

Probabilistic Flooding for Efficient Information Dissemination in Random Graph Topologies¹

Konstantinos Oikonomou^{a,*}, Dimitrios Kogias^b,
Ioannis Stavrakakis^b

^a*Ionian University, Department of Informatics, Corfu, Greece*

^b*National and Kapodistrian University of Athens,
Department of Informatics & Telecommunications, Athens, Greece*

Abstract

Probabilistic flooding has been frequently considered as a suitable dissemination information approach for limiting the large message overhead associated with traditional (full) flooding approaches that are used to disseminate globally information in unstructured peer-to-peer and other networks. A key challenge in using probabilistic flooding is the determination of the *forwarding probability* so that global network outreach is achieved while keeping the message overhead as low as possible. In this paper, by showing that a *probabilistic flooding network*, generated by applying probabilistic flooding to a connected random graph network, can be (asymptotically) “bounded” by properly parameterized random graph networks and by invoking random graph theory results, asymptotic values of the *forwarding probability* are derived guaranteeing (probabilistically) successful *coverage*, while significantly reducing the message overhead with respect to traditional flooding. Asymptotic expressions with respect to the average number of messages and the average time required to complete network coverage are also derived, illustrating the benefits of the properly parameterized probabilistic flooding scheme. Simulation results support the claims and expectations of the analytical results and reveal certain aspects of probabilistic flooding not covered by the analysis.

Key words: Probabilistic Flooding, Information Dissemination, Random Graphs.

* Corresponding Author. Address: Ionian University, Department of Informatics, Tsirigoti Square 7, 49100 Corfu, Greece. Tel: +30 26610 87708, Fax: +30 26610 48491.

Email addresses: okon@ionio.gr (Konstantinos Oikonomou),
dimkog@di.uoa.gr (Dimitrios Kogias), ioannis@di.uoa.gr (Ioannis Stavrakakis).

¹ This work has been supported in part by the project ANA (Autonomic Network

1 Introduction

A plethora of network protocols requiring information dissemination (such as routing broadcasting, content discovery etc) are based on *traditional flooding* since its early introduction, [1]. Traditional flooding is capable of sending messages (one message for each neighbor node apart from the node that the message arrived from) and traversing *all network links* and *all network nodes* in any connected network topology in a small number of (time) steps and it is simple and easy to implement in a distributed fashion. The inherent simplicity of traditional flooding as well as its guaranteed success in disseminating information everywhere, have been two of the main reasons for its wide-spread protocol instantiations. However, by traversing all network links, the *number of messages* forwarded in the network is (asymptotically) of the order of the number of the network links (actually, more than the number of network links and less than twice this number). Therefore, even though *termination time* (that is, the time required to outreach all network nodes) is small (equal to the *eccentricity* of each node and upper bounded by the *network diameter*), the aforementioned increased number of messages is prohibitive in modern network environments that are typically large-scale with respect to the number of nodes and network links.

Typical examples of a large-scale modern network environment are *Peer-to-Peer* (P2P) networks, [2], which, following the Napster legacy, [3], offer to network users huge amounts of content of, basically, delay tolerant services (e.g., files). While, the underlying structure present in *structured* P2P networks facilitates reaching the nodes with relatively low delay and message overhead, [2], [4], [5], [6], the problem of *disseminating information* is a major challenge in *unstructured* P2P networks, (e.g. Gnutella, [7]), as there is no structure to take advantage of and support the design of an effective scheme. As a result, a brute-force approach is followed, typically implemented through resource wasteful approaches such as traditional flooding and some of its message reducing variations, [2], [7], [8], [9], [10], [11]. For example, in Gnutella, [7], *controlled flooding* is employed for searching purposes restricting message flooding to a small number of L hops around the node that has initiated the search. Such an approach – referred to hereafter as *L-flooding* – is scalable in terms of the number of messages for small values of L but, at the same time, it reduces the covered network area; the use of this approach would reduce the probability of success for the searching process in [7], particularly in large-scale networks.

Architecture) (IST-27489), SOCIALNETS (IST-217141) and 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%).

Message reduction during an information dissemination process is a key goal in many protocol designs, [12], [13]. The idea of reducing the messages of traditional flooding by forwarding a message to neighbors *probabilistically* (with some *forwarding probability*) and not *deterministically* lays behind *probabilistic flooding*, [14], [15], [16], [17], [18], [19], [20]. There is clearly a trade-off between the induced total number of messages and the network *coverage* (i.e., the fraction of network nodes reached by such messages): the smaller the forwarding probability, the smaller the message overhead and the smaller the network coverage.

The work presented here investigates probabilistic flooding when the underlying network is a *random graph*, [21], and aims at designing such a scheme in a way that the aforementioned trade-off is effectively managed. That is, achieve high network coverage with a relatively small number of messages. Analytical tools and results, borrowed from *random graph theory*, [21], [22], [23], [24], [25], [26], are considered for analyzing probabilistic flooding and comparing its performance against traditional flooding. Even though it has been shown that some networks do not follow the random graph model but rather power-law models, e.g., [23], [24], it is still valid that many useful results may be extracted after studying random graph topologies, e.g., [17], [27], [28].

One of the main contributions of this work is establishing a connection between random graphs and the *probabilistic flooding network*; the latter is defined to be the network consisted of the (sub)set of links and nodes of the underlying random graph network that are traversed by the messages under the probabilistic flooding. This allows for the subsequent study of probabilistic flooding in random graph topologies based on analytical tools borrowed from random graph theory.

Another contribution of this work is the derivation of (asymptotic) analytical expressions on the *appropriate* value of the forwarding probability, defined to be the value for which the probabilistic flooding network consists (with high probability) of a certain number of network nodes (e.g., all network nodes) using the smallest possible number of messages. The case of reaching all network nodes – which is equivalent to full network coverage or *global network outreach* – is extensively studied here.

The (asymptotic) analysis has shown a significant decrease with respect to the number of messages, reduced from the order of $\Theta(N^2)$ (i.e., traditional flooding) to the order of $\Theta(N \ln(N))$ (i.e., probabilistic flooding) in random graph topologies of N nodes, for the case of global network outreach. Furthermore, as it is analytically shown in this paper, the price paid for this reduction is (a) an increase of the time required to cover a certain network area; (b) any network coverage (e.g., global node outreach) is achieved with high probability under probabilistic flooding as opposed to certainty under traditional flooding.

These analytical findings are supported by simulation results.

The analytical study has been extended to the case of L -coverage, achieved when no node is away from a reached node by more than L -hops. This notion of coverage (yielding overhead savings) is applicable to various practical cases such as when supplementing L -hop flooding (for small L) allows every node to obtain the disseminating information, as it is the case under Gnutella mentioned before. The cases of $L = 0$, $L = 1$ and $L = 2$ are studied here (the case of $L = 0$ is equivalent to global network outreach mentioned before) and it is analytically shown that there is a significant reduction with respect to the number of messages when probabilistic flooding is employed and an L -coverage is targeted. Given the *small-world phenomenon* in random graphs (i.e., significantly small diameters even for large network sizes), [22], [26], larger values of L would have been significantly close to the network diameter for network sizes of 10000 nodes (as it will be mentioned in Section 6, the simulation program had memory and performance limitations for network sizes greater than 10000 nodes). For example, when L is equal to the network diameter, information dissemination is obsolete since all nodes are by definition less than or equal to L -hops away from any node, and thus from the source node as well. Such a case is investigated at the simulations section.

Simulation results confirm the analytical findings of this paper and at the same time explore various aspects not covered by the analysis. In particular, it is shown that the probabilistic flooding network is (asymptotically) “bounded” by two random graphs and that savings under probabilistic flooding are significant when compared to traditional flooding (or controlled flooding in the case of $L = 1, 2$ for *fairness* issues as it is explained in more detail in the simulations section). On the other hand, it is shown that termination time does increase (even though at a smaller rate compared to message savings) as it is also expected by the analysis. However, due to the “small-world phenomenon” in random graphs (i.e., small network diameters even for large network sizes), when L is comparable to the network diameter, simulation results deviate from the analytical ones. For the simulation scenarios of 10^4 nodes considered in this work, $L = 2$ is comparable to the network diameter. However, larger network sizes (e.g., 10^8 network nodes) could not be considered due to computational and memory limitations. This issue is further explained in the simulations section.

In the following section, past related works are presented and particularly those mostly related to this paper. Section 3 summarizes important results from random graph theory that will be used throughout this work. Section 4 presents the probabilistic flooding scheme and discusses its relevance to random graphs. Analytical (asymptotic) results are presented in Section 5, and simulation results in Section 6. The conclusions are drawn in Section 7.

2 Past Related Work

The effectiveness of most of the works (including the one presented in this paper) attempting to reduce the number of messages sent in a network, is frequently established through comparison with traditional flooding. Approaches, like random walkers, e.g., [8], [10], may be seen as an entirely different class which -while typically achieving low message overhead- results in large network cover times. Hybrid probabilistic schemes (e.g., a local flooding process initiated after a random walk) have also been proposed and analyzed, [10], as well as other schemes that adapt the employed values of L for L -flooding in a probabilistic manner, [11]. Another modification, [9], allows for network nodes to *selectively* forward messages to their neighbors, thus significantly reducing the number of messages in the network. Other approaches for filling the gap between random walkers and traditional flooding have also been proposed. For example, in [29] replicated random walker strategies have been studied for various network topologies. Gossip strategies, under which messages are forwarded based on local information, e.g., [12], [30], [31], [32], [33], [34], [35], have also been on the edge of this research area for many years, e.g., for the design of network protocols.

Probabilistic flooding has attracted considerable attention since its introduction, basically for providing scalable dissemination information approaches in modern large-scale environments like P2P networks, e.g., [14], [15], [16]. Even though – as already mentioned – its main disadvantage with respect to traditional flooding is its probabilistic nature, it may be employed in cases that probabilistic guarantees are affordable given the subsequent savings with respect to the number of messages. These trade-offs have been the focus of recent studies, e.g., [18], [19], [20].

The most relevant works to the work presented here are those in [19] and [20]. In [19], Stauffer and Barbosa compare probabilistic flooding against a proposed *heuristic flooding* that instead of employing a fixed forwarding probability for all network nodes – as it is the case under probabilistic flooding – it adapts accordingly based on a heuristic rule. In [20], Crisostomo, Schilcher, Bettstetter and Barros, consider random graph topologies and – based on dominating set properties – provide for suitable values of global network outreach based on both analytical and simulation results.

Both [19] and [20] mention the relevance between random graphs and probabilistic flooding in random graph topologies. However, none of these approaches either investigate this further or exploit this in their work, as it is the case in this paper. Furthermore, the work in this paper focuses in more general cases than global network outreach with respect to covering (notion of L -coverage), that allow for further reduction of the messages forwarded in

the network.

