P H Y S I C A L   R E V I E W   L E T T E R S

Unfolding Accessibility Provides a Macroscopic Approach to Temporal Networks

Hartmut H. K. Lentz

Institute for Physics, Humboldt-University of Berlin, Newtonstrasse 15, 12489 Berlin, Germany and Institute of Epidemiology, Friedrich-Loeffler-Institute, Seestrasse 55, 16868 Wusterhausen, Germany

Thomas Selhorst

Institute of Epidemiology, Friedrich-Loeffler-Institute, Seestrasse 55, 16868 Wusterhausen, Germany

Igor M. Sokolov

Institute for Physics, Humboldt-University of Berlin, Newtonstrasse 15, 12489 Berlin, Germany
(Received 8 October 2012; published 11 March 2013)

An accessibility graph of a network contains a link wherever there is a path of arbitrary length between two nodes. We generalize the concept of accessibility to temporal networks. Building an accessibility graph by consecutively adding paths of growing length (unfolding), we obtain information about the distribution of shortest path durations and characteristic time scales in temporal networks. Moreover, we define causal fidelity to measure the goodness of their static representation. The practicability of our proposed methods is demonstrated for three examples: networks of social contacts, livestock trade, and sexual contacts.

Many real-world systems can be described in terms of networks [1], i.e., sets of nodes, or vertices, and edges connecting node pairs. Prominent examples are social contacts, flow of people between cities, trade between countries, or the World Wide Web [2–6]. In many cases, a static network description is sufficient: an edge is drawn whenever a connection between the corresponding nodes was established during the whole data acquisition time. Mathematically, static networks are considered as graphs and can be represented by adjacency matrices. These contain unit elements in the entries corresponding to the indices of nodes connected by an edge, all other elements being zeros. This representation is used in a large number of algebraic methods for network analysis [7,8].

Every route across the network connecting two vertices along the network's edges is called a path. For a macroscopic view of a network, it is often sufficient to know whether the nodes are connected by a path of whatever length, i.e., whether one node is accessible starting from another one. Accessibility can be mapped onto a single mathematical object, the accessibility matrix of the network, containing unit entries for indices corresponding to the pairs of the nodes connected by paths of arbitrary length [7,8]. This matrix is the adjacency matrix of the accessibility graph. To build this graph, one starts with the network itself and first adds edges between the nodes connected by paths of length 2. Then one subsequently adds longer paths as edges step by step (Fig. 1) until there are no more new paths to add. It turns out that valuable insights into the network structure are not only given by the accessibility graph itself, but also by the graphs obtained on intermediate stages of its step-by-step construction.

We call this step-by-step construction the unfolding of the accessibility graph.

Although the concept of accessibility is plausible in static networks, it is still not clear how to generalize it to networks, where edges are not constant over time. Many systems have to be represented as temporal networks: their nodes being fixed, but edges, representing connections between them, are only present during certain periods of time [9–14]; see Refs. [15,16] for a review. Temporal networks are not graphs and cannot in general be represented by single matrices. Therefore, most attention on temporal networks has focused on data-driven approaches [10,12,17]. Coevolutionary models have been proposed to understand the mechanisms behind network formation [18,19]. Work concerning macroscopic properties of temporal networks is rare. This is partly due to the absence of an established formalism to treat these objects [15,20]. A time-aggregated, static representation of a temporal network may contain paths which do not follow a causal sequence of connections between nodes, i.e., paths which
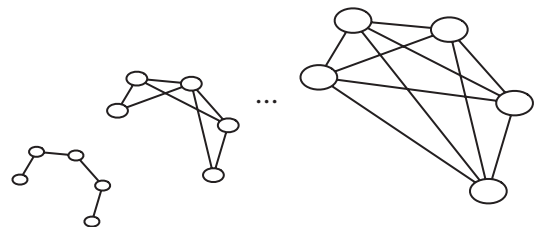

FIG. 1. Unfolding of an accessibility graph. Starting with a network (left) we unfold its accessibility graph by adding direct edges for paths of length 2 (center) and so forth. The fully unfolded accessibility graph is shown on the right.

cannot be taken in a real system. Therefore, a static representation may be insufficient for the description of a temporal network.

