

Flow motifs reveal limitations of the static framework to represent human interactions

Luis E. C. Rocha* and Vincent D. Blondel

Department of Mathematical Engineering, Université catholique de Louvain, Louvain-la-Neuve, Belgium

(Received 22 March 2013; published 22 April 2013)

Networks are commonly used to define underlying interaction structures where infections, information, or other quantities may spread. Although the standard approach has been to aggregate all links into a static structure, some studies have shown that the time order in which the links are established may alter the dynamics of spreading. In this paper, we study the impact of the time ordering in the limits of flow on various empirical temporal networks. By using a random walk dynamics, we estimate the flow on links and convert the original undirected network (temporal and static) into a directed flow network. We then introduce the concept of flow motifs and quantify the divergence in the representativity of motifs when using the temporal and static frameworks. We find that the regularity of contacts and persistence of vertices (common in email communication and face-to-face interactions) result on little differences in the limits of flow for both frameworks. On the other hand, in the case of communication within a dating site and of a sexual network, the flow between vertices changes significantly in the temporal framework such that the static approximation poorly represents the structure of contacts. We have also observed that cliques with 3 and 4 vertices containing only low-flow links are more represented than the same cliques with all high-flow links. The representativity of these low-flow cliques is higher in the temporal framework. Our results suggest that the flow between vertices connected in cliques depend on the topological context in which they are placed and in the time sequence in which the links are established. The structure of the clique alone does not completely characterize the potential of flow between the vertices.

DOI: [10.1103/PhysRevE.87.042814](https://doi.org/10.1103/PhysRevE.87.042814)

PACS number(s): 89.75.Hc, 89.75.Fb, 02.50.Ey

I. INTRODUCTION

Networks have been used to describe and model natural and artificial systems from different scientific domains. The common practice has been to construct a network by aggregating all available information into a single graph structure, eventually containing weights or directions on the links to reflect attributes of the system [1,2]. The main underlying assumption is that links remain fixed during the sampling interval such that all information of the system is summarized into the network topology. Although static structures provide several insights about the patterns of interaction, neglecting the timing of events results in an incomplete picture of the interaction dynamics. The time order of events may reveal cyclic behavior [3,4], bursts of activity [3,5,6], causality [7,8], and long-term correlations [6,9]. One typical illustration of the ordering effect is the absence of paths between vertices A and C if B and C are linked before A and B. In the static version, however, this path would exist either way. The temporal constraints significantly reduce the number of paths and increase the path-length between two vertices [10,11], directly affecting the centrality of vertices [12–14].

The network structure defines potential pathways for the transmission of infection, information, or other quantities. In the context of dynamical processes, several studies suggest that temporal patterns of link activity reflect into the spread by speeding up or slowing down epidemics [15–18], information [19–21], random walks [22–24], and consensus dynamics [25]. Most of the studies about dynamics on temporal networks focus on global statistics and on detecting the timing for emergent phenomena. There is little understanding, however, on local effects of the time ordering. Consider, for example,

the triangles in Fig. 1(a). If vertices A and B connect (at $t = 8$) after B and C ($t = 5$), A may infect C directly (at $t = 2$), but not through B. Paths get more complicated when another link is added to A [Figs. 1(b)–1(d)]; depending on the time-stamp of the new link, different routes are available for an infection to reach B and C (if $t = 1$), or to not reach them at all (if $t = 9$). In temporally large networks (i.e., several days or weeks), a particular sequence may not affect much the dynamics because links are likely to repeat; e.g., the time order of events in Fig. 1(a) may repeat every day and an infection would pass from A to B in one day and then to C at the day after (via B). Although intuitively one may expect that the infection or information flowing between adjacent vertices always change if time constraints are introduced, these changes in fact depend on the particularities of the system.

In this paper, we estimate the limits of flow between vertices in empirical temporal networks and compare with the respective static version. To do that, we let a random walker explore the network (both the temporal and the static versions) and measure the frequency of hops between two vertices to estimate the local flow. Since the walker moves according to the availability of links in the temporal network, the flow captures the interplay between temporal and topological constraints. In the static network, however, links are available at any moment and, thus, only the topology affects the dynamics. This procedure converts the original simple network into a flow network containing information about the dynamics on the links. The concept of flow networks has been exploited before, in the context of static networks, to characterize the community structure [26,27] and the spreading dynamics [28]. Since this measure of flow is globally optimized, in order to quantify relative changes between the temporal and the static frameworks, we split the links into two classes of high- and low-flow. To quantify the impact of temporal constraints in the

*Luis.Rocha@uclouvain.be

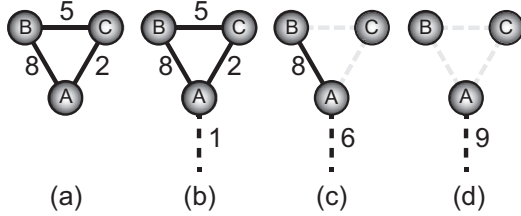


FIG. 1. Vertices A, B, and C are connected by links and the numbers correspond to the time stamps, i.e., the time in which the link is available. Given the temporal subgraph in (a), the other three sketches represent the potential paths between the vertices in (a). These potential paths vary according to the time stamp of the black dotted link. For example, in (b) there are different potential paths connecting B and C to the rest of the network, while in (d) they appear completely disconnected.

flow, we measure the representativity of flow subgraphs (or flow motifs) in temporal networks and compare with the static version. In this context, motifs are convenient because they summarize the information about the system-wide flow into small subunits that can be studied individually but collectively provide a fingerprint of the network [7,29–34].

II. MATERIALS AND METHODS

In this section, we describe the empirical networks used in this paper, the procedure to estimate the flow using the random walker dynamics on temporal and static networks, and the methodology to obtain flow subgraphs and flow motifs.

