A Study of Information Dissemination Under Multiple Random Walkers and Replication Mechanisms

Konstantinos Oikonomou
Dept. of Informatics
Ionian University
Corfu, Greece
okon@ionio.gr

Dimitrios Kogias
Dept. of Informatics &
Telecommunications
National and Kapodistrian
University of Athens
Athens, Greece
dimkog@di.uoa.gr

Ioannis Stavrakakis
Dept. of Informatics and
Telecommunications
National and Kapodistrian
University of Athens
Athens, Greece
ioannis@di.uoa.gr

ABSTRACT

Random walk-based approaches are suitable for information dissemination in mobile and opportunistic environments that are inherently dynamic and typically large-scale. *Multiple random walkers* have been proposed in the recent past as a suitable mechanism to significantly reduce *termination time* when covering a network topology. In this paper, the multiple random walkers mechanism is studied and analyzed under a different and novel perspective that allows for further understanding of network *coverage* when employing multiple random walkers in a network. Given that under this mechanism the number of movements is significantly increased (proportional to the number of random walkers) – and sometimes not as effective as expected due to frequent *revisits* that do not improve coverage as shown in this paper – a *replication* mechanism is introduced that is capable of reducing the number of movements by reducing initial revisits. Simulation results considering geometric random graph topologies – which are suitable for modeling mobile environments – are also presented supporting the analytical findings and shedding more light on various aspects of both the aforementioned mechanisms.

Categories and Subject Descriptors

C.2 [**Computer-Communication Networks**]: Miscellaneous

General Terms

Algorithms, Performance, Theory

Keywords

Dissemination and Replication Techniques, Multiple Random Walkers, Coverage

1. INTRODUCTION

Information dissemination is a challenging problem in networking environments like mobile and opportunistic networks. These networks, inherently dynamic and typically large-scale, suffer from scalability problems if traditional approaches are followed (e.g., spanning tree formation, flooding-based approaches). For example, traditional flooding [11], under which information is deterministically disseminated to all network nodes, is known to require a large *number of messages*. In a mobile environment this corresponds to an increased amount of consumed energy that may severely affect vital aspects of the network operation (e.g., network lifetime). On the other hand, traditional flooding is efficient in terms of *termination time* (i.e., time until all network nodes are visited) known to be upper bounded by the network diameter.

Under a different prespective, *random walks* [8], [5] have been proposed for disseminating information to deal with the idiosyncrasies of these network environments. A random walker, carrying a piece of information to be disseminated, probabilistically chooses the next neighbor node to visit independently of any previous visit. This mechanism is expected to probabilistically *cover* a network (i.e., reach all network nodes) using significantly less messages (one random walker movement corresponds to one message) than under traditional flooding at the expense of increased termination time.

In order to fill the gap between random walks and traditional flooding, several ideas have recently emerged attempting to exploit the good properties of random walks (i.e., relatively small number of messages) and trying to reduce termination time. These approaches among others include random walkers with jumps [14], random walkers replication [7], multiple random walkers [3], hybrid approaches [5]. Further information about various dissemination information approaches is given in Section 2.

The work presented in this paper is motivated by the work presented in [3] and [7]. In particular, it is shown in [3] that multiple random walkers, starting from the same network node, are capable of accelerating the information dissemination process and reduce termination time by a factor equal to the number of random walkers, for a wide range of topologies. On the other hand, as the number of random walkers increases, the number of messages sent increases proportionally to the number of random walkers. Moreover, since random walkers start from the same network node, it is ex-

pected for some initial movements to partially overlap (thus, not improving coverage) due to visits to already visited network nodes (i.e., *revisits*). This motivates the adoption of a *replication* approach – under which replicas of random walkers are probabilistically created after each movement – so as to avoid initiating all of them at the same time as under the multiple case [7]. A simple replication mechanism is proposed here capable of covering larger network areas than multiple random walkers for the same number of random walkers and allowed number of messages.

The main contribution of this paper is the study of multiple random walkers from a different prespective than the one presented in [3]. The analytical part in this paper initially assumes a fully connected network topology which allows for the derivation of an analytical expression that confirms the results presented in [3] allowing also for further understanding of various aspects of information dissemination under multiple random walkers. To the best of the authors' knowledge this is the first time that such an analysis is carried out for multiple random walkers. The analysis continues capturing coverage in less dense topologies and an analytical expression is derived showing how coverage is affected by frequent random walk revisits.

Simulation results, derived considering geometric random graph topologies that are suitable for modelling mobile environments [10], are shown to be in accordance with the analytical findings. In addition, the proposed replication mechanism is compared against the multiple random walkers mechanism and it is shown that the former is capable of covering larger network areas than the latter under certain conditions also explored in this paper.

Section 2 briefly covers recent works related to information dissemination. In Section 3 the multiple random walkers mechanism is studied analytically and in Section 4 the replication mechanism is presented. Section 5 includes the simulation results and the conclusions are drawn in Section 6.

2. PAST RELATED WORK

Many variations of traditional flooding have been proposed in the past for information dissemination purposes attempting to reduce the (increased) number of messages. One such approach is probabilistic flooding [4], [13], [9], [12] that probabilistically ensures global network outreach with high probability (meaning that some nodes may not receive the particular message) and significantly reduces the number of messages (when compared against traditional flooding), even though termination time slightly increases (again compared against traditional flooding). Other flooding-based variations have been proposed like controlled flooding (information is disseminated only a few hops away). This approach is used, among others, in the Gnutella Peer-to-Peer system [1]. Obviously, under controlled-flooding, a small number of hops allow for both reduced number of messages and termination time. On the other hand, an inappropriately small number of hops may result in information being locally disseminated and consequently, in an ineffective information dissemination scheme.

