PHYSICAL REVIEW E **76**, 046106 (2007)

Analyzing trails in complex networks

Luciano da Fontoura Costa,* Francisco A. Rodrigues,† and Gonzalo Travieso‡
Instituto de Física de São Carlos, Universidade de São Paulo, P. O. Box 369, São Carlos, São Paulo 13560-970, Brazil

(Received 24 April 2007; revised manuscript received 1 August 2007; published 11 October 2007)

Even more interesting than the intricate organization of complex networks is the dynamical behavior of systems underlain by such structures. Among the many types of dynamics, one particularly interesting category involves the evolution of trails left by moving agents progressing through random walks and dilating processes in a complex network. The emergence of trails is present in many dynamical process, such as pedestrian traffic, information flow, and metabolic pathways. Important problems related to trails include the reconstruction of the trail and the identification of its source, when complete knowledge of the trail is missing. In addition, the following of trails in multiagent systems represents a particularly interesting situation related to pedestrian dynamics and swarming intelligence. The present work addresses these three issues while taking into account permanent and transient marks left in the visited nodes. Different topologies are considered for trail reconstruction and trail source identification, including four complex network models and four real networks, namely, the Internet, the U.S. airlines network, an email network, and the scientific collaboration network of complex network researchers. Our results show that the topology of the network influences trail reconstruction, source identification, and agent dynamics.

DOI: 10.1103/PhysRevE.76.046106 PACS number(s): 89.75.Hc, 89.75.Fb, 89.70.+c

I. INTRODUCTION

Complex networks have become one of the leading paradigms in science thanks to their ability to represent and model highly intricate structures (e.g., [1–4]). However, as a growing number of works have shown (e.g., [2,3]) the dynamics of systems whose connectivity is defined by complex networks is often even more complex and interesting than the connectivity of the networks themselves. One particularly interesting type of nonlinear dynamics involves the evolution of *trails* left by moving agents during random walks or dilation processes along the network . The term "dilation" refers to the progressive visiting of neighboring nodes after starting from one or more nodes. For instance, starting from node i, at each subsequent time the neighbors of i are visited, then their unvisited neighbors, and so on, defining a hierarchical system of neighborhoods (e.g., [5–7]). Although the dynamics is described as agents visiting network sites, it can be considered also as the evolution of activity in the nodes of the network, where each network edge represents the possibility of activity propagation between the corresponding nodes. Another important related problem involves attempts to recover incomplete trails. In other words, in cases in which only partial evidence is available to observation, it becomes important to try to infer the full set of visited nodes.

The emergence of trails has been studied as representing an interesting type of self-organizational system. Helbing *et al.* [8] proposed a model of pedestrian motion in order to explore the evolution of trails in urban green areas. Also, trails have been considered in swarming intelligence analysis [9,10], not only as a means to understand animal behavior [11], but also as a source of insights for new optimization and routing algorithms [12,13]. These works considered the evolution of trails in regular grids. However, the communication structures where the trail can be defined are not homogeneous in many cases. Many systems, such as the Internet [5], social relationships [14], the distribution of streets in cities [15], and the connections between airports [16], are defined by an irregular topology—more specifically, most of these systems are represented by scale-free networks [4]. Here, we study the influence of different topologies in trail recovery, source identification, and agent dynamics.

The analysis of trails left in complex networks can have many useful applications. For instance, in information networks the recovery of the trail left by a spreading virus on the Internet can be useful to identify the source of contamination and propose strategies for computer immunization. Similarly, the identification of the origin of rumors, diseases, fads, and opinion formation [17] is important to understand human communication dynamics. Another relevant problem is related to traffic improvement and security. In the former case, identification of the covered trails by packages exchanged between computers can help the development of optimal routing paths. In the latter, the source of terrorism strategies and drug trafficking can be determined by analysis of clues identified in social and airline networks. The analysis of trails can also have useful applications in biology. For instance, in ecology, trail analysis can be applied to quantify the interference of human activity in animal behavior and to identify a focus of pollution. In paleontology, the recovery of the trails of animal displacement by fossil analysis can help the understanding of diversification among species. In epidemiology, the identification of disease source can help to stop the spreading process as well as to devise effective prevention strategies.

In order to properly represent trails occurring in complex networks, we associate state variables with each node i, $i=1,2,\ldots,N$, of the network. The trail is then defined by marking such variables along the corresponding dynamical

---

*Fax: +55 16 3371 3616. luciano@if.sc.usp.br
†Fax: +55 16 3371 3616. francisco@ifc.usp.br
‡Fax: +55 16 3371 3616. gonzalo@ifsc.usp.br

process. Only trails generated by self-avoiding random walks and dilations are considered in the current work; these are characterized by the fact that a node is never visited more than once. We restrict our attention to binary trails, characterized by binary state variables.[1] The types of trails can be further classified by considering the marks to be permanent or transient. In the latter case, the mark associated with a node can be deleted after the visit. While many different transient dynamics are possible, we restrict our attention to the following two types: (i) *Poissonian*, where each mark has a fixed probability of being removed after the visit; and (ii) *evanescent*, where the only observable portion of the trail corresponds to the node(s) being currently visited.

The current work addresses the problem of recovering trails in complex networks and identifying their origin, while considering permanent and transient binary marks in four different network models, namely, the Erdős-Rényi, Watts-Strogatz, Barabási-Albert, and Dorogovtsev-Mendes-Samukhin models; and four real networks: the Internet at the autonomous system level, the U.S. airlines network, an email network from the University Rovira i Virgili, and the scientific collaboration network of complex network researchers. We also consider the analysis of agent propagation using the four network models. The next sections start by presenting the basic concepts in complex networks and trails and follow by reporting the simulation results, with corresponding discussion.

II. BASIC CONCEPTS IN COMPLEX NETWORKS AND TRAILS

An undirected complex network (or graph) G is defined as $G=(V,Q)$, where V is the set of N nodes and Q is the set of E edges of the type $\{i,j\}$, indicating that nodes i and j are bidirectionally connected. Such a network can be completely represented in terms of its *adjacency matrix* K, such that the presence of the edge $\{i,j\}$ is indicated as $K(i,j)=K(j,i)=1$ [otherwise $K(i,j)=K(j,i)=0$]. The degree of a node i corresponds to the number of edges connected to it, which can be calculated as $k(i)=\sum_{j=1}^{N}K(i,j)$. The clustering coefficient is related to the presence of triangles (cycles of length 3) in the network [18]. The clustering coefficient of a node i is given by the ratio between the number of edges among the neighbors of i and the maximum possible number of edges among these neighbors; the clustering coefficient of the network is the average of the clustering coefficient of its nodes.

This paper considers four theoretical network models and four real complex networks. The network models are (a) Erdős-Rényi (ER) [19], (b) Watts-Strogatz (WS) [18], (c) Barabási-Albert (BA), [1] and (d) Dorogovtsev-Mendes-Samukhin (DMS) [20]. In the first model, networks are constructed by considering a constant probability λ of connection between any pair of nodes; in the second, networks start with a regular topology, whose nodes are connected in a ring to a defined number κ of neighbors in each direction, and later the edges are rewired with a fixed probability; networks of the third and fourth models are grown by starting with m_0 nodes and progressively adding new nodes with m edges, which are connected to the existing nodes with probability proportional to their degree (e.g., [1]). The DMS model differs from the BA model by adding an initial attractiveness k_0 to each node, independent of its degree. When $k_0=0$, the DMS model is similar to the BA model [20]. All simulations considered in this work assume that the networks have the same number of nodes $N=1000$ and average degree $\langle k\rangle=2m=\lambda(N-1)=2\kappa=4$. The real networks considered in this work are the Internet at the level of autonomous systems,[2] the U.S. airlines [21], the email network from the University Rovira i Virgili (Tarragona) [22], and the scientific collaboration of complex network researchers.[3]

Trails are generated as subsets of the nodes V are visited during the evolution of random walks or dilations through the network. We assume that just one trail is allowed at any time in a complex network. We consider only self-avoiding random walks, in which no node is visited more than once. At each node, the agent chooses a new node to be visited at random among the not yet visited neighbors of the node. To understand the dilation process, consider $\nu(i)$, the set of neighbors of node i. Starting with i_0, the initial node of the propagation (origin), all nodes in $\nu(i_0)$ are visited; after that, for all $j\in\nu(i_0)$, the nodes in $\nu(j)$ not yet visited are recursively visited; this process is repeated for a given number of neighborhood hierarchies (e.g., [5–7,23]); see Fig. 1.

