Parsimonious Flooding in Dynamic Graphs*

Hervé Baumann
University Paris Diderot
herve.baumann@liafa.jussieu.fr

Pierluigi Crescenzi
University of Florence

Pierre Fraigniaud
CNRS and Univ. Paris Diderot
pierre.fraigniaud@liafa.jussieu.fr

ABSTRACT

An edge-Markovian process with birth-rate p and death-rate q generates sequences of graphs $(G_0, G_1, G_2, \ldots)$ with the same node set $[n]$ such that G_t is obtained from G_{t-1} as follows: if $e \notin E(G_{t-1})$ then $e \in E(G_t)$ with probability p, and if $e \in E(G_{t-1})$ then $e \notin E(G_t)$ with probability q. Clementi et al. (PODC 2008) analyzed thoroughly information dissemination in such dynamic graphs, by establishing bounds on their flooding time — flooding is the basic mechanism in which every node becoming aware of an information at step t forwards this information to all its neighbors at all forthcoming steps $t' > t$. In this paper, we establish tight bounds on the complexity of flooding for all possible birth rates and death rates, completing the previous results by Clementi et al. Moreover, we note that despite its many advantages in term of simplicity and robustness, flooding suffers from its high bandwidth consumption. Hence we also show that flooding in dynamic graphs can be implemented in a more parsimonious manner, so that to save bandwidth, yet preserving efficiency in term of simplicity and completion time.

For a positive integer k, we say that the flooding protocol is k-active if each node forwards an information only during the k time steps immediately following the step at which the node receives that information for the first time. We define the *reachability threshold* for the flooding protocol as the smallest integer k such that, for any source $s \in [n]$, the k-active flooding protocol from s completes (i.e., reaches all nodes), and we establish tight bounds for this parameter. We show that, for a large spectrum of parameters p and q, the reachability threshold is by several orders of magnitude smaller than the flooding time. In particular, we show that it is even constant whenever the ratio $p/(p+q)$ exceeds $\log n/n$. Moreover, we also show that being active for a number of steps equal to the reachability threshold (up to a multiplicative constant) allows the flooding protocol to complete in *optimal* time, i.e., in asymptotically the same number of steps as when being perpetually active. These results demonstrate that flooding can be implemented in a practical and efficient manner in dynamic graphs. The main ingredient in the proofs of our results is a reduction lemma enabling to overcome the time dependencies in edge-Markovian dynamic graphs.

Categories and Subject Descriptors

C.2.1 [**Computer-Communication Networks**]: Network Architecture and Design

General Terms

Algorithms, Performance, Theory

Keywords

Gossip protocol, epidemic protocol, evolving graphs, broadcasting

1. INTRODUCTION

Gossip protocols have been identified as simple, efficient, and robust mechanisms for various network and system tasks, such as, e.g., multicast [18, 25], resource location [23, 24], and distributed databases management [11, 22]. In parallel, the epidemiology community has considered several models for the analysis of the spreading of an infection in a population. These models include the famous SIR (Susceptible-Infective-Removed) and SIS (Susceptible-Infective-Susceptible) models, aiming at capturing the way a virus disseminates in a population as a function of the reaction to the virus experienced by every people [4]. Gossip (and epidemic) protocols tolerate a high degree of dynamism in their running environment. It is therefore of premier importance to evaluate the precise impact of this dynamism on the efficiency of the gossip protocols. It is indeed known that network dynamics can have a tremendous impact in certain circumstances. This impact can be quite positive whenever the network evolution is ergodic (e.g., when measuring the global bandwidth of an ad hoc radio network [17]), but also quite negative whenever the network evolution is arbitrary (e.g., when measuring the cover time of random walks [2]).

In the framework of gossip protocols and epidemiology, one way to handle dynamism is to assume that the network

*A part of this work was done during the stay of the second author at University Paris Diderot as Invited Professor. All authors were supported by COST Action 295 "DYNAMO". The first and third authors received additional supports from ANR project ALADDIN, and from INRIA project GANG. The second author received additional supports from the EU under the EU/IST Project 15964 AEOLUS.

PODC'09, August 10–12, 2009, Calgary, Alberta, Canada.

evolves with time as a sequence $(G_0, G_1, G_2, \dots)$ of graphs on the same set of vertices, where the graph G_t considered at time t is an Erdös-Renyi random graph drawn in $\mathcal{G}_{n,p}$. In this model, several investigations have been recently performed to measure the impact of the network evolution on the performances of algorithms. For instance, it was proved that radio broadcasting performs efficiently, even for p below the connectivity threshold of $\mathcal{G}_{n,p}$ (see [9]). In [2] it is proved that the cover time of a random walk remains polynomial for any $p > 0$. Threshold phenomena have also been identified; for instance, it was proved that SIR epidemic only contaminates a constant number of nodes if $p = c/n$ with $c < 1$, but contaminates a constant fraction of the nodes if $c > 1$ (see [12]).

In term of modeling real world networks, the elementary random process $(G_0, G_1, G_2, \dots)$ with $G_t \in \mathcal{G}_{n,p}$ suffers from the absence of time dependencies. The network does evolve but the structure of the network at time t is independent from its structure at time $t' < t$. This does not precisely reflect what is observed in many contexts such as wireless networks (the connection between two nodes is highly correlated to the previous existence of this connection) and P2P networks (the occurrence of an information exchange between two participants is highly correlated to the existence of previous exchanges) [29]. A more evolved model capturing time-dependencies has been recently considered in [8, 10]. Similarly to the elementary random process, the model, called *edge-Markovian* process, and denoted $\mathcal{M}_{n,p,q}$, generates a random sequence of graphs $(G_0, G_1, G_2, \dots)$ on the same node set $[n]$. This sequence is set based on a birth-rate p and death-rate q as follows[1]: G_0 is an Erdös-Renyi random graph in $\mathcal{G}_{n,\widehat{p}}$ where $\widehat{p} = p/(p+q)$, and, for any $t > 0$, a non-existing edge $e \notin E(G_{t-1})$ exists in $E(G_t)$ with probability p, while an existing edge $e \in E(G_{t-1})$ disappears from $E(G_t)$ with probability q.

In their companion papers [8, 10], Clementi et al. analyzed the flooding protocol in edge-Markovian dynamic graphs, i.e., in graph sequences generated by the edge-Markovian process. Flooding in dynamic graphs is the basic mechanism in which every node becoming aware of an information at step t forwards this information to all its neighbors at all forthcoming steps $t' > t$. Flooding is a core mechanism for information dissemination in contexts in which the network topology is highly dynamic and unknown, such as P2P networks, mobile networks, or any network susceptible to faults, and several variants of flooding designed to limit the bandwidth consumption have been proposed [6, 16, 26]. Clementi et al. produced several bounds on the flooding time in edge-Markovian dynamic graphs. In particular they proved that there is a wide class of dynamic graphs in which the flooding time does not (asymptotically) depends on the edge death-rate q.

Despite the interests of flooding in term of simplicity and robustness, this protocol suffers from a severe drawback in dynamic graphs: it requires every node, upon reception of a source message, to forward this message during *all* forthcoming time steps. This results in a waste of resources in terms both of link bandwidth and node computation. Of course, if one knows that flooding completes in T time steps, then all nodes can be bounded to be active for only this amount of time. Nevertheless, T is typically growing with the size n of the network, hence being active T steps still results in a significant waste of resources. Our objective is to force flooding to perform more parsimoniously, by limiting the number of steps during which every node is active in forwarding a message to its neighbors, yet allowing the message to eventually reach all nodes in short time. Parsimonious flooding enables to save bandwidth and computational resources, and potentially energy as well, the latter parameter being known to be crucial for ad hoc and sensor networks.

1.1 Parsimonious flooding

For a positive integer k, we say that a flooding protocol (in dynamic graphs) is *k-active* if each node forwards a source message only during the k time steps immediately following the step at which the node receives that message for the first time. For instance, the 1-active flooding protocol is the standard flooding protocol for static networks: a message is forwarded only once, at the step immediately following its reception. However, in dynamic networks, the flooding protocol may have to be active for $k > 1$ steps in order for the message to reach all nodes. The smaller the parameter k, the lesser the resource consumption by the protocol.

Our objective is to determine the minimum k for which the k-active flooding protocol completes correctly, i.e., the message eventually reaches all nodes. We define the *reachability threshold* for the flooding protocol in $\mathcal{M}_{n,p,q}$ as the smallest integer k such that, for any source $s \in [n]$, the k-active flooding protocol from s completes correctly almost surely[2]. Clearly if flooding completes in T steps in $\mathcal{M}_{n,p,q}$, then the reachability threshold of flooding in $\mathcal{M}_{n,p,q}$ is at most T. But in fact, we will show that, for a large spectrum of parameters p and q, the reachability threshold is just $o(T)$, and often even just constant. Moreover, we will also show that being active for a number of steps equal to the reachability threshold (up to a multiplicative constant) is sufficient for the flooding protocol to complete in *optimal* time, i.e., in asymptotically the same number of steps as when being perpetually active.