The first step for the introduction of a formalism for temporal networks will thus be taking into account the causality of paths. In the present Letter we introduce accessibility graphs of temporal networks and use them for a description of their macroscopic properties. In addition, we define the measure of causal fidelity quantifying the goodness of a static representation of a temporal network. We first briefly review the basic concept of accessibility in static networks and introduce the method of unfolding accessibility for the static case. Then we discuss generalizations of the concept of accessibility to temporal networks. Finally, we analyze three different data sets with respect to their characteristic time scales and compare their temporal and static representations.

Let us consider a static network $G = (V, E)$, where V is the set of vertices and E is the set of edges. The network is represented by an adjacency matrix $\mathbf{A}$, with $(\mathbf{A})_{uv} = 1$ if there is an edge between nodes u and v, and $(\mathbf{A})_{uv} = 0$ otherwise. The accessibility graph (transitive closure) of a static network is denoted by $G^* = (V, E^*)$. It contains an edge $(u, v) \in E^*$ if a path of arbitrary length exists between the nodes u and v. Given an adjacency matrix $\mathbf{A}$, the elements of the matrix $\mathbf{A}^n$ are the numbers of paths of length n connecting pairs of nodes of the graph. Consequently, the cumulative matrix $\mathbf{C}_n = \mathbf{A} + \mathbf{A}^2 + \cdots + \mathbf{A}^n$ has a nonzero entry $(\mathbf{C}_n)_{uv} \neq 0$ if the nodes u and v are connected by a path of whatever length smaller than or equal to n. Summing up all powers of the adjacency matrix, i.e., building a matrix $\mathbf{C}_{n=N} = \sum_{i=1}^{N} \mathbf{A}^i$ (where N is the total number of nodes in the network), is related to the transitive closure of the network.

Provided that we are interested only in the number of nodes that can be reached from each initial node within n steps, we record whether the elements of $\mathbf{C}_n$ are zero or not. Changing every nonzero element of $\mathbf{C}_n$ to unity yields the accessibility matrix $\mathbf{P}_n$. The matrix $\mathbf{P}_n$ (with binary entries) can be treated as a Boolean matrix. Alternatively, one can consider the adjacency matrix $\mathbf{A}$ as a Boolean matrix from the very beginning. The product of two Boolean matrices is defined by using normal algebra, but Boolean arithmetics, interpreting multiplication as a logical AND ($\wedge$) operation and addition as a logical OR ($\vee$) operation. The product of two Boolean matrices is $(\mathbf{AB})_{ij} = \bigvee_{k=1}^{N} a_{ik} \wedge b_{kj}$. In this notation, the accessibility matrix $\mathbf{P}_n$ of path length n becomes [21]

$$\mathbf{P}_n = \bigvee_{i=1}^{n} \mathbf{A}^i, \tag{1}$$

where the ith power of the Boolean matrix $\mathbf{A}$ is computed using Boolean matrix products. A brief discussion of some properties of the fully exploited accessibility graph $\mathbf{P}_{n=N}$ is given in the Supplemental Material [22]. The number of

edges in $\mathbf{P}_1$ gives the number of paths of length 1, and at every step $n \to n + 1$ the network is traversed by paths with one more edge. Thus, Eq. (1) corresponds to a breadth-first search (BFS) in the network. This relation is also used in the Floyd-Warshall algorithm [21,23] for the computation of shortest path lengths. When the BFS reaches the diameter D of the network, i.e., there are no more shortest paths to traverse, the accessibility matrix saturates, $\mathbf{P}_N = \mathbf{P}_D$.

