Random Walk with Jumps in large-scale Geometric Random Graphs ⋆

Leonidas Tzevelekas [a,∗], Konstantinos Oikonomou [b]
Ioannis Stavrakakis [a]

[a] *National & Kapodistrian University of Athens, Department of Informatics & Telecommunications, Athens, Greece*

[b] *Ionian University, Department of Informatics, Corfu, Greece*

Abstract

The *information dissemination* problem in large scale network environments like wireless sensor and ad hoc networks is studied here considering geometric random graphs and random walk-based approaches. The traditional random walk case is studied and an analytical expression with respect to *coverage* (i.e., the proportion of the network nodes visited by a random walk *agent*) as a function of the number of the agent movements is derived. It is observed that the cover time of the RW agent is large in random geometric graphs of low degree (as it is commonly the case is wireless environments). As this inefficiency is attributed (as discussed in the paper) to the inability of the RW agent to move away from already likely covered areas, a mechanism for directional movement (i.e. jumping) of the agent is proposed and studied, that allows the agent to jump to different network areas, most likely not covered yet. The proposed mechanism is studied analytically and via simulations and the parameters (of the network topology and the mechanism) under which the proposed scheme outperforms the RW are determined.

Key words: RW Agent, J-RW Agent, Cover, Partial Cover, Random Geometric Graphs.

⋆ This work has been supported in part by the project ANA (Autonomic Network Architecture) (IST-27489), the PENED 2003 program of the General Secretariat for Research and Technology (GSRT) co-financed by the European Social Funds (75%) and by national sources (25%) and the NoE CONTENT (IST-384239).
∗ Corresponding Author. Address: National & Kapodistrian University, Department of Informatics & Telecommunications, Ilissia, 15 784 Athens, Greece. Tel: +30 210 727 5341, Fax: +30 210 727 5333.

Email addresses: ltzev@di.uoa.gr (Leonidas Tzevelekas), okon@ionio.gr (Konstantinos Oikonomou), ioannis@di.uoa.gr (Ioannis Stavrakakis).

1 Introduction

One of the main challenges associated with large-scale, unstructured and dynamic networking environments is that of *efficiently* reaching out to all or a portion of the network nodes (i.e., *disseminating information*) in order to provide, e.g., software updates or announcements of new services or queries. The high dynamicity and the sheer size of such networking topologies ask for the adoption of decentralized approaches to information dissemination [10], [11], [12], [13]. In this paper the problem of efficiently disseminating information (or queries) across a large-scale, resource-limited, ad-hoc-structured wireless network, such as wireless sensor network is considered. One of the simplest approaches employed for disseminating information in such environments, is the traditional *flooding* approach. Under flooding ([1], [2], [3], [5]), each time a node receives a message for the first time from some node, it forwards it to all its neighbors except from that node. Despite its simplicity and speed (typically achieving the shortest cover time possible, upper bounded by the network diameter), the associated large message overhead is a major drawback.

As flooding is considered not to be an option for large scale, wireless networking environments due to strict energy limitations of individual sensor nodes, approaches based on random walks are viewed as reasonable choices [14] [4], [15], [16]. Random walks possess several good characteristics such as simplicity, robustness against dynamic failures or changes to the network topology, and lack of need for knowledge of the network physical and topological characteristics. The *Random Walk agent* (RW agent) employed within a network of wireless sensors moves from neighbor node to neighbor node in a random manner, frequently revisiting previously covered nodes in a circular manner, even without backtracking (returning to the node it just came from is not allowed); these revisits constitute overhead and impact negatively on the cov er time [8]. Such a poor behavior of the RW agent is attributed to the pure random manner of its movement, combined with some problematic topological characteristics of large scale wireless ad hoc networks, such as cliques and bottlenecks.

Large-scale geometric random graphs have been studied in the past in relationship with percolation theory, statistical physics and hypothesis testing [19]. Recently, the $G(N, r_c)$ random geometric graph has received significant attention due to its applicability in modelling wireless ad hoc and sensor networks, where N is the number of network nodes and r_c is the *connectivity radius*. The network connectivity of random geometric graphs depends on (a) the connectivity radius r_c; and (b) the geometric position of nodes. In particular, any nodes having geometric (euclidean) distance below the connectivity radius r_c are considered to be bi-directionally connected. Naturally, connectivity radius

r_c should be large enough, such that the network is connected, i.e. there are no isolated nodes within the network. Such a network connectivity model is substantially different from the classic Erdos-Renyi network model for Internet topologies [18], where any two nodes have a fixed non-zero probability of being connected (power-law graphs). Existence of long-haul links (links connecting nodes residing far apart in physical distance), although perfectly valid in Erdos-Renyi graphs, do not appear in random geometric graphs, due to physical limitations associated with the connectivity radius r_c.

In this paper the Jumping Random Walk (J-RW) mechanism is proposed as an efficient alternative against the RW for information dissemination/ retrieval in large scale environments, like wireless sensor networks. The proposed scheme exploits the benefits of the RW mechanism (simple, decentralized, robust to topology changes) while providing a 'boost' in performance, i.e. accelerating the coverage process within the network. The latter is achieved by introducing a second state of operation to the RW agent in which the random movement paradigm is replaced by a non-random "directional" movement paradigm. It turns out that this improves significantly the cover time by "creating" long links in topologies that lack them. It should be noted that the RW agent corresponds to a special case (parameter setting) of the proposed J-RW agent.

2 The RW Agent in various topologies

A credible alternative to flooding for disseminating information in an unstructured environment, is the RW agent. In RW-based approaches, the initiator node employs an agent that will move randomly in the network, one hop/ node per time slot, informing (or querying) all the nodes in its path. In particular, large-scale, highly decentralized networking topologies like peer to peer (P2P) or wireless sensor networks have employed RW-based information dissemination/ data quering approaches. Authors in [6] proposed a number of algorithms for RW-based searching in unstructured P2P networks, whereas probability-based information dissemination has been investigated for use in sensor networks [12], with data routing as the main consideration. Random walk in large-scale P2P nets has been shown to possess a number of good properties for searching and/or distributing of information within the network [7].

The overhead of RW-based solutions is considered to be much smaller than that of the flooding approaches, at the expense of a significant increase of cover time. *Cover time* is the expected time taken by a random walk to visit all nodes of a network. The generally relatively large (compared to flooding) cover time achieved under RW-based approaches depends on the network topology. For instance, it is $O(N\ln(N))$ for the fully connected graph (best-case scenario) and $O(N^3)$ for clique topologies (worst case scenario) [8], [9]. Random walks on random geometric graphs $G(N, r_c)$ have been shown to have optimal cover time $\Theta(N \ln(N))$ and optimal partial cover time $\Theta(N)$ with high propability given that the connectivity radius of each node r_c fullfills a certain threshold property, i.e. given that $r_c^2 \geq \frac{c8\ln(N)}{N}$ [17]. Generally, it has been shown that cover time is lower for high connectivity network topologies, such as complete graphs, and it is higher in network topologies presenting bottlenecks. In the latter case, the number of revisits of already covered nodes (which affect the induced overhead and cover time) becomes particularly high.

