J. Parallel Distrib. Comput. 74 (2014) 2433–2437

Contents lists available at ScienceDirect

J. Parallel Distrib. Comput.

journal homepage: www.elsevier.com/locate/jpdc

Research note

Flooding in dynamic graphs with arbitrary degree sequence

Hervé Baumann [a], Pierluigi Crescenzi [b,*], Pierre Fraigniaud [a]

[a] CNRS and University of Paris Diderot, France
[b] Università degli Studi di Firenze, Italy

HIGHLIGHTS

- We study the flooding time in dynamic random graphs with arbitrary degree sequence.
- In the case of power-law degree sequences, the flooding time is almost surely $\log(n)$.
- In the general case, upper bounds depend on specific properties of the sequence.

ARTICLE INFO

Article history:
Received 28 February 2013
Received in revised form
21 December 2013
Accepted 17 January 2014
Available online 30 January 2014

Keywords:
Flooding protocol
Dynamic graphs
Power-law graphs
Distributed computing

ABSTRACT

This paper addresses the flooding problem in dynamic graphs, where flooding is the basic mechanism in which every node becoming aware of a piece of information at step t forwards this information to all its neighbors at all forthcoming steps $t' > t$. We show that a technique developed in a previous paper, for analyzing flooding in a Markovian sequence of Erdös–Rényi graphs, is robust enough to be used also in different contexts. We establish this fact by analyzing flooding in a sequence of graphs drawn independently at random according to a model of random graphs with given expected degree sequence. In the prominent case of power-law degree distributions, we prove that flooding takes almost surely $O(\log n)$ steps even if, almost surely, none of the graphs in the sequence is connected. In the general case of graphs with an arbitrary degree sequence, we prove several upper bounds on the flooding time, which depend on specific properties of the degree sequence.

1. Introduction

One of the basic communication tasks that has been extensively studied in the context of communication networks is the *broadcast* operation: one distinguished node of the network (the source node) aims at sending a message to all the other nodes of the network. The simplest communication process that implements such an operation is the *flooding* mechanism, according to which (1) the source node is initially informed, and (2) when a node v, which is not informed yet, has an informed neighbor, then v becomes informed itself at the next time step. The question is how long does it take to get all the nodes informed, i.e. what is the speed of information spreading? Clearly, this question has an easy answer in the case of *static* networks: the flooding worst-case completion time is just equal to the diameter of the network. In a *dynamic* network, however, nodes and edges can appear and disappear over time (several emerging networking technologies such as ad hoc wireless, sensor, mobile networks, and peer-to-peer networks are inherently dynamic). In dynamic networks, it is actually not even clear whether the flooding completion time is finite. In order to investigate such an issue, the concept of evolving graph has been introduced in the literature. An *evolving graph* is a sequence of graphs $(G_t)_{t\geq 0}$ with the same set of nodes but potentially different set of edges. As far as we know, this definition has been formally introduced for the first time by Ferreira [13]. This concept is clearly general enough to enable modeling essentially any kind of dynamicity, ranging from *adversarial* evolving graphs (see, for example, [10]) to *random* evolving graphs (see, for example, [4,14]). In this latter case, the *edge-Markovian* model has been recently studied with respect to the flooding completion time. In particular, Clementi et al. [11] show some tight upper bounds on the flooding completion time in the case of edge-Markovian evolving graphs, while Baumann et al. [2] extended and adapted these results to the case in which every node forwards the message only for a limited number of time steps after message reception. In order to prove these latter results, the authors make use of the so-called *reduction lemma*, which intuitively shows that the flooding completion time of an edge-Markovian evolving graph is equally

* Corresponding author.
E-mail address: pierluigi.crescenzi@unifi.it (P. Crescenzi).

