I. INTRODUCTION

Many real-world systems can be studied by using the network paradigm [1,2]. Though in many of these situations the connections between network nodes are constant in time and it is sufficient to adopt a static network description, this approach may not be always suitable and a temporal network description is required [3]. Relevant examples are epidemic spreading, human communication, or transportation networks, i.e., systems where the strong time variability plays a crucial role in determining connections and interactions [4–6]. Also, approaching continuous dynamical systems from a network perspective [7,8] requires a spatiotemporal discretization that, in the nonautonomous case, often determines a marked time-dependence of the resulting networks. This feature is clearly recognizable, in particular, in climate networks [9–12] and in networks describing connectivity by fluid flows [13,14].

Prominent connectivity patterns in networks can be revealed by introducing the concept of paths [1,2]. While its definition is simple and intuitive for static or aggregated networks, for the temporal case paths between nodes can suddenly appear and disappear in time [3,15–22]. Thus, there is recently a focus toward the definition and characterization of paths in temporal networks. The concept of shortest path in static network analysis has been generalized to include information on the time necessary to establish a space-time connection between nodes. This was the motivation behind the development of fastest path analysis, specially for unweighted and undirected networks [3,20]. Although it is relevant to study the time required to build a path among two nodes, it is equally crucial to understand how to quantify the importance and the distribution of such paths. This issue becomes essential when one tries to exploit this information in order to define and evaluate global network properties, for example, betweenness centrality.

In this paper we extend the concept of most probable path (MPP) [23–27] to the case of temporal, weighted, and directed networks. We quantify the relative importance of the MPP with respect to the whole set of paths and to the subset of highly probable paths (HPPs) incorporating most of the connection probability. Using such sets of paths we are able to define a betweenness centrality measure for temporal weighted networks. The approach presented here is applied to a flow network describing ocean water transport by surface currents in the Mediterranean Sea. In this example we demonstrate that information contained in MPPs (or in small subsets of HPPs) suffices to describe all major transport properties, despite the number of such paths being just a very small fraction of the full set. The MPPs correspond to the main carriers of ocean mass transport (showing connectivity patterns) and MPP betweenness to a measure for a clear identification of the main avenues of water transport in the Mediterranean Sea.

II. MOST PROBABLE PATHS

The analysis is restricted to a time interval \([t_0, t_M]\) in which \(M + 1\) snapshots of the state of the network are taken at times \(t_l = t_0 + l\tau, l = 0, 1, \ldots, M\), with \(\tau\) the time between them. We consider a temporal, directed, and weighted network of \(N\) nodes. Its time-dependent connectivity is described by a set of weighted adjacency matrices \(A^{(l)}\), \((l = 1, \ldots, M)\), in which the matrix element \(A^{(l)}_{ij} \geq 0\) specifies the strength of connectivity from \(i\) to \(j\) during the time interval \([t_{l-1}, t_l]\). A convenient way to analyze the system is using time-ordered graphs (TOGs) [19]. Formally, the TOG can be considered a static network of \(N \times (M + 1)\) nodes with directed and causal links. For each snapshot \(l\), a group \(V(t_l)\) of \(N\) nodes replicating the nodes of the original network can be defined. Links are then established only from nodes at successive times, i.e., from \(i_{l-1} \in V(t_{l-1})\) to \(j_l \in V(t_l)\) with the weights given by those in the original temporal network: \(A^{(l)}_{i_{l-1}, j_l}\).