According to the typical RW-based information dissemination paradigm, each node receiving the RW agent or packet, chooses a forwarding neighbor arbitrarily based on the uniform distribution. For a network modelled as a graph $G(V, E)$ (where V and E denote its vertices and edges, respectively) a node $v \in V$ with *connectivity degree* $\delta(v)$ chooses each one of its next hop neighbors $u \in V$ with probability $p_u = 1/\delta(v)$ if $(v, u) \in E$ and $p_u = 0$ for all other $u \in V$. An important (for this paper) variant of RW-based propagation of information is the RW *without backtracking*. Under this scheme the RW visiting a node $v \in V$ will choose the next hop node $u \in V$ arbitrarily among the neighbors of node v with probability $p_u = 1/(\delta(v) - 1)$ if $(v, u) \in E$, where u is any neighbor of v except the one from last hop, and $p_u = 0$ for all other $u \in V$.

For the rest of this paper it is assumed that $G(V, E)$ is a connected network. Let N be the number of network nodes (equivalently, the size of set V). In such a network, a RW agent moving according to the previously described mechanism, will eventually visit or *cover* all network nodes after some time (cover time). Let $C_r(t)$ be the fraction of network nodes covered (or visited) after t time units or movements of the RW agent (i.e., the RW agent start moving at $t = 0$), for a particular realization (sample path) of the walk and for a given intitiator node. $C_r(t)$ will be referred to hereafter as the *coverage* at time t. Clearly, $C_r(t)$ depends on the network size, the network topology, the initiator node and other factors. If T_r denotes the cover time then $C_r(T_r) = 1$; clearly $C_r(0) = 0$. As time increases, the RW agent is expected either to move to a node that hasn't been covered previously (thus, $C_r(t)$ increases) or to move to an already covered node (thus, $C_r(t)$ remains the same). Therefore, $C_r(t)$ is a non-decreasing function ($C_r(t_1) \leq C_r(t_2)$, for $t_1 < t_2$).

The number of movements of RW-based solutions is much smaller than that under flooding approaches (where one movement of an agent corresponds to one message transmission), at the expense of a significant increase in cover time. For example, for the case of a fully connected network, a number of movements of the order of $N \ln(N)$, [9], is required under the RW mechanism, while under flooding approaches the corresponding number of movements (or messages) is of the order of N^2. On the other hand, cover time under the RW mechanism is of the order of $N \ln(N)$, while under flooding it is upper bounded by the network diameter plus 1 (e.g., for the case of a complete graph cover time under flooding is 2).

3 The J-RW Agent

3.1 Motivation

Figure 1.a illustrates a RW agent movement path initiated from the initiator node depicted inside the dotted ellipsis. The random walker spends some time revisiting nodes in the depicted "upper-left" network part, while nodes in other network parts are left unvisited. Suppose now that after a few time units – long enough to "cover" a certain network part – the RW agent moves to a "new" (most likely uncovered) network part ("bottom right" network part in Figure 1.b). It is more likely than before to cover nodes that have not been visited previously by the agent, and therefore, accelerate the overall network cover process.

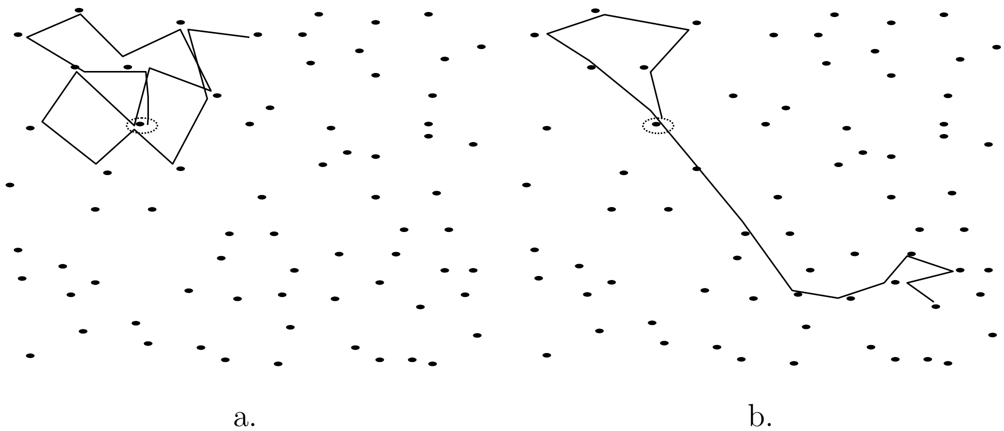


Fig. 1. An agent (starting from the initiator node depicted in dotted ellipsis) under: (a) the RW mechanism and (b) the J-RW mechanism.

One possible way for the agent to move away from a certain network region would be to carry out a number of consecutive directional movements, implementing a jump. This directional movement mechanism or *jumping*, initially proposed in [20], can be realized by switching occasionally away from the RW operation and engaging an operation implementing a directional movement. That is, such a RW agent (to be referred to as the Jumping Random Walk (J-RW)) operates under two states: *State* 0 under which it implements the typical RW mechanism without backtracking, and *State* 1 during which the directional move is implemented; the time spent in state 1 (freezing state) will be referred to as the freezing (the direction) period.

The J-RW mechanism moves the agent – at the end of the freezing period – to networks that are expected (due to the directional freeze) to be geographically more distant than those reached by the RW agent after the same number of movements. That is, the introduction of the freezing state implements in essence jumps, defined as the physical distance between the nodes hosting the RW agent at the beginning and the end of the freezing period.

The improvement in the cover time may be viewed as a consequence of "sampling" the network more uniformly, by moving the "sampling" agent into remote and likely new (not yet sampled) areas, as opposed to keeping the agent wandering around a certain locality according to the RW mechanism and (over)sampling predominately a certain locality. When a network graph has long links (that can take an agent into a remote network region in a topological sense), it has been shown in [7] that a RW agent produces a uniform sampling of the network nodes.

