Robust reconstruction of Barabási-Albert networks in the broadcast congested clique model

Pablo Moisset de Espanés[1], Ivan Rapaport[1,2], Daniel Remenik [1,2], and Javiera Urrutia[2]

[1]CMM (UMI 2807 CNRS), Universidad de Chile
[2] Departamento de Ingeniería Matemática, Universidad de Chile

Abstract

In the broadcast version of the congested clique model, n nodes communicate in synchronous rounds by writing $\mathcal{O}(\log n)$-bit messages on a whiteboard, which is visible to all of them. The joint input to the nodes is an undirected n-node graph G, with node i receiving the list of its neighbors in G. Our goal is to design a protocol at the end of which the information contained in the whiteboard is enough for reconstructing G. It has already been shown that there is a one-round protocol for reconstructing graphs with bounded degeneracy. The main drawback of that protocol is that the degeneracy m of the input graph G must be known *a priori* by the nodes. Moreover, the protocol fails when applied to graphs with degeneracy larger than m. In this paper we address this issue by looking for *robust* reconstruction protocols, that is, protocols which always give the correct answer and work efficiently when the input is restricted to a certain class. We introduce a very simple, two-round protocol that we call Robust-Reconstruction. We prove that this protocol is robust for reconstructing the class of Barabási-Albert trees with (expected) message size $\mathcal{O}(\log n)$. Moreover, we present computational evidence suggesting that Robust-Reconstruction also generates logarithmic size messages for arbitrary Barabási-Albert networks. Finally, we stress the importance of the preferential attachment mechanism (used in the construction of Barabási-Albert networks) by proving that Robust-Reconstruction *does not* generate short messages for random recursive

trees.

keywords: broadcast congested clique model; Barabási-Albert networks; distributed, parallel, cluster and local computing; bounded communication; graph reconstruction; message passing.

1 Introduction

The $\mathcal{CONGEST}$ model is a synchronous, message-passing model of distributed computation in which each one of n nodes can send $\mathcal{O}(\log n)$ bits along each of its incident communication links in each round [24]. In the particular case where the communication network is a complete graph all the information distributed in the nodes becomes local. Therefore, the only obstacle to perform any task is due to congestion. The main theoretical purpose of this model, known as congested clique [4, 9, 10, 11, 14, 16, 19, 20, 23], is to serve as a basic model for understanding the role played by congestion in distributed computation. Besides, there are interesting connections between the congested clique and popular models such as MapReduce [15]. Typically, the joint input to the n nodes in the congested clique model is an undirected n-node graph G, with node i receiving the list of its neighbors in G. Each node can send, in each round, $\mathcal{O}(\log n)$ bits along each of its $n-1$ communication links.

In the much more restricted, broadcast version of the congested clique model, each node can only broadcast a single $\mathcal{O}(\log n)$-bit message over all its links in each round [11]. This setting –which is the one we consider in this paper– is equivalent to the multi-party, number-in-hand computation model, where communication takes place in a shared whiteboard [1, 2, 5, 6, 7, 11, 13, 17, 18]. Writing a message $\mathcal{M}$ on the whiteboard is equivalent to broadcasting $\mathcal{M}$.

We assume that the ID of each node is a unique number between 1 and n and that the only information each node has, besides n and its own ID, is the list of IDs of its neighbors in G. At the end of the protocol the whiteboard must contain enough information to answer some question which is usually related to the topology of G. Typical goals are the following: (i) determine whether G contains a particular subgraph H, (ii) decide whether G is connected, (iii) reconstruct G.

There are two classical complexity measures:

1. Round complexity: number of rounds, where in each round all nodes write simultaneously one message on the whiteboard.

2. Message size complexity: number of bits of the longest message written on the whiteboard during the process.

If there is no restriction on the message size then there is a trivial one-round protocol that reconstructs any graph: given an arbitrary graph G and given an arbitrary assignment of IDs to each of the n nodes of G, every node writes on the whiteboard the 0-1 vector $x \in \{0,1\}^n$ corresponding to the indicator function of its neighborhood. With this information on the whiteboard, every node can easily reconstruct G.