Random walks, on the other hand, may achieve global network outreach after a significant amount of time depending on the topology characteristics (e.g., density, bottleneck links). Given that movements are random, the network is eventually probabilistically covered (any guarantee is given

with high probability and not certainty). In order to deal with the main drawback of random walkers (i.e., revisits), various approaches like random walks with jumps [14] (allows for random walkers to move to different network areas increasing the probability to visit not previously visited nodes) and random walks replication [7] have been proposed. Hybrid probabilistic schemes (e.g., a controlled flooding process initiated after a random walk) have also been proposed [5]. Other modifications like the one in [6], allow for network nodes to forward messages to their neighbors in a random manner, thus significantly reducing the number of messages in the network.

As already mentioned, the most related work to this paper is [3]. Alon et. al analyze termination time until certain network topologies are covered by multiple random walkers starting from the same network node. They show that this approach is faster than the single random walker case and proportional to the number of multiple random walkers for a wide range of topologies. The work presented in this paper is different from the one in [3] since: i) the analytical part is different even though the basic results are in accordance; ii) the analytical expressions have a rather tractable form that can be used for further understanding and exploring multiple random walkers; iii) a replication mechanism is proposed and studied in order to reduce initial revisits and improve coverage.

3. MULTIPLE RANDOM WALKERS

Assume that a network is represented by a bidirectional graph $G(V, E)$, where V is the set of network nodes and E the set of (bidirectional) links among nodes. For the rest N denotes the number of network nodes, or $N = |V|$, where $|\cdot|$ denotes the number of elements of a certain set.

A random walker may be seen as an agent moving in the network disseminating (or collecting) information. Assuming time t to take discrete values starting from 0, each time unit corresponds to one movement for all random walkers. For the case of m multiple random walkers, each time unit corresponds to m movements (or messages). A basic assumption is that all m random walkers are initially placed at the same node [3], referred to hereafter as the *initiator* node.

The multiple random walkers mechanism: Having started with m random walkers from the same initiator node, each random walker moves to one of its neighbor nodes being selected randomly and independently among the set of neighbor nodes, provided that the chosen node is not the previously visited node unless this is the only neighbor node.

Let us $C_m(t)$ denote *coverage* or the fraction of the network nodes visited by any of the m random walkers at time t. $C_m(t)$ is an increasing function of t taking values between $1/N$ (i.e., the case when only one node is visited) and 1 (i.e., all nodes are visited). It is easy to derive that $C_m(0) = 1/N$, which corresponds to the case that all m random walkers are located at the initiator node. Note that for the typically large-scale network environments considered in this work, N is significantly large such that $C_m(0) = 1/N \to 0$. As time increases, the random walkers start moving, visiting network nodes and increasing coverage. Under the assumption that the network is connected, after some time all network nodes will be visited by (at least) one of the m random walkers. For the rest *termination time*, denoted by T_m, is defined as the smallest value of t such that $C_m(t) = 1$.

Depending on the case (for example sparse topologies of many bottleneck links), T_m may become significantly large, thus, sometimes, failing to provide a satisfactory insight on the dissemination information characteristics. Alternatively, it is frequently convenient to consider the *asymptotic termination time* $\hat{T}_m$ which is defined as the smallest value of t such that $\lim_{N \to +\infty} C_m(t) = 1$.

Assume that the underlying topology is fully connected (i.e., all nodes are connected to all other nodes and thus, $|E| = N\frac{N-1}{2}$). Even though this is not a realistic topology scenario it is useful for providing insight on the behavior of multiple random walkers (less dense topologies are also considered later in this work).

THEOREM 1. *In a fully connected network topology of N nodes and m random walkers, coverage $C_m(t)$ as a function of time t is given by:*

$$C_m(t) = 1 - e^{-\frac{m}{N}t}. \tag{1}$$

The proof can be found in Appendix A.

Figure 1 depicts coverage $C_m(t)$ as given by Equation (1) as a function of t, for $N = 1000$ (zoom in different ranges of t). It is clear that as $t \to +\infty$, $C_m(t) \to 1$. It is evident that as m increases, coverage increases as well.
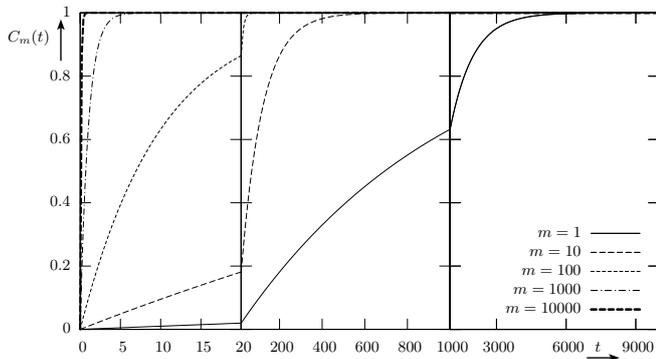

Figure 1: Coverage $C_m(t)$ as a function of t for a network of $N = 1000$ nodes and various values of m (zoom in different ranges of t).

Based on Equation (1) an analytical expression with respect to the asymptotic termination time $\hat{T}_m$ can be derived. It is enough to satisfy that $\lim_{N \to +\infty} C_m(\hat{T}_m) = 1$. This is satisfied by the following:

$$\hat{T}_m = \frac{N}{m} \ln(N). \tag{2}$$

From Equation (2) it is evident that for the case of a complete graph, the (asymptotic) termination time decreases proportionally by the number of multiple random walkers m which is in accordance with [3]. In the following section, the fully connected topology assumption is relaxed and in the simulations section presented later, simulation results are shown to be in accordance with the analytical findings of this section.

