Expansion and Flooding in Dynamic Random Networks with Node Churn*

Luca Becchetti
Sapienza Università di Roma
Rome, Italy
becchetti@dis.uniroma1.it

Andrea Clementi
Università di Roma Tor Vergata
Rome, Italy
clementi@mat.uniroma2.it

Francesco Pasquale
Università di Roma Tor Vergata
Rome, Italy
pasquale@mat.uniroma2.it

Luca Trevisan
Università Bocconi
Milan, Italy
l.trevisan@unibocconi.it

Isabella Ziccardi
Università dell'Aquila
L'Aquila, Italy
isabella.ziccardi@graduate.univaq.it

Abstract

We study expansion and flooding in evolving graphs, when nodes and edges are continuously created and removed. We consider a model with Poisson node inter-arrival and exponential node survival times. Upon joining the network, a node connects to $d = \mathcal{O}(1)$ random nodes, while an edge disappears whenever one of its endpoints leaves the network. For this model, we show that, although the graph has $\Omega_d(n)$ isolated nodes with large, constant probability, flooding still informs a fraction $1 - \exp(-\Omega(d))$ of the nodes in time $\mathcal{O}(\log n)$. Moreover, at any given time, the graph exhibits a "large-set expansion" property. We further consider a model in which each edge leaving the network is replaced by a fresh, random one. In this second case, we prove that flooding informs all nodes in time $\mathcal{O}(\log n)$, with high probability. Moreover, the graph is a vertex expander with high probability.

Keywords: Dynamic Networks; Random Evolving Graphs; Node Churn; Vertex Expansion; Flooding.

*An extended abstract of this paper has been presented at the *IEEE 41st International Conference on Distributed Computing Systems (ICDCS)* 2021. LT's work on this project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 834861). LB's work on this project was Partially supported by the ERC Advanced Grant 788893 AMDROMA "Algorithmic and Mechanism Design Research in Online Markets", the EC H2020RIA project "SoBigData++" (871042), and the MIUR PRIN project ALGADIMAR "Algorithms, Games, and Digital Markets".

1 Introduction

In this paper, we investigate *information diffusion* in *dynamic networks*, with a focus on *flooding*, the basic mechanism whereby each informed node in turn relays the information received to each of its neighbors. Flooding represents the fastest protocol for *broadcast*, a fundamental communication primitive in distributed systems.

We use the term *dynamic network* to denote communication networks that change over time, as nodes enter or leave the system and links between nodes are created or destroyed. Several important cases of information diffusion occur in networks that evolve over time, such as social or peer-to-peer networks.

Information diffusion in dynamic networks has been the focus of extensive previous work, surveyed in Section 2. We are interested in models that i) exhibit *node churn* (that is, in which nodes enter and leave the network over time) and ii) in which edges are created randomly, rather than according to sophisticated distributed algorithms. Our motivation is that a satisfactory modeling of network formation in social networks and peer-to-peer networks will have to meet both requirements. To the best of our knowledge, information diffusion in dynamic networks with node churn and with random, uniform edge creation was not studied before.

We consider two models, both of which generate *sparse* networks, meaning that these networks contain $\mathcal{O}(n)$ edges, if n denotes the number of nodes. In both models, when a node is "born," i.e., it enters the network, the node connects to $d = \mathcal{O}(1)$ nodes, chosen at random among those currently in the network, while the two models differ in the way nodes react to the failure of incident edges resulting from neighbours leaving the network. We show that, as simple as they are, these dynamic random graphs present interesting expansion properties and that flooding informs all or most nodes (depending on details of the model) in $\mathcal{O}(\log n)$ time.

We kept our modeling choices as simple as possible, defining models that are described by few parameters, in order to identify qualitative features that we believe might prove robust across different scenarios. If, as a result, our models may be too simplistic to reflect all properties of real networks, one of the models we consider (the Poisson model with edge regeneration defined below) bears a certain resemblance to the way peer-to-peer networks such as bitcoin are self-organize and evolve [24, 19].

1.1 Modeling networks that change with time

To define a dynamic network model in the framework outlined above, we have to specify how nodes enter and exit the network, and how edges are generated and destroyed.

Modeling node churn. In this paper, we study a continuous-time model introduced in [22], in which the number of births within each time unit follows a Poisson distribution with mean λ, and where the lifetime of each node is independently distributed as an exponential distribution with parameter μ, so that the average lifetime of a node is $1/\mu$ and the average number of nodes in the network at any given time is λ/μ. In order to reduce the number of parameters, we assume that the time it takes to send a message along an edge is the same, or the same order as, the typical time between node births, which is λ. We choose time units, so that $\lambda = 1$, and we let $n = 1/\mu$.

Modeling edge creation and destruction. When a node enters the network, we assume that it connects to $d = \mathcal{O}(1)$ nodes chosen uniformly at random among those currently in the network. Once an edge (u, v) is created, it remains active as long as both u and v are alive. We study two models: one *without edge regeneration* and one *with edge regeneration*. In the former, edges are created only when a new node joins the network, while in the latter a node creates its outgoing edges not only when it joins the network, but also every time it loses an outgoing edge due to one of its neighbors leaving the network, so that its out-degree is always equal to d.

Although the assumption that a node can pick its neighbors uniformly at random among all nodes currently in the network is unrealistic in many settings, the edge creation and regeneration processes that describe our models resemble the way in which some unstructured peer-to-peer networks maintain a "random" topology. For example, each *full-node* of the Bitcoin network running the Bitcoin Core implementation has a "target out-degree value" and a "maximum in-degree value" (respectively 8 and 125, in the default configuration) and it locally stores a large list of (ip addresses of) "active" nodes. Such list is initially started with nodes received in response to queries to some DNS seeds. Whenever the number of its current neighbors falls below the configured target value, a full-node tries to establish new connections with nodes sampled from its list. The list stored by a full-node is periodically advertised to its neighbors and updated with the lists advertised by its neighbors. Hence, in the long run each full-node samples its out-neighbors from a list consisting a "sufficiently random" subset of all nodes in the network.

1.2 Results and techniques

Informing most nodes in the models without edge regeneration. At first, a negative result for the model without edge regeneration is Lemma 4.2, stating that, with high probability, $\Omega_d(n)$ vertices of the network are isolated at any given time. A vertex v becomes isolated if all the d edges created at birth were to nodes that have meanwhile died, and v was never been chosen as neighbor by younger nodes. Because of the presence of such isolated nodes, we can show (Theorem 4.4) that broadcasting a message to all nodes is not possible, or at least it takes at least $\Omega_d(n \log n)$ time. Furthermore, there is a constant probability that a broadcast dies out after reaching only $\mathcal{O}(1)$ vertices.

At the same time however, we prove Theorem 4.5, stating that a large constant probability exists (tending to 1 as $d \to \infty$ as $1 - \exp(-\Omega(d))$) that a broadcast (starting at least $\Omega(n)$ round after the creation of the network) will reach, say, 90% of the nodes (in general, a constant fraction that tends to 1 as $d \to \infty$ as $1 - \exp(-\Omega(d))$) after $\mathcal{O}(\log n)$ time.

To prove this fast convergence we establish two results. The first one (Lemma 4.6) states that, within $\mathcal{O}(\log n / \log d)$ time, a broadcast reaches $\Omega(n/d)$ nodes. To prove this, we argue that, while the number of informed nodes is less than $\mathcal{O}(n/d)$, there is a good probability that the number of informed nodes grows by a constant factor at each round (and the probability that the above condition fails after exactly t rounds decreases exponentially with t, so that we can take a union bound over all t). The basic idea of this proof is to apply the principle of deferred decision to the d edges chosen by each vertex, and assume that those edges are chosen after the vertex is informed, so that the "frontier" of newly informed vertices keeps growing. There are two difficulties with this approach. One is that older nodes are likely to have chosen neighbors that have meanwhile died, and so older nodes are unlikely to significantly contribute to the number of nodes that will be newly informed at the next round. The second difficulty is that a node may become informed by a message coming from one of the d neighbors chosen at birth, so that we cannot really apply deferred decision in the way that we would like.

To overcome these difficulties, we only consider nodes that are informed through special kinds of paths from the source node (this will undercount the number of informed nodes and make our result true for a stronger reason). Specifically, we define an "onion-skin" process that only considers paths that alternate between "young" nodes whose age is less than the median age and "old" nodes whose age is more than the median age. Furthermore, this process arbitrarily splits the d edges chosen by each node at birth into $d/2$ "type-A" edges and $d/2$ "type-B" edges, and only considers paths that, besides alternating between young nodes and old nodes, also alternate between type-A edges and type-B edges. With this restrictions and conventions in place, we can study what happens for every pair of consecutive rounds by applying deferred decision.

As sketched above, we are thus able to show that $\Omega(n/d)$ nodes are informed within $\mathcal{O}(\log n / \log d)$ rounds. To complete the argument, we show (Lemma 4.3) that, if d is a sufficiently large constant, all sets of at least $n/10$ vertices have constant vertex expansion[1], which leads to informing at least $.9n$ nodes after another $\mathcal{O}(1)$ rounds. Above, $1/10$ can be replaced by $\exp(-\Omega(d))$. This tradeoff is best possible because, as we argued above, there likely are $\Omega_d(n)$ isolated vertices that we will not be able to inform.

Informing all nodes in the models with edge regeneration. A first positive result for the model with edge regeneration is Theorem 5.3, stating that at each time, the graph has constant vertex expansion with high probability. This in turn implies (Theorem 5.7) that, despite the presence of node churn, broadcast reaches all nodes within $\mathcal{O}(\log n)$ rounds.

A standard approach to prove vertex expansion in random graphs works as follows: we bound the probability that a fixed set of k vertices fails to have constant vertex expansion, then we take a union bound by multiplying by $\binom{n}{k}$ and then by summing over k. A first difficulty in our dynamic setting is in characterizing the probability that an edge exists between a pair of vertices u, v, because such probability is a non-trivial function of the age of u and v. The main difficulty is that, in order to compute the probability that an edge (u, v) exists, we need to know the age of u and v and so we have to take a union bound over all subsets of vertices of all possible ages. But, at any given time, there are

[1]Informally, the vertex expansion of a subset S of the vertices of a graph with vertex set V is the average number of distinct vertices in $V \setminus S$ that each vertex in S is connected to. It is a measure of how well S is connected to the rest of the graph. This notion is formally given in Definition 3.1.

		Poisson dynamic graphs	
		without Edge Regeneration	**with Edge Regeneration**
Expansion properties	Negative Results	There is a constant fraction of isolated nodes w.h.p. (Lemma 4.2)	—
	Positive Results	$\Theta(1)$-Expansion of big-size node subsets w.h.p. (Lemma 4.3)	$\Theta(1)$-Expansion w.h.p. (Theorem 5.3)
Flooding	Negative Results	Flooding may not complete, with probability $\Theta(1)$ (Theorem 4.4)	—
	Positive Results	Flooding informs a fraction $1 - e^{-d/20}$ of the nodes in $\mathcal{O}(\log n)$ time, with probability $1 - 2e^{-d/576}$ (Theorem 4.5)	Flooding time is $\mathcal{O}(\log n)$ w.h.p. (Theorem 5.7)

Table 1: Summary of our results.

nodes of age up to $n \log n$, and so we end up with $\binom{n \log n}{k}$ cases in our union bound for sets of size k, while the probability that one such set is non-expanding is as high as $1/\binom{n}{k}^{\mathcal{O}(1)}$ for sets that contain mostly young vertices. The point is that most of the $\binom{n \log n}{k}$ possible ways of choosing k nodes of all possible ages involve choices of several old nodes, which are unlikely to have all survived. In order to carefully account for the "demographics" of all possible sets of edges in our union bound, we look at the logarithm of the probability that a certain set fails to expand, interpret it as the KL divergence of two appropriately defined distributions, and then use inequalities about KL divergence.

We finally remark that some technical care is required when applying the above expansion properties to bound flooding time and thus prove Theorem 5.7, since the analysis requires handling the presence of a random number of node insertion/deletions during every 1-hop message transmission.

Table 1 summarizes our positive and negative results for the models we consider, providing references to the corresponding theorems and lemmas that are proven in the sections that follow.

Roadmap. The remainder of this paper is organized as follows. We discuss related and previous work in more detail in Section 2. In Section 3, we define the Poisson models of dynamic graphs, we state our results for such models and provide their full proofs. In Section 4, we analyse the Poisson model without edge regeneration, while Section 5 is devoted to the analysis of the model version with edge regeneration. Section 6 provides some further overall remarks about our contribution and poses an open question. Finally, some mathematical tools and technical lemmas we used in the analysis are given in the Appendix.

2 Related work

A first, rough classification of dynamic graphs can be made according to an important feature: whether or not the set of nodes keeps the same along all the graph process. In the affirmative case, we have an *edge-dynamic graph* $\{G_t = (V, E_t),\ t \geqslant 0\}$ where the topology dynamics defines the way the edges of a fixed set V of participant nodes change over time. For this class of dynamic graphs, several models, such as worst-case adversarial changes [13, 14, 18] and Markovian evolving graphs [5, 6], have been introduced, their basic connectivity properties have been derived, and, fundamental distributed tasks, such as broadcast and consensus, have been rigorously analyzed.

In contrast, much less analytical works are currently available when (even) the set of participant nodes can change over time. This class of dynamic graphs $\{G_t = (V_t, E_t),\ t \geqslant 0\}$ are often called *dynamic networks with churn* [1]: in this framework, the specific graph dynamics describe both the node insertion/deletion rule for the time sequence V_t and the edge updating rule for the time sequence E_t. The number of nodes that can join or leave the network at every round is called *churn rate*. For brevity's sake, in what follows we will only describe those previous analytical results on dynamic networks with churn which are related to the models we studied in this paper. In particular, we mainly focus on previous work where some connectivity properties of a dynamic networks with churn have been rigorously proved.

As remarked in the introduction, to the best of our knowledge, previous analytical studies focus on distributed algorithms that are suitably designed to maintain topologies having good connectivity properties.

Pandurangan et al. [22] introduced a partially-distributed protocol that constructs and maintains a bounded-degree graph which relies on a centralized cache of a constant number of nodes. In more detail, their protocol ensures the network is connected, has logarithmic diameter, and has always bounded degree. The protocol manages a central cache which maintains a subset of the current set of vertices. When joining the network, a new node chooses a constant number of nodes in the cache. The insertion/deletion procedures for the central cache follows rather complex rules which take $\mathcal{O}(\log n)$ overhead and delays, w.h.p.

In [11], Duchon et al presented ad-hoc protocols that maintain a given distribution of random graphs under an arbitrary sequence of vertex insertions and deletions. More in detail, given that the graph G_t is random uniform over the set of k-out-degree graphs with n nodes, they provide suitable distributed randomized protocols that can insert (respectively delete) a node such that the graph G_{t+1} at round t is again random uniform over the set of k-out-degree graphs with $n+1$ (respectively, $n-1$) nodes. They do not assume a centralized knowledge of the whole graph but, instead, their protocol relies on some random primitives to sample arbitrary-sized subsets of nodes uniformly at random. For instance, once a new node u is inserted, a random subset of nodes is selected (thanks to one of such centralized primitives), and each of them is forced to delete one of its link and to deterministically connect to u. The basic versions of their insertion/deletion procedures require each node to communicate with nodes at distance 2, while their more refined version (achieving optimal performance) requires communications over longer paths.

An important and effective approach to keep a dynamic graph with churn having good expansion properties is based on the use of ID random walks. Roughly speaking, this approach let every participating node start k independent random walks of tokens containing its ID and all the other nodes collaborate to perform such random walks for enough time so that the token is well-mixed over the network. Once a token is mature, it can be used by any node that, in that step, needs a new edge by simply asking to connect to it. The probabilistic analysis then typically shows two main, correlated invariants: on one hand, the edge set, arising from the above random-walk process, form a random graph having good expansion properties. On the other hand, after a small number of steps, the random walks are well-mixed.

Cooper et al [7] consider two deterministic churn processes: in the first one, at every round a new node is inserted while no nodes leave the network, while, in the second process, the size n of the graph does never change since, at every round, a new node is inserted and the oldest node leaves the graph (this is in fact the streaming model we study in this paper). They provide a protocol where each node v starts $c \cdot m$ independent random walks (containing the ID-label of v) until they are picked up, m at a time, by new nodes joining the network. The new node connects to the m peers that contributed the tokens it got. The resultant dynamic topology is shown to keep diameter $\mathcal{O}(\log n)$, and to be fault-tolerant against adversarial deletion of both edges and vertices. We remark that the tokens in the graph must be constantly circulated in order to ensure that they are well-mixed. Moreover, the rate at which new nodes can join the system is limited, as they must wait while the existing tokens mix before they can use them.

Law and Siu [15] provide a distributed algorithm for maintaining a regular expander in the presence of limited number of insertions/deletions. The algorithm is based on a complex procedure that is able to sample uniformly at random from the space of all possible $2d$-regular graphs formed by d Hamiltonian circuits over the current set of alive nodes. They present possible distributed implementations of this sample procedure, the best of which, based on random walks, have $\mathcal{O}(\log n)$ overhead and time delay. Such solutions cannot manage frequent node churn.

Further distributed algorithms with different approaches achieving $\mathcal{O}(\log n)$ overhead and time delay in the case of slow node churn are proposed in [4, 12, 16, 23].

In [2], Augustine et al present an efficient randomized distributed protocol that guarantees the maintenance of a bounded degree topology that, with high probability, contains an expander subgraph whose set of vertices has size $n - o(n)$, where n is the stable network size. This property is preserved despite the presence of a large oblivious adversarial churn rate — up to $\mathcal{O}(n/\text{polylog}(n))$. In more detail, it is preserved under an oblivious churn adversary [3] that: can remove any set of nodes up to the churn limit in every round, and, at the same time, it should add (an equal amount of) nodes to the network with the following constraints. A new node should be connected to at least one existing node and the number of new nodes added to an existing node should not exceed a fixed constant (thus, all nodes have constant bounded degree). The expander maintenance protocol is efficient even though it is rather complex and the local overhead for maintaining the topology is polylogarithmic in n. A complication of the protocol follows from the fact that, in order to prevent the growth of large clusters of nodes outside the expander subgraph, it uses special criteria to "refresh" the links of some nodes, even when the latter have not been involved by any edge deletion due to the node churn.

