Smart random walkers: the cost of knowing the path

Juan I. Perotti and Orlando V. Billoni

1Facultad de Matemática, Astronomía y Física, Universidad Nacional de Córdoba and
Instituto de Física Enrique Gaviola (IFEG-CONICET), Ciudad Universitaria, 5000 Córdoba, Argentina

(Dated: February 28, 2012)

In this work we study the problem of targeting signals in networks using entropy information measurements to quantify the cost of targeting. We introduce a penalization rule that impose a restriction to the long paths and therefore focus the signal to the target. By this scheme we go continuously from fully random walkers to a walker biased to the target. We found that the optimal degree of penalization is mainly determined by the topology of the network. By mean of the analysis of several examples we have found that a small amount of penalization reduce considerably the typical walk length, and from this we conclude that a network can be efficiently navigated with restricted information.

PACS numbers: 89.70.Hj, 89.70.Cf, 05.40.Fb
Keywords: Searchability, Complex Networks, Random Walk

I. INTRODUCTION

In the problem of targeted signaling or targeted navigability in a network, a message or vehicle begins a journey at a given source vertex with the intention of reaching another target vertex in the most efficient way possible— it is implicitly assumed that the message or vehicle is restricted to jump from vertex to vertex along the edges available in the network. Among the many applications of this area it can be mentioned, the distant communication in complex systems, problems related to the traffic in cities, etc., all this problems make the area an active field of research. The efficiency in solving the problem is measured in terms of a cost which is associated to each possible path the message/vehicle can follow in its journey from the source to the target. Two main issues have to be accounted in defining the cost of targeting, on one side the length of the paths and, on the other side how difficult is to identify a set of convenient paths which connect the source with the target. Once the cost is defined, the efficiency in the task of targeted signaling/navigability can be improved by choosing a convenient searching strategy that minimize the cost.

The most basic strategy for searching the target is the non–biased or fully random walker. The message or vehicle moves randomly without bias through the network with the hope that eventually will reach the target. In this case there is no cost associated in choosing the appropriated paths from the source to the target. Now the problem is focused in finding a reasonable strategy for searching the target, the election of a particular strategy is driven by the minimization of the length of the journey from the source to the target. Among the different strategies one can mention: self–avoiding random walks, intermittent random walks on lattices, the consideration of local topological features, and greedy strategies which are used in the case of networks with spatial embedding. A common feature of all the above mentioned methods is that they use knowledges about the topological structure for the design of the strategies.

A different approach to the problem consist in evaluating the difficulty of choosing the appropriated paths. The difficulty can be quantified in terms of the amount of information or knowledge required to follow these paths. In some cases the information is measured ad–hoc, for example using a fixed information cost per vertex traversed. In general the information required for choosing the right direction in the network depends on the local topological details, for instance; the information required to take the right direction grows with the number of available options. Entropy measures provides a natural way of quantifying information, in fact this measurements have been applied successfully in complex networks before. In particular, these methods were applied recently for quantifying information in the problem of targeted signaling/navigability. This is the approach we adopted in our work to quantify the difficulty of choosing the appropriate paths connecting the source with the target. We will focus on strategies that may be adapted to any topology and in which the message/vehicle is represented by a biased random walker. These strategies will allow us the interpolation between a fully random walker which uses no information to reach its destiny and a directed walker that travels along the shortest paths using all the available information to orient itself. In this regard, there are several antecedents with strategies that interpolate to some extend the random and the biased regime.
In this work we will follow the line of previous works\cite{19,21} in the sense that an information measure is used to regulate how directed the walks are. We extend the ideas of Refs. \cite{22,20} since in these works the information is measured considering only the shortest paths and the optimal penalization strategy for the achievement of relatively short paths is at expenses of limited information\cite{21}. We introduce a formalism for measuring the amount of information used by a biased random walker to reach its target. Using this formalism we develop a method which depends of one parameter that regulates how biased is the random walk. In our method the overall information is increased each time the walker perform a step, so longer walks results in larger penalization. Optimizing the walker’s information forces the walker to travel along increasingly shorter paths or equivalently, the paths are biased to the target.