4. A REPLICATION MECHANISM

In topologies less dense than fully connected ones, it is expected that random walkers originating from a common initiator node to frequently revisit network nodes not only due to the probabilistic nature of the random walk mechanism (as it is the case for a fully connected topology), but also due to the topology characteristics. For example, in networks where bottleneck links are common, random walkers are expected to revisit certain network nodes more frequently than in topologies of no bottleneck links. Therefore, in such a network it is expected m random walkers to cover an almost overlapping network area (frequent revisits) at the beginning, before moving to distant (and likely not previously visited) areas. Therefore, instead of m distinct movements corresponding to the m random walkers, a macroscopic observer (most likely) would observe a number of distinct movements less than m, increasing (on average) as time increases.

In order to exploit this observation and proceed with a qualitative analysis, it is assumed that the underlying topology is a fully connected network (as before), in which random walkers move (and overlap) as it would have been the case if the underlying topology was not a fully connected one. The fully connected topology assumption is useful in order to simplify the analysis reusing results derived when proving Theorem 1. For the rest, $f(t)$ denotes the (average) fraction of random walkers seen by the macroscopic observer at time t. $mf(t)$ corresponds to the (average) number of distinct movements of random walkers in the network. In general, $f(0)$ is expected to be rather small and $f(t)$ to be close to 1 for large values of t. Assume that $f(t) = 1 - e^{-\alpha t}$, where α is a constant that varies depending on the characteristics of each environment (e.g., number of nodes, density, bottleneck links).

LEMMA 1. *In a fully connected network topology of N nodes and $mf(t)$ random walkers, coverage as a function of time t is given by:*

$$C_m(t) = 1 - e^{-\frac{m}{N}\left(t - \frac{1}{\alpha}\left(1 - e^{-\alpha t}\right)\right)}. \tag{3}$$

The proof can be found in Appendix B.

Figure 2 depicts $C_m(t)$ as given by Equation (3) and the first derivative of $C_m(t)$, $\frac{dC_m(t)}{dt}$ (derived in Appendix B). It is interesting to observe from Figure 2.a that $C_m(t)$ increases as time increases (as expected) but not that quickly. In particular, for small values of t, $C_m(t)$ increases slowly, then it reaches an inflection point at some point $t = t_0$. This is more clearly seen when observing Figure 2.b. It is obvious that depending on the value of α, $\frac{dC_m(t)}{dt}$ increases, assumes a maximum (the inflection point of $C_m(t)$) and then decreases. According to the analysis presented in Appendix B, the inflection point of $C_m(t)$ is assumed at $t_0 = -\frac{\ln \frac{2 + \frac{\alpha N}{m} - \sqrt{\left(2 + \frac{\alpha N}{m}\right)^2 - 4}}{2}}{\alpha}$ and $\lim_{\alpha \to +\infty} \frac{dC_m(t)}{dt}\Big|_{t=t_0} = \frac{m}{N}$.

The existence of the inflection point (later also confirmed by simulation results), is the basic motivation behind the introduction of a simple replication mechanism in the sequel. It is evident that due to revisits, a large number of random walkers may not always allow for significant coverage improvement, while at the same time an increased number of network resources (i.e., random walker movements) are wasted. Under replication, a small number of $m_0 \ll m$ random walkers is initially released at the initiator node and
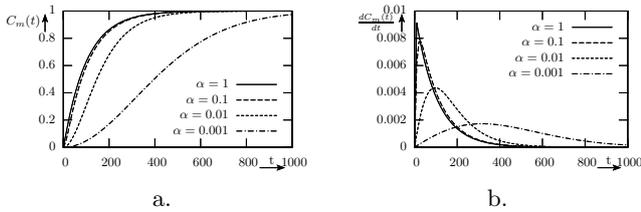
Figure 2: $C_m(t)$ **and** $\frac{dC_m(t)}{dt}$ **as a function of** t **for** $\frac{m}{N} = 0.01$ **and various values of** α.

afterwards, more random walkers are created by replicating the existing ones [7]. Special care is taken for the total number of random walkers in the network not to exceed m. Note that values of m_0 comparable to m eventually do not make any difference with respect to the problem of revisits since they reduce the replication mechanism to the multiple random walkers mechanism.

The replication mechanism: Having started with m_0 random walkers, for each random walker a replica is created with probability $1/q$ after each movement. All random walkers move in the network according to the multiple random walkers mechanism.

Parameter $q \geq 1$ corresponds to the (average) number of time units before a replica is created for each random walker. If q is small then random walkers are replicated fast, while large values of q permit random walkers to move further before replicated. The obvious gain of replication – as it will also be demonstrated in the following section using simulation results – is movement savings for as long as the number of random walkers remains smaller than m. It is easy to calculate that replication – taking place every (on average) q time units – doubles the number of random walkers in the network. Given that the total number should not exceed m and m_0 is the initial number of random walkers, at the n-th replication there will be $m_0 2^n$ random walkers in the network. Note that $m_0 2^n < m$, or $n < \log_2(m/m_0)$. From the previous it is concluded that replication will stop (on average) when $t = q \log_2(m/m_0)$. After this point, the number of movements is equal to m. However, until $t = q \log_2(m/m_0)$ there are fewer movements under replication than under the multiple random walkers mechanism. In particular, under the latter mechanism $mq \log_2(m/m_0)$, while under the former mechanism $m_0 q \sum_{i=0}^{\log_2(m/m_0)} 2^i$ movements take place, assuming that m/m_0 is an integer power of 2. For the rest, let $R(m/m_0)$ be the ratio of the previously mentioned number of movements, or, $R(m/m_0) = \frac{m_0 q \sum_{i=0}^{\log_2(m/m_0)} 2^i}{mq \log_2(m/m_0)}$,

$$R\left(\frac{m}{m_0}\right) = \frac{\sum_{i=0}^{\log_2 \frac{m}{m_0}} 2^i}{\frac{m}{m_0} \log_2 \frac{m}{m_0}}. \qquad (4)$$

Figure 3 depicts numerical results with respect to $R(m/m_0)$ as a function of m/m_0. If $m/m_0 = 1$ there is no gain ($R(m/m_0) = 1$). Note that $R(m/m_0) \leq 1$ for $m/m_0 \geq 1$.