3 The Random Graph Model

A network topology is typically represented by a graph \mathbb{G} defined in terms of a set of nodes $\mathcal{V}(\mathbb{G})$ and a set of (bidirectional) links $\mathcal{E}(\mathbb{G})$ connecting pairs of nodes in $\mathcal{V}(\mathbb{G})$. Let the number of nodes in the network be denoted by N ($N = |\mathcal{V}(\mathbb{G})|$, where $|X|$ corresponds to the number of elements of a certain set X). Random graphs, mainly introduced by the pioneering work of P. Erdős and A. Rényi, [21], have certain properties that are useful in shedding light on the particulars of various networks like the Internet, social contact networks, biological networks, web-page networks etc., [23], [24]. For most of the cases, random graphs have been considered for studying topological aspects of these networks (e.g., the average number of neighbors, the diameter of the network, the emergence of the giant component etc.). In this work, random graphs and their properties are used for understanding and analyzing the information dissemination network created after performing probabilistic flooding in random graph network topologies.

In the random graph literature there are two basic models for representing a random graph of N nodes: the *binomial* and the *uniform*. The first model – denoted as $\mathbb{G}(N, p(N))$ – assumes a certain probability $p(N)$ for selecting each link out of the $\binom{N}{2}$ possible links to become part of the particular random graph $\mathbb{G}(N, p(N))$ (in order to simplify the notation, in the sequel $\mathbb{G}(N, p)$ will be used instead of $\mathbb{G}(N, p(N))$). Under the second model, a graph of N nodes and M links is uniformly chosen out of the $\binom{N}{M}$ possible different graphs of N nodes and M links and is denoted as $\mathbb{G}(N, M)$. Both models are asymptotically equivalent provided that $p(N)\binom{N}{2}$ is close to M , [26], [22]. For the rest of this paper, the binomial model $\mathbb{G}(N, p)$ is considered. For this model, the *expected number of links* is equal to $p(N)\binom{N}{2}$ and therefore, in order to ensure the equivalence of both previous models, it is assumed that $p(N)\binom{N}{2} \approx M$. Furthermore, it is also assumed that N is significantly large (e.g., such that $p(N)(N-1) \approx p(N)N$ is satisfied), which is a basic assumption in the literature of random graphs, e.g., [22], [26].

Based on the previously mentioned model description, it is easily derived that for $p(N) = 0$, there are no links in the resulting graph ($M = 0$) and $\mathbb{G}(N, p)$ consists only of N nodes with no links among them (Figure 1.a), whereas for $p(N) = 1$, the resulting graph is the *complete graph* ($M = \binom{N}{2}$). For values of $p(N)$ slightly higher than 0 (always under the assumption that N is fairly large), some links start to be created among some nodes, creating *small components* of a few nodes (Figure 1.b). Note that Figure 1 illustrates

the various stages of a random graph of N nodes – as probability $p(N)$ (or the number of links M) increases – using 2-dimension drawings for better visualization. In general, a random graph is not expected to be a 2-dimensional graph, [22].

For many properties of random graphs (and particularly for those useful for the subsequent analysis of probabilistic flooding), it has been shown in the literature that there exists a *critical probability* $p_Q(N)$ – for some property Q – such that if $p(N)$ grows faster than $p_Q(N)$ as N increases, then $\mathbb{G}(N, p)$ has property Q , otherwise it does not. More formally, [23], [25],

$$\lim_{N \rightarrow \infty} Pr\{\mathbb{G}(N, p) \text{ has property } Q\} = \begin{cases} 0 & \text{if } \frac{p(N)}{p_Q(N)} \rightarrow 0, \\ 1 & \text{if } \frac{p(N)}{p_Q(N)} \rightarrow +\infty. \end{cases} \quad (1)$$

In the sequel and based on this definition, any argument that a random graph has property Q is made *with high probability* (w.h.p.). For the rest of this paper, the following well-known asymptotic notations will be used:

- a. $p(N) = O(p_Q(N))$, if $p(N) \leq Cp_Q(N)$ for some constant $C > 0$ (equivalently, $\lim_{N \rightarrow \infty} \frac{p(N)}{p_Q(N)} \rightarrow 0 \Leftrightarrow O(p(N)) < O(p_Q(N))$, [37]);
- b. $p(N) = \Omega(p_Q(N))$, if $p(N) \geq cp_Q(N)$ for some constant $c > 0$ (equivalently, $\lim_{N \rightarrow \infty} \frac{p(N)}{p_Q(N)} \rightarrow +\infty \Leftrightarrow O(p(N)) > O(p_Q(N))$, [37]);
- c. $p(N) = \Theta(p_Q(N))$, if $Cp_Q(N) \geq p(N) \geq cp_Q(N)$ for some constants $C, c > 0$.

Note that for some property Q and $p(N)$ such that $p(N) = \Theta(p_Q(N))$, $\lim_{N \rightarrow \infty} \frac{p(N)}{p_Q(N)} = c'$, [37], where c' is a positive constant ($0 < c' < +\infty$). Therefore, the case of $p(N) = \Theta(p_Q(N))$ is not covered by the definition given by Equation (1). The basic reason is that for $p(N) = \Theta(p_Q(N))$, the graph passes a *phase transition* period, [22], [26], [23], during which it “moves” from a state in which it does not have property Q to a state in which it does have property Q and therefore, it is not known for certain (not even with high probability) whether it has or not property Q . For the rest of this paper, notation $p(N) = \Theta(p_Q(N))$ is used to indicate such a phase transition with respect to property Q (e.g., the emergence of the giant component that indicates that the network becomes connected).

In this paper the interest is on three properties (Q_0 , Q_1 and Q_2) of random graphs: (a) Q_0 : the graph is connected; (b) Q_1 : the graph consists of a *giant component* and some isolated nodes; and (c) Q_2 : the giant component has just emerged. The corresponding critical probabilities ($\frac{\ln(N)}{N}$, $\frac{\ln(N)}{2N}$, and $\frac{1}{N}$, respectively) as a function of N for each property are shown in Table 1, [22], [26]. Note that, $\frac{\ln(N)}{N} > \frac{\ln(N)}{2N} > \frac{1}{N}$, for large values of N . The analysis of

probabilistic flooding in random graph topologies presented in the following sections is based on the emergence of these three properties.

Table 1

Critical Probability Functions for Properties Q_0 , Q_1 , and Q_2 .

Critical Function	Probability	Property Description
$p_{Q_2}(N) = \frac{1}{N}$		Q_2 : Emergence of giant component. Some components are also present along with some isolated nodes.
$p_{Q_1}(N) = \frac{\ln(N)}{2N}$		Q_1 : Giant component and some isolated nodes.
$p_{Q_0}(N) = \frac{\ln(N)}{N}$		Q_0 : Connected network.

Figures 1.c and 1.d depict a 2-dimensional illustration of these properties. Note that in general random graphs are not 2-dimensional as depicted here for illustration purposes. In Figure 1.c, where $O(p(N)) < O(p_{Q_2}(N))$, the graph consists of numerous isolated components. These isolated components become larger as $p(N)$ increases and at some point, the largest components are so large that it is likely a new link to merge two of them resulting in a new, larger (merged) component which, in turn, is likely to be further increased by a new link, etc. Eventually, for $p(N) = \Theta(p_{Q_2}(N))$ the giant component emerges, as depicted in Figure 1.d. As $p(N)$ increases further such that $O(p(N)) > O(p_{Q_2}(N))$ and $O(p(N)) < O(p_{Q_1}(N))$, the giant component increases further by connecting to the smaller components (i.e., graph has property Q_2 and not property Q_1 w.h.p.). For $p(N) = \Theta(p_{Q_1}(N))$, there exist only the giant component and some isolated nodes, as depicted in Figure 1.e. Further increase of $p(N)$ such that $O(p(N)) > O(p_{Q_1}(N))$ and $O(p(N)) < O(p_{Q_0}(N))$ connects the giant component to the isolated nodes (i.e., graph has property Q_1 and not property Q_0 w.h.p.). Finally, the network becomes connected w.h.p. for $p(N) = \Theta(p_{Q_0}(N))$, as shown in Figure 1.f, and remains connected thereafter, for $O(p(N)) > O(p_{Q_0}(N))$.

The *diameter* (i.e., the maximum shortest path of all pairs of nodes in the same component) of a random graph $\mathbb{G}(N, p)$, denoted as $\mathcal{D}(\mathbb{G}(N, p))$, is also important for the subsequent analysis of probabilistic flooding. For $O(p(N)) > O(p_{Q_0}(N))$ is known, [23], [36], to be proportional to:

$$D(\mathbb{G}(N, p)) = \frac{\ln(N)}{\ln(p(N)N)}. \quad (2)$$

For the rest of this paper, it is assumed that $\mathcal{D}(\mathbb{G}(N, p)) = \Theta\left(\frac{\ln(N)}{\ln(p(N)N)}\right)$, when $O(p(N)) > O(p_{Q_0}(N))$. For the case of $O(p(N)) < O(p_{Q_0}(N))$ and $O(p(N)) > O(p_{Q_2}(N))$, the network diameter is equal to the diameter of the giant component, (the network is not connected and there exists a giant component w.h.p.). Since the number of nodes in the giant component are

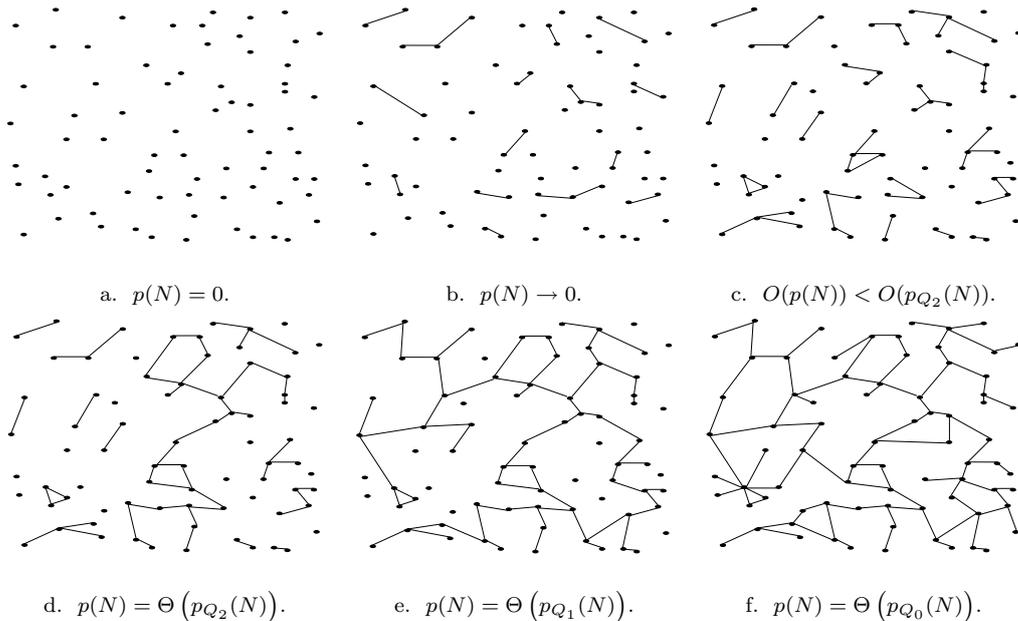


Fig. 1. An example of a random graph $\mathbb{G}(N, p)$ (depicted in two dimensions).

of the order of $O(N)$, [26], [23], a very rough idea about the diameter of the giant component is given by $O(\ln(N))$ (for this case the giant component looks mostly like a tree even though some cycles do exist).