Recently, the flooding process has been analytically studied over dynamic graph models with churn in [3, 1]. Here, the authors consider the model analysed in [2], that we discussed above. Using the expansion property proved in [2], they show that, for any fixed churn rate $C(n) \leqslant n/\text{polylog}n$ managed by an oblivious worst-case adversary, there is a set S of size $n - \mathcal{O}(C(n))$ of nodes such that, if a source node in S starts the flooding in round t, then all except $\mathcal{O}(C(n))$ nodes get informed within round $t + \mathcal{O}(\log(n/C(n)) \log n)$, w.h.p.

Our models are inspired by the way some unstructured P2P networks maintain a "well-connected" topology, despite nodes joining and leaving the network, small average degree and almost fully decentralized network formation. For example, after an initial bootstrap in which they rely on DNS seeds for node discovery, full-nodes of the Bitcoin network [19] running the Bitcoin Core implementation turn to a fully-decentralized policy to regenerate their neighbors when their degree drops below the configured threshold [8]. This allows them to pick new neighbors essentially at random among all nodes of the network [24]. Notice also that the real topology of the Bitcoin network is hidden by the network formation protocol and discovering the real network structure has been recently an active subject of investigations [9, 20].

3 Definitions and preliminaries

A *dynamic graph* $\mathcal{G}$ is a family of graphs $\mathcal{G} = \{G_t = (V_t, E_t) : t \in \mathbb{T}\}$, where $\mathbb{T}$ is a totally-ordered set (in this paper we consider *continuous* dynamic graphs, i.e., with $\mathbb{T} = \mathbb{R}^+$). If $\{V_t\}_t$ and $\{E_t\}_t$ are families of random sets we call the corresponding random process a *dynamic random graph.* We call G_t the *snapshot* of the dynamic graph at time t. For a set of nodes $S \subseteq V_t$, we denote with $\partial^t_{out}(S)$ the outer boundary of S in snapshot G_t; we omit superscript t when it is clear from the context. For any set S, we denote with $|S|$ its size.

In this paper, we study the *vertex expansion* of the snapshots (see Definition 3.1) and the *flooding time* (see Definitions 3.2 and 3.3) for two continuous-time dynamic graph models in which nodes' arrivals follow a Poisson process and their lifetimes obey an exponential distribution (see Definitions 4.1 and 5.1).

The *vertex expansion* of a snapshot of a dynamic graph is the vertex expansion of a static graph. We recall here the definition.

Definition 3.1 (Vertex expansion). *The* vertex isoperimetric number $h_{out}(G)$ *of a graph* $G = (V, E)$ *is*

$$h_{out}(G) = \min_{\substack{S \subset V: \\ 0 \leqslant |S| \leqslant |V|/2}} \frac{|\partial_{out}(S)|}{|S|},$$

where $\partial_{out}(S)$ *is the outer boundary of* S*, i.e.,* $\partial_{out}(S) = \{v \in V \setminus S : \{u, v\} \in E \text{ for some } u \in S\}$*. Given a constant* $\varepsilon > 0$*, a graph* G *is a* (vertex) ε-expander *if* $h_{out}(G) \geqslant \varepsilon$*.*

Flooding is a popular epidemic process in which a *source* node s sends a message M_s to all its neighbors that, in turn, forward M_s to all their neighbors, and so on. The *flooding time* is the time it takes a message to arrive to all reachable nodes. There are different ways in which the flooding process can be formalized in a continuous dynamic graph model, depending on how the time it takes a message to flow from a node to its neighbors relates to the frequency of changes in the topology of the graph. We here choose the following definition of "asynchronous" flooding, in which a message takes one unit of time to flow from an informed node to its neighbors.

Definition 3.2 ("Asynchronous" flooding). *Let* $\mathcal{G} = \{G_t = (V_t, E_t) : t \in \mathbb{R}^+\}$ *be a dynamic (random) graph. The flooding process over* $\mathcal{G}$ *starting at time* t_0 *from vertex* $v_0 \in V_{t_0}$ *is the sequence of (random) sets of nodes* $\{I_t : t \in \mathbb{R}^+\}$ *where,* $I_t = \emptyset$ *for all* $t < t_0$*,* $I_{t_0} = \{v_0\}$ *and,*[2] *for every* $t \geqslant t_0$*,* I_t *contains all nodes in* V_t *that were neighbors of some node in* I_{t-1} *in snapshot* G_{t-1}*, in addition to all previously informed nodes*

$$I_t = \left(\left(\bigcup_{t' < t} I_{t'} \right) \cup \partial^{t-1}_{out}(I_{t-1}) \right) \cap V_t .$$

The subset I_t *is the set of nodes that, at time* t*, are in the* informed *state. We further say that the flooding* completes *the broadcast if a time* t *exists such that* $I_t \supseteq V_t$*, in which case* $t - t_0$ *is the* flooding time *of the source message.*

In order to analyze the flooding process in Definition 3.2, it will be convenient to study a discretized version of the above process, in which nodes are informed only at discrete time steps.

[2]In this paper we thus assume that I_0 contains the node joining the network at round t_0.

Definition 3.3 (Discretized" flooding)**.** *Let $\mathcal{G} = \{G_t = (V_t, E_t) : t \in \mathbb{R}^+\}$ be a continuous dynamic (random) graph. The* discretized *flooding process over $\mathcal{G}$ starting at time $t_0 \in \mathbb{R}^+$ from vertex $v_0 \in V_{t_0}$ is the sequence of (random) sets of nodes $\{I_t : t \in \mathbb{N}\}$ where, $I_t = \emptyset$ for all $t < t_0$, $I_{t_0} = \{v_0\}$ and, for every t of the form $t_0 + m$ with integer m, I_t contains all nodes in I_{t-1} that did not die in the time interval $(t-1, t)$ and all nodes in V_t that were neighbors of some node in I_{t-1} in the whole time interval $(t-1, t)$:*

$$I_t = \left(I_{t-1} \cup \partial_{out}^{t-1}(I_{t-1} \cap V_t)\right) \cap V_t \, .$$

The subset I_t is the set of nodes that, at time t, are in the informed *state. We further say that the flooding* completes *the broadcast if a* round *t exists such that $I_t \supseteq V_{t-1}$, in which case $t - t_0$ is the* flooding time *of the source message.*

Notice that, since in the discretized process informed nodes have to wait until the next discrete time before relaying the message, the flooding time of the discretized process can only be larger than the flooding time of the asynchronous one. We will prove upper bounds on the flooding time of the discretized process, that thus are also upper bounds on the flooding time of the asynchronous one.

3.1 Node churning

We study dynamic graphs in which the set of nodes is governed by the following random process $\{V_t\}_t$.

Definition 3.4 (Poisson node churn [22])**.** *Let $\lambda, \mu \in \mathbb{R}^+$. A Poisson node churn is a random process $\{V_t : t \in \mathbb{R}\}$ such that:*

1. *$V_0 = \emptyset$;*
2. *The arrival of new nodes in V_t is a Poisson process with rate λ;*
3. *Once a node is in V_t, its lifetime has exponential distribution with parameter μ.*

This subsection provides some useful properties of the above random node churn process. First observe that, according to Definition 3.4, the time interval between two consecutive node arrivals is an exponential random variable of parameter λ, while the number of nodes joining the network in a time interval of duration τ is a Poisson random variable with expectation $\tau \cdot \lambda$. Moreover $\{V_t : t \in \mathbb{R}^+\}$ is clearly a continuous Markov process.

A first important fact our analysis relies on is that we can bound the number of active nodes at every time. In particular, it is easy to show that $\mathbf{E}\left[|V_t|\right] \to \lambda/\mu$ and, moreover, the following bound in concentration holds.

Lemma 3.5 (Pandurangan et al. [22] - Number of nodes in the network)**.** *For every pair of parameters λ and μ such that $n = \lambda/\mu$ is sufficiently large, consider the Poisson node churn $\{V_t : t \in \mathbb{R}^+\}$ in Definition 3.4. Then, for every fixed real $t \geqslant 3n$, w.h.p. $|V_t| = \Theta(n)$ and, more precisely,*

$$\mathbf{Pr}\left(0.9n \leqslant |V_t| \leqslant 1.1n\right) \geqslant 1 - 2e^{-\sqrt{n}} \, .$$

Leveraging Lemma 3.5, in our analysis of the Poisson node churn process we set $\lambda = 1$, without loss of generality, and we define the key parameter $n = 1/\mu$, representing the expected number of nodes in the network at each time step, in the long run. Moreover, since the probability that two or more churn events occur at the same time is zero, the points of change of the dynamic graph yield a discrete-time sequence of *events*. Thus, we can restrict ourselves to observe and prove properties of the dynamic graph only when one graph-changing event occurs, i.e., at the arrival of a new node or at the death of an existing one.

Definition 3.6. *Let $\{V_t : t \in \mathbb{R}^+\}$ be a Poisson node churn as in Definition 3.4. We define the infinite sequence of random variables* steps *$\{T_r : r \in \mathbb{N}\}$ (with parameters λ and μ) as follows:*

$$T_0 = 0 \quad \textit{and} \quad T_{r+1} = \inf\{t > T_r : V_t \neq V_{T_r}\} \, , \; \textit{for } r = 0, 1, 2, \ldots .$$

It is worth mentioning that, since $\{V_t : t \in \mathbb{R}^+\}$ is a continuous Markov process, the stochastic process $\{V_{T_r} : r \in \mathbb{N}\}$ defined above is a discrete Markov chain.

The proofs of the next lemmas of this subsection are based on standard probabilistic arguments and, thus, they are given in Appendix C. In particular, the proof of the next lemma is given in Appendix C.3.

Lemma 3.7. *For every sufficiently large n, consider the Markov chain $\{V_{T_r}, r \in \mathbb{N}\}$ in Definition 3.6 with parameters $\lambda = 1$ and $\mu = 1/n$. Then, for every fixed integer $r \geqslant n \log n$,*

$$0.47 \leqslant \mathbf{Pr}\left(|V_{T_{r+1}}| = |V_{T_r}| - 1\right) \leqslant 0.53 \tag{1}$$

and

$$0.47 \leqslant \mathbf{Pr}\left(|V_{T_{r+1}}| = |V_{T_r}| + 1\right) \leqslant 0.53\,. \tag{2}$$

Moreover, for each $v \in N_{T_r}$,

$$\frac{1}{2.2n} \leqslant \mathbf{Pr}\left(v \notin V_{T_{r+1}} \;\middle|\; v \in V_{T_r}\right) \leqslant \frac{1}{1.8n}\,. \tag{3}$$

Lemma 3.8 provides a useful bound on the lifetime of any node in the network. The proof is in Appendix C.4.

Lemma 3.8 (Nodes' lifetimes)**.** *Consider the Markov chain $\{V_{T_r}, r \in \mathbb{N}\}$ in Definition 3.6 with parameters $\lambda = 1$ and $\mu = 1/n$. For every sufficiently large n and for every fixed integer $r \geqslant 7n \log n$, with probability at least $1 - 2/n^{2.1}$, each node in V_{T_r} was born after step $T_{r-7n\log n}$.*

Finally, we give upper bounds on the number of nodes that join and leave the network in a time interval of $\log n$ steps. The proof is in Appendix C.5.

Lemma 3.9. *Let $t_0 = T_{r_0}$ for some $r_0 \geqslant 7n \log n$. For every sufficiently large n, in the interval $[t_0, t_0 + \log n]$:*

1. *With probability at least $1 - 1/n$, at most $4 \log n$ nodes join the network;*
2. *With probability at least $1 - 1/n$, at most $4 \log n$ nodes leave the network.*

4 Graphs without edge regeneration

In this subsection, we consider the first variant of the Poisson model, in which new edges are created only when new nodes join the network.

Definition 4.1 (Poisson dynamic graphs without edge regeneration)**.** *A* Poisson Dynamic Graph without edge regeneration *(for short, PDG) $\mathcal{G}(\lambda, \mu, d)$ is a continuous dynamic random graph $\{G_t = (V_t, E_t) : t \in \mathbb{R}^+\}$ where the set of nodes V_t evolves according to Definition 3.4, while the set E_t of edges is updated according to the following topology dynamics:*

1. *When a new node appears, it makes d independent connection* requests*, each one to a* destination *node chosen uniformly at random among the nodes in the network, and, each of such requests is accepted by the destination so that the edge is activated (i.e. it is included in E_t). The d independent requests will be indexed in the range $\{1, \ldots, d\}$;*
2. *When a node dies, all its incident edges disappear.*

4.1 On the topology of the snapshots

While static d-regular random graphs show good expansion properties even for small, constant values of d (we give a simple proof of this well-known fact in Appendix C.1), the snapshots generated by the Poisson model without edge regeneration exhibit a linear fraction of isolated nodes, even for large values of d.

Lemma 4.2 (Isolated nodes)**.** *For every positive constant d and for every sufficiently large n, let $\{G_t = (V_t, E_t) : t \in \mathbb{R}^+\}$ be a PDG sampled from $\mathcal{G}(\lambda, \mu, d)$ with $\lambda = 1$ and $\mu = 1/n$. For every fixed integer $r \geqslant 7n \log n$, w.h.p., G_{T_r} contains a subset S of nodes such that: i) $|S| \geqslant \frac{1}{30} n e^{-3d}$; ii) each node in S is isolated at T_r and remains isolated over its entire lifetime.*

Proof. Let $r \geqslant 7n \log n$. Define the event

$$L_r = \{\text{each node in } V_{T_r} \text{ is born after time } T_{r-7n\log n}\} \cap \{|V_{T_i}| \in [0.9n, 1.1n] \text{ with } i = r - 7n \log n, \ldots, r\}.$$

Then, $\mathbf{Pr}(L_r) \geqslant 1 - 1/n^2$ follows from Lemma 3.5, Lemma 3.8 and a union bound. Conditioned on the event L_r, when the generic node in V_{T_r} joined the network, the network itself has at least $0.9n$ and at most $1.1n$ nodes. Let ε be an arbitrary value with $1/10 < \varepsilon \leqslant 1/3$: we denote by H the set of the εn oldest nodes in V_{T_r}. Moreover, consider the subset A of all nodes in H that survive at most $2n$ further steps from step T_r onwards, i.e.,

$$A = \{v \in H \text{ s.t. } v \notin V_{T_{r+2n}}\}.$$

Next, using (3) from Lemma 3.7,

$$\mathbf{E}[|A|] = \sum_{v \in H} \mathbf{Pr}\left(v \notin V_{T_{r+2n}}\right) = \sum_{v \in H} (1 - \mathbf{Pr}(v \in T_{r+2n})) \geqslant \varepsilon n \left(1 - \left(1 - \frac{1}{1.8n}\right)^{2n}\right) \geqslant (1 - e^{-1})\varepsilon n,$$

where, in the first inequality, we use that $\mathbf{Pr}(v \in T_{r+2n}) = \mathbf{Pr}(v \in T_{r+1}) \prod_{i=1}^{2n} \mathbf{Pr}(v \in T_{r+i+1} \mid T_{r+i})$ and the fact that $|H| = \varepsilon n$. So, for large enough n, we can apply a standard concentration argument (Theorem A.1) to obtain

$$\mathbf{Pr}\left(|A| \geqslant \frac{\varepsilon n}{4}\right) \geqslant 1 - e^{-\varepsilon n/16} \geqslant 1 - \frac{1}{n}.$$

Now, consider the random variable

$$X = \{\text{number of nodes in } A \text{ that are isolated at time } T_r \text{ and for the rest of their lifetime}\}.$$

We next show that, w.h.p., $X \geqslant \frac{1}{10}\varepsilon n e^{-3d}$. First, to bound the expectation of X, for each $v \in V_{T_r}$, we introduce the following random variables

$$\Delta_v^{in} = \{\text{maximum in-degree of node } v \text{ from } T_r \text{ for the end of its lifetime}\},$$
$$\Delta_v^{out} = \{\text{out-degree of the node } v \text{ at time } T_r\}.$$

Since $\Delta_v^{out} = 0$ implies that node v will have no out-edges from step T_r onwards, we can then write X as a function of Δ_v^{in} and Δ_v^{out} for $v \in A$:

$$X = \sum_{v \in A} \mathbb{1}_{\{\Delta_v^{in}=0\}} \mathbb{1}_{\{\Delta_v^{out}=0\}}.$$

As each node establishes its links independently to the others,

$$\mathbf{E}[X \mid L_r] = \sum_{v \in A} \mathbf{Pr}\left(\Delta_v^{in} = 0 \mid L_r\right) \mathbf{Pr}\left(\Delta_v^{out} = 0 \mid L_r\right). \tag{4}$$

The probability of a node $v \in V_{T_r}$ to have in-degree 0 over its entire lifetime, conditioned on L_r, is

$$\mathbf{Pr}\left(\Delta_v^{in} = 0 \mid L_r\right) = \left(1 - \frac{1}{0.9n}\right)^{d(2n+\varepsilon n)} \geqslant e^{-5d/2},$$

since each node $v \in A$ has lifetime at most $2n$ steps and there are at most εn nodes younger than him in H. The probability of a node $v \in A$ having no out-edges in the current step (conditioned on L_r) is

$$\mathbf{Pr}\left(\Delta_v^{out} = 0 \mid L_r\right) = \left(1 - \frac{\varepsilon n}{0.9n}\right)^{d} \geqslant e^{-d/2},$$

since a node $v \in A$ has no out-edges at time T_r if its requests are not towards the nodes in A (the oldest in V_{T_r}), and the last inequality holds since $\varepsilon \leqslant 1/3$. Therefore, since $|A| \geqslant \varepsilon n/4$ with probability at least $1 - 1/n$, (4) implies

$$\mathbf{E}[X \mid L_r] \geqslant \mathbf{E}\left[X \;\middle|\; L_r, |A| \geqslant \frac{\varepsilon n}{4}\right] \mathbf{Pr}\left(|A| \geqslant \frac{\varepsilon n}{4}\right) \geqslant \frac{\varepsilon n}{4} e^{-3d} \left(1 - \frac{1}{n}\right) \geqslant \frac{\varepsilon n}{5} e^{-3d}.$$

To get concentration results, we now write X as a function of $2n \cdot d$ independent random variables and, then, use the method of bounded differences. Let Y_j^v the random variables that indicates if the j-th request of a node $v \in \cup_{i=0}^{2n} V_{T_{r+i}}$ is towards the set A. By definition, the random variables $\{Y_v^j : v \in V_{T_r} \cup V_{T_{r+1}} \cdots \cup V_{T_{r+2n}}, j \in \{1, \ldots, d\}\}$ are mutually independent. Denote by $\mathbf{Y}$ the vector of these random variables. We can easily express X as a function of $\mathbf{Y}$ as $X = f(\mathbf{Y})$. Notice that if the vectors $\mathbf{Y}$ and $\mathbf{Y}'$ only differ in one coordinate, then $|f(\mathbf{Y}) - f(\mathbf{Y}')| \leqslant 2$. This is because, in the worst case, an isolated node can change its destination from a dead node to another isolated node, so that the number of isolated nodes can only decrease (or increase) by at most 2 units. We can thus apply Theorem A.2 and get

$$\mathbf{Pr}\left(X \leqslant \mu - M \mid L_r\right) \leqslant e^{-\frac{2M^2}{4nd}},$$

where μ is any lower bound to $\mathbf{E}[X \mid L_r]$. Taking $\mu = \frac{1}{5}\varepsilon n e^{-3d}$ and $M = \frac{1}{10}\varepsilon n e^{-3d}$ yields

$$\mathbf{Pr}\left(X \leqslant \frac{1}{10}\varepsilon n e^{-3d} \,\middle|\, L_r\right) \leqslant e^{-n\frac{\varepsilon^2 e^{-6d}}{100d}}. \tag{5}$$

Hence, the number of isolated nodes is $X \geqslant \frac{1}{10}\varepsilon n e^{-3d}$ w.h.p. If n is large enough, (5) and the law of total probability imply

$$\mathbf{Pr}\left(X \leqslant \frac{1}{10}\varepsilon n e^{-3d}\right) \leqslant \mathbf{Pr}\left(X \leqslant \frac{1}{10}\varepsilon n e^{-3d} \,\middle|\, L_r\right) + \frac{1}{n^2} \leqslant e^{-n\frac{\varepsilon^2 e^{-2d}}{72d}} + \frac{1}{n^2} \leqslant \frac{2}{n^2}.$$

The lemma then follows by setting $\varepsilon = 1/3$. □

The lemma that follows highlights weak expansion properties of the Poisson model without edge regeneration. In particular, we show that, for any sufficiently large t, all subsets of V_t including a sufficiently large, constant fraction of the nodes exhibit good expansion properties.