A. Temporal networks

A temporal network of size N may be defined as an ordered sequence of triplets (i, j, t) . Each triplet corresponds to a single interaction event at time t ($t = 0, 1, 2, \dots$) between a pair (i, j) of vertices i and j ($i, j = 1, 2, \dots, N$). The total number of triplets (or the number of contacts or single event interactions) is given by C and the number of unique pairs (i, j) (i.e., links in the static network) is given by E . An empirical temporal network is a set of interaction events recorded every δt (the sampling resolution) during an interval $T_i \leq t \leq T_f$. The sample interval ΔT is, thus, $T_f - T_i$. The corresponding static network is obtained by aggregating all interactions events within ΔT into a single graph structure and using the number of contacts per link as weights.

We use a selection of empirical temporal networks to represent different classes of human communication and proximity contacts. These networks represent diverse underlying structures relevant to the spread of information and of infections. The first dataset corresponds to a network of online communication between members of a dating site (POK) [35]. The second corresponds to email communication within a major university (EMA) [36]. The third corresponds to sexual contacts between sex-buyers and -sellers (SEX) and is fully bipartite by definition [6]. The other datasets correspond to face-to-face interactions between delegates in a scientific conference (SPC) and to visitors of a museum exposition (SPM) [4]. All networks are assumed undirected. Table I

TABLE I. Characteristics of the samples of the empirical networks: number of vertices (N), number of single event interactions (C), number of unique pairs of connected vertices (E), the network resolution (δt), and the sample interval ΔT .

	N	C	E	δt	ΔT (day)
POK	13 225	167 909	40 821	1 min	100
EMA	3 180	301 815	31 664	1 min	80
SEX	12 157	34 060	26 383	1 day	1 000
SPC	113	20 818	2 196	20 s	2.5
SPM	72	6 980	691	20 s	1

summarizes the fundamental characteristics of the network samples used hereafter.

B. Random walker dynamics

The random walk is the simplest dynamic process one may study in a network and conveniently gives a measure of flow within the network [26–28,37]. In a time-ordered network, the concept of flow given by the random walk dynamics remains intuitive, but the implementation of the process is not as simple as in the static network because time introduces a few new constraints. In particular, the dataset is limited within a time interval and, thus, border effects appear in the beginning and ending of the interval; the interactions are sparse and, thus, the walker may get trapped in a vertex for a while; and at last, there is a time scale.

In our model, the walker starts its network tour in a uniformly chosen vertex and only moves forward (jump or hop) whenever there is a contact active, choosing uniformly if more than one contact is available (e.g., a vertex contacts two other vertices at the same time). The walker performs a large number of hops ($H \geq 10^7 \gg C$) in order to explore as much as possible the network and reach the stationary state. We count the number of hops h_{ij} from vertex i to j and normalize by the total number of hops H to obtain an estimation of the flow from i to j ; i.e., $f_{ij} = h_{ij}/H$. Note that for the static network, the stationary solution of the flow in the links can be calculated directly from a modified version of the google matrix (see Ref. [27]).

For simplicity, the walker step corresponds to one time unit (which is set as the temporal resolution). After each hop, time is incremented. In static networks, there is always the possibility to move except in the case of isolated vertices. On the other hand, in temporal networks the walker may have to wait in a vertex until a contact becomes available. Since this waiting time is not counted, the walker dynamics is not affected in case of absence of contacts during some periods of time (e.g., overnight) because the walker remains still, in the vertex, during the period of inactivity. Nevertheless, if the vertex has only one contact, the walker may get trapped in that vertex because after the arrival it cannot move forward. To avoid such situation, at each time step, the walker at vertex i teleports to a uniformly chosen vertex j ($i \neq j$) at a uniformly chosen time t ($T_i \leq t \leq T_f$) with probability τ or hops to an adjacent vertex with probability $1 - \tau$. If the walker reaches the last time step, it teleports with $\tau = 1$. The teleportation

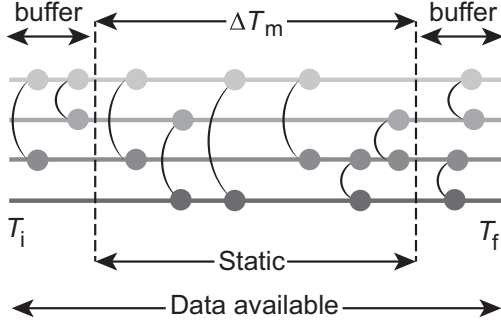


FIG. 2. The sample of the empirical network is available between times T_i and T_f . The random walker moves in, out, and within the buffer regions but only hops within ΔT_m are counted in the flow. The respective static network contains only the links made within ΔT_m .

makes the walk ergodic and is not counted in the flow to improve its estimation [37].

To remove the finite-size effects due to the limited sampling window, we adopt open boundary conditions. This is done by defining two buffer intervals given, respectively, by $[T_i, T_i + T_{\text{buffer}}]$ and $[T_f - T_{\text{buffer}}, T_f]$, i.e., one in the beginning and one in the end of the network (Fig. 2). The walker may move in, out, and within these regions but only hops between $(T_i + T_{\text{buffer}}) < t < (T_f - T_{\text{buffer}})$ are used to estimate the flow. The length of the buffer is $T_{\text{buffer}} = 1.5/\tau$ time steps (with contacts). We use $\tau = 0.1$, which implies on 15 time steps lost on each side. This choice guarantees that a walker reaching the end of the interval ΔT_m continues moving (within the buffer and thus without the interference of the border) until it teleports due to τ (one expects on average $1/\tau$ hops before teleportation, thus, $1.5/\tau$ is chosen to account for the fluctuations). The same reasoning applies on the left border of the network. Such open boundaries imply that the walker does not see the borders of the network. One can intuitively understand the meaning of open boundaries by imagining an infinite system where the walker can hop everywhere but only hops within the interval ΔT_m are counted in the statistics (Fig. 2).

