Components in time-varying graphs

Vincenzo Nicosia[1,2], John Tang[3], Mirco Musolesi[4], Giovanni Russo[2,5], Cecilia Mascolo[3], Vito Latora[1,2]
[1] *Dipartimento di Fisica e Astronomia, Università di Catania and INFN, Via S. Sofia, 64, 95123 Catania, Italy*
[2] *Laboratorio sui Sistemi Complessi, Scuola Superiore di Catania, Via Valdisavoia 9, 95123 Catania, Italy*
[3] *Computer Laboratory, University of Cambridge, 15 JJ Thomson Av., Cambridge CB3 0FD, UK*
[4] *School of Computer Science, University of St Andrews, Fife KY16 9SX, UK and*
[5] *Dipartimento di Matematica ed Informatica, Università di Catania, Via S. Sofia, 64, 95123 Catania, Italy*
(Dated: June 21, 2011)

Real complex systems are inherently time-varying. Thanks to new communication systems and novel technologies, it is today possible to produce and analyze social and biological networks with detailed information on the time of occurrence and duration of each link. However, standard graph metrics introduced so far in complex network theory are mainly suited for static graphs, i.e., graphs in which the links do not change over time, or graphs built from time-varying systems by aggregating all the links as if they were concurrent in time. In this paper, we extend the notion of connectedness, and the definitions of node and graph components, to the case of *time-varying networks*, which are represented as time-ordered sequences of graphs defined over a fixed set of nodes. We show that the problem of finding strongly connected components in a time-varying graph can be mapped into the problem of discovering the maximal-cliques in an opportunely constructed static graph, which we name the *affine graph*. It is therefore a NP-complete problem. As a practical example, we have performed a temporal component analysis of time-varying graphs constructed from Reality Mining, a dataset of human contacts recorded at MIT over a period of six months. The results show that taking time into account in the definition of graph components allows to capture important features of the system. In particular, we observe a large variability in the size of node temporal in- and out-components. This is due to intrinsic differences in the activity patterns of the individuals, which cannot be detected by static graph analysis.

PACS numbers: 89.75.Hc,89.75.-k,89.75.Fb

I. INTRODUCTION

Complex network theory has proved to be a versatile framework to represent and analyze biological, social and man-made complex systems [1, 2]. Typically, a complex system is inherently dynamic. Social interactions and human activities are intermittent [3–6], the neighborhood of individuals moving over a geographic space evolves over time [7, 8], links appear and disappear in the World-Wide-Web [9], in patterns of interactions among genes from microarray experiments [10, 11] and in functional brain networks [12, 13]. In all these networks, time plays a central role: links exist only for certain time periods, and are often recurrent. Despite this fact, most of the classic studies in complex networks theory are based on the analysis of the topological properties of *static graphs*. These are graphs in which the links do not change over time, or graphs built from time-varying systems as the result of the aggregation of all interactions, as if these were all concurrent in time. The evolution of linking patterns over time, when considered, has been usually studied by creating a series of graphs, each graph containing all the links appeared in a certain time interval. Then, each standard graph measure is evaluated for the static graph obtained at each time window, and plotted as a function of time [14, 15]. Today, thanks to modern technologies, for the first time we have the opportunity to study large social and biological networks with precise temporal information on the appearance, duration and frequency of links among a set of nodes, and many other similar databases will be produced, at an ever increasing rate, in the near future. These datasets demand for new network measures and models that can take account of the richness introduced by detailed temporal information. Some recent works have analyzed large interconnected systems with fluctuating interactions [16–18], and some graph measures have been already extended to the case of connection patterns which evolve over time [19]. In previous works [20–22] we have shown that static analysis of aggregated graphs is not able to capture the real dynamic behavior and time correlations of complex networks which evolve over time. Moreover, if temporal ordering of links is ignored, static analysis overestimates the number of available links at each time and, therefore, underestimates actual lengths of walks and paths. In Ref. [23] we have formalized the concept of *time-varying graph* and we have introduced a measure of path lengths for time-varying graphs which takes into account the actual time ordering, duration and correlations between links which appear at different times. The concept of walk has been generalized to the case of time-varying graphs in Refs. [24, 25], while alternative definitions of path lengths have been proposed in Refs. [26, 27]. Recent works have also studied the onset of synchronization in populations of agents interacting through time-evolving topologies [28].

In this paper, we focus our attention on two important concepts in graph theory, namely those of *connectedness* and *connected components* of a graph. These concepts have been used in complex networks to study the reachability of pairs of nodes [29], and to characterize the resilience of networks to attacks [30]. The paper is organized as follows. In Section II, we briefly review the concepts of connectedness and components in static graphs, while in Section III we extend them to the case of time-varying graphs. In particular, we define the in- and out-component of a given node, and we give the definition of weakly and strongly connected components. In Section IV we show that the problem of finding components in a time-varying graph can be mapped into the maximal-clique problem for an opportunely constructed graph, which we call *affine graph*. An affine graph is a static graph which incorporates all the information on the temporal reachability of couples of nodes, and is a useful tool to analyze the components of the corresponding time-varying graph. Thanks to this mapping, we prove also that finding strongly connected components in time-varying graphs is a NP-complete problem. Finally, in Section V we analyze time-varying graphs constructed from a dataset of human interactions, and we show that the analysis of temporal strongly connected components reveals interesting variations in the pattern of contacts, which cannot be observed by using an aggregate static graph.

II. COMPONENTS IN STATIC GRAPHS

Let us consider a graph G with N nodes and K links. From now on we will refer to it as to a *static graph*. We will consider the case of undirected and directed static graphs separately. An *undirected static graph* G can be represented by a symmetric adjacency matrix, a $N \times N$ matrix A whose each entry a_{ij} is equal to one if and only if there is a link between i and j, and is equal to zero otherwise. In order to define graph components, we need to introduce the concept of connectedness, first for pairs of nodes, and then for the whole graph. Two nodes i and j of an undirected graph G are said to be *connected* if there exists a path between i and j. G is said to be *connected* if all pairs of nodes in G are connected, otherwise it is said to be unconnected or disconnected. A *connected component* of G associated to node i is the maximal connected induced subgraph containing i, i.e., it is the subgraph induced by all nodes connected to node i. If an undirected graph is not connected, it is always possible to find a partition of the graph into a set of disjoint connected components. It is straightforward to prove that this partition is unique.

A *directed static graph* G is described in general by a non-symmetric adjacency matrix, a $N \times N$ matrix A whose each entry a_{ij} is equal to one if and only if there is a directed link from i to j, and is equal to zero otherwise. Defining connectedness for pairs of nodes in a directed graph is more complex than in an undirected graph, because a directed path may exist through the network from vertex i to vertex j, but this does not guarantee that any path from j to i does actually exist. Consequently, we have two different definitions of connectedness between two nodes, namely *weak* and *strong* connectedness. In particular, we can define the *weakly* and the *strongly connected components* of a directed graph as follows [31]. Two nodes i and j of a directed graph G are said *strongly connected* if there exists a path from i to j and a path from j to i. A directed graph G is said *strongly connected* if all pairs of nodes (i, j) are strongly connected. A *strongly connected component* of G associated to node i is the maximal strongly connected induced subgraph containing node i, i.e., it is the subgraph which is induced by all nodes which are strongly connected to node i. A *weakly connected component* of G is a component of its *underlying undirected graph* G^u, which is obtained by removing all directions in the edges of G. Two nodes i and j of G are *weakly connected* if they are connected in G^u, and a directed graph G is said to be *weakly connected* if the underlying undirected graph G^u is connected. Hence, the components of a directed graph can be of two different types, namely weakly and strongly connected. It is useful to review also the definitions of components *associated to a node* of a directed graph. We have four different definitions:

1. The *out-component of node* i, denoted as $\mathrm{OUT}(i)$, is the set of vertices j such that there exists a directed path from i to $j, \forall j$.