4 Probabilistic Flooding and Random Graphs

4.1 On Probabilistic Flooding

Under traditional flooding, messages arrive to all nodes having traversed all network links, by sending one message to each neighbor node each time a node receives the forwarding message for the first time (excluding the particular neighbor node that the message has arrived from). The total number of messages under traditional flooding can be greater than the number of links of the network topology since either one or – at most – two messages is possible to be forwarded over any link. Therefore, the *number of messages* under traditional flooding over a network topology \mathbb{G} is of the order of $\Theta(|\mathcal{E}|)$ (greater than $|\mathcal{E}|$ and less than $2|\mathcal{E}|$). This number turns out to be fairly large for large $N = |\mathcal{V}(\mathbb{G})|$ resulting in the scalability problems of traditional flooding. For instance, for a network topology modelled as a random graph, $\mathbb{G}(N, p)$, the number of messages under traditional flooding is of the order of $\Theta\left(p(N)\binom{N}{2}\right) = \Theta\left(p(N)N\frac{N-1}{2}\right)$ which becomes very large for large networks. This large message overhead leads, on the other hand, to relatively short *termination time* (defined as the number of time steps until the message

dissemination is completed) bounded by the network diameter $\mathcal{D}(\mathbb{G}(N, p))$ (i.e., $O(\mathcal{D}(\mathbb{G}(N, p)))$).

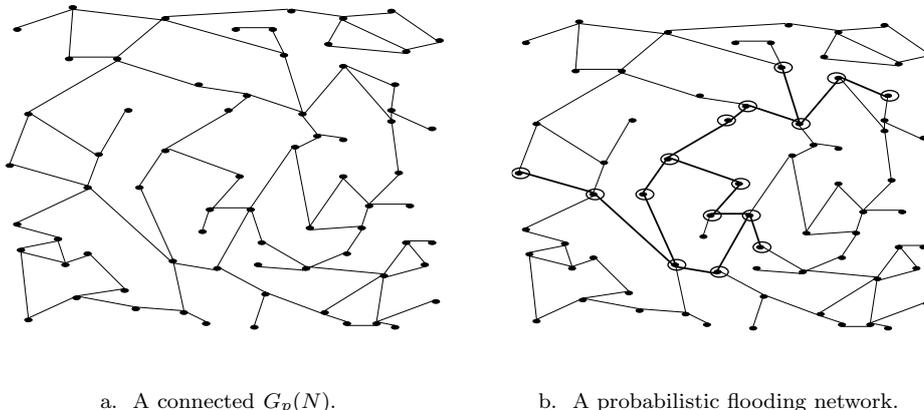


Fig. 2. An example of nodes (marked with ellipses) reached by probabilistic flooding messages and the corresponding links (marked with thick lines) that these messages have been forwarded over.

Under probabilistic flooding, [14], the *initiator* node (i.e., the source of the information to be disseminating) sends a message to each of its neighbor nodes with an (independent) constant *forwarding probability* $p_f(N)$. Any node receiving such a message forwards it to its neighbor nodes (apart from the one the message has arrived from) with probability $p_f(N)$. Clearly, for $p_f(N) = 0$, no messages are sent in the network, while for $p_f(N) = 1$, probabilistic flooding reduces to traditional flooding. As a result of probabilistic flooding, a network can be defined that consists: (a) of the *set of nodes* that have been reached by the messages plus the initiator node; and (b) of the *set of links* over which these messages have been forwarded. This particular network will be referred to hereafter as the *probabilistic flooding network*; an example of such a network is depicted in Figure 2. It is trivial to show (based on the definition of probabilistic flooding) that a probabilistic flooding network is actually a connected network each link of which is created as a result of message forwarding over the corresponding network link. Let $\mathbb{P}(\mathbb{G}(N, p), p_f)$ denote the probabilistic flooding network generated by applying probabilistic flooding with probability $p_f(N)$ over network $\mathbb{G}(N, p)$.

An observation, that will be utilized later, is that some of the links of $\mathbb{G}(N, p)$ become part of $\mathbb{P}(\mathbb{G}(N, p), p_f)$ with probability $p_f(N)$ if only one of the end nodes receives the message from an other link and take a forwarding decision, some with probability $\tilde{p}(N)$ if both end nodes of the link receive the message from a different than their common link and, thus, both make an (assumed independent) decision that each fails with probability $1 - p_f(N)$ and some with probability 0 if none of end nodes receives the message.

From the above it can be proved that $\tilde{p}(N)$ equals:

$$\tilde{p}(N) = 1 - (1 - p_f(N))(1 - p_f(N)) = 2p_f(N) - p_f^2(N). \quad (3)$$

4.2 *L-Coverage of a network*

The network coverage is typically defined as the fraction of network nodes that are forwarded a message, or “covered ” by an information dissemination process. In this paper (as also in [29]) the notion of coverage is extended to that of *L-Coverage* (L an integer and $L \geq 0$) referring to the portion of network nodes which are at most L hops away from a node that has received the message and denoted by $C(L)$. Clearly, $C(0)$ corresponds to the portion of nodes that have actually received the message and if $C(0) = 1$ a global outreach has been achieved. Notice that $C(L)$ is non-decreasing with L . Notice, also, that for traditional flooding $C(0) = 1$ and, thus, $C(L) = 1$ for any $L \geq 0$ as well.

Since random graph networks have small diameters (known as the small-world phenomenon), [23], [26], the interest is basically for coverage of $L = 0$, $L = 1$ and $L = 2$ and the main focus is to reduce as much as possible the number of messages required to achieve $C(L) = 1$, for such values of L . It should be mentioned that probabilistic flooding employs a probabilistic mechanism and therefore any claim that “a certain coverage $C(L) = 1$ ” is always made *with high probability*.

4.3 *Probabilistic Flooding in Random Graphs*

The objective when constructing a probabilistic flooding network is to keep the number of its links (i.e., number of forwarded messages) as low as possible, while striving to achieve a certain coverage $C(L) \rightarrow 1$. Clearly (and assuming always that $\mathbb{G}(N, p)$ is connected), for $p_f(N) = 1$ the probabilistic flooding network contains all network nodes and all network links of $\mathbb{G}(N, p)$. That is, $\mathbb{P}(\mathbb{G}(N, p), 1) = \mathbb{G}(N, p)$. The interest in the sequel of this section is on determining a value of $p_f(N)$ (to keep the number of messages under probabilistic flooding low) such that $C(L) = 1$ w.h.p., for $L = 0, 1, 2$.

Given that each node of a connected network is associated with at least one link and most likely with several, removing a link from a network does not necessarily disconnect (or “remove”) an associated node as well. In other words, the decrease in the number of nodes in a network as a result of a decrease in the number of links is expected to be lower than the decrease in the number of links. Consequently, it is conceivable that all network nodes continue to be

included in a network (i.e., be connected) despite the removal of a number of links w.h.p. This observation suggests that a probabilistic flooding network with sufficiently high forwarding probability may still keep all the nodes “connected” and in the network, despite a potentially significant removal of links due to a decision not to forward a message over such links.

In view of the above discussion it is evident that as $p_f(N)$ decreases, the number of links in $\mathbb{P}(\mathbb{G}(N, p), p_f)$ decreases as well, while the number of nodes in $\mathbb{P}(\mathbb{G}(N, p), p_f)$ decreases at a lower rate. Consequently, for a small reduction in $p_f(N)$ below the value of 1, it is expected that all network nodes still be included in $\mathbb{P}(\mathbb{G}(N, p), p_f)$, or $C(0) = 1$ w.h.p. Further reduction in the value of $p_f(N)$ will result in a further decrease in the number of network links, until a small number of nodes of $\mathbb{G}(N, p)$ do not belong in $\mathbb{P}(\mathbb{G}(N, p), p_f)$ (see also Section 3). It is thus expected that there is a certain value for the forwarding probability, denoted by $p_{f,0}(N)$, such that: (a) if $O(p_f(N)) < O(p_{f,0}(N))$, then the probabilistic flooding network does not include all network nodes w.h.p.; (b) if $O(p_f(N)) > O(p_{f,0}(N))$, then the probabilistic flooding network does include all network nodes w.h.p.

The goal in the sequel is to derive an asymptotic expression for $p_{f,0}(N)$ by utilizing properties of random graphs (those already described in Section 3), taking into consideration that since $C(0) = 1$ w.h.p all network nodes are expected to be part of the probabilistic flooding network and, in view of the latter and as discussed in 4.1, each of the links of $\mathbb{G}(N, p)$ will be included in $\mathbb{P}(\mathbb{G}(N, p), p_f)$ either with probability $p_f(N)$ or $\tilde{p}(N)$. Thus, it is evident that $\mathbb{P}(\mathbb{G}(N, p), p_f)$ contains on average more links than $\mathbb{G}(N, p \times p_f)$. Consequently, when $\mathbb{G}(N, p \times p_f)$ is connected w.h.p., then $\mathbb{P}(\mathbb{G}(N, p), p_f)$ is also connected w.h.p. and, thus, includes all network nodes w.h.p. Note also that since $p_f(N) \leq 2p_f(N) - p_f^2(N) = \tilde{p}(N)$ (the equality holds for $p_f(N) = 1$), $\mathbb{G}(N, p \times \tilde{p})$ contains (on average) more links than $\mathbb{G}(N, p \times p_f)$ and when the latter network is connected the former is also connected w.h.p.

Based on the previous two observations, and assuming a certain value for $p(N)$, as $p_f(N)$ increases, it is expected that there will be some probability value for $p_f(N)$ for which $\mathbb{G}(N, p \times \tilde{p})$ becomes connected w.h.p. As $p_f(N)$ increases further, $\mathbb{P}(\mathbb{G}(N, p), p_f)$ becomes connected (equivalently, $C(0) = 1$) w.h.p. For further increment, $\mathbb{G}(N, p \times p_f)$ also becomes connected. Consequently, the particular value of $p_f(N)$ (i.e., $p_{f,0}(N)$ for $C(0) = 1$) for which probabilistic flooding disseminates information to all network nodes (global network outreach) is “between” the values of $p_f(N)$ for which $\mathbb{G}(N, p \times \tilde{p})$ and $\mathbb{G}(N, p \times p_f)$ become connected. This is demonstrated later in Section 6 using simulation results.