The paper is organized as follow. In section II we introduce the theoretical background, defining the measures of information used by the random walker in going from the source to the target. We also introduce the penalization rules used to interpolate between the random and the directed regimes, in particular the optimal penalization is defined. In section III we analyze simple examples that can be solved analytically which are useful to understand how the method works in different topological environments, analyzing some limiting cases. In section IV we applied the method by using numerical simulations in more complex networks, like a random network and a Barabási–Albert scale–free network model. Finally, in section V final remarks, conclusions and possible extensions to our work are discussed.

II. THE MODEL

Consider a non–directed network with $N$ vertices and $M$ links where a random walker jumps at a given time step from a vertex $i$ to a neighbor vertex $j$ with probability $q_{ij}$. For each vertex $i$ in the network the transition probabilities $q_{ij}$ satisfy the normalization condition,

$$ \sum_{j \in nn_i} q_{ij} = 1, \quad (1) $$

where $nn_i$ is the set of all nearest neighbors vertices of vertex $i$, and $q_{ij} = 0$ for all $i$, it means the walker is forced to move at each time step. The amount of information given to the walker—for taking an exit from a given vertex $i$ to one of its nearest neighbors—is the information cost defined by\cite{20}

$$ \ln(k_i) - \left[ \sum_{j \in nn_i} q_{ij} \ln q_{ij} \right], \quad (2) $$

which is the difference between the maximum entropy in the space of events of taking one of the exits minus the entropy the exits of vertex $i$ already has associated. Here $k_i$ denotes the degree of vertex $i$. Let us consider now the walker starts its journey at a source vertex named $s$ and ends the trip at a vertex we call the target $t$; furthermore, the walker pass during the journey by the vertex $i$. We want to obtain an expression for the information needed in going from $s$ to $t$ given a distribution of probabilities $q_{ij}$. From the information cost defined above one can derive a recursive expression for the amount of information $S(i \rightarrow t)$ used by the walker in going from vertex $i$ to $t$; accordingly this information cost is expressed as:

$$ S(i \rightarrow t) = \ln(k_i) + \sum_{j \in nn_i} q_{ij} \ln q_{ij} + \sum_{j \in nn_i} q_{ij} S(j \rightarrow t), \quad (3) $$

hence, with the constrain that $S(t \rightarrow t) = 0$, i.e., no information is needed by the walker once the target is reached, a set of linear equations with unknowns $\{S(i \rightarrow t)\}_{i=1,...,N}$ can be defined and solved provided the probabilities $\{q_{ij}\}$ are known. A similar approach was used by Rosvall et al.\cite{20} to quantify the amount of information needed by a walker which is restricted to walk only the shortest paths. In the case that the random walker can step back during the walk the amount of information is:

$$ S_{sp}(s \rightarrow t) = -\ln \left( \sum_{\pi \in \Pi(s,t)} \frac{1}{k_s} \prod_{j \in \pi} \frac{1}{k_j} \right), \quad (4) $$

where $\Pi(s,t)$ denotes the set of all shortest paths $\pi$ between $s$ and $t$, and $\pi$ denotes the set of interior vertices of the shortest path $\pi$.

The minimum for the information $S(s \rightarrow t)$ introduced in Eq. (3), regarding the transition probabilities $q_{ij}$, correspond to a fully random walker with probabilities defined by,

$$ q_{ij}^* = \frac{1}{k_i}, \quad \forall i,j. \quad (5) $$

In this case $S(i \rightarrow t) = 0$ for all $i$. As expected, in finite networks the fully random walker needs no information for reaching the target but has the drawback of very long walks on average. Since we are interested in targeted signaling the results obtained above are of little utility. In order to fix this problem, we will introduce a penalization rule that weights the paths favoring the shortest paths to the target. This penalization will modify the transition probabilities that minimizes the information required to reach the target; the longer walks will be rejected and then a random walker that search the network using this probabilities will be biased to the target.