We now consider a flow or transport process by releasing independent random walkers in each node of the network. Their motion is assumed to be Markovian and is defined by single-step transition probabilities proportional to the entries in the adjacency matrices. Specifically the probability of reaching node \(k\) at time \(t_l\) under the condition of being at \(k_{l-1}\) at time \(t_{l-1}\) is

\[
T^{(l)}_{k_{l-1}, k_l} = \frac{A^{(l)}_{k_{l-1}, k_l}}{s^{(l)}_{out}(k_{l-1})}.
\]

(1)

Here \(s^{(l)}_{out}(k) = \sum_j A^{(l)}_{kj}\) is the out-strength of node \(k\) during time step \(l\), so that \(\sum_j T^{(l)}_{kj} = 1\). A generic \(M\)-step path \(\mu\)
between two nodes $I$ and $J$ is defined as a $(M + 1)$-uplet $\mu \equiv \{I, k_1, ..., k_{M-1}, J\}$ providing a sequence of nodes crossed to reach $J$ at time $t_M$ from $I$ at time $t_0$. Under the Markovian assumption the probability for a random walker to take the path $\mu$ under the condition of starting at $I$ is

$$<p_{IJ}^{M}\mu> = T_{I_k}^{(1)} \prod_{l=2}^{M-1} T_{k_{l-1}k_l}^{(l)} T_{k_{M-1}J}^{(M)}.$$ (2)

The MPP $\mu_{IJ}^{M}$ is the path that maximizes Eq. (2) with respect to the intermediate nodes $k_2, ..., k_{M-1}$. Its probability is denoted by $P_{IJ}^{M} = \max_{\mu \in HPP_{IJ}} <p_{IJ}^{\mu}>$. The exact maximization of Eq. (2) can be obtained iteratively by noting that in the first step the maximum probability to reach a given node $k_1$ is simply $P_{Ik_1}^{1} = T_{I_k}^{(1)}$ and then using the recurrence

$$P_{Ik_{l+1}}^{l+1} = \max_k \{ P_{Ik_l}^{l} T_{k_{l+1}k}^{(l+1)} \}$$ (3)

for $l = 1, 2, ..., M - 1$ until reaching node $k_M = J$. This type of iterative optimization is similar (taking logarithms of the probabilities involved) to the one used to find optimal configurations of directed polymers in random media [28] and can be considered as an adaptation of the classical Dijkstra algorithm [29] to the layered and directed structure of the TOG. The computational cost of the maximization is strongly reduced by calculating first accessibility matrices [21] and restricting the maximization search to the set of nodes that are accessible from $I$ and for which $J$ results accessible as well.

To assess whether the MPP alone is a good representation of the transport dynamics we introduce the quantity $\xi_{IJ}^{M}$, the accessibility restricting the maximization search to the set of nodes that are accessible within a time interval $[t_0, t_M]$ and the maximization search to the set of nodes that are accessible from $I$ and for which $J$ results accessible as well.

The MPP $\mu_{IJ}^{M}$ is the path that maximizes Eq. (2) with respect to the intermediate nodes $k_2, ..., k_{M-1}$. Its probability is denoted by $P_{IJ}^{M} = \max_{\mu \in HPP_{IJ}} <p_{IJ}^{\mu}>$. The exact maximization of Eq. (2) can be obtained iteratively by noting that in the first step the maximum probability to reach a given node $k_1$ is simply $P_{Ik_1}^{1} = T_{I_k}^{(1)}$ and then using the recurrence

$$P_{Ik_{l+1}}^{l+1} = \max_k \{ P_{Ik_l}^{l} T_{k_{l+1}k}^{(l+1)} \}$$ (3)

for $l = 1, 2, ..., M - 1$ until reaching node $k_M = J$. This type of iterative optimization is similar (taking logarithms of the probabilities involved) to the one used to find optimal configurations of directed polymers in random media [28] and can be considered as an adaptation of the classical Dijkstra algorithm [29] to the layered and directed structure of the TOG. The computational cost of the maximization is strongly reduced by calculating first accessibility matrices [21] and restricting the maximization search to the set of nodes that are accessible from $I$ and for which $J$ results accessible as well.

To assess whether the MPP alone is a good representation of the transport dynamics we introduce the quantity $\xi_{IJ}^{M}$, the accessibility restricting the maximization search to the set of nodes that are accessible within a time interval $[t_0, t_M]$ and the maximization search to the set of nodes that are accessible from $I$ and for which $J$ results accessible as well.