From the discussion in Section 3 it follows that random graph $\mathbb{G}(N, p \times p_f)$ becomes connected (or has property Q_0) w.h.p. when $p(N)p_f(N) = \Theta(p_{Q_0}(N))$,

or $p_f(N) = \Theta\left(\frac{p_{Q_0}(N)}{p(N)}\right)$. On the other hand, random graph $\mathbb{G}(N, p \times \tilde{p})$ becomes connected when $p(N)\tilde{p}(N) = \Theta(p_{Q_0}(N))$, or $\tilde{p}(N) = \Theta\left(\frac{p_{Q_0}(N)}{p(N)}\right)$, or $p_f(N) = \Theta\left(1 - \sqrt{1 - \frac{p_{Q_0}(N)}{p(N)}}\right)$ (since $\tilde{p}(N) = 2p_f(N) - p_f^2(N)$ and $O(p(N)) > O(p_{Q_0}(N))$) in order to ensure connectivity of $\mathbb{G}(N, p)$, [18], [37].

An interesting result is that even though $\mathbb{G}(N, p \times \tilde{p})$ becomes connected for smaller values of $p_f(N)$ when compared to $\mathbb{G}(N, p \times p_f)$ (as already mentioned $\tilde{p} > p_f$ for $0 < p_f < 1$), these values have the same asymptotical order and eventually, or $p_{f,0}(N) = \Theta\left(\frac{p_{Q_0}(N)}{p(N)}\right)$ as it is shown in Lemma 1. This asymptotic result suffices for the analysis purposes in the following section and the subsequent simulation results.

Lemma 1 *Random graph $\mathbb{G}(N, p \times \tilde{p})$ and random graph $\mathbb{G}(N, p \times p_f)$ become connected for values of $p_f(N)$ of the same asymptotic order, provided that random graph $\mathbb{G}(N, p)$ is connected.*

PROOF. Given that random graph $\mathbb{G}(N, p)$ is connected, based on Equation 1, this is equivalent to $\frac{p(N)}{p_{Q_0}(N)} \rightarrow +\infty$, or $\frac{p_{Q_0}(N)}{p(N)} < 1$. It is easy to show that $\frac{p_{Q_0}(N)}{p(N)} \geq 1 - \sqrt{1 - \frac{p_{Q_0}(N)}{p(N)}}$ is satisfied. For this, it is enough to show that $\sqrt{1 - \frac{p_{Q_0}(N)}{p(N)}} \geq 1 - \frac{p_{Q_0}(N)}{p(N)}$ is satisfied. The latter expression is valid since $\sqrt{1-x} \geq 1-x$ for $0 \leq x \leq 1$, which in this case for $x = \frac{p_{Q_0}(N)}{p(N)}$.

The next step is to show that there exists some constant $0 < c \leq 1$, such that $1 - \sqrt{1 - \frac{p_{Q_0}(N)}{p(N)}} \geq c \frac{p_{Q_0}(N)}{p(N)}$ is satisfied. Let $x = \frac{p_{Q_0}(N)}{p(N)}$. The case of $x = 0$ and $x = 1$ is trivial so the focus next is on $0 < x < 1$. It is sufficient to show that there exists a constant $0 < c \leq 1$, such that $1 - \sqrt{1-x} \geq cx$, or $1 - cx \geq \sqrt{1-x}$, or, $c \leq \frac{1-\sqrt{1-x}}{x}$. Let $f(x) = \frac{1-\sqrt{1-x}}{x}$. The first derivative $f'(x) = \frac{2-2\sqrt{1-x}-x}{2\sqrt{1-x}} = 0$, has no solution in $0 < x < 1$. That means that for $0 < x < 1$, $f(x)$ either monotonically increases or monotonically decreases. For $x \rightarrow 1$, $\lim_{x \rightarrow 1} f(x) = \lim_{x \rightarrow 1} \frac{1-\sqrt{1-x}}{x} = 1$. For $x \rightarrow 0$, $\lim_{x \rightarrow 0} f(x) = \lim_{x \rightarrow 0} \frac{1-\sqrt{1-x}}{x} = \lim_{x \rightarrow 0} \frac{(1-\sqrt{1-x})'}{(x)'} = \lim_{x \rightarrow 0} \frac{\frac{1}{2\sqrt{1-x}}}{1} = \lim_{x \rightarrow 0} \frac{1}{2\sqrt{1-x}} = 0.5$. Since $\lim_{x \rightarrow 0} f(x) < \lim_{x \rightarrow 1} f(x)$ -and in view of the monotonicity of $f(x)$ argued above- it is implied that $f(x)$ monotonically increases and therefore, $f(x) \geq 0.5$, for $0 < x < 1$.

Eventually, for any $c \leq 0.5$ it is satisfied that $c \leq f(x) = \frac{1-\sqrt{1-x}}{x}$, for $0 < x < 1$, or, $1 - \sqrt{1 - \frac{p_{Q_0}(N)}{p(N)}} \geq c \frac{p_{Q_0}(N)}{p(N)}$ is satisfied. Since $\frac{p_{Q_0}(N)}{p(N)} \geq 1 - \sqrt{1 - \frac{p_{Q_0}(N)}{p(N)}} \geq c \frac{p_{Q_0}(N)}{p(N)}$ is satisfied, then $1 - \sqrt{1 - \frac{p_{Q_0}(N)}{p(N)}} = \Theta\left(\frac{p_{Q_0}(N)}{p(N)}\right)$.

Consequently, $p_{f,0}(N) = \Theta\left(\frac{p_{Q_0}(N)}{p(N)}\right)$, and the lemma is proved. \square

So far, the case of global network outreach (i.e., $C(L) = 1$ for $L = 0$) w.h.p. has been studied. The study for $L = 1$ and $L = 2$ follows the previously described steps with respect to properties Q_1 and Q_2 . For instance, if $\mathbb{G}(N, p \times p_f)$ has property Q_1 w.h.p., it is expected $\mathbb{P}(\mathbb{G}(N, p), p_f)$ to include the majority of the network nodes (as it is the case for the giant component of $\mathbb{G}(N, p \times p_f)$ that consists of the majority of the network nodes) and those nodes that are not part of $\mathbb{P}(\mathbb{G}(N, p), p_f)$, to be at most 1 hop away from at least one node of $\mathbb{P}(\mathbb{G}(N, p), p_f)$ (since an increase in $p_f(N)$ could result in the creation of a new link between these nodes and a node of the giant component). To support this claim, in [26] it is also mentioned that when the property Q_1 exists (w.h.p.), then apart from the giant component there exist isolated nodes in the network for which any new link will connect them to the giant component w.h.p. Assuming $p_{f,1}(N)$ to be the particular probability such that $C(1) = 1$ is satisfied w.h.p., it is derived that $p_{f,1}(N) = \Theta\left(\frac{p_{Q_1}(N)}{p(N)}\right)$. The same applies for $L = 2$ and property Q_2 . As already mentioned, in [26] when property Q_2 exists, the giant component is present in the network along with small components of one, two or three nodes (w.h.p.) for which any new link will connect them to the giant component (thus some of these being at distance 2 hops away from nodes already within the giant component) w.h.p. Therefore, assuming $p_{f,2}(N)$ to be the particular probability such that $C(2) = 1$ is satisfied w.h.p., it is derived that $p_{f,2}(N) = \Theta\left(\frac{p_{Q_2}(N)}{p(N)}\right)$. Figure 3 depicts asymptotic numerical values of $p_{f,L}(N)$ as a function of N for $L = 0, 1, 2$, using the results depicted in Table 1. For this particular case, the underlying network topology is considered fixed, and therefore $p(N)$ is a positive constant.

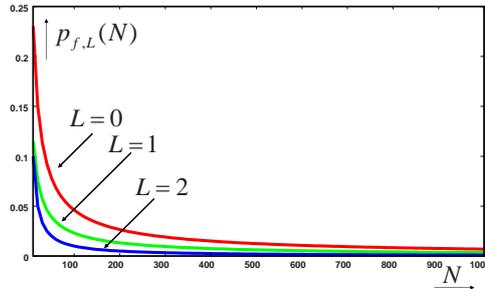


Fig. 3. $p_{f,L}(N)$ as a function of N for $L = 0, 1, 2$.

5 Asymptotic Analysis

The asymptotic expressions derived in the previous section are considered next to asymptotically analyze probabilistic flooding and compare it against traditional flooding, with respect to the number of messages required and

termination time (upper bounds) until a certain coverage $C(L) = 1$ is achieved.

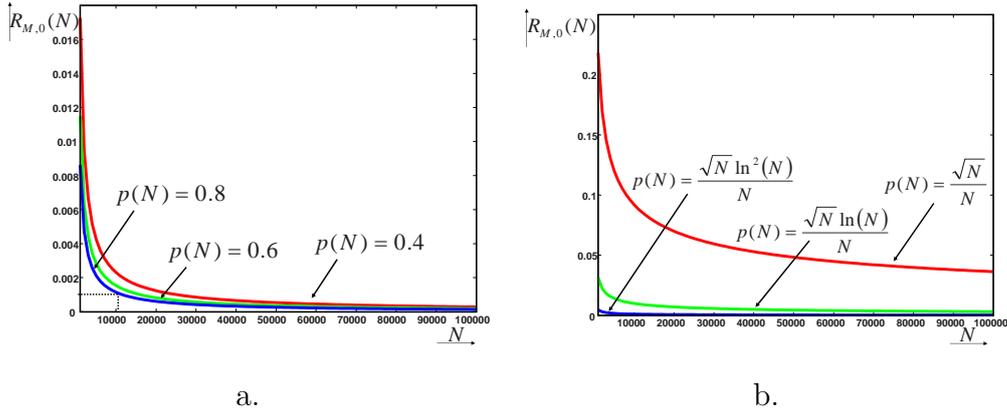


Fig. 4. $R_{M,0}(N)$ as a function of N for various values and asymptotic orders of $p(N)$.