In order to represent trails, we associate two binary state variables $v(i)$ and $s(i)$ with each node i, which can take the values 0 (not yet visited) or 1 (visited). The state variables $v(i)$ indicate the real visits to each node but are available only to the moving agents; the state variables $s(i)$ are the "marks" of the visits yet available for observation, providing not necessarily complete information about the visits. The structure of the network is assumed to be known to the observer and possibly also to the moving agent(s). Such a situation corresponds to many real problems. For instance, if the trail is being defined as an exploring agent moves through unknown territory, the agent may keep some visited places marked with physical signs (e.g., flags or stones) which are accessible to observers, while keeping a complete map of visited sites available only to her/himself. Trails are here classified as *permanent* or *transient*. In the case of permanent trails, $s(i)=v(i)$, i.e., all visited nodes are known. In the transient type, the state variables $s(i)$ of each node i can be reset to zero after being visited. Transient trails can be further subdivided into (i) Poissonian, characterized by the fact that each visited node has a fixed probability γ of not being observed, i.e., for nodes with $v(i)=1$, $s(i)$ is 1 with probability

---

[1] In other words, a node can be marked as either already visited (1) or not (0). Graded states, e.g., indicating the time of the visit, are considered only on analysis of dynamical agent propagation in Sec. VI.

[2] The data considered in our work are available at the website of the National Laboratory of Applied Network Research [25]. We used the data collected in February 1998.

[3] The scientific collaboration of complex network researchers was compiled by Newman from the bibliographies of two review articles on networks [2,3].

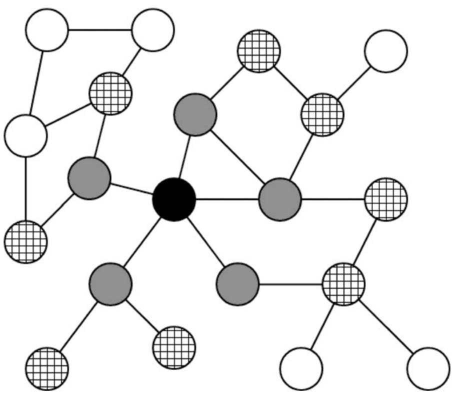

FIG. 1. Dilating trail with two levels in a simple network. The origin of this two-hierarchy trail is the black node, whose immediate neighbors are marked in gray. The nodes with the crossed pattern correspond to the neighbors of the neighbors of the source of the trail. The corresponding evanescent trail would include only the crossed nodes. A Poissonian version of this trail would imply a ratio γ of unmarked (and unobservable) nodes.

$1-\gamma$ and 0 with probability γ [nodes with $v(i)=0$ always have $s(i)=0$]; and (ii) evanescent, where only the last visited nodes are accessible to the observer. Figure 2 shows a classification of the main types of trails considered in this work.

The *real extension* of a trail is defined as being equal to the sum of the state variables $v(i)$. The *observable extension* of a trail is equal to the sum of the state variables $s(i)$. Given a trail, we can define the *observation error* as being equal to

$$\epsilon=\sum_{i=1}^{N}\left[1-\delta(v(i),s(i))\right], \tag{1}$$

where $\delta(a,b)$ is the Kronecker delta function, yielding 1 when $a=b$ and zero otherwise. Note that this error measures the incompleteness of the information provided to the observer. It is also possible to normalize this error by dividing it by N, so that $0\leq\epsilon\leq 1$; this normalization is not used in this work.

It is assumed that the observer will try to recover the original, complete, trail from its observation. In this case, the observer applies some heuristic in order to obtain a recovered trail specified by an additional set of state variables $r(i)$ [$r(i)=1$ if node i is in the recovered trail]. Such a heuristic may take into account the *overlap error* between the observable states $s(i)$ and the recovered values $r(i)$, defined as

$$\xi=\sum_{i=1}^{N}\left[1-\delta(s(i),r(i))\right]. \tag{2}$$

Note that as the observer has no access to $v(i)$, the recovery error has to be estimated using $s(i)$. The actual *recovery error*, which can be used to infer the quality of the recovery, is given by

$$\rho=\sum_{i=1}^{N}\left[1-\delta(v(i),r(i))\right]. \tag{3}$$

Figure 3 illustrates the three state variables related to each network node and the respectively defined errors.

When using recovery heuristics based on the evaluation of the overlap error, it may happen that two or more different recovered trails yield the same overlap error. In this case, it is interesting to consider two additional parameters in order to quantify the quality of the recovery: (i) the number M of estimated trajectories corresponding to the minimum overlap error; and (ii) the fraction f of times that the correct source can be found among the M recovered trails. When average values of M and f are close to 1, it means that the recovery strategy is precise.

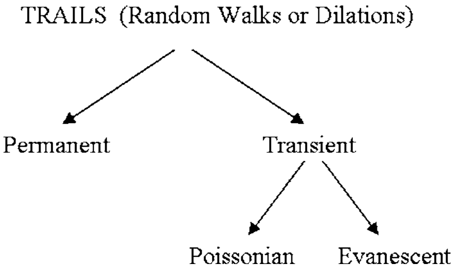

FIG. 2. Trails, including those defined by random walks and dilations, can be subdivided as being permanent or transient. The latter type can be further subdivided into Poissonian and evanescent.

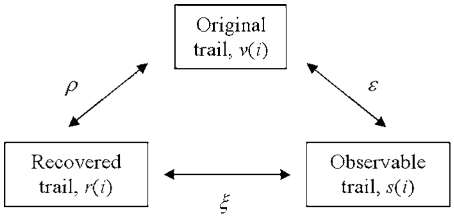

FIG. 3. The three state variables associated with each network node i and the defined errors ϵ, ξ, and ρ.

III. PROBLEMS CONSIDERED

Although the problem of trail analysis in complex networks is potentially very rich and can be extended to many possible interesting situations, for simplicity's sake we restrict our interest to the three following cases.

Poissonian trails from random walks. Because the consideration of permanent and evanescent trails left by random walks is trivial,[4] we concentrate our attention on the problem of recovering Poissonian trails left by single moving agents during random walks. Once such a trail is recovered, its source can be estimated as corresponding to one of its two

---

[4] Permanent trails left by random walks require no recovery, while their source should necessarily correspond to either of its two extremities. Evanescent trails defined by random walks are meaningless, as only the current position of the single agent is available to the observer.

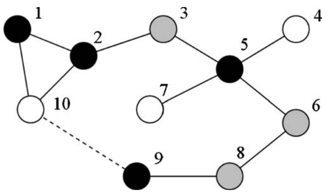

FIG. 4. Example of simple Poissonian trail in a network. The black nodes correspond to s; the original trail included the black and gray nodes.

extremities; we do not consider the problem of source identification for this kind of trail. The recovery error is used to measure the quality of the reconstructed trail.

