinzivillo, A. Sîrbu, D. Pedreschi, and F. Giannotti, NDlib: A Python library to model and analyze diffusion processes over complex networks, Int. J. Data Sci. Anal. **5**, 61 (2017).
[28] M. J. Keeling and P. Rohani, *Modeling Infectious Diseases in Humans and Animals* (Princeton University Press, Princeton, 2008).

[29] W. Dong, B. Lepri, and A. S. Pentland, Modeling the co-evolution of behaviors and social relationships using mobile phone data, in *Proceedings of the 10th International Conference on Mobile and Ubiquitous Multimedia, MUM '11* (ACM, New York, 2011), pp. 134–143.

[30] D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat, and M. Boguñá, Hyperbolic geometry of complex networks, Phys. Rev. E **82**, 036106 (2010).

[31] An agent's κ is its average degree per time slot.

[32] R. Leenders, Longitudinal behavior of network structure and actor attributes: modeling interdependence of contagion and selection, in *Evolution of Social Networks* (Gordon and Breach Publishers, New York, 1997), pp. 165–184.

[33] S. H. Lee and P. Holme, Navigating temporal networks, Physica (Amsterdam) **513A**, 288 (2019).

[34] M. Boguñá, D. Krioukov, and K. C. Claffy, Navigability of complex networks, Nat. Phys. **5**, 74 (2009).