In wireless environments like sensor networks, the physical and the network topologies are typically correlated: a long path between two nodes in the network topology corresponds to a large physical distance between these nodes in the physical topology. In essence, the proposed J-RW mechanism applied over a network with no long (physical) links (like a wireless sensor network) creates virtual long links in this network and results in an environment that is equivalent to that of applying the RW agent over a network with some long links. Thus, the proposed J-RW mechanism is expected to result in a more uniform sampling of the network nodes, which – as argued earlier – leads to a better cover time.

Besides the improved cover time, the increased uniformity of the (node) sampling under the J-RW mechanism may be on its own another important property of the proposed information dissemination scheme when considered in conjunction with certain specific and fairly common applications such as those related to sensing the environment. In such applications and due to the typically high spatial correlation of nearby nodes, a dissemination of a query on the state of the field may target only a portion of the network nodes to conserve energy, [8]. Since the J-RW mechanism possesses the uniform sampling capability as argued earlier, it is expected that the query dissemination and collected responses would better represent the state of the sensor field and contain less redundant information. For such environments it is reasonable to base the evaluation of information dissemination schemes on the partial cover time as opposed to the 100% cover time.

3.2 Description

The proposed J-RW mechanism is based on two underlying states. When in *State* 0, the J-RW agent operates as the already described RW agent. When in *State* 1, it implements a directional walk, by selecting as the next node to visit to be the neighbor of the current node that is the closest to the line connecting the current node and the node visited by the agent in the previous discrete time, in the direction away from the previously visited node. The directional walk may be easily implemented through a simple look up table involving the geographic locations of the neighbours of a node; this table determines the next node to forward the agent to under the directional walk, given that the agent came to this node from a given neighbour. The geographic information can be easily retrieved either at the time of deployment in the case of a static sensor field (with provisions for second, third, etc. choices when lower order choices are not available due to battery depletion), or after the deployment of the field with the help of a localization protocol run occasionally.

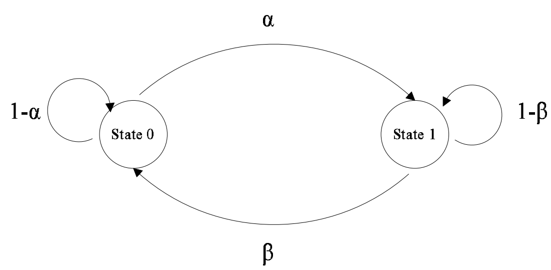


Fig. 2. Markov chain mechanism for controlling J-RW agent movements.

State transitions of the J-RW agent are assumed to occur at the discrete times according to a simple 2-state Markov chain, as shown in Figure 2; let α (β) denote the transition probabilities from State 0 to State 1 (State 1 to State 0) and let $T_0 = 1/\alpha$ ($T_1 = 1/\beta$) denote the mean time (in discrete times of our reference time, or number of visits to nodes) that the agent spends in State 0 (State 1). Clearly, β (or, T_1) determines the length of time over which the directional walk is continuously in effect and, thus, the mean length of the induced jump. Similarly, α (or, T_0) determines the length of time over which the RW mechanism is continuously in effect. It should be noted that α and β should be carefully selected so that the mix of the two distinct operations is effectively balanced. β should be such that the implemented jump is sufficiently large to move the agent away from the current locality that is likely to be covered by the operation at State 0, and on the other hand, it should not be too large in which case it would leave uncovered large areas or require the random walk operation to operate long enough (at the increased cost of revisits) to cover the large areas between the start and the end of the jump. Similarly, α should be such that the time spent at State 0 be balanced so as to not over-cover or under-cover the current locality.

As previously for the RW mechanism, coverage and cover time under the J-RW mechanism may be defined in a similar manner, denoted by $C_j(t)$ and T_j respectively. $C_j(t)$ is a non-decreasing function of t taking values between 0 (for $t = 0$) and 1 (for $t \geq T_j$).

4 Coverage Analysis

This section analyzes coverage under the RW mechanism in order to extract useful information regarding the performance of the RW agent and consequently to use these results for further understanding of the particulars of the J-RW mechanism. The analysis followed in this section is different than any other previous analysis in the best of the authors' knowledge.

4.1 Coverage under the RW mechanism

The main aim here is to derive an analytical expression for $C_r(t)$, which will serve as a tool for further understanding of random walk based information dissemination. Lets assume that the network topology is fully connected (i.e., all nodes are connected to all other nodes). This is actually the case for large values of r_c in geometric random graphs. For example, for nodes scattered in the $[0, 1] \times [0, 1]$ 2-dimensional plane, any value of $r_c \geq \sqrt{2}$ ensures that there is a link among any pair of nodes.

In such a network, each time the RW agent decides to move to a new neighbor node at time t (thus, arriving at time $t + 1$), coverage $C_r(t)$: (a) may increase ($C_r(t + 1) = C_r(t) + 1/N$), provided that the new node has not been covered previously; or (b) remain the same ($C_r(t + 1) = C_r(t)$), provided that the new node has already been covered. Note that at time t, in a fully connected network the RW agent may select one out of $N - 2$ network nodes (i.e., all network nodes excluding the one the agent came from and the one that is currently located at). Since $1/N$ corresponds to the coverage contribution of the node the agent came from and $1/N$ to the coverage contribution of the node that is currently located at, then $C_r(t) - 2/N$ is the coverage corresponding to the remaining $N - 2$ nodes and eventually, $(N - 2) \times (C_r(t) - 2/N)$ corresponds to the number of nodes that have already been visited by the agent (excluding the one the agent came from and the one that is currently located at). For large values of N (which is typically the case considered in this paper), $(N - 2) \times (C_r(t) - 2/N) \approx NC_r(t)$. Consequently, the probability to choose a node that has not been visited previosuly (and consequently increase coverage) is equal to the probability of selecting one out of the $N - NC_r(t)$ nodes that have not been visited previsouly, or $N(1 - C_r(t))/N = 1 - C_r(t)$. It is easily derived now that (on average) the increment of the coverage after a RW agent moves at time t, is given by the probability $1 - C_r(t)$ that it moves to a node not visited before multiplied by $1/N$ which is the contribution to coverage by each node that is visited for the first time. That is,

$$C_r(t+1) - C_r(t) = \frac{1}{N}\Big(1 - C_r(t)\Big). \tag{1}$$

Equation (1) can be expressed in a more convenient form by switching from discrete to continuous time. Let t be continuous and let $\tilde{C}_r(t)$ denote the corresponding continuous and increasing function of $C_r(t)$. The difference $C_r(t+1) - C_r(t)$ can be approximated by $\frac{\tilde{C}_r(t_1) - \tilde{C}_r(t_0)}{t_1 - t_0} = \frac{\mathrm{d}\tilde{C}_r(t)}{\mathrm{d}t}$, for $t_0 < t_1$. Based on Equation (1),

$$\frac{\mathrm{d}\tilde{C}_r(t)}{\mathrm{d}t} = \frac{1}{N}\Big(1 - \tilde{C}_r(t)\Big). \tag{2}$$

The derivative $\frac{\mathrm{d}\tilde{C}_r(t)}{\mathrm{d}t}$ corresponds to the rate at which $\tilde{C}_r(t)$ increases. Obviously, for $t = 0$ (i.e., the RW agent is about to start moving in the network), $\frac{\mathrm{d}\tilde{C}_r(t)}{\mathrm{d}t} = 1$, since during its first step the RW agent will move to a node definitely not covered previously. $\frac{\mathrm{d}\tilde{C}_r(t)}{\mathrm{d}t}$ will eventually become zero, when all nodes are covered or, $\tilde{C}_r(t) = 1$.