Poissonian trails from dilations. In this case, only a fraction of the nodes visited by the dilating process is available to the observer. Two problems are of interest here, namely, recovering the trail and identifying its origin. To quantify the quality of the recovery, we evaluate the average values of the number of trails with minimal overlap error $\langle M \rangle$ and the fraction of correct source identifications $\langle f \rangle$.

Evanescent trails from dilations. In this type of problem, only the currently visited nodes are available to the observer, who is requested to reconstruct the trail and infer its possible origin. This corresponds to potentially the most challenging of the considered situations. Note that this case too is subject to random removal of marks, i.e., the values of $s(i)$ are not only of the evanescent type but also can be randomly changed to 0. The results are evaluated by computing $\langle M \rangle$ and $\langle f \rangle$.

IV. STRATEGIES FOR RECOVERY AND SOURCE IDENTIFICATION

Several heuristics can possibly be used for recovering a trail from the information provided by K and $s(i)$. In this work, we consider a strategy based on the topological proximity on the network between nodes with $s(i)=1$ that are not connected. In the case of trails left by random walks, the following algorithm is used.

(1) Initialize a list r as being equal to s.

(2) For each node i with $s(i)=1$: (a) Identify the node j with $r(i)=1$ which is connected to at most one other node with $r(i)=1$ and is closest to i (in the sense of the shortest topological path, but excluding shortest paths with length 0 or 1 in the network); (b) obtain the list L of nodes linking i to j through the corresponding shortest path (if more than one shortest path exists, one of them is chosen at random); and (c) for each node k in L, make $r(k)=1$.

After all nodes with $s(i)=1$ have been considered, the recovered trail will be given by the nodes with $r(i)=1$.

Figure 4 illustrates a simple Poissonian random walk trail, where the black nodes are those in s. The original trail is composed of the nodes in s plus the gray nodes. It can be easily verified that the application of the above reconstruction heuristic will properly recover the original trail in this particular case. More specifically, we would have the following sequence of operations.

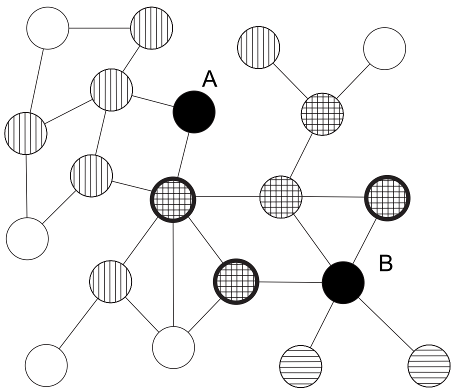

FIG. 5. Simple illustration of the source of ambiguity in trail source determination. See text for explanation.

Step 1. Node 1 is connected to node 5 through the shortest path (1, 2, 3, 5).

Step 2.: Node 2 is connected to node 5 (no effect).

Step 3. Node 5 is connected to node 2 (no effect).

Step 4. Node 9 is connected to node 5 through the shortest path (9, 8, 6, 5).

However, if the dashed edge connecting nodes 9 and 10 were included into the network, a large recovery error would have been obtained because the algorithm would link node 9 to node 1 or 2 and not to node 5.

A different strategy is used for recovery and source identification in the case of dilation trails, which involves repeating the dilation dynamics while starting from each of the network nodes. The most likely recovered trails are those corresponding to the smallest obtained overlap error. Note that more than one trail may correspond to the smallest error. Also, observe that the possible trail sources are simultaneously determined by this algorithm. Actually, it is an interesting fact that complete recovery of the trail is automatically guaranteed once the original source is properly identified. This is an immediate consequence of the fact that the recovery strategy involves the reproduction of the original dilation, so that the original and obtained trails for the correct source will necessarily be identical.

Some additional remarks are required in order to clarify the reason why more than one trail can be identified as corresponding to the minimal overlap error in Poissonian dilation trails. Figure 5 illustrates a simple network with two trails extending through two hierarchies, one starting from the source A and the other from B, which are respectively identified by the vertical and horizontal patterns. Note that some of the nodes are covered by both trails, being therefore represented by the crossed pattern. Now, assume that the original trail was left by A but that the Poissonian version incorporated only the three nodes with thick border (i.e., all the other nodes along this trail were deleted before presentation to the observer). Because the three nodes are shared by both trails, the same overlap error will be obtained by starting at node A or B. It is expected that the higher the value of γ, the more ambiguous the source identification becomes.

When many possible recovered trails with the same overlap error are found, i.e., when $M>1$, the identification of the source is ambiguous. To take this fact into account, in those

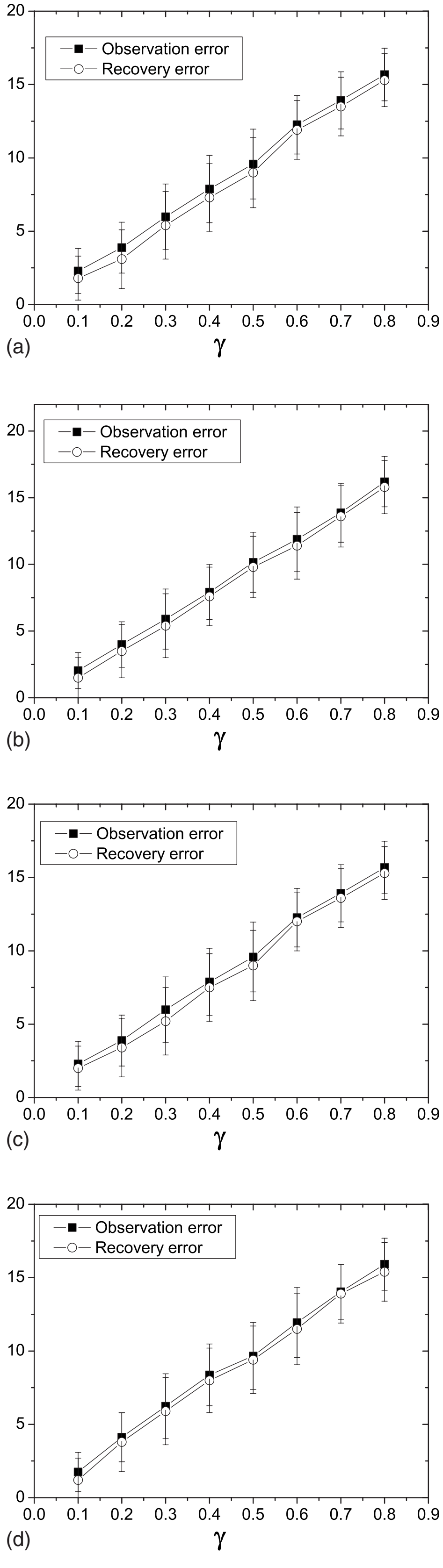

FIG. 6. Observation error (black squares) and recovery error (white circles) obtained by using the recovery algorithm for Poissonian trails from random walks in the (a) ER, (b) WS, (c) BA, and (d) DMS network models.