It is interesting to observe from Equation (4) that $R(m/m_0)$ does not depend on q. This indicates that movement savings of the replication mechanism over the multiple random walkers mechanism depend only on how smaller is m_0 than m.
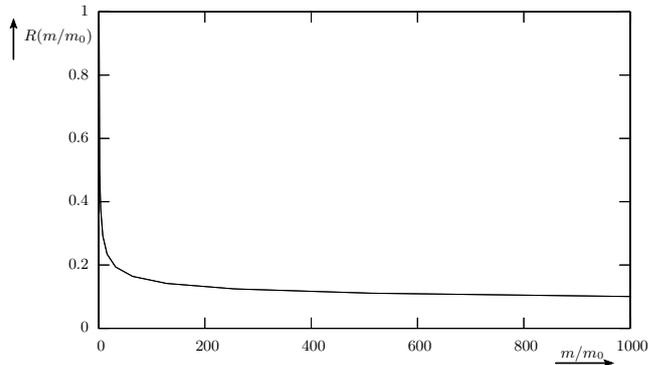

Figure 3: $R(m/m_0)$ **as a function of** m/m_0.

5. SIMULATION RESULTS

A simulation program exploiting the capabilities of the Omnet++ simulation platform [2] was created for the simulation purposes. The aim of the simulation results presented in this section is threefold: (i) to confirm the analytical findings of the previous sections; (ii) to shed more light on the specifics of the replication mechanism; and (iii) to compare it against the multiple random walkers mechanism for cases not covered by the analysis. The results derived in the sequel are averaged values of ten independent runs for network topologies of 1000 nodes unless otherwise mentioned.

The underlying topology considered during simulations is modelled as a geometric random graph [10]. It is constructed by uniformly scattering N nodes in the area $[0,1] \times [0,1]$ creating a link among two nodes if the euclidean distance among them is smaller than or equal to a certain constant r_c. Depending on the particular value of r_c, the resulting geometric random graph may be a connected one or not. For example, for $r_c \geq \sqrt{2}$, the resulting graph is the complete graph (all nodes connected to all other nodes). As r_c decreases, the number of links decreases until a certain threshold value $r_{c,0}$ such that if $r_c < r_{c,0}$, the geometric random graph is not connected anymore (even though for $r_c = r_{c,0}$, it is connected).

5.1 Multiple Random Walkers Mechanism

In order to evaluate the analytical results with respect to the multiple random walkers mechanism studied in Section 3 and particularly Theorem 1, r_c is set to $1.5 > \sqrt{2}$ in order to ensure that the underlying topology is fully connected. Figure 4 presents simulation results with respect to coverage $C_m(t)$ for various values of m. Note that the numerical results (based on the analysis) depicted in Figure 1 are almost identical to the simulation results depicted in Figure 4 (almost impossible to identify differences despite the zoom).

Figure 5 illustrates both simulation and numerical results (the latter after Equation (1)) regarding coverage for $m = 100$ and various network sizes. As before, it is obvious that the analysis captures the behavior observed by the simulations (lines correspond to numerical results based on the analysis and points to simulation results).

Figure 6.a depicts both simulation and numerical results with respect to the asymptotic termination time $\hat{T}_m$ as a function of m. The numerical results (depicted as a line) are derived based on Equation (2). Simulation results (depicted
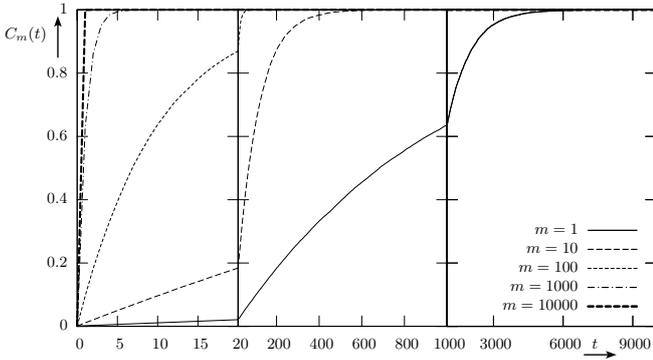
Figure 4: Simulation results for coverage ($C_m(t)$) as a function of t (zoom in different ranges of t) for a network of $N = 1000$ nodes and various values of m.
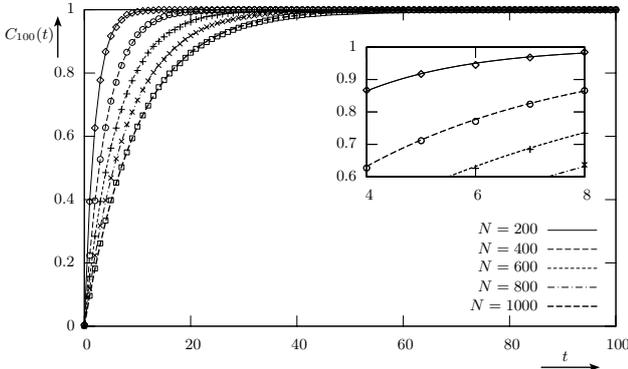

Figure 5: Simulation and numerical results for coverage as a function of t for $m = 100$ and various values of N.

as a line with points) are derived when $C_m(t) = 1 - \frac{1}{N}$ (all but one nodes covered). It is easy to see in the zoom area that both curves are significantly close. This is also the case with respect to both numerical and simulation results presented in Figure 6.b, where $\hat{T}_m$ is depicted as a function of N for various values of m. As before, it is clearly illustrated that the analytical findings capture the behavior observed by the simulations.