1.2 Previous work

There is a vast body of literature on broadcast and gossip protocols in static networks (see, e.g., the surveys [14, 19, 20]). Broadcasting in random graphs $\mathcal{G}_{n,p}$ has been analyzed in [13, 15, 27], and randomized gossip protocols in specific metrics have been analyzed in [23, 24]. In all these cases, there are no time-dependencies induced by any evolution of the network structure.

Several papers tackle information spreading problems in the context of wireless networks (see, e.g., [5, 28]) but the time-dependencies are either ignored, or overcame by assuming sufficiently long time slots for these dependencies to become negligible. In fact, the only information spreading work we are aware of, dealing explicitly with time-dependencies regarding the presence of the links, is [8, 10]. In the former, the model is the one considered in this paper, i.e., edge-Markovian dynamic graphs. In the latter, the authors consider other types of Markovian dynamics, including the *geo-*

[1] The general definition in [8] assumes that G_0 can be arbitrary, and the definition of edge-Markovian evolving graphs given here is called *stationary* edge-Markovian evolving graphs in [10].

[2] A series of events $\mathcal{E}_n$ holds almost surely (a.s.) if $\Pr[\mathcal{E}_n] \to 1$ when $n \to \infty$, i.e., $\Pr[\mathcal{E}_n] = 1 - o(1)$. These events holds with high probability (w.h.p.) if $\Pr[\mathcal{E}_n] \geq 1 - O(\frac{1}{n^\alpha})$ for some $\alpha > 0$.

metric Markovian evolving graphs, and give general bounds on the flooding time in Markovian dynamic graphs satisfying certain expansion properties.

More specifically, it is proved in [8] that, for any initial graph G_0 (i.e., not necessarily $G_0 \in \mathcal{G}_{n,\widehat{p}}$), and any birth-rate and death-rate $0 < p, q < 1$, the flooding time in edge-Markovian graphs is at most $O(\log n / \log(1 + np))$. For the stable graph G_0 (i.e., $E(G_0) = \emptyset$), it is proved that, for any $0 < p, q < 1$, the flooding time is at least $\Omega(\log n/np)$), and if $p \geq c \log n/n$ for $c > 1$ then the flooding time is at least $\Omega(\log n/ \log(1 + np))$.

The case $G_0 \in \mathcal{G}_{n,\widehat{p}}$ is considered in [10], where it is proved that for $\widehat{p} \geq c \log n/n$ with c large enough, the flooding time is at most

$$O\left(\frac{\log n}{\log n\widehat{p}} + \log\log n\widehat{p}\right)$$

and at least

$$\Omega\left(\frac{\log n}{\log n\widehat{p}}\right).$$

1.3 Our results

In this paper, we first revisit the results in [8, 10], and we give tight bounds on the flooding time (without any bound on the activity constraints) for all possible values of $p, q \in (0, 1)$ whenever $G_0 \in \mathcal{G}_{n,\widehat{p}}$. These bounds are summarized in Table 1, where $\widehat{p} = \frac{p}{p+q}$. For $\widehat{p} \geq \frac{c \log n}{n}$ with $c > 1$, flooding performs, a.s, in $\Theta(\frac{\log n}{\log(n\widehat{p})})$ rounds. For $0 < \widehat{p} \leq \frac{c}{n}$ with $c > 0$, flooding a.s. performs in $\Theta(\frac{\log n}{np})$ rounds. In between, the situation is more complex, and depends on the relative values of $\widehat{p}$ and p (see Figure 1). If $np \geq \log n\widehat{p}$ then the flooding time is $\Theta(\frac{\log n}{\log(n\widehat{p})})$, whereas if $np \leq \log n\widehat{p}$ then the flooding time is $\Theta(\frac{\log n}{np})$.

In parallel to the computation of the bounds on the flooding time, we have established tight bounds on the reachability threshold. If $\widehat{p} \geq \frac{c \log n}{n}$ with $c > 1$, then this parameter is equal to 1. That is, just one step of activity is enough to make sure that, a.s., the message reaches all nodes. If $\widehat{p} \leq \frac{c \log n}{n}$ with $c < 1$, then the reachability threshold is $\Theta(\frac{\log n}{np})$. (Note that the condition is on $\widehat{p}$ while the value for the reachability threshold depends on p).

Interestingly enough, for any $p, q \in (0, 1)$, we also prove that if k is the activity threshold, then, a.s., an $O(k)$-active flooding protocol completes in the same time as flooding without constraints on the activity, up to a multiplicative constant. In other words, the reachability threshold for the flooding protocol is essentially of the same order of magnitude as the activity threshold for this protocol to complete in optimal time, up to multiplicative constants. In particular, for $\widehat{p} \geq \frac{c \log n}{n}$ with $c > 1$, one step of activity is sufficient for flooding to complete in optimal time. Similarly, for $\frac{1}{n} \ll \widehat{p} \leq \frac{c \log n}{n}$ with $c < 1$, and $np \geq \log n\widehat{p}$, the reachability threshold $\Theta(\frac{\log n}{np})$ is significantly smaller than the optimal flooding time $\Theta(\frac{\log n}{\log(n\widehat{p})})$, yet being active for $O(\frac{\log n}{np})$ steps is sufficient for flooding to complete in asymptotically optimal time $O(\frac{\log n}{\log(n\widehat{p})})$. For all the remaining cases, the thresholds for completion and for optimality coincide, up to multiplicative constants.

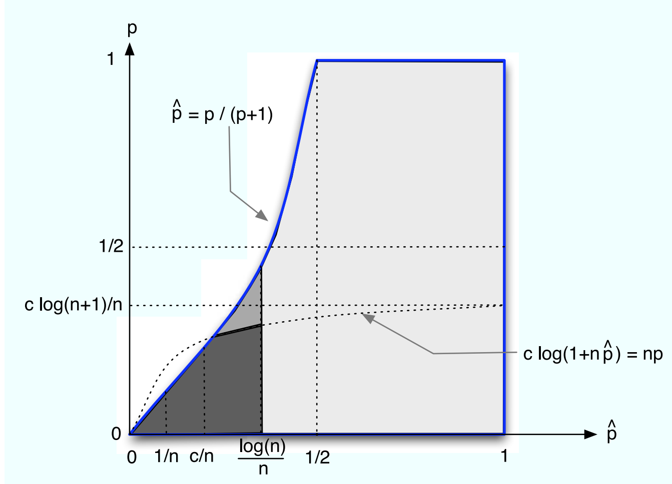

Figure 1: Graphical visualization of our results. The light grey zone corresponds to combinations of parameters p and $\widehat{p}$ where 1 step of activity insures optimal flooding time. The medium grey zone is where the reachability threshold is significantly smaller than optimal flooding time. Finally, the dark grey zone is where the reachability threshold is of the same order of magnitude as the optimal flooding time.

2. MODEL AND NOTATIONS

Let us first recall the definition of the edge-Markovian dynamic graphs, introduced in [8, 10]. In the following, n will denote a positive integer, and p and q will denote two real values such that $0 < p < 1$, and $0 < q < 1$. We set $\widehat{p} = \frac{p}{p+q}$. An edge-Markovian process $\mathcal{M}_{n,p,q}$ with birth-rate p and death-rate q generates sequences of graphs $(G_0, G_1, G_2, \ldots)$ with the same node set $[n] = \{1, \ldots, n\}$ such that (1) $G_0 \in \mathcal{G}_{n,\widehat{p}}$ is an Erdös-Renyi random graph, and (2), for any $t > 0$, G_t is obtained from G_{t-1} by using the Markovian rule with parameters p and q defined by:

- if $e \notin E(G_{t-1})$ then $e \in E(G_t)$ with probability p and $e \notin E(G_t)$ with probability $1 - p$;
- if $e \in E(G_{t-1})$ then $e \notin E(G_t)$ with probability q and $e \in E(G_t)$ with probability $1 - q$.

Remark. Since $G_0 \in \mathcal{G}_{n,\widehat{p}}$, we have $G_t \in \mathcal{G}_{n,\widehat{p}}$ for any $t \geq 0$. However, the distribution of G_t knowing $G_{t'}$ for $t' \leq t - 1$ is more complex.