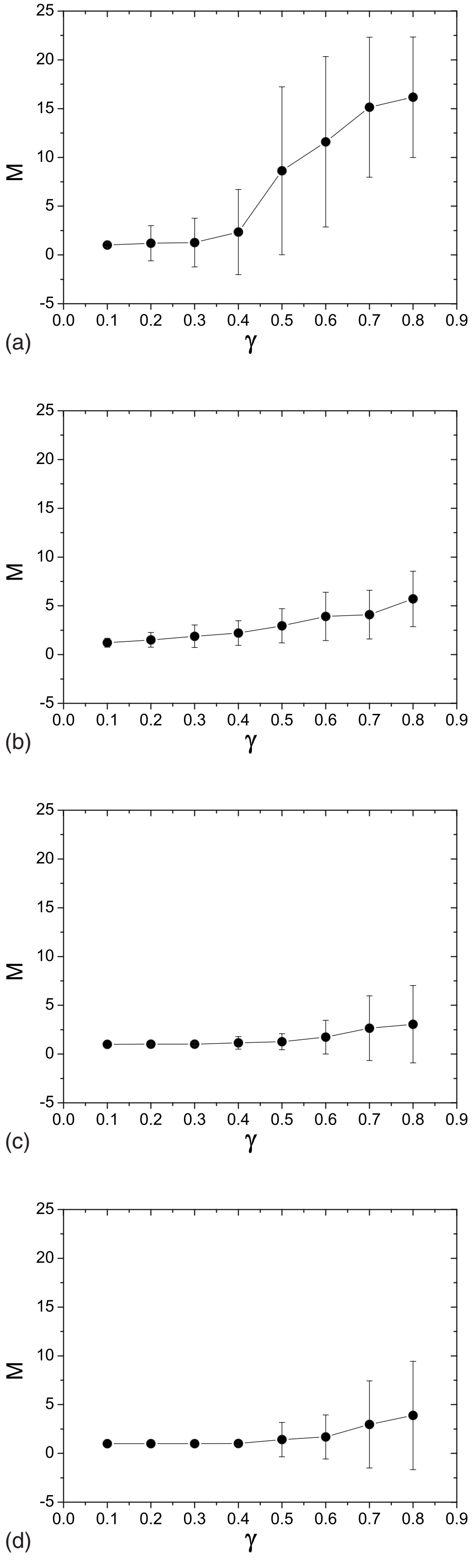

FIG. 7. Average and standard deviation, in terms of γ, of the number M of detected trails corresponding to the minimal overlap error with respect to Poissonian dilation trails obtained for ER (a), WS (b), BA (c), and DMS (d) network models.

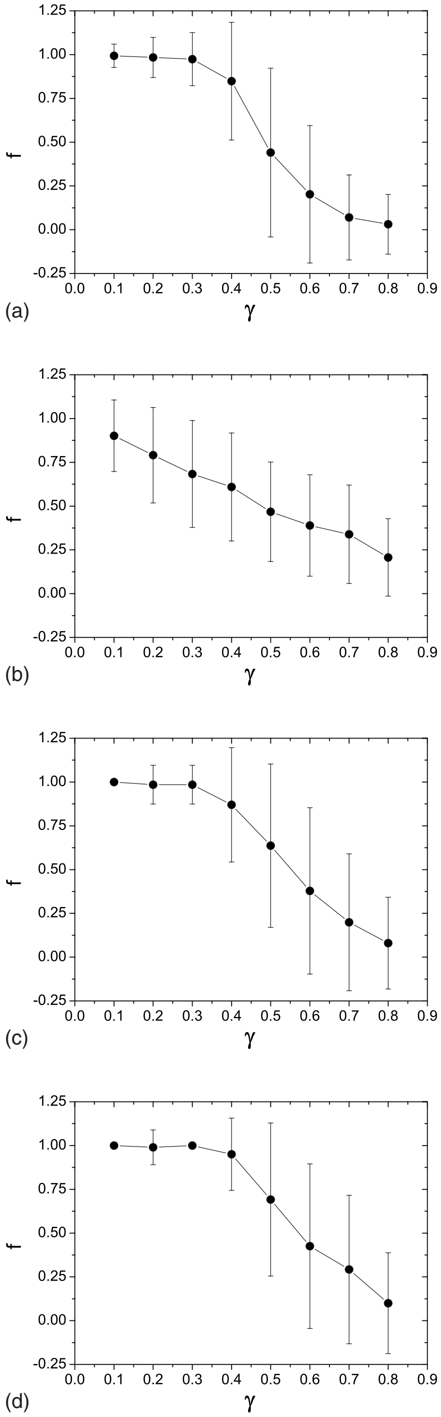

FIG. 8. Average and standard deviation of the flag f indicating that the correct source has been identified among the detected trails with minimal overlap error ξ in the recovery of Poissonian dilation trails for ER (a), WS (b), BA (c), and DMS (d) network models.

cases we consider that each of the possible sources is as good as the other, and therefore can be used as the evaluated source; therefore we make $f=1/M$.

V. SIMULATION RESULTS AND DISCUSSION

To evaluate the recovery strategies under different topologies, randomly generated trails are studied in the ER, WS,

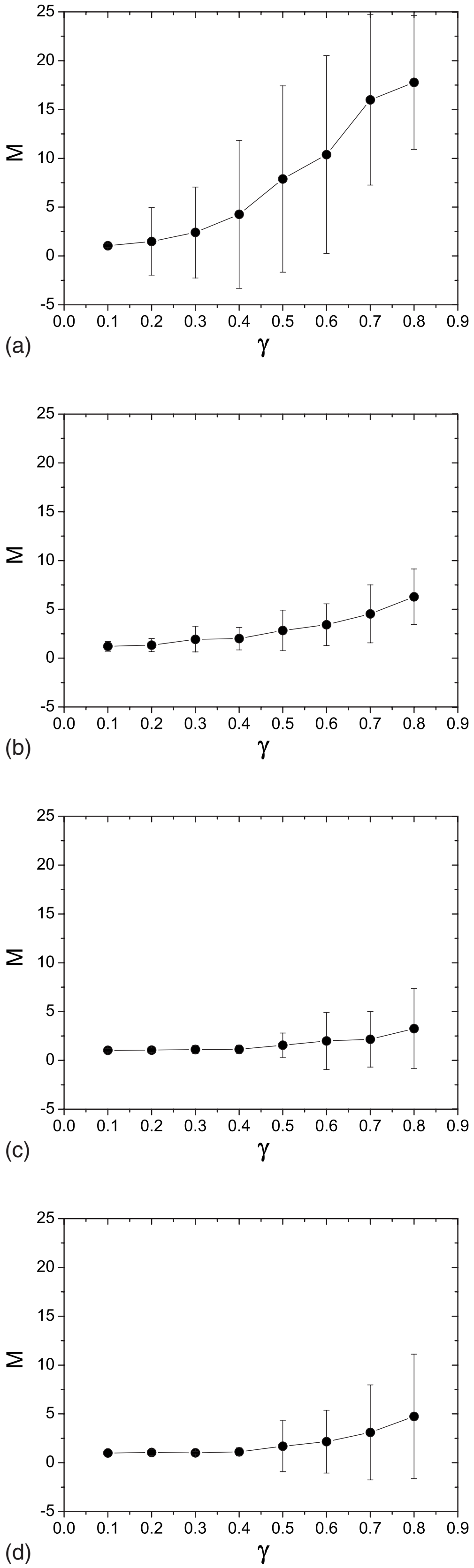

FIG. 9. Average and standard deviation in terms of γ, of the number M of detected evanescent trails corresponding to the minimal overlap error obtained for ER (a), WS (b), BA (c), and DMS (d) network models.