Lemma 4.3 (Expansion of large subsets). *For every constant $d \geqslant 20$ and for every sufficiently large n, let $\{G_t = (V_t, E_t) : t \in \mathbb{R}^+\}$ be a PDG sampled from $\mathcal{G}(\lambda, \mu, d)$ with $\lambda = 1$ and $\mu = 1/n$. Then, for every fixed integer $r \geqslant 7n \log n$, with probability at least $1 - 2/n^2$, the snapshot G_{T_r} satisfies*

$$\min_{ne^{-d/20} \leqslant |S| \leqslant 1.1n/2} \frac{|\partial_{out}(S)|}{|S|} \geqslant 0.1\,.$$

Proof. We prove that, for any two disjoint sets $S, T \subseteq V_{T_r}$, such that $ne^{-d/20} \leqslant |S| \leqslant 1.1n/2$, $|T| = 0.1|S|$, the event $A_{S,T} = \{\partial_{out}(S) \subseteq T\}$ occurs with negligible probability. To this purpose, consider again the event

$$L_r = \{\text{each node in } V_{T_r} \text{ was born after time } T_{r-7n\log n}\} \cap \{|V_{T_i}| \in [0.9n, 1.1n] \text{ with } i = r - 7n\log n, \ldots, r\},$$

and note again that Lemma 3.5 and Lemma 3.8 imply $\mathbf{Pr}(L_r) \geqslant 1 - 1/n^2$. From the law of total probability,

$$\mathbf{Pr}\left(\min_{ne^{-d/20} \leqslant |S| \leqslant 1.1n/2} \frac{|\partial_{out}(S)|}{|S|} \leqslant 0.1\right) \leqslant \sum_{\substack{ne^{-d/20} \leqslant |S| \leqslant 1.1n/2 \\ |T| = 0.1|S|}} \mathbf{Pr}\left(A_{S,T} \mid L_r\right) + \frac{1}{n^2}. \tag{6}$$

To upper bound $\mathbf{Pr}(A_{S,T} \mid L_r)$ let $P = V_{T_r} - S - T$ and note that $|P| \geqslant 0.9n - 1.1|S|$ by definition. The event $A_{S,T}$ implies that all the edges originating from S must have destinations in T: this is equivalent to have no edge between S and P. Since

$$|\{(a,b) \mid a \in S,\ b \in P\}| = |S| \cdot |P|,$$

two cases may arise:

Case 1: $|\{(a,b) \mid a \in S,\ b \in P, a \text{ younger than } b\}| \geqslant |S| \cdot |P|/2$;

Case 2: $|\{(a,b) \mid a \in S,\ b \in P, b \text{ younger than } a\}| \geqslant |S| \cdot |P|/2$.

For each $a \in S$, denote by N_a the number of nodes in P that are older than a. In the first case, we clearly have $\sum_{a \in S} N_a \geqslant |S| \cdot |P|/2$, so that

$$\mathbf{Pr}\left(A_{S,T} \mid L_r\right) \leqslant \prod_{a \in S}\left(1 - \frac{N_a}{1.1n}\right)^d \leqslant e^{-d\sum_{a \in S} N_a/(1.1n)} \leqslant e^{-d|S|\cdot|P|/2.2n}. \tag{7}$$

To derive the first inequality above, for each $a \in S$, we considered the probability that a fixed request from node a has not a destination in P that is older than a. Moreover, we used the fact that, conditioned on the event L_r, the probability that a node a chooses any fixed, older node $v \in V_{T_r}$ at least $1/1.1n$. Using a symmetric argument, we get the same claim for the second case. Hence, plugging (7) into (6) and considering the number of possible ways to choose S and T disjoint and of appropriate sizes, we obtain

$$\mathbf{Pr}\left(\min_{ne^{-d/20}\leqslant|S|\leqslant 1.1n/2} \frac{|\partial_{out}(S)|}{|S|} \leqslant 0.1\right) \leqslant \sum_{s=ne^{-d/20}}^{1.1n/2} \binom{1.1n}{s}\binom{1.1n-s}{0.1s} e^{-ds\frac{0.9n-1.1s}{2.2n}} + \frac{1}{n^2} \leqslant \frac{2}{n^2}\,, \quad (8)$$

where, to prove the last inequality, for a large enough n and for any $d \geqslant 20$, we bound each binomial coefficient via $\binom{n}{k} \leqslant \left(\frac{n\cdot e}{k}\right)^k$ and then, we calculate the derivative of the function $f(s)$ that represents each term of the sum: it turns out that each of such terms reaches its maximum at the extremes, i.e. in $s = 1$ or in $s = 1.1n/2$. So, we get that the sum in (8), if $d \geqslant 20$, is bounded by $2/n^2$.

□

4.2 Flooding

The negative result of Lemma 4.2 implies that flooding has non-negligible chances to fail in rapidly informing the entire network.

Theorem 4.4 (Flooding). *For every positive constant d, for every sufficiently large n and for every fixed $r_0 \geqslant 7n\log n$, the flooding process over a PDG sampled from $\mathcal{G}(\lambda,\mu,d)$ with $\lambda = 1$, $\mu = 1/n$ and starting at $t_0 = T_{r_0}$ satisfies the following claims:*

1. *With probability at least e^{-5d^2}, for every $t \geqslant t_0$, I_t contains at most $d+1$ nodes;*
2. *W.h.p., the flooding time is $\Omega(n)$ rounds.*

Proof. Since the goal here is to prove a negative result, we will adopt the original Definition 3.2 of asynchronous flooding. Let s_0 be the source node, joining the network at time $t_0 = T_{r_0}$. Consider the event

$$C_{r_0}^{r_0+n} = \{|N_{T_i}| \in [0.9n, 1.1n] \text{ with } i = r_0, \ldots, r_0+n\}\,,$$

and notice that Lemma 3.5 implies $\mathbf{Pr}\left(C_{r_0}^{r_0+n}\right) \geqslant 1 - 1/n^2$. We next consider the following events:

$A =$ "s_0 has all its out-edges to nodes that are isolated at time t_0 and for the rest of their lifetimes",

$B =$ "at least $\dfrac{ne^{-3d}}{30}$ nodes in V_{t_0} are isolated at time t_0 and for the rest of their lifetimes".

For sufficiently large n, Lemma 4.2 implies $\mathbf{Pr}(B) \geqslant 0.9$, whence

$$\mathbf{Pr}\left(B \mid C_{r_0}^{r_0+n}\right) = \frac{\mathbf{Pr}\left(B \cap C_{r_0}^{r_0+n}\right)}{\mathbf{Pr}\left(C_{r_0}^{r_0+n}\right)} \geqslant \mathbf{Pr}(B) + \mathbf{Pr}\left(C_{r_0}^{r_0+n}\right) - 1 \geqslant 0.9 - \frac{1}{n^2} \geqslant \frac{1}{2},$$

where in the second inequality we use the simple inequality $\mathbf{Pr}(P \cap Q) \geqslant \mathbf{Pr}(P) + \mathbf{Pr}(Q) - 1$. Then,

$$\begin{aligned}\mathbf{Pr}\left(A \mid C_{r_0}^{r_0+n}\right) &\geqslant \mathbf{Pr}\left(A \cap B \mid C_{r_0}^{r_0+n}\right) = \mathbf{Pr}\left(A \mid B \cap C_{r_0}^{r_0+n}\right)\mathbf{Pr}\left(B \mid C_{r_0}^{r_0+n}\right) \\ &\geqslant \left(\frac{ne^{-3d}}{30}\frac{1}{1.1n}\right)^d \cdot \frac{1}{2} = \left(\frac{e^{-3d}}{3.3}\right)^d,\end{aligned} \quad (9)$$

where the third inequality follows since $C_{r_0}^{r_0+n}$ implies that at time t_0 there are at most $1.1n$ nodes in the network. Define the event

$$E = \text{"}s_0 \text{ is the destination of no in-edges over its lifetime"}\,.$$

Then, the event $A \cap E$ implies that all informed nodes $s_0, s_1, \ldots, s_d$ are isolated along their entires lifetimes, whence

$$\mathbf{Pr}\left(|I_t| \leqslant d+1 \text{ for all } t \geqslant t_0\right) \geqslant \mathbf{Pr}(A \cap E)\,. \quad (10)$$

Denote by D_{s_0} the lifetime (in steps) of node s_0. Since, for (3) from Lemma 3.7, s_0 survives for n steps with probability at most $(1 - 1/(2.2n))^n$, whenever n is large enough,

$$\mathbf{Pr}\left(D_{s_0} \leqslant n \mid C_{r_0}^{r_0+n}\right) = \frac{\mathbf{Pr}\left(\{D_{s_0} \leqslant n\} \cap C_{r_0}^{r_0+n}\right)}{\mathbf{Pr}\left(C_{r_0}^{r_0+n}\right)} \geqslant 1 - \left(1 - \frac{1}{2.2n}\right)^n - 1/n^2 \geqslant \frac{1}{3}\,, \quad (11)$$

where we again used $\mathbf{Pr}(P \cap Q) \geqslant \mathbf{Pr}(P) + \mathbf{Pr}(Q) - 1$. Moreover,

$$\mathbf{Pr}\left(E \mid C_{r_0}^{r_0+n}\right) \geqslant \mathbf{Pr}\left(E \mid C_{r_0}^{r_0+n} \cap D_{s_0} \leqslant n\right) \mathbf{Pr}\left(D_{s_0} \leqslant n \mid C_{r_0}^{r_0+n}\right). \tag{12}$$

Recalling that the destination node of each connection request is chosen uniformly at random over the nodes currently in the network, and since at most n nodes join the network in n steps,

$$\mathbf{Pr}\left(E \mid C_{r_0}^{r_0+n} \cap \{D_{s_0} \leqslant n\}\right) \geqslant \left(1 - \frac{1}{0.9n}\right)^{dn} \geqslant e^{-2d}. \tag{13}$$

Combining (13) and (11) with (12),

$$\mathbf{Pr}\left(E \mid C_{r_0}^{r_0+n}\right) \geqslant \frac{e^{-2d}}{3}. \tag{14}$$

From (10), from the independence between A and E, from (9) and (14),

$$\mathbf{Pr}\left(|I_t| \leqslant d+1 \text{ for all } t \geqslant t_0\right) \geqslant \mathbf{Pr}\left(A \cap E \mid C_{r_0}^{r_0+n}\right) \mathbf{Pr}\left(C_{r_0}^{r_0+n}\right) \geqslant \left(\frac{e^{-3d}}{3.3}\right)^d \cdot \frac{e^{-2d}}{3} \geqslant e^{-5d^2}.$$

As for the second claim of the theorem, the linear lower bound on the flooding time τ can be easily shown as follows. Lemma 4.2 implies the presence of at least $\frac{1}{30}e^{-3d}n = \Omega(n)$ (when d is a constant) isolated nodes at T_{r_0}, that will remain such over the rest of their lifetimes. As a consequence, to see all nodes informed, one has to at least wait for all these nodes to leave the network: this requires $\tau = \Omega(n)$ time, w.h.p.

□

We next complement the negative results above by showing that, following the arrival of an informed node at some time t, a fraction $1 - e^{-\Omega(d)}$ of the vertices of the network will become informed within the following $O(\log n)$ flooding rounds, with probability $1 - e^{-\Omega(d)}$. In the proof of the result that follows, we consider the discretized flooding process from Definition 3.3 for the sake of the analysis.

Theorem 4.5 (Flooding completes for a large fraction of nodes)**.** *For every constant $d \geqslant 112$, for every sufficiently large n and for every fixed $r_0 \geqslant 7n\log n$, there is a $\tau = \mathcal{O}(\log n/\log d + d)$, such that the flooding process over a PDG sampled from $\mathcal{G}(\lambda, \mu, d)$, with $\lambda = 1$, $\mu = 1/n$ and starting at $t_0 = T_{r_0}$, satisfies*

$$\mathbf{Pr}\left(|I_{t_0+\tau}| \geqslant (1 - e^{-\frac{d}{25}})n\right) \geqslant 1 - 3e^{-\frac{d}{56}}.$$

The proof of the above theorem is a simple consequence of the next two lemmas.

Lemma 4.6 (Flooding reaches $\Omega(n)$ nodes, part 1)**.** *Under the hypotheses of Theorem 4.5, there is $\tau_1 = \mathcal{O}(\log n/\log d)$ such that*

$$\mathbf{Pr}\left(|I_{t_0+\tau_1}| \geqslant \frac{2n}{d}\right) \geqslant 1 - 2e^{-\frac{d}{56}}.$$

Lemma 4.7 (Flooding reaches almost-all nodes, part 2)**.** *Under the hypotheses of Theorem 4.5, for $\tau_1 = \mathcal{O}(\log n/\log d)$ as in Lemma 4.6 and for some $\tau_2 = \mathcal{O}(d)$,*

$$\mathbf{Pr}\left(|I_{t_0+\tau_1+\tau_2}| \geqslant (1 - e^{-\frac{d}{25}})n \mid |I_{t_0+\tau_1}| \geqslant \frac{2n}{d}\right) \geqslant 1 - \frac{1}{n^{0.8}}.$$

The proofs of Lemma 4.6 and Lemma 4.7 above are given in Subsections 4.2.1 and 4.2.2, respectively. In particular, Theorem 4.5 follows by taking $\tau = \tau_1 + \tau_2$, and noticing that

$$\begin{aligned}\mathbf{Pr}\left(|I_{t_0+\tau}| \geqslant (1 - e^{-\frac{d}{25}})n\right) &\geqslant \mathbf{Pr}\left(|I_{t_0+\tau_1}| \geqslant \frac{2n}{d}\right) \mathbf{Pr}\left(|I_{t_0+\tau_1+\tau_2}| \geqslant (1 - e^{-\frac{d}{25}})n \mid |I_{t_0+\tau_1}| \geqslant \frac{2n}{d}\right) \\ &\geqslant \left(1 - 2e^{-\frac{d}{56}}\right)\left(1 - \frac{1}{n^{0.8}}\right) \geqslant 1 - 3e^{-\frac{d}{56}},\end{aligned}$$

where the last inequality holds for sufficiently large n given that d is a constant.

4.2.1 Proof of Lemma 4.6

In this lemma we analyze the evolution of the flooding process from the onset, till a constant fraction of the nodes has been informed. To this purpose, we consider all nodes that are in the system at time t_0, when the informed node joins the network. On the other hand, we completely disregard nodes that were born in the interval $[t_0, t_0+\log n]$ (i.e., we don't count them as hits, though we keep into account that they remove probability mass from destinations in V_{t_0}), since their number is negligible with high probability from Lemma 3.9. Conversely, we do need to consider nodes that die in the interval $[t_0, t_0 + \log n]$, since their failure might adversely affect flooding. Removing them at the onset or at the end may be tricky, since in both cases we need to argue that their removal has no significant topological effects (we cannot simply assume an adversarial removal). To sidestep these challenges, we remove each of them with probability $\log n/n$ (this is an upper bound on the probability that a node dies in the interval $[t_0, t_0 + \log n]$) the first time it is reached by the flooding process. Moreover, to simplify the analysis we introduce the *Onion-Skin process*, which iteratively builds a connected, bipartite graph, corresponding to alternating paths, in which "young" nodes only connect to "old" ones. In particular, each realization of this process generates a subset of the edges generated by the original topology dynamics, so that each iteration of the process corresponds to a partial flooding in the original graph, in which a new layer of informed nodes is added to the subset of already informed ones.