By slightly abusing terminology, we will say that a node belongs to $\mathcal{M}_{n,p,q}$ in the sense that it belongs to all graphs G_t, $t \geq 0$. For a sequence of graphs $S = (G_0, G_1, G_2, \ldots)$, we will note $S \in \mathcal{M}_{n,p,q}$ to specify that the probability space for S is the one of all sequences generated by $\mathcal{M}_{n,p,q}$.

Our interest is the flooding protocol in edge-Markovian graphs, as analyzed in [8] and [10]. The flooding protocol in $\mathcal{M}_{n,p,q}$ from an arbitrary source $s \in [n]$ is performed by all nodes in $\mathcal{M}_{n,p,q}$, in synchronous steps, as follows. At every time step, every node which has received the source message during a previous step forwards this message to all its neighbors. More formally, initially there is a single informed node $s \in [n]$. Let I_t be the set of all nodes which hold the message at time t. Hence $I_0 = \{s\}$. For $t \geq 0$, I_{t+1} is defined from I_t by $I_{t+1} = I_t \cup N_t$ where N_t is the set of

	$0 < \widehat{p} \le \frac{c}{n}$, $c > 0$	$\frac{1}{n} \ll \widehat{p} \le \frac{c \log n}{n}$, $c < 1$ $np \le \log n\widehat{p}$	$np \ge \log n\widehat{p}$	$\widehat{p} \ge \frac{c \log n}{n}$, $c > 1$
Flooding time	$\Theta(\frac{\log n}{np})$	$\Theta(\frac{\log n}{np})$	$\Theta(\frac{\log n}{\log(n\widehat{p})})$	$\Theta(\frac{\log n}{\log(n\widehat{p})})$
Reachability threshold	$\Theta(\frac{\log n}{np})$	$\Theta(\frac{\log n}{np})$	$\Theta(\frac{\log n}{np})$	1

Table 1: Summary of results

all nodes that are neighbors in G_t of at least one node in I_t. Flooding always reaches all nodes eventually, since, for any node $u \in [n]$, the probability that no edge exists between u and s during t steps is $(1-\widehat{p})(1-p)^{t-1} \to 0$ as $t \to \infty$. Of course, node u may receive the message much before an edge $\{s, u\}$ exists, through intermediate nodes and longer paths.

Given n, p, q, the random variable $T_s = T_s(n, p, q)$ is the flooding time from s in $\mathcal{M}_{n,p,q}$, defined by

$$T_s = \min\{t \ge 0 \mid I_t = [n]\}.$$

The flooding protocol as described above assumes that all nodes are perpetually active. That is, upon reception of the source message, every node will forward this message to all its neighbors during all forthcoming time steps. We are interested in bounding the time interval during which a node is active, while still permitting flooding to complete. Let k be a positive integer. A k-active flooding protocol is a flooding protocol in which each node forwards the source message only during the k time steps immediately following the step at which the node receives the message for the first time. That is, by setting $I_t^{(k)}$ as the set of all nodes which hold the message at time t of the execution of the k-active flooding protocol, we have $I_{t+1}^{(k)} = I_t^{(k)} \cup N_t^{(k)}$ where $N_t^{(k)}$ is the set of all nodes that are neighbors in G_t of at least one node in $I_t^{(k)} \setminus I_{t-k}^{(k)}$. Hence, if $k = \infty$, the k-active flooding protocol coincides with the standard flooding protocol analyzed in [8] and in [10], while if $k = 1$ the k-active flooding protocol corresponds to the case in which each node forwards the source message only once (at the step immediately following the message reception), as for flooding in a static network.

We are interested in k-active flooding protocols that are sufficiently active for insuring that the source message reaches all nodes. For that purpose, for the k-active flooding protocol, we define the random variable

$$T_s^{(k)} = \min\{t \ge 0 \mid I_t^{(k)} = [n]\}.$$

We are interested in the following parameter:

Definition 1. *The* reachability threshold *for the flooding protocol in $\mathcal{M}_{n,p,q}$ is the smallest integer k such that $T_s^{(k)} < \infty$ almost surely*[3]*, for any $s \in [n]$.*

In the forthcoming sections, we will give tight lower and upper bounds on the reachability threshold and on the flooding time, for all ranges of birth- and death-rates.

[3]Formally, the value of the reachability threshold depends on the guarantee placed on the event $T_s^{(k)} < \infty$, and may vary depending on the setting of the term $o(1)$ in the definition of "almost surely".

3. REDUCING COMPLETION TIME TO DIAMETER

One of the main difficulties in the analysis of dynamic graphs is to handle time dependencies. In this section, we introduce a tool that will be shown quite useful for reducing the problem of determining the completion time of a flooding protocol in dynamic graphs into the problem of computing the diameter of a family of random weighted graphs. The benefit of this tool is twofold. First, the analysis of flooding is reduced to computing diameters (the two parameters coincide in the static case, but are a priori quite different in the dynamic setting). Second, the analysis of sequences of time-dependent random graphs is reduced to the analysis of a single random graph.

We reduce the analysis of flooding in $\mathcal{M}_{n,p,q}$ into computing the diameter of *weighted* random graphs defined as follows. Let a and b be two real numbers such that $0 < a < 1$, and $0 < b < 1$. Let Z be the random variable with values in the set of positive integers, defined by

$$\Pr\{Z = z\} = \begin{cases} a & \text{if } z = 1, \\ (1-a)(1-b)^{z-2}b & \text{if } z > 1. \end{cases}$$

Given two random variables X and X', we denote by $X \sim X'$ the fact that X and X' are identically distributed (i.e., $\Pr\{X = x\} = \Pr\{X' = x\}$ for any x). An n-node clique G belongs to the family of weighted random graphs $\mathcal{G}_{n,a,b}$ whenever each of its $\binom{n}{2}$ possible edges e is given a random positive integral weight equal to Z_e, independently from the other edges, where $Z_e \sim Z$. To capture the activity of flooding protocols, we define the family of k-*bounded* weighted random graphs $\mathcal{G}_{n,a,b}^{(k)}$ as follows: the node set is $[n]$, and an edge e is present only if $Z_e \le k$, and whenever the edge e is present it receives the weight Z_e.

By definition, $\mathcal{G}_{n,a,b} = \mathcal{G}_{n,a,b}^{(\infty)}$. Moreover, each edge e of $\mathcal{G}_{n,a,b}^{(k)}$ is present with probability $\varrho = \sum_{z=1}^{k} \Pr\{Z = z\}$, independently from the other edges. Hence, if one ignores the edge weights, then $\mathcal{G}_{n,a,b}^{(k)} = \mathcal{G}_{n,\varrho}$. More generally, by definition, $\mathcal{G}_{n,a,b}^{(1)} = \mathcal{G}_{n,a}$ for any b. Therefore, by several aspects, the k-bounded weighted random graphs generalize the (standard) random graphs.

The distance between two nodes in a weighted graph is the minimum weight of any path connecting these nodes in the graph, where the weight of a path is the sum of the weights of its edges. The eccentricity of a node is the largest distance between this node and any other node in the graph. For any $s \in [n]$, let $X_s^{(k)}$ be the random variable equal to the eccentricity of s in the k-bounded weighted random graph $\mathcal{G}_{n,\widehat{p},p}^{(k)}$. The following lemma relates the random variable $X_s^{(k)}$ to the flooding time $T_s^{(k)}$ in edge-Markovian processes, and states that these two variables have the same probability distribution. ($\mathbb{N}^+$ denotes the set of positive integers).

Lemma 1. (Reduction Lemma) *For any $n \geq 1$, any $0 < p, q < 1$, any $k \in \mathbb{N}^+ \cup \{\infty\}$, and any $s \in [n]$, we have $T_s^{(k)} \sim X_s^{(k)}$, that is*

$$\Pr(T_s^{(k)} = x) = \Pr(X_s^{(k)} = x), \quad \forall x \geq 0,$$

where the probability space for the l.h.s. of the equality is $\mathcal{M}_{n,p,q}$, while the one for the r.h.s. is $\mathcal{G}^{(k)}_{n,\widehat{p},p}$.