Equation (2) is a first class differential equation, and the solution satisfying the previous properties of $\tilde{C}_r(t)$ (e.g., increasing, $0 \le \tilde{C}_r(t) \le 1$), is given by,

$$\tilde{C}_r(t) = 1 - e^{-\frac{t}{N}}. \tag{3}$$

For convenience of the presentation, very frequently in the sequel, the normalized version $\tilde{C}_r(t/N) = 1 - e^{-t}$ will be used instead of $\tilde{C}_r(t)$.

In the literature, a well known analytical result for the complete graph case, is that cover time T_r is of the order of $N\ln(N)$, [9]. Using Equation (3) it is now easy to derive that $\tilde{C}_r(N\ln(N)) = 1 - e^{-\frac{N\ln(N)}{N}} = 1 - \frac{1}{N}$. For large values of N it is clear that $\tilde{C}_r(N\ln(N)) \to 1$. Another interesting result is that for $t = N$, $\tilde{C}_r(N)) = 1 - e^{-\frac{N}{N}} = 1 - e^{-1} = 0.632$, which basically means that for movements equal to the number of the network nodes, on average 63.2% of the network nodes are visited by the RW agent. These analytical findings are confirmed later by simulation results presented in Section 5.

Equation (3) was derived assuming a fully connected network. By reducing r_c in geometric random graphs, the number of neighbor nodes decreases and therefore a RW agent has fewer choices to move than before. Therefore, the fraction of nodes that have (not) been visited previously, is expected to deviate from $C_r(t)$ (or $1 - C_r(t)$). As r_c decreases even further, it is getting more and more difficult for the agent to move to a node not previously visited (due to the fewer choices of movement), thus increasing the number of revisits and eventually, decreasing the actual rate under which $C_r(t)$ increases. For example, bottlenecks tend to appear in the network topology, [19], likely "forcing" a RW agent to keep (re)visiting a comparably small number of nodes for a long time.

In order to account for the aforementioned decrease in the increase rate of $\tilde{C}_r(t)$, it is assumed in the sequel that derivative $\frac{\mathrm{d}\tilde{C}_r(t)}{\mathrm{d}t}$ is given by Equation (4),

$$\frac{\mathrm{d}\tilde{C}_r(t)}{\mathrm{d}t} = \frac{k}{N}\Big(1 - \tilde{C}_r(t)\Big), \tag{4}$$

where $k \leq 1$ is a positive constant related to the topology characteristics (i.e., connectivity radius r_c and number of nodes N). The solution of Equation (4) is given by,

$$\tilde{C}_r(t) = 1 - e^{-\frac{k}{N}t}. \tag{5}$$

The case of $k = 1$ corresponds to the fully connected network topology (i.e., large values of r_c) as it is concluded from Equation (3). Therefore, smaller values of r_c (however large enough for the network to be connected) should result in smaller values of k. This will be further explored and evaluated using simulation results presented later in Section 5.

Figure 3.a depicts $\tilde{C}_r(t/N)$ as it is given by Equation (5) as a function of t/N for various values of k. It is interesting to note that as k decreases, cover time increases. For example, for $t = \frac{N}{k}\ln(N)$, $\tilde{C}_r(\frac{N}{k}\ln(N)) = 1 - \frac{1}{N}$ which tends to zero for large values of N. Let $\tilde{T}_r = \frac{N}{k}\ln(N)$ be referred to hereafter as the *asymptotic* cover time for the RW agent (i.e., $\lim_{N\to+\infty}\tilde{C}_r(\tilde{T}_r) = 1$). Given that $k < 1$, it is evident that as r_c decreases (the network has fewer links), k decreases and therefore, the asymptotic cover time $\tilde{T}_r$ increases by a factor $\frac{1}{k}$.

This observation is graphically presented in Figure 3. It is evident that the first derivative of coverage is high at the beginning and then it becomes significantly small, particularly for values of k close to 1. Along with $\tilde{C}_r(t/N)$ in Figure 3.a, a (dotted) line corresponding to $\tilde{C}_r(t/N) = t/N$ is also depicted. This line corresponds to the best (even though frequently not realistic) dissemination information scenario (i.e., the – artificial – case in which a node not previously visited is reached after each step). It is interesting to note that the particular

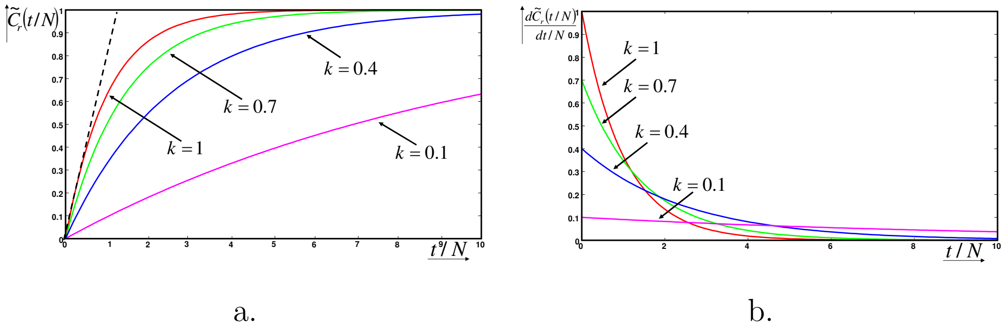


Fig. 3. $\tilde{C}_r(t/N)$ and $\frac{\mathrm{d}\tilde{C}_r(t/N)}{\mathrm{d}t/N}$ as a function of t/N for various values of k.

case of $k = 1$ is the one most close to this best scenario. This is basically the case since for some time at the beginning (small values of t/N) $\frac{\mathrm{d}\tilde{C}_r(t/N)}{\mathrm{d}t/N}$ remains close to 1 for $k = 1$, as it is also depicted in Figure 3.b. Afterwards, $\frac{\mathrm{d}\tilde{C}_r(t/N)}{\mathrm{d}t/N}$ decreases (more rapidly than for cases of smaller k) due to frequent revisits of the RW agent. Simulation results are presented later in Section 5 demonstrating the accuracy of the previous analysis.