The total number of nonzero elements of the accessibility matrix $\mathbf{P}_n$ is associated with the distribution of shortest path lengths. We define the density $\rho(\mathbf{A})$ of an $N \times N$ matrix $\mathbf{A}$ with N^2 entries and $\mathrm{nnz}(\mathbf{A})$ nonzero elements as

$$\rho(\mathbf{A}) = \frac{\mathrm{nnz}(\mathbf{A})}{N^2}. \tag{2}$$

When $\mathbf{A}$ is an adjacency matrix, $\rho(\mathbf{A})$ defines the edge density of the network, giving the probability to find an edge between two nodes chosen at random. In analogy, the probability to find a path of length $l \leq n$ between two randomly chosen nodes in a connected network is given by

$$F_n = P(l \leq n) \equiv \rho(\mathbf{P}_n). \tag{3}$$

Consequently, the path density $\rho(\mathbf{P}_n)$ gives the cumulative distribution of shortest path lengths in the network. The corresponding probability to find the shortest path of length n is given by $f_n = (F_n - F_{n-1})$, with $F_0 = 0$. The distribution saturates when n reaches the diameter D of the network. For a single-component network $F_D = \rho(\mathbf{P}_D) = 1$. For networks composed of disjoint components $\rho(\mathbf{P}_D) < 1$, since their accessibility matrices have less than N^2 nonzero entries [22]. In this case we consider F_n as an "improper" probability distribution, which is not normalized to unity, and define its median as the value of n, where F_n approaches one-half of its saturation value F_D.

Temporal networks allow edges to vary over time, and are represented by a sequence of adjacency matrices $\mathcal{A} = \{\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_T\}$, where each matrix $\mathbf{A}_t$ is a snapshot of the system at time t. The increment corresponds to the temporal resolution and T is the maximal time given in the data. The accessibility graph $\mathcal{G}^* = (V, E^*)$ of a temporal network has to allow for only causal paths (time respecting paths), in which the temporal order of edges has to be taken into account: it contains an edge $(u, v) \in E^*$ if and only if there is a causal path between the nodes u and v. A time-aggregated, static representation follows by summing up all (Boolean) matrices in the sequence, $\mathbf{A} = \vee_{t=1}^{T} \mathbf{A}_t$.

As we have seen for the static case, the product of two identical adjacency matrices gives information about the number of paths of length 2. If we multiply different matrices, i.e., two matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ in the sequence $\mathcal{A}$, we obtain nonzero entries wherever nodes receive edges at time 1 and cast forth edges at time 2. It is important to emphasize that the entries $(\mathbf{A}_1\mathbf{A}_2)_{ij}$ vanish if an edge received by node i at time 1 is not cast forth to

node j at exactly time 2. Using products of different adjacency matrices in the sequence $\mathcal{A}$, we could derive a dynamic accessibility matrix exactly as it was done in the static case. Many systems, however, show significant waiting times between node activities [24]. Therefore, this straightforward approach does not give useful results because it does not include the possibility of node waiting times (see Supplemental Material [22]).

To overcome this drawback, we allow for temporal shortcuts in the accessibility matrix by taking into account products of adjacency matrices corresponding to remote time steps. As an example, if many nodes in the system receive links at time 5 and cast forth links at time 100, the product $\mathbf{A}_5\mathbf{A}_{100}$ should make a large contribution to the accessibility matrix. To introduce shortcuts, we define the accessibility matrix of a temporal network by adding an identity matrix $\mathbf{1}$ to each matrix $\mathbf{A}_i$ in the sequence. This yields all ordered products of snapshot matrices in time: $\mathcal{C}_n = \prod_{i=1}^{n}(\mathbf{1} + \mathbf{A}_i)$. The addition of an identity matrix can be interpreted as the ability of each node to keep edge information over time; i.e., it introduces memory into the system. We have to point out that this analysis focuses on systems with infinite memory.