By using periodic boundary conditions, we would not compromise any links (in the buffer), but we introduce undesired biases due to the artificial periodicity. For example, imagine that vertex B is connected to A and C; if vertex A and B interact before B and C, the walker can move from A to B and then from B to C in two hops, but if the links are artificially periodic, the walker may move back from C to B and then from B to A in two hops as well. We expect that such biases are more evident in temporally sparse networks. A fine analysis of the impact of boundary conditions is out of the scope of our study and we therefore simply adopt the open boundaries for the reasons stated above. Finally, the static version of the empirical network is obtained by collecting only the links made within the interval ΔT_m and is thus weighted. The link weights define the probability of hopping to specific neighbors.

C. Flow motifs

By coupling the estimated flow to the links, we convert the original network structure into a flow network that

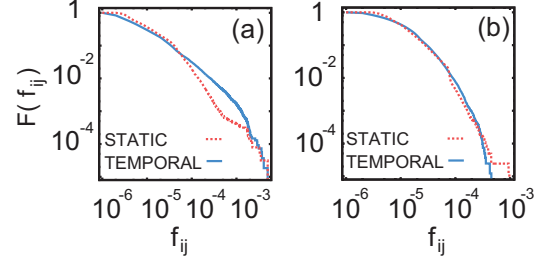


FIG. 3. (Color online) Cumulative distribution of link flow f_{ij} for all links in the network. The curves in (a) correspond to the EMA network and in (b) to the POK network. Significant differences in the distributions appear on links with moderate to high flow.

simultaneously contain information about the topology and about the diffusion dynamics. In order to compare the limits of flow due to the time ordering and in the absence of time ordering, we take a coarse grained approach and split the links into two categories of high and low flow. The splitting point is given by the median of the estimated flow such that an equal number of links are labeled red (dark gray) for high flow and blue (light gray) for low flow. The median is statistically more robust than, e.g., the mean to summarize the central tendency of the data in the case of skewed distributions, since it is not affected by the asymmetry of the data (e.g., very large values bias the mean toward higher values). We show in Fig. 3 that the (cumulative) distributions of flow in the links (f_{ij}) are broad and right skewed in both temporal and static frameworks, which reflects the fact that few links concentrate most of the flow. In our methodology, any other choice of splitting point would significantly unbalance the number of low- and high-flow links and, thus, bias the results. For example, higher (than the median) thresholds would create a large number of low-flow links and emphasize the links with very large flow (the opposite for lower than the median thresholds). This procedure defines high and low flow relatively to all links in the same network and independently of the adopted framework (temporal or static). The median is statistically optimal, for increasing (or decreasing) threshold, the difference between the temporal and static frameworks tends to decrease on average. Since the flow is not necessarily symmetric between the same pair of vertices, the flow network has directed links.

To study the flow network, we analyze its various motifs. Motifs (antimotifs) are subgraphs containing a number of vertices and combinations of connected pairs of vertices, observed more (less) often than one expects by chance [31]. In some contexts, motifs are believed to correspond to single regulatory units designed to perform specific functions and tasks [29–32], while in other contexts, they are used to identify representative sequences of events [7,34]. All together, motifs create a fingerprint of the network [30]. We define a flow motif by measuring the statistical representativity of the subgraphs. This is done by comparing the frequency of the subgraph in the flow network (μ_{ori}) and in randomized versions of the same network. The randomized networks are generated by uniformly selecting two links and switching one of the contacts of each link. Due to the peculiar structure of the flow network (multiplex network), different null models are possible and

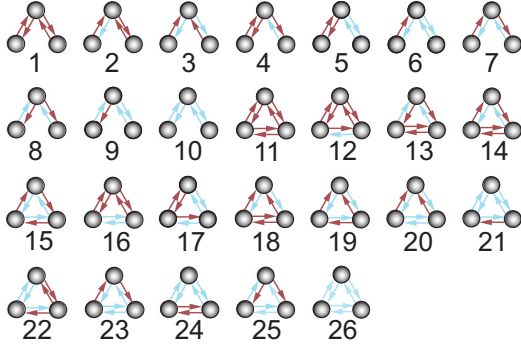


FIG. 4. (Color online) All 26 types of connected triads, i.e., subgraphs with 3 vertices. There are 10 triplets (subgraphs 1 to 10) and 16 triangles (subgraphs 11 to 26). Red (dark gray) links correspond to high flow and blue (light gray) links to low flow. Bipartite networks only have triplets.

the choice of a particular model depends on which constraints one wants to study. In the present work, we fix the in and out degree, the bipartivity (if applicable), and the labels (i.e., the total in and out flow) of the vertex [38]. This particular null model conserves various dynamic and topologic features of the network but breaks down the correlations of the flow between topologically close links. Other choices of null models, e.g., conserving separately the in and out degree for each of the low- and high-flow networks, would also be effective to study the contribution of these structures to generate specific flow subgraphs. The representativity is thus given by the Z score, i.e., $Z = (\mu_{\text{ori}} - \langle \mu_{\text{rand}} \rangle) / \sigma$, where $\langle \mu_{\text{rand}} \rangle$ is the average frequency of the subgraph in 1000 random ensembles and σ is the respective standard deviation [29,31]. We expect that relaxing the constraints during the randomization (constraints of the null model) tends to increase (decrease) the Z score of the most relevant motifs (antimotifs). To illustrate the concept of flow networks, Fig. 4 shows all 26 types of 3-vertex flow subgraphs.

III. RESULTS

In this section, we compare the flow properties of temporal networks in respect to the static network and discuss the characteristics and importance of the most abundant flow motifs.