The Onion-Skin process and its analysis. Let $m = |V_{t_0}|$. From Lemma 3.5, we know that $m \in [0.9n, 1.1n]$ with probability at least $1 - 1/n^2$. We now build a map $h : S \to [m]$, so that for $v \in S$, $h(v) = i$ if v is the i-th youngest node in the system at time t_0. Note that $h(s) = 1$ by definition. We next define $Y = \{v \in V_{t_0} : h(v) \leqslant m/2\}$ as the subset of *young* nodes and $O = \{v \in V_{T_{r_0}} : h(v) \geqslant m/2 + 1\}$ as the subset of *old* nodes.

Starting from s, the Onion-Skin process builds a connected, bipartite graph, so that young nodes are only connected to old ones. The process unfolds over a suitable number k of phases, with $k = \mathcal{O}(\log n/ \log d)$. Each phase consists of two steps, in each of which a suitable subset of young nodes attempt to establish links toward old nodes. In the following, we denote by $Y_k \subseteq Y$ and $O_k \subseteq O$ the subsets of young and old nodes that are informed by the end of phase k, respectively. In the remainder, we let $O_{-1} = \emptyset$ for notational convenience and, without loss of generality, we use the interval $[d]$ to number the links established by each vertex. The Onion-Skin process is formally described in Figure 4.2.1. It should be noted that each realization of this process generates a subset of the edges generated by the original topology dynamics. Moreover, each iteration of the process corresponds to a partial flooding in the original graph. Flooding is partial since i) the network uses a subset of the edges that would be present in the original graph and ii) every newly informed node tosses a coin and dies with probability $\log n/n$, which is an upper bound on the overall probability of that node dying in the interval $[T_{r_0}, T_{r_0} + \log n]$. This last fact trivially follows since the lifetime of a node obeys an exponential distribution with parameter $\mu = 1/n$.

In the remainder of this proof, all events are conditioned on the following event, whose definition is given in the proof of Lemma 4.2 and is repeated here to keep the proof self-contained:

$$L_{r_0} = \{\text{each node in } V_{T_r} \text{ was born after time } T_{r_0 - 7n\log n}\} \cap \{|V_{T_i}| \in [0.9n, 1.1n] \text{ for } i = r_0 - 7n\log n, \dots, r_0\}.$$

As already noted in the proof of Lemma 4.2, $\mathbf{Pr}(L_{r_0}) \geqslant 1 - 1/n^2$ from Lemma 3.5, Lemma 3.8 and a union bound.

It should be noted that in the Onion-Skin process, we are using the principle of deferred decisions, delaying decision as to the establishment of a link (u, v) to the moment one of its endpoints is informed in the flooding process. On the other hand, this means that the probability that the j-th connection originating from u has v as destination is equal to $1/|V_{T_r}|$ if u's arrival in the network corresponds to the r-th event, with $r \leqslant r_0$. The conditionings above (holding with probability at least $1-2/n^2$) ensure that the above probability fell in the interval $\left[\frac{1}{1.1n}, \frac{1}{0.9n}\right]$ for all nodes that were in the network at time T_{r_0}. Note also that we are completely disregarding nodes that join the network in the interval $[T_{r_0}, T_{r_0}+\log n]$. For these nodes, we are using Lemma 3.9 to conclude that their number is at most $4\log n$, w.h.p.

We next analyze Phase 0 and the generic Phase k separately. For each $i = 1, \dots, d$, for each node $v \in V_{t_0}$ and for each set $A \subseteq V_{T_{r_0}}$, we define the Bernoulli random variable $R_{v,A}$ as follows:

$$R_{v,A} = \begin{cases} 1 & \text{if } x \geqslant 1 \text{ connections from } v \text{ have destination in } A \\ 0 & \text{otherwise} \end{cases}$$

Onion-Skin process:

Phase 0**:** Let $Y_0 := \{s\}$ and let s establish d links. Let $Z_0 \subset O$ the subset of old nodes that are destinations of these links. Links with endpoints in Y are discarded. O_0 is obtained from Z_0 by removing each vertex in Z_0 (and the just established links) independently, with probability $\log n/n$.

Phase k $\geqslant$ 1:

Step 1. Each node in $Y - Y_{k-1}$ establishes links $\{1, \ldots, d/2\}$. Let $W_k \subseteq Y - Y_{k-1}$ the subset of nodes with at least one link to nodes in $O_{k-1} - O_{k-2}$. Links to nodes in Y are again discarded. $Y_k - Y_{k-1}$ is obtained from W_k by removing each vertex in W_k (and the just established links) independently, with probability $\log n/n$.

Step 2. Each node in $Y_k - Y_{k-1}$ establishes links $\{d/2+1, \ldots, d\}$. Let Z_k be the subset of nodes in $O - O_{k-1}$ that are reached by at least one such link. Links to nodes in Y are discarded. $O_k - O_{k-1}$ is obtained from Z_k by removing each node in Z_k (and the links just established) independently, with probability $\log n/n$.

Figure 1: The Onion-Skin process

Moreover, for each $i = 1, \ldots, d$ and for every $v \in V_{T_{r_0}}$ that joined the network at step $\hat{r} \leqslant r_0$, define the variable $A_v^{(i)} \in V_{T_{\hat{r}}}$ as

$$A_v^{(i)} = w, \text{ where } w \in V_{T_{\hat{r}}} \text{ is the destination of the } i\text{-th connection request of } v.$$

As for phase 0 of the Onion-Skin process, we prove the following claim.

Claim 4.8 (Analysis of Phase 0)**.** *At the end of Phase* 0*, it holds*

$$\mathbf{Pr}\left(|O_0| \geqslant \frac{d}{16}\right) \geqslant \left(1 - \frac{2\log n}{n}\right)\left(1 - 2e^{-\frac{d}{36}}\right).$$

Proof. For each $v \in O$,

$$\mathbf{Pr}\left(\exists i \in [d] : A_s^{(i)} = v\right) \geqslant 1 - \left(1 - \frac{1}{1.1n}\right)^d \geqslant 1 - e^{-\frac{d}{1.1n}},$$

which implies, since $\frac{d}{1.1n} \leqslant 0.5$,

$$\mathbf{E}[Z_0] = |O|\left(1 - e^{-\frac{d}{1.1n}}\right) \geqslant \frac{0.9n}{2} \cdot \frac{d}{1.43n} \geqslant \frac{d}{4},$$

where the last equality follows since $|O| \geqslant 0.9n/2$. We next bound the probability that Z_0 is smaller than $d/8$. To this purpose, we cannot simply apply a Chernoff bound to the binary variables that describe whether or not a node $v \in O$ was the recipient of at least one link originating from s, since these are not independent. We instead resort to Theorem A.2. In particular, we define the function $f(A_s^{(1)}, \ldots, A_s^{(d)}) = |Z_0|$. Clearly, f is well-defined and it satisfies the Lipschitz condition with values $\beta_1 = \cdots = \beta_d = 1$, since changing the destination of one link can affect the value of $|Z_0|$ by at most 1. We can thus apply Theorem A.2 to obtain:

$$\mathbf{Pr}\left(|Z_0| < \frac{d}{8}\right) \leqslant \mathbf{Pr}\left(|Z_0| < \mathbf{E}[|Z_0|] - \frac{d}{8}\right) \leqslant e^{-\frac{d}{32}}.$$

We next argue about $|O_0|$. If $|Z_0| = x$ and B is the number of nodes in Z_0 that are removed,

$$\mathbf{E}[B] \leqslant \frac{x \log n}{n}.$$

Applying Markov's inequality, for sufficiently large n,

$$\mathbf{Pr}\left(B \geqslant \frac{x}{2}\right) \leqslant \frac{2\log n}{n},$$

whence the thesis immediately follows. □

We next analyze Phase k of the Onion-Skin process. We first examine Step 1 of this phase, i.e., the number of nodes in $Y - Y_{k-1}$ that connect to nodes in $O_{k-1} - O_{k-2}$ using connection requests that belong to the subset $\{1, \ldots, d/2\}$.

Claim 4.9 (Analysis of Phase k - Step 1)**.** *Assume that $|O_{k-1}| \leqslant n/d$ and $|Y_{k-1}| \leqslant n/d$. For sufficiently large n, at the end of Phase k, it holds*

$$\mathbf{Pr}\left(|Y_k - Y_{k-1}| \geqslant \frac{yd}{28} \;\middle|\; |O_{k-1} - O_{k-2}| = y\right) \geqslant \left(1 - \frac{2\log n}{n}\right)\left(1 - e^{-\frac{yd}{56}}\right). \tag{15}$$

Proof. It should be noted that: i) no connections originating from nodes in $Y - Y_{k-1}$ have been established so far (otherwise, the involved nodes would belong to Y_{k-1}, a contradiction), and ii) as a consequence, for each $u \in Y - Y_{k-1}$, none of its connections in $\{1, \ldots, d/2\}$ had destination in O_{k-2}. From the definition of $|W_k|$,

$$|W_k| = \sum_{v \in Y - Y_{k-1}} R_{v, O_{k-1} - O_{k-2}}.$$

Then, if $v \in Y - Y_{k-1}$,

$$\mathbf{Pr}\left(R_{v, O_{k-1} - O_{k-2}} = 1 \;\middle|\; |O_{k-1} - O_{k-2}| = y\right) \geqslant 1 - \left(1 - \frac{y}{1.1n}\right)^{\frac{d}{2}} > 1 - e^{-\frac{yd}{2.2n}},$$

since at most $1.1n$ nodes were present when v joined the network. Since $\frac{yd}{2.2n} \leqslant 0.5$,

$$1 - e^{-\frac{yd}{2.2n}} > \frac{yd}{3n}.$$

As a consequence,

$$\mathbf{E}\left[|W_k| \mid |O_{k-1} - O_{k-2}| = y\right] \geqslant |Y - Y_{k-1}| \cdot \frac{yd}{3n} \geqslant \frac{yd}{7},$$

where the last inequality follows from the fact that $|Y - Y_{k-1}| \geqslant (0.45n - n/d)$ and $d \geqslant 112$. On the other hand, the $R_{v, O_{k-1} - O_{k-2}}$'s are mutually independent, so we can apply a standard Chernoff bound (Theorem A.1) to obtain

$$\mathbf{Pr}\left(|W_k| < \frac{yd}{14} \;\middle|\; |O_{k-1} - O_{k-2}| = y\right) \leqslant e^{-\frac{yd}{56}}.$$

If B denotes the number of vertices removed from W_k, we can proceed as in the analysis of Step 2 of Phase 0 to obtain (15).

□

We now examine Step 2 of Phase k, i.e., we consider nodes in $Y_k - Y_{k-1}$ that connect to nodes in $O - O_{k-1}$ using the requests labeled in $\{d/2 + 1, \ldots, d\}$.

Claim 4.10 (Analysis of Phase k - Step 2)**.** *Assume that $|O_{k-1}| \leqslant n/d$ and $|Y_{k-1}| \leqslant n/d$. For sufficiently large n, at the end of Phase k, it holds*

$$\mathbf{Pr}\left(|O_k - O_{k-1}| \geqslant \frac{xd}{28} \;\middle|\; |Y_k - Y_{k-1}| = x\right) \geqslant \left(1 - \frac{2\log n}{n}\right)\left(1 - e^{-\frac{xd}{56}}\right). \tag{16}$$

Proof. For every $v \in O - O_{k-1}$, it holds

$$\mathbf{Pr}\left(\exists u \in Y_k - Y_{k-1}, \exists i \in [d/2] : A_u^{(i)} = v \;\middle|\; |Y_k - Y_{k-1}| = x\right) \geqslant 1 - \left(1 - \frac{1}{1.1n}\right)^{\frac{xd}{2}} \geqslant 1 - e^{-\frac{xd}{2.2n}}.$$

since at most $1.1n$ nodes were present when v joined the network. Since $\frac{xd}{2.2n} \leqslant 0.5$, we have that $1 - e^{-\frac{xd}{2.2n}} \geqslant \frac{xd}{3n}$. From the facts above and from the definition of Z_k in the Onion-Skin process,

$$\mathbf{E}\left[|Z_k| \mid |Y_k - Y_{k-1}| = x\right] \geqslant |O - O_{k-1}| \cdot \frac{xd}{3n} \geqslant \frac{xd}{7},$$

where, similarly to Claim 4.9, we used $|O - O_{k-1}| \geqslant 0.45n - n/d$ and $d \geqslant 112$.

Differently from Claim 4.9, we cannot simply concentrate, since the events $\{A_u^{(i)} = v\}$ are negatively correlated as v varies over $O - O_{k-1}$. We again resort to Theorem A.2. In this case, we have $\frac{xd}{2}$ connections that are established (independently of each other) from vertices in $Y_k - Y_{k-1}$. As for the set of $xd/2$ variables $\{A_u^{(i)}\}_{u \in Y_k - Y_{k-1}, i \in [d/2]}$, we notice that the domain of $A_u^{(i)}$ is the set V_t if u joined the system at time t and remind that we are conditioning to $0.9n \leqslant |V_t| \leqslant 1.1n$, as explained at the beginning of the proof. We next define the function

$$f\left(\{A_u^{(i)}\}_{u \in Y_k - Y_{k-1}, i \in [d/2]}\right) = |Z_k|,$$

and, like in the analysis of Step 1 of Phase 0, we observe that f satisfies the Lipschitz condition with constants $\beta_1 = \cdots = \beta_{\frac{xd}{2}} = 1$. We can thus apply Theorem A.2 to obtain

$$\mathbf{Pr}\left(|Z_k| < \frac{xd}{14} \;\middle|\; |Y_k - Y_{k-1}| = x\right) \leqslant e^{-\frac{xd}{56}}.$$

Finally, we remove nodes from Z_k exactly as we did in Phase 0 and in Step 1 of Phase k. The analysis proceeds exactly the same, so that we can get (16).

□

Thanks to Claims 4.8, 4.9 and 4.10, we can prove the following conclusive result for the Onion-Skin process that easily implies Lemma 4.6.

Claim 4.11. *A positive integer $k = \mathcal{O}(\log n / \log d)$ exists such that*

$$\mathbf{Pr}\left(|Y_k \cup O_k| \geqslant \frac{2n}{d}\right) \geqslant 1 - 2e^{-\frac{d}{56}}.$$

Proof. Consider the generic k-th phase and assume that $|Y_{k-1}| \leqslant n/d$ and $|O_{k-1}| \leqslant n/d$. Then, Claims 4.8, 4.9 and 4.10 combined with the chain rule imply

$$\mathbf{Pr}\left(\left\{|Y_k - Y_{k-1}| \geqslant \left(\frac{d}{28}\right)^{2k+1}\right\} \cap \left\{|O_k - O_{k-1}| \geqslant \left(\frac{d}{28}\right)^{2k}\right\}\right) \geqslant \prod_{i=0}^{2k}\left(1 - e^{-\left(\frac{d}{28}\right)^i \frac{d}{56}}\right)\left(1 - \frac{2\log n}{n}\right)^{2k+1}.$$

So, for some $k = \frac{\log(n/d)}{2\log(d/28)}$, at the end of Phase k,

$$\mathbf{Pr}\left(\left\{|Y_k - Y_{k-1}| \geqslant \frac{n}{d}\right\} \cap \left\{|O_k - O_{k-1}| \geqslant \frac{n}{d}\right\}\right) \geqslant \prod_{i=0}^{2k}\left(1 - e^{-\left(\frac{d}{28}\right)^i} \frac{d}{56}\right)\left(1 - \frac{2\log n}{n}\right)^{2k+1}. \tag{17}$$

For the expression

$$P = \prod_{i=0}^{2k+1}\left(1 - e^{-\left(\frac{d}{28}\right)^i \frac{d}{56}}\right),$$

we next show that $P \geqslant 1 - e^{-\frac{d}{56}}$. We start with the following bound

$$-\log P = -\sum_{i=0}^{2k+1} \log\left(1 - e^{-\left(\frac{d}{28}\right)^i \frac{d}{56}}\right) \leqslant \sum_{i=0}^{2k+1} \log\left(\frac{1}{1 - e^{-\left(\frac{d}{28}\frac{d}{56}\right)^i}}\right) \leqslant 2\sum_{i=0}^{2k+1} e^{-\left(\frac{d}{28}\right)^i \frac{d}{56}} \leqslant e^{-\frac{d}{56}},$$

where: i) the third inequality holds because, for each $x \in [0,1]$, $x \geqslant \log\left(\frac{2}{2-x}\right)$, and ii) the last inequality follows for any $d \geqslant 112$ and from the fact that $\sum_{i=0}^{\infty} e^{-\left(\frac{d}{28}\right)^i} \leqslant \frac{1}{2}$. So,

$$P = \frac{1}{e^{\log P}} \geqslant e^{-e^{d/56}} \geqslant 1 - e^{-\frac{d}{56}}, \tag{18}$$

where the last inequality follows since $e^{-x} \geqslant 1 - x$ for each $x \geqslant 0$. Finally, from (17) and (18),

$$\mathbf{Pr}\left(\left\{|Y_k - Y_{k-1}| \geqslant \frac{n}{d}\right\} \cap \left\{|O_k - O_{k-1}| \geqslant \frac{n}{d}\right\}\right) \geqslant \left(1 - e^{-\frac{d}{56}}\right)\left(1 - \frac{2\log n}{n}\right)^{2k+1} \geqslant 1 - 2e^{-\frac{d}{56}},$$

where the last inequality follows for n sufficiently large and since $k = \frac{\log(n/d)}{2\log(d/28)}$.