2. The *in-component of a node* i, denoted as $\mathrm{IN}(i)$, is the set of vertices j such that there exists a directed path from j to $i, \forall j$.

3. The *weakly connected component of a node* i, denoted as $\mathrm{WCC}(i)$, is the set of vertices j such that there exists a path from i to $j, \forall j$ in the underlying undirected graph G^u.

4. The *strongly connected component of a node* i, denoted as $\mathrm{SCC}(i)$, is the set of vertices j such that there exists a directed path from i to j and also a directed path from j to $i, \forall j$.

We have already used the last two concepts for the definitions of weakly and strongly connected components of a directed graph given above. In fact, the property of weakly and strongly connectedness between two nodes is reflexive, symmetric and transitive, i.e., in mathematical terms, it is an *equivalence relation*. Therefore, it is possible to define weakly and strongly connected components of a graph by means of the weakly and strongly connected components associated to the nodes of the graph: a strongly (weakly) connected component of node is also a strongly (weakly) connected component of the whole graph.

Conversely, the definitions of out-component and in-component of a node are not based on *equivalence relations*. In fact, the symmetry property does not yield: $i \in \mathrm{OUT}(j)$ does not imply $j \in \mathrm{OUT}(i)$. This means that out- and in-components can be associated only to nodes, and cannot be directly extended to the entire graph. In

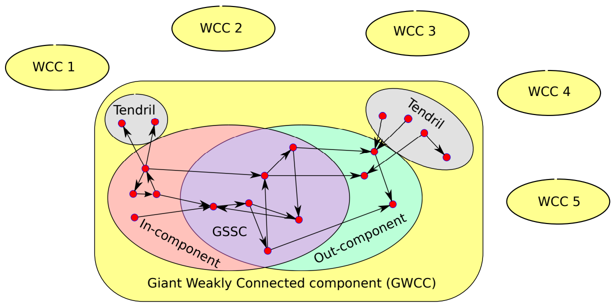

FIG. 1. A directed graph can be partitioned into a set of disjoint weakly connected components (in yellow). Furthermore, each of these components has a rich internal structure, as shown for the GWCC.

practice, we cannot partition a graph into a disjoint set of in- or out-components, while it is possible to identify a partition of a static graph into a disjoint set of weakly or strongly connected components. However, the in- and out-components of the nodes of a graph can be used to define the strongly connected components of the graph. From the above definitions, we observe that $i \in \mathrm{OUT}(j)$ if and only if $j \in \mathrm{IN}(i)$. Furthermore, we notice that i and j are strongly connected if and only if $j \in \mathrm{OUT}(i)$ and, at the same time, $i \in \mathrm{OUT}(j)$. Or equivalently, if and only if $j \in \mathrm{OUT}(i)$ and $j \in \mathrm{IN}(i)$. Therefore, the strongly connected component of node i is the intersection of $\mathrm{IN}(i)$ and $\mathrm{OUT}(i)$.

We are now ready to describe in detail the rich interplay between the various concepts of connectedness in a directed static graph. In the most general case, as shown in Fig. 1, a directed graph can be decomposed in a set of disjoint weakly connected components. In a large graph, one component will be larger than all the others. This component is usually called the *giant weakly connected component* GWCC of the graph. If we treat each link in the GWCC as bidirectional, then every node in the GWCC is reachable from every other node in the GWCC. As shown in Fig. 1, the GWCC contains the *giant strongly connected component* GSCC, consisting of all sites reachable from each other following directed links. All the sites reachable from the GSCC are referred to as the *giant OUT component*, and the sites from which the GSCC is reachable are referred to as the *giant IN component*. The GSCC is the intersection of the giant IN- and OUT-components. All sites in the GWCC, but not in the IN- and OUT-components, are referred to as “tendrils”.

III. COMPONENTS IN TIME-VARYING GRAPHS

In this paper, we consider *time-varying graphs*, which are graphs characterized by fluctuating links among a fixed set of nodes. A time-varying network can be described as an ordered sequence of graphs, i.e., an ordered set $\{G_1, G_2, \ldots, G_M\}$ of M graphs defined over N nodes, where each graph G_m in the sequence represents the state of the network, i.e., the configuration of links, at time t_m, where $m = 1, \ldots, M$. In this notation, the quantity $t_M - t_1$ is the temporal length of the observation period. The graphs in the sequence can be uniformly distributed over time, i.e., $t_{m+1} = t_m + \Delta t,\ \forall m = 1, \ldots, M-1$ [23], or in general they can correspond to any ordered sequence of times such that $t_1 < t_2 < \ldots < t_M$ [24]. In compact notation, we denote the graph sequence as $\mathcal{G} \equiv \mathcal{G}_{[t_1,t_M]}$. Each graph in the sequence can be either undirected or directed. Consequently, the time-varying graph $\mathcal{G}$ can be described by means of a time-dependent adjacency matrix $A(t_m),\ \ m = 1, \ldots, M$, where $a_{ij}(t_m)$ are the entries of the adjacency matrix of the graph at time t_m. This matrix is in general non-symmetric. If we discard the time-ordering of the links of a time-varying graph $\mathcal{G}$, and consider all links as concurrent in time, we obtain its corresponding *aggregate static graph*. In Fig. 2 we report a simple time-varying graph $\mathcal{G}_{[t_1,t_4]}$ with $N = 5$ nodes and $M = 4$ graphs (panel a) and the corresponding aggregate static graph (panel b). It is worth noticing that the aggregate graph discards most of the richness of the original time-varying graph. For instance, three paths exist between node 1 and 5 in the static aggregate graph, namely 1-4-5, 1-2-5 and 1-2-4-5, while in the time-varying graph there is no temporal path from 1 to 5. The problem of defining connectedness and components in time-varying graphs looks more similar to the case of directed static graphs than to the case of undirected static graphs. In fact, even if each graph G_m, $m = 1, \ldots, M$ in the sequence is undirected, the temporal ordering of the graphs naturally introduces a directionality. Imagine a time-varying graph composed by a sequence of 24 undirected graphs $\{G_1, G_2, \ldots G_{24}\}$, where each graph reports the contacts observed among a set of N individuals during one hour of the day. Let us suppose that there exists no link between nodes i and j in any of the 24 graphs. Moreover, imagine that i meets a third node l at 9:30am (i.e., $a_{il}(t_{10}) = 1$) while l meets j at 3:15pm (i.e., $a_{lj}(t_{16}) = 1$). This means that in the time-varying graph there exists a path i-l-j from i to j , and in practice i can use this path to send a message to j through l. However, this does not imply that j can also send a message to i. Due to the temporal ordering of links, the path i-l-j is different from the path j-l-i, even if all the 24 graphs in the sequence are undirected. The existence of the first path does not imply the existence of the second one, and vice-versa. For instance, in the time-varying graph $\mathcal{G}_{[t_1,t_4]}$ reported in Fig. 2, there exists a path which connects node 5 to node 1 (i.e., the link a_{52} at time t_1 and the link a_{21} at time t_3) but there is no path which connects node 1 to node 5. An immediate consequence of this fact is that node 5 can send a message to node 1 at time t_1, while node 1 cannot send a message to node 5.