The simplest way of introducing a penalization is by paying a cost each time the walker pass through a vertex. This information cost is not used by the walker when is travelling the network—unlike the information associated to the $q_{ij}$—but it allows the evaluation of intrinsic properties of the paths to the target, taking into account the whole network. For instance, depending of the degree of
penalization needed for reaching an optimal set of path one can estimate how difficult is the task of finding the paths in a given network. Once the penalization term is introduced in equation (6) it becomes,

\[ F_\gamma(i \to t) = \ln \gamma + \ln(k_i) + \sum_{j \in n_i} q_{ij} \ln q_{ij} + \sum_{j \in n_i} q_{ij} F_\gamma(j \to t), \]

where the term, \( \ln \gamma \), with \( \gamma \geq 1 \) is the penalization term. Now since \( \gamma > 1 \), \( F(s \to t) = 0 \) is not anymore a minimum and hence a random walking does not minimize the information. As will be shown in the next section, minimizing \( F(s \to t) \) with respect to \( \{q_{ij}\} \) keeping \( \gamma \) fixed leads to a biased walk which becomes more directed to its target as \( \gamma \) increases. In fact, the fully random walker correspond to \( \gamma = 1 \) and, in the other extreme when \( \gamma \to \infty \), the walker is forced to walk along the shortest paths. The quantity \( F_\gamma(s \to t) \) stands for the amount of information the walker uses in going from \( s \) to \( t \) plus the \textit{intrinsic} information related to the penalization.

To clarify the role of \( \gamma \) let us introduce a quantity that will allow to define an optimal value for the penalization. First of all, we name by \( \{q_{ij}\} \) the probabilities \( \{q_{ij}\} \) that minimizes \( F_\gamma(s \to t) \) at a given fixed value of \( \gamma \), and \( F^*_\gamma(s \to t) \) the function evaluated at these values, i.e. the minimum. Furthermore, \( S^*(s \to t) \) is the value of \( S(s \to t) \) evaluated on \( \gamma^*_\gamma \). We compute the amount of information introduced by \( \gamma \) as \( F^*_\gamma(s \to t) - S^*(s \to t) \) which is related to intrinsic properties of the network as we mentioned above. Then the relative amount of intrinsic information in going from \( s \) to \( t \) is,

\[ R_\gamma(s \to t) = \frac{F^*_\gamma(s \to t) - S^*(s \to t)}{F^*_\gamma(s \to t)}. \]

The quantity \( R_\gamma(s \to t) \) lies in \([0,1]\) reaching its maximum value 1 when \( \gamma \to 1 \) or \( \gamma \to \infty \). It has a minimum value \( R^*(s \to t) \) at \( \gamma^* \in (1, \infty) \) which we define as the optimal value of \( \gamma \). At \( \gamma^* \) the walker minimizes the relative amount of intrinsic information with respect to the whole information. It means that, up to minimum point, the information the walker gain above the paths to the target is preponderating. An increase of \( \gamma \) further \( \gamma^* \) certainly implies a gain of useful information, but at lower pace than the intrinsic information. Then, the value \( \gamma^* \) gives insights above the searchability of a network in relation with its topology.

III. SIMPLE EXAMPLES SOLVED ANALYTICALLY

To further clarify the formal ideas introduced in the above section let us analyze some simple examples which can be solved analytically. We named each example analyzed in order to facilitate the discussion (see figure 1). In addition, these examples will provide some insights on the problem of targeted delivery of information or navigation. In particular, the last two examples correspond to extreme cases in which remarkable different topological patterns prevails. On the one extreme, the case where only one right path to the target exist; being all the other alternative paths dead ends, and in the other extreme, the case where there are a lot of similar paths to the target—not all of them optimal—with a few shortest paths. The penalization scheme behaves differently in each case serving as an indicator of which kind of topological pattern could prevail on real networks or network models.

A. The Unique–way path

This example is outlined in figure 1(a). For the sake of clarity let us simplify the notation redefining \( F_\gamma(i \to t) \) by \( F_s(S(i \to t) \) by \( S_i \) and \( R_\gamma(i \to t) \) by \( R_i \). In case this set of equations (6) takes the form:

\[ F_s = \ln \gamma + \ln 2p + p \ln (1 - p) \ln (1 - p) + (1 - p) F_i, \]

\[ F_i = \ln \gamma + F_s. \]  

Solving the equation for \( F_s \) and minimizing with respect to \( p \) fixing \( \gamma \) one obtain the following expression for the critical \( p \)

\[ p^* = \frac{1}{2} \frac{2 \gamma^2 - 1}{\gamma^2}. \]  