Proof. Fix $s \in [n]$ and $k \in \mathbb{N}^+ \cup \{\infty\}$. Let $S = (G_0, G_1, \ldots)$ be an arbitrary realization of the edge-Markovian process $\mathcal{M}_{n,p,q}$. We define a weighted graph H in which the eccentricity of s will be proved to be equal to the completion time of the k-active flooding protocol in S, applied from s. For any node $u \in [n]$, let $t(u)$ denote the time step at which u receives the message for the first time, i.e.:

$$t(u) = \min\{t \mid u \notin \cup_{i=0}^{t-1} I_i^{(k)} \wedge u \in I_t^{(k)}\}.$$

(Recall that $I_t^{(k)}$ denotes the set of nodes informed during the t first steps of the k-active flooding protocol, for any $t \geq 0$, with $I_0^{(k)} = \{s\}$). Note that for $k < \infty$, we may have $t(u) = \infty$. For any edge $e = \{u, v\}$ with $u, v \in [n]$ and $u \neq v$, let

$$t(e) = \begin{cases} \min\{t(u), t(v)\} & \text{if } t(u) < \infty \text{ or } t(v) < \infty; \\ -1 & \text{otherwise.} \end{cases}$$

The -1 in the above definition is for edges whose extremities have not received the message, and is somewhat arbitrary. Also, for any edge e, let

$$t_{\text{after}}(e) = \min\{t \mid t > t(e) \wedge e \in E(G_t)\}.$$

Thus, if $t(e) \geq 0$, $t_{\text{after}}(e)$ denotes the first step at which the edge e appears after one of its two endpoints has been informed for the first time. Otherwise, $t_{\text{after}}(e)$ denotes the first step at which the edge e appears. We define the weighted graph H as follows. $V(H) = [n]$, and, for any two distinct nodes $u, v \in [n]$:

$$e = \{u, v\} \in E(H) \iff t_{\text{after}}(e) - t(e) \leq k$$

Moreover, if $e \in E(H)$, then the weight of e is equal to

$$w(e) = t_{\text{after}}(e) - t(e).$$

We denote by ϕ the mapping that, given S, returns H.

For any $u \in V(H)$, let $d(u)$ be the weighted distance from s to u in H. One important property of ϕ is the following :

Claim 1. Let $T_s(S)$ be the completion time of the k-active flooding protocol from s in S, and let $ecc_H(s)$ be the eccentricity of s in $H = \phi(S)$. For any $u \in [n]$, we have $d(u) = t(u)$. As a consequence, we have $T_s(S) = ecc_H(s)$.

Proof. Let $u \in [n]$. Let us first show that $t(u) \leq d(u)$. If u is not connected to s in H, then $d(u) = \infty$ and thus $t(u) \leq d(u)$. So assume that u is connected to s in H. Let $u_0, u_1, u_2, \ldots, u_{\ell-1}, u_\ell$ be a shortest path in H from $s = u_0$ to $u = u_\ell$. Assume for the purpose of contradiction that there exists an edge e of this path with $t(e) = -1$. Then let $i > 0$ be the smallest index such that $t(\{u_i, u_{i+1}\}) = -1$. Node u_i has not received the message in S, though u_{i-1} has received the message. Since the flooding is k-active, this implies that $t_{\text{after}}(\{u_{i-1}, u_i\}) - t(\{u_{i-1}, u_i\}) > k$, and thus $\{u_{i-1}, u_i\} \notin E(H)$, a contradiction. Therefore all edges of the path satisfies $t(\{u_i, u_{i+1}\}) \geq 0$. As a consequence, for every $i = 0, \ldots, \ell-1$, we have $t(u_{i+1}) \leq t(u_i) + w(\{u_i, u_{i+1}\})$. The weight of the path is thus equal to

$$\begin{aligned} d(u) &= \sum_{i=0}^{\ell-1} w(\{u_i, u_{i+1}\}) \\ &\geq \sum_{i=0}^{\ell-1} (t(u_{i+1}) - t(u_i)) \\ &= t(u_\ell) - t(u_0) = t(u) - t(s) = t(u). \end{aligned}$$

Let us now show that $t(u) \geq d(u)$. If $t(u) = \infty$, then this holds. So assume that $t(u) < \infty$. There exists a sequence of nodes $u_0, u_1, u_2, \ldots, u_{\ell-1}, u_\ell$ such that $s = u_0$, $u = u_\ell$, and, for any i with $1 \leq i \leq \ell$, u_i has been informed by u_{i-1}. This implies that, for any i with $1 \leq i \leq \ell$, the edge $e_i = \{u_{i-1}, u_i\}$ belongs to $E(H)$ because the flooding protocol is k-active. Moreover,

$$w(e_i) = t_{\text{after}}(e_i) - t(e_i) = t_{\text{after}}(e_i) - t(u_{i-1}) = t(u_i) - t(u_{i-1}).$$

Hence, there is a path from s to u whose weight $\sum_{i=1}^{\ell} w(e_i)$ is equal to $t(u) - t(s) = t(u)$. Thus, for any node u, the distance $d(u)$ from s to u in H is at most $t(u)$. This completes the proof of the claim. ◇

According to Claim 1, the completion time of the k-active flooding protocol from s in S can be computed by evaluating the eccentricity of s in the graph $H = \phi(S)$. This graph H is clearly a random graph (which depends on the realization S of the edge-Markovian process $\mathcal{M}_{n,p,q}$). On the other hand, H can also be viewed as one realization of $\mathcal{G}^{(k)}_{n,\widehat{p},p}$, simply because the weight of an edge is between 1 and k. Hence, in particular, ϕ can be viewed as a mapping from $\mathcal{M}_{n,p,q}$ to $\mathcal{G}^{(k)}_{n,\widehat{p},p}$. Let $\sim$ be the equivalence relation in $\mathcal{M}_{n,p,q}$ defined by

$$S \sim S' \iff \phi(S) = \phi(S').$$

By this definition, ϕ becomes a one-to-one mapping from $\mathcal{M}_{n,p,q}/\!\sim$ to $\mathcal{G}^{(k)}_{n,\widehat{p},p}$. In fact, ϕ is also onto. To see why, let H be a realization of $\mathcal{G}^{(k)}_{n,\widehat{p},p}$. For any pair of nodes $e = \{u, v\}$, $u, v \in [n]$, $u \neq v$, let

$$d(e) = \begin{cases} \min\{d(u), d(v)\} & \text{if } d(u) < \infty \text{ or } d(v) < \infty; \\ -1 & \text{otherwise.} \end{cases}$$

Define $S = (G_0, G_1, \ldots)$ as any realization of $\mathcal{M}_{n,p,q}$ in which, for any edge e, and for any $t \in [0, kn)$, we have that $e \in E(G_t)$ if and only if $e \in E(H)$ and $t = d(e) + w(e)$. (For $t \geq kn$, S is arbitrary). By construction, $\phi(S) = H$. Therefore,

$$\begin{array}{cccc} \phi: & \mathcal{M}_{n,p,q}/\!\sim & \rightarrow & \mathcal{G}^{(k)}_{n,\widehat{p},p} \\ & S & \mapsto & H \end{array}$$

is a one-to-one and onto mapping. We now show that ϕ is a morphism, that is it preserves probability :

Claim 2. For any $H \in \mathcal{G}^{(k)}_{n,\widehat{p},p}$, $Pr\{\phi^{-1}(H)\} = Pr\{H\}$, where the probability space for the l.h.s. of the equality is $\mathcal{M}_{n,p,q}$, while the one for the r.h.s. is $\mathcal{G}^{(k)}_{n,\widehat{p},p}$.

Proof. Let $H \in \mathcal{G}^{(k)}_{n,\widehat{p},p}$. By definition of ϕ, the set $\phi^{-1}(H)$ is equal to the set of all sequences $S = (G_0, G_1, \ldots) \in \mathcal{M}_{n,p,q}$ such that, for any pair of nodes e, we have:

- if $e \in E(H)$ then $\left(\bigwedge_{i=1}^{w(e)-1}\left(e \notin E(G_{t(e)+i})\right)\right) \wedge \left(e \in E(G_{t(e)+w(e)})\right)$;
- if $e \notin E(H)$ then $\bigwedge_{i=1}^{k}\left(e \notin E(G_{t(e)+i})\right)$.