In order to define node connectedness for a time-varying graph, we first need to introduce a mathematical definition of *reachability* for an ordered pair of nodes i and j. We say that i can reach j if i can send a mes-

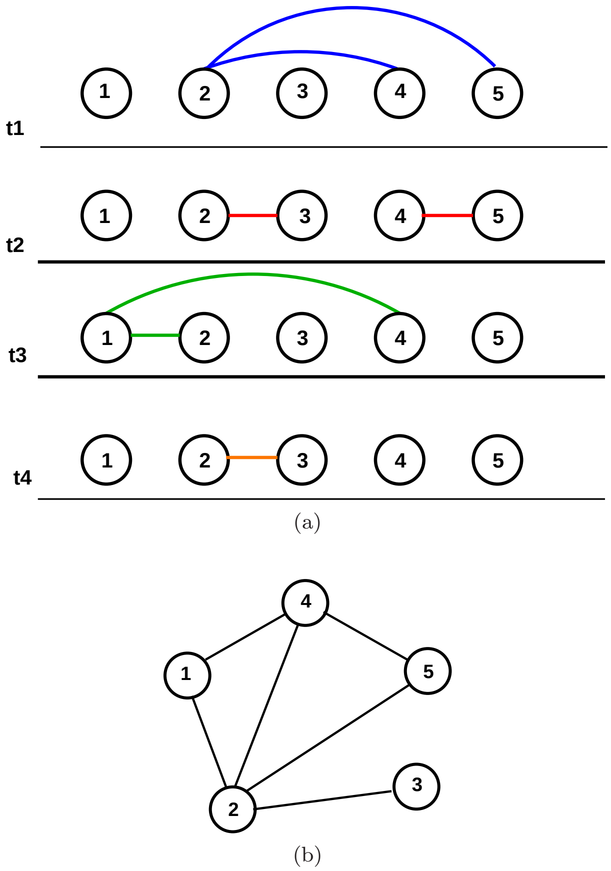

FIG. 2. A time-varying graph $\mathcal{G}$ consisting of a sequence of $M = 4$ graphs with $N = 5$ nodes (panel a) and its corresponding aggregate static graph (panel b). The static representation of graphs discards time ordering of links and time correlations of paths. In the aggregate graph node 1 and node 2 are neighbors, but in the original time-varying graph they are directly connected only in one of the four graphs of the sequence, namely in G_{t_3}. Moreover, in the aggregate graph a path exists from node 1 to node 5 and vice-versa, while in the time-varying graph there exists a temporal path from 5 to 1 but there are no temporal paths from 1 to 5.

sage to j directly or through a time-ordered sequence of contacts. In mathematical terms this implies the existence of a walk connecting i to j. In [24] the concept of walk has been generalized to the case of time-varying graphs. In a time-varying graph, a *walk*, also called *temporal walk*, from node i to node j is defined as a sequence of L edges $[(n_{r_0}, n_{r_1}), (n_{r_1}, n_{r_2}), \ldots, (n_{r_{L-1}}, n_{r_L})]$, with $n_{r_0} \equiv i$, $n_{r_L} \equiv j$, and an increasing sequence of times $t_{r_1} < t_{r_2} < \ldots < t_{r_L}$ such that $a_{n_{r_{l-1}}, n_{r_l}}(r_l) \neq 0 \quad l = 1, \ldots, L$. A *path*, also called *temporal path*, of a time-varying graph is a walk for which each node is not visited more than once. For instance, in the time-varying graph of Fig. 2, the sequence of edges $[(5, 2), (2, 1)]$ together with the sequence of times t_1, t_3 is a temporal path of the graph. This path starts at node 5 at time t_1 and arrives at node 1 at time t_3. The concept of temporal path from a node to another, together with a measure of temporal node distance, was first introduced in [23]. Alternative measures of temporal distance have also been proposed in [26, 27]. Given the definitions of temporal walk and path, we can introduce the concepts of temporal connectedness (in a weak and in a strong sense) for a pair of nodes.

A node i of a time-varying graph $\mathcal{G}_{[t_1,t_M]}$ is *temporally connected* to a node j if there exists in $[t_1, t_M]$ a temporal path going from i to j. This relation is not symmetric: if node i is temporally connected to node j, in general node j can be either temporally connected or disconnected to i. In the graph $\mathcal{G}_{[t_1,t_4]}$ of Fig. 2, node 5 is temporally connected to 1 but node 1 is not connected to node 5. For this reason, we introduce the definition of *strong connectedness*, which enforces symmetry:

Definition 1 (Strong connectedness) *Two nodes i and j of a time-varying graph are* strongly connected *if i is temporally connected to j and also j is temporally connected to i.*

Strong connectedness is a reflexive and symmetric relation, so that if i is strongly connected to j, then j is strongly connected to i. However this definition of strong connectedness lacks transitivity, and therefore it is not an equivalence relation. In fact, if i and j are strongly connected and j and l are strongly connected, nothing can be said, in general, about the connectedness of i and l. In the example shown in Fig. 2, node 5 and 2 are strongly connected and also 2 and 1 are strongly connected, but nodes 5 and 1 are not strongly connected, since there exists no temporal path which connects node 1 to node 5. It is also possible to introduce the concept of weak connectedness for a pair of nodes. Similarly to the case of static directed graphs, given a time-varying graph $\mathcal{G}$, we construct the underlying undirected time-varying graph $\mathcal{G}^u$, which is obtained from $\mathcal{G}$ by discarding the directionality of the links of all the graphs $\{G_m\}$, while retaining their time ordering.

Definition 2 (Weak connectedness) *Two nodes i and j of a time-varying graph are* weakly connected *if i is temporally connected to j and also j is temporally connected to i in the underlying undirected time-varying graph $\mathcal{G}^u$.*

Also weak connectedness is a reflexive and symmetric relation, but it is not transitive. This definition of weak connectedness is quite similar, but not identical, to that given for directed static graphs. In fact, two nodes in $\mathcal{G}$ can be weakly connected even if there is no temporal directed path which connects them, but the temporal ordering of links breaks the transitivity so that if i and j are weakly connected and j and l are weakly connected, then nothing can be said about the weak connectedness of i and l. All these subtleties are due to the fact that time-varying graphs have a much richer structure compared to static graphs, so that the existence of a temporal path between two nodes crucially depends on the time ordering of links, and does not guarantee the existence of the backward path. Notice that the definitions of strong and weak connectedness given above for time-varying graph are consistent with those given for static graphs, so that if two nodes are strongly (weakly) connected in a time-varying graph, then they are also strongly (weakly) connected in the corresponding aggregate static graph. The vice-versa is trivially not true, so that two nodes which are strongly connected in the aggregate graph can be temporally disconnected in the time-varying graph.