For the case of global network outreach or $C(0) = 1$, it is known that the number of messages under traditional flooding (M_{trad}) is of the order of:

$$M_{trad} = \Theta\left(p(N) \frac{N-1}{2} N\right). \quad (4)$$

Under probabilistic flooding, a probabilistic flooding network $\mathbb{P}(\mathbb{G}(N, p), p_f)$ is created, and therefore, the number of messages is of the order of the number of links of this network, or $\Theta(\mathcal{E}(\mathbb{P}(\mathbb{G}(N, p), p_f)))$. Based on the study presented in the previous section, the number of messages under probabilistic flooding (M_{prob}) for global network outreach are of the order of:

$$\begin{aligned} M_{prob} &= \Theta\left(p_{Q_0}(N) \frac{N(N-1)}{2}\right) = \Theta\left(\frac{N-1}{2} \ln(N)\right) \\ &= \Theta(N \ln(N)). \end{aligned} \quad (5)$$

Assuming $p(N)$ to be constant note that under traditional flooding, the number of messages is of the order of $\Theta(N^2)$ while under probabilistic flooding are of the order of $\Theta(N \ln(N))$, which is a significant reduction compared to traditional flooding that becomes more significant as N increases, as illustrated in Figure 4.a. Let $R_{M,L}(N)$ denote the (asymptotic) fraction of messages under probabilistic flooding over those under traditional flooding for some L , or, $R_{M,L}(N) = \frac{\mathcal{E}(\mathbb{P}(\mathbb{G}(N,p), p_f))}{\mathcal{E}(\mathbb{G}(N,p))}$. For the case of $L = 0$,

$$R_{M,0}(N) = \frac{\ln(N)}{p(N)N}. \quad (6)$$

Obviously, $R_{M,0}(N) \rightarrow 0$, when $N \rightarrow +\infty$ ($O(p(N)) > O(p_{Q_0}(N)) = O(\frac{\ln(N)}{N})$). Note that in strict terms, $R_{M,0}(N) = \Theta\left(\frac{\ln(N)}{p(N)N}\right)$. However, in order to simplify the notations in the sequel $\Theta(\cdot)$ will not appear for both $R_{M,L}(N)$ and $R_{T,L}(N)$ (to be defined later).

As already mentioned, since $\mathbb{G}(N, p)$ is a connected network w.h.p., then $O(p(N)) > O(p_{Q_0}(N)) = O(\frac{\ln(N)}{N})$. For $p(N) = \Theta\left(\frac{\ln(N)}{N}\right)$ (which means that $\mathbb{G}(N, p)$ has just become connected), it is interesting to see that $R_{M,0}(N) = \Theta(1)$, which apparently demonstrates the fact that there is no advantage under probabilistic flooding when compared to traditional flooding for this case (the number of messages is the same under both probabilistic flooding and traditional flooding). Actually, this particular case is -asymptotically- equivalent to $p_f(N) = 1$, for which probabilistic flooding reduces to traditional flooding. In order to explain further this observation, note that for the case of $p(N) = \Theta\left(\frac{\ln(N)}{N}\right)$, $\mathbb{G}(N, p)$ has just become connected w.h.p. which apparently means that the number of “redundant” links (links over which traditional flooding forwards messages and probabilistic flooding “saves” by probabilistically “avoiding” to do so) is significantly reduced. The shape of the network – even though it contains cycles – looks mostly like a tree, [22], [26], and therefore, the ability of probabilistic flooding to “avoid” forwarding messages over “redundant” links is reduced.

As $p(N)$ increases asymptotically, the potential advantage of probabilistic flooding increases, as it is shown in Figure 4.b for various functions of $p(N)$ of different asymptotic order (i.e., $\frac{\sqrt{N}}{N}$, $\frac{\sqrt{N}\ln(N)}{N}$, $\frac{\sqrt{N}\ln^2(N)}{N}$), [37], in order to illustrate in more depth how $p(N)$ affects $R_{M,0}(N)$. Note that as $p(N)$ increases ($\frac{\sqrt{N}}{N} < \frac{\sqrt{N}\ln(N)}{N} < \frac{\sqrt{N}\ln^2(N)}{N}$), $R_{M,0}(N)$ decreases which means that the savings in messages under probabilistic flooding increase when compared to traditional flooding. However, assuming large values of N but not that large such that $N \rightarrow +\infty$, and $p(N) = 1$, $R_{M,0}(N) \rightarrow \frac{\ln(N)}{N} > 0$, which is the asymptotic lower bound with respect to N . Still, as $N \rightarrow +\infty$, $R_{M,0}(N) \rightarrow 0$.

It should be mentioned that the tightest lower bound with respect to the number of messages for global network outreach is $N - 1$ for any (connected) network of N nodes, assuming that there is global knowledge available regarding the topology of the network in order to derive the appropriate paths for the forwarded messages in the network (e.g., creating a spanning tree). Such global knowledge cannot be assumed for the large-scale and dynamic network environments considered in this work. What is interesting is that the number of messages under probabilistic flooding – that assumes *no global knowledge* for the network topology $\mathbb{G}(N, p)$ – is (asymptotically) close to the optimal one – that requires global knowledge – in the order of $\Theta(\ln(N))$ ($\Theta(N \ln(N))$) and $N - 1$ messages, respectively).

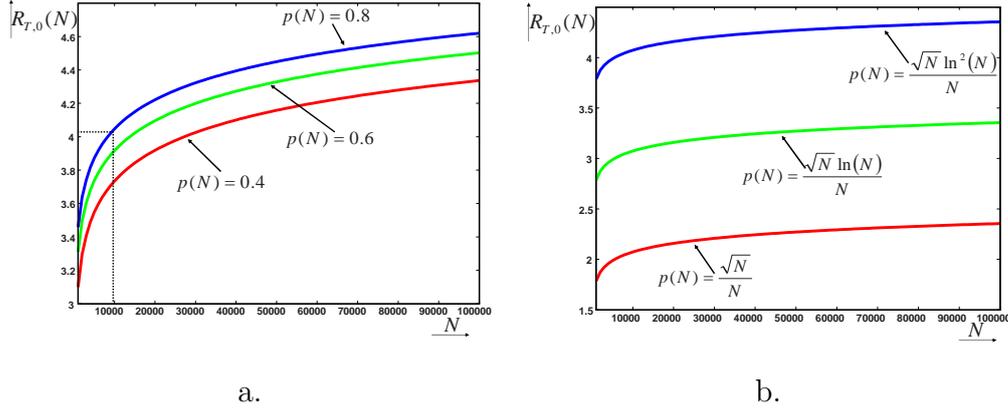


Fig. 5. $R_{T,0}(N)$ as a function of N for various values and asymptotic orders of $p(N)$.

The reduced number of messages under probabilistic flooding is achieved at the expense of larger termination delays. This is shown by comparing the network diameter of $\mathbb{G}(N, p)$ and $\mathbb{P}(\mathbb{G}(N, p), p_f)$, for $p_f(N) = \Theta\left(\frac{\ln(N)}{N}\right)$ (the upper bound of termination time corresponds to the network diameter). Let $R_{T,L}(N)$ denote the (asymptotic) fraction of the network diameter of the probabilistic flooding network over the network diameter minus L (for fairness issues as it will be explained in the sequel in Section 6), for some $L = 0, 1, 2$, or $R_{T,L}(N) = \frac{\mathcal{D}(\mathbb{P}(\mathbb{G}(N, p), p_f))}{\mathcal{D}(\mathbb{G}(N, p)) - L}$. Given that for $L = 0$, $\mathcal{D}(\mathbb{G}(N, p)) = \Theta\left(\frac{\ln(N)}{\ln(p(N)N)}\right)$, $\mathcal{D}(\mathbb{P}(\mathbb{G}(N, p), p_f)) = \Theta(\mathcal{D}(\mathbb{G}(N, p * p_f))) = \Theta\left(\frac{\ln(N)}{\ln(p(N)p_f(N)N)}\right)$ according to Lemma 1, and $p(N)p_f(N) = \Theta\left(\frac{\ln(N)}{N}\right)$, it follows that,

$$R_{T,0}(N) = \frac{\ln(p(N)N)}{\ln(\ln(N))}. \quad (7)$$

Figure 5 depicts $R_{T,0}(N)$ for various values and asymptotic orders of $p(N)$. It is observed that the upper bound of termination time does not increase as rapidly as the number of messages decreases. For instance, for $N = 10000$ and $p(N) = 0.8$, as it is depicted in Figure 4.a, $R_{M,0}(N) \approx 0.1$, and as depicted in Figure 5.a, $R_{T,0}(N) \approx 4.03$. So, in asymptotic terms and the particular depicted case, probabilistic flooding employs 0.1% of the messages required under traditional flooding, at the expense of increasing the (average) upper bound of termination time 4.03 times. Given that diameters in random graph topologies are basically small (as already mentioned it is known as the “small world” phenomenon, [22], [23]) a time delay upper bound of the order of $4.03 \times \mathcal{D}(\mathbb{G}(10000, 0.8))$ may be viewed as small and thus not be prohibitive for employing probabilistic flooding.

Apart from the aforementioned increment of the (average) upper bound of termination time, another drawback of probabilistic flooding is the (unavoidable) employment of a probabilistic process instead of a deterministic one as it is the case under traditional flooding. For example, reduction of the num-

ber of messages under probabilistic flooding takes place at the cost of being achieved only with high probability, whereas under traditional flooding it is guaranteed.

So far in this section the global network outreach case (i.e., $L = 0$ or $C(0) = 1$) has been studied. The cases corresponding to $L = 1$ and $L = 2$ are naturally expected to yield smaller number of messages under probabilistic flooding – compared to traditional flooding – since the particular values of $p_f(N)$ are expected to be (on average) smaller than those ensuring global network outreach (i.e., $C(0) = 1$) w.h.p. Given that the case of $L = 1$, corresponds to property Q_1 , as it can be seen from Table 1, $p_{Q_1}(N) = \frac{\ln(N)}{2N} = \Theta\left(\frac{\ln(N)}{N}\right)$. It is evident that the asymptotic analysis that has been previously followed for $L = 0$ applies to this particular case as well. Note, that the resemblance is only asymptotic and savings with respect to the number of messages are greater for the case of $L = 1$ than for the case of $L = 0$ under probabilistic flooding, as it will be demonstrated by simulations in the following section.

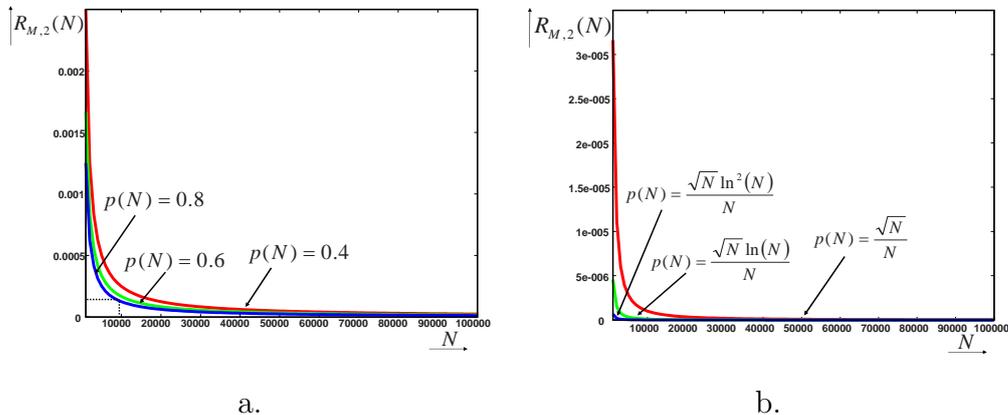


Fig. 6. $R_{M,2}(N)$ as a function of N for various values and asymptotic orders of $p(N)$.