Let us define the corresponding event, with $d(e)$ replacing $t(e)$ in the above formula:

$$\begin{aligned}\mathcal{E}(e) &= \left[(e \in E(H)) \wedge \left(\bigwedge_{i=1}^{w(e)-1}\left(e \notin E(G_{d(e)+i})\right)\right)\right.\\ &\quad \left.\wedge \left(e \in E(G_{d(e)+w(e)})\right)\right]\\ &\quad \bigvee \left[(e \notin E(H)) \wedge \left(\bigwedge_{i=1}^{k}\left(e \notin E(G_{d(e)+i})\right)\right)\right]\end{aligned}$$

From Claim 1, we know that if $S \in \phi^{-1}(H)$, then $t(u) = d(u)$ for any node $u \in [n]$, and hence for any pair of nodes e, $d(e) = t(e)$. Therefore,

$$\Pr\left\{\phi^{-1}(H)\right\} = \Pr\left\{\bigwedge_{e}\Big(\mathcal{E}(e) \wedge (t(e) = d(e))\Big)\right\}.$$

Let us sort the events $\mathcal{E}(e)$ in increasing order of the distance of e from s in H. So, for $e = \{u, v\}$, let $d_{min}(e) = \min\{d(u), d(v)\}$. Note that whenever $d_{min}(e) < \infty$, we have $d_{min}(e) = d(e)$. We get a list of events $\mathcal{E}(e_1), \ldots, \mathcal{E}(e_m)$ with $m = \binom{n}{2}$, where $0 \le d_{min}(e_i) \le d_{min}(e_{i+1})$ for $i = 1, \ldots, m-1$. We show, by induction on z, that

$$\Pr\left\{\bigwedge_{i=1}^{z}\Big(\mathcal{E}(e_i) \wedge (t(e_i) = d(e_i))\Big)\right\} = \prod_{i=1}^{z}\mathcal{P}(e_i)$$

where, for $e \in E(H)$,

$$\mathcal{P}(e) = \begin{cases} \widehat{p} & \text{if } w(e) = 1,\\ (1-\widehat{p})(1-p)^{w(e)-2}p & \text{otherwise}\end{cases}$$

and, for $e \notin E(H)$,

$$\mathcal{P}(e) = (1-\widehat{p})(1-p)^{k-1}.$$

This property trivially holds for $z = 0$. Assume that it holds for $z \ge 0$. By induction hypothesis, we have

$$\begin{aligned}&\Pr\left\{\bigwedge_{i=1}^{z+1}\Big(\mathcal{E}(e_i) \wedge (t(e_i) = d(e_i))\Big)\right\}\\ &= \Pr\left\{\Big(\mathcal{E}(e_{z+1}) \wedge (t(e_{z+1}) = d(e_{z+1}))\Big)\right.\\ &\qquad \left.\mid \bigwedge_{i=1}^{z}\Big(\mathcal{E}(e_i) \wedge (t(e_i) = d(e_i))\Big)\right\} \cdot \prod_{i=1}^{z}\mathcal{P}(e_i).\end{aligned}$$

Now we show that the probability of the left product member is equal to $\mathcal{P}(e_{z+1})$. For this purpose, let us rewrite this expression as

$$\begin{aligned}&\Pr\left\{\mathcal{E}(e_{z+1}) \mid (t(e_{z+1}) = d(e_{z+1}))\right.\\ &\qquad \left.\wedge \bigwedge_{i=1}^{z}\Big(\mathcal{E}(e_i) \wedge (t(e_i) = d(e_i))\Big)\right\}\\ &\times \Pr\left\{t(e_{z+1}) = d(e_{z+1}) \mid \bigwedge_{i=1}^{z}\Big(\mathcal{E}(e_i) \wedge (t(e_i) = d(e_i))\Big)\right\}\end{aligned}$$

We compute separately the two terms of the product above.

Let us concentrate on the first conditional event. By the edge-independence of the edge-Markovian process, the event $\mathcal{E}(e_{z+1})$ is independent from the event $\bigwedge_{i=1}^{z}\mathcal{E}(e_i)$ because they are different edges, and because the $d(e_i)$'s are fixed parameters (there is thus no dependent timing issues, just independent existence or non existence of edges).

We claim that the event $\mathcal{E}(e_{z+1})$ is also independent from the event $\bigwedge_{i=1}^{z+1}(t(e_i) = d(e_i))$. The intuition for this claim is that the behavior of an edge after a given time t is independent from events related to the reception of the message by nodes at time $t' < t$. And the behavior of an edge whose extremities have not received the message is independent from the reception of the message by nodes at finite times, or the non reception of the message by some other nodes. More formally, let us focus on events of the form "$t(e) = d(e)$" for some $e = \{u, v\}$, $u \ne v$. For this purpose, let $\mathcal{T}(e, t)$ denote the event "$t(e) = t$".

We have

$\mathcal{T}(e, 0) = \Big((u = s) \vee (v = s)\Big)$, and

$\mathcal{T}(e, 1) = \overline{\mathcal{T}(e, 0)} \wedge \Big((\{s, u\} \in E(G_1)) \vee (\{s, v\} \in E(G_1))\Big)$.

More generally, for $t \ge 2$, we have

$$\mathcal{T}(e, t) = \left(\bigwedge_{t'=0}^{t-1}\overline{\mathcal{T}(e, t')}\right)\bigwedge \mathcal{T}'(e, t)$$

where

$$\begin{aligned}\mathcal{T}'(e, t) =& \Big((\{s, u\} \in G_t \wedge t \le k) \vee (\{s, v\} \in G_t \wedge t \le k)\Big)\\ &\vee \bigvee_{\ell=1}^{t-1}\ \bigvee_{\substack{(w_1,\ldots,w_\ell)\\ \in([n]\setminus\{u,v\})^\ell}}\ \bigvee_{\substack{(t_1,\ldots,t_{\ell+1})\in\{1,\ldots,k\}^{\ell+1}\\ \sum_{r=1}^{\ell+1} t_r = t}}\\ &\left[\Big(\{s, w_1\} \in E(G_{t_1})\Big)\right.\\ &\wedge \left(\bigwedge_{j=1}^{\ell-1}\Big(\{w_j, w_{j+1}\} \in G_{\sum_{r=1}^{j+1} t_r}\Big)\right)\\ &\left.\wedge \Big(\{w_\ell, u\} \in G_{\sum_{r=1}^{\ell+1} t_r} \vee \{w_\ell, v\} \in G_{\sum_{r=1}^{\ell+1} t_r}\Big)\right]\end{aligned}$$

Finally, for $t = \infty$,

$$\mathcal{T}(e, \infty) = \bigwedge_{t \ge 0}\overline{\mathcal{T}(e, t)}.$$

Therefore, the event $\mathcal{T}(e, t)$ can be expressed as a Boolean formula including solely logical expressions of the type "$u = s$" or "$v = s$" or "$e' \in E(G_{t'})$", where $e' \ne e$, and $t' \le t$.

Most importantly, by defining $t_{min}(e') = \min\{t(w), t(w')\}$ for any edge $e' = \{w, w'\}$, we get that the events in $\mathcal{T}(e,t)$ does not relate to edges e' satisfying $t_{min}(e') < t$.

As a consequence, for $e = e_{z+1}$ and $t = d_{min}(e_{z+1})$, we get that the event $\bigwedge_{i=1}^{z+1}(t(e_i) = d(e_i))$ depends only on edges e' with $t_{min}(e') < d_{min}(e_{z+1})$. (This is because $d_{min}(e_i) \leq d_{min}(e_{z+1})$ for any $i = 1, \ldots, z$). Thus the event $\bigwedge_{i=1}^{z+1}(t(e_i) = d(e_i))$ does not depend on $\mathcal{E}(e_{z+1})$ because this latter event specifies the behavior of e_{z+1} at times at least $d_{min}(e_{z+1})$. Therefore,

$$\begin{aligned}\Pr\Big\{&\mathcal{E}(e_{z+1}) \mid (t(e_{z+1}) = d(e_{z+1})) \\ &\wedge \bigwedge_{i=1}^{z}\Big(\mathcal{E}(e_i) \wedge (t(e_i) = d(e_i))\Big)\Big\} \\ = \ &\Pr\{\mathcal{E}(e_{z+1})\}.\end{aligned}$$

Now just remains to compute

$$\Pr\left\{t(e_{z+1}) = d(e_{z+1}) \mid \bigwedge_{i=1}^{z}\Big(\mathcal{E}(e_i) \wedge (t(e_i) = d(e_i))\Big)\right\}.$$

Let $H' = \cup_{1\leq i\leq z} e_i$ be the (weighted) graph induced by edges e_i, $1 \leq i \leq z$. Whenever the event $\bigwedge_{i=1}^{z}\Big(\mathcal{E}(e_i) \wedge (t(e_i) = d(e_i))\Big)$ holds, the weights of these edges e_i in H' are the same as their weights $w(e_i)$ in H (by giving infinite weights to edge not in H). Therefore, for every node u in H', the following two facts hold: (1) the distance $d'(u)$ from u to s in H' is the same as the distance $d(u)$ from u to s in H, and (2) the time $t'(u)$ at which node u receives the message from s in the flooding protocol restricted to the edges of H' is the same as the time $t'(u)$ at which node u receives the message from s in S. As a consequence, $t'(u) = d'(u)$ for every node in H'. On the other hand, from our analysis of the event $\mathcal{T}$, we get that the event "$t(e_{z+1}) = d(e_{z+1})$" depends only of edges e with $t(e) < d(e_{z+1})$, hence only of edges e_i with $1 \leq i \leq z$. In other words, the event "$t(e_{z+1}) = d(e_{z+1})$" is fully determined by H'. Therefore, at least one extremity of e_{z+1} belongs to H' and we have $t(e_{z+1}) = t'(e_{z+1}) = d'(e_{z+1}) = d(e_{z+1})$. As a consequence,

$$\Pr\left\{t(e_{z+1}) = d(e_{z+1}) \mid \bigwedge_{i=1}^{z}\Big(\mathcal{E}(e_i) \wedge (t(e_i) = d(e_i))\Big)\right\} = 1.$$

Therefore

$$\begin{aligned}\Pr\Big\{&\Big(\mathcal{E}(e_{z+1}) \wedge (t(e_{z+1}) = d(e_{z+1}))\Big) \\ &\mid \bigwedge_{i=1}^{z}\Big(\mathcal{E}(e_i) \wedge (t(e_i) = d(e_i))\Big)\Big\} \\ = \ &\Pr\{\mathcal{E}(e_{z+1})\}.\end{aligned}$$

Since $\Pr\{\mathcal{E}(e_{z+1})\} = \mathcal{P}(e_{z+1})$, we finally obtain

$$\Pr\left\{\bigwedge_{i=1}^{z+1}\Big(\mathcal{E}(e_i) \wedge (t(e_i) = d(e_i))\Big)\right\} = \prod_{i=1}^{z+1}\mathcal{P}(e_i)$$

which completes the proof of the induction.