Figure 7 depicts simulation results with respect to coverage for a single random walker in the network for various values of r_c. It is interesting to observe that there exists a wide range of values of r_c for which the depicted curves are very close to the fully connected topology (i.e., $r_c = 1.5$). For example, the curve corresponding to $r_c = 1.0$ is difficult to be distinguished from the one for which $r_c = 1.5$. As r_c decreases further, it is evident that coverage increases slowly, thus resulting in larger termination times. This is expected since for small values of r_c topologies contain bottleneck links that force random walkers to increased revisits.

Figure 8 illustrates the same scenario as in Figure 7 for $m = 10, 100, 1000$ and 10000. The interesting part is that as m increases, $C_m(t)$ has an inflection point some time after the beginning after which random walkers are capable of moving in the network covering network nodes with increas-
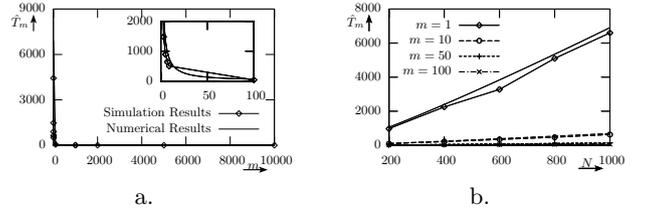

a.   b.

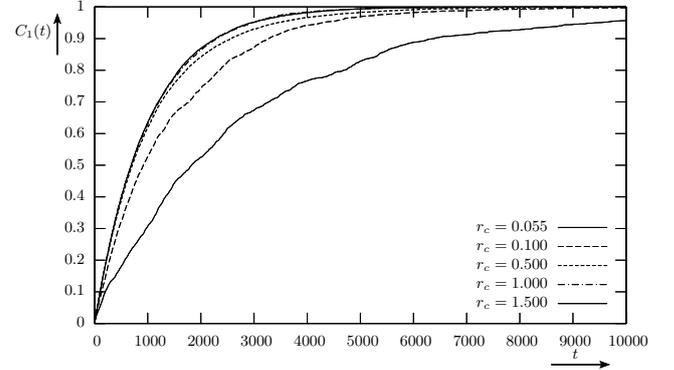**Figure 6: $\hat{T}_m$ as a function of (a) m and (b) r_c.**

Figure 7: $C_1(t)$ as a function of t for various values of r_c.

ing rate. This is in accordance with the analysis presented in Section 4 and particularly Equation (3), especially after observing the resemblance of the curves between Figure 8 and Figure 2.a. A direct comparison of the aforementioned figures does allow for further conclusions since Equation (3) has been derived under different approximations (e.g., no geometric random graph topologies). Future work will focus further on this aspect.

5.2 Replication Mechanism

The existence of the inflection point has been the motivation behind the introduction of replication in Section 4. In the sequel, assume random walkers move until a certain predefined number of movements is exhausted. This number of movements will be referred to hereafter as *budget*, denoted by H. For simplicity, assume H to be equal to $N \ln(N)$ for the rest of this work. Recall from Section 3 that $N \ln(N)$ corresponds to the asymptotic termination time for the case of one random walker in a fully connected topology. Given that multiple random walkers require less time to cover a fully connected topology, $N \ln(N)$ amounts to a significantly large budget for covering purposes. Under the multiple random walkers mechanism, this budget is divided equally among the m initial random walkers. Under the replication mechanism, it is divided equally among the m_0 initial random walkers and every time a replication occurs for a certain random walker its budget is equally divided among itself and the new one. For the rest $q = 1$ and $m_0 = 1$. That means that replication takes place too fast; on average after each random walker movement. Further investigation on various values of H, q and m_0 will be the focus of future work.
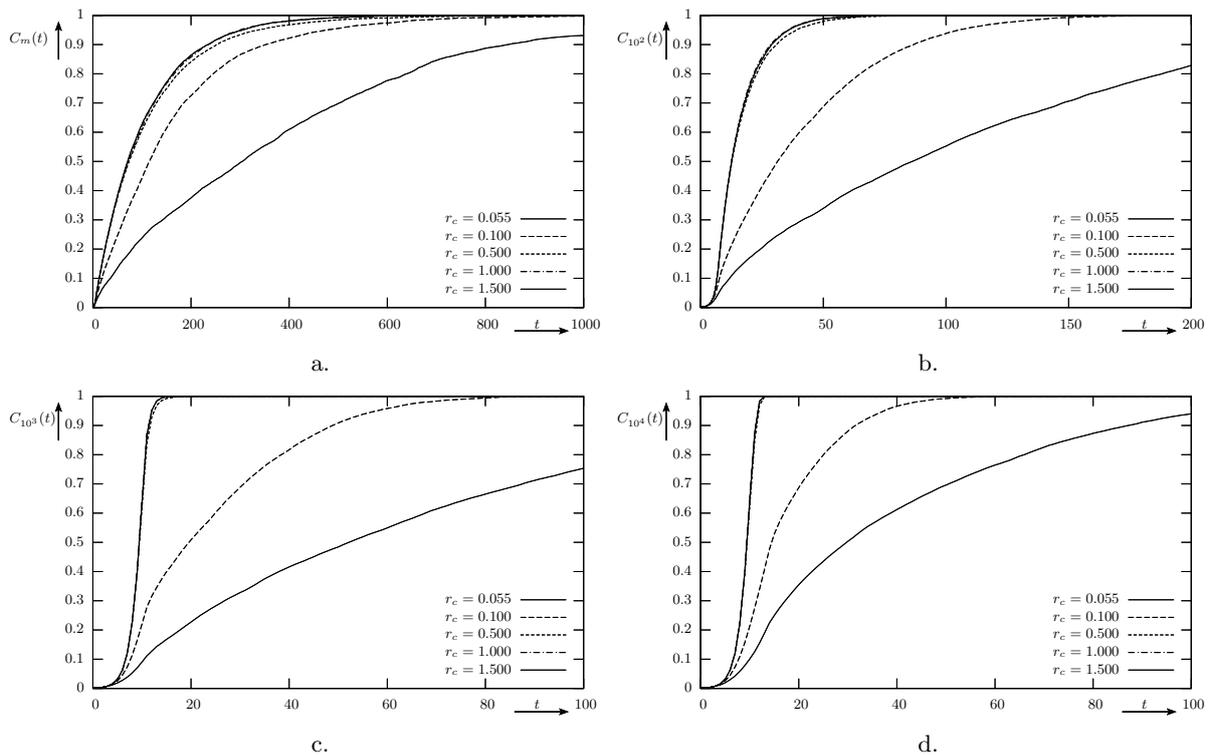
Figure 8: $C_m(t)$ $(m = 10, 10^2, 10^3 \text{ and } 10^4)$ **as a function of** t **for various values of** r_c.