4.2 Coverage under the J-RW mechanism

There are basically two issues that need to be explored with respect to the J-RW mechanism presented in Section 3: (a) for how long should the J-RW agent stay in State 0 (governed by α); and (b) for how long should the J-RW agent stay at State 1 (governed by β).

Let's assume that α is close to 0. This is a trivial case in which the J-RW mechanism resembles the RW mechanism. This choice is expected to be a suitable one in topologies of highly connected nodes (large number of neighbors) such that the agent is allowed to move to nodes residing away from the already covered network parts. This may happen due to its own probabilistic movement within the network. An obvious example is the fully connected topology, but as it will be shown later using simulation results, this is also applied for topologies of smaller connectivity (i.e., smaller values of r_c) than the fully connected topology.

On the other hand, values of α close to 1, mean that the agent will not stay long in State 0. This eventually means that the agent will not be allowed a large time period to cover a certain network part. In this case, the agent will be mostly "jumping" (depending on the value of β) to different network parts thus failing to exhaustively cover network areas. This reassembles a RW agent moving over an overlay topology of longer links.

The connectivity of the topology (which for the case of the geometric random graphs is related to r_c) plays an important role for the investigation of the appropriate values of β with respect to coverage under the J-RW mechanism. Highly connected topologies (i.e., comparably high values of r_c but not necessarily close to the particular value that the topology is fully connected) are characterized by significantly small diameters, [19]. In such networks, a RW agent is not expected to be limited within certain network parts and therefore, jumping would not improve coverage. In such cases, β should be close to 1. On less connected topologies (i.e., topologies of small values of r_c but large enough for the network to be connected) it is expected the RW agent to frequently revisit nodes due to the topology's structure (e.g., bottlenecks). In such a case, β should be small enough to allow the agent to move to different network parts. On the other hand, too small values of β (e.g., β close to 0) will eventually result to an agent mostly operating in State 1 (i.e., jumping) at the expense of exploring more thoroughly the visiting part of the network.

As it can be concluded from the previous discussion, coverage under the J-RW mechanism is related to r_c, α and β. However, an analytical expression for the coverage considering r_c, α and β is difficult to be derived and its further investigation will be based on simulations presented in the following section. Let $C_j(t)$ ($\tilde{C}_j(t)$) denote the coverage under the J-RW mechanism in a similar way as $C_r(t)$ ($\tilde{C}_r(t)$) denotes coverage under the RW mechanism. After a long time, it is reasonable to assume that the J-RW agent will have moved to all different network parts and have covered (on average) the same proportion of network nodes within each visited network part. Therefore, at time t it is expected that (on average) the fraction of non-visited neighbor nodes of the node that the agent is located at, will be $N \times C_j(t)$. Following an analysis similar to the one presented for the case of the RW mechanism, the following analytical expression for coverage under the J-RW can be written

$$\tilde{C}_j(t) = 1 - e^{-\frac{k'}{N}t}, \tag{6}$$

where k' is a positive constant depending on the particulars of the topology (i.e., r_c) and the J-RW mechanism (i.e., α and β), as it will be also shown in the following section using simulation results.

5 Simulation Results and Evaluation

A simulation program exploiting the capabilities of the Omnet++ simulation platform, [21], was created for the simulation purposes. The aim of the simulation results presented is this section is twofold: to confirm the analytical findings of the previous section and to shed more light on the behavior of the J-RW mechanism (mostly in comparison to the RW mechanism) for cases not covered by the analysis.

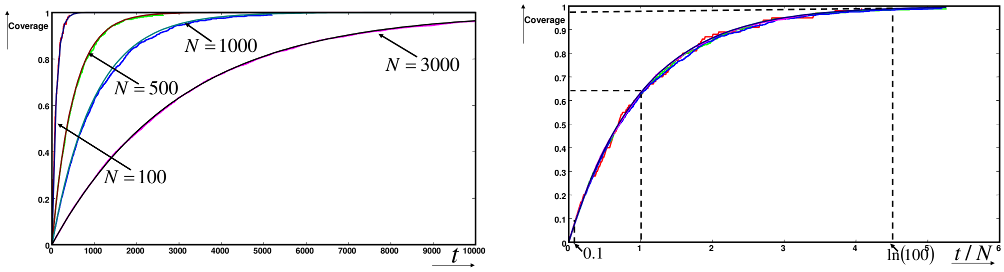
There are multiple simulation runs executed under specific sets of parameters for the network and the investigated schemes. During each simulation run there is a large-scale node set up, with node population varying from 100 to 3000 nodes depending on the case. The nodes are placed at random locations on a square plain $[0, 1] \times [0, 1]$. The random positions (x_u, y_u) of each node $u \in V$ are chosen within the set $[0, 1]$ using the uniform probability distribution. Each node u is aware about its own position: (x_u, y_u). Each node is connected to some other node if the euclidean distance among them is less or equal to r_c. Clearly, for $r_c \geq \sqrt{2}$, the resulting network is fully connected. Depending on N (the size of the network), the lower bound of r_c for which the topology remains connected varies (typically decreases as N increases). Four different values of r_c (0.05, 0.1 0.5 and 1.0) are considered in the sequel for those topologies of $N = 1000$. Note that all four values are less than $\sqrt{2} \approx 1.4$, since a fully connected network is not representative of the wireless environment (e.g., wireless sensor networks) that is considered here.

Coverage is the main focus in the result to be presented. These results correspond to one experimentation instance (apart from some cases which are explicitly mentioned) and not averaged values. Averaging would have given a macroscopic view but it would also hide important details.

5.1 The RW Mechanism

An important contribution in the analytical part of this paper, is the derivation of the coverage in a fully connected network (i.e., $r_c \geq \sqrt{2}$), shown in equation 3. Figure 4.a presents coverage as a function of t (i.e., the number of movements of the random walker) in fully connected network topologies of 100, 500, 1000 and 3000 nodes. For each network topology, the analytically derived value $\tilde{C}_r(t)$ is also depicted. It is important to note that the analytical findings are almost identical to the simulation results. Figure 4.b presents (normalized) coverage as a function of t/N. As before, it is demonstrated that $\tilde{C}_r(t/N)$ is almost identical to $C_r(t/N)$ as it is concluded by observing simulation results for the case of a fully connected topology.