The case of $L = 2$, is different than the case of $L = 1$ and $L = 0$, since $p_{Q_2}(N) = \frac{1}{N}$. Following the same steps as for the case of $L = 0$, it is concluded that,

$$R_{M,2}(N) = \frac{1}{p(N)N}. \quad (8)$$

$R_{M,2}(N)$ for various values of $p(N)$ is depicted in Figure 6. Note that for $N = 10000$ and $p(N) = 0.8$ as depicted in Figure 6.a, the (asymptotic) savings with respect to the number of messages is about 0.01% as opposed to 0.1% depicted in Figure 4.a.

6 Simulation Results and Evaluation

The purpose of the simulation results presented in this section is two-fold: (a) to support the claims and expectations of the analysis; and (b) to provide for further insight into various aspects of probabilistic flooding in random graphs. A simulation program was developed in C programming language and random graph topologies up to 10000 nodes were created, while traditional and probabilistic flooding was considered over these topologies. The presented results are averaged values of 100 independent runs. For each run the initiator node was randomly selected and different seeds with respect to pseudorandom sequences (governing the selection of initiator nodes, the creation of topologies and message forwarding under probabilistic flooding) were used. In many cases $p(N) = 0.0008$, that is a value large enough for $\mathbb{G}(N, p)$ to be connected and at the same time small enough in order for the resulting topology to be of reduced number of links which corresponds to a worst-case scenario under probabilistic flooding when compared against traditional flooding. Let $\bar{C}(L)$ denote the averaged (coverage) fraction of nodes that belong to either the probabilistic flooding network or the giant component of the under consideration random graph, depending on the case. Let $\bar{R}_{M,L}(N)$ ($\bar{R}_{T,L}(N)$) denote the averaged fraction of the number of messages (termination time) under probabilistic flooding over that under traditional flooding for $L = 0, 1, 2$.

6.1 Bounded Probabilistic Flooding

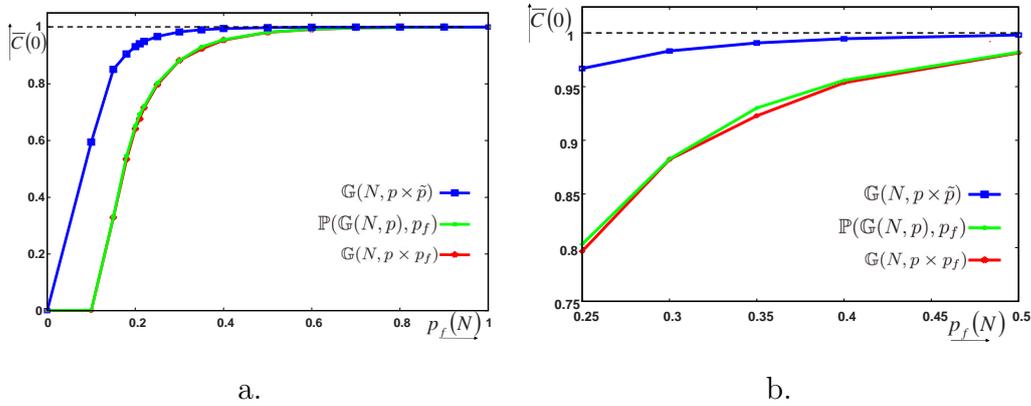


Fig. 7. $\bar{C}(0)$ as a function of $p_f(N)$ for $N = 10000$ and $p(N) = 0.0008$.

Figure 7.a depicts simulation results with respect to coverage for $L = 0$. It is observed that probabilistic flooding's performance lays between the performance of $\mathbb{G}(N, p \times \tilde{p})$ and $\mathbb{G}(N, p \times p_f)$, as was expected from the analysis. In particular, as $p_f(N)$ increases, $\mathbb{P}(\mathbb{G}(N, p), p_f)$ includes all network nodes after $\mathbb{G}(N, p \times \tilde{p})$ has achieved global network outreach ($p_f(N) \approx 0.4$ in Figure 7) and slightly before $\mathbb{G}(N, p \times p_f)$ ($p_f(N) \approx 0.5$ in Figure 7). Global

network outreach within the simulation section is considered to be achieved when network coverage is equal to 98%.

It is also interesting to note that $\bar{C}(0)$ for the probabilistic flooding network is closer to the particular value of $\bar{C}(0)$ for $\mathbb{G}(N, p \times p_f)$ than to $\bar{C}(0)$ for $\mathbb{G}(N, p \times \tilde{p})$. This is more clearly depicted in Figure 7.b that is a zoom version of Figure 7.a in the range of values $p_f(N) \in [0.25, 0.5]$. However, by the time $\mathbb{G}(N, p \times p_f)$ achieves global network outreach (i.e., $p_f(N) = 0.5$ according to the simulation results) all three curves are close ($\bar{C}(0) = 99.8\%$, 98.2% and 98.14% , for $\mathbb{G}(N, p \times \tilde{p})$, $\mathbb{P}(\mathbb{G}(N, p), p_f)$ and $\mathbb{G}(N, p \times p_f)$, respectively).

The analysis presented in Section 4.3 have shown the asymptotic “equivalence” among the previously discussed graphs with respect to $L = 0$ (i.e., Lemma 1). As it is depicted in Figure 8, this is also the case for $L = 1$ and $L = 2$ as well (cases not covered by the analysis in Section 4.3). Obviously, coverage under probabilistic flooding is between coverage of both (i.e., $\mathbb{G}(N, p \times \tilde{p})$ and $\mathbb{G}(N, p \times p_f)$) random graphs for $L = 1, 2$ as well. The sudden increase in the performance of the probabilistic flooding is due to the “phase transition phenomenon” that takes place. As it was mentioned in Section 3, there is a critical probability for larger values of which a property of the graph exists while for lower values it does not. Therefore, coverage $C(L)$ increasing suddenly by the time property Q_L is satisfied.

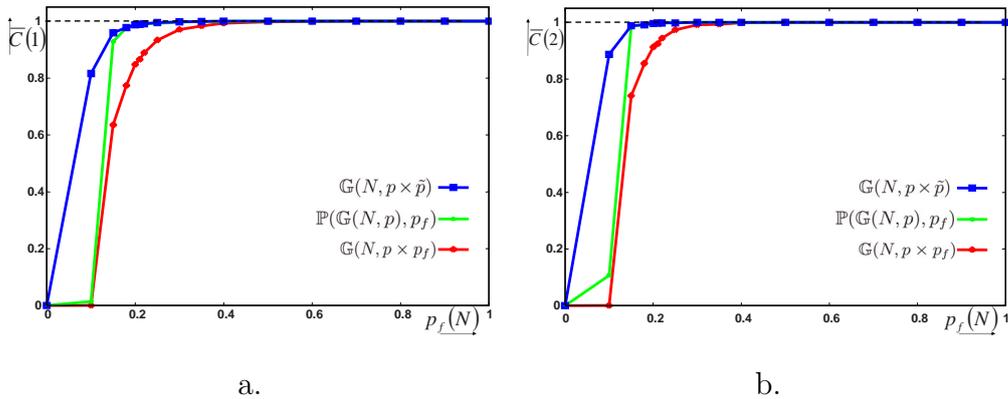


Fig. 8. $\bar{C}(1)$ and $\bar{C}(2)$ as a function of $p_f(N)$ for $N = 10000$ and $p(N) = 0.0008$.

Figure 9 depicts the particular values of $p_{f,L}(N)$ for which $\bar{C}(L) = 0.98$, for $p(N) = 0.0008$ and different values of L . $\bar{C}(L) = 0.98$ is considered instead of $\bar{C}(L) = 1$ due to the probabilistic nature of both random graphs and probabilistic flooding (e.g., sometimes for networks of 10000 nodes it is unreasonably time consuming for $\bar{C}(L) = 1$ to be satisfied). As L increases, the particular value of $p_{f,L}(N)$ decreases as expected, since smaller fractions of the network nodes are needed in order for probabilistic flooding to cover the network for the particular value of L . At the same time, the particular value of $p_{f,L}(N)$ significantly decreases, as the network size N increases. This is due to the fact that as N increases and given that $p(N)$ is fixed, the number of redundant

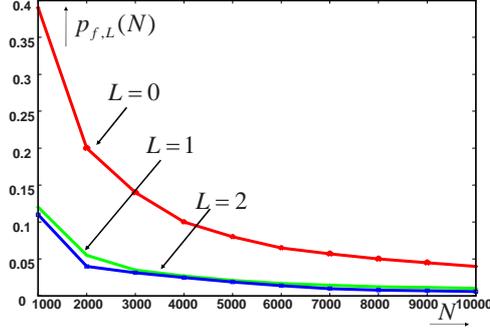


Fig. 9. $p_{f,L}(N)$ for which $\bar{C}(L) = 0.98$ as a function of N for $L = 0, 1, 2$ in a $\mathbb{G}(10000, 0.0008)$ network topology.

links significantly increases (for example, if 10000 nodes are already present in the network, the addition of a new node for $p(N) = 0.0008$ – a comparably small value of $p(N)$ – corresponds to on average 8 new links), thus allowing for the coverage of the network under probabilistic flooding for small values of $p_{f,L}(N)$. It should be also noted that the shape of the curves in Figure 9 is similar (in asymptotic terms) to the curves depicted in Figure 3 corresponding to the asymptotic values of $p_{f,L}(N)$ as derived by the analysis.

6.2 Links and Messages

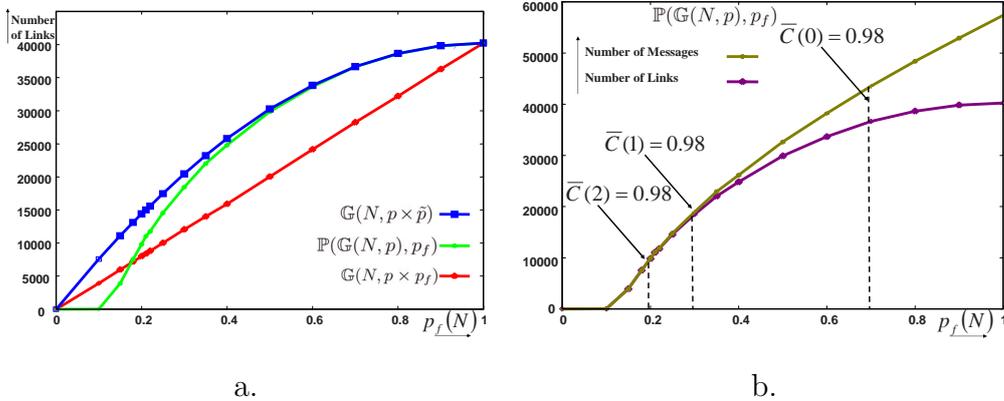


Fig. 10. Number of links and messages for $N = 10000$ and $p(N) = 0.0008$.