A. Deviation from the static network

We start our analysis by counting the number of links with the same label [high and low flow, or red (dark gray) and blue (light gray), respectively] in both temporal and static networks. We identify that the proportion of links similar in both frameworks varies according to the category of the network, the lowest being the SEX network with 63.1% and the highest being the SPM network with 86.6%. This means that about 37% and 13% of the links change the category of flow intensity (high to low, or low to high) if the time information is included. This discrepancy is understandable since the SEX network is temporally sparse and vertices are mostly active during relatively short intervals [6], in such a scenario, the time order plays an important role because cycles or repetitions of

contacts are unlikely. In the SPM network, however, vertices correspond to individuals visiting a museum in well-defined time slots [4], which means that a group of vertices remains well connected during the visit and loosely connected with the group of the next (or previous) time slot. In such case, one expects that the static network closely represents the real interaction dynamics. The other cases contain intermediate results consistent with the category of the networks, i.e., 72.2% (POK), 74.4% (SPC), and 79.13% (EMA).

The proportion of links with same label in both frameworks indicates that, to different extents, time order alters the flow on the links on each category of network. One may expect that the matching of colors on both frameworks decreases if more than one link is compared simultaneously, e.g., the label of all links connected to a vertex. Here, however, we want to study how the system-wide characteristics of the flow change from one framework to another. Therefore, we look into the statistics of the subgraphs. To simplify our analysis, we focus on two categories of networks, POK and EMA, corresponding to the largest datasets available (Table I) and provide summary statistics for the other networks. The networks POK and EMA have the same resolution δt and approximately the same ΔT (Table I).

If we count the frequency of the subgraphs in the networks POK and EMA, we identify that triplets (subgraphs 1 to 10; see Fig. 4) are in general significantly more common than triangles irrespective of the framework adopted (Fig. 5). This simply reflects the sparsity of the networks. The four subgraphs that together account for more than 60% of the occurrences are, in order of decreasing frequency, subgraphs 1, 10, 5, and 4.

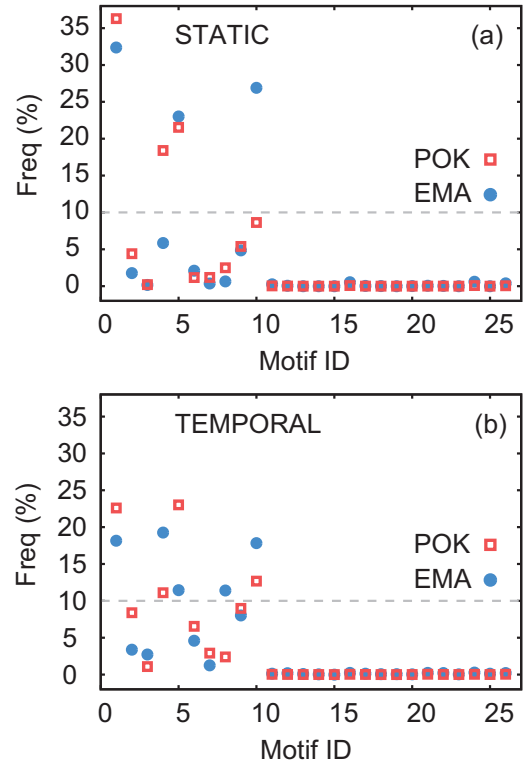


FIG. 5. (Color online) Relative frequency of the subgraphs in the (a) static and (b) temporal networks for the POK and EMA networks.

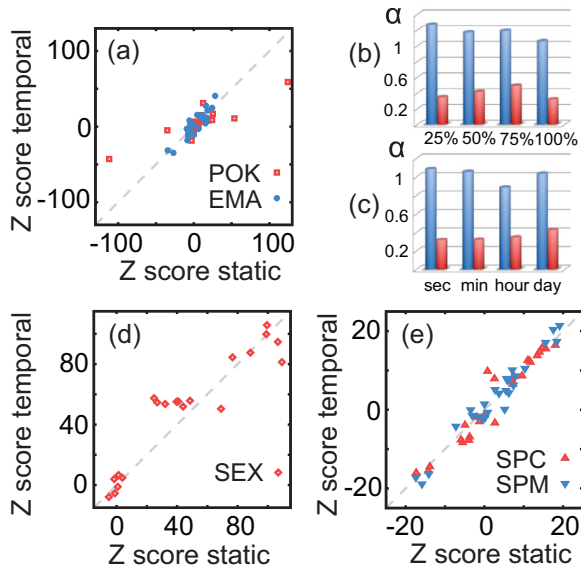


FIG. 6. (Color online) (a,d,e) Relation between the Z scores of the 3-vertex subgraphs for the static and for the temporal version of the same network. The dashed line corresponds to $\alpha = 1$ in $f(x) = \alpha x + \beta$; (b) The slope α of the linear regression between the Z scores in the static and in the temporal frameworks for POK [red (dark gray), shorter bars] and EMA [blue (light gray), higher bars]. Each group corresponds to different fractions of the original network (given in percent); (c) The slope α of the linear regression for various temporal resolutions.

With the exception of subgraph 4, the other subgraphs reflect symmetric relations between the central vertex and the two contacts. Subgraph 4 typically represents the periphery of the network where one of the leaf vertices receives high flow but transmits low flow. In both networks, the frequency of subgraph 1 decreases by about 40% in the temporal version in comparison to the static case. Subgraph 10 also decreases by the same amount in the EMA network but increases its frequency in the case of POK.

Although the frequency of subgraphs illustrates some general properties, the Z score provides a robust measure of significance, since in this case, the frequency of the subgraph is compared to what one expects by chance. If we compare the Z score of the 3-vertex subgraphs detected in the temporal network and in the respective static version, we observe a positive correlation between the two cases (Fig. 6 and Table II). By fitting a straight line [$f(x) = \alpha x + \beta$, where

TABLE II. Statistics of the relation between the temporal and static frameworks. The slope α of the linear regression between the Z scores in the static and in the temporal frameworks. The square of the sample correlation R^2 and the respective p value.