□

4.2.2 Proof of Lemma 4.7

We suppose $|I_{t_0+\tau_1}| \geqslant 2n/d$ and, for the sake of simplicity, we omit this conditioning for the rest of the proof. We then prove that, for $t \geqslant t_0 + \tau_1$, the size of the set of informed nodes grows by a constant factor in each flooding round, reaching $(1 - e^{-d/10})n$ within τ_2 rounds, with $\tau_2 = \Theta(d)$. To this aim, we note that Lemma 4.3 implies that, for $d \geqslant 20$, the graph $G_t = (V_t, E_t)$ is w.h.p. an expander for sets of large size, i.e.,

$$\min_{\substack{S \subseteq V_t: \\ ne^{-d/20} \leqslant |S| \leqslant 1.1n/2}} \frac{|\partial_{out}(S)|}{|S|} \geqslant 0.1\,.$$

Indeed, for $t \geqslant t_0 + \tau_1$, $|I_t|$ grows by a constant factor in each round, as long as $|I_t| \leqslant 1.1n/2$. To see this, recall that, from Definition 3.3,

$$I_{t+1} = (I_t \cup \partial^t_{out}(I_t \cap V_{t+1})) \cap V_{t+1}.$$

Moreover, from Lemma 3.9, in each time unit, with probability exceeding $1-1/n$, at most $4 \log n$ informed nodes leave the network. As a consequence, as long as $|I_t| \leqslant 1.1n/2$,

$$\mathbf{Pr}\left(|I_{t+1}| \geqslant 1.1|I_t| - 5 \log n\right) \geqslant 1 - \frac{2}{n},$$

since this is the probability that G_t is an expander for large subsets, and that in $[t, t+1]$ at most $4 \log n$ nodes leave the network. Since $2n/d \leqslant |I_t| \leqslant 1.1n/2$, we have $|I_{t+1}| \geqslant 1.05|I_t|$ for sufficiently large n, with probability at least $1 - 2/n$. This implies that, within $\tau_2' = \Theta(\log d)$ rounds, we will get $|I_{t_0+\tau_1+\tau_2'}| \geqslant 1.1n/2$, with probability at least $1-2d/n$. For $t \geqslant t_0+\tau_1+\tau_2'$, we turn to the set $S_t = V_t - I_t$ of non-informed nodes in the graph at time t, showing that its size decreases by a constant factor in each flooding round, as long as $|S_t| \geqslant ne^{-d/20}$. To begin, we note that

$$\partial_{out}(S_{t+1} \cap V_t) \subseteq S_t - (S_{t+1} \cap V_t)\,, \tag{19}$$

since $\partial_{out}(S_{t+1} \cap V_t)$ are nodes reachable in one edge from the non-informed nodes, and so they were not informed in the previous time. So, from (19) and Lemma 4.3, since $|S_{t+1} \cap V_t| \leqslant 1.1n/2$, we get that, as long as $|S_{t+1} \cap V_t| \leqslant ne^{-d/20}$,

$$\mathbf{Pr}\left(|S_t| \geqslant 1.1|S_{t+1} \cap V_t|\right) \geqslant 1 - \frac{1}{n}\,, \tag{20}$$

(this is indeed the probability that G_t is an expander for large subsets). Since, from Lemma 3.9, at each step, at most $4 \log n$ nodes join the network with probability $1 - 1/n$, we have that $|S_{t+1}| \leqslant |S_{t+1} \cap V_t| + 4 \log n$. This fact, together with (20) implies that, as long as $|S_{t+1} \cap V_t| \leqslant ne^{-d/20}$,

$$\mathbf{Pr}\left(|S_{t+1}| \leqslant 0.91|S_t| + 4 \log n\right) \geqslant 1 - \frac{2}{n}\,.$$

This implies that in $\tau_2 = \Theta(d)$ rounds, we will get

$$|S_{t_0+\tau_1+\tau_2} \cap V_{t_0+\tau_1+\tau_2}| \leqslant ne^{-d/20}$$

with probability at least $1 - 1/n^{0.9}$. So, again from Lemma 3.9, $|S_{t_0+\tau_1+\tau_2}| \leqslant ne^{-d/25}$, with probability at least $1 - 1/n^{0.8}$.

5 Graphs with edge regeneration

In this section we will study Poisson dynamic graphs with continuous edge regeneration that can be formalized as follows.

Definition 5.1 (Poisson dynamic random graphs with edge regeneration)**.** *A* Poisson Dynamic Graph with edge Regeneration $\mathcal{G}(\lambda, \mu, d)$ *(for short, PDGR) is a continuous dynamic random graph* $\{G_t = (V_t, E_t) : t \in \mathbb{R}^+\}$*, where the set of nodes* V_t *evolves according to Definition 3.4, while the set of edges* E_t *evolves according to the following topology dynamics:*

1. *When a new node appears, it makes d independent connection* requests*, each one to a* destination *node chosen uniformly at random among the nodes in the network, and, each of such requests is accepted by the destination so that the edge is activated (i.e. it is included in E_t). The d independent requests will be indexed in the range $\{1, \dots, d\}$.*

2. When a node dies, all its incident edges disappear (i.e. they are removed from E_t);
3. When a node has one of its d outgoing edges disappearing, it makes a new connection request to a destination node chosen uniformly at random among all the nodes in the network and the destination accepts the request thus activating the corresponding edge.

5.1 Preliminary properties

A first immediate property of the PDGR model is that every node always has degree at least d.

We next provide an upper bound on the probability that a fixed node chooses any other active node in the network as destination of one of its d requests. We remark that the bound $O(1/n)$, we used in the analysis of the model without edge regeneration, does not hold in this setting essentially because of the edge regeneration and the presence of "very old" nodes (i.e. those nodes having age $\omega(n)$). Thanks to a more refined analysis, the next lemma shows an upper bound as function of the age of the nodes.

Lemma 5.2. *For every constant $d \geqslant 20$ and for every sufficiently large n, let $\{G_t = (V_t, E_t) : t \in \mathbb{R}^+\}$ be a PDGR sampled from $\mathcal{G}(\lambda, \mu, d)$ with $\lambda = 1$ and $\mu = 1/n$. Then, for every fixed integer $r = \Omega(n \log n)$, consider the snapshot G_{T_r}. Let $u \in V_{T_r}$ be a node born in step T_{r-i} for some integer $i \leqslant r$. Then, if a node $v \in V_{T_r}$ is born before u, the probability that a fixed request of u has destination v is at most*

$$\frac{1}{0.8n}\left(1 + \frac{i}{1.7n}\right). \tag{21}$$

While, if v is born after u, the probability that a fixed request of u has destination v is always $\leqslant \frac{1}{0.8n}$.

Proof. We first bound the probability that a fixed request of u has destination v when v is younger than u. We define the event[3]

$$A_{u,v} = \{\text{a fixed request of } u \text{ has destination } v \text{ at time } T_r\},$$

and the event

$$L_r = \{\text{each node in } V_{T_r} \text{ is born after time } T_{r-7n\log n}\} \cap \{|V_{T_i}| \in [0.9n, 1.1n] \text{ with } i = r - 7n\log n, \dots, r\}.$$

From Lemma 3.5, Lemma 3.8 and an union bound we get that $\mathbf{Pr}(L_r) \geqslant 1 - 1/n^2$. We notice that the event L_r means that, when each node in V_{T_r} joined the network, the network has at least $0.9n$ nodes and at most $1.1n$ nodes. From the law of total probability,

$$\mathbf{Pr}(A_{u,v}) \leqslant \mathbf{Pr}(A_{u,v} \mid L_r) + \frac{1}{n^2} \leqslant \frac{1}{0.9n} + \frac{1}{n^2} \leqslant \frac{1}{0.8n},$$

where $\mathbf{Pr}(A_{u,v} \mid L_r) \leqslant 1/(0.9n)$ since u can choose v only after a death of one of its neighbors, being v younger than u.

We now analyze the case in which v is older than u, where u is born at step T_{r-i}. For the law of total probability,

$$\mathbf{Pr}(A_{u,v}) \leqslant \mathbf{Pr}(A_{u,v} \mid L_r) + \frac{1}{n^2}. \tag{22}$$

So, we need to evaluate $\mathbf{Pr}(A_{u,v} \mid L_r)$. For each $k \geqslant 1$ and $w \in V_{T_k}$, define the event

$$D_{w,k} = \{w \text{ dies at time } T_k\}.$$

To bound $\mathbf{Pr}(D_{w,k} \mid L_r)$, for each $k = r-i, \dots, r$ and $w \in N_{T_k}$, we use Lemma 3.7 to get $\mathbf{Pr}(D_{w,k}) \leqslant 1/(1.8n)$, and, hence, for the definition of conditional probability,

$$\mathbf{Pr}(D_{w,k} \mid L_r) = \frac{\mathbf{Pr}(D_{w,k} \cap L_r)}{\mathbf{Pr}(L_r)} \leqslant \frac{\mathbf{Pr}(D_{w,k})}{1 - 1/n^2} = \frac{1/1.8n}{1 - 1/n^2} \leqslant \frac{1}{1.7n}. \tag{23}$$

Now, for each $j = r-i, \dots, r$, define the following events

$$A^j_{u,v} = \{\text{a fixed request of } u \text{ connects to } v \text{ at time } T_j\},$$

[3]We here avoid to index the specific requests of u since the considered graph process is perfectly symmetric w.r.t. the d random requests of every node.

and write $A_{u,v} = \cup_{j=r-i}^{r} A_{u,v}^{j}$. Notice that, for $j = r - i$, it holds that

$$\mathbf{Pr}\left(A_{u,v}^{r-i} \mid L_r\right) \leqslant \frac{1}{0.9n}, \tag{24}$$

since this is the probability that the request of u has destination v at the time of u's arrival (since v is older than u). On the other hand, for each $j = r - i + 1, \ldots, r$, for the law of total probability, we have[4]

$$\mathbf{Pr}\left(A_{u,v}^{j} \mid L_r\right) \leqslant \mathbf{Pr}\left(A_{u,v}^{j} \,\middle|\, L_r, \cap_{j'=r-i}^{j-1}\left(A_{u,v}^{j'}\right)^{C}\right), \tag{25}$$

since any fixed request of u can choose v as destination at step T_j only if, at step T_{j-1}, the same request was not connected to v, so the second term derived from the law of total probability is equal 0, and we omitted it. We notice that, conditioning on the event that u, at time T_{j-1}, is not connected to v, the event $A_{u,v}^{j}$ is the intersection between the two events "the node connected to the fixed request of u dies at time T_j" and "the fixed request of u is re-connected to v at time T_j". So, for the definition of conditional probability,

$$\mathbf{Pr}\left(A_{u,v}^{j} \,\middle|\, L_r, \cap_{j'=r-i}^{j-1}\left(A_{u,v}^{j'}\right)^{C}\right) \leqslant \frac{1}{1.7n} \cdot \frac{1}{0.9n}, \tag{26}$$

where the first factor $1/(1.7n)$ in the r.h.s. of (26) is an upper bound on the probability (conditional to L_r from (23)) that the node to which u is connected dies at time T_j. Moreover, $1/(0.9n)$ is the probability, conditional to L_r, that the request of u connects to v at time T_j, if its neighbour is died at time T_j. So, recalling that $A_{u,v} = \cup_{j=r-i}^{r} A_{u,v}^{j}$, from (24), (25) and (26),

$$\mathbf{Pr}\left(A_{u,v} \mid L_r\right) \leqslant \sum_{j=r-i}^{r} \mathbf{Pr}\left(A_{u,v}^{j} \mid L_r\right) \leqslant \frac{1}{0.9n}\left(1 + \frac{i}{1.7n}\right). \tag{27}$$

Finally, since conditional to L_r, we get $i \leqslant 7n \log n$, using (27) in (22), the proof is completed.

□

5.2 Expansion properties

The expansion property satisfied by the Poisson model with edge regeneration can be formalized as follows.

Theorem 5.3 (Expansion)**.** *For every constant $d \geqslant 35$ and for every sufficiently large n, let $\{G_t = (V_t, E_t) : t \in \mathbb{R}^+\}$ be a PDGR sampled from $\mathcal{G}(\lambda, \mu, d)$ with $\lambda = 1$ and $\mu = 1/n$. Then, for every fixed integer $r \geqslant 7n \log n$, with probability at least $1 - 1/n$, the snapshot G_{T_r} is a vertex ε-expander with parameter $\varepsilon \geqslant 0.1$, i.e.,*

$$\min_{\substack{S \subseteq V_{T_r} \\ 0 \leqslant |S| \leqslant 1.1n/2}} \frac{|\partial_{out}(S)|}{|S|} \geqslant 0.1.$$

The proof proceeds analyzing three different size ranges of the vertex subset $S \subseteq V_{T_r}$, the expansion of which has to be shown. We start by analyzing the most interesting case, namely, the expansion of middle-size subsets.

Expansion of middle-size subsets. We consider subsets of size in the range $n/\log^2 n \leqslant |S| \leqslant n/14$, the analysis of which definitely represents one of the key technical contributions of this paper. Indeed, departing from the other size ranges, we cannot use any rough, worst-case counting argument: for instance, assuming that all nodes in the considered subset S have age $\mathcal{O}(n \log n)$ and applying the corresponding edge-probability bound given by (21) would lead to a useless, too large union bound for the probability of non-expansion for some subset S.

To cope with this technical issue, we need to partition and classify the subsets S and T according to their *age profile*. More in detail, we first define a sequence of $\Theta(\log n)$ *slices* of possible nodes ages and then we provide an effective age profile of each subset S (and T) depending on how large its intersection is with each of these slices. Thanks to the properties of the exponential distributions of the life of every

[4]The term E^C denotes the complement of the event E.

node in the Poisson model (see (3) in Lemma 3.7), we show that the existence of a given subset in a given time has a probability that essentially depends on its profile. Roughly speaking, the more is the number of old nodes in S, the less is the probability of the presence of S in V_{T_r}.

Then, combining this profiling with a more refined use of the parameterized bound on the edge probability in (21), we get a mathematical expression (see (45)) that, in turn, we show to be dominated by the KL divergence of two suitably defined probability distributions. Finally, our target probability bound, stated in the next lemma, is obtained by the standard KL divergence inequality (see Theorem A.4). The arguments above allow us to prove the following result.

Lemma 5.4 (Expansion of middle-size subsets)**.** *Under the hypothesis of Theorem 5.3, for subsets S of V_{T_r}, with probability of at least $1 - 2/n^2$,*

$$\min_{n/\log^2 n \leqslant |S| \leqslant n/14} \frac{|\partial_{out}(S)|}{|S|} \geqslant 0.1.$$

Proof. From Lemma 3.8, all the nodes in V_{T_r} are born after time $T_{r-7n\log n}$ with probability of at least $1 - 1/n^2$. So, if we define the event

$$L_r = \{\text{each node in } V_{T_r} \text{ is born after time } T_{r-7n\log n}\},$$

we get that $\mathbf{Pr}(L_r) \geqslant 1 - 1/n^2$: through the rest of this proof, we will condition to this event. So, if we denote i as the node that joined the network at step T_{r-i+1} (i.e. the node has age i), conditioning on L_r,

$$V_{T_r} \subseteq \{1, 2, 3, \ldots, 7n\log n\}.$$

We have to show that any two disjoint sets $S, T \subseteq \{1, 2, \ldots, 7n\log n\}$, such that $n/\log^2 n \leqslant |S| \leqslant n/14$, $|T| = 0.1|S|$, $S, T \subseteq V_{T_r}$, and $\partial_{out}(S) \subseteq T$, may exist only with negligible probability. To this aim, we define the following event

$$A_{S,T} = \{\partial_{out}(S) \subseteq T\} \cap \{S, T \subseteq V_{T_r}\}.$$

For the law of total probability,

$$\mathbf{Pr}\left(\min_{n/\log^2 n \leqslant |S| \leqslant n/14} \frac{|\partial_{out}(S)|}{|S|} \leqslant 0.1\right) \leqslant \sum_{\substack{n/\log^2 n \leqslant |S| \leqslant n/2, |T| = 0.1|S| \\ S,T \subseteq \{1,2,\ldots,7n\log n\}}} \mathbf{Pr}(A_{S,T} \mid L_r) + \frac{1}{n^2}, \tag{28}$$

and, hence, our next goal is to upper bound the quantity $\mathbf{Pr}(A_{S,T} \mid L_r)$. For each $i \in S$, let B_i be the event " each of the d requests of node i has destination in $S \cup T$ ". Then, we can write

$$A_{S,T} = \bigcap_{i \in S} B_i \cap \{S, T \in V_{T_r}\},$$

and, for Bayes' rule,

$$\mathbf{Pr}(A_{S,T} \mid L_r) = \mathbf{Pr}(\cap_{i \in S} B_i \mid S, T \subseteq V_{T_r}, L_r)\, \mathbf{Pr}(S, T \subseteq V_{T_r} \mid L_r). \tag{29}$$

From Lemma 5.2, conditional to the event $\{S, T \subseteq V_{T_r}\}$,

$$\mathbf{Pr}(B_i \mid S, T \subseteq V_{T_r}, L_r) = \left[\frac{|S \cup T|}{0.8n}\left(1 + \frac{i}{1.7n}\right)\right]^d. \tag{30}$$

Since we will use (29) to bound $\mathbf{Pr}(A_{S,T} \mid L_r)$, we need an upper bound for $\mathbf{Pr}(S, T \subseteq V_{T_r} \mid L_r)$. To this aim, we can use the bound in (3) of Lemma 3.7. However, according to the definition of *step* in Definition 3.6, we know that the death of one node in one single step is not independent of the death of the others. Indeed, if we know that, in a given step, the node v dies, we will also know that in this step no other event occurs, and, so, none of the other nodes dies. Moreover, if we know that one node does not die in a given step, the probability to die of the other nodes will be larger. To cope with this issue, we consider the probability that a fixed set of node survives in one step. From Lemma 3.7, for an arbitrary set of k nodes, it holds

$$\mathbf{Pr}\left(v_1, \ldots, v_k \in V_{T_r} \mid v_1, \ldots, v_k \in V_{T_{r-1}}, L_r\right) \leqslant 1 - \frac{k}{2.2n} \leqslant \left(1 - \frac{1}{2.2n}\right)^k, \tag{31}$$

where the last inequality follows from the binomial inequality, i.e. $\forall x \in \mathbb{R}$ s.t. $x \geqslant -1$ and $\forall k \in \mathbb{N}$ it holds $(1+x)^k \geqslant 1+kx$. We notice that (31) is the probability that the next step is not characterized by the death of any of the k considered nodes. So, thanks to (31) and to the Markov property of $\{V_{T_r} : r \in \mathbb{N}\}$,

$$\mathbf{Pr}\left(S, T \subseteq V_{T_r} \mid L_r\right) \leqslant \prod_{i \in S \cup T} \left(1 - \frac{1}{2.2n}\right)^i \leqslant \prod_{i \in S \cup T} e^{-i/2.2n}, \tag{32}$$

where, in (32) we used the fact that, from (31), each node contributes in the product with a factor $1-1/(2.2n)$ for each step of its life. Since each node chooses the destination of its out-edges independently of the other nodes, combining (30) and (32) with (29), we obtain

$$\mathbf{Pr}\left(A_{S,T} \mid L_r\right) \leqslant \prod_{i \in S \cup T} e^{-i/2.2n} \cdot \prod_{i \in S} \min\left\{1, \left[\frac{|S \cup T|}{0.8n}\left(1 + \frac{i}{1.7n}\right)\right]^d\right\}. \tag{33}$$

For each set $H \subseteq V_{T_r}$, we define the sequence $(K_1^H, \ldots, K_L^H)$ (where $L = 7 \log n$), whose goal is to classify the nodes of the set according to their *age profile*:

$$\begin{aligned}
K_1^H &= |H \cap \{1, 2, \ldots, n\}|, \\
K_2^H &= |H \cap \{n+1, \ldots 2n\}|, \\
&\ldots \\
K_L^H &= |H \cap \{(L-1)n+1, \ldots, Ln\}|.
\end{aligned}$$

Notice that, if $|H| = h$ and $K_1^R = h_1, \ldots, K_L^R = h_L$, then it must hold $\sum_{m=1}^L h_m = h$. For each set $H \subseteq V_{T_r}$, we denote the vector of random variables $(K_1^H, \ldots, K_L^H)$ as $\mathbf{K}^H$. According to this definition, by setting $\mathbf{k} = (k_1, \ldots, k_L)$ and $\mathbf{h} = (h_1, \ldots, h_L)$, we can rewrite (28) as follows

$$\mathbf{Pr}\left(\min_{n/\log^2 n \leqslant |S| \leqslant n/14} \frac{|\partial_{out}(S)|}{|S|} \leqslant 0.1\right) \tag{34}$$

$$\leqslant \sum_{k=n/\log^2 n}^{n/14} \sum_{\substack{k_1+\cdots+k_L=k \\ h_1+\cdots+h_L=0.1k}} \sum_{\substack{S,T:\ \mathbf{K}^S=\mathbf{k} \\ \mathbf{K}^T=\mathbf{h}}} \mathbf{Pr}\left(A_{S,T} \text{ s.t. } \mathbf{K}^S = \mathbf{k},\ \mathbf{K}^T = \mathbf{h} \;\middle|\; L_r\right) + \frac{1}{n^2}.$$

Indeed, we have to sum over all the possible size $k = n/\log^2 n, \ldots, n/14$ of the set S, all the possible vectors $\mathbf{k}$ and $\mathbf{h}$ whose sum of the elements is equal to k and $0.1k$, respectively (i.e. the characterization of the age profiles of S and T with $|S| = k$ and $|T| = 0.1k$), and, finally, over all the possible sets S, T characterized by $\mathbf{K}^S = \mathbf{k}$ and $\mathbf{K}^T = \mathbf{h}$, respectively.