The fact that all curves follow the same pattern as it is illustrated in Figure 4.b allows for certain observations. First, it is obvious that for about $0.1N$ movements (see dotted line at $t/N = 0.1$ depicted in Figure 4.b), the random walker has a very good performance (the number of revisits remains small) in the sense that a new movement most likely results in visiting a node that has not been visited before. At time $t = N$, it is interesting to see that for almost all cases, 63.2% of the total number of network nodes has been covered



a. b.

Fig. 4. Coverage and normalized coverage under the RW mechanism for fully connected topologies of various network sizes.

as it was expected from the analysis presented in the previous section. For $N = 100$, $t = N \ln(N)$ corresponds to $t/N = \ln(100) \approx 4.6$ in Figure 4.b. As expected from the analysis (and also depicted in Figure 4.b), coverage is about $1 - 1/N = 99\%$, which is very close to that depicted using simulation results.

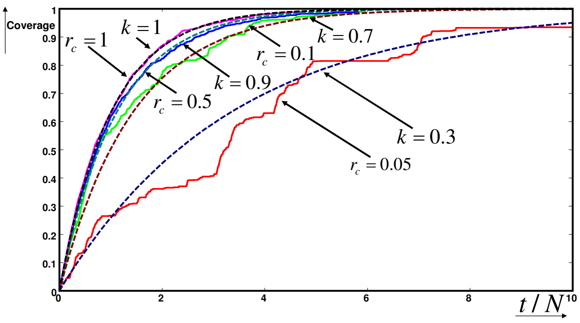


Fig. 5. Coverage for topologies of 1000 nodes and various values of r_c (0.05, 0.1, 0.5 and 1.0). $\tilde{C}_r(t)$ is also depicted for corresponding values of k (0.3, 0.7, 0.9 and 1.0).

Figure 5 presents simulation results for various topologies derived for $r_c = 0.05$, 0.1, 0.5 and 1.0. The first observation is that for the appropriate value of k (i.e. $k = 0.3$ for $r_c = 0.05$, $k = 0.7$ for $r_c = 0.1$, $k = 0.9$ for $r_c = 0.5$ and $k = 1.0$ for $r_c = 1.0$), the analytical expression for coverage, given by Equation (5) (i.e., $\tilde{C}_r(t) = 1 - e^{-kt}$) approximates well the simulation results. Another interesting observation is that for $r_c > 0.1$, the appropriate value of k is very close to 1.0, This is basically due to the fact that the number of neighbors (on average) increases much faster than r_c (it is $\pi r_c^2 N$ on average) and, thus, a small increase in r_c results in a large increase of the node degree (number of neighbors). For $N = 1000$ and $r_c = 0.1$, there are (on average) about 31 neighbor nodes for each node which implies that the topology is highly connected. As r_c increases further, it is interesting to observe that the coverage approaches closely $1 - e^{-t}$, even for cases for which r_c is significantly smaller that $\sqrt{2} \approx 1.4$.

Figure 6 presents simulation results under the J-RW mechanism for a network of 1000 nodes and various values of r_c and α. β has been kept constant and equal to 0.4, which means that as soon as State 1 is assumed (i.e., directional movement) the agent moves (on average) for 2-3 nodes towards a certain direction (more details are provided in the description of the J-RW mechanism in Section 3) before State 0 is assumed. In Figure 6.a, coverage under the RW mechanism is clearly depicted and it is less than the coverage under J-RW for any value of α (e.g., 0.2, 0.4, 0.6 and 0.8). Note that the case depicted in Figure 6.a corresponds to a topology that is not highly connected ($r_c = 0.05$), thus even a relatively small value of $\beta = 0.4$ results in the J-RW doing significantly long jumps to get a perfomance improvement.

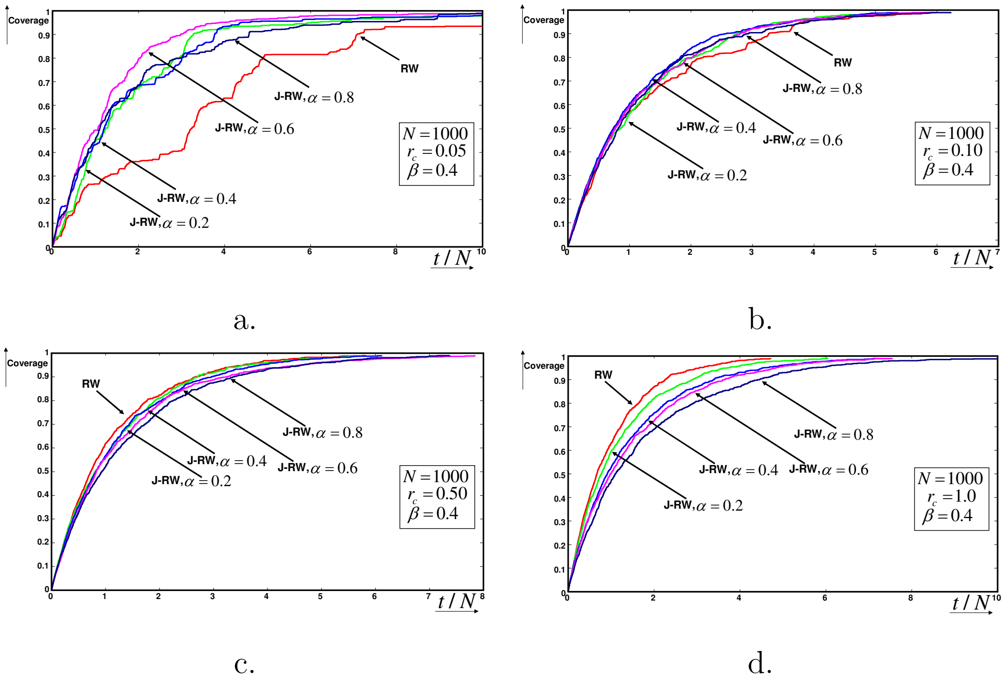


Fig. 6. Coverage for various values of α ($\beta = 0.4$) as a function of t/N.

As the topology becomes more connected (r_c increases), the advantage of the J-RW mechanism is less obvious. For example, for the case depicted in Figure 6.b ($r_c = 0.1$), coverage under the RW mechanism is still smaller than that under the J-RW mechanism (for any value of α), even though not that smaller as before, while for the case depicted in Figure 6.c ($r_c = 0.5$), coverage under the RW mechanism is now larger than that under the J-RW mechanism (for any value of α). As r_c increases further, coverage under the RW mechanism is clearly higher than that under the J-RW mechanism for the specific combination of values of α and β. This is clearly depicted in Figure 6.d for the case of $r_c = 1.0$ and can be attributed to the fact that the RW mechanism can now fully exploit the increased connectivity of the graph. In particular, subsequent movements of the RW agent in highly connected graphs are similar to J-RW agent movements in low connectivity graphs (in terms of how far in physical distance the agent moves). Thus, there is no coverage benefit when introducing J-RW agent in highly connected graphs (as compared with RW agent), on the contrary the J-RW mechanism of 'locking' in state 1 tends to push the agents away towards the physical boundaries of the examined network and thus result in an actual decrease in network coverage.