	α	R^2	p value
POK	0.32 ± 0.05	0.640	<0.01
SEX	0.86 ± 0.08	0.867	<0.01
SPC	1.04 ± 0.06	0.930	<0.01
SPM	1.05 ± 0.04	0.958	<0.01
EMA	1.06 ± 0.07	0.897	<0.01

x is the Z score static and $f(x)$ is the Z score temporal], we see that the results for POK network deviate significantly from $\alpha = 1$ (dashed line) while in the EMA network, $\alpha = 1.06$ [Fig. 6(b)]. In the case of POK, highly represented motifs (and antimotifs) in the static network are significantly less represented in the temporal version ($\alpha = 0.32$). On the other hand, for EMA, motifs (and antimotifs) with large absolute Z score in the static network, which is the reference network, are similarly represented in the temporal network. This linear relation, characterized by the slope α , is robust to variations in the sample size of the networks [Fig. 6(b)] as well as to different temporal resolutions [Fig. 6(c)]. For decreasing resolution, the temporal network should converge to the static case. We see that tuning the resolution from seconds to day, α slowly increases for the POK network. The results suggest that the discrepancy between the temporal and static frameworks is indeed characteristic of these systems. The SPC and SPM networks have similar characteristics to EMA, but SEX has the same tendency as the POK network with a relatively higher α [Figs. 6(d) and 6(e) and Table II]. In the SEX case, however, a number of subgraphs (with Z scores static between 20 and 70), corresponding to highly asymmetric combinations (e.g., subgraphs 2, 4, 6, and 9 in Fig. 4), are unrelated in the static and in the temporal frameworks.

The POK and EMA networks represent two distinct classes of temporal networks; in the POK network, the vertex turnover is much higher than in the EMA network. If we measure the fraction of vertices both in the initial and in the final 5% of active links, we identify that only 1.5% of the vertices belong to this group in the POK network (and 1.7% in the SEX), while this number is 47.9% for the EMA network (and 38.9% in the SPC). A high vertex turnover rate means that vertices enter and leave the system frequently, such that contacts (or partnerships) vary in time. This is not surprising since in the dating site (POK), members are active for some period and then leave the community either because they found a date or because they are consistently unsuccessful (similar dynamics happen in the network of sex-buyers and -sellers, where there is a strong tendency to not repeat partners and to not stay long in the system [6]). As a consequence, the collapse of the temporal network into a static structure mixes links between vertices made at different moments. In the context of spreading processes, this mixing homogenizes the flow between close vertices because it removes the time constraint. On the other hand, low turnover is associated to increased network stability in the sense that individuals tend to interact with roughly the same partners in time. This is expected in the email exchange context (EMA), where individuals tend to interact with about the same people within a period of a few months (or during a conference [4]). Due to the regularity of the contacts, the transmission potential of a particular link is not significantly affected. The SPM consists of a particular case because, within the same time slot, individuals interact similarly to EMA or SPC cases (i.e., adjacent vertices remain the same and interactions are persistent) and only few links connect visitors of different time slots (in the static framework, this effect is reflected into a strong community structure [4]). In other words, the static framework is adequate to estimate the limits of flow in the case of regularity and stability of contacts, typically with low vertex turnover; on the other hand, the static

framework poorly approximates the contact patterns in cases of high vertex turnover and temporal variation of partnerships (temporal mixing).

B. Individual flow motifs

In this part, we compare the representativity of the flow subgraphs for two classes of networks represented by POK and EMA networks. In the context of static directed networks, it is known that the representativity of each subgraph varies for different categories of networks. For example, the feed-forward motif is commonly observed in transcription networks but poorly represented in the networks of adjacent words in written texts [30]. In order to compare networks of different sizes, it is necessary to normalize the vector of Z scores to length 1 by using $NZ_i = Z_i / (\sum_j Z_j^2)^{0.5}$. Motifs in large networks display higher Z scores than in small networks [29]; the normalization corrects that by measuring the relative significance of subgraphs.

Figure 7(a) (bottom) shows the normalized Z scores for the static POK and EMA networks. Since there is no threshold to define a motif or antimotif, the dashed lines at ± 0.2 are simply reference lines. In the POK network, subgraphs 26 and 11 are, respectively, the most represented motifs, and subgraph 10 is the most antimotif. On the other hand, in the EMA network, subgraphs 24, 4, and 26 are the most represented motifs, and subgraphs 5 and 10 are antimotifs. This reflects both the clustering structure (since motifs 11, 24, and 26 are triangles) and the difference in the contact patterns in both networks. In fact, taking the difference of the normalized Z scores between the POK and EMA networks [Fig. 7(a) (top)], we identify that motifs 24 and 26, and antimotif 5 are significantly different in both networks.

In the case of temporal networks, the representativity of subgraphs is more distinguishable between both networks [Fig. 7(b)]. Most subgraphs are poorly represented in the POK network (NZ score close to zero), but in the case of EMA, the triplets are typically well represented (NZ score close to 0.2) while triangles are under-represented (NZ score close to -0.1). In the temporal framework, for both networks, subgraphs 24 and 26 are motifs and subgraphs 5 and 10 are antimotifs. The largest differences are observed, respectively, for subgraphs 26 and 5.

These observations indicate that triangle structures (i.e., the fact that two friends are also friends themselves) do not necessarily imply on high-flow between the vertices in the context of temporal networks. In fact, in a triangle, it is more likely to find low-flow between all three pairs of vertices (subgraph 26) or high-flow only between one pair of vertices (subgraph 24) than high-flow between the three pairs of vertices (subgraph 11). In the static network, on the other hand, the high-flow triangle (subgraph 11) is relatively more represented in both networks, but still, not as much as the low-flow triangles (subgraphs 26 and 24). In the case of triplets, the symmetric subgraphs are significantly antirepresented (subgraphs 5 and 10). The symmetric high-flow subgraphs are not significantly represented (subgraph 1). This antirepresentativity of symmetric triplets is also observed in directed networks of online social networks and web pages [30].