We are now ready to give the definitions of components associated to a node of a time-varying graph $\mathcal{G}$:

1. The *temporal out-component of node* i, denoted as $\mathrm{OUT}_T(i)$, is the set of vertices which can be reached from i in the time-varying graph $\mathcal{G}$.

2. The *temporal in-component of a node* i, denoted as $\mathrm{IN}_T(i)$, is the set of vertices from which i can be reached in the time-varying graph $\mathcal{G}$.

3. The *temporal weakly connected component of a node* i, denoted as $\mathrm{WCC}_T(i)$, is the set of vertices which i can reach, and from which i can be reached, in the underlying undirected time-varying graph $\mathcal{G}^u$.

4. The *temporal strongly connected component of a node* i, denoted as $\mathrm{SCC}_T(i)$, is the set of vertices from which vertex i can be reached, and which can be reached from i, in the time-varying graph $\mathcal{G}$.

Differently from the case of directed static graphs, it is not possible to define the strongly (weakly) connected components of a time-varying graph starting from the definition of connectedness for pairs of nodes. As we explained above, this is because the relation of strongly (weakly) connectedness for couples of nodes is not an equivalence relation. For this reason, we give the following definition of strongly connected component of a time-varying graph:

Definition 3 (Strongly connected component)
A set of nodes of a time-varying graph $\mathcal{G}$ is a temporal strongly connected component of $\mathcal{G}$ if each node of the set is strongly connected to all the other nodes in the set.

Similarly, a set of nodes is a *weakly connected component* if each node in the set is weakly connected to all the other nodes in the set. The definitions of strongly and weakly connected components enforce transitivity, but the check of strong (weak) connectedness has to be directly performed for every couple of nodes. Suppose for instance that we want to verify if the five nodes in the graph $\mathcal{G}$ shown in Fig. 2 form a strongly connected component. In the static aggregate graph this check has $O(K)$ computational complexity, where K is the total number of links in the graph. In fact, we have only to check that 2, 3, 4 and 5 are connected to 1, which can be done by a *depth first* visit of the graph started at node 1, since node connectedness is an equivalence relation for static graphs, and a component of a node is also a component for the whole graph. On the contrary, for a time-varying graph we should check the connectedness of all the possible couples of nodes, so that a procedure to verify that a set of N nodes forms a strongly connected component has computational complexity $O(N^2)$, for every check, instead of $O(K)$. Moreover, while static directed graphs admit only one partition into strongly connected components, for a time-varying graph there exists in general more than one possible partition, as we shall see in the next section.

IV. THE AFFINE GRAPH OF A TIME-VARYING GRAPH

We show in the following that the problem of finding the strongly connected components of a time-varying graph is equivalent to the well-known problem of finding the maximal-cliques of an opportunely constructed static graph [32]. We call such a static graph the *affine graph* corresponding to the time-varying graph. It is defined as follows:

Definition 4 (Affine graph of $\mathcal{G}$)
Given a time-varying graph $\mathcal{G} \equiv \mathcal{G}_{[t_1,t_M]}$, the associated affine graph $G_{\mathcal{G}}$ is an undirected static graph with the same nodes as $\mathcal{G}$, and such that two nodes i and j are linked in $G_{\mathcal{G}}$ if i and j are strongly connected in $\mathcal{G}$.

In practice, the affine graph of a time-varying graph can be obtained by computing the temporal shortest paths between any two pairs of nodes, and then adding a link between two nodes i and j of the affine graph only if the temporal distance from i to j and the temporal distance from j to i are both finite. Another method to construct the affine graph makes use of the out-components of all the nodes. We start by considering the out-component of the first node, let us say $i = 1$, and then we check, one by one, if for each node $j \in OUT_T(i), j > i$ then also $i \in OUT_T(j)$. If this is true, we put a link between i and j in the affine graph. We then repeat this procedure for the second node, $i = 2$, for the third node, $i = 3$ and so on. We obtain the affine graph by iterating over the out-components of all the nodes. In Fig. 3 we report the affine graph corresponding to the time varying graph shown in Fig. 2. In this graph, node 1 is directly connected to nodes $\{2, 3, 4\}$, since it is temporally strongly connected to them in the time-varying graph. Similarly, node 2 is connected to nodes $\{1, 3, 4, 5\}$, node 3 is connected to $\{1, 2\}$, node 4 is connected to $\{1, 2, 5\}$ and node 5 is connected to $\{2, 4\}$. Hence, the affine graph $G_{\mathcal{G}}$ has only 7 of the 10 possible links, each link representing strong connectedness between two nodes.

We briefly report here some definitions about graph cliques. Given an undirected static graph, a *clique* is a complete subgraph, i.e., a subgraph in which all the nodes are directly linked to each other. A *maximal-clique* is a clique that is not included in any larger clique, while a

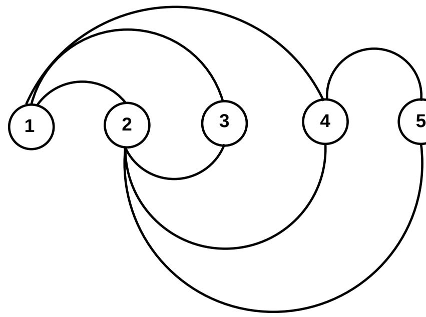

FIG. 3. The affine graph $G_{\mathcal{G}}$ associated to the time-varying graph $\mathcal{G}$ reported in Fig. 2. The affine graph is static and undirected, and each of its maximal-cliques correspond to a strongly connected component of the original time-varying graph $\mathcal{G}$.

maximum-clique is a *maximal-clique* whose size is equal or larger than those of all the other cliques [33].

By construction, a clique of the affine graph $G_{\mathcal{G}}$, contains nodes which are strongly connected to each other, so that the *maximal-cliques* of the affine graph, i.e., all the cliques which are not contained in any other clique, are temporal strongly connected components (SCC_T) of $\mathcal{G}$. Similarly, all the *maximum-cliques* of the affine graph $G_{\mathcal{G}}$, i.e., its largest maximal-cliques, are the largest temporal strongly connected components (LSCC_T) of $\mathcal{G}$. Therefore, the affine graph can be used to study the connectedness of a time-varying graph, and the properties of the strongly connected components of a time-varying graphs can be obtained from known results about maximal-cliques on static graphs. For instance, the problem of finding a partition of $\mathcal{G}$ that contains the minimum number of disjoint strongly connected components is equivalent to the well–known problem of finding a partition of the corresponding affine graph $G_{\mathcal{G}}$ in the smallest number of disjoint maximal-cliques [32]. Unfortunately, this problem is known to be NP–complete, and in practice can be exactly solved only for small graphs. In the case of the affine graph in Fig. 3, it is possible to check by hand that there are only three possible partitions of $G_{\mathcal{G}}$ into maximal-cliques, namely:

1. $\{1,2,3\} \bigcup \{4,5\}$
2. $\{1,2,4\} \bigcup \{3\} \bigcup \{5\}$
3. $\{2,4,5\} \bigcup \{1,3\}$

Notice that the second partition contains two isolated nodes, which are indeed degenerated maximal-cliques. Therefore, the original time-varying graph admits only two different partitions into a minimal number of non-degenerated strongly connected components, namely into two components containing at least two nodes each. One possible partition of our network $\mathcal{G}_{[t_1,t_4]}$ is composed of the components $\{1,2,3\}$ and $\{4,5\}$, while the other partition consists of $\{2,4,5\}$ and $\{1,3\}$. If we discard the temporal ordering of links, we obtain different results. In fact, the aggregate static graph shown in Fig. 2 has only one connected component, which includes all the five nodes.