From (33), we get

$$\mathbf{Pr}\left(A_{S,T} \text{ s.t. } \mathbf{K}^S = \mathbf{k},\ \mathbf{K}^T = \mathbf{h} \;\middle|\; L_r\right)$$

$$\leqslant p(\mathbf{k}, \mathbf{h}) = \prod_{m=1,\ldots,L} \left(e^{-0.4(m-1)(k_m+h_m)} \min\left\{1, \left[\frac{|S \cup T|}{0.8n}(1+0.6m)\right]^{dk_m}\right\}\right). \tag{35}$$

The number of subsets $S, T \subseteq \{1, 2, \ldots, 7n \log n\}$ such that $(K_1^S, \ldots, K_L^S) = (k_1, \ldots, k_L)$ and $(K_1^T, \ldots, K_L^S) = (h_1, \ldots, h_L)$ is bounded by

$$n(\mathbf{k}, \mathbf{h}) = \prod_{\substack{m=1 \\ k_m \neq 0}}^L \binom{n}{k_m} \cdot \prod_{\substack{\ell=1 \\ h_\ell \neq 0}}^L \binom{n}{h_\ell}. \tag{36}$$

So, we introduce the quantity $s(\mathbf{k}, \mathbf{h})$ and, from (35) and (36),

$$s(\mathbf{k}, \mathbf{h}) = \sum_{\substack{S,T:\mathbf{K}^S=\mathbf{k} \\ \mathbf{K}^T=\mathbf{h}}} \mathbf{Pr}\left(A_{S,T} \text{ s.t. } \mathbf{K}^S = \mathbf{k},\ \mathbf{K}^T = \mathbf{h}\right) \leqslant n(\mathbf{k}, \mathbf{h}) \cdot p(\mathbf{k}, \mathbf{h}). \tag{37}$$

Combining (36) and (35) with (37) and, since $|S \cup T| = 1.1k$,

$$s(\mathbf{k}, \mathbf{h}) \leqslant \prod_{\substack{\ell=1 \\ h_\ell \neq 0}}^L \binom{n}{h_\ell} e^{-0.4(\ell-1)h_\ell} \cdot \prod_{\substack{m=1 \\ k_m \neq 0}}^L \left(\binom{n}{k_m} e^{-0.4(m-1)k_m} \min\left\{1, \left(\frac{1.1k}{0.8n}(1+0.6m)\right)^{dk_m}\right\}\right).$$

The next step is to prove that $s(\mathbf{k},\mathbf{h}) \leqslant 2^{-0.15k}$ and, to this aim, we split $s(\mathbf{k},\mathbf{h})$ in two factors, $s_1(\mathbf{k},\mathbf{h})$ and $s_2(\mathbf{k},\mathbf{h})$, defined as follows:

$$s_1(\mathbf{k},\mathbf{h}) = \prod_{\substack{m=1\\h_m\neq 0}}^{L} \binom{n}{h_m} e^{-0.4(m-1)h_m},$$
$$s_2(\mathbf{k},\mathbf{h}) = \prod_{\substack{m=1\\k_m\neq 0}}^{L} \binom{n}{k_m} e^{-0.4(m-1)k_m} \min\left\{1, \left(\frac{1.1k(1+0.6m)}{0.8n}\right)^{dk_m}\right\}.$$

To give an upper bound on $s(\mathbf{k},\mathbf{h})$, we provide separate upper bounds for $\log(s_1(\mathbf{k},\mathbf{h}))$ and $\log(s_2(\mathbf{k},\mathbf{h}))$. In particular, we want to show that

$$\log(s(\mathbf{k},\mathbf{h})) \leqslant -0.15k, \tag{38}$$

which implies that

$$s(\mathbf{k},\mathbf{h}) \leqslant 2^{-0.15k}. \tag{39}$$

We will start bounding $\log(s_1(\mathbf{k},\mathbf{h}))$. Using $\binom{n}{k} \leqslant \left(\frac{n\cdot e}{k}\right)^k$,

$$\log(s_1(\mathbf{k},\mathbf{h})) \leqslant \sum_{\substack{m=1\\h_m\neq 0}}^{L} h_m \log\left(\frac{n}{h_m} e^{-0.4m+1.4}\right). \tag{40}$$

Since $\log(x)$ is a concave function, we can apply Jensen's inequality in Theorem A.3 taking $a_m = h_m$ $x_m = \frac{n}{h_m} e^{-0.4m+1.4}$ and, recalling that $\sum_{m=1}^{L} h_m = 0.1k$, we obtain

$$\sum_{\substack{m=1\\h_m\neq 0}}^{L} h_m \log\left(\frac{n}{h_m} e^{-0.4m+1.4}\right) \leqslant \sum_{m=1}^{L} h_m \log\left(\frac{n}{0.1k} \sum_{m=1}^{L} e^{-0.4m+1.4}\right). \tag{41}$$

Since $\sum_{m=1}^{L} e^{-0.4m+1.4} \leqslant 7$, combining (40) with (41) and since $k \leqslant n/14$, we get

$$\log(s_1(\mathbf{k},\mathbf{h})) \leqslant 0.1k \log\left(\frac{7n}{0.1k}\right) \leqslant k \log\left(\frac{n}{7k}\right), \tag{42}$$

where the last inequality follows by a simple calculation.

As for $\log(s_2(\mathbf{k},\mathbf{h}))$,

$$\log(s_2(\mathbf{k},\mathbf{h})) \leqslant \sum_{\substack{m=1\\k_m\neq 0}}^{L} k_m \log\left(\frac{n}{7k} \cdot \frac{n\cdot e}{k_m} e^{-0.4(m-1)} \left(\min\left\{1, \frac{1.1k(0.6m+1)}{0.8n}\right\}\right)^d\right) - k \log\left(\frac{n}{7k}\right). \tag{43}$$

Then, since $\log(s(\mathbf{k},\mathbf{h})) = \log(s_1(\mathbf{k},\mathbf{h})) + \log(s_2(\mathbf{k},\mathbf{h}))$, from (42) and (43),

$$\log(s(\mathbf{k},\mathbf{h})) \leqslant \sum_{\substack{m=1\\k_m\neq 0}}^{L} k_m \log\left(\frac{0.6n^2}{k\cdot k_m} e^{-0.4m} \left(\min\left\{1, \frac{1.1k(0.6m+1)}{0.8n}\right\}\right)^d\right).$$

So, from the above inequality,

$$-\frac{\log(s(\mathbf{k},\mathbf{h}))}{k} \geqslant \sum_{\substack{m=1\\k_m\neq 0}}^{L} \frac{k_m}{k} \log\left(\frac{k_m}{k}\frac{9}{10} \cdot \frac{k^2}{0.6n^2} e^{0.4m} \left(\min\left\{1, \frac{1.1k(0.6m+1)}{0.8n}\right\}\right)^{-d}\right) + \log(10/9) \tag{44}$$

Now, notice that, if we prove that

$$\sum_{\substack{m=1\\k_m\neq 0}}^{L} \frac{k_m}{k} \log\left(\frac{k_m}{k}\frac{9}{10} \cdot \frac{k^2}{0.6n^2} e^{0.4m} \left(\min\left\{1, \frac{1.1k(0.6m+1)}{0.8n}\right\}\right)^{-d}\right) \geqslant 0, \tag{45}$$

then, from (44), we would get (38), since $\log(10/9) \geqslant 0.15$. So, we want to prove (45). Thanks to the *KL divergence inequality* (see Theorem A.4), it is sufficient to show that the following functions are density mass functions over $\{1, 2, \dots, L\}$:

$$p_m = \frac{k_m}{k} \quad \text{and} \quad q_m = \frac{10}{9} \cdot \frac{0.6n^2}{k^2} e^{-0.4m} \min\left\{1, \left(\frac{1.1k(0.6m+1)}{0.8n}\right)^d\right\}.$$

Notice that $\sum_{m=1}^{L} p_m = 1$, and

$$\begin{aligned}\sum_{m=1}^{L} q_m &= \sum_{m=1}^{0.9\frac{n}{k}-1} \frac{10}{9}\frac{0.6n^2}{k^2} e^{-0.4m} \left(\frac{9}{10}\right)^{d-2} \left(\frac{1.1k(0.6m+1)}{0.8n}\right)^2 + \sum_{r=0.9\frac{n}{k}}^{L} \frac{10}{9}\frac{0.6n^2}{k^2} e^{-0.4m} \\ &\leqslant 1.1 \left(\frac{1.1}{0.8}\right)^2 \left(\frac{9}{10}\right)^{d-3} + \frac{10}{9} \cdot \frac{0.6n^2}{k^2} \cdot e^{-0.36\frac{n}{k}} \leqslant 1,\end{aligned}$$

where the last inequality holds taking d large enough ($d \geqslant 30$) and $k \leqslant \frac{n}{14}$. So, we have proved that q_m and p_m are density mass functions over $\{1, 2, \dots, L\}$ and so, thanks to Theorem A.4, (45) holds and implies (39).

Combining (37) with (34) and using (39),

$$\mathbf{Pr}\left(\min_{n/\log^2 \leqslant |S| \leqslant n/14} \frac{|\partial_{out}(S)|}{|S|} \leqslant 0.1\right) \leqslant \sum_{k=n/\log^2 n}^{n/14} \sum_{\substack{k_1+\dots+k_L=k \\ h_1+\dots+h_L=0.1k}} s(\mathbf{k}, \mathbf{h}) + \frac{1}{n^2} \leqslant \frac{2}{n^2},$$

where the last inequality holds since the number of integral sequences $k_1, \dots, k_L$ that sum up k is bounded by $\binom{k+L}{L}$ (and the same holds for h_m). Hence, from simple calculations and recalling that $L = 7n \log n$,

$$\sum_{k=n/\log^2 n}^{n/14} \sum_{\substack{k_1+\dots+k_L=k \\ h_1+\dots+h_L=0.1k}} s(\mathbf{k}, \mathbf{h}) \leqslant \sum_{k=n/\log^2 n}^{n/14} \binom{L+0.1k}{L} \binom{L+k}{L} 2^{-0.15k} \leqslant \frac{1}{n^2}.$$

□

Expansion of small subsets. We next study vertex subsets of size $\mathcal{O}(n/\log^2 n)$. Their number is small enough to get the desired expansion property by apply a standard counting argument combined with our bound on the edge probability in Lemma 5.2 without using the "age profile" argument we required for the analysis of the middle-size subsets.

Lemma 5.5 (Expansion of small subsets)**.** *Under the hypothesis of Theorem 5.3, for subsets S of V_{T_r}, with probability of at least $1 - 2/n^2$,*

$$\min_{0 \leqslant |S| \leqslant n/\log^2 n} \frac{|\partial_{out}(S)|}{|S|} \geqslant 0.1.$$

Proof. We show that two disjoint sets $S, T \subseteq V_t$, with $|S| \leqslant n/\log^2 n$ and $|T| = 0.1|S|$, such that $\partial_{out}(S) \subseteq T$, exist with negligible probability. For each $S, T \subseteq V_{T_r}$, we call such event

$$A_{S,T} = \{\partial_{out}(S) \subseteq T\}.$$

Then, as for the event

$$L_r = \{\text{each node in } V_{T_r} \text{ is born after time } T_{r-7n\log n}\} \cap \{|V_{T_r}| \in [0.9n, 1.1n]\},$$

from Lemma 3.5, Lemma 3.8 and an union bound, we obtain $\mathbf{Pr}(L_r) \geqslant 1 - 1/n^2$. So, for the law of total probability,

$$\mathbf{Pr}\left(\min_{0 \leqslant |S| \leqslant n/\log^2 n} \frac{|\partial_{out}(S)|}{|S|} \leqslant 0.1\right) \leqslant \sum_{\substack{|S| \leqslant n/\log^2 n \\ |T| = 0.1|S|}} \mathbf{Pr}(A_{S,T} \mid L_r) + \frac{1}{n^2}. \tag{46}$$

The next step of the proof is to upper bound $\mathbf{Pr}\left(A_{S,T} \mid L_r\right)$. From Lemma 5.2 and since L_r implies that all the active nodes were born after time $T_{r-7n\log n}$,

$$\mathbf{Pr}\left(A_{S,T} \mid L_r\right) \leqslant \left(\frac{|S \cup T|}{0.8n}\left(1+\frac{7n\log n}{1.7n}\right)\right)^{d|S|} \leqslant \left(\frac{|S \cup T|}{0.8n}\left(1+5n\log n\right)\right)^{d|S|}. \tag{47}$$

Notice that, since $|S| \leqslant n/\log^2 n$, the above equation offers a sufficiently small bound. So, combining (47) with (46), we obtain

$$\mathbf{Pr}\left(\min_{0\leqslant|S|\leqslant n/\log^2 n} \frac{|\partial_{out}(S)|}{|S|} \leqslant 0.1\right) \leqslant \sum_{s=1}^{n/\log^2 n} \binom{1.1n}{s}\binom{1.1n-s}{0.1s}\left(\frac{1.1s}{0.8n}(1+5n\log n)\right)^{ds} + \frac{1}{n^2}. \tag{48}$$

In the equation above, we bounded each binomial coefficient with the inequality $\binom{n}{k} \leqslant \left(\frac{n\cdot e}{k}\right)^k$ for each $k \leqslant n$ and $n \geqslant 2$. Then, by calculating the derivative of the function $f(s)$ that represents each term of the sum, we derive that each of such terms reaches its maximum at the extremes, i.e. in $s = 1$ or in $s = n/\log^2 n$. So, we get that the sum in (48), if $d \geqslant 35$, is bounded by $2/n^2$.

□

Expansion of big subsets. The last case of our analysis of the vertex expansion of the PDGR model considers subsets of big size $|S| \geqslant n/14$. It analysis proceeds exactly as the proof of Lemma 4.3 about the expansion of large subsets in the PDG model. Indeed, in both the PDG and PDGR models, we use the fact that any node $u \in V_{T_r}$ chooses any fixed older node $v \in V_{T_r}$ with probability $\geqslant 1/1.1n$ (thanks to Lemma 3.5). The proof is omitted since it is identical to that of Lemma 4.3.

Lemma 5.6 (Expansion of large subsets)**.** *Under the hypothesis of Theorem 5.3, for subsets S of V_{T_r}, with probability of at least $1 - 2/n^2$,*

$$\min_{n/14\leqslant|S|\leqslant|V_{T_r}|/2} \frac{|\partial_{out}(S)|}{|S|} \geqslant 0.1.$$

5.3 Flooding

In this section, we study the flooding process over the dynamic graph model PDGR with edge regeneration we introduced in Definition 3.2. The vertex expansion property, shown in Theorem 5.3, is here exploited to obtain a logarithmic bound on the time required by this process to inform all the nodes of the graph. Notice that, according to the considered topology dynamics, if there is a time in which all the alive nodes are informed, then every successive snapshot of the dynamic graph will have all its nodes informed as well, w.h.p.

Theorem 5.7 (Flooding)**.** *For every constant $d \geqslant 35$, for every sufficiently large n and for every fixed $r_0 \geqslant 7n\log n$, consider the flooding process in Definition 3.3 over a PDGR sampled from $\mathcal{G}(\lambda, \mu, d)$, with $\lambda = 1$ and $\mu = 1/n$, and starting at $t_0 = T_{r_0}$. Then, w.h.p., the flooding time is $\mathcal{O}(\log n)$.*

As remarked in Section 2, in dynamic networks without node churn, it has already been shown that the good vertex expansion of every snapshot implies fast flooding time (see for instance [6]). In the Poisson models, the presence of random node churn requires to consider some new technical issues. Indeed, once we observe the set of informed nodes I_t at a given snapshot $G_t = (V_t, E_t)$, the expansion of I_t refers to topology E_t while the 1-hop message transmissions take one unit of time. So, during this time interval, some topology changes may take place affecting the expansion observed at time t. To cope with this issue, our analysis splits the process into three consecutive phases and prove they all have logarithmic length, w.h.p.