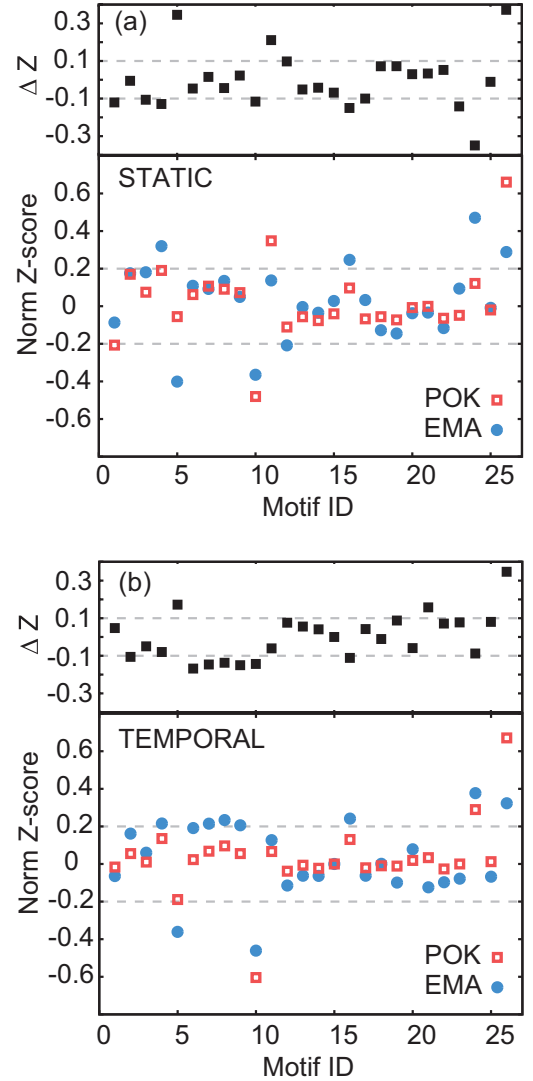


FIG. 7. (Color online) (a) Normalized Z scores of the subgraphs for the static networks; (b) Normalized Z scores of the subgraphs for the temporal networks. On top of (a) and (b), the difference, $\Delta Z = NZ_{\text{POK}} - NZ_{\text{EMA}}$, is presented; therefore, $\Delta Z > 0$ means that a particular subgraph is more abundant in the POK than in the EMA network.

C. Higher-order motifs

Using the flow network defined in Sec. II C, it is possible to extract information about subgraphs with any number of vertices. Higher-order motifs provide detailed information of the flow within higher neighborhoods [39], but the number of possible subgraphs grows significantly for each added vertex. Furthermore, the computational cost makes the detection of higher-order motifs impractical for networks of moderate size and link density such that other methods are more adequate to study the flow properties of large groups of vertices [26,27]. We therefore restrict our analysis to subgraphs with 4 vertices.

We measure the Z scores of the 4-vertex subgraphs by using 100 realizations of the randomized version. Figure 8 shows the frequency of subgraphs with a certain value of Z score for both temporal and static frameworks. In the static framework, subgraphs typically have Z scores concentrated

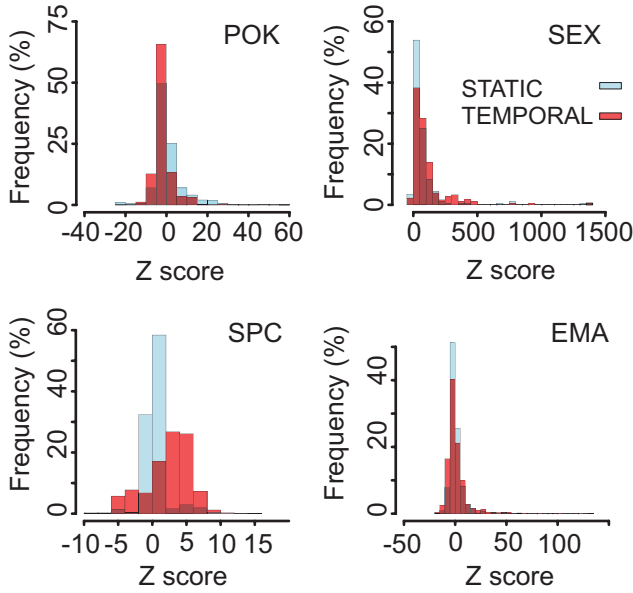


FIG. 8. (Color online) Frequency of subgraphs with a certain Z score, for the static and temporal versions of the empirical networks. The results for SPM are qualitatively similar to EMA and thus omitted.

around zero, while a larger number of subgraphs have higher Z scores in the temporal framework. These results indicate that the time order significantly constrains the flow such that more types of subgraphs become representative whenever the temporal framework is adopted; these subgraphs provide finer details about the relevance of the routes of spreading. In the SEX and EMA networks, there are highly represented motifs in both temporal and static frameworks, while in the POK and SPC cases, motifs are more distinguishable in the temporal and static frameworks, respectively.