It is easily to verify that \( p^* \to 1/2 \) when \( \gamma \to 1 \) and that \( p^* \to 1 \) when \( \gamma \to \infty \) therefore one obtain the expected limiting cases. The un- penalized case, \( \gamma \to 1 \), correspond to a fully random walker, and the other case \( \gamma \to \infty \), correspond to a walker fully biased towards the shortest path. One can see that \( p^* \) increases when \( \gamma \) goes from 1 to \( \infty \) indicating that the bias in the walk grows with the penalization \( \gamma \), in other words a larger penalization.
leads to a shorter walk. The relative amount of intrinsic information corresponding to this example,

$$R_s = \frac{(2\gamma^2 + 1) \ln \gamma}{(2\gamma^2 + 1) \ln \gamma - \ln \gamma^2 + (2\gamma^2 - 1) \ln \left(\frac{2\gamma^2 + 1}{2}\right)}$$

is plotted as a function of $\gamma$ in Fig. 2. One can see a minimum which correspond to an optimal penalization $\gamma^*$. Notice that $R_s$ is large even at $\gamma^*$; more than half of the total information $F_s^*$ is due to the information introduced by $\gamma$. The inset shows the information used by the walker to reach the target $S^*$ as function of $\gamma$. It increases as $\gamma$ increases and converges asymptotically to the case of the shortest path $S_{sp} = \ln 2$ as $\gamma \to \infty$.

B. The Diamond

This is a bit more complicated example (1b). In this case the set of equations (9) takes the form:

$$F_i = \ln \gamma + \ln 3 + u \ln u + w \ln w,$$

$$+(1 - u - w) \ln(1 - u - w) + u F_s + w F_i,$$

$$F_s = \ln \gamma + F_i,$$

(11)

symmetry arguments where used here to consider that $F_i = F_j$ and $p = 1/2$. Solving for $F_s$ and minimizing leads to

$$w^* = \frac{1}{3\gamma} \quad \text{and} \quad u^* = \frac{1}{3\gamma^2}.$$  

(12)

These probabilities tend to the right limits $w^*, u^* \to 1/3$ for $\gamma \to 1$ and $w^*, u^* \to 0$ for $\gamma \to \infty$ obtaining the fully random walker and the fully directed walker respectively. Notice that the probability $u$ of step away from $t$ decreases faster with $\gamma$ than the probability $w$ of just taking a longer path. An optimal penalization $\gamma^* \approx 2.18$ exist where $R_s^* \approx 0.77$ which again implies a considerable fraction of intrinsic information. Also in this case the walker’s information about the shorter paths $S^*(s \to t)$ tends asymptotically to the shortest paths information $S_{sp}(s \to t) = \ln 3$ as $\gamma$ increases. It must be remarked that if instead of equation (11) we use the information of the shortest paths where stepping back is avoided, then $S^*(s \to t)$ may become even larger. Anyway, the right magnitude to compare with is that of equation (11).

C. The star web

The star like network (see Fig. 4) represent the extremal case of only one direct path to the target and a large number of dead ends. At variance with the previous examples which contains a fixed number of vertices, this example has no restrictions in the number of vertices, allowing the study of quantities that scales with the network’s size. Like the previous examples, this new one can be analytically solved by using the particular symmetries of this network, whatever the size of the network. Equations (11) in this case reduces to

$$F_i = \ln \gamma + \ln(n + 2) + p \ln p + (1 - p - w) \ln(1 - p - w),$$

$$+w \ln \frac{u}{n} + p F_s + w F_i,$$

$$F_j = \ln \gamma + F_i,$$

$$F_s = \ln \gamma + F_i,$$

(13)

where $F_j := F_{j_1} = ... = F_{j_n}$. The critical probability of $F_s$ is,

$$p^* = \frac{1}{(n + 2)\gamma^2} \quad \text{and} \quad w^* = \frac{n}{(n + 2)\gamma^2},$$

(14)