Other interesting results stem from the mapping into affine graphs and from the following well known results for cliques in graphs.

1. Checking if a graph contains a clique of a given size k has polynomial computational complexity, and precisely $O(N^k k^2)$ [34].

2. The *clique decision problem*, i.e., the problem of testing whether a graph contains a clique larger than a given size $\overline{k}$, is NP–complete [32]. Therefore, any algorithm which verifies if a time-varying graph has a strongly connected component whose size is larger than a fixed value $\overline{k}$, has exponential computational complexity.

3. Listing all the maximal-cliques of a graph has exponential computational complexity, namely $O(3^{N/3})$ on a graph with N nodes [35, 36]. Consequently, finding all strongly connected components of a time-varying graph with N nodes, requires an amount of time which exponentially grows with N.

4. The problem of finding a maximum-clique for an undirected graph is known to be hard–to–approximate [37–39], and an algorithm that finds maximum-cliques requires exponential time. The best algorithm works in $O(\sim 1.2^N)$ for a graph with N nodes [40, 41].

5. The problem of determining if a graph can be partitioned into K different cliques is NP–complete, and consequently also the problem of finding the minimum number of cliques that cover a graph, known as the *minimum clique cover*, is NP–complete [32]. This means that there exists no efficient algorithm to find a partition of a time-varying graph made by a set of disjoint strongly connected components. Moreover, there are in general more than one partition of a graph into maximal-cliques, so that a time-varying graph cannot be uniquely partitioned into a set of disjoint strongly connected components.

The existence of a relation between the strongly connected components of a time-varying graph and the maximal-cliques of its affine graph implies that it is practically unfeasible to find all the strongly connected components of large time-varying graphs. The problem can be exactly solved only for relatively small networks, for which it is computationally feasible to enumerate all the maximal-cliques of the corresponding affine graphs. Even if, in many practical cases, it is possible to find only the maximal-cliques up to a certain size $\overline{k}$, we can still obtain some information about the maximum value of $\overline{k}$ to be checked. First of all, in order to have a clique of size $\overline{k}$ the graph should have at least $\overline{k}$ nodes having at least $\overline{k}$ links. Moreover, each clique of order $\overline{k} > 3$ has exactly

$\binom{\overline{k}}{3}$ sub–cliques of order 3, so that in order for a subgraph to be a clique of order $\overline{k}$, the graph should have at least $\binom{\overline{k}}{3}$ triangles. This means that there is a relation between the number of triangles of the affine graph and the size of its maximum-cliques. In particular, the number of existing triangles in the affine graph fixes an upper bound for the size of the largest admissible maximal-cliques of the graph.

V. RESULTS

As a practical example, in this section we extract and analyze node and graph components of real time-varying social networks. Namely, we consider time-varying networks constructed from the Reality Mining dataset, which records contacts among students and faculty members at the Massachussets Institute of Technology [42]. The information on contacts were collected by providing each participant with a Bluetooth-enabled mobile phone. A Bluetooth device is able to detect similar devices within a limited range (5 to 10 meters), and participants were asked to bring the mobile phone with them all the time. Each device performs a scan procedure to detect other devices every 5 minutes, and stores the list of other detected devices at each scan. We make the reasonable assumption that two individuals are co–located, i.e., they are at the same place, at a given time, if their respective devices detect each other. The dataset includes co–location information among one hundred individuals during six months, from the end of June 2004 to the end of December 2004. At each time t, a co–location graph can be obtained by connecting through undirected links all the nodes which are co–located at that time. We constructed several time-varying graphs, i.e., sequences of co–location graphs obtained every 5 minutes.

Here we report results of component analysis performed on *a)* graphs corresponding to the first half and to the second half of a week, *b)* graphs corresponding to different days of a week and *c)* graphs corresponding to different weeks. In particular, we will focus our attention on the Fall term (namely from start of September to mid of December), which corresponds to weeks from 10 to 19 in the dataset. We chose this dataset for two very simple reasons. First of all, due to the relatively small number of nodes, it is possible to extract all the maximal-cliques of the corresponding affine graphs by using a limited amount of computational resources. Secondly, this dataset represents a real human interaction network and, as we shall see in the following, the approximation made representing it as a static graph, i.e., considering all the links as concurrent in time, is a very poor and unrealistic representation of the system.

In Fig. 4 we consider week 11. For each node, we report the size of temporal in-component (panel a) and temporal out-component (panel b) during the beginning of the week (WB), namely from Monday 12:00am to Thursday 12:00pm (red circles), and during the end of

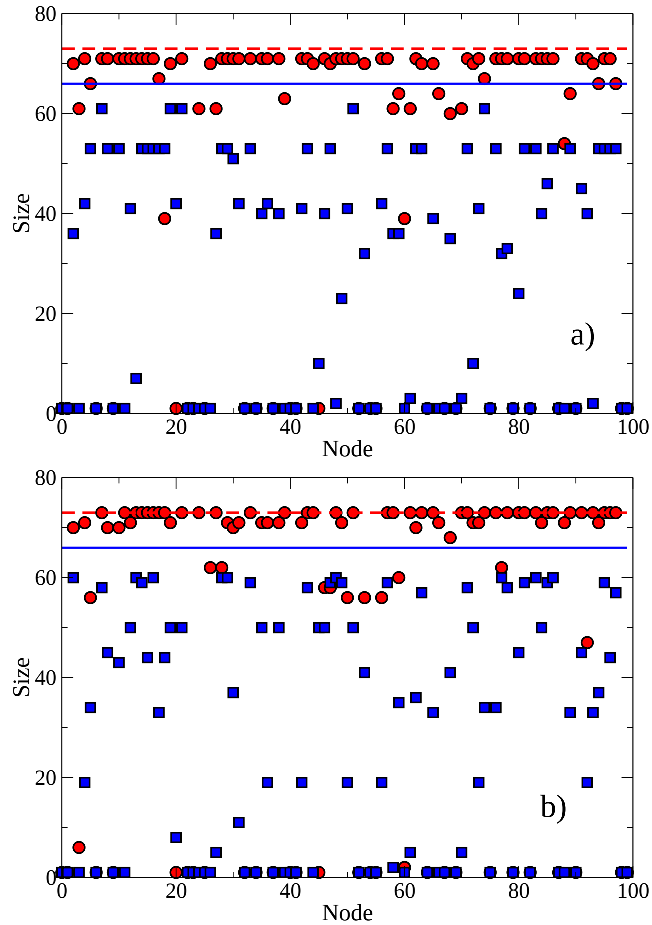

FIG. 4. Size of the temporal in-component (a) and out-component (b) for each of the $N = 100$ individuals during week 11 of the Reality Mining dataset. Red circles and blue squares correspond, respectively, to the beginning of the week (WB) and to end of the week (WE). For comparison, the size of the largest connected component of the corresponding aggregate static graph are reported as dashed red line (WB) and solid blue line (WE), respectively.