Figure 9: $C_m(t)$ **as a function of** t **for (a)** $m = 50$ **and (b)** $m = 100$, **when** $H = 7000 \approx N \ln(N)$ **for** $N = 1000$ **and** $r_c = 0.055$.

Figure 10: $C_m(t)$ **as a function of** m **when** $H = 7000 \approx N \ln(N)$ **for** $N = 1000$ **and various values of** r_c.

In Figure 9, $C_m(t)$ is depicted for $H = 7000 \approx N \ln(N)$ ($N = 1000$), $r_c = 0.055$ and (a) $m = 50$ and (b) $m = 100$. These simulation results correspond to a single run (not averaged as before). It is interesting to see that for $m = 50$ (Figure 9.a), coverage under the replication mechanism is smaller than that under the multiple random walkers mechanism. The case depicted for $m = 100$ (Figure 9.b) is different since coverage under the replication mechanism eventually becomes larger than that under the multiple random walkers mechanism.

This behavior is more clearly depicted in Figure 10. Obviously, rather small values of m (e.g., $m = 1$) reduce replication to the multiple random walkers mechanism. As m increases (for $r_c = 0.055$ and $r_c = 0.100$) coverage under the replication mechanism becomes smaller that than under the multiple random walkers mechanism. However, as m increases further, the latter observation is reversed and
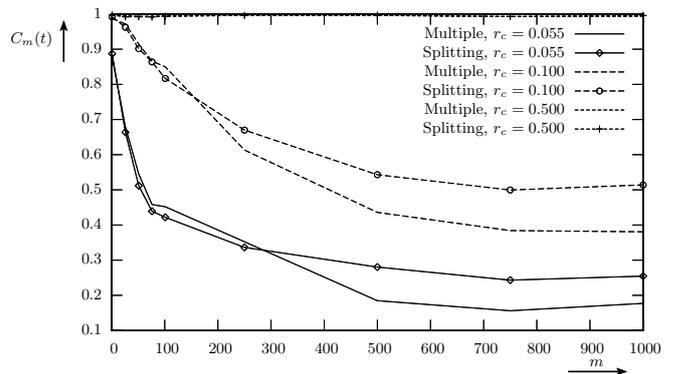
coverage under the replication mechanism becomes larger. This cannot be observed for $r_c = 0.500$ (or larger values of r_c) since coverage is close to 1. For these cases, the selected value of H is significantly large allowing for global network outreach under both mechanism.

6. CONCLUSIONS

In this paper information dissemination under multiple random walkers mechanisms has been studied and analyzed aiming at movements reduction thus preserving valuable network resources in mobile and opportunistic environments. The multiple random walkers mechanism [3] was studied and

analyzed from a novel perspective showing that multiple random walkers reduce termination time by a factor equal to their number in fully connected topologies. The derived analytical expressions were shown to be in accordance to the results presented in [3] and at the same time to provide for a more tractable form for further investigation.

For topologies less dense than fully connected ones, the topology characteristics (e.g., bottleneck links) force random walkers to increased revisits in addition to those due to the probabilistic nature of their movement. This effect is more obvious at the beginning of their movements and before random walkers are moved to different (and most likely not covered) network areas. The analytical results derived in this paper capture these observations revealing a certain inflation point after which coverage increment increases more rapidly due to reduced revisits. In order to reduce initial revisits, a simple replication mechanism was introduced allowing for movement savings.

Simulation results were also presented in order to evaluate the analytical findings. In addition, it was shown that under certain conditions studied here, coverage under the replication mechanism may become higher than that under the multiple random walkers mechanism.

Further study on the replication mechanism and its various parameters will be considered in future work in the area. Furthermore, a derivation of an analytical expression with respect to coverage under the multiple random walkers mechanism in geometric random graphs is expected to shed additional light on coverage with respect to topologies less dense than fully connected ones.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] Gnutella rfc.
http://rfc-gnutella.sourceforge.net/, 2002.

[2] Omnet++ simulation program.
http://www.omnetpp.org, 2009.

[3] N. Alon, C. Avin, M. Koucky, G. Kozma, Z. Lotker, and M. R. Tuttle. Many random walks are faster than one. In *SPAA '08: Proceedings of the twentieth annual symposium on Parallelism in algorithms and architectures*, pages 119–128, New York, NY, USA, 2008. ACM.

[4] F. Banaei-Kashani and C. Shahabi. Criticality-based analysis and design of unstructured peer-to-peer networks as "complex systems". In *CCGRID '03: Proceedings of the 3st International Symposium on Cluster Computing and the Grid*, page 351, Washington, DC, USA, 2003. IEEE Computer Society.

[5] C. Gkantsidis and M. Mihail. Hybrid search schemes for unstructured peer-to-peer networks. In *In Proceedings of IEEE INFOCOM*, pages 1526–1537, 2005.