Figure 10 depicts simulation results for the number of links of the probabilistic flooding network and the number of messages under probabilistic flooding that reveal some interesting aspects of the probabilistic flooding process not captured by the analysis. In particular, in Figure 10.a, the number of links is depicted for the previously mentioned random graphs and the probabilistic flooding network. It is interesting to observe that for small values of $p_f(N)$ (e.g., smaller than 0.18), the number of links of the probabilistic network $\mathcal{E}(\mathbb{P}(\mathbb{G}(N, p), p_f))$ is smaller than the number of links for either random graph. This is not a contradiction to the analytical results, since the analysis

was based on the existence of the giant component. For such small values of $p_f(N) < 0.18$ (and given the small value of $p(N) = 0.0008$), the giant component is not present. This can also be observed by looking at the value of the $\bar{C}(0)$ in Figure 7.a, where the larger existing component that contains – at least 50% of the network nodes – is created for $p_f \geq 0.18$ (when 50% of the nodes belong to a certain component, it is safe to assume that a giant component starts to emerge, [26]). As soon as the giant component emerges (for $p_f(N) > 0.18$), the number of links of the probabilistic flooding network lays between the number of links of the other two random graph topologies, as it is observed from Figure 10.a and expected from the analysis.

Figure 10.b presents a comparison between the number of the links of the probabilistic flooding network and the number of messages sent under probabilistic flooding considering a $\mathbb{G}(N, p)$ network topology. The vertical dotted lines correspond to the particular values of $p_f(N)$ for which $\bar{C}(L) = 0.98$, for $L = 0, 1, 2$ (e.g., as also observed from Figure 7 and Figure 8). It is interesting to note that for small values of $p_f(N)$ the number of links and the number of messages is about the same, particularly for those values of $p_f(N)$ for which $\bar{C}(2) = 0.98$ and $\bar{C}(1) = 0.98$. However, after the particular value of $p_f(N)$ for which $\bar{C}(1) = 0.98$ and as $p_f(N)$ increases – and before $\bar{C}(0) = 0.98$ – the number of messages gradually deviates (increases) from the number of links. The maximum difference takes place for $p_f(N) = 1$, (around 20000 more messages than links in a network of 10000 network nodes) which is the case when probabilistic flooding reduces to traditional flooding. Even for $p_f(N) = 0.7$ (i.e., $\bar{C}(0) = 0.98$) the difference is about 10000 more messages than links. This means that for a certain link of the probabilistic flooding network, in many cases two messages were forwarded over it (instead) of one, resulting in a significant number of redundant links particularly for $\bar{C}(0) = 0.98$ (i.e., global network outreach). Such redundancy is not observed for $L = 1$ and $L = 2$ as it is also demonstrated in Figure 10.b. The main reason is that both $\bar{C}(1) = 0.98$ and $\bar{C}(2) = 0.98$ are achieved for small values of $p_f(N)$ (thus, redundancy is limited).

6.3 Global Network Outreach ($L = 0$)

Figure 11 depicts simulation results with respect to $\bar{R}_{M,0}(N)$ and $\bar{R}_{T,0}(N)$ for three different values of $p(N)$ (0.004, 0.005 and 0.01). As it can be observed from Figure 11.a, the savings on the number of messages increase with the number of network nodes N , for the same value of $p(N)$, as it is expected by the analysis and particularly Equation (6). There is a clear resemblance between the simulation results presented in Figure 11.a and the (asymptotic) analytical results presented in Figure 4. The savings on the number of messages increase as $p(N)$ increases. This is due to the fact that large values for $p(N)$

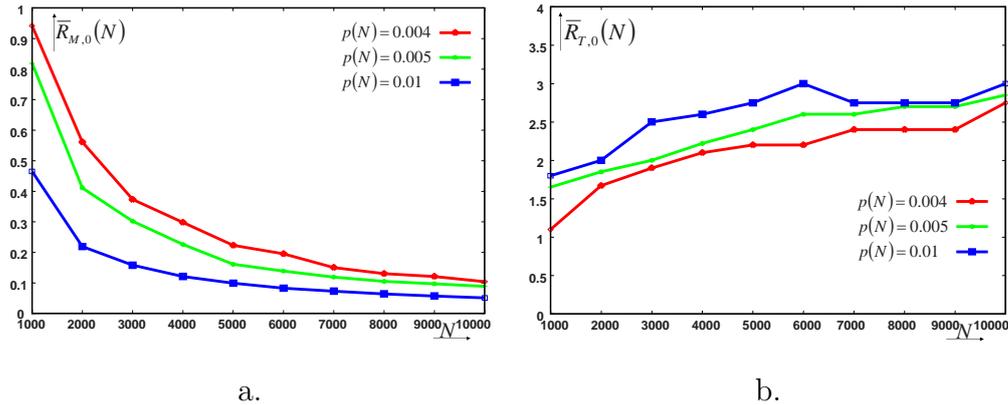


Fig. 11. $\bar{R}_{M,0}(N)$ and $\bar{R}_{T,0}(N)$ as a function of N for various values of $p(N)$

provide for an increase on the number of network links and, therefore, an increase on the number of redundant links that are used under traditional flooding to forward messages.

Figure 11.b depicts simulation results with respect to termination time ratio $\bar{R}_{T,0}(N)$. $\bar{R}_{T,0}(N)$ increases by the increase on the network's size, as it is already mentioned in the analysis (Section 5). It is observed that the larger the value for $p(N)$, the larger the increase of $\bar{R}_{T,0}(N)$ as it is also expected given Equation (7). Note that an increase on $p(N)$ corresponds to networks of more redundant links, smaller diameters and smaller termination times under traditional flooding, which leads to an increase in $\bar{R}_{T,0}(N)$. As before there is a clear resemblance between the simulation results presented in Figure 11.b and the (asymptotic) analytical results presented in Figure 5.a.

6.4 The case of $L = 1$ and $L = 2$

Note that the comparison between probabilistic flooding and traditional flooding with respect to coverage, number of messages and termination time, for $L = 1$ and $L = 2$ is not a fair one since traditional flooding deterministically achieves $\bar{C}(L) = 1$ for $L = 0, 1, 2$, having “spent” as many/much messages/time as required for $\bar{C}(0) = 0.98 \approx 1$ (global network outreach). A fairer comparison – as it is the case in the sequel – would be one under which controlled flooding is used for flooding information ($r(u) - L$) hops away from any initiator node u , where $r(u)$ is the maximum number of hops between any node in the network and node u over a shortest path ($r(u)$ is usually referred to as the *radius* or *eccentricity* of node u and is upper bounded by the network diameter).

Figure 12 depicts simulation results for $\bar{R}_{M,1}(N)$ and $\bar{R}_{T,1}(N)$. The curve corresponding to $\bar{R}_{M,1}(N)$ (Figure 12.a) is similar to the one expected from the

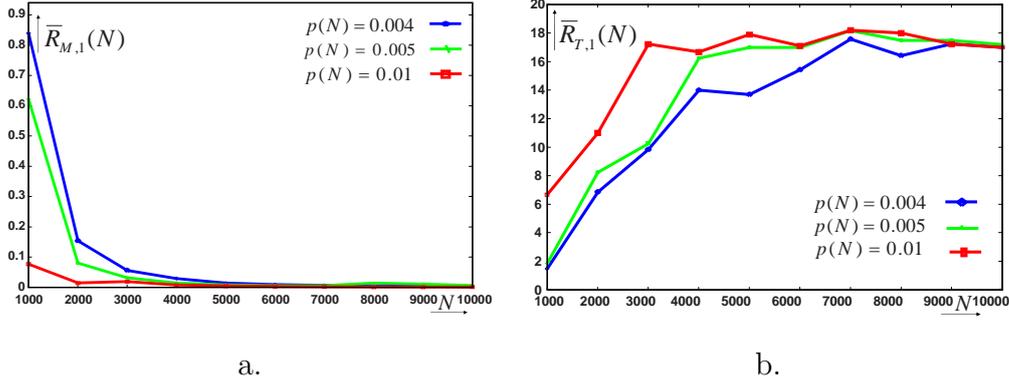


Fig. 12. $\bar{R}_{M,1}(N)$ and $\bar{R}_{T,1}(N)$ as a function of N for various values of $p(N)$

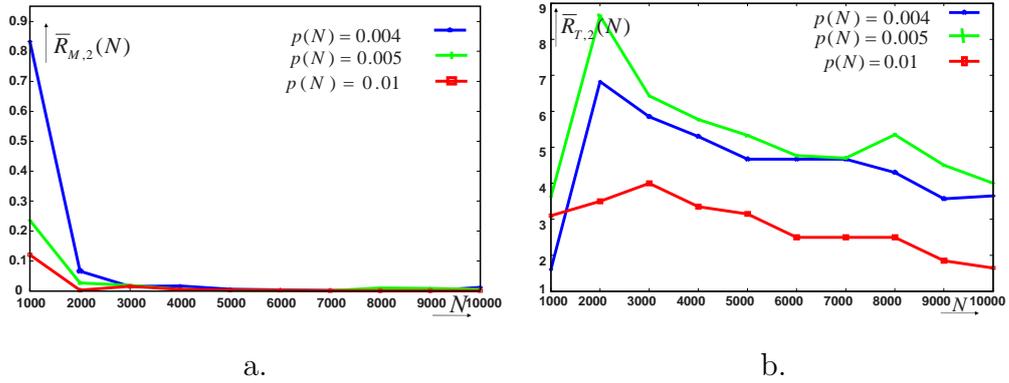


Fig. 13. $\bar{R}_{M,2}(N)$ and $\bar{R}_{T,2}(N)$ as a function of N for various values of $p(N)$

analysis (i.e., asymptotically similar to $\bar{R}_{M,0}(N)$ as also observed from Figure 11.a). In fact, $\bar{R}_{M,1}(N)$ is slightly better than $\bar{R}_{M,0}(N)$ ($\bar{R}_{M,1}(N) < \bar{R}_{M,0}(N)$) because of the smaller values for $p_f(N)$ required for coverage to become equal to 0.98 for either case ($L = 0$ and $L = 1$). $\bar{R}_{T,1}(N)$ is depicted in Figure 12.b and – as it is the case for $\bar{R}_{M,1}(N)$ – the corresponding graph is similar to $\bar{R}_{T,0}(N)$.

Figure 13 depicts simulation results for both $\bar{R}_{M,2}(N)$ and $\bar{R}_{T,2}(N)$. It can be observed from Figure 13.a that $\bar{R}_{M,2}(N)$ is slightly smaller than $\bar{R}_{M,1}(N)$, as expected (as already mentioned smaller values of $p_f(N)$ are capable of achieving $\bar{C}(2) = 0.98$ rather than $\bar{C}(1) = 0.98$).