http://dx.doi.org/10.1016/j.jpdc.2014.01.007
0743-7315/

distributed as the diameter of a suitably defined weighted random graph. As a consequence of this lemma, the analysis of the flooding time reduces to the analysis of the connectivity properties of such a random weighted graph. In this research note, we show how the reduction lemma can be applied to the analysis of the flooding completion time in the case of another interesting kind of random evolving graphs in which, at each time step $t \geq 0$, the graph G_t is chosen uniformly at random in a family of graphs with an arbitrary degree sequence. Indeed, even if the Erdös–Rényi random graphs [12] share many similar aspects with large scale real-world graphs, it has been repeatedly observed that there are several significant differences between these two families of graphs. One of the most well-known differences concerns the fact that, in "real" networks, there are few edges, and there can be vertices with very large degrees (those networks are often called *power-law* graphs). Instead, all nodes in the Erdös–Rényi random graphs basically have the same degree. Hence, a general random graph model has been introduced in the literature for producing random graphs with arbitrary given sequence of expected degrees (for a definitive reference to this model, see [7]). We adopt this model in this paper.

Our results: by applying the technique of Baumann et al. [2], we first show that the flooding completion time of a random evolving graph $(G_t)_{t\geq 0}$ is bounded by $kD + 2C$, where, intuitively: (1) k is the smallest time necessary for the appearance of a giant component in each random graph; (2) D is the diameter of the giant component; and (3) C is the time required for the nodes outside the giant component to eventually get an edge connecting them to the giant component. Based on this result, we develop a general methodology for analyzing flooding in sequences of random graphs. We apply this methodology to the case of power-law evolving graphs (that is, sequences of mutually independent random graphs such that the number y of nodes of degree x distributes like $1/x^\beta$ for some $\beta > 0$), and to the case of an arbitrary given degree sequence $\mathbf{w}$. In the former case we prove that the flooding completion time is almost surely (a.s.) logarithmic in the number of nodes, while in the latter case we show several bounds, which depend on some specific properties of the degree sequence. We believe that these results provide an interesting first step towards the complete analysis of the flooding completion time in the case of more realistic evolving graph models (a preliminary short version of the results included in this paper has been presented in [3]).

2. The model

Given a random graph model $\mathcal{G}$, we are considering sequences $\mathcal{S} = (G_t)_{t\geq 0}$ of mutually independent random graphs in $\mathcal{G}$ on the same set $[n] = \{1, \ldots, n\}$ of nodes. We are then interested in the time it takes for a rumor initiated at an arbitrary node $i \in [n]$ to reach all nodes. The rumor propagates to the nodes according to the aforementioned flooding mechanism. Apart from ill-defined random graph models in which a node, or a set of nodes, remains perpetually disconnected from the others, the rumor will eventually reach all nodes. We define the random variable $f_{\mathcal{G}}(\mathcal{S}, i)$ as the number of steps it takes for the flooding protocol to propagate a rumor initiated at node i in a sequence $\mathcal{S}$ of random graphs picked from $\mathcal{G}$. (When it is clear from the context, the subscript $\mathcal{G}$ will be omitted). We are actually focusing on $f(\mathcal{S}) = \max_{i\in[n]} f(\mathcal{S}, i)$, that is the maximum flooding time, taken over all possible sources. We now describe the families of dynamic graphs we are interested in.

Arbitrary degree sequence: given a list $\mathbf{w} = (w_1, \ldots, w_n)$ of n positive reals with $\max_{1\leq i\leq n} w_i^2 < \sum_{i=1}^n w_i$, this paper is dealing with the random graph model $\mathcal{G}_{\mathbf{w}}$ as defined by Chung and Lu [5]. The probability of existence of an edge $e_{i,j}$ between nodes i and j (not necessarily distinct) is set to $p_{i,j} = \frac{w_i w_j}{\sum_{k=1}^n w_k}$. Hence, the expected degree d_i of node i in $\mathcal{G}_{\mathbf{w}}$ is precisely the "weight" w_i of node i. Moreover, the expected average degree of a graph G in $\mathcal{G}_{\mathbf{w}}$ is $d = \frac{1}{n}\sum_{i=1}^n w_i$ and its second order average degree is $\tilde{d} = \frac{\sum_{i=1}^n w_i^2}{\sum_{i=1}^n w_i}$. The analysis of flooding finds its main interest when each of the graphs G_t is unlikely to be connected. In such a framework, the way the information floods is indeed entirely governed by the dynamics of the graphs in the sequence, where new edges produce opportunities for the information to propagate, while disappearing edges cause delays in the message propagation through the (dynamic) network.