Another interesting aspect is the shape of the curve corresponding to the J-RW mechanism. As it can be observed in 6, it clearly follows the pattern of the analytical expression $1 - e^{-\frac{k'}{N}t}$, for suitably selected values of k', as it was already mentioned in Section 4. This is more clearly seen in the simulation results depicted in Figure 7 along with plots for $\tilde{C}_r(t/N)$ and $\tilde{C}_j(t/N)$. It is obvious that simulation results follow the same pattern for the J-RW mechanism as it is also the case (also shown before) for the RW mechanism.

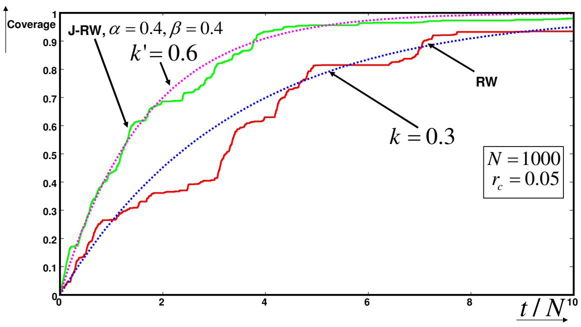


Fig. 7. Simulation and analytical results for a topologies of 1000 nodes under the RW and the J-RW mechanism.

The particular value of β that was used for the simulation results depicted in Figure 6 was fixed ($\beta = 0.4$). Figure 8 presents simulation results (coverage) as a function of both α and β (their values in the range $[0.2, 0.8]$) at time $t = N$ and Figure 9 at time $t = N \ln(N)$.

The results depicted in Figure 8.a correspond to $r_c = 0.05$. About 25% of the network is covered under the RW mechanism while coverage for this case under the J-RW mechanism varies from 30% to 55% depending on the particular values of α and β. For $r_c = 0.1$, as depicted in Figure 8.b, coverage under the RW mechanism is again smaller than that under the J-RW mechanism for any selection of α and β. In Figure 8.c ($r_c = 0.5$), it is interesting to see that coverage under the RW mechanism is greater than that under the J-RW mechanism, apart from those cases that β is high (e.g., 0.8) and α is small (e.g., 0.2). In such cases the high value of β results in relatively small jumps within the network, which are small enough to be effective in such a highly connected network. Any smaller value of *beta* would result in larger jumps within the highly connected graph, 'pushing' again the J-RW agents towards the boundaries of the network and reducing performance. Thus the

J-RW agent is able to move now between different 'neighbor areas' and explore them efficiently.

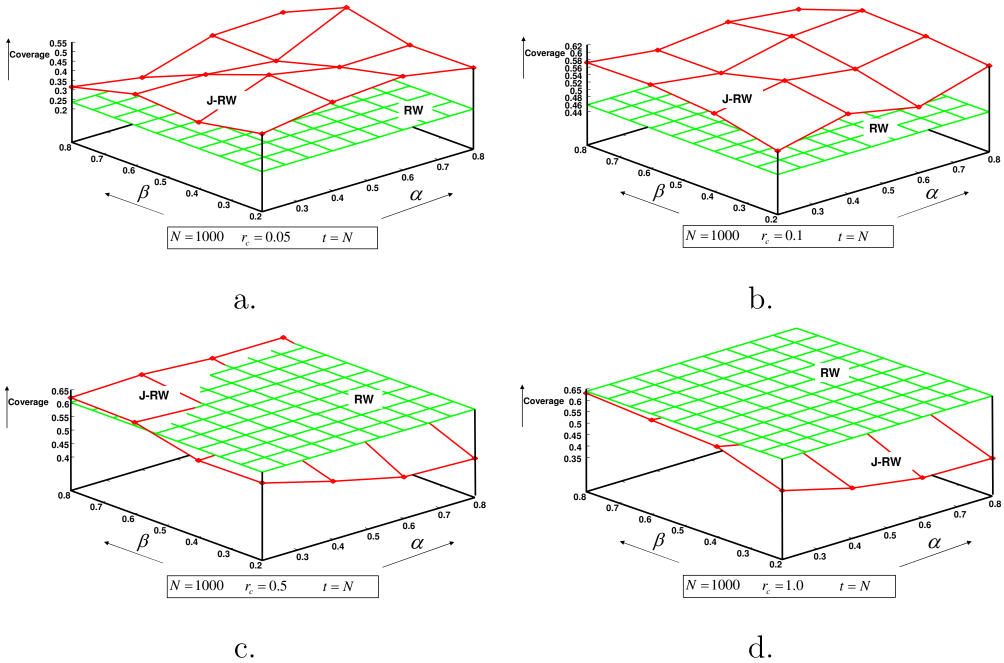


Fig. 8. Coverage for various values of α and β at $t = N$.

As r_c increases further, as it is the case in Figure 8.c for $r_c = 1.0$, it is obvious that there is no combination of α and β in the given range of values such that coverage under the J-RW mechanism be greater than coverage under the RW mechanism. At this point it is important to note that the J-RW mechanism becomes equivalent to the RW mechanism for $\alpha = 0$ and $\beta = 1$ (which means that the mechanism always stays at State 0).

Simulation results for $t = N \ln(N)$ are depicted in Figure 9. The observations are identical to those regarding Figure 8 apart from the fact that coverage is close to 1 (which is expected given the analytical results). Similarly to the case of $t = N$, for $t = N \ln(N)$ coverage under the J-RW mechanism is larger than that under the RW for topologies of small values of r_c. It is interesting to observe in Figure 9.a that for large values of both α and β (e.g., 0.8) coverage under the J-RW can be close to 98%, which is a significant improvement when compared to coverage under the RW mechanism that is close to 86%. In all other cases and particularly those for which coverage under the RW mechanism is close to 100%, coverage under the J-RW mechanism is close to 100% for values of β not that small (long jumps should be avoided in such highly connected topologies). Note that values for α and β around 0.5, appear to be an appropriate selection, as it is observed from the simulation results depicted in both Figure 8 and Figure 9, since: (a) for topologies of small r_c, coverage under the RW mechanism is greater than coverage under the J-RW mechanism; and (b) for all other cases, coverage is almost the same under