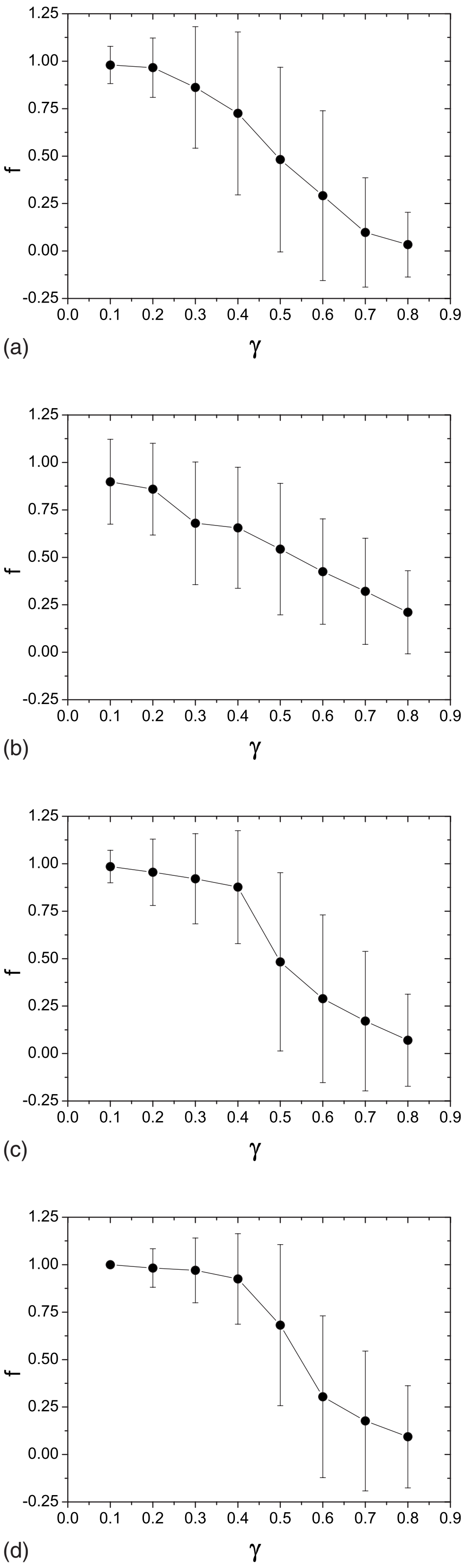

FIG. 10. Average and standard deviation of the flag *f* indicating that the correct source has been identified among the detected evanescent trails with minimal overlap error ξ for ER (a), WS (b), BA (c), and DMS (d) network models.

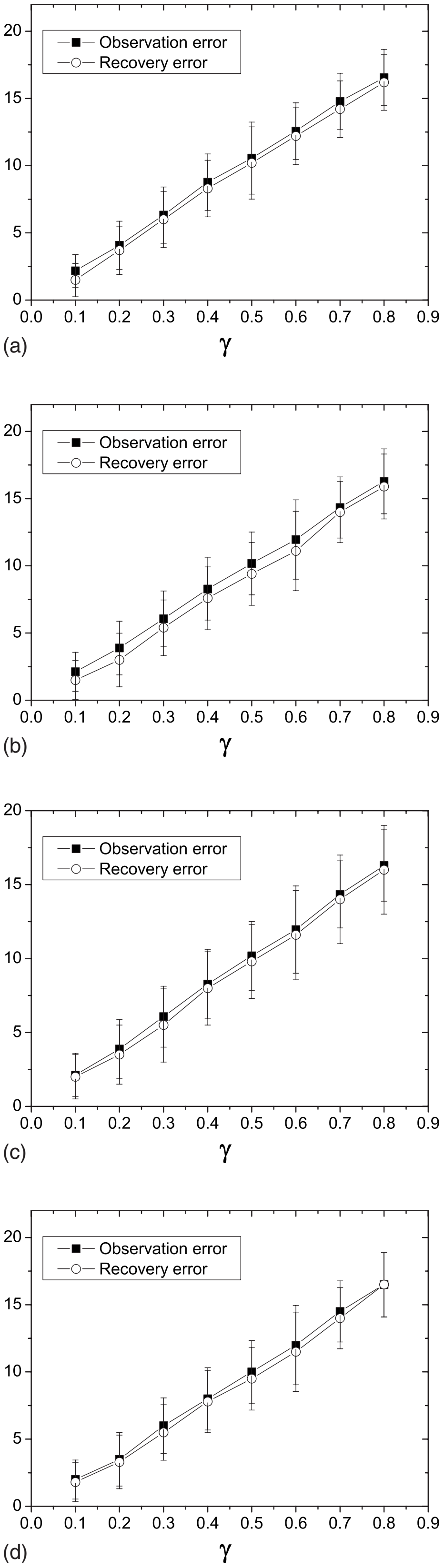

FIG. 11. Observation (black squares) and recovery (white circles) errors obtained by using the recovery algorithm for a Poissonian trail from random walks, for (a) the Internet, (b) the U.S. airlines, (c) the email network from the University Rovira i Virgili, and (d) the scientific collaboration of complex network researchers.

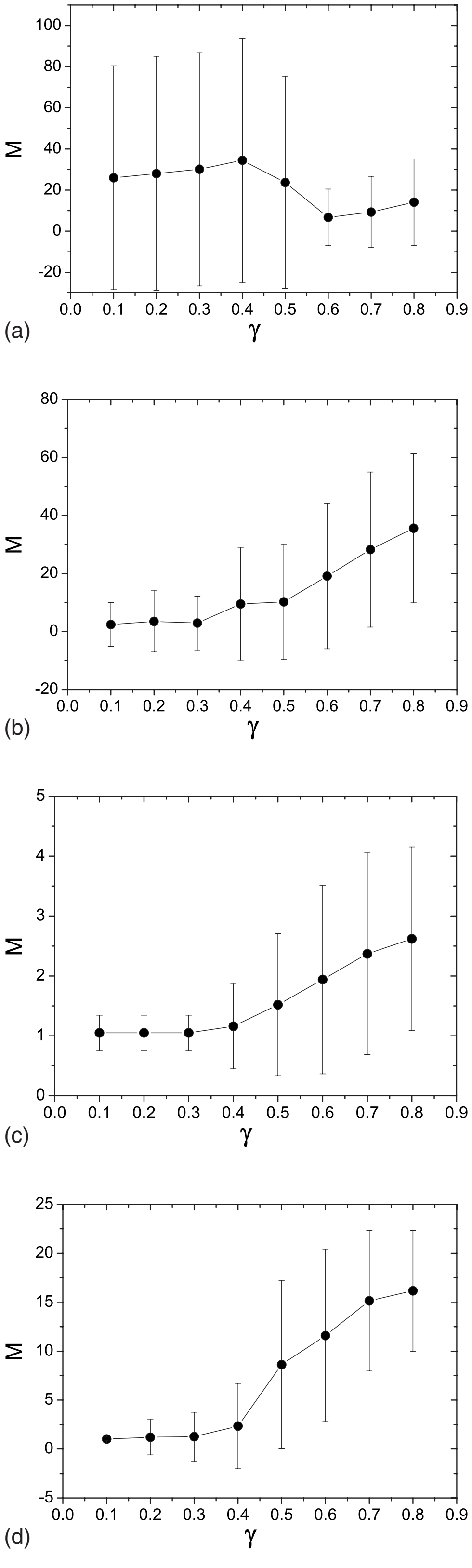

FIG. 12. Average and standard deviation, in terms of γ, of the number M of detected trails corresponding to the minimal overlap error obtained in the case of Poissonian dilation trails for (a) the Internet, (b) the U.S. airlines, (c) the email network from the University Rovira i Virgili, and (d) the scientific collaboration of complex network researchers.