which satisfy the expected limiting cases: $p^* \to 1/(n + 2)$, $w^* \to n/(n + 2)$ for $\gamma \to 1$ and $p^*, w^* \to 0$ for $\gamma \to \infty$. In this example it is interesting to compute the amount of information related to shortest paths $S_{sp}$ and the optimal information $S^*$ at $\gamma^*$ as function of the number of vertices $n$. The following expression is obtained for the first, $S_{sp}(s \to t) = \ln(n + 2)$, while the second ($S^*$) is obtained numerically. Figure 5 shows these quantities as function of $(n + 2)$ in a linear log plot. One can see that also $S^*$ scales logarithmically with the network’s size, however this amount of information is always smaller than the information related to the shortest path and the difference between them increases with $n$. This implies that, as far as the optimal walks defined by $\{q^*_i\}$ are convenient, it is useful to relax the restriction of walking the shortest paths, because it is cheaper in terms of information to walk along paths which are not so short. Furthermore, the relative amount of intrinsic information $R_s^*$ decreases with the system size (see the inset of figure 5). It means
that the walker’s information about the shorter paths in the network eventually becomes predominant. This is also consistent with the decreasing in the value of optimal penalization $\gamma^*$ as function of $n$, which is required to learn the shorter paths (see the inset). This indicates that in this topology the walker can learn efficiently the ways to the target.

D. The Equivalent paths

This example (see figure 3) allows to visualize one of the main motivations to generalize the approach which measures the information considering only shortest paths $S_{sp}$ to a measure that includes all the possible paths. Specifically, this example allows the study of a case in which the walker has many alternatives consisting of equivalent paths, that are not much longer than the shortest ones. Due to its particular topology this example can also be solved analytically for arbitrary network’s sizes. Accordingly, applying Eq. 3 to this particular example the following set of equations turns out,

\[
F_s = \ln \gamma + \ln (n+1) + p \ln p + (1-p) \ln \left(\frac{1-p}{n}\right) + (1-p) F_i
\]
\[
F_i = \ln \gamma + \ln 2 + u \ln u + (1-u) \ln (1-u) + u F_a,
\]

where $F_i := F_{i_1} = \ldots = F_{i_n}$. As usual we solve for $F_s$ and minimize, leading to

\[
p^* = \frac{2\gamma^2(n+1) - n}{(\gamma(n+1)+1) \left(n + \frac{2\gamma^2(n+1) - n}{\gamma(n+1)+1}\right)}
\]

which satisfies the right limits $p^* \rightarrow 1/(n+1)$, $u^* \rightarrow 1/2$ for $\gamma \rightarrow 1$ and $p^* \rightarrow 1$, $u^* \rightarrow 0$ for $\gamma \rightarrow \infty$. Like in the Star web, in this example the optimal information $S^*$ and the shortest paths information $S_{sp}$ scales logarithmically with the network’s size (see figure 4), and also $S_{sp} > S^*$ but at variance with Star web the difference between them remains almost constant. Here, the optimal penalization $\gamma^*$ grows with system size $n$ and the relative amount of intrinsic information $R^*_s \rightarrow 0.5$ for all the sizes (red dashed line). The orange dotted line indicates $R^*_s = 0.5$.

\[
u^* = 1 - \frac{1}{2\gamma(n+1)+1} \frac{2\gamma^2(n+1) - n}{\gamma(n+1)+1}
\]
IV. NUMERICAL SIMULATIONS RESULTS

In this section we apply the ideas introduced in the previous sections to more complex network topologies. We choose two paradigmatic cases such as the random and scale-free networks. Since these systems cannot be solved analytically, all the results we show here are obtained by numerical simulations. We performed an optimization procedure minimizing $F_{\gamma}$ with respect to $\{q_{ij}\}$ for a sequence $\gamma = \delta, 2\delta, 3\delta, ...$ where $\delta$ is a small quantity ($\delta \in [0.005, 0.05]$). This is a convenient procedure since $\{q_{ij}^*\}$ varies smoothly with $\gamma$. We start using the values given by equation (4) as the initial guess for $\gamma = \delta$, and then we use the last minimum obtained for the subsequent values of $\gamma$. We performed the minimization using the implementation of the SLSQP algorithm provided by SciPy as a part of the Sage Mathematics Software. If $\delta$ is too large the minimization algorithm fails to converge since the initial guess is too far away from the minimum, even in small networks. The average walk length is obtained by a Monte Carlo procedure using the set $\{q_{ij}^*\}$ for the transition probabilities, this quantity is analyzed as a function of the penalization and the shortest path length between the sources and the target.