The depicted simulation results with respect to $\bar{R}_{T,2}(N)$ in Figure 13.b allow for an interesting observation. Apart from an increase at the beginning for small values of N , as N increases it is clear that $\bar{R}_{T,2}(N)$ decreases, in contrast to both the $\bar{R}_{T,0}(N)$ and $\bar{R}_{T,1}(N)$ or $R_{T,0}(N)$ and $R_{T,1}(N)$. In order to explain this observation it is important to get into more details regarding the case of $L = 2$. For a certain (fixed) value of $p(N)$ as N increases, the number of links among nodes increases and eventually the diameter of the network decreases, [26]. For the conducted simulations of network sizes between 1000

and 10000 nodes (the simulator had performance and memory limitations running for topologies larger than 10000 nodes), the considered values of $p(N)$ (selected large enough such that all networks to be connected w.h.p.), were large enough to result in small network diameters (for $\mathbb{G}(10000, 0.0005)$ the diameter is 4) and therefore, comparable to 2. For such case, it is evident that the performance under the probabilistic flooding is comparable to that under controlled flooding ($r(u) - L$ hops away for the initiator node u). As N increases further, the network diameter becomes smaller, thus allowing for probabilistic flooding to faster cover the network. Note, however, that this is a limited case since for a network diameter comparable to L , information dissemination is becoming obsolete (e.g., for L equal to the network diameter all nodes are at most L hops away from any network node by definition).

7 Conclusions

The problem of limited information dissemination in large, unstructured networks employing probabilistic flooding is studied and analyzed here. Global network outreach is available in a network if the employed information dissemination scheme is capable of taking a message from any originating node to any other network node. Such network outreach is needed in order to support routing protocols, advertise a new service, search for some information, etc. Traditional flooding schemes achieve global network outreach in unstructured networks with certainty (deterministically, for a connected network) at a large message overhead cost. In this paper, probabilistic flooding schemes have been considered in order to reduce their associated large overhead, at the price of providing probabilistic global network outreach guarantees. It is shown here that the network created by probabilistic flooding over a random graph network asymptotically lays between two random graph networks – which are determined - facilitating this way the derivation of asymptotic analytical expressions on the value of the forwarding probability that results in a fairly decreased (compared to the traditional flooding) number of messages, while covering the network with high probability (w.h.p.). Apart from global network outreach (i.e., $L = 0$ case), the cases of $L = 1$ and $L = 2$ have also been analytically studied.

A comparison of probabilistic flooding under the derived forwarding probability with the traditional flooding is carried out. For the case of global network outreach it is shown that the number of messages under probabilistic flooding asymptotically increases as $N \ln(N)$ as opposed to N^2 under traditional flooding, for any fixed value of the forwarding probability. Thus, significant message overhead reduction can be achieved, especially for large networks (large N). The relative message overhead (compared to that under traditional flooding) is also derived and shown to yield substantial message overhead savings even for

low to medium values of N . However, as it was mentioned, the global network outreach achieved under traditional flooding with certainty is now achieved only with high probability. Termination time is also studied and as expected, it is slightly higher under probabilistic flooding (compared to message reduction) but this is practically insignificant if one considers typical time scales of flooding and usage of network information. Extensive simulation results support the analytical claims and expectation and in addition reveal certain interesting aspects with respect to probabilistic flooding.

References

- [1] A. Segall, "Distributed network protocols," *IEEE Trans. Inform. Theory*, vol. IT-29, Jan. 1983.
- [2] Lua, E.K., Crowcroft, J., Pias, M., Sharma, R., Lim, S. "A Survey and Comparison of Peer-to-Peer Overlay Network Schemes", *IEEE Communications Survey and Tutorial*, March (2004).
- [3] Napster. [Online]. Available: <http://www.napster.com/>
- [4] I. Stoica, R. Morris, D. Karger, M.F. Kaashoek, and H. Balakrishnan, "Chord: A scalable peer-to-peer lookup service for internet applications," in *Proceedings of ACM SIGCOMM '01*, 2001.
- [5] AB.Y. Zhao, J. Kubiatowicz, and A. Joseph, "Tapestry: An infrastructure for fault-tolerant wide-area location and routing," *Tech. Rep. UCB/CSD-01-1141*, UCB, 2001.
- [6] A. Rowstron and P. Druschel, "Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer systems," in *IFIP/ACM International Conference on Distributed Systems Platforms (Middleware)*, 2001, pp. 329 - 350.
- [7] Gnutella RFC, <http://rfc-gnutella.sourceforge.net/>, 2002.
- [8] C. Lv, P. Cao, E. Cohen, K. Li, and S. Shenker, "Search and Replication in Unstructured Peer-to-Peer Networks," *ICS 2002*, 2002.
- [9] V. Kalogeraki, D. Gunopoulos and D. Zeinalipour-Yazti, "A Local Search Mechanism for Peer-to-Peer Networks," in *CIKM (International Conference on Information and Knowledge Management)*, 2002.
- [10] C. Gkantsidis, M. Mihail and A. Saberi, "Hybrid Search Schemes for Unstructured Peer-to-Peer Networks," *IEEE Infocom 2005*, 2005.
- [11] N. B. Chang and M. Liu, "Optimal Controlled Flooding Search in a Large Wireless Network," *Third International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt'05)*, 2005, pp. 229 - 237.

- [12] D. Kempe, J. Kleinberg, A. Demers, “Spatial gossip and resource location protocols,” Proc. 33rd ACM Symposium on Theory of Computing, 2001.
- [13] V. V. Dimakopoulos and E. Pitoura, “On the Performance of Flooding-Based Resource Discovery,” IEEE Transactions on Parallel and Distributed Systems, Vol. 17 (11), 2006, November, pp/ 290-297.
- [14] F. Banaei-Kashani and C. Shahabi, “Criticality-based Analysis and Design of Unstructured Peer-to-Peer Network as “Complex Systems,” in Proceedings of the Third International Symposium on Cluster Computing and the Grid, 2003, pp. 51 - 358.
- [15] Yoav Sasson, David Cavin, and André Schiper, “Probabilistic Broadcast for Flooding in Wireless Mobile Ad hoc Networks,” Swiss Federal Institute of Technology (EPFL), Technical Report IC/2002/54.
- [16] D. Tsoumakos and N. Roussopoulos, “Adaptive Probabilistic Search for Peer-to-Peer Networks,” 3rd IEEE International Conference on P2P Computing, 2003.
- [17] B. Wu, and A. D. Kshemkalyani, “Modeling message propagation in random graph networks,” Computer Communications, Elsevier, Vol. 31, N. 17, pp. 4138–4148, 2008.
- [18] K. Oikonomou, and I. Stavrakakis, “Performance Analysis of Probabilistic Flooding Using Random Graphs,” The First International IEEE WoWMoM Workshop on Autonomic and Opportunistic Communications (AOC 2007), Helsinki, Finland, 18 June, 2007.
- [19] A.O. Stauffer, and V.C. Barbosa, “Probabilistic Heuristics for Disseminating Information in Networks,” Networking, IEEE/ACM Transactions on Volume 15, Issue 2, Page(s):425 – 435, April 2007.
- [20] S. Crisostomo, U. Schilcher, C. Bettstetter, and J. Barros, “Analysis of Probabilistic Flooding: How do we Choose the Right Coin?,” Proc. IEEE International Conference on Communications, Dresden, Germany, June 2009.
- [21] P. Erdős and A. Rényi, “On Random Graphs,” Publicationes Mathematicae Debrecen 6, pp. 290-297.
- [22] B. Bollobás, “Random Graphs,” Cambridge University Press, Second Edition, 1999.
- [23] R. Albert and A. Barabási, “Statistical mechanics of complex networks,” Rev. Mod. Phys., 2002.
- [24] M. E. J. Newman, “The structure and function of complex networks,” SIAM Review, 2003, Vol. 45, pp. 167-256.
- [25] P. V. Mienghem, “Performance Analysis of Communication Networks and Systems,” Cambridge University Press, 2006.

- [26] A. R. S. Janson, T. Luczak, "Random Graphs," Wiley Interscience Series in Discrete Mathematics and Optimization, 2000.
- [27] R. Gaetaa, and M. Sereno, "Random graphs as models of hierarchical peer-to-peer networks," *Performance Evaluation*, Volume 64, Issues 9-12, October 2007, Pages 838–855, 2007.
- [28] S. Tewari, and L. Kleinrock, "Analysis of Search and Replication in Unstructured Peer-to-Peer Networks," *Proceedings of the 2005 ACM SIGMETRICS international conference on Measurement and modeling of computer systems*, June 06-10, 2005, Banff, Alberta, Canada.
- [29] D. Kogias, K. Oikonomou, and I. Stavrakakis, "Study of Randomly Replicated Random Walks For Information Dissemination Over Various Network Topologies," *Sixth Annual Conference on Wireless On demand Network Systems and Services (WONS)*, February 2–4, 2009. Snowbird, Utah, USA.
- [30] B. Williams, D. Mehta, T. Camp and W. Navidi, "Predictive Models to Rebroadcast in Mobile Ad Hoc Networks," *IEEE Transactions on Mobile Computing (TMC)*, 2004, Vol. 3, pp. 295-303.
- [31] B. Williams and T. Camp, "Comparison of broadcasting techniques for mobile ad hoc networks," *ACM Symposium on Mobile Ad Hoc Networking and Computing (MOBIHOC)*, 2002, pp. 194-205.
- [32] R. Karp and C. Schindelhauer and S. Shenker and B. Vocking, "Randomized rumor spreading," *Proceedings of the 41st Annual Symposium on Foundations of Computer Science, FOCS '00*, 2000.
- [33] S. Boyd, A. Ghosh, B. Prabhakar and D. Shah, "Randomized gossip algorithms," *IEEE/ACM Trans. Netw.*, 2006, Vol. 14, pp. 2508-2530.
- [34] A. Ganesh, L. Massoulie, D. Towsley, "The effect of network topology on the spread of epidemics," 13-17 March 2005, *INFOCOM 2005*, Vol. 2, 2005, pp. 1455-1466.
- [35] Z. J. Haas, J. Y. Halpern and L. Li, "Gossip-based ad hoc routing," *IEEE/ACM Trans. Netw.*, 2006, Vol. 14, pp. 479-491.
- [36] F. Chung and L. Lu, "The Diameter of Sparse Random Graphs," *Adv. in Appl. Math.*, 2001, Vol. 26(4), pp. 257-279.
- [37] K. A. Berman, and J. L. Paul, "Algorithms: Sequential, Parallel, and Distributed," *Course Tecnology, Thomson Learning*, 2005.