For $z = m$, we get

$$\begin{aligned}\Pr\{\phi^{-1}(H)\} &= \Pr\left\{\bigwedge_{i=1}^{m}\Big(\mathcal{E}(e_i) \wedge (t(e_i) = d(e_i))\Big)\right\} \\ &= \prod_{i=1}^{m}\mathcal{P}(e_i).\end{aligned}$$

This probability is precisely the probability of $H \in \mathcal{G}^{(k)}_{n,\widehat{p},p}$. This completes the proof of the claim. ◇

We have now all ingredients to establish the lemma. Let $x \geq 0$. Let

$$\mathcal{S} = \{S \in \mathcal{M}_{n,p,q} \mid T_s^{(k)}(S) = x\},$$

and let

$$\mathcal{H} = \{H \in \mathcal{G}^{(k)}_{n,\widehat{p},p} \mid \mathrm{ecc}_H(s) = x\},$$

so that

$$\Pr(T_s^{(k)} = x) = \sum_{S\in\mathcal{S}}\Pr(S) \text{ and } \Pr(X_s^{(k)} = x) = \sum_{H\in\mathcal{H}}\Pr(H).$$

From Claim 1, ϕ maps $\mathcal{S}$ to $\mathcal{H}$. Moreover since the mapping $\phi : \mathcal{S}/\!\sim\, \to \mathcal{H}$ is one-to-one and onto, we get

$$\Pr(T_s^{(k)} = x) = \sum_{H\in\mathcal{H}}\Pr(\phi^{-1}(H)).$$

Thus, by Claim 2,

$$\Pr(T_s^{(k)} = x) = \sum_{H\in\mathcal{H}}\Pr(H),$$

that is

$$\Pr(T_s^{(k)} = x) = \Pr(X_s^{(k)} = x),$$

which completes the proof of the lemma. □

We complete this section by pointing out a direct corollary of the Reduction Lemma. That is this version of the Lemma that we shall actually use in the rest of the paper. Let $\mathrm{diam}(\mathcal{G}^{(k)}_{n,a,b})$ denote the random variable equal to the diameter of the weighted random graph $\mathcal{G}^{(k)}_{n,a,b}$. The corollary below follows after noticing the simple fact that the diameter of a graph is at least the eccentricity of each of its nodes, and at most twice this eccentricity.

Corollary 1. *For any $n \geq 1$, any $0 < p, q < 1$, any $k \in \mathbb{N}^+ \cup \{\infty\}$, and any $s \in [n]$, we have*

$$\begin{aligned}\forall x \geq 0, \Pr\Big(\mathit{diam}(\mathcal{G}^{(k)}_{n,\widehat{p},p}) \leq x\Big) &\leq \Pr\Big(T_s^{(k)} \leq x\Big) \\ &\leq \Pr\Big(\mathit{diam}(\mathcal{G}^{(k)}_{n,\widehat{p},p}) \leq 2x\Big).\end{aligned}$$

Proof. Let $x \geq 0$. The Reduction Lemma states that $\Pr(T_s^{(k)} = x) = \Pr(X_s^{(k)} = x)$. We have:

$$\mathrm{diam}(\mathcal{G}^{(k)}_{n,\widehat{p},p}) \leq x \ \Rightarrow\ X_s^{(k)} \leq x \ \Rightarrow\ T_s^{(k)} \leq x$$

and

$$T_s^{(k)} \leq x \ \Rightarrow\ X_s^{(k)} \leq x \ \Rightarrow\ \mathrm{diam}(\mathcal{G}^{(k)}_{n,\widehat{p},p}) \leq 2x.$$

□

The Reduction Lemma and the corollary above are the main ingredient we use for computing the reachability threshold of flooding. This is done in the next section.

4. REACHABILITY THRESHOLD AND OPTIMAL FLOODING TIME

In this section, we compute the reachability threshold of the flooding protocol in edge-Markovian graphs $\mathcal{M}_{n,p,q}$, for the whole spectrum of $p, q \in (0,1)$. We show that the flooding time remains of the same order of magnitude as the optimal whenever flooding is bounded to be k-active, where k is (up to a multiplicative constant) equal to the reachability threshold. We distinguish the two standard regimes: beyond the connectivity threshold of $\mathcal{G}_{n,\widehat{p}}$, and below its connectivity threshold.

4.1 Beyond the connectivity threshold

If $\widehat{p} \geq c\,\frac{\log n}{n}$ with $c > 1$, then, a.s, a graph in any sequence generated by $\mathcal{M}_{n,p,q}$ is connected[4]. It is thus expected that flooding in $\mathcal{M}_{n,p,q}$ would complete at least at fast as in $\mathcal{G}_{n,\widehat{p}}$. This was established in [10] where it was proved that if $\widehat{p} \geq c\,\frac{\log n}{n}$ for a sufficiently large constant c, then, w.h.p., $T_s = O(\frac{\log n}{\log(n\widehat{p})} + \log\log(n\widehat{p}))$ and $T_s = \Omega(\frac{\log n}{\log(n\widehat{p})})$. The Reduction Lemma allows us to extend these results by proving that, a.s., $T_s = \Theta(\frac{\log n}{\log(n\widehat{p})})$ whenever $n\widehat{p} - \log n \to \infty$. On the other hand, the Reduction Lemma also allows us to prove that, if $n\widehat{p} - \log n \to \infty$ then, a.s., only *one* step of activity is sufficient for flooding to complete in $O(\frac{\log n}{\log(n\widehat{p})})$ steps. Hence, the reachability threshold for flooding in edge-Markovian graphs with $n\widehat{p} - \log n \to \infty$ is just 1, and this single step of activity is sufficient to flood in optimal time, i.e., asymptotically as quickly (up to a multiplicative constant) as when the activity is unbounded.

THEOREM 1. *For any birth-rate p and death-rate q such that $n\widehat{p} - \log n \to \infty$, the completion time of the 1-active flooding protocol in $\mathcal{M}_{n,p,q}$ from any source $s \in [n]$ is, a.s., $O(\frac{\log n}{\log(n\widehat{p})})$. Moreover, this completion time is asymptotically equal to the optimal flooding time.*

PROOF. By the Reduction Lemma, and Corollary 1, $\Pr\{T_s^{(1)} \leq x\} \leq \Pr\{\mathrm{diam}(\mathcal{G}^{(1)}_{n,\widehat{p},p}) \leq 2x)\}$. Now, $\mathcal{G}^{(1)}_{n,\widehat{p},p} = \mathcal{G}_{n,\widehat{p}}$. Moreover, if $n\widehat{p} - \log n \to \infty$, then $\mathcal{G}_{n,\widehat{p}}$ is a.s.connected [3]. In fact, it has been proved (see [7]) that the diameter[5] of $\mathcal{G}_{n,\widehat{p}}$ is, a.s., $O(\frac{\log n}{\log(n\widehat{p})})$. Therefore, a.s., $T_s^{(1)} \leq O(\frac{\log n}{\log(n\widehat{p})})$.

As mentioned before in the paper, the fact that the (fully active) flooding protocol cannot complete faster than $\Omega(\frac{\log n}{\log(n\widehat{p})})$ step has been proved in [10] for $\widehat{p} \geq c\,\frac{\log n}{n}$ for a sufficiently large constant c. The fact that this lower bound holds on the whole range of values for $\widehat{p}$ such that $n\widehat{p} - \log n \to \infty$ will be proved later in the proof of Theorem 2. □

Remark. Most of our results are given with guarantee "almost surely" (a.s.). This is to fit with the statements in the literature regarding random graphs. However, by the Reduction Lemma, any result regarding the connectivity or the diameter of random graphs that holds with high probability directly translates to our framework with the same probabilistic guarantee.