The values for the probabilities $\{q_{ij}^*\}$ depend on the target vertex $t$ but are independent of the source vertex $s$ for each value of $\gamma$. We tested this analytically in the examples of figure 1 and numerically on a small random network. Let’s define the vector $\vec{q}^*$ whose components are the no null values of $\{q_{ij}^*\}$. In figure 2 we plot the ratio between the dispersion of $\vec{q}^*$ with respect to $s$, $D(\vec{q}^*) = 1/(n-2)\sum_{s \neq t} |q_s^2(s) - \langle q_s^2 \rangle|^2$, and the norm, $\langle q_s^2 \rangle = 1/(n-1)\sum_{s \neq t} q_s^2(s)$, as a function of $\gamma$ and for different targets $t$. It is shown that this ratio is much smaller than unity confirming the independency of $\{q_{ij}^*\}$ with respect to the source.

Since information is an additive quantity, the relative amount of intrinsic information for the overall network can be defined, given target $t$. It considers all the possible sources and hence the paths to the target $t$. Consequently we have,

$$R_{\gamma}(t) = \frac{\sum_{i \neq t} F_{\gamma}(i \rightarrow t) - S(i \rightarrow t)}{\sum_{i \neq t} F_{\gamma}(i \rightarrow t)},$$

then the overall optimal penalization $\gamma_t^*$ associated to the target $t$ can be obtained from this expression. For the sake of brevity let’s omit the reference to $t$ so we will refer to $R_{\gamma}(t)$ by $R_{\gamma}$ and its optimal version $R^*(t)$ by $R^*$.

A. Random networks

Let us first analyze the case of a random network, all the calculations in this section were performed using a network of $N = 100$ vertices with an average degree $\langle k \rangle = 3$ and using a target chosen at random. Care was taken to obtain a random network that consist of only one connected component. Here we show results corresponding to a single realization for the target and the network, since similar results were obtained using different realizations. Figure 3 shows $R_{\gamma}$ as function of $\gamma$ obtained for this network. From this curve we obtained the optimal overall penalization $\gamma^*_t \simeq 1.105$. In this figure we also show $\langle R \rangle_L$, which is the relative information average over different sources that are at a fixed distance $L$ from the target. We obtained from these curves the optimal penalization $\gamma^*_L$; the dependence of $\gamma^*_L$ on $L$ is shown in the inset. We observe that $\gamma^*_L$ varies with $L$ but in every case $\gamma^*_L$ is of the same order of magnitude that $\gamma^*_t$.

In order to analyze the role of the penalization in the restricted walks toward the target, we analyze typical walk lengths as function of $\gamma$. We first obtain the transition probabilities associated to the target $t$, for a given $\gamma$, and then using this probabilities we implement a Monte Carlo process for obtaining a set of trajectories corresponding to random walkers which are biased to the target. The random walkers start its journey at every possible source available in the network. Averaging over these trajectories we obtain the average walk length $\langle wl \rangle$. In figure 4 we plot $\langle wl \rangle$ corresponding to the same realization for the network and target we used for obtaining the results of figure 3. When $\gamma$ approaches its minimal physical value $\gamma = 1$ the walk is fully random and the average walk length between all the possible sources and the fixed target is $\langle wl \rangle \simeq 260$ which is at least an order of magnitude larger than the typical distance between vertices. Then as $\gamma$ is increased the average walk length decreases drastically. We denote by $\langle L \rangle_t$ the averaging of all the shortest path that reach the target (note that in this calculation all sources are included). At the optimal penalization $\gamma^*_t$ the difference between the average walk
length and the average shortest paths $\langle L \rangle_t$ is nearly two times $\langle L \rangle_t$, i.e. the average walk length is of the same order of magnitude than the average of the shortest paths. We also computed at the optimal penalization the average walk length $\langle wL \rangle_L$ that correspond to averages in trajectories restricted to start at sources that are at a distance $L$ from the target. We plotted this quantity as function $L$ in inset of Fig. 7. The average walk length $\langle wL \rangle_L$ increases linearly with the shortest path length $L$, but the relative excess $(\langle wL \rangle_L - L)/L$ is almost constant taking a value close to 2 (see inset of figure 7). In order to explore the dependency of the walker’s information $S$ on the penalization we plot in figure 8 the average, $\langle S \rangle$, over all the sources as a function of $\gamma$. It can be seen that $\langle S \rangle$ grows with $\gamma$ but it is always much smaller than the averaged shortest path information $\langle S_{sp} \rangle$, in particular at the overall optimal penalization $\gamma^*$. In the present approach the walker uses less information than in the shortest paths approach $S_{sp}$ but there is a price to pay for; the path length to the target $\langle wL \rangle$ is longer than the shortest path $L$. In addition, according to the current approach the amount of information the walker learns on average $\langle S \rangle_L$ does not depend on $L$ (inset in figure 8) as far as $L$ is greater than 3. One can think that there is a distance horizon ($L = 3$) that define two regimes. At short distances the walker can improve its information about the paths to the target as the distance grows. Whereas for targets far away the sources the amount of information cannot be improved. In other words, the searchability at short distance is favored.