The MPP $\mu_{IJ}^{M}$ is the path that maximizes Eq. (2) with respect to the intermediate nodes $k_2, ..., k_{M-1}$. Its probability is denoted by $P_{IJ}^{M} = \max_{\mu \in HPP_{IJ}} <p_{IJ}^{\mu}>$. The exact maximization of Eq. (2) can be obtained iteratively by noting that in the first step the maximum probability to reach a given node $k_1$ is simply $P_{Ik_1}^{1} = T_{I_k}^{(1)}$ and then using the recurrence

$$P_{Ik_{l+1}}^{l+1} = \max_k \{ P_{Ik_l}^{l} T_{k_{l+1}k}^{(l+1)} \}$$ (3)

for $l = 1, 2, ..., M - 1$ until reaching node $k_M = J$. This type of iterative optimization is similar (taking logarithms of the probabilities involved) to the one used to find optimal configurations of directed polymers in random media [28] and can be considered as an adaptation of the classical Dijkstra algorithm [29] to the layered and directed structure of the TOG. The computational cost of the maximization is strongly reduced by calculating first accessibility matrices [21] and restricting the maximization search to the set of nodes that are accessible from $I$ and for which $J$ results accessible as well.

To assess whether the MPP alone is a good representation of the transport dynamics we introduce the quantity $\xi_{IJ}^{M}$, the accessibility restricting the maximization search to the set of nodes that are accessible within a time interval $[t_0, t_M]$ and the maximization search to the set of nodes that are accessible from $I$ and for which $J$ results accessible as well.

To assess whether the MPP alone is a good representation of the transport dynamics we introduce the quantity $\xi_{IJ}^{M}$, the accessibility restricting the maximization search to the set of nodes that are accessible within a time interval $[t_0, t_M]$ and the maximization search to the set of nodes that are accessible from $I$ and for which $J$ results accessible as well.

To assess whether the MPP alone is a good representation of the transport dynamics we introduce the quantity $\xi_{IJ}^{M}$, the accessibility restricting the maximization search to the set of nodes that are accessible within a time interval $[t_0, t_M]$ and the maximization search to the set of nodes that are accessible from $I$ and for which $J$ results accessible as well.

To assess whether the MPP alone is a good representation of the transport dynamics we introduce the quantity $\xi_{IJ}^{M}$, the accessibility restricting the maximization search to the set of nodes that are accessible within a time interval $[t_0, t_M]$ and the maximization search to the set of nodes that are accessible from $I$ and for which $J$ results accessible as well.

To assess whether the MPP alone is a good representation of the transport dynamics we introduce the quantity $\xi_{IJ}^{M}$, the accessibility restricting the maximization search to the set of nodes that are accessible within a time interval $[t_0, t_M]$ and the maximization search to the set of nodes that are accessible from $I$ and for which $J$ results accessible as well.

To assess whether the MPP alone is a good representation of the transport dynamics we introduce the quantity $\xi_{IJ}^{M}$, the accessibility restricting the maximization search to the set of nodes that are accessible within a time interval $[t_0, t_M]$ and the maximization search to the set of nodes that are accessible from $I$ and for which $J$ results accessible as well.

To assess whether the MPP alone is a good representation of the transport dynamics we introduce the quantity $\xi_{IJ}^{M}$, the accessibility restricting the maximization search to the set of nodes that are accessible within a time interval $[t_0, t_M]$ and the maximization search to the set of nodes that are accessible from $I$ and for which $J$ results accessible as well.

To assess whether the MPP alone is a good representation of the transport dynamics we introduce the quantity $\xi_{IJ}^{M}$, the accessibility restricting the maximization search to the set of nodes that are accessible within a time interval $[t_0, t_M]$ and the maximization search to the set of nodes that are accessible from $I$ and for which $J$ results accessible as well.