the week (WE), namely from Thursday 12:00pm to Sunday 11:59pm (blue squares). During WB almost all nodes have temporal in-components and out-components of similar sizes. In fact, the majority of nodes have in-component of size 72 and out-component of size 74. Conversely, during WE, we observe a wider distribution of the sizes of temporal in- and out-components. In particular, in panel (a) we notice a group of nodes having an in-component of size 53, another group whose in-component contains around 40 nodes, and other nodes with in-component of size smaller than 30. Similarly, in panel (b), there is a group of nodes whose out-component contains around 60 nodes, a second group of nodes with out-component sizes between 40 and 50, and many other nodes having out-component with less than 40 nodes. The observed small variability in the size of node components during WB, is due to the fact that students and faculty members have more opportunities to meet and interact at lectures during WB. Even if not all students attend the same classes, and not all professors teach to all the students, there is a high probability that two individuals would be connected by longer temporal paths. Conversely, during WE, the students usually meet other students in small groups, and they usually do not meet professors and lecturers, except for the classes held on Thursday afternoon and on Friday. As a result, the size of the in- and out-components during WE exhibits large differences from node to node. Such fluctuations are lost in a static graph description, which aggregates all the links independently of their time ordering. In fact, the two static aggregate graphs corresponding respectively to WB and WE, have only one giant connected component, which contains the majority of the nodes, while the remaining nodes are isolated. As comparison, the size of the giant component of the aggregate static graphs for WB and WE are also reported in Fig. 4, respectively as dashed red line and solid blue line. Notice that the static aggregate graph corresponding to a co–location time-varying graph is intrinsically undirected. Therefore, the in- and out-components of a node in this graph coincide and correspond to the component to which the node belongs. Moreover, in a static aggregate graph all the links (and consequently also all the paths) are always available, so that all the nodes in the same connected component have the same component size. As a result, the variability in the node connectedness of the time-varying network, which is evident from the distribution of circles and squares in Fig. 4, is flattened down in the aggregate static graph. In the latter case, all information about network connectedness is represented by a single value, namely the size of the largest connected component, which does not provide any information about the mutual reachability of two generic nodes of such a component. In particular, the size of the giant connected component of the static aggregate graph is equal to 74 during WB and to 66 during WE, despite the fact that in the same intervals the majority of nodes have much smaller temporal in- and out-components.

Monday Week #	K	T	N_s	$\langle s \rangle$	S	N_S	N_U	N_I	C
10	646	4341	22	10.3	27	1	27	27	62
11	554	4414	15	9.1	29	1	29	29	54
12	105	307	11	4.1	13	1	13	13	22
13	772	8322	16	10.6	36	1	36	36	59
14	815	6481	20	12.7	27	1	27	27	62
15	1485	22096	23	23.7	44	1	44	44	67
16	1022	9033	22	16.5	29	1	29	29	70
17	1284	15572	19	22.3	38	1	38	38	67
18	1417	18430	16	20.7	44	1	44	44	67
19	1106	13531	13	20.9	38	2	42	34	60

TABLE I. Structural properties of the affine graph corresponding to the time-varying graph of the first 24 hours of the week (Monday), for each week of the Fall term: number of links (K), number of triangles (T), number of maximal cliques (N_s), average size of maximal cliques ($\langle s \rangle$), size of the largest maximal clique (S), number of largest maximal cliques (N_S), number of nodes in the union (N_U) and in the intersection (N_I) of all largest maximal cliques. The size of the giant component of the corresponding static aggregate graph (C) is reported in the rightmost column.

In Table I we report some relevant structural properties of the affine graphs. We consider and compare the time-varying graphs constructed in the first 24 hours (Monday) of ten consecutive weeks (from week 10 to week 19). We observe large fluctuations in the measured values. The number of links K ranges from 105 in week 12 to 1485 in week 15, while the number of triangles T is in the range $[307, 22096]$, with a mean value around 10000 and a standard deviation equal to 6932. This variance is due to the fact that, even if the daily activity of each individual is, on average, almost periodic, in a particular day we can observe a peculiar temporal pattern of connections, for instance because some students decide to skip a class or because the lessons are suspended for public holidays. In particular, this is exactly what happens on week 12. Monday of week 12 is September 11^{th} 2004, and corresponds to the *Patriot Day*, a national holiday introduced in the US in October 2001, designated in memory of the 2977 killed in the September 11^{th}, 2001 attacks. Therefore, we observe the minimum connectivity and the minimum number of triangles on week 12, because all teaching activities were suspended, and students did not participate to lessons as usual.Also the number N_s and the average size $\langle s \rangle$ of maximal cliques of the affine graphs change from one week to another. In particular, during weeks 10 to 14 we observe relative smaller values of N_s and $\langle s \rangle$ than in weeks 15 to 19, which is probably due to the relatively lower number of links and triangles. Conversely, if we consider the size S of the largest strongly connected component (i.e., the largest maximal-clique of the affine graph), we notice that it is not strongly correlated with K and T. For instance, the size of the largest strongly connected component found at week 11 ($S = 29$) is equal to that observed at week 16. However at week 11 the affine graph has a much smaller number of links and triangles than at week 16. Moreover, on Monday of week 14 we have a maximal-clique of size 27, even if the number of links and triangles is higher than on Monday of week 11. These results confirm that the size of the largest strongly connected component of a time-varying graph is mainly due to the actual configuration of links and triangles of the corresponding affine graph, and not only to their relative number. We notice also that every affine graph reported in Table I admits a single LSCC_T, except at week 19 where two LSCC_Ts of size $S = 38$ emerge. For this reason, we also looked at the number of nodes N_U which participate to *at least one* LSCC_T, and at the number N_I of nodes which participate to *all* LSCC_Ts. These numbers correspond, respectively, to the number of nodes found in the union and in the intersection of all LSCC_Ts. An interesting result is that $N_I = 34$ on week 19, so that 34 nodes participate to both maximal 42-node cliques. These 34 nodes play a very im-

Week #	K	T	N_s	$\langle s\rangle$	S	N_S	N_U	N_I	C
10	2200	45428	10	44.0	61	1	61	61	69
11	2506	54500	12	46.8	64	1	64	64	75
12	2598	57913	12	43.5	66	1	66	66	77
13	2965	71561	9	62.5	69	1	69	69	79
14	2590	56826	15	39.3	64	1	64	64	79
15	3321	85348	9	54.7	74	1	74	74	85
16	2927	69452	9	53.2	70	1	70	70	80
17	2802	66247	10	57.9	69	1	69	69	77
18	2298	47429	12	40.0	61	2	62	60	73
19	2966	70963	13	53.8	69	3	72	68	81

TABLE II. Structural properties of the affine graph corresponding to the time-varying graph of the whole week, for each week of the Fall term. Legend as in Table I.

portant role in the structure of the network. If we remove just one of them, then the resulting affine graph does not have a clique of size 42 any more, and consequently the size of the LSCC_T of the remaining time-varying graph is smaller than 42. At the same time, removing all these N_I nodes will cause a significant reduction in the size of LSCC_Ts, in the number of triangles of the affine graph and, consequently, in the number of SCC_Ts. The nodes that participate to at least one LSCC_T are important in the diffusion of information throughout time-varying graphs. In fact, it is sufficient to pass a message to one of the nodes in a LSCC_T early in the morning, to assure that at least N_U nodes will receive the message before the end of the day.