[6] V. Kalogeraki, D. Gunopulos, and D. Zeinalipour-Yazti. A local search mechanism for peer-to-peer networks. In *CIKM '02: Proceedings of the eleventh international conference on Information and knowledge management*, pages 300–307, New York, NY, USA, 2002. ACM.

[7] D. Kogias, K. Oikonomou, and I. Stavrakakis. Study of randomly replicated random walks for information dissemination over various network topologies. In *Sixth Annual Conference on Wireless On demand Network Systems and Services (WONS)*, February 2-4 2009.

[8] Q. Lv, P. Cao, E. Cohen, K. Li, and S. Shenker. Search and replication in unstructured peer-to-peer networks. In *ICS '02: Proceedings of the 16th international conference on Supercomputing*, pages 84–95, New York, NY, USA, 2002. ACM.

[9] K. Oikonomou and I. Stavrakakis. Performance analysis of probabilistic flooding using random graphs. In *The First International IEEE WoWMoM Workshop on Autonomic and Opportunistic Communications (AOC 2007)*, pages 1–6, 18 June 2007.

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

[11] A. Segal. Distributed network protocols. *IEEE Trans. on Information Theory*, pages 23–35, 1983.

[12] A. O. Stauffer and V. C. Barbosa. Probabilistic heuristics for disseminating information in networks. *IEEE/ACM Trans. Netw.*, 15(2):425–435, 2007.

[13] D. Tsoumakos and N. Roussopoulos. Adaptive probabilistic search for peer-to-peer networks. In *P2P '03: Proceedings of the 3rd International Conference on Peer-to-Peer Computing*, page 102, Washington, DC, USA, 2003. IEEE Computer Society.

[14] L. Tzevelekas and I. Stavrakakis. Improving partial cover of random walks in large-scale wireless sensor networks. In *Proceedings of the 3rd IEEE WoWMoM AOC Workshop*, 15 June 2009.

APPENDIX

A.

In order to prove Theorem 1, let us consider two consecutive time instances t_1 and t_2 such that $t_2 = t_1 + 1$. Note that $C_m(t_2) \geq C_m(t_1)$ is satisfied. Between t_1 and t_2 all m random walkers have moved from one node to another neighbor node. Without loss of generality, let us name each random walker after a unique integer i within range $i \in [1, m]$. Let us now assume that the previously mentioned m individual random walker movements take place one by one within the $[t_1, t_2]$ interval and particularly at time $t_{1,1}, \ldots, t_{1,i}, \ldots, t_{1,m}$ ($t_1 = t_{1,1} < \cdots < t_{1,m} = t_2$), for each random walker $i \in [1, m]$. Let us $\tilde{C}_m(t)$ denote coverage within $[t_1, t_2]$, under the previous assumption that random walkers move in sequence. Given that any random walker movement is independent of any other one, eventually $\tilde{C}_m(t_{1,m}) = C_m(t_{1,m}) = C_m(t_2)$.

Under the previous assumption, at time $t_{1,1}$ random walker 1 moves from some node u to some neighbor node v. This movement takes place instantaneously at time $t_{1,1}$ such that at $^-t_{1,1}$ (i.e., just before $t_{1,1}$) random walker 1 is still at node u, while at $^+t_{1,1}$ it has moved at node v. Before moving at $t_{1,1}$, random walker 1 has $N - 2$ equivalent options: to visit

all network nodes minus the one that is currently located at and the one that it came from. Some of these $N - 2$ nodes may have already been visited by another random walker. The number of these (already visited) nodes is on average $(N - 2)(C_m(t_1) - 2/N)$ ($2/N$ corresponds to coverage regarding the current node and the one that the random walker came from). If random walker 1 moves to any one of these nodes, coverage will not increase. If, on the other hand, it moves to one of the rest $(N - 2)(1 - C_m(t_1))$ nodes, the coverage will increase by $1/N$. Therefore, on average random walker 1 will move to a previously not visited node with probability $(N - 2)/(N - 2)(1 - C_m(t_1)) = 1 - C_m(t_1)$. Eventually, $\tilde{C}_m(t_{1,1}) - C_m(t_1) = 1/N(1 - C_m(t_1))$. For random walker 2, $\tilde{C}_m(t_{1,2}) - \tilde{C}_m(t_{1,1}) = 1/N(1 - \tilde{C}_m(t_{1,1}))$. It is easy to derive that for random walker $i \in [1, m]$:

$$\tilde{C}_m(t_{1,i}) - \tilde{C}_m(t_{1,i-1}) = \frac{1}{N}\left(1 - \tilde{C}_m(t_{1,i-1})\right).$$

Recall that $\tilde{C}_m(t_{1,1}) = C_m(t_1)$ and $\tilde{C}_m(t_{1,m}) = C_m(t_2)$. By summing up for all values $i \in [1, m]$:

$$C_m(t_2) - C_m(t_1) = \frac{1}{N}\sum_{i=1}^{m}\left(1 - \tilde{C}_m(t_{1,i-1})\right).$$