4.2 Below the connectivity threshold

We now consider the case where $\mathcal{G}_{n,\widehat{p}}$ is not likely to be connected, that is $\widehat{p} \leq c\,\frac{\log n}{n}$ with $c < 1$. (In particular, p and $\widehat{p}$ both go to zero when n goes to infinity). Let us start by establishing a condition for a k-active flooding to complete almost surely.

LEMMA 2. *For any birth-rate p and death-rate q such that $p \to 0$ and $\widehat{p} \to 0$ when $n \to \infty$, the reachability threshold from any $s \in [n]$ in $\mathcal{M}_{n,p,q}$ is, a.s., at least $\Omega(\frac{\log n - n\widehat{p}}{np})$.*

PROOF. We compute a lower bound on the weight $w(e)$ of an edge in $\mathcal{G}^{(k)}_{n,\widehat{p},p}$ for flooding to complete. By definition of $\mathcal{G}^{(k)}_{n,\widehat{p},p}$, we have $\Pr\{w(e) \leq k\} = 1-(1-\widehat{p})(1-p)^{k-1}$. For flooding to complete a.s, the Reduction Lemma states that this probability must be greater than $\frac{\log n}{n}$ because otherwise the random graph induced by all the edges with weight at most k would, a.s., be not connected. Therefore, $k \geq 1 + \frac{\log(1-\frac{\log n}{n})-\log(1-\widehat{p})}{\log(1-p)} = 1 + \frac{\log n - n\widehat{p}}{np}(1+o(1))$. □

THEOREM 2. *Let us consider any birth-rate p and death-rate q such that $\frac{1}{n} \ll \widehat{p} \leq \frac{c\log n}{n}$ with $c < 1$. Then the reachability threshold from any $s \in [n]$ in $\mathcal{M}_{n,p,q}$ is, a.s., equal to $\Theta(\frac{\log n}{np})$. Moreover the optimal time of flooding from any $s \in [n]$ in $\mathcal{M}_{n,p,q}$ is, a.s., equal to*

$$\Theta(\frac{\log n}{np} + \frac{\log n}{\log n\widehat{p}}),$$

and the k-active flooding protocol from any $s \in [n]$ in $\mathcal{M}_{n,p,q}$ with $k = \Omega(\frac{\log n}{np})$ completes a.s. in optimal time.

PROOF. By Lemma 2, and using the fact that the optimal flooding time is at least equal to the reachability threshold, we get that the flooding time is at least $\Omega(\frac{\log n}{np})$ whenever $\widehat{p} \leq \frac{c\log n}{n}$ with $c < 1$. Let us show that, also, the flooding time is at least $\Omega(\frac{\log n}{\log(\widehat{p}n)})$. (This latter result holds for any $\widehat{p} \geq 1/n$, and hence holds in the context of Theorem 1). Recall that I_t denotes the set of nodes that are aware of the message at time step t during the execution of the flooding protocol. We assume that, instead of only one source s, a set I_0 of sources initiates flooding of the same message, with $|I_0| = \beta\log n$ for some $\beta > 1$. Flooding from I_0 performs at least at fast as flooding from a single source. Let $\tilde{p} = \max\{p, \widehat{p}\}$. Since $\widehat{p} \gg \frac{1}{n}$, we have $\tilde{p}n \geq 1$. For $U \subseteq [n]$, let $\Gamma_t(U) = \{e = \{u,v\} \in E(G_t) \mid u \in U, v \notin U\}$. Hence we have

$$\Pr\{|I_{t+1}| > (1+e\tilde{p}n)|I_t|\} \leq \Pr\{|\Gamma_t(I_t)| > e\tilde{p}n|I_t|\}.$$

Now, an edge connecting $u \in I_t$ to $v \notin I_t$ exists in G_t with probability p if $u \in I_{t-1}$, and with probability $\widehat{p}$ if $u \notin I_{t-1}$. Indeed, $u \in I_{t-1}$ and $v \notin I_t$ implies that the edge was not existing in G_{t-1}. And if $u \notin I_{t-1}$, then, as seen in the proof of the Reduction Lemma, the existence of the edge $\{u,v\}$ in G_t is independent from the past, and hence occurs in G_t with probability $\widehat{p}$. All these edges in $\Gamma_t(I_t)$ are mutually independent. Therefore, if $B(a,b)$ denotes a binomial random variable with parameters a and b (i.e, the sum of a independent Bernoulli trials of probability b), we get

$$\Pr\{|\Gamma_t(I_t)| > e\tilde{p}nx \mid |I_t| = x\}$$
$$\leq \Pr\{B((n-x)x, \tilde{p}) > e\tilde{p}nx \mid |I_t| = x\}$$

[4]All logarithms are in the natural base e.

[5]The diameter of a disconnected graph is the maximum diameter (in the usual sense) of its connected components.

for any $x \geq |I_0|$. By Chernoff bound, we derive that

$$\Pr\{B(nx, \tilde{p}) > e\tilde{p}nx \mid |I_t| = x\} \leq e^{-\tilde{p}nx}.$$

Since $\tilde{p} \geq 1/n$ and $x \geq |I_0| \geq \beta \log n$, we get that

$$\Pr\{|I_{t+1}| > (1 + e\tilde{p}n)x \mid |I_t| = x\} \leq n^{-\beta}.$$

Therefore

$$\Pr\{|I_{t+1}| > (1 + e\tilde{p}n)|I_t|\} \leq n^{-\beta}.$$

By union-bound, we derive

$$\Pr\left\{|I_t| > (1 + e\tilde{p}n)^t |I_0|\right\} \leq \sum_{i=0}^{t-1} \Pr\{|I_{i+1}| > (1 + e\tilde{p}n)|I_i|\} \leq t/n^\beta.$$

For $t < \frac{\log n - \log(\beta \log n)}{\log(1+e\tilde{p}n)}$, we have $(1+e\tilde{p}n)^t|I_0| > n$, and thus

$$\Pr\{|I_{t+1}| > n\} \leq \frac{\log n - \log|I_0|}{\log(1 + e\tilde{p}n)} n^{-\beta} \leq \frac{\log n}{\log(1 + e)} n^{-\beta} = o(1/n).$$

From this we conclude that whenever $\widehat{p} \geq 1/n$, the flooding time is, w.h.p., at least $\frac{\log n - \log|I_0|}{\log(1+e\tilde{p}n)}$, that is at least $\Omega(\frac{\log n}{\log(\widehat{p}n)})$. Thus, independently from p, if $\frac{1}{n} \ll \widehat{p} \leq \frac{c \log n}{n}$ with $c < 1$, then, a.s.,

$$T_s \geq \Omega\left(\frac{\log n}{np} + \frac{\log n}{\log(n\widehat{p})}\right).$$

To compute an upper bound on the flooding time, we apply the Reduction Lemma, i.e., we focus on $\mathrm{diam}(\mathcal{G}^{(k)}_{n,\widehat{p},p})$. Since $\widehat{p} \geq c/n$ for some $c > 1$, $\mathcal{G}_{n,\widehat{p}}$ a.s. contains a giant component of size at least $g(c)n$, where $g(c) \in (0, 1)$ is the unique positive root of $1 - x = e^{-cx}$ (see, e.g., [1] or [21]). Moreover, the diameter of this giant component is, a.s., at most $O(\frac{\log n}{n\widehat{p}} + \frac{\log n}{\log(n\widehat{p})})$ (see [7]). Hence, the subgraph of $\mathcal{G}^{(k)}_{n,\widehat{p},p}$ induced by all edges of weight 1 contains, a.s, a giant component C of size at least $g(c)n$ and diameter at most $O(\frac{\log n}{n\widehat{p}} + \frac{\log n}{\log(n\widehat{p})})$. For each node $u \notin C$, we now estimate the weight of an edge connecting this node to C. Let $\mathcal{E}_u$ be the event "node $u \notin C$ has no edge to C in $\mathcal{G}^{(k)}_{n,\widehat{p},p}$". We have

$$\Pr\{\mathcal{E}_u\} \leq (1 - p)^{(k-1)\cdot g(c)\cdot n}.$$

Let $\mathcal{E} = \cup_u \mathcal{E}_u$. By union bound,

$$\Pr\{\mathcal{E}\} \leq (1 - g(c))n(1 - p)^{(k-1)\cdot g(c)\cdot n}.$$

Let us consider the following activity:

$$k = 1 + \frac{2 \log n}{g(c)np}.$$

With this setting of the activity, we get $\Pr\{\mathcal{E}\} \leq 1/n$, and thus the diameter of $\mathcal{G}^{(k)}_{n,\widehat{p},p}$ is, a.s., at most $O(\frac{\log n}{np} + \frac{\log n}{\log(n\widehat{p})} + \frac{\log n}{n\widehat{p}})$. Since $\widehat{p} \geq p/2$, we get that, a.s.,

$$T^{(k)}_s \leq O\left(\frac{\log n}{np} + \frac{\log n}{\log(n\widehat{p})}\right).$$