Finally, in the rightmost column of Table I we report the size C of the giant component of the corresponding static aggregate graph. Notice that for any of the ten weeks under consideration, the value of C is much larger than S, as a consequence of the fact that the static representation of the time-varying graph systematically overestimates node connectedness and paths availability. In panel (a) of Fig. 5 we plot the value of S and C for each Monday of the Fall term. We notice that both C and S are able to capture the anomalous behavior at Monday of week 12 (*Patriot Day*). If we focus our attention on the period from week 13 to week 19, the size of the giant connected component of the aggregate static graph is in the range $[59, 70]$, while the size of the LSCC_T of the time-varying graphs in the same interval exhibits wider fluctuations between $S = 27$ (week 14) and $S = 44$ (week 15 and week 18). This variability is due to the intrinsic fluctuations observed in human contact networks. For instance, some of the students which attended a given class on Monday of week 13, might have decided to remain at home on week 14, and this eventually had an impact on the availability of links and paths, producing smaller strongly connected components. This intrinsic variability is somehow flattened down if we use the standard static component analysis and compute the largest connected component of a static graph which aggregates all the links of one day. Furthermore, we notice the lack of correlation between C and S. (the linear correlation coefficient between C and S from week 13 to week 19 is equal to $r = 0.12$). For instance, at Monday of week 16 we observe the maximum value of C, namely $C = 70$, while the time-varying graph has a largest strongly connected component of size $S = 29$, which is relatively small compared to the other weeks. Conversely, at Monday of week 13 we observe a relatively small giant component, with $C = 59$ nodes, while the size of the largest strongly connected component is $S = 36$.

In order to show the results of our analysis when applied at a larger temporal scale (weeks instead of days), we have reported in Table II the structural properties of the affine graphs constructed from the contacts observed during a whole week. As in Table I, we compare the 10 weeks in the Fall term. We observe a variance in the number of links and triangles: K is in the range $[2200, 3321]$ and T is in the range $[45428, 85348]$, and still there is no appreciable correlation between the average size $\langle s\rangle$ of SCC_Ts and K or T. If we look at panel (b) of Fig. 5, where we report S and C for the time-varying graph corresponding to the whole week, we notice that the size of the LSCC_T at each week is still lower than the size of the giant component of the corresponding aggregate graph. Differently from the case of single days, at a scale of the entire week we observe a clear correlation between S and C. The linear correlation coefficient between C and S, from week 10 to week 19, is now equal to $r = 0.89$. These results confirm that the number and size of strongly connected components in time-varying graphs depend on the length of the period during which we observe the system. In our system at the scale of a week, almost all the nodes are in the largest strongly connected component because longer temporal paths appear, so that the affine graphs at different weeks are more similar to each other and the information extracted from a temporal analysis is similar to that obtained by plotting static measures on aggregated graphs as function of time. On the contrary, at the scale of a day, our system has affine graphs which are disconnected or similar to trees, with very few triangles and relatively small cliques. In this case, as shown in Fig. 5, a temporal component analysis of time-varying graphs reveals interesting details about the dynamics of contacts, which cannot be detected by a static graph analysis.

Finally, in Fig. 6 we show the temporal evolution of S and C during the week. In particular we compare week 13 and week 16. A point of the plot at time t is obtained by considering the time-varying graph constructed from the events occurred in the interval $[0, t]$, where $t = 0$ corresponds to Monday at 00:00. For each of these time-varying graphs we construct the corresponding affine graph to compute $S(t)$, and then we consider the static aggregate graph to obtain $C(t)$. We observe that $S(t)$ is always smaller than $C(t)$, $\forall t$. In particular, until Tuesday at midnight the size of the largest strongly connected component in week 16 is around $S = 30$, which is less than 50% of the size reached on Sunday. Moreover, at Wednesday midnight the maximal-clique contains $S = 54$ nodes, and the size continues to grow until

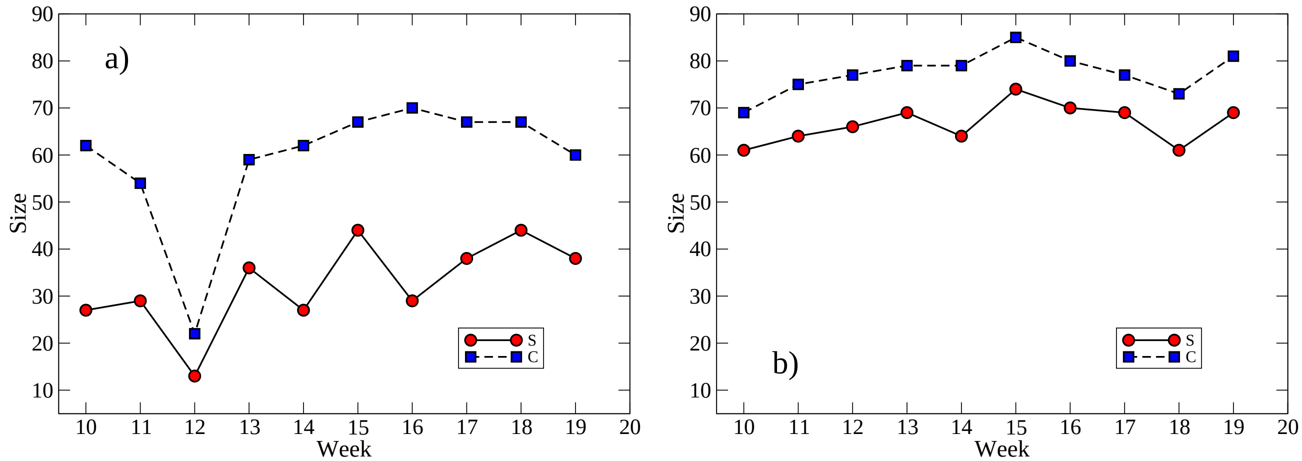

FIG. 5. Panel a): size of the $LSCC_T$ of the time-varying graph on Monday (red circles) and of the giant component of the corresponding static aggregate graph (blue squares). Panel b): the same as panel a) but for the time-varying graph corresponding to the whole week.