Two important parameters will impact the efficiency of flooding in dynamic networks. One is the *size* (i.e., number of nodes) of the largest connected component, and the other is the *volume* (i.e., twice the number of edges) of the "fattest" connected component. More specifically, following the notation of Chung and Lu [5], the size of a node set S is denoted by $|S|$, and its volume is denoted by $\mathrm{vol}(S) = \sum_{i\in S} w_i$. Hence, the volume of S is the sum of the expected degrees of the nodes in S, and the expected value of $\mathrm{vol}(G) = \mathrm{vol}(V(G))$ (where $V(G)$ denotes the set of nodes of G) is equal to nd. In what follows, we assume, w.l.o.g., that $w_1 \leq w_2 \leq \cdots \leq w_n$.

Power-law networks: an important case of degree sequence is the power-law degree distribution, that is graphs for which the number y of nodes of degree x distributes like $1/x^\beta$ for some $\beta > 0$. There are actually several models aiming at describing tractable families of random power-law graphs (see, for example, [1,6,15]). In this paper, we consider the power-law random graph model $\mathcal{G}_{\alpha,\beta}$ introduced and analyzed by Lu [15]. Roughly, this model has two parameters, α and β, the same way $\mathcal{G}_{n,p}$ (in which a graph with n nodes is constructed by including in it each edge with probability p, independently from every other edge) has two parameters. When the degree distribution is plotted in log–log scale, α is the y-intercept, while $-\beta$ is the slope. More specifically, a graph in $\mathcal{G}_{\alpha,\beta}$ has a degree distribution such that the expected number of nodes with degree k is equal to $\left\lfloor \frac{e^\alpha}{k^\beta} \right\rfloor$. In this model, the maximum degree is $\Delta = \left\lfloor e^{\alpha/\beta} \right\rfloor$, and the number of nodes is $n = \sum_{k=1}^{\Delta} \left\lfloor \frac{e^\alpha}{k^\beta} \right\rfloor$. We have

$$n \simeq \begin{cases} \zeta(\beta)\ e^\alpha & \text{if } \beta > 1, \\ \alpha\ e^\alpha & \text{if } \beta = 1, \\ c\ e^{\alpha/\beta} & \text{if } 0 < \beta < 1, \end{cases}$$

where $\zeta(x) = \sum_{k=1}^\infty \frac{1}{k^x}$ is the zeta function, and $\beta/(1-\beta) \leq c \leq 1/(1-\beta)$. To pick at random a graph in $\mathcal{G}_{\alpha,\beta}$, one fixes a sequence $\mathbf{w} = (w_1, \ldots, w_n)$ satisfying that for any $k \in [\Delta]$, $|\{i : k \leq w_i \leq k+1\}| = \left\lfloor \frac{e^\alpha}{k^\beta} \right\rfloor$, and then uses $p_{i,j} = \frac{w_i w_j}{\sum_{k=1}^n w_k}$ to pick the edges (see [15] for the details of this construction).

3. A general methodology

We first briefly recall the technique introduced by Baumann et al. [2], which allows us to reduce the analysis of flooding to computing diameters (the two parameters coincide in the static case, but are *a priori* quite different in the dynamic setting). Given a random graph model $\mathcal{G}$ for n vertices, let $p_{i,j}$ denote the probability that there exists an edge $e_{i,j}$ between node i and node j in any of the graphs in the sequence. Assume that all edges are mutually independent. We define a weighted graph H as follows. H is the n-node clique whose edges have weights defined as follows. The weight $\mathrm{weight}(e_{i,j}) \geq 1$ of edge $e_{i,j}$ between node i and node j is drawn at random according to the geometric distribution of parameter $p_{i,j}$. That is, $\Pr\left\{\mathrm{weight}(e_{i,j}) = k\right\} = p_{i,j}(1-p_{i,j})^{k-1}$. For $k \in \mathbb{N}^+ \cup \{\infty\}$,