We complete the proof by noticing that Lemma 2 states that an activity of $1 + \frac{2 \log n}{g(c)np}$ is of the same order of magnitude as the reachability threshold. □

THEOREM 3. *Let us consider any birth-rate p and death-rate q such that $0 < \widehat{p} \leq \frac{c}{n}$ for some constant $c > 0$. Then, a.s., both the reachability threshold and the optimal flooding time from any $s \in [n]$ in $\mathcal{M}_{n,p,q}$ are equal to $\Theta(\frac{\log n}{np})$.*

PROOF. Two cases are considered depending on whether $\mathcal{G}_{n,\widehat{p}}$ is likely to have a giant component or not. Let us first assume that $\widehat{p} = c/n$ for some constant $c > 1$, that is, a.s, there is a giant component in $\mathcal{G}_{n,\widehat{p}}$. Then, by the same arguments as in the proof of Theorem 2, we get that, a.s., $T_s \geq \Omega(\frac{\log n}{np} + \frac{\log n}{\log(n\widehat{p})})$ and $T^{(k)}_s \leq O(\frac{\log n}{np} + \frac{\log n}{\log(n\widehat{p})})$ for $k = 1 + \frac{2 \log n}{g(c)np}$.

Assume now that $\widehat{p} < 1/n$, i.e., $\mathcal{G}_{n,\widehat{p}}$ is not likely to have a giant component. We can reduce this case to the previous one by performing an appropriate number of flooding steps in order to increase the probability of each edge to exist. More precisely, let $d > 1$ be an arbitrarily constant, and let $\gamma = \frac{d}{n}$. Let us compute the minimum value k such that, for any edge e, $\Pr\{w(e) \leq k\} \geq \gamma$. We have

$$\Pr\{w(e) \leq k\} = 1 - (1 - \widehat{p})(1 - p)^{k-1}$$

and hence

$$k \geq 1 + \frac{\log(1 - \gamma) - \log(1 - \widehat{p})}{\log(1 - p)} = 1 + \frac{d - n\widehat{p}}{np}(1 + o(1)) = \Theta\left(\frac{1}{np}\right).$$

For such a $k = \Theta(1/np)$, we insure that $\mathcal{G}^{(k)}_{n,\widehat{p},p}$ has a giant component of size at least $g(d)n$ and diameter bounded by $O(k\frac{\log n}{\log n\varrho}) \leq O(k \log n)$. That is, after $O(\frac{\log n}{np})$ time steps, we can apply the result of the previous case, and, after additional $\frac{\log n}{np}$ steps, flooding is complete. The lower bound $\Omega(\frac{\log n}{np})$ is given by Lemma 2. □

5. ACKNOWLEDGEMENTS

All the authors are thankful to Andrea Clementi, Francesco Pasquale, and Riccardo Silvestri, for many helpful comments on previous versions of this paper.

6. REFERENCES

[1] N. Alon, J. H Spencer. The Probabilistic Method. *Wiley* (2000).

[2] C. Avin, M. Koucký, and Z. Lotker. How to Explore a Fast-Changing World. In 35th International Colloquium on Automata, Languages and Programming (ICALP), pages 121–132, 2008.

[3] B. Bollobás. Random Graphs. *Cambridge University Press* (2001).

[4] F. Brauer, P. van den Driessche, and J. Wu (Eds). Mathematical Epidemiology. Lecture Notes in Mathematics, subseries in Mathematical Biosciences Subseries, Vol. 1945, 2008.

[5] D. Chakrabarti, J. Leskovec, C. Faloutsos, S. Madden, C. Guestrin, and M. Faloutsos. Information survival threshold in sensor and P2P networks. In 26th IEEE International Conference on Computer Communications (INFOCOM), pages 1316–1324, 2007.

[6] N. Chang and M. Liu. Controlled flooding search in a large network. IEEE/ACM Transactions on Networking 15(2):436-449, 2007.

[7] F. Chung, and L. Lu. The Diameter of Sparse Random Graphs. *Advances in Applied Math.* 26:257-279, 2001.

[8] A. Clementi, C. Macci, A. Monti, F. Pasquale, and R. Silvestri. Flooding Time in Edge-Markovian Dynamic Graphs. In 27th ACM Symp. on Principles of Distributed Computing (PODC), pages 213–222, 2008.

[9] A. Clementi, A. Monti, F. Pasquale, and R. Silvestri. Communication in Dynamic Radio Networks. In 26th ACM Symp. on Principles of Distributed Computing (PODC), pages 205–214, 2007.

[10] A. Clementi, A. Monti, F. Pasquale, and R. Silvestri. Information Spreading in Stationary Markovian Evolving Graphs. In 23rd IEEE International Parallel and Distributed Processing Symposium (IPDPS), 2009.

[11] A. Demers, D. Greene, C. Hauser, W. Irish, J. Larson, S. Shenker, H. Sturgis, D. Swinehart, and D. Terry. Epidemic Algorithms for Replicated Database Maintenance. In 6th ACM Symposium on Principles of Distributed Computing (PODC), pages 1–12, 1987.

[12] M. Draief, A. Ganesh, and L. Massoulié. Thresholds for virus spread on networks. Ann. Appl. Probab. 18(2):359-378, 2008.

[13] U. Feige, D. Peleg, P. Raghavan, and E. Upfal. Randomized broadcast in networks. In Int. Symposium on Algorithms (SIGAL), LNCS 450, pages 128–137, Springer, 1990.

[14] P. Fraigniaud and E. Lazard. Methods and Problems of Communication in Usual Networks. Discrete Applied Mathematics 53:79–133, 1994.

[15] A. Frieze and G. Grimmett. The shortest-path problem for graphs with random arc-lengths. Discrete Applied Math. 10:57-77, 1985.

[16] Gnutella RFC. `http://rfc-gnutella.sourceforge.net/`

[17] M. Grossglauser and D. Tse. Mobility increases the capacity of ad-hoc wireless networks. IEEE/ACM Transactions on Networking 10(4):477-486, 2002.

[18] I. Gupta, A.-M. Kermarrec, and A. Ganesh. Efficient Epidemic-Style Protocols for Reliable and Scalable Multicast. In 21st Symposium on Reliable Distributed Systems (SRDS), pages 180–189, 2002.

[19] S. Hedetniemi, S. Hedetniemi, and A. Liestman. A survey of gossiping and broadcasting in communication networks. Networks 18:319–349, 1986.

[20] J. Hromkovič, R. Klasing, B. Monien, and R. Peine. Dissemination of information in interconnection networks (broadcasting and gossiping). Combinatorial Network Theory, pages 125–212. Kluwer Academic, D.-Z. Du and D. Hsu (eds), 1995.

[21] S. Janson , T. Luczak, and A. Rucinski. Random Graphs. *Wiley* (2000).

[22] D. Kempe, A. Dobra, and J. Gehrke. Computing Aggregate Information using Gossip. In 44th Annual IEEE Symposium on Foundations of Computer Science (FOCS), pages 482–491, 2003.

[23] D. Kempe, J. Kleinberg, and A. Demers. Spatial gossip and resource location protocols. In 33rd ACM Symposium on Theory of Computing (STOC), pages 163–172, 2001.

[24] D. Kempe and J. Kleinberg. Protocols and impossibility results for gossip-based communication mechanisms. In 43st IEEE Symp. on Foundations of Computer Science (FOCS), pages 471–480, 2002.

[25] J. Luo, P. Eugster, and J.-P. Hubaux. Route driven gossip: probabilistic reliable multicast in ad hoc networks. In 22nd Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM), pages 2229–2239, 2003.

[26] Q. Lv, P. Cao, E. Cohen, K. Li, and S. Shenker. Search and replication in unstructured peer-to-peer networks. In 16th ACM Int. Conference on Supercomputing (ICS), pages 84–95, 2002.

[27] B. Pittel. On spreading a rumour. SIAM J. Applied Math., 47:213-223, 1987.

[28] A. Sarwate and A. Dimakis. The Impact of Mobility on Gossip Algorithms. In 28th Conference on Computer Communications (INFOCOM), 2009.

[29] K. Sripanidkulchai, B. Maggs, and H. Zhang. Efficient content location using interest-based locality in peer-to-peer systems. In 22nd Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM), pages 2166–2176, 2003.