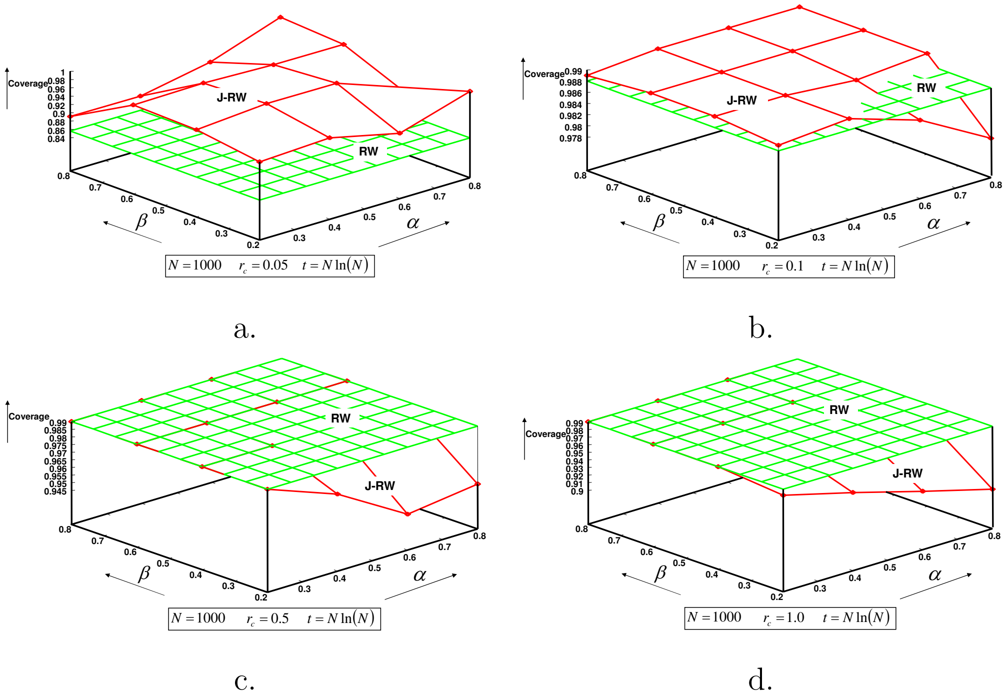


Fig. 9. Coverage for various values of α and β at $t = N \ln(N)$.

either mechanism.

6 Conclusions

Random walk based information dissemination or search is a non trivial problem in large scale, wireless network environments like wireless sensor and ad hoc networks. The simple random walk (RW) methodology, already proposed for use in such environments, is associated with high inefficiencies due to frequent already covered nodes revisits. An acceleration mechanism for covering the network with a RW agent, the Jumping Random Walk, has been proposed to accelerate coverage of the disseminating agent within the network. Since the traditional random walk mechanism requires a significant amount of time before covering the entire network in sparse network topologies, as it is commonly the case is wireless environments, the jumping random walk can provide a significant 'boost' in the cover process. Both analytical and experimental results provide for the correctness of our approach and exhibit the significant advantages in terms of coverage percentages provided by the jumping random walk agent.

References

[1] A. Segal, Distributed network protocols, IEEE Trans. on Information Theory IT (29) (1983) 23–35.

[2] B.Williams, T.Camp, Comparison of broadcasting techniques for mobile ad hoc networks, in: MOBIHOC 2002, 2002, pp. 194–205.

[3] K.Oikonomou, I.Stavrakakis, Performance analysis of probabilistic flooding using random graphs, in: The First International IEEE WoWMoM Workshop on Autonomic and Opportunistic Communications (AOC 2007),, 2007.

[4] D.Tsoumakos, N.Roussopoulos, Adaptive probabilistic search for peer-to-peer networks, in: 3rd IEEE International Conference on P2P Computing, 2003.

[5] M. O.-K. M.Bani Yassein, S.Papanastasiou, On the performance of probabilistic flooding in mobile ad hoc networks, in: 11th International Conference on Parallel and Distributed Systems (ICPAD'05), 2005.

[6] A. L.-O. R.Dorrigiv, P.Pralat, search algorithms for unstructured peer-to-peer networks, in: LCN 07: Proceedings of the 32nd IEEE Conference on Local Computer Networks, 2007, pp. 343–352.

[7] M. M. C. Gkantsidis, A. Saberi, Random walks in peer-to-peer networks, in: Proc. INFOCOM'04, Hong Kong, 2004.

[8] C. Avin, C. Brito, Efficient and robust query processing in dynamic environments using random walk techniques, in: Proc. IPSN 2004, Berkeley, California, 2004.

[9] L. Lov'asz, Random walks on graphs: a survey, Combinatorics, Paul Erdos is Eighty, J. Bolyai Math. Soc., Vol. II 2 (1996) 353–397.

[10] E. S. S. Dolev, J. Welch, Random walk for self-stabilizing group communication in ad hoc networks, IEEE Trans. on Mobile Computing 5 (7).

[11] C. Avin, B. Krishnamachari, The power of choice in random walks: An empirical study, in: Proc. 9th ACM/IEEE International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems, (MSWiM), Malaga, Spain, 2006.

[12] D. Braginsky, Rumor routing algorithm for sensor networks, in: Proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Applications, 2002, pp. 22–31.

[13] U. Feige, A spectrum of time-space trade-offs for undirected s-t connectivity, Journal of Computer and System Sciences 54 (2) (1997) 305–316.

[14] A. A. Abouzeid. Nabhendra Bisnik, Optimizing random walk search algorithms in p2p networks, Comnet 51 (6) (2006).

[15] H. W. T. Lin, Search performance analysis in peer to peer networks, in: Third International Conference on Peer To Peer Computing (P2P'03), 2003, pp. 82–83.

[16] E. C. K. L. S. S. Q. Lv, P. Cao, Search and replication in unstructured peer-to-peer networks, in: ACM SIGMETRICS'02, 2002, pp. 258–259.

[17] C. Avin, G. Ercal, On the cover time of random geometric graphs, in: ICALP. (2005), pp. 677–689.

[18] M. Medina, I. Matta, J. Byers, On the origin of power laws in Internet topologies, SIGCOMM Comput. Commun. Rev. 30 (2) (2000).

[19] M. Penrose, Random Geometric Graphs, Oxford studies in probability, Oxford University Press, 2003.

[20] L. Tzevelekas, I. Stavrakakis, Improving Partial Cover of Random Walks in Large Scale Wireless Sensor Networks, 3rd IEEE WoWMoM Workshop on Autonomic and Opportunistic Communications (AOC'09), 15 June 2009, Kos, Greece.

[21] Omnet++ Simulation Program. `http://www.omnetpp.org`