Since we are not interested in the actual number of paths between two nodes, we can either set any nonzero element of $\mathcal{C}_n$ to unity or, equivalently, use the Boolean formulation

$$\mathcal{P}_n = \bigwedge_{i=1}^{n}(\mathbf{1} \vee \mathbf{A}_i)$$
$$= \mathbf{1} \vee \mathbf{A}_1 \vee \mathbf{A}_2 \vee \mathbf{A}_3 \vee \cdots \vee \mathbf{A}_n \vee \mathbf{A}_1\mathbf{A}_2$$
$$\vee \mathbf{A}_2\mathbf{A}_3 \vee \mathbf{A}_1\mathbf{A}_3 \vee \mathbf{A}_1\mathbf{A}_2\mathbf{A}_3 \vee \ldots. \quad (4)$$

Linear terms in (4) correspond to the aggregated network. Higher order products (always respecting a causal sequence of indices) represent all possible direct connections or temporal shortcuts. It should be noted that, due to the causality of the paths, the temporal accessibility graph of an undirected network is in general directed. The unfolding of the accessibility matrix given in (4) can now be used to compute shortest-path distributions F_n for temporal networks in the same manner as for the static case. Note that the index n now has the meaning of the duration of the path (total number of time steps necessary to traverse the system, regarding waiting times). It defines the time (in units determined by the temporal resolution of the used data) which, say, the passengers spent in motion. Thus, a temporal network may still show a small-world property as long as its time-aggregated representation is a small-world network. However, if the the links between nodes are activated only during distant periods in time, the spread over the network would be slow: even a small world can be a ''slow world.''

The fully unfolded accessibility graph can in principle be used to detect temporal connected components in the network. Therefore, it is sufficient to detect cliques in the

TABLE I.   Temporal network data sets.

Data set	Size	Snapshots	t resolution	Type	Source
Conference	113	10 618	20 s	Undirected	[26]
Pig trade	89 745	465	1 day	Directed	[28]
Sexual contacts	16 730	2232	1 day	Undirected	[27]

symmetric part of $\mathcal{P}_{n=N}$ [25]. As mentioned in Ref. [25], finding cliques is known to be NP complete even in static networks and cliques are in general degenerated in temporal systems.

We have to point out that the identity matrix on the right-hand side of (4) is an artifact of the introduced memory term. It does not make any difference in undirected networks, whose accessibility matrices always exhibit nonvanishing diagonal entries for $n \geq 2$ [22], but may make a difference in directed networks. However, since the number of diagonal elements is small compared to the total number of elements of a large matrix, this deviation is negligible in large networks.

Temporal networks are often approximated by their time-aggregated counterparts. However, some paths in a static network representation do not follow a chronological order and are never taken. In order to quantify the quality of a static representation, we define the causal fidelity c as the fraction of the number of paths in a static network which can also be taken in a temporal one:

$$c = \frac{\rho(\mathcal{P}_T)}{\rho(\mathbf{P}_T)}. \quad (5)$$

The values of c lay in the interval $0 \leq c \leq 1$, where $c = 1$ means that the path density of a static representation does not differ from the temporal system. Low values of c indicate that the majority of paths in the static approximation do not follow a causal sequence of edges and thus are not present in the temporal one.

As examples, we consider several real-world networks showing different time scales ranging from 3 days [26] to 6 years [27]; see Table I.
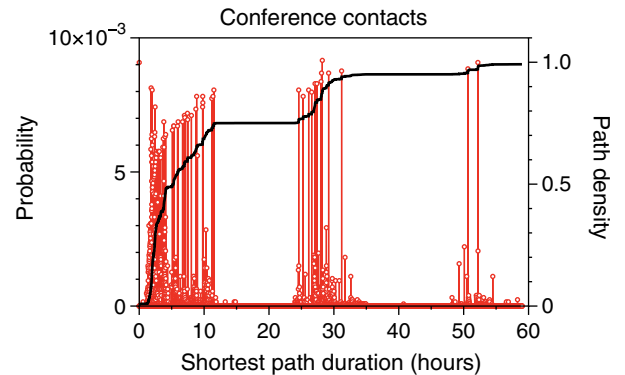