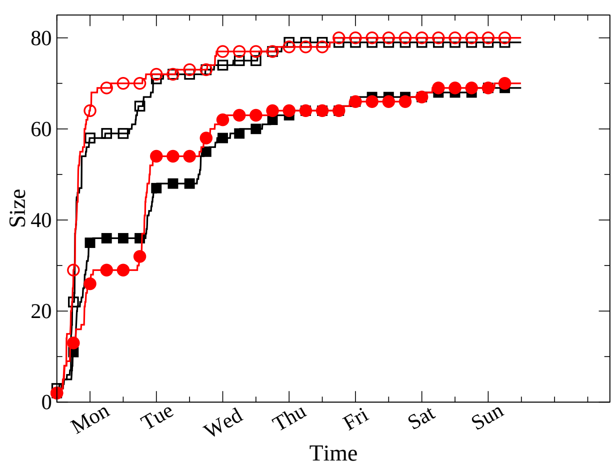

FIG. 6. Size of the largest strongly connected component of the time-varying graph (lines with filled symbols) and size of the giant component of the corresponding static aggregate graph (lines with empty symbols). The black lines with squares correspond to week 13, while red lines with circles correspond to week 16. Large ticks on the x-axis indicate 12:00pm of each day.

the end of the week. Conversely, the size of the giant component of the corresponding aggregate graph on Tuesday at midnight is $C = 73$, which is more than twice larger than the largest strongly connected component at the same time and corresponds to 90% of the size of the giant component at the end of the week. On Friday at midnight the size of the giant component has already reached its maximal value, and does not change any more until the end of Sunday. Notice that the temporal evolution of the size of the giant component over the week looks similar in the two cases, while we observe interesting differences in the temporal evolution of the size of largest strongly connected component. In fact, the size of the $LSCC_T$ at the end of Monday of week 13 is $S = 36$, while at the same time the size of $LSCC_T$ for week 16 is $S = 29$. This indicates that during Monday of week 13 there has been a higher number of contacts than during Monday of week 16. On the contrary, at the end of Tuesday the size of $LSCC_T$ of week 13 is $S = 48$, which is smaller than the value observed at the same time in week 16, i.e., $S = 54$. All these variations, which are due to the temporal correlation and fluctuations in the individuals' connection patterns, disappear in an aggregate static representation.

VI. CONCLUSIONS

Conventional definitions of connectedness and components proposed so far have only considered aggregated, static topologies, neglecting important temporal information such as time order, duration and frequency of links. In this work we have extended the concepts of connectedness to the case of time-varying graphs, and we have introduced definitions of node and graph components which take into account times of appearance and temporal correlations of links. The proposed temporal measures are able to capture variations and fluctuations in the linking patterns, typical of many real social and biological systems. As a first application we have studied a database of human contacts, showing that variations in the pattern of connections among nodes produce relevant differences in the size and number of temporal strongly connected components. We pointed out the important role played by nodes which belong to many strongly connected components at the same time, and we have also analyzed how temporal strongly connected components evolve over time. We hope that our formalism would be useful to analyze other datasets of time-varying networks that will be available in the near future, and to better characterize dynamical processes that take place on these networks, such as diffusion of information and spreading of diseases.

[1] S. Boccaletti, et al., Phys. Rep. **424**, 175 (2006).
[2] A. Arenas, Albert Díaz-Guilera, J. Kurths, Y. Moreno, C. Zhou, Phys. Rep. **469**, 93 (2008).
[3] P. Holme, Phys. Rev. E **71**, 046119 (2005).
[4] J. Kleinberg, Commun. ACM, **51**, 66 (2008).
[5] G. Kossinets, J. Kleinberg, and D. Watts. The Structure of Information Pathways in a Social Communication Network. In *Proc. of ACM SIGKDD '08*, 435 (2008).
[6] B. Tadić and S. Thurner, Physica A **332**, 566 (2004).
[7] M. C. González, C.A. Hidalgo and A.-L. Barabási, Nature **453**, 779 (2008).
[8] A. Buscarino, L. Fortuna, M. Frasca, V. Latora, Europhys. Lett. **82**, 38002 (2008).
[9] B. Tadić, Physica A **293**, 273 (2001)
[10] C. Caretta-Cartozo, P. De Los Rios, F. Piazza, P. Liò, PLoS Computational Biology, **3**, e103 (2007).
[11] A. Ahmed and E. P. Xing, Proc. Natl. Acad. Sci. USA **106**, 11878 (2009).
[12] M. Valencia, J. Martinerie, S. Dupont, and M. Chavez, Phys. Rev. E. **77**, 050905R (2008).
[13] F. De Vico Fallani et al., Journ. Phys. A: Math. Theor. **41**, 224014 (2008).
[14] A.-L. Barabási, H. Jeong, R. Ravasz, Z. Neda, T. Vicsek, and A. Schubert, Physica A **311**, 590 (2002).
[15] A. Clauset and N. Eagle. Persistence and Periodicity in a Dynamic Proximity Network. In *Proc. of DIMACS Workshop on Computational Methods for Dynamic Interaction Network* (2007).
[16] G. Miritello, E. Moro, and R. Lara, Phys. Rev. E **83**, 045102(R) (2011).
[17] J. Sthele, A. Barrat and G. Bianconi, Phys. Rev. E **81**, 035101 (2010).
[18] P. Grindrod, D. J. Higham, Proc. R. Soc. A **466**, 753 (2010).
[19] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela, Science **328**, 876 (2010).
[20] Tang, J., Musolesi, M., Mascolo, C. and Latora, V., Procs. of ACM WOSN09 31 (2009).doi:10.1145/1592665.1592674
[21] J. Tang, M. Musolesi, C. Mascolo, V. Latora, and V. Nicosia, In Procs. of ACM WSNS'10 (2010).
[22] J. Tang, C. Mascolo, M. Musolesi and V. Latora, In Proc. of 12^{th} IEEE Int. Symp. WOWMOM (2011).
[23] J. Tang, S. Scellato, M. Musolesi, C. Mascolo, and V. Latora, Phys. Rev. E **81**, 055101 (2010).
[24] P. Grindrod, D. J. Higham, M. C. Parsons, E. Estrada, Phys. Rev. E **83**, 046120 (2011).
[25] D. J. Higham, P. Grindrod, E. Estrada, "Mathematics Faces up to Facebook", University of Strathclyde Mathematics and Statistics Research Report **19** (2010).
[26] V. Kostakos, Physica A **388**, 1007 (2009).
[27] R. K. Pan and J. Saramäki, arXiv:1101.5913
[28] N. Fujiwara, J. Kurths and A. Díaz-Guilera, Phys. Rev. E **83**, 025101(R) (2011).
[29] Broder, A. et al., Computer Networks **33**, 309 (2000).
[30] Albert, R., Jeong, H. and Barabasi, A., Nature **406**, 378 (2000).
[31] Dorogovtsev, S.N., Mendes, J.F.F. and Samukhin, A.N., Phys. Rev. E **64**, 025101 (2001).
[32] R. M. Karp, Complexity of Computer Computations. New York: Plenum. 85, (1972).
[33] D. B. West, "Introduction to graph theory", Prentice Hall, 2nd ed., ISBN 0130144002 (2001)
[34] R. G. Downey and M. R. Fellows, Theoretical Computer Science **141** (12), 109 (1995).
[35] J. W. Moon, L. Moser, Israel Journal of Mathematics **3**, 23 (1965).
[36] C. Bron, J. Kerbosch, Communications of the ACM **16** (9), 575 (1973).
[37] U. Feige, S. Goldwasser, L. Lovász, S. Safra, M. Szegedy, Proc. 32nd IEEE Symp. on Foundations of Computer Science, 2 (1991).
[38] S. Arora, S. Safra, Journal of the ACM **45** (1), 70 (1998).
[39] S. Arora, C. Lund, R. Motwani, M. Sudan, M. Szegedy, Journal of the ACM **45** (3), 501 (1998).
[40] R.E. Tarjan, A.E. Trojanowski, SIAM Journal on Computing **6**, 537 (1977).
[41] Robson, J. M. (1986), Journal of Algorithms **7** (3), 425 (1986).
[42] N. Eagle A. Pentland, "Reality mining: sensing complex social systems", Pers. Ubiquit. Comput. 10, 255 (2006).