The next step is to move from the discrete world to the continuous world, as it is common practice in the literature with respect to such problems. Note that $C_m(t_2) - C_m(t_1)$ can be written as $C_m(t + \Delta t) - C_m(t) = \frac{C_m(t+\Delta t) - C_m(t)}{\Delta t}$, where $\Delta t = t_2 - t_1 = 1$. By definition, the first derivative of $C_m(t)$, $\frac{dC_m(t)}{dt} = \lim_{\Delta t \to 0}\frac{C_m(t+\Delta t) - C_m(t)}{\Delta t}$. Therefore, $\frac{dC_m(t)}{dt} = \frac{1}{N}\lim_{\Delta t \to 0}\sum_{i=1}^{m}\left(1 - \tilde{C}_m(t_{1,i-1})\right)$, or, $\frac{dC_m(t)}{dt} = \frac{1}{N}\sum_{i=1}^{m}\lim_{\Delta t \to 0}\left(1 - \tilde{C}_m(t_{1,i-1})\right)$. Given that $\Delta t \to 0$, $t_{1,i} - t_{1,i-1} \to 0$. For such small time intervals that tend to be close to zero (i.e., $\Delta t \to 0$), it is reasonable to assume that $\lim_{\Delta t \to 0}\tilde{C}_m(t_{1,i-1})$ does not change (remains constant) for all $i \in [1, m]$ and eventually, $\lim_{\Delta t \to 0}\tilde{C}_m(t_{1,i-1}) = C_m(t)$, for any $i \in [1, m]$. Eventually,

$$\frac{dC_m(t)}{dt} = \frac{m}{N}\left(1 - C_m(t)\right).$$

Given that it is required $C_m(0) = 0$ to be satisfied and $\lim_{t \to +\infty} C_m(t) = 1$, coverage $C_m(t)$ is given by $C_m(t) = 1 - e^{-\frac{m}{N}t}$, and the theorem is proved.

B.

In order to prove Lemma 1, certain elements of the proof of Theorem 1 presented in Appendix A, are reused. In particular, with respect to the first derivative of coverage it can be shown that:

$$\frac{dC_m(t)}{dt} = \frac{m}{N}\left(1 - e^{-\alpha t}\right)\left(1 - C_m(t)\right).$$

The solution of this first order differential equation, given that $C_m(0) = 0$ and $\lim_{t \to +\infty} C_m(t) = 1$, is

$$C_m(t) = 1 - e^{-\frac{m}{N}\left(t - \frac{1}{\alpha}\left(1 - e^{-\alpha t}\right)\right)}.$$

In order to analyze further $C_m(t)$, the second derivative is derived and studied in the sequel. In particular, $\frac{d^2 C_m(t)}{d^2 t} = \frac{m}{N}\left(\alpha e^{-\alpha t} - \frac{m}{N}\left(1 - e^{-\alpha t}\right)^2\right)e^{-\frac{m}{N}\left(t - \frac{1}{\alpha}\left(1 - e^{-\alpha t}\right)\right)}$, which becomes zero, when $\alpha e^{-\alpha t} - \frac{m}{N}\left(1 - e^{-\alpha t}\right)^2 = 0$. The latter expression is equivalent to solving the following trinomial (with

respect to $e^{-\alpha t}$), $\left(e^{-\alpha t}\right)^2 - \left(2 + \frac{\alpha N}{m}\right)e^{-\alpha t} + 1 = 0$. There are two solution for this trinomial: $\frac{2 + \frac{\alpha N}{m} + \sqrt{\left(2 + \frac{\alpha N}{m}\right)^2 - 4}}{2}$ and $\frac{2 + \frac{\alpha N}{m} - \sqrt{\left(2 + \frac{\alpha N}{m}\right)^2 - 4}}{2}$. Given that $t \geq 0$ and $\alpha > 0$, $-\alpha t \leq 0$ and therefore, $e^{-\alpha t} \leq e^0 = 1$. It is easy to show that the first solution (i.e., $\frac{2 + \frac{\alpha N}{m} + \sqrt{\left(2 + \frac{\alpha N}{m}\right)^2 - 4}}{2}$) does not satisfy this inequality. In particular, it is enough to show that $\frac{2 + \frac{\alpha N}{m} + \sqrt{\left(2 + \frac{\alpha N}{m}\right)^2 - 4}}{2} > 1$, or $2 + \frac{\alpha N}{m} + \sqrt{\left(2 + \frac{\alpha N}{m}\right)^2 - 4} > 2$, or $\frac{\alpha N}{m} + \sqrt{\left(2 + \frac{\alpha N}{m}\right)^2 - 4} > 0$, that is always satisfied. In a similar manner it is easy to show that the second solution (i.e., $\frac{2 + \frac{\alpha N}{m} - \sqrt{\left(2 + \frac{\alpha N}{m}\right)^2 - 4}}{2}$) does satisfy $e^{-\alpha t} \leq 1$. It is enough to show that $\frac{2 + \frac{\alpha N}{m} - \sqrt{\left(2 + \frac{\alpha N}{m}\right)^2 - 4}}{2} > 1$, is never satisfied, or $\frac{\alpha N}{m} - \sqrt{\left(2 + \frac{\alpha N}{m}\right)^2 - 4} > 0$, or $\left(\frac{\alpha N}{m}\right)^2 > \left(2 + \frac{\alpha N}{m}\right)^2 - 4$, or $\left(\frac{\alpha N}{m}\right)^2 > 4 + \left(\frac{\alpha N}{m}\right)^2 + 4\frac{\alpha N}{m} - 4$, or $0 > 4\frac{\alpha N}{m}$, which is not satisfied. Eventually, $\frac{d^2 C_m(t)}{d^2 t} = 0$ when $t = t_0 = -\frac{\ln\frac{2 + \frac{\alpha N}{m} - \sqrt{\left(2 + \frac{\alpha N}{m}\right)^2 - 4}}{2}}{\alpha}$. Note that $\lim_{t \to t_0^-}\frac{d^2 C_m(t)}{d^2 t} > 0$ and $\lim_{t \to t_0^+}\frac{d^2 C_m(t)}{d^2 t} < 0$, therefore, at $t = t_0$, $\frac{dC_m(t)}{dt}$ assumes a maximum. $\frac{dC_m(t)}{dt}\big|_{t=t_0}$ is a function of α and it can be derived that $\lim_{\alpha \to +\infty}\frac{dC_m(t)}{dt}\bigg|_{t=t_0} = \frac{m}{N}$.