As remarked in Section 3, the discretized version of the flooding process (Definition 3.3) is always slower than the original one (Definition 3.2), so our proof of Theorem 5.7 will proceed by analyzing the discretized process along three consecutive phases.

Phase 1: The bootstrap. The first phase lasts until the source information reaches a subset of size n^ϵ, for some constant $\epsilon < 1$ (to make some calculations simpler, in our analysis we fix $\epsilon = 0.1$).

Lemma 5.8 (Phase 1: The bootstrap)**.** *Under the hypotheses of Theorem 5.7 there is $\tau_1 = \mathcal{O}(\log n)$ such that, with probability at least $1 - 1/n^{0.6}$,*

$$|I_{t_0+\tau_1}| \geqslant n^{0.1}.$$

Proof. Let t be such that $t \geqslant t_0$ and $|I_t| \leqslant n^{0.1}$. We want to prove that, w.h.p., $|I_{t+1}| \geqslant 0.1|I_t|$. According to Definition 3.3, we write

$$I_{t+1} = (I_t \cup \partial^t_{out}(I_t \cap V_{t+1})) \cap V_{t+1}. \tag{49}$$

We first estimate the probability that $I_t \cap V_{t+1} = I_t$, i.e. the probability that all the nodes in I_t survive for a time interval equal to 1. Since the life of a node follows an exponential distribution with parameter $1/n$ and since $|I_t| \leqslant n^{0.1}$,

$$\mathbf{Pr}\left(I_t = I_t \cap V_{t+1}\right) \geqslant e^{-|I_t|/n} \geqslant 1 - \frac{1}{n^{0.9}}. \tag{50}$$

Then, we need to estimate the following probability

$$\begin{aligned}&\mathbf{Pr}\left(|(I_t \cup \partial^t_{out}(I_t)) \cap V_{t+1}| \geqslant 1.1|I_t|\right) \geqslant \\ &\mathbf{Pr}\left(|(I_t \cup \partial^t_{out}(I_t)) \cap V_{t+1}| \geqslant 1.1|I_t| \;\middle|\; |I_t \cup \partial^t_{out}(I_t)| \geqslant 1.1|I_t|\right) \mathbf{Pr}\left(|I_t \cup \partial^t_{out}(I_t)| \geqslant 1.1|I_t|\right)\end{aligned} \tag{51}$$

From Theorem 5.3,

$$\mathbf{Pr}\left(|I_t \cup \partial^t_{out}(I_t)| \geqslant 1.1|I_t|\right) \geqslant \mathbf{Pr}\left(G_t \text{ is an expander of parameter } 0.1\right) \geqslant 1 - \frac{1}{n}, \tag{52}$$

since, if G_t is an expander of parameter 0.1, $|\partial^t_{out}(I_t)| \geqslant 0.1|I_t|$. Moreover,

$$\mathbf{Pr}\left(|(I_t \cup \partial^t_{out}(I_t)) \cap V_{t+1}| \geqslant 1.1|I_t| \mid |I_t \cup \partial^t_{out}(I_t)| \geqslant 1.1|I_t|\right) \geqslant e^{-1.1|I_t|/n} \geqslant 1 - \frac{1.1}{n^{0.9}}, \tag{53}$$

since this is the probability that $1.1|I_t|$ fixed nodes in $I_t \cup \partial^t_{out}(I_t)$ survive for a time interval equal to 1. So, combining (52) and (53) with (51), we get that, for a sufficiently large n,

$$\mathbf{Pr}\left(|(I_t \cup \partial^t_{out}(I_t)) \cap V_{t+1}| \geqslant 1.1|I_t|\right) \geqslant \left(1 - \frac{1}{n}\right) \cdot \left(1 - \frac{1.1}{n^{0.9}}\right) \geqslant 1 - \frac{1}{n^{0.8}}. \tag{54}$$

Thanks to (49), for any t such that $|I_t| \leqslant n^{0.1}$,

$$\begin{aligned}\mathbf{Pr}\left(|I_{t+1}| \geqslant 1.1|I_t|\right) &= \mathbf{Pr}\left(|(I_t \cup \partial^t_{out}(I_t)) \cap V_{t+1}| \geqslant 1.1|I_t| \text{ and } I_t = I_t \cap V_{t+1}\right) \\ &\geqslant \mathbf{Pr}\left(|(I_t \cup \partial^t_{out}(I_t)) \cap V_{t+1}| \geqslant 1.1|I_t|\right) + \mathbf{Pr}\left(I_t = I_t \cap V_{t+1}\right) - 1 \\ &\geqslant 1 - \frac{1}{n^{0.8}} - \frac{1}{n^{0.9}} \geqslant 1 - \frac{1}{n^{0.7}},\end{aligned} \tag{55}$$

where the first inequality follows from the fact that, for any two events A, B, it holds $\mathbf{Pr}\left(A \cap B\right) \geqslant \mathbf{Pr}\left(A\right) + \mathbf{Pr}\left(B\right) - 1$ and the second inequality follows from (50) and (54).

We can then easily conclude from (55) and from the chain rule that, after a phase of length $\tau_1 = \mathcal{O}(\log n)$ rounds, we get $|I_{t_0+\tau_1}| \geqslant n^{0.1}$, with probability at least $1 - 1/n^{0.6}$. □

Phase 2: Exponential growth of the informed nodes. In the next lemma we show that, after the bootstrap, the flooding process yields an exponential increase of the number of informed nodes until it reaches half of the nodes in the network.

Lemma 5.9 (Phase 2)**.** *Under the hypotheses of Theorem 5.7 and assuming the claim of Lemma 5.8 holds for some $\tau_1 = \mathcal{O}(\log n)$, with probability at least $1 - 1/n^{0.8}$, a time $\tau_2 = \mathcal{O}(\log n)$ exists such that*

$$|I_{t_0+\tau_1+\tau_2}| \geqslant \frac{1.1n}{2}.$$

Proof. Let $t \geqslant t_0 + \tau_1$ be such that such that $n^{0.1} \leqslant |I_t| \leqslant 1.1n/2$. We want to prove that for such t we have that $|I_{t+1}| \geqslant 1.09|I_t|$ with probability at least $1 - 1/n^{0.9}$. Observe first that, from claim 2 of Lemma 3.9, in any time interval of unit length, at most $4\log n$ nodes leave the network, with probability at least $1 - 1/n$. If $D_{t,t+1}$ is the number of nodes that leave the network in the time interval $[t, t+1]$,

$$\mathbf{Pr}\left(D_{t,t+1} \leqslant 4\log n\right) \geqslant 1 - \frac{1}{n} \tag{56}$$

We recall that from Definition 3.3 the set of infected nodes at time $t+1$ is

$$I_{t+1} = (I_t \cup \partial^t_{out}(I_t \cap V_{t+1})) \cap V_{t+1}.$$

In Theorem 5.3, we have shown that, with probability at least $1 - 1/n$, the graph G_t is such that each subset of nodes S such that $|S| \leqslant 1.1n/2$ satisfies $|\partial_{out}(S)| \geqslant 0.1|S|$. So, Theorem 5.3 with (56) imply

$$\begin{aligned}&\mathbf{Pr}\left(|I_{t+1}| \geqslant 1.1|I_t| - 3\log n\right) \geqslant \mathbf{Pr}\left(|I_{t+1}| \geqslant |I_t| + 0.1\left(|I_t| - 4\log n\right) - 4\log n\right)\\ &\geqslant \mathbf{Pr}\left(\{G_t \text{ is an expander of parameter } 0.1\} \cap \{D_{t,t+1} \leqslant 4\log n\}\right) \geqslant 1 - \frac{2}{n} \geqslant 1 - \frac{1}{n^{0.9}}.\end{aligned}$$

So, since for each $t \geqslant t_0 + \tau_1$ we have that $|I_{t_0+\tau_1}| \geqslant n^{0.1}$, with probability at least $1 - 1/n^{0.9}$ we have that

$$|I_{t+1}| \geqslant 1.09|I_t|.$$

We thus have an exponential growth of the set of the informed nodes and, for the chain rule, there exists a time $\tau_2 = O(\log n)$ such that $|I_{t_0+\tau_1+\tau_2}| \geqslant 1.1n/2$, with probability at least $1 - 1/n^{0.8}$ for large enough n. □

Phase 3: Exponential decrease of the non-informed nodes. The analysis of this phase considers the subset $S_t \subseteq V_t$ of the non-informed nodes. More precisely, we prove that S_{t+1} w.h.p. decreases by a constant factor despite the node churn.

Lemma 5.10 (Phase 3)**.** *Under the hypotheses of Theorem 5.7, assume the claims of Lemma 5.8 and of Lemma 5.9 hold for some $\tau_1 = \mathcal{O}(\log n)$ and $\tau_2 = \mathcal{O}(\log n)$, respectively. Then, with probability at least $1 - 1/n^{1/3}$, there is a time $\tau_3 = \mathcal{O}(\log n)$ such that*

$$I_{t_0+\tau_1+\tau_2+\tau_3} = V_{t_0+\tau_1+\tau_2+\tau_3}.$$

Proof. For any $t \geqslant t_0 + \tau_1 + \tau_2$, we now consider the set $S_{t+1} \subseteq V_{t+1}$ of non informed nodes at time $t+1$, i.e. $S_{t+1} = V_{t+1} - I_{t+1}$. We recall that the set of informed nodes at time $t+1$ is

$$I_{t+1} = (I_t \cup \partial^t_{out}(I_t \cap V_{t+1})) \cap V_{t+1}.$$

It follows that, since every node v in $\partial^{t+1}_{out}(S_{t+1} \cap V_t)$ is reachable in 1-hop by the set of non-informed nodes S_t at time $t+1$, v was not informed at time t. This implies that

$$\partial^{t+1}_{out}(S_{t+1} \cap V_t) \subseteq S_t - (S_{t+1} \cap V_t).$$

Since $t \geqslant t_0 + \tau_1 + \tau_2$, we have that $|S_t| \leqslant |V_t| - |I_t| \leqslant 1.1n/2$ and, since $S_{t+1} \cap V_t \subseteq S_t$, for Theorem 5.3, we have that $|\partial_{out}(S_{t+1} \cap V_t)| \geqslant 0.1|S_{t+1} \cap V_t|$ with probability at least $1 - 1/n$. Consider the random variable $J_{t,t+1}$ that indicates the number of nodes that join the network in the time interval $[t, t+1]$. Then, the above considerations imply that with probability at least $1 - 1/n$

$$|S_{t+1}| \leqslant \frac{1}{1.1}|S_t| + J_{t,t+1}.$$

Claim 1 of Lemma 3.9 imply that $J_{t,t+1} \leqslant 4\log n$ with probability at least $1 - 1/n$. So, with probability at least $1 - 2/n$, we will have

$$|S_{t+1}| \leqslant 0.91|S_t| + 4\log n.$$

Moreover, if t is such that $\log^2 n \leqslant |S_t| \leqslant 1.1n/2$, we have that for large enough n, with probability at least $1 - 2/n$,

$$|S_{t+1}| \leqslant 0.95|S_t|.$$

This implies that with probability at least $1 - 1/n^{1/2}$, there exists a $\tau_3' = O(\log n)$ such that $|S_{t_0+\tau_1+\tau_2+\tau_3'}| \leqslant \log^2 n$.
After the process reaches a state with the above small number of non-informed nodes, we consider the set of non-informed nodes at time t without including the set of nodes that join the network after time $t_0 + \tau_1 + \tau_2 + \tau_3'$: we call the latter set as S^*_t, for each $t \geqslant t_0 + \tau_1 + \tau_2 + \tau_3'$. Similar to the first part of the proof, we get that $\partial^{t+1}_{out}(S^*_{t+1}) \subseteq S^*_t - S^*_{t+1}$. From Theorem 5.3, it holds, with probability at least $1 - 1/n$, $|\partial_{out}(S^*_{t+1})| \geqslant |S^*_{t+1}|$, which implies

$$|S^*_{t+1}| \leqslant 0.91|S^*_t|.$$

Since $|S^*_{t_0+\tau_1+\tau_2+\tau_3'}| \leqslant \log^2 n$, there is a $\tau_3 = \mathcal{O}(\log n)$ such that, $|S^*_{t_0+\tau_1+\tau_2+\tau_3}| < 1$ with probability at least $1 - 1/n^{1/2}$.
In conclusion, let J_{τ_3',τ_3} be the number of nodes that join the network from time $t_0 + \tau_1 + \tau_2 + \tau_3'$ to time $t_0 + \tau_1 + \tau_2 + \tau_3$. Since the arrival of the nodes during an interval of length $\tau_3 - \tau_3'$ is a Poisson process of mean $\tau_3 - \tau_3'$, for the tail bound for the Poisson distribution (Theorem B.4), w.h.p. $J_{\tau_3',\tau_3} = \mathcal{O}(\log n)$. Moreover, from Lemma 3.5, each of these new nodes connect to the set of informed nodes with probability at least $(1 - (2\log^2 n/n)^d)(1 - 1/n^2)$. Moreover, each informed node to which the new nodes have connected survive for the 1-hop transmission with probability $e^{-1/n} \geqslant 1 - \frac{1}{n}$. So, each node that joins the network after time $t_0 + \tau_1 + \tau_2 + \tau_3'$ gets informed within time $t_0 + \tau_1 + \tau_2 + \tau_3$, with probability at least $1 - 1/n^{1/2}$. We conclude that there exists a time $\tau_3 = O(\log n)$ such that $|I_{t_0+\tau_1+\tau_2+\tau_3}| = |V_{t_0+\tau_1+\tau_2+\tau_3}|$ with probability at least $1 - 1/n^{1/3}$. □

6 Overall remarks and open questions

We studied two models of fully-random dynamic networks with node churns. We analysed their expansion properties and gave results about the performances of the flooding process. We essentially show that such important aspects depend on the specific adopted topology dynamic, namely, on whether or not, edge regeneration takes place along the time process.

We remark that the Poisson model with edge regeneration bears a certain similarity to the way peer-to-peer networks such as Bitcoin are formed. In particular, although the random choices over the current node set V_t the nodes make to establish connections is not the connection mechanism adopted in standard Bitcoin implementations, the set of IP addresses of the active full-nodes of the Bitcoin network can be easily discovered by a crawler (see, e.g., [24]). This implies that, potentially, nodes can implement a good approximation of the fully-random strategy by picking random elements from such on-line table.

We see an interesting future research direction related to our work. The topology dynamics we considered yield sparse graphs at every round, however, the maximum node degree can be of magnitude $\mathcal{O}(\log n)$. For some real applications this bound is too large, and finding natural, fully-random topology dynamics that yield bounded-degree snapshots of good expansion properties is a challenging issue which has strong theoretical and practical motivations [25, 1, 17].

References

[1] John Augustine, Gopal Pandurangan, and Peter Robinson. Distributed algorithmic foundations of dynamic networks. *ACM SIGACT News*, 47(1):69–98, 2016.

[2] John Augustine, Gopal Pandurangan, Peter Robinson, Scott T. Roche, and Eli Upfal. Enabling robust and efficient distributed computation in dynamic peer-to-peer networks. In Venkatesan Guruswami, editor, *IEEE 56th Annual Symposium on Foundations of Computer Science, FOCS 2015, Berkeley, CA, USA, 17-20 October, 2015*, pages 350–369. IEEE Computer Society, 2015.

[3] John Augustine, Gopal Pandurangan, Peter Robinson, and Eli Upfal. Towards robust and efficient computation in dynamic peer-to-peer networks. In Yuval Rabani, editor, *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2012, Kyoto, Japan, January 17-19, 2012*, pages 551–569. SIAM, 2012.

[4] Baruch Awerbuch and Christian Scheideler. The hyperring: a low-congestion deterministic data structure for distributed environments. In J. Ian Munro, editor, *Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2004, New Orleans, Louisiana, USA, January 11-14, 2004*, pages 318–327. SIAM, 2004.

[5] Andrea E. F. Clementi, Claudio Macci, Angelo Monti, Francesco Pasquale, and Riccardo Silvestri. Flooding time of edge-markovian evolving graphs. *SIAM J. Discret. Math.*, 24(4):1694–1712, 2010.

[6] Andrea E. F. Clementi, Angelo Monti, Francesco Pasquale, and Riccardo Silvestri. Information spreading in stationary markovian evolving graphs. *IEEE Trans. Parallel Distrib. Syst.*, 22(9):1425–1432, 2011.

[7] Colin Cooper, Martin E. Dyer, and Catherine S. Greenhill. Sampling regular graphs and a peer-to-peer network. *Comb. Probab. Comput.*, 16(4):557–593, 2007.

[8] Bitcoin Core. Bitcoin Core 0.11 (ch 4): P2P Network. https://en.bitcoin.it/wiki/Bitcoin_Core_0.11_(ch_4):_P2P_Network. Accessed: 2020-07-15.

[9] Sergi Delgado-Segura, Surya Bakshi, Cristina Pérez-Solà, James Litton, Andrew Pachulski, Andrew Miller, and Bobby Bhattacharjee. Txprobe: Discovering bitcoin's network topology using orphan transactions. In Ian Goldberg and Tyler Moore, editors, *Financial Cryptography and Data Security - 23rd International Conference, FC 2019, Frigate Bay, St. Kitts and Nevis, February 18-22, 2019, Revised Selected Papers*, volume 11598 of *Lecture Notes in Computer Science*, pages 550–566. Springer, 2019.

[10] Devdatt Dubhashi and Alessandro Panconesi. *Concentration of Measure for the Analysis of Randomized Algorithms*. Cambridge University Press, USA, 1st edition, 2009.

[11] Philippe Duchon and Romaric Duvignau. Local update algorithms for random graphs. In Alberto Pardo and Alfredo Viola, editors, *LATIN 2014: Theoretical Informatics - 11th Latin American Symposium, Montevideo, Uruguay, March 31 - April 4, 2014. Proceedings*, volume 8392 of *Lecture Notes in Computer Science*, pages 367–378. Springer, 2014.