To assess whether the MPP alone is a good representation of the transport dynamics we introduce the quantity $\xi_{IJ}^{M}$, the accessibility restricting the maximization search to the set of nodes that are accessible within a time interval $[t_0, t_M]$ and the maximization search to the set of nodes that are accessible from $I$ and for which $J$ results accessible as well.

To assess whether the MPP alone is a good representation of the transport dynamics we introduce the quantity $\xi_{IJ}^{M}$, the accessibility restricting the maximization search to the set of nodes that are accessible within a time interval $[t_0, t_M]$ and the maximization search to the set of nodes that are accessible from $I$ and for which $J$ results accessible as well.

To assess whether the MPP alone is a good representation of the transport dynamics we introduce the quantity $\xi_{IJ}^{M}$, the accessibility restricting the maximization search to the set of nodes that are accessible within a time interval $[t_0, t_M]$ and the maximization search to the set of nodes that are accessible from $I$ and for which $J$ results accessible as well.
and 3 months, a time interval during which the horizontal-flow assumption remains a good approximation). We build ten temporal networks, each one having \( t_0 \) as January 1 of each of the years available in the simulation database (2002–2011).

V. RESULTS

In Fig. 1 (left panel) we show on map the set of all the MPPs of \( M = 9 \) time steps starting from a given node in one of our temporal networks (the one corresponding to 2011), and we notice how many different connections are possible from a single starting node. The \( P^{M}_{I,J} \) values span several orders of magnitude and this behavior is typical for the distribution of probability across MPPs. We stress here that MPPs do not coincide in general with fastest paths: the fastest connection among two nodes is not always the most probable one (see Appendix), stressing the importance of a weighted description of the network.

To assess how representative of the whole dynamics are MPPs such as the ones shown in Fig. 1, we show in Fig. 2 the distribution of \( \lambda^{M}_{I,J} \) and \( \lambda^{M}_{I,J}(\epsilon) \) for two values of \( M \). The distributions are collected from the median values of the whole set of accessible pairs \( (I,J) \) in our ten temporal networks. For small \( M \) most of the MPP have significant \( \lambda \) values, but as \( M \) increases the peak in the distribution of \( \lambda^{M}_{I,J} \) shifts toward zero (we have checked that exponentially as a consequence of the dramatic increase in the number of available paths between two nodes. Then, it becomes important to consider larger sets of paths such as \( K^{M}_{I,J}(\epsilon) \). For the cases plotted, i.e., \( M = 5,6 \) and \( \epsilon = 0.1 \), the mean values of \( \lambda^{M}_{I,J}(\epsilon) \) are around 0.5. This means that, despite \( K^{M}_{I,J}(\epsilon) \) may not contain the full set of paths that individually carry a probability larger than \( \epsilon P^{M}_{I,J} \), it is large enough so that, for most of the \( (I,J) \) pairs, it contains globally over 50% of the connection probability. This result further gains meaning if we consider that the number of paths in \( K^{M}_{I,J}(\epsilon) \) is on average well below 1% of the total number of paths of \( M = 5 \) and 6 steps. Hence, despite the strong particle dispersion characterizing our flow networks it is true that only a small subset of paths contribute significantly to the main transport features. This conclusion will also show up when studying global network properties, such as the betweenness centrality measure.

In Fig. 3 we show the multistep MPP betweenness \( B^{[6,9]}_{I,J,K} \), averaged over our ten networks. We have noticed that the distribution of the betweenness decreases exponentially at large betweenness values so that there are not strong hubs in the network. Spatial patterns determined by the transport dynamics of the flow are clearly evident in the figure where high betweenness areas are organized in one-dimensional-like structures corresponding to the main corridors of transport, i.e., narrow pathways that connect different regions of the ocean. Main paths of the Mediterranean Sea are observed like Cyprus and Rhodes Gyres, surrounding the Ionian basin, the Algerian current, and those along the Sicily strait, etc. Note that because of the ten-years average, individual short-lived mesoscale features (eddies and fronts) are averaged out. We note, however, that despite this interyear average, MPP
betweenness maps remain dependent on the starting date \(t_0\). These observations highlight the seasonal and interannual variability of the flow and justifies further the need of a time-dependent network approach as opposed to a fully averaged static network description.