FIG. 2 (color online).   Density of the accessibility graph (black line) and distribution of shortest path durations (red line with circles) for a network of face-to-face contacts at a conference.
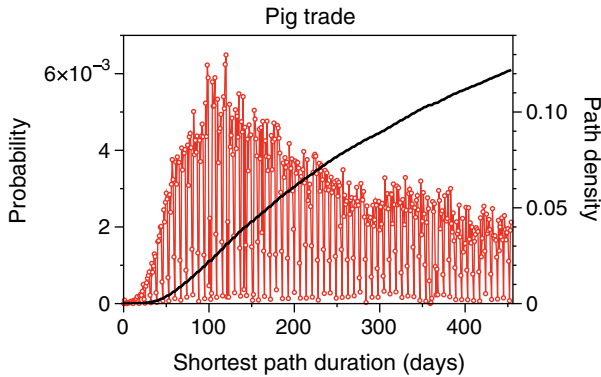
FIG. 3 (color online).   Density of the accessibility graph (black line) and distribution of shortest path durations (red line with circles) for a network of pig trade in Germany.
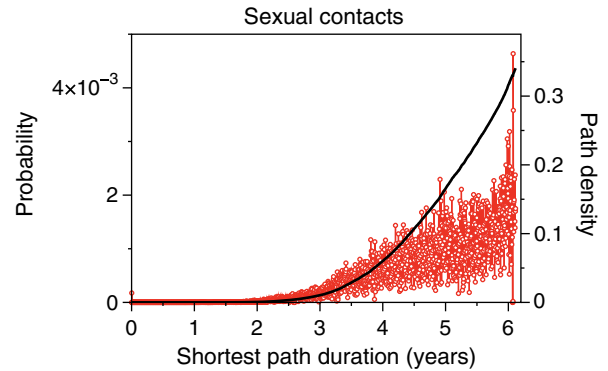

FIG. 4 (color online).   Density of the accessibility graph (black line) and distribution of shortest path durations (red line with circles) for a network of sexual contacts.

We unfold the accessibility graphs of these networks and obtain path densities and distributions of shortest path durations. Hereby, we treat the path density as a (improper) cumulative distribution. The results are shown in Figs. 2–4. In addition, we compute the causal fidelity c for each network following Eq. (5).

Figure 2 shows results for a network of conference attendees. Every face-to-face contact between two persons is recorded as an undirected edge. The data set covers a time span of 3 days, separated by time spans of no activity (nights). The density of the accessibility graph in Fig. 2 shows clear and fast saturation at a value of $\rho = 0.9917$, indicating that after only 3 days a large part of the network is causally connected. More than 70% of the possible paths are traversed within only 1 day (black line). The median of the distribution is reached in less than 6 hours. An aggregated network representation gives a good approximation of the system, which is also reflected in the high causal fidelity of $c \approx 0.99$ of its static representation.

Figure 3 shows the results for the network of pig trade in Germany over a time span of 455 days. Every node in the network is an agricultural holding and an edge is present when a holding sells livestock animals to another one. In contrast to the two other data sets, the pig trade network is directed. This is reflected in relatively low values of the path density. The shortest path lengths show a broad distribution (red). Although the temporal diameter of the pig trade network is by far larger than the observation period of 455 days, the distribution shows a global maximum (i.e., possesses a mode) at around 120 days. This means that the typical spreading time scales are of the order of 100 days. The comblike structure of the graph reflects the time inhomogeneity of trade, which is not so frequent on Saturdays and almost absent on Sundays. The static representation of the network captures its features sufficiently well ($c = 0.72$). A similar behavior can be expected for pig trade networks of other countries because these systems are determined by the typical lifetime of a livestock pig.

Figure 4 shows the distribution of shortest path durations for a network of sexual contacts. The network reflects physical contacts between prostitutes and their customers. Each contact is recorded as an undirected link. Here, no saturation of the path density is reached within the study period, and the distribution of shortest path durations does not show a clear maximum. The figure demonstrates that the majority of shortest paths takes more than 2 years to traverse. Even if time scales cannot be resolved from the data, the observation is useful to define a minimum duration of future field observations and may be of use for the implementation of vaccination strategies. The figure shows that it takes at least 3 years to obtain a considerable density of the accessibility graph. The sexual contact network is only poorly represented by its aggregated counterpart, as reflected by low causal fidelity of this representation ($c = 0.38$). A static network would significantly overestimate the number of transmission paths, and the system has to be analyzed from a temporal network perspective [27]. Our examples suggest that a primary criterion for a high causal fidelity is that the data acquisition time should be longer than the typical shortest path duration of the system, which can be inferred from $\rho(\mathcal{P}_n)$. An analysis of the causal fidelity for different systematically randomized models of the same networks is provided in the Supplemental Material [22].