Rather than looking to several individual subgraphs or motifs, we turn our attention to a special structure, the fully connected 4-vertex subgraph (4-vertex clique). By measuring both the Z scores and the relative frequency, we identify that the low-flow (and not the high-flow) subgraphs are the most significant in all studied datasets (Fig. 9). The relative frequency of low-flow subgraphs is typically one order of magnitude higher than the high-flow cases on both temporal and static frameworks. With few exceptions, the Z scores are also significantly higher for the low-flow subgraphs. In the SEX network, which is bipartite, high-flow subgraphs are more frequent, but the Z scores are significantly higher for the low-flow case as well. In accordance with the results for 3-vertex subgraphs (more specifically, triangles), these symmetric low- and high-flow subgraphs are also more represented in the static versions of the empirical networks, reinforcing the conclusions that time order reduces the availability of certain paths and consequently increases the variability of flow between adjacent vertices.

The results suggest that single structures alone (i.e., cliques of 3 and 4 vertices) are insufficient to properly quantify the flow, or the ability to propagate information or infections, between adjacent vertices. In fact, in static networks, the spreading processes are constrained by the connectivity of the

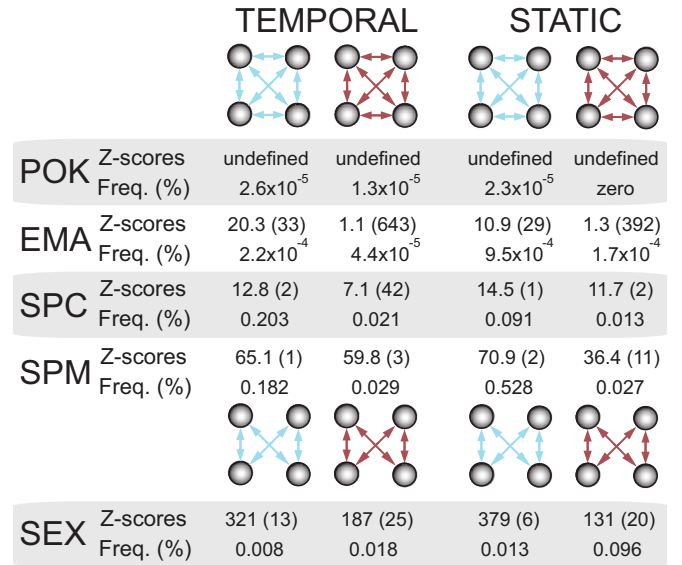


FIG. 9. (Color online) The Z score and the frequency of fully connected 4-vertex motifs where all links are either low flow or high flow, for the temporal and static frameworks. The number in parentheses indicates the rank of the subgraph in terms of its representativity (Z score) in respect to the other subgraphs of the same network and framework. The SEX case is the only purely bipartite network, which means that connections between different classes of vertices are nonexistent.

vertices linked to these single structures; i.e., the flow within a structure depends on the context in which the structure is placed. In the temporal framework, on the other hand, not only the topological context but also the time sequence of the links, within and outside the structure, play a major role to define possible routes of propagation and thus regulate the flow between adjacent vertices.

IV. CONCLUSIONS

The increasing availability of high-resolution temporal network data has encouraged researchers to study the effects of temporal constraints in spreading processes on networks. While several studies suggest that temporal constraints regulate the speed of the spreading processes, it is unclear how these constraints affect the spreading between vertices. To address this problem, we have proposed a methodology to quantify the limits of flow between adjacent vertices by using a random walker dynamics and motif analysis. The normalized frequency of hops between two vertices gives an estimate of the overall flow between these vertices. By labeling the links into high- and low-flow classes, we convert the original undirected network into a more complex network (multiplex network with two types of links and directions) containing information about the dynamics. We introduce the concept of flow motifs as particular subgraphs observed significantly more often than one expects by chance.

We compare the abundance of 3-vertex subgraphs in the temporal networks in respect to the static case and identify that in the case of emails exchange and face-to-face interactions, the regularity of contacts and persistence of vertices result on insignificant changes in the limits of flow whenever the

static or temporal framework is adopted. On the other hand, in the case of communication between members of a dating site and of a sexual contact network, the flow between adjacent vertices changes significantly in the temporal framework such that the static approximation poorly represents the structure of contacts. This happens because in this category of networks, the vertex turnover is high and contacts vary between active members such that links active at different times in the temporal network are mixed in the aggregated static structure. We have also observed that triangles of connected vertices do not necessarily correspond to high flow between the vertices forming the triangular structure. In fact, for the temporal network, the most representative triangle subgraph corresponds to symmetric low flow between the three vertices and the second most representative triangle to high flow between a pair of vertices and low flow in the other two links. Similar results are obtained when using 4-vertex subgraphs. In that case, on both temporal and static frameworks, subgraphs with all low-flow links are more often and more representative (higher Z score) than subgraphs with all high-flow links. These results suggest that low-order cliques (i.e., fully connected structures with 3 or 4 vertices) do not necessarily imply high flow between the associated vertices. The flow, and possibly other spreading processes, within those structures depends not only on the topological context in which the structure is placed

but also in the time order that links are established in and around the structure.

The proposed methodology is general and can be applied to study other processes, as for example, epidemics, opinion dynamics, or traffic. It provides a summarized description of the potential of communication, influence, or transmission between small groups of vertices. This description can be used to identify vertices and links relevant to regulate spreading processes not only in the context of temporal networks but also in static networks.

ACKNOWLEDGMENTS

The authors thank Martin Rosvall for discussions about random walker models, Taro Takaguchi for pointing out missing references, Petter Holme and the SocioPattern consortium for sharing, respectively, the dating community and the face-to-face contacts datasets. L.E.C.R. acknowledges support as a FSR fellow of the Académie Universitaire Louvain, cofunded by the Marie Curie Actions of the European Commission. Computational resources have been provided by the supercomputing facilities of the Université Catholique de Louvain (CISM/UCL) and the Consortium des Équipements de Calcul Intensif en Fédération Wallonie Bruxelles (CECI) funded by FRS-FNRS.