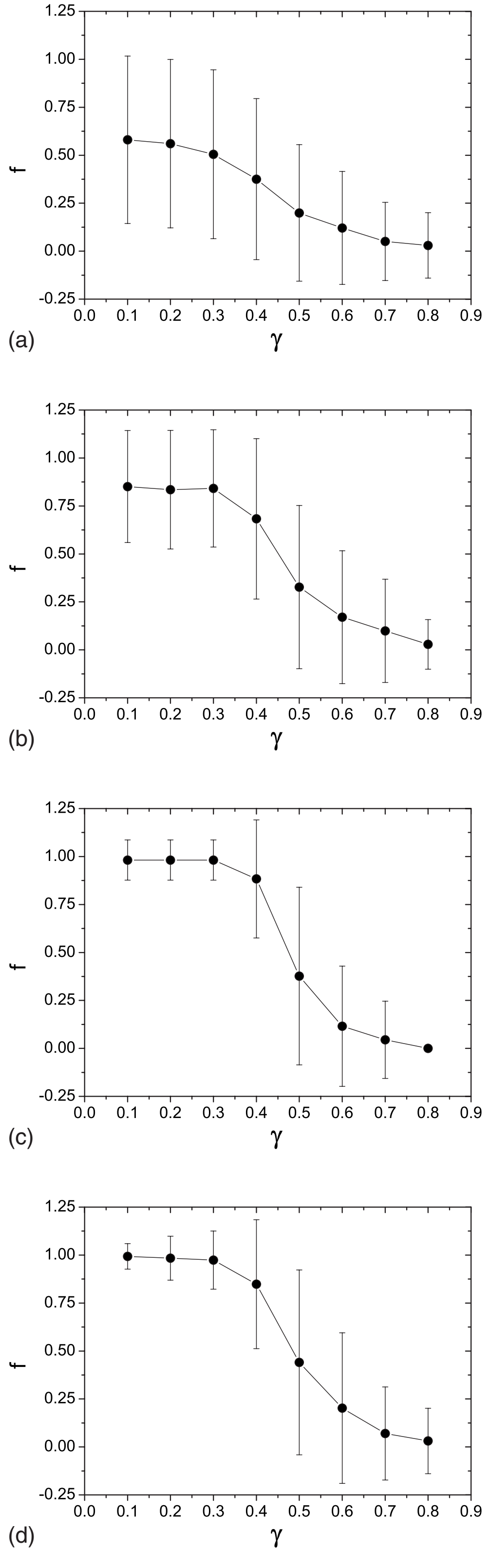

FIG. 13. Average and standard deviation of the flag f indicating that the correct source has been identified among the detected Poissonian trails with minimal overlap error ξ for (a) the Internet, (b) the U.S. airlines, (c) the email network from the University Rovira i Virgili, and (d) the scientific collaboration of complex network researchers.

BA, and DMS network models and the networks of the Internet [autonomous system (AS)], U.S. airlines, email, and scientific collaboration, as indicated previously. The following sections present and discuss those results.

A. Network models

Each considered network model was formed from $N = 1000$ nodes and average degree $\langle k \rangle = 4$. All random walk trails were Poissonian with real extent equal to 20 nodes and $\gamma = 0.1, 0.2, \ldots, 0.8$. All dilation trails took place along two hierarchies, while the corresponding Poissonian and evanescent cases assumed $\gamma = 0.1, 0.2, \ldots, 0.8$. In order to provide statistically significant results, each configuration (i.e., type of network, trail, and γ) was simulated 100 times. The rewiring probability in the WS model is the same as in the ER model, i.e., $p = \langle k \rangle / (N-1)$. The initial connectivity in DMS network models is $k_0 = 5$.

Figure 6 shows the average observation and recovery errors, with respective standard deviations, obtained for the Poissonian random walk trails in the four considered network models. The figure indicates an almost linear increase of the recovery error with γ. Such a monotonic increase is explained by the fact that the higher the value of γ, the more incomplete the observable states become. As the recovery of trails with more gaps will necessarily imply more wrongly recovered patches, the corresponding error therefore will increase with γ. Also, as can be seen by a comparison between observation and recovery errors, the adopted recovery heuristic allowed moderate results for all considered network models, without significant differences among the models, which suggests that this recovery strategy is independent of the network topology.

Figure 7 gives the average and standard deviation of M for Poissonian dilation trails corresponding to the minimal overlap error ξ for the ER, WS, BA, and DMS networks. In all of these models, the average and standard deviation values of M tend to increase with γ, starting at $\langle M \rangle = 1$. This effect is a consequence of the fact that the more sparse the information about the real trail, the more likely it is to cover the observable states s with dilations starting from different nodes. Interestingly, the increase of $\langle M \rangle$ is substantially more accentuated for ER networks, and BA networks are the least subject to source determination ambiguities.

For the Poissonian dilation trails, the average $\langle f \rangle$ (and standard deviation) of the flag f is given in terms of γ in Fig. 8 for the ER, WS, BA, and DMS networks. It is clear from these results that the average number of times, throughout the realizations, in which the correct source is identified among those trails corresponding to the minimal overlap error ξ tends to decrease with increasing γ. This is a direct consequence of the fact that higher values of γ imply substantial distortions to the original trail, ultimately leading to shifts in the identification of the correct source. The behavior of $\langle f \rangle$ is similar for the ER, BA, and DMS network models, with a sharp decrease for $\gamma \gtrsim 0.3$. For WS networks, on the other hand, $\langle f \rangle$ has a smooth decrease. The sources of the trails are best identified for ER, BA, and DMS models when $\gamma \lesssim 0.3$. For higher values of γ, the sources are best identified for WS network models.

Finally, we turn our attention to transient dilation trails of the evanescent category. Recall that in this type of trail only the current position of the trail (i.e., its border) is available to the observer. Figure 9 presents the average and standard deviation of M obtained, in terms of γ, for the ER, WS, BA, and DMS network models. The result is similar to the case of Poissonian trails (Fig. 7), with the recovery strategy having the worst results for ER networks, and similar results among the other models. But for the evanescent trails M grows more gradually than for Poissonian trails.

Figure 10 shows the average and standard deviation of the values of the flag f in terms of γ obtained for the same models. Again, the results are similar to those obtained for the Poissonian trails (Fig. 8), but with a more gradual decrease of f for the ER model.

Remarkably, though retaining less information about the original trail than the Poissonian counterparts, the evanescent trails tend to allow a similar identification of the source of the trail and the original trail.

TABLE I. Statistical measurements for the considered real networks. N is the number of nodes, $\langle k \rangle$ is the average degree, and CC is the average clustering coefficient.

Network	N	$\langle k \rangle$	CC
Internet	3522	3.59	0.19
U.S. airlines network	332	12.81	0.62
Collaboration in science	1589	3.45	0.02
Email network	1133	19.24	0.19

B. Real networks

We considered four different networks in our simulations, namely: the Internet at the level of autonomous systems, the U.S. airlines [21], the email network from the University Rovira i Virgili (Tarragona) [22], and the scientific collaboration of complex networks researchers. Table I presents some information about these networks. All random walk trails were Poissonian with real extent equal to 20 nodes and all dilation trails took place along two hierarchies, with $\gamma = 0.1, 0.2, \ldots, 0.8$. Figure 11 shows average recovery errors obtained for the Poissonian random walk trails in the four considered real networks. Again, as we observed for the network models, the recovery error increases almost linearly with γ, being only slightly smaller than the observation error. The adopted recovery method achieves slightly better results for the U.S. airlines network than for the other networks.