B. Scale–free networks

Scale free networks are characterized by a power law distributions in the connectivity of the vertices and even small networks shows the presence of highly connected vertices, when compared to the mean value of their connectivity. Therefore, although in our case the size of the network is small, a scale-free topology will allow us the study of how the different quantities are affected by
the vertex’s degree. We performed the calculations on a Barabási–Albert network model\(^1\) with \(N = 100\) vertices and \((k) \simeq 4\). Figure 9 shows that the walk length \(<wl>\) and the shortest path length \(<L>\) decreases with target’s degree and the decrease of the former is more pronounced. The hubs can be found easily having a walk length much closer to the shortest paths, whereas in poorly connected vertices significantly longer walks are required. According to the present approach the hubs are favored, both regarding the number of steps and regarding the information that is needed; as far as the optimal condition is easily achievable. Like the case of the random network, in the scale–free networks the walker’s information \(<S>\) (full magenta circles) and the shortest path information \(<S_{sp}>\) (full cyan squares).

This implies that it is more expensive to find the optimal walking pattern in highly connected targets.

V. DISCUSSION AND CONCLUSIONS

In this work we have introduced an approach for measuring the amount of information used by a biased random walker that moves to a target. In this framework we extend the ideas of Rosvall et al.\(^3\) because we consider not only the shortest paths but all the possibilities paths to the target. Based in this approach we propose a penalization rule, depending on one parameter, that bias a random walker to the target, provided the walker can use less information as possible. The basic idea was that each step that makes the walker is penalized, hence this leads to an overall penalization that tends to reduce the walk lengths. Our approach is consistent since the two main quantities that determines the cost associated to the task of targeted signaling are counterbalanced; a shortening of the walk length through penalization implies an increase of the information required, and vice-versa. At this point, it is important to stress that in this scheme the penalization operates globally limiting the overall available information, unlike other approaches\(^2\) in which the information at each vertex is limited. This has the advantage of overcoming the undesired effect of affecting mostly the highly connected vertices, provided the vertex’s degrees are taken into account.\(^3\)

We also introduced the idea of intrinsic information, in order to define an optimal penalization. We have shown through some network models that in practice a small amount of penalization is enough to drastically reduce the typical walk length, and then a network can be efficiently navigated with restricted information. On the other hand, once the optimal penalization is reached it is highly expensive to further reduce the typical walk length, in particular an infinite penalization is required to restrict the path lengths to the shortest ones. The typical walk length in a random network was analyzed and compared to the corresponding shortest path at the optimal penalization. It is found that the difference between these lengths grows linearly with the shortest path length. This is connected with a trade–off at which the amount of information does not increases with the length of the shortest paths. In addition, from the trend of \(<S>_{L}\) a distance horizon can be identified which define a range

\[\langle R^* \rangle_{k}, \langle S^*_{sp} \rangle_{k} \]

\[\langle S \rangle \approx \langle S_{sp} \rangle \]

\[\langle R \rangle \approx \langle R_{sp} \rangle \]

\[\gamma \]

\[\langle R \rangle = \langle R_{sp} \rangle \]

\[\gamma^* \]

\[\langle S \rangle \approx \langle S_{sp} \rangle \]

\[\langle R^* \rangle \approx \langle R_{sp} \rangle \]

\[\gamma^* \]

\[\langle S \rangle \approx \langle S_{sp} \rangle \]

\[\langle R \rangle \approx \langle R_{sp} \rangle \]

\[\gamma^* \]

\[\langle S \rangle \approx \langle S_{sp} \rangle \]

\[\langle R \rangle \approx \langle R_{sp} \rangle \]

\[\gamma^* \]
of efficient searchability.