The robustness of our methodology has been tested by checking the stability of our results on MPP betweenness under different conditions. First, dealing with temporal networks it is important to understand how much the results are affected by the choice of the time-step duration \(\tau\) [34]. We checked this issue considering different \(M\) and \(\tau\) values but the same total duration. Results for the MPP betweenness remain nearly unchanged when keeping a constant total duration \(M\tau\), confirming robustness under variations of the temporal resolution. Second, we noted that the MPP betweenness does not significantly change when computed with just the 50% of MPPs with larger values of \(\lambda\) (i.e., when using a threshold to retain only the most significant MPPs). Finally, to support our interpretation of most probable paths as main carriers of connectivity, we considered also betweenness calculated from HPP subsets so that \(k_{M,K}^{ij}\) is now the number of times node \(K\) appears in the set \(K_{M,I}^{ij}(\epsilon)\) of HPPs between \(I\) and \(J\), with \(\epsilon = 0.1\). We did not appreciate relevant differences between this calculation and the one involving only MPPs. Indeed the Pearson correlation coefficient between the two betweennesses is larger than 0.9. Hence, despite the MPPs represent a small portion of the paths in the \(k_{M,I}^{ij}(\epsilon)\) subsets (between 3 and 10% for \(\epsilon = 0.1\), depending on the value of \(M\)), which is itself a very small fraction of the full set of paths in the network, they seem to be representative of the main spatiotemporal structures describing the global dynamics. Indeed, center and right panels of Fig. 1 show that most of the relevant paths remain spatially close to the MPP. This observation is confirmed by calculations of the spatial dispersion between paths in \(k_{M,I}^{ij}(\epsilon)\), whose average turns out to be of the order of the size of the boxes defining the nodes.

VI. CONCLUSIONS

In this paper we introduced tools to compute highly probable paths in weighted temporal networks and to evaluate their relative importance. Betweenness centrality measures based on them have also been introduced. We applied this approach to characterize connectivity in the Mediterranean Sea from a network-theory perspective. Here, MPPs correspond also to the set of paths that maximize the fraction of transported mass, giving therefore a clear physical interpretation of connection probabilities. Despite MPPs representing only a small fraction of the whole set of paths, we found that they suffice to highlight the main transport pathways across our network, since most of the HPPs remain geographically close to them. This means that paths followed by fluid masses connecting two regions are organized in elongated narrow tubes centered on the MPP.

We believe that the study of fluid transport as a network will provide new tools and insights complementing standard Lagrangian methods. Indeed most of these are devoted to the identification of barriers to transport or coherent regions with small fluid exchange with the surroundings. Here we are instead addressing the opposite question: how to detect regions and pathways that maximize fluid interchange across the network. Even though in principle pathways are simply given by trajectories, it is almost impossible to extract clear and significant patterns from the complex sets of trajectories that arise in all, except the most simple, time-dependent flows. Our approach allows us to quantify explicitly transport among two subregions of the domain, highlighting the optimal path connecting them. In this sense MPP analysis differs from simply studying the evolution in time of tracer concentrations seeded in a given region.

Beyond the fluid dynamics context, MPPs and the MPP-betweenness measure here introduced could be easily transferred to other kinds of weighted temporal networks. This could be relevant, for instance, in defining vulnerability metrics in disease spreading processes, individuate critical nodes in biological and ecological networks, or in detecting bottlenecks of reaction chains in metabolic networks.

ACKNOWLEDGMENTS

We acknowledge financial support from FEDER and MINECO (Spain) through the ESCOLA (Grant No. CTM2012-39025-C02-01) and INTENSE@COSYP (Grant No. FIS2012-30634) projects, and from European Commission Marie-Curie ITN program (FP7-320 PEOPLE-2011-ITN) through the LINC project (Grant No. 289447). We acknowledge helpful discussions with P. Fleurquin and V. Rossi.