Figure 12 gives the average and standard deviation of M for trails corresponding to the minimal overlap error ξ for Poissonian dilation trials in the considered real networks. The value of $\langle M \rangle$ tends to increase with γ for all networks. For the Internet, $\langle M \rangle$ has two distinct behaviors: (i) for $\gamma \lesssim 0.4$ and $\gamma \gtrsim 0.6$, $\langle M \rangle$ increases slowly; (ii) for $0.4 \lesssim \gamma \lesssim 0.6$, $\langle M \rangle$ decreases; in the region $\gamma \lesssim 0.5$, M has high standard deviations. In the case of the U.S. airlines and the sci-

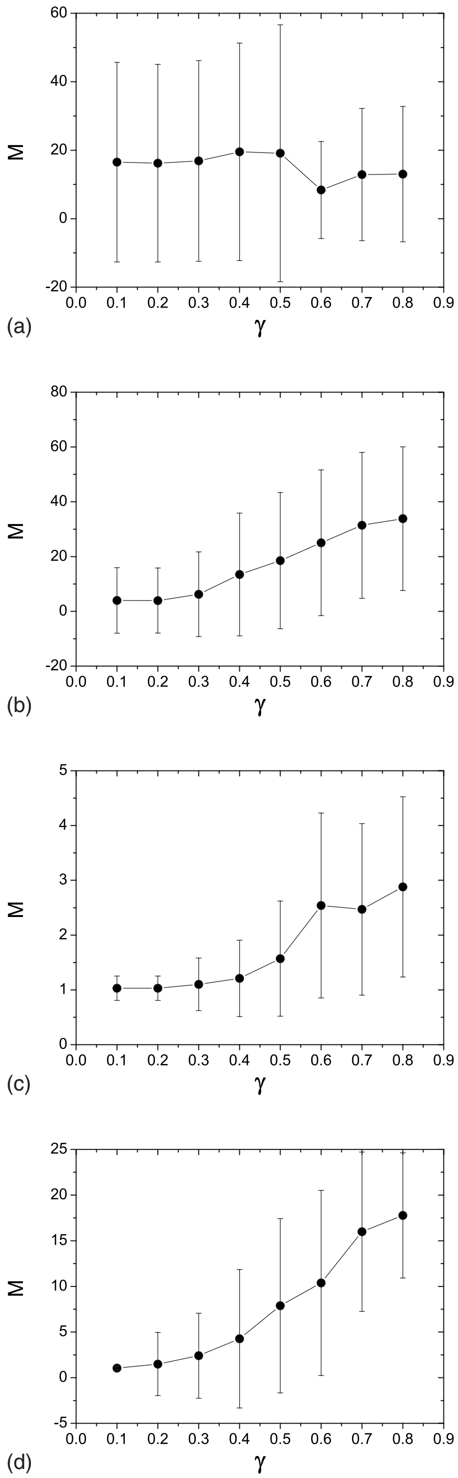

FIG. 14. Average and standard deviation, in terms of γ, of the number M of detected evanescent trails corresponding to the minimal overlap error obtained for (a) the Internet, (b) the U.S. airlines, (c) the email network from the University Rovira i Virgili, and (d) the scientific collaboration of complex network researchers.

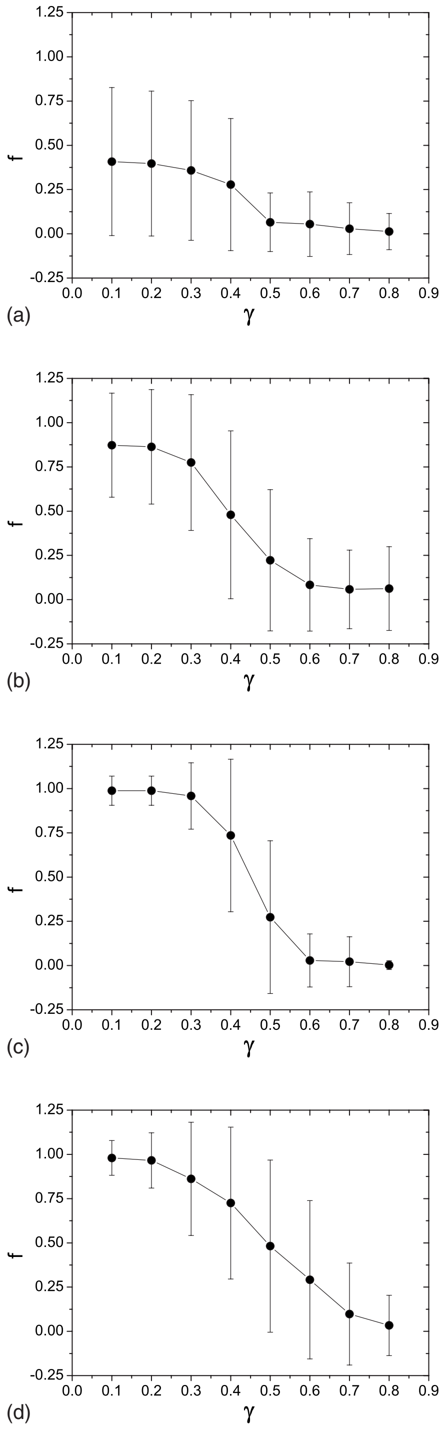

FIG. 15. Average and standard deviation of the flag f indicating that the correct source has been identified among the detected evanescent trails with minimal overlap error ξ for (a) the Internet, (b) the U.S. airlines, (c) the email network from the University Rovira i Virgili, and (d) the scientific collaboration of complex network researchers.

entific collaboration networks, $\langle M \rangle$ has a similar behavior, but has larger values than from the U.S. airlines. The smallest values of $\langle M \rangle$ are obtained for the email network. Therefore, trails can be better recovered in this type of network, which is an important discovery because it has implications for the identification of the source of spreading of virus or rumors, among other cases.

The average $\langle f \rangle$ of the correct source identification flag (and standard deviation) is given in terms of γ in Fig. 13 for

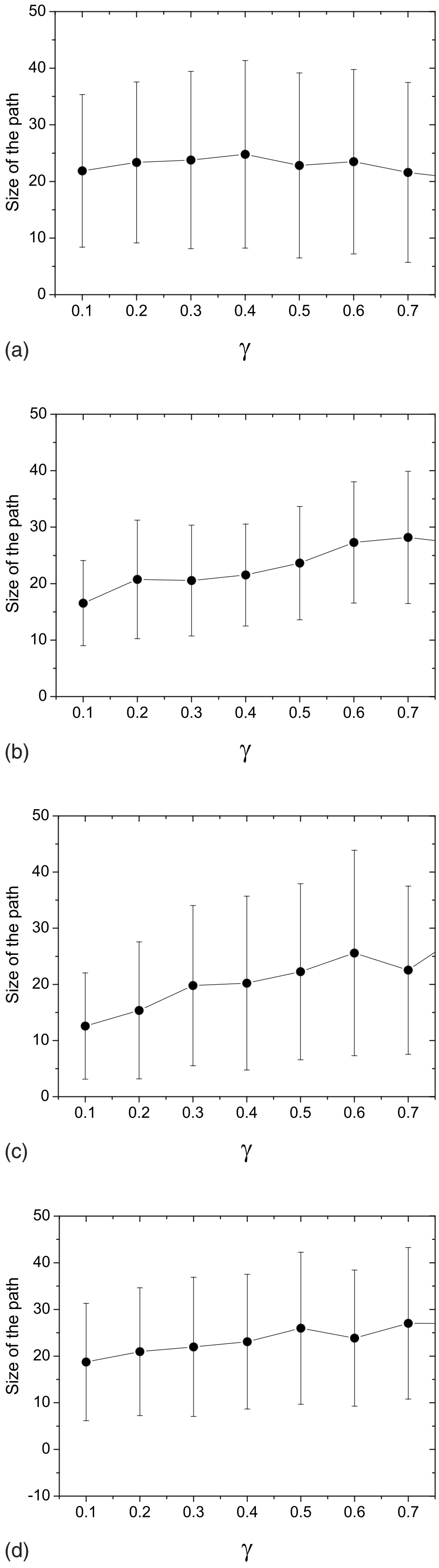

FIG. 16. Average and standard deviation in terms of the length of the path covered by the second agent obtained for ER (a), WS (b), BA (c), and DMS (d) network models. Each point is an average of 500 realizations.

the considered real networks. The source identification is worst for the Internet.