[12] Riko Jacob, Andrea Richa, Christian Scheideler, Stefan Schmid, and Hanjo Täubig. A distributed polylogarithmic time algorithm for self-stabilizing skip graphs. In *Proceedings of the 28th ACM Symposium on Principles of Distributed Computing*, PODC '09, page 131–140, New York, NY, USA, 2009. Association for Computing Machinery.

[13] Fabian Kuhn, Nancy Lynch, and Rotem Oshman. Distributed computation in dynamic networks. In *Proceedings of the Forty-Second ACM Symposium on Theory of Computing*, STOC '10, page 513–522, New York, NY, USA, 2010. Association for Computing Machinery.

[14] Fabian Kuhn and Rotem Oshman. Dynamic networks: Models and algorithms. *SIGACT News*, 42(1):82–96, March 2011.

[15] Ching Law and K-Y Siu. Distributed construction of random expander networks. In *Proceedings of the 22nd Annual Joint Conference of the IEEE Computer and Communications (INFOCOM 2003)*, volume 3, pages 2133–2143. IEEE, 2003.

[16] Peter Mahlmann and Christian Schindelhauer. Peer-to-peer networks based on random transformations of connected regular undirected graphs. In *Proceedings of the Seventeenth Annual ACM Symposium on Parallelism in Algorithms and Architectures*, SPAA '05, page 155–164, New York, NY, USA, 2005. Association for Computing Machinery.

[17] Yifan Mao, Soubhik Deb, Shaileshh Bojja Venkatakrishnan, Sreeram Kannan, and Kannan Srinivasan. Perigee: Efficient peer-to-peer network design for blockchains. In *Proc. of the ACM Symposium on Principles of Distributed Computing, to appear*, 2020. Technical report available at https://arxiv.org/abs/2006.14186.

[18] Othon Michail. An introduction to temporal graphs: An algorithmic perspective. *Internet Mathematics*, 12(4):239–280, 2016.

[19] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system. https://bitcoin.org/bitcoin.pdf, 2008.

[20] Till Neudecker, Philipp Andelfinger, and Hannes Hartenstein. Timing analysis for inferring the topology of the bitcoin peer-to-peer network. In *2016 Intl IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld), Toulouse, France, July 18-21, 2016*, pages 358–367. IEEE Computer Society, 2016.

[21] James R. Norris. *Markov Chains*. Cambridge University Press, 1998.

[22] Gopal Pandurangan, Prabhakar Raghavan, and Eli Upfal. Building low-diameter peer-to-peer networks. *IEEE J. Sel. Areas Commun.*, 21(6):995–1002, 2003.

[23] Gopal Pandurangan and Amitabh Trehan. Xheal: A localized self-healing algorithm using expanders. *Distrib. Comput.*, 27(1):39–54, February 2014.

[24] Addy Yeow. Global Bitcoin Nodes Distribution. https://bitnodes.io/. Accessed: 2020-07-15.

[25] Zeyuan Allen Zhu, Aditya Bhaskara, Silvio Lattanzi, Vahab S. Mirrokni, and Lorenzo Orecchia. Expanders via local edge flips. In Robert Krauthgamer, editor, *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016*, pages 259–269. SIAM, 2016.

Appendix

A Mathematical tools

Theorem A.1 (Chernoff Bound, [10]). *Let $X_1, \ldots, X_n$ be independent Poisson trials such that* $\mathbf{Pr}(X_i = 1) = p_i$. *Let $X = \sum_{i=1}^n X_i$, $\mu = \mathbf{E}[X]$ and suppose $\mu_L \leqslant \mu \leqslant \mu_H$. Then, for all $0 < \varepsilon \leqslant 1$ and $t \geqslant 0$ the following bounds hold*

$$\mathbf{Pr}(X \geqslant (1+\varepsilon)\mu_H) \leqslant e^{-\frac{\varepsilon^2}{3}\mu_H}$$

$$\mathbf{Pr}(X \leqslant (1-\varepsilon)\mu_L) \leqslant e^{-\frac{\varepsilon^2}{2}\mu_L}$$

$$\mathbf{Pr}(X \geqslant \mathbf{E}[X] + t), \mathbf{Pr}(X \leqslant \mathbf{E}[X] - t) \leqslant e^{-2\frac{t^2}{n}}.$$

Theorem A.2 (Method of bounded differences,[10]). *Let $\mathbf{Y} = (Y_1, \ldots, Y_m)$ be independent random variables, with Y_j taking values in the set A_j. Suppose the real-valued function f defined on $\prod_j A_j$ satisfies the Lipschitz condition with coefficients β_j, i.e.*

$$|f(\mathbf{y}) - f(\mathbf{y}')| \leqslant \beta_j$$

whenever vectors $\mathbf{y}$ $\mathbf{y}'$ differs only in the j-th coordinate. Then, for any $M > 0$

$$\mathbf{Pr}(f(\mathbf{Y}) \geqslant \mathbf{E}[f(\mathbf{Y})] + M) \leqslant e^{-\frac{2M^2}{\sum_{j=1}^m \beta_j^2}},$$

and

$$\mathbf{Pr}(f(\mathbf{Y}) \leqslant \mathbf{E}[f(\mathbf{Y})] - M) \leqslant e^{-\frac{2M^2}{\sum_{j=1}^m \beta_j^2}}.$$

Theorem A.3 (Jensen's inequality). *For any real concave function φ, numbers $x_1, \ldots, x_L$ in its domain and positive weights $a_1, \ldots, a_L$, the following holds:*

$$\frac{\sum_{m=1}^L a_m \varphi(x_m)}{\sum_{m=1}^L a_m} \leqslant \varphi\left(\frac{\sum_{m=1}^L a_m x_m}{\sum_{m=1}^L a_m}\right).$$

Theorem A.4 (Kullback-Leibler divergence inequality). *Let p_m and q_m be two discrete probability mass functions, with $m \in \{1, \ldots, L\}$. We have that*

$$\sum_{r=1}^L p_m \log_2\left(\frac{p_m}{q_m}\right) \geqslant 0.$$

B Useful tools for Poisson processes

Definition B.1 (Counting process). *The stochastic process $\{X(t), t \geqslant 0\}$ is said to be a* counting process *if $X(t)$ represents the total number of events that occurred up to time t.*

Definition B.2. *Let $\{X(t), t \geqslant 0\}$ be a counting process. It is a* Poisson process *if*

1. *$X(0) = 0$;*
2. *$X(t)$ has independent increments;*
3. *the number of events in any interval of length t has a Poisson distribution with mean λt. That is, for all $s, t \geqslant 0$,*

$$\boldsymbol{Pr}(X(t+s) - X(s) = n) = e^{-\lambda t}\frac{(\lambda t)^n}{n!}, \quad n \geqslant 0.$$

Theorem B.3. *Let $\{X(t), t \geqslant 0\}$ be a Poisson process. Then, given $X(t) = n$, the n arrival times $S_1, \ldots, S_n$ have the same distribution as the order statistics corresponding to n independent random variables uniformly distributed in the interval $(0, t)$.*

Theorem B.4 (Tail bound for the Poisson distribution)**.** *Let X have a Poisson distribution with mean λ. Then, for each $\varepsilon > 0$,*

$$\mathbf{Pr}\left(|X - \lambda| \geqslant x\right) \leqslant 2e^{-\frac{x^2}{2(\lambda+x)}}$$

Theorem B.5 ([21])**.** *Let I be a countable set and let T_k, $k \in I$, be independent exponential random variables of parameter q_k. Let $0 < q = \sum_{k \in I} q_k \leqslant \infty$. Set $T = \inf_k T_k$. Then this infimum is attained at a unique random value K of k, with probability 1. Moreover, T and K are independent with T exponential of parameter q and $\mathbf{Pr}(K = k) = q_k/q$.*

C Omitted proofs

C.1 Static random graphs

Lemma C.1. *The static random graph in which each node has d neighbors, each chosen uniformly at random and independently of the others, is a $\Theta(1)$-expander w.h.p., for each $d \geqslant 3$.*

Proof. Consider the static random graph $G = (N, E)$, where E is generated randomly as stated in the claim of the Lemma. Let $S \subset N$ a subset of the nodes, with $|S| = s$ and $T \subseteq N - S$ a second arbitrary subset of the nodes, disjoint from S, with $|T| = 0.1s$. We know that in this model, an edge originating from a node v has destination u with probability $\frac{1}{n-1}$. Since all edges are established independently from each other, the probability that all edges originating from nodes in S have endpoints $S \cup T$ is

$$\left(\frac{|S \cup T|}{n-1}\right)^{ds}.$$

So, the probability that the outer boundary of S is all in T is

$$\mathbf{Pr}\left(\partial_{out}(S) \subseteq T\right) \leqslant \left(\frac{1.1s}{n-1}\right)^{ds}.$$

From a union bound over all sets T disjoint from S and with $|T| = 0.1s$, all sets S with s elements and all possible sizes $s = 1, \ldots, n/2$ of s we obtain:

$$\mathbf{Pr}\left(G \text{ is not an expander}\right) \leqslant \sum_{s=1}^{n/2} \binom{n}{s} \binom{n-s}{0.1s} \left(\frac{1.1s}{n-1}\right)^{ds}.$$

Standard calculus allows proving that the r.h.s. of the inequality above is upper bounded by $1/n^{d-2}$ for $d \geqslant 3$. To prove this, we use $\binom{n}{s} \leqslant \left(\frac{n \cdot e}{s}\right)^s$ and we compute the derivative of the function $f(s)$ that corresponds to each term in the resulting sum, obtaining that its maximum is attained when $s = 1$ or $s = n/2$. □

C.2 Jump processes

Thanks to Theorem B.5, we can easily analyze the random variables corresponding to times at which new events occur in the Poisson node churn model.

Lemma C.2 (Jump process)**.** *Consider, the discrete Markov chain $\{V_{T_r}, r \in \mathbb{N}\}$ from Definition 3.6. For every fixed integer $r \geqslant 0$ and for every integer $N \geqslant 0$, conditioned to the event "$|V_{T_r}| = N$", T_{r+1} is a random variable with exponential distribution and parameter $N\mu + \lambda$. Moreover,*

$$\mathbf{Pr}\left(|V_{T_{r+1}}| = |V_{T_r}| - 1 \;\middle|\; |V_{T_r}| = N\right) = \frac{N\mu}{N\mu + \lambda}, \tag{57}$$

$$\mathbf{Pr}\left(|V_{T_{r+1}}| = |V_{T_r}| + 1 \;\middle|\; |V_{T_r}| = N\right) = \frac{\lambda}{N\mu + \lambda}. \tag{58}$$

Finally, for every fixed node $v \in V_{T_r}$, the probability that the next event is v leaving the network is

$$\mathbf{Pr}\left(v \notin V_{T_{r+1}} \;\middle|\; v \in V_{T_r}, |V_{T_r}| = N\right) = \frac{\mu}{N\mu + \lambda}. \tag{59}$$

Proof. Conditional to the event "$|V_{T_r}| = N$", we consider the N random variables with exponential distribution and parameter μ that represent the lifetimes of nodes in the network at time T_r. We further consider the random variable with exponential distribution and with parameter λ that represents the arrival time of the next node entering the network. We notice that i) these $N+1$ random variables are independent and ii) T_{r+1} is the minimum among their values. Hence, Theorem B.5 implies that T_{r+1} has an exponential distribution with parameter $N\mu + \lambda$. Moreover, conditional to "$|V_{T_r}| = N$", the event "$|V_{T_{r+1}}| = |V_{T_r}| - 1$" is the event in which the minimum T_{r+1} is attained by one of the N random variables representing the lifetimes of the nodes that are in the system at T_r. This and Theorem B.5 imply (57). A similar argument leads to (58) and (59). □

C.3 Proof of Lemma 3.7

The lemma easily follows from Lemma C.2 and from concentration of the number of nodes as stated in Lemma 3.5. To begin, for each $r \geqslant n \log n$, we define the following event:

$$C_r = \{|V_{T_r}| \in [0.9n, 1.1n]\}.$$

Lemma 3.5 immediately implies $\mathbf{Pr}(C_r) \geqslant 1 - 1/n^2$. We next show $\mathbf{Pr}\left(|V_{T_{r+1}}| = |V_{T_r}| + 1\right) \leqslant 0.53$ in (1). For each $r \geqslant n \log n$, we write:

$$\begin{aligned} &\mathbf{Pr}\left(|V_{T_{r+1}}| = |V_{T_r}| - 1\right) \\ &= \mathbf{Pr}\left(|V_{T_{r+1}}| = |V_{T_r}| - 1 \;\middle|\; C_r\right) \mathbf{Pr}(C_r) + \mathbf{Pr}\left(|V_{T_{r+1}}| = |V_{T_r}| - 1 \;\middle|\; C_r^C\right) \mathbf{Pr}\left(C_r^C\right). \end{aligned}$$

The equation above and the law of total probability in turn imply:

$$\mathbf{Pr}\left(|V_{T_{r+1}}| = |V_{T_r}| - 1\right) \leqslant \sum_{N=0.9n}^{1.1n} \mathbf{Pr}\left(|V_{T_{r+1}}| = |V_{T_r}| - 1 \;\middle|\; |V_{T_r}| = N\right) \mathbf{Pr}\left(|V_{T_r}| = N \mid C_{T_r}\right) + \frac{1}{n^2} \tag{60}$$

From Lemma C.2, with $\lambda = 1$ and $\mu = 1/n$, we also have:

$$\mathbf{Pr}\left(|V_{T_{r+1}}| = |V_{T_r}| - 1 \;\middle|\; |V_{T_r}| = N\right) = \frac{N/n}{N/n + 1}. \tag{61}$$

Hence, combining (61) and (60),

$$\mathbf{Pr}\left(|V_{T_{r+1}}| = |V_{T_r}| - 1\right) \leqslant \sum_{N=0.9n}^{1.1n} \frac{N}{N+n} \mathbf{Pr}\left(|V_{T_r}| = N \mid C_{T_r}\right) + \frac{1}{n^4} \leqslant 0.53\,.$$

To show $0.47 \leqslant \mathbf{Pr}\left(|V_{T_{r+1}}| = |V_{T_r}| + 1\right)$ in (2), we can use the inequality above:

$$\mathbf{Pr}\left(|V_{T_{r+1}}| = |V_{T_r}| + 1\right) = 1 - \mathbf{Pr}\left(|V_{T_{r+1}}| = |V_{T_r}| - 1\right) \geqslant 1 - 0.53 = 0.47.$$

Following the very same line of argument, we prove the lower bound for $\mathbf{Pr}\left(|V_{T_{r+1}}| = |V_{T_r}| - 1\right)$ in (1), the upper bound for $\mathbf{Pr}\left(|V_{T_{r+1}}| = |V_{T_r}| + 1\right)$ in (2), and the two inequalities in (3) for $\mathbf{Pr}\left(v \notin V_{T_{r+1}} \;\middle|\; v \in V_{T_r}\right)$.

C.4 Proof of Lemma 3.8

Proof. Let $r \geqslant 7n \log n$. Denote by C_{T_r} the event $\{|V_{T_r}| \in [0.9n, 1.1n]\}$. Lemma 3.5 implies $\mathbf{Pr}(C_{T_r}) \geqslant 1 - 1/n^2$. Furthermore, from Lemma 3.7 and the memoryless property of the exponential distribution:

$$\mathbf{Pr}\left(v \in V_{T_r} \;\middle|\; v \in V_{T_{r-7n\log n}}\right) \leqslant \left(1 - \frac{1}{2.2n}\right)^{7n \log n} \leqslant e^{-3.1 \log n} = \frac{1}{n^{3.1}}. \tag{62}$$

We next show that every node in V_{T_r} joined the network after time $T_{r-7n\log n}$, with high probability. To this purpose, we condition on the event C_{T_r}, which holds with high probability has shown above.

$$\begin{aligned} &\mathbf{Pr}\left(\text{there exists a node } v \in V_{T_r} \text{ born before } T_{r-7n\log n}\right) \\ &\leqslant \mathbf{Pr}\left(\text{there exists a node } v \in V_{T_r} \text{ born before } T_{r-7n\log n} \mid C_{T_r}\right) + \frac{1}{n^2} \end{aligned}$$

and, since C_{T_r} guarantees that the nodes in the network at time T_r are at most $1.1n$, from equation (62) we get the lemma. □

C.5 Proof of Lemma 3.9

Proof. As for the first claim, consider the random variable $J_{t_0,t_0+\log n}$ that indicates the number of nodes that join the network in the time interval $[t_0, t_0 + \log n]$. Definitions 3.4 and B.2 (Poisson Process) imply that $J_{t_0,t_0+\log n}$ is a Poisson random variable with mean $\log n$. The tail bound for the Poisson distribution (Theorem B.4) than yields:

$$\mathbf{Pr}\left(J_{t_0,t_0+\log n} \leqslant 4\log n\right) \geqslant 1 - \frac{1}{n}.$$

As for the second claim, we indicate with L the number of nodes that have left the network in the time interval $[t_0, t_0 + \log n]$. Consider the generic node $i \in V_{t_0}$ and let $L_i(\tau)$ be the binary random variable such that $L_i(\tau) = 1$ if i is not alive at time $t_0 + \tau$ and $L_i(\tau) = 0$ otherwise. Clearly,

$$L = \sum_{i \in V_{t_0}} L_i(\log n).$$

Since the death process follows the exponential distribution with parameter $\mu = 1/n$,

$$\mathbf{Pr}\left(L_i(\log n) = 1\right) = 1 - e^{-\frac{\log n}{n}} \leqslant \frac{\log n}{n}.$$

Since each node leaves the network independently of the other nodes, from the Chernoff's bound (Theorem A.1)

$$\mathbf{Pr}\left(L \geqslant 4\log n \mid |V_{t_0}| \in [0.9n, 1.1n]\right) \leqslant \frac{1}{n^2},$$

and the thesis follows from Bayes' rule and from Lemma 3.5, that guarantees that $|V_{t_0}| \in [0.9n, 1.1n]$ with probability at least $1 - 1/n^2$, when $t_0 = T_{r_0}$ with $r_0 \geqslant 7n\log n$. □