-
- [1] M. Newman, *Networks: An Introduction* (Oxford University Press, New York, 2010).
 - [2] L. da F. Costa, O. N. Oliveira Jr., G. Travieso, F. A. Rodrigues, P. R. V. Boas, L. Antiqueira, M. P. Viana, and L. E. C. Rocha, *Adv. Phys.* **60**, 329 (2011).
 - [3] P. Holme, *Europhys. Lett.* **64**, 427 (2003).
 - [4] L. Isella, J. Stehlé, A. Barrat, C. Cattuto, J.-F. Pinton, and W. V. den Broeck, *J. Theor. Biol.* **271**, 166 (2011).
 - [5] A.-L. Barabási, *Nature (London)* **435**, 207 (2005).
 - [6] L. E. C. Rocha, F. Liljeros, and P. Holme, *Proc. Natl. Acad. Sci. USA* **107**, 5706 (2010).
 - [7] L. Kovanen, M. Karsai, K. Kaski, J. Kertész, and J. Saramäki, *J. Stat. Mech.* (2011) P11005.
 - [8] H. H. K. Lentz, T. Selhorst, and I. M. Sokolov, *Phys. Rev. Lett.* **110**, 118701 (2013).
 - [9] D. Rybski, S. V. Buldyrev, S. Havlin, F. Liljeros, and H. A. Makse, *Proc. Natl. Acad. Sci. USA* **106**, 12640 (2009).
 - [10] J. Tang, S. Scellato, M. Musolesi, C. Mascolo, and V. Latora, *Phys. Rev. E* **81**, 055101 (2010).
 - [11] R. K. Pan and J. Saramäki, *Phys. Rev. E* **84**, 016105 (2011).
 - [12] P. Grindrod, M. C. Parsons, D. J. Higham, and E. Estrada, *Phys. Rev. E* **83**, 046120 (2011).
 - [13] T. Takaguchi, N. Sato, K. Yano, and N. Masuda, *New J. Phys.* **14**, 093003 (2012).
 - [14] M. Korschake, H. H. K. Lentz, F. J. Conraths, P. Hövel, and T. Selhorst, *PLoS ONE* **8**, e55223 (2013).
 - [15] L. E. C. Rocha and V. D. Blondel, *PLOS Comput. Biol.* **9**, e1002974 (2013).
 - [16] T. Takaguchi, N. Masuda, and P. Holme, *Bursty Communication Patterns Facilitate Spreading in a Threshold-Based Epidemic Dynamics* (2012), pp. 1–15. [arXiv:1206.2097](https://arxiv.org/abs/1206.2097).
 - [17] L. E. C. Rocha, F. Liljeros, and P. Holme, *PLOS Comput. Biol.* **7**, e1001109 (2011).
 - [18] J. Stehlé, N. Voirin, A. Barrat, C. Cattuto, V. Colizza, L. Isella, C. Régis, J.-F. Pinton, N. Khanafer, W. V. den Broeck, and P. Vanhems, *BMC Med.* **9**, 87 (2011).
 - [19] G. Miritello, E. Moro, and R. Lara, *Phys. Rev. E* **83**, 045102(R) (2011).
 - [20] M. Karsai, M. Kivelä, R. K. Pan, K. Kaski, J. Kertész, A.-L. Barabási, and J. Saramäki, *Phys. Rev. E* **83**, 025102(R) (2011).
 - [21] L. Tabourier, A. Stoica, and F. Peruani, *How to Detect Causality Effects on Large Dynamical Networks: A Case Study, Communication Systems and Networks (COMSNETS)*, pp. 1–7 (2012).
 - [22] M. Starnini, A. Baronchelli, A. Barrat, and R. Pastor-Satorras, *Phys. Rev. E* **85**, 056115 (2012).
 - [23] J. O. Haerter, B. Jamtveit, and J. Mathiesen, *Phys. Rev. Lett.* **109**, 168701 (2012).
 - [24] N. Perra, A. Baronchelli, D. Mocanu, B. Gonçalves, R. Pastor-Satorras, and A. Vespignani, *Phys. Rev. Lett.* **109**, 238701 (2012).
 - [25] T. Takaguchi and N. Masuda, *Phys. Rev. E* **84**, 036115 (2011).
 - [26] M. Rosvall and C. T. Bergstrom, *Proc. Natl. Acad. Sci. USA* **105**, 1118 (2008).
 - [27] Y. Kim, S.-W. Son, and H. Jeong, *Phys. Rev. E* **81**, 016103 (2010).
 - [28] R. Lambiotte, R. Sinatra, J.-C. Delvenne, T. S. Evans, M. Barahona, and V. Latora, *Phys. Rev. E* **84**, 017102 (2011).

- [29] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, *Science* **298**, 824 (2002).
- [30] R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, and U. Alon, *Science* **303**, 1538 (2004).
- [31] U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits* (Chapman and Hall/CRC, London, 2006).
- [32] J. F. Knabe, C. L. Nehaniv, and M. J. Schilstra, *BioSystems* **94**, 68 (2008).
- [33] P. R. V. Boas, F. A. Rodrigues, G. Travieso, and L. da F. Costa, *Phys. Rev. E* **77**, 026106 (2008).
- [34] R. Sinatra, D. Condorelli, and V. Latora, *Phys. Rev. Lett.* **105**, 178702 (2010).
- [35] P. Holme, C. R. Edling, and F. Liljeros, *Soc. Net.* **26**, 155 (2004).
- [36] J.-P. Eckmann, E. Moses, and D. Sergi, *Proc. Natl. Acad. Sci. USA* **101**, 14333 (2004).
- [37] R. Lambiotte and M. Rosvall, *Phys. Rev. E* **85**, 056107 (2012).
- [38] S. Wernicke and F. Rasche, *Bioinformatics* **22**, 1152 (2006).
- [39] L. da F. Costa and L. E. C. da Rocha, *Eur. Phys. J. B* **50**, 237 (2006).