For transient dilation trails of the evanescent category, the results are shown in Fig. 14 (for M) and Fig. 15 (for f). As for the models, the results are close to those obtained considering Poissonian dilation trails, despite the fact that the evanescent category provides less information for trail recovery.

VI. MULTIAGENTS

We considered the dynamics of multiagents in trail evolution using four complex network models: ER, WS, BA, and DMS. Each considered network model is formed from $N=1000$ nodes and average degree $\langle k\rangle=4$. The process is defined as follows: (i) the first agent leaves a gradient trail—the current position has the strongest mark and the source the weakest—by self-avoiding random walks, (ii) the path is erased with a probability γ (Poissonian trail as before), and (iii) the second agent tries to reach the target (the last vertex of the trail) by following preferentially the strongest marks, at each immediate neighborhood, left by the first agent. When the second agent does not find any mark, it performs a random walk until another mark is found. This process is performed, for example, by ants in searching for food—the first agent can represent an ant that leaves a trail of pheromone which will be followed by the second ant. The objective of our investigation is to determine the influence of the topology in target identification efficiency, as well as possible overall trajectory minimization, by measuring the length of the path covered by the second agent. All random walk trails were Poissonian with real extent equal to 20 nodes and $\gamma=0.1,0.2,\ldots,0.8$. Figure 16 presents the length of the path covered by the second agent as a function of the erasure rate γ. As can be clearly seen, when $\gamma<0.5$ the second agent covers the smallest paths for the BA, WS, and DMS network models, followed by the ER. This suggests that the topology of the network is fundamental for trajectory following. Indeed, the hubs present in BA and DMS network models provide shortcuts through the network. Enhanced efficiency was also found for the WS network models, but the high clustering coefficient was identified as being fundamental in this case. While the length of the path followed by the second agent stays almost constant as γ increases for the ER network model, it increases in the other models. For $\gamma>0.5$, the length of the path for the ER network model reaches its smallest value. Therefore, when the trail is almost complete, the BA, WS, and DMS topologies provide the best performances, but when the trail is sparse, ER allows the shortest paths. Thus, it was verified that the topology strongly influences agent dynamics.

VII. CONCLUDING REMARKS

A great part of the interest in complex networks has stemmed from their ability to represent and model intricate natural and human-made structures ranging from the Internet to protein interaction networks. There is a growing interest in the study of dynamics in such systems (e.g., [2,3,24]).

Among the many types of interesting dynamics which can occur on complex networks, we have the evolution of trails left by moving agents during random walks and dilations. In particular, given one of such (possibly incomplete) trails, immediately implied problems involve the recovery of the full trail and the identification of its possible source. Such problems are particularly important because they are directly related to a large number of practical and theoretical situations, including fad and rumor spreading, epidemiology, exploration of new territories, and transmission of messages in communications, among many other possibilities.

The important problem of analyzing trails left in networks by moving agents during random walks and dilations has been formalized and investigated by using two heuristic algorithms in the present paper. We considered four models of complex networks, namely, the Erdős-Rényi, Barabási-Albert, Watts-Strogatz, and Dorogovtsev-Mendes-Samukhin models, and four different real networks: the Internet at the level of autonomous systems, the U.S. airlines, the email network from the University Rovira i Virgili (Tarragona), and the scientific collaboration of complex network researchers. Also, we considered two types of trail: permanent and transient. Particular attention was given to trails with transient marks. In the case of random walk trails, we investigated how incomplete Poissonian trails can be recovered by using a shortest path approach. The recovery and identification of the source of dilation trails was approached by reproducing the dilating process for each of the network nodes and comparing the trails obtained with the observable state variables.

It has been shown through simulation that both such strategies are potentially useful for trail reconstruction and source identification. In addition, a series of interesting results and trends have been identified. First, it was found that the shortest path approach for recovery of trails left by random walks provides similar results for all considered networks and network models, which suggests that this strategy is independent of the network topology. Second, for dilatation trails it was found that the Poissonian and evanescent types of trail allow similar efficiency in the identification of sources, despite the fact that the latter trails incorporate less information than the former.

The analysis of multiagents on networks showed that the topology strongly influences the performance. When the trail is almost complete, the Barabási-Albert, Watts-Strogatz, and Dorogovtsev-Mendes-Samukhin network models provide the best performance. On the other hand, when the information about the trail is sparse, the final point of the trail is reached faster for the Erdős-Rényi network model.

It is believed that the suggested methods and experimental results have paved the way for important related work, including investigation of the scaling of the effects and trends identified in the present work to other network sizes, average node degrees, and network models. At the same time, it would be interesting to consider graded state variables, more than a single trail occurring simultaneously in a network, other types of random walk (e.g., preferential [24]), as well as alternative recovery and source identification strategies. One particularly promising future possibility regards the recovery of diffusive dynamics in complex networks.

ACKNOWLEDGMENTS

L.d.F.C. is grateful to CNPq (Grants No. 308231/03-1 and No. 301303/06-1) and FAPESP (Grant No. 05/00587-5) for financial support. F.A.R. acknowledges FAPESP support (Grant No. 04/00492-1).

[1] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).

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

[3] S. Boccaletti, V. Latora, Y. Moreno, M. Chaves, and D.-U. Hwang, Phys. Rep. **424**, 175 (2006).

[4] L. da Fontoura Costa, F. A. Rodrigues, G. Travieso, and P. R. V. Boas, Adv. Phys. **56**, 167 (2007).

[5] M. Faloutsos, P. Faloutsos, and C. Faloutsos, Comput. Commun. Rev. **29**, 251 (1999).

[6] L. da Fontoura Costa, Phys. Rev. Lett. **93**, 098702 (2004).

[7] L. da Fontoura Costa and F. Silva, J. Stat. Phys. **125**, 841 (2006).

[8] D. Helbing, J. Keltsch, and P. Molnár, Nature (London) **388**, 47 (1997).

[9] E. Bonabeau, M. Dorigo, and G. Theraulaz, *Swarm Intelligence: From Natural to Artificial Systems* (Oxford University Press, London, 1999).

[10] J. Kennedy and R. Eberhart, *Swarm Intelligence* (Morgan Kaufmann, San Francisco, CA, 2001).

[11] D. Helbing, F. Schweitzer, J. Keltsch, and P. Molnár, Phys. Rev. E **56**, 2527 (1997).

[12] M. Dorigo and L. Gambardella, Biosystems **43**, 73 (1997).

[13] M. Dorigo and T. Stützle, *Ant Colony Optimization* (MIT Press, Cambridge, MA, 2004).

[14] M. E. J. Newman and J. Park, Phys. Rev. E **68**, 036122 (2003).

[15] M. Rosvall, A. Trusina, P. Minnhagen, and K. Sneppen, Phys. Rev. Lett. **94**, 028701 (2005).

[16] R. Guimerá and L. Amaral, Eur. Phys. J. B **38**, 381 (2004).

[17] F. A. Rodrigues and L. da F. Costa, Int. J. Mod. Phys. C **16**, 937 (2005).

[18] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).

[19] P. Erdős and A. Rényi, Acta Math. Acad. Sci. Hung. **12**, 261 (1961).

[20] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Phys. Rev. Lett. **85**, 4633 (2000).

[21] V. Batagelj and A. Mrvar, http://vlado.fmf.uni-lj.si/pub/networks/data

[22] R. Guimerà, L. Danon, A. Díaz-Guilera, F. Giralt, and A. Arenas, Phys. Rev. E **68**, 065103(R) (2003).

[23] L. da F. Costa and L. E. C. da Rocha, Eur. Phys. J. B **50**, 237 (2006).

[24] G. Travieso and L. da Fontour Costa, Phys. Rev. E **74**, 036112 (2006).

[25] National Laboratory of Applied Network Research, http://www.nlanr.net