The ideas introduced in this paper were applied to undirected and unweighted networks. But, for instance, in the study of traffic on cities directed and weighted network are needed, since streets have different capacities and directions. The extension of the formalism to include directed networks is straightforward, but care must be taken to ensure that each vertex is accessible from each other vertex, otherwise the analysis have to be restricted to each strongly connected components of the network. Also the penalization scheme may be generalized to be vertex dependent. In the case of the dual representation of network’s cities where vertices are streets and edges are road intersections, the traffic congestion on each street can be used to regulate the amount of penalization in order to avoid a traffic jam. Although the proposed formalism requires the solution of an optimization problem, that has a large computational cost in a serial implementation, the algorithm is specially suited for a parallel implementation since each target \( t \) can be treated separately.

The authors thank Sergio Cannas and Daniel Stariolo for useful discussions. This work was partially supported by grants from CONICET (Argentina), Agencia Córdoba Ciencia (Argentina), SeCyT, Universidad Nacional de Córdoba (Argentina).

\* Electronic address: perotti@famaf.unc.edu.ar
\† Electronic address: billoni@famaf.unc.edu.ar

1. K. Sneppen, A. Trusina, and M. Rosvall, Europhys. Lett. 69, 853 (2005).
2. M. Rosvall, A. Trusina, P. Minnhagen, and K. Sneppen, Phys. Rev. Lett. 94, 028701 (2005).
3. S. Scellato, L. Fortuna, M. Frasca, J. Gómez-Gardeñes, and V. Latora, Euro. Phys. J. B 73, 303 (2010).
4. J. D. Noh and H. Rieger, Phys. Rev. Lett. 92, 118701 (2004).
5. L. d. F. Costa and G. Travieso, Phys. Rev. E 75, 016102 (2007).
6. Y. Moreno and A. Vázquez, Eur. Phys. J. B 31, 265 (2003).
7. J. Balthrop, S. Forrest, M. E. J. Newman, and M. M. Williamson, Science 304, 527 (2004).
8. S.-J. Yang, Phys. Rev. E 71, 016107 (2005).
9. G. Oshanin, K. Lindenberg, and H. Wio, Journal of Physics A 42, 434008 (2009).
10. L. A. Adamic, R. M. Lukose, A. R. Puniyani, and B. A. Huberman, Phys. Rev. E 64, 046135 (2001).
11. J. M. Kleinberg, Nature 406, 845 (2000). ISSN 0028-0836.
12. A. P. S. de Moura, A. E. Motter, and C. Grebogi, Phys. Rev. E 68, 036106 (2003).
13. D. O. Cajueiro, Phys. Rev. E 79 (2009).
14. D. O. Cajueiro, Physica A 389, 1945 (2010).
15. R. Solé and S. Valverde, in Complex Networks, edited by E. Ben-Naim, H. Frauenfelder, and Z. Toroczkai (Springer Berlin / Heidelberg, 2004), vol. 650 of Lecture Notes in Physics, pp. 189–207.
16. L. Demetrius and T. Manke, Physica A 346, 682 (2005).
17. G. Bianconi, Phys. Rev. E 79, 036114 (2009).
18. Y. Lin, K. C. Desouza, and S. Roy, Applied Mathematics and Computation 216, 2824 (2010).
19. E. Capobianco, Journal of Computational Science 2, 144 (2011).
20. M. Rosvall, A. Grönlund, P. Minnhagen, and K. Sneppen, Phys. Rev. E 72 (2005).
21. M. Rosvall, P. Minnhagen, and K. Sneppen, Phys. Rev. E 71 (2005).
22. V. Sood and P. Grassberger, Phys. Rev. Lett. 99, 098701 (2007).
23. A. Fronczak and P. Fronczak, Phys. Rev. E 80, 016107 (2009).
24. D. Kraft, Technical Report DFVLR-FB 88-28, Institut für Dynamik der Flugsysteme, Oberpfaffenhofen (1988).
25. D. Kraft, ACM Trans. Math. Softw. 20, 262 (1994).
26. E. Jones, T. Oliphant, P. Peterson, et al., SciPy: Open source scientific tools for Python (2001–), http://www.scipy.org/.
27. W. Stein et al., Sage Mathematics Software (Version 4.6.2), The Sage Development Team (2011), http://www.sagemath.org.
28. M. Rosvall and K. Sneppen, International Journal of Bifurcation and Chaos 17, 2509 (2007).
29. A.-L. Barabási and R. Albert, 286, 509 (1999).