APPENDIX: COMPARING FASTEST AND MOST PROBABLE PATHS

In the study of temporal networks, the concept of fastest path has been put forward as a natural extension of the shortest path of static networks. In our work we define and analyze a different type of relevant path, which is the MPP. It is important to address the differences between most probable and fastest paths, and we do so in this Appendix.

The MPP refers to the path transporting the maximum fraction of water (or of probability) between two nodes, and the fastest path to the pathway linking the two nodes in the shortest time. This second concept cannot be implemented when the number of time steps is fixed. However, we can reclaim the concept of fastest path within a multistep approach, i.e., by looking at a time window specified by a range of values for the number...
FIG. 4. (Color online) We show the fastest MPP (dashed blue line) and the absolute MPP (continuous red line), between an origin node $I$ (black star) and a destination node $J$ (black triangle). The considered full set of paths ranges The fastest MPP reaches the destination node in four steps of $\tau = 10$ days while the absolute MPP needs eight steps, i.e., 40 days more. The probability associated to the fastest MPP is $5.9 \times 10^{-7}$ and the probability of the absolute MPP is $6.7 \times 10^{-6}$.

of time steps $M$. We can then define the set $M^{|M_{\text{max}} - M_{\text{min}} + 1|}$ of $(M_{\text{max}} - M_{\text{min}} + 1)$ MPPs for the pair $I, J$ for $M \in [M_{\text{min}}, M_{\text{max}}]$, and the fastest MPP as the MPP in $M^{|M_{\text{max}} - M_{\text{min}} + 1|}$ corresponding to the smaller $M$. On the other side we can also define an absolute MPP, i.e., the MPP in $M^{|M_{\text{max}} - M_{\text{min}} + 1|}$ having the highest probability. By comparing the set of absolute MPPs with the set of fastest MPPs we can address the question: is the fastest path necessarily the most probable?

In Fig. 4 we show that the fastest MPP among two nodes of the network is different from the absolute MPP and that its probability, in several cases, can be orders of magnitude smaller. We considered for this example paths ranging from $M = 3$ to $M = 9$ steps of ten days (i.e., a time scale of 1–3 months) with starting date January 1, 2011. The results show the importance of distinguishing between the connections realized in the shortest time and the connections that carry most of the transported mass (the most probable).

To display in a more systematic way the differences between fastest and absolute MPPs across the network, we study the rank plot of the whole set of paths during ten years (2002–2011) ranging from three to nine steps of ten days starting at January 1 of each year (see Fig. 5). The rank plot displays the probabilities of each path in one of the sets sorted in decreasing order. We see a gap in probabilities between the two sets of about one order of magnitude in most of the range displayed. The fastest MPPs have probabilities significantly smaller than absolute MPPs.

Finally, we also evaluate how these differences are reflected in the betweenness measures. We define the relative difference among the betweenness computed using the set of fastest MPPs and absolute MPPs for the node $K$ as

$$\Delta_K = \frac{2B_{abs}^K - B_{fast}^K}{B_{abs}^K + B_{fast}^K},$$

(A1)

where $B_{abs}^K$ is the betweenness computed using absolute MPPs and $B_{fast}^K$ the betweenness computed using fastest MPPs. We consider once more paths ranging from three to nine steps of ten days with starting date January 1, 2011, and we compute the spatial average (i.e., the average over nodes $K$) for the absolute value of the relative difference finding $\langle|\Delta_K|\rangle_K = 0.32$. This means that, on average, the difference between the two measures is about 30%.

We stress that all the comparisons above are among paths that are already MPPs linking a pair of nodes. Considering still fastest paths (for example, the one by which the very first particle from one node reaches the other) will lead to much stronger differences. In summary, the results show the importance of distinguishing between the connection realized in the shortest time and the connection that carries most of the transported mass. This gains even more relevance when considering possible applications, such as rescue operations, pollutant-spreading, or biological connectivity.