To conclude, we introduced temporal accessibility graphs as a formalism to analyze temporal network data in terms of causal paths. The method is the temporal generalization of the transitive closure of a static network. It gives information about the distribution of time scales in the system and also provides an instrument for quantifying the quality of their simpler, static representation. The examples of real networks considered demonstrate the practicability and usefulness of this approach.

[1] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).

[2] S. Wasserman and K. Faust, *Social Network Analysis* (Cambridge University Press, Cambridge, England, 1994).

[3] L. Hufnagel, D. Brockmann, and T. Geisel, Proc. Natl. Acad. Sci. U.S.A. **101**, 15 124 (2004).

[4] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).

[5] F. Liljeros, C. Edling, L. Amaral, and H. Stanley, Nature (London) **411**, 907 (2001).

[6] C. Hidalgo, B. Klinger, A.-L. Barabási, and R. Hausmann, Science **317**, 482 (2007).

[7] M. E. J. Newman, *Networks: An Introduction* (Oxford University Press, New York, 2010).

[8] J. M. Aldous and R. J. Wilson, *Graphs And Applications: An Introductory Approach* (Springer, New York, 2000).

[9] M. C. Vernon and M. J. Keeling, Proc. R. Soc. B **276**, 469 (2009).

[10] P. Bajardi, A. Barrat, F. Natale, L. Savini, and V. Colizza, PLoS ONE **6**, e19869 (2011).

[11] J. Stehlé, N. Voirin, A. Barrat, C. Cattuto, V. Colizza, L. Isella, C. Régis, J.-F. Pinton, N. Khanafer, W. Van den Broeck, and P. Vanhems, BMC Med. **9**, 87 (2011).

[12] L. E. C. Rocha, F. Liljeros, and P. Holme, PLoS Comput. Biol. **7**, e1001109 (2011).

[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] M. Konschake, H. H. K. Lentz, F. J. Conraths, P. Hövel, and T. Selhorst, PLoS ONE **8**, e55223 (2013).

[15] P. Holme and J. Saramäki, Phys. Rep. **519**, 97 (2012).

[16] A. Casteigts, P. Flocchini, W. Quattrociocchi, and N. Santoro, Int. J. Parallel, Emergent Distributed Syst. **27**, 387 (2012).

[17] R. K. Pan and J. Saramäki, Phys. Rev. E **84**, 016105 (2011).

[18] M. G. Zimmermann, V. M. Eguiluz, and M. San Miguel, Phys. Rev. E **69**, 065102 (2004).

[19] T. Gross and B. Blasius, J. R. Soc. Interface **5**, 259 (2008).

[20] L. Danon, A. P. Ford, T. House, C. P. Jewell, M. J. Keeling, G. O. Roberts, J. V. Ross, and M. C. Vernon, Interdiscip. Perspect. Infect. Dis. **2011**, 284909 (2011).

[21] S. Warshall, J. Assoc. Comput. Mach. **9**, 11 (1962).

[22] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.110.118701 for unfolding accessibility of static networks, an alternative approach to temporal network accessibility and randomized models.

[23] R. W. Floyd, Commun. ACM **5**, 345 (1962).

[24] J. L. Iribarren and E. Moro, Phys. Rev. Lett. **103**, 038702 (2009).

[25] V. Nicosia, J. Tang, M. Musolesi, G. Russo, C. Mascolo, and V. Latora, Chaos **22**, 023101 (2012).

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

[27] L. E. C. Rocha, F. Liljeros, and P. Holme, Proc. Natl. Acad. Sci. U.S.A. **107**, 5706 (2010).

[28] EUR-Lex, Directive 2000/15/EC of the European Parliament and the Council of 10 April 2000 amending Council Directive 64/432/EEC on health problems affecting intracommunity trade in bovine animals and swine; http://www.hi-tier.de, data access restricted to competent authorities in Germany.

[29] http://www.sociopatterns.org.