On the other hand, if we restrict the message size then reconstructing G becomes much more difficult. Despite this, in [6] it was proved that if the degeneracy m of G is bounded and known in advance, then it is possible to reconstruct G with a one-round protocol of $\mathcal{O}(\log n)$ message size. The degeneracy m of the graph is defined as follows: G is *m-degenerate* if one can remove from G a vertex r of degree at most m, and then proceed recursively on the resulting graph $G' = G - r$, until obtaining an empty graph; the *degeneracy* of G is the smallest m such that G is m-degenerate. Note that many graph classes such as planar graphs and bounded treewidth graphs have bounded degeneracy. For instance, the degeneracy of trees is 1.

In the one-round protocol of [6], the information that each node v writes in the whiteboard corresponds to the following $(m+2)$-tuple:

- its identifier $ID(v)$.
- its degree $d_G(v)$ in G.
- for each integer p, $1 \leq p \leq m$, the quantity $\sum_{w \in N_G(v)} (ID(w))^p$ (i.e., the sum of p's powers of the identifiers of the neighbors).

We stress that this protocol always fails when applied to graphs with degeneracy larger than m. In other words, the drawback of previous protocol is that it is not robust. A protocol is said to be *robust* if it always gives the correct answer and it works efficiently when the input is restricted to a certain class.

The main purpose of this paper is to address this robustness issue in the broadcast congested clique model. We will present a two-round protocol that always reconstructs the input graph G and is guaranteed to be efficient if G is a Barabási-Albert tree. This type of random tree is a particular case of a Barabási-Albert network, which is a scale-free random graph model of bounded degeneracy which represents many real-world situations ranging from the genome to the Internet [8]. We also report on simulations which strongly suggest that our robust protocol not only reconstructs efficiently Barabási-Albert trees, but also any Barabási-Albert network.

Our approach was inspired by the work of Raghavan and Spinrad [25] in the non-distributive, centralized setting. The authors in [25] motivated their work by saying that "it is often not easy to determine whether the input is of the form for which the algorithm is designed; the recognition problem for the input class may be open or even NP-hard or worse." They illustrate this by studying the problem of finding the maximum independent set of well covered graphs (these are graphs for which every maximal independent set is also maximum). Obviously, there is a polynomial time algorithm for finding a maximum independent set if the input is *restricted* to well covered graphs. Nevertheless, in [25] Raghavan and Spinrad prove that there is no polynomial time *robust* algorithm for finding a maximum independent set for well covered graphs unless P=NP.

2 Preliminaries

Definition 1 *Let $\mathcal{G}$ be a class of (possibly randomly generated) graphs. We say that a protocol $\mathcal{P}$ is robust and reconstructs $\mathcal{G}$ with message size $\mathcal{O}(f(n))$ if and only if*

- *$\mathcal{P}$ is deterministic and reconstructs every graph G.*
- *If $G = (V, E) \in \mathcal{G}$ (is generated by some random mechanism) then, when $\mathcal{P}$ is applied to G, for every node i the (expected) size of the longest message broadcasted by node i is bounded above by $\mathcal{O}(f(|V|))$.*

The following simple proposition states that if we want to design robust protocols with $\mathcal{O}(\log n)$ message size, then they need to have at least two rounds.

Proposition 1 *Suppose that $\mathcal{P}$ is a one-round protocol that reconstructs trees with message size $\mathcal{O}(f(n))$. Then, if $\mathcal{P}$ is robust, we have $f(n) = \Omega(n)$.*

Proof Suppose that $\mathcal{P}$ is a robust one-round protocol. Since the class of all labeled graphs with n vertices has cardinality $2^{\frac{n(n-1)}{2}}$, there must be some $n \in \mathbb{N}$ and a graph G_n of size n for which some messages have at least $\frac{1}{n}\frac{n(n-1)}{2} = \frac{n-1}{2}$ bits. Now suppose that v is the node of G_n that writes the longest message. It is always possible to design a tree T_n of size n with a node v having the same neighborhood in both T_n and G_n. □

In this paper we define a very simple, two-round robust protocol that generates short messages when applied to Barabási-Albert networks, which are defined in Section 2.2. The protocol, which we call Robust-Reconstruction, is defined as follows. Let $G = (V, E)$ be an arbitrary graph and let $V = \{v_0, v_1, v_2, \ldots, v_{n-1}\}$.

Robust-Reconstruction

- `Round 1.` Each node v_i writes on the whiteboard its own ID and its degree $d_G(v_i)$.
- `Round 2.` Each node v_i writes the IDs of its neighbors having degree greater than or equal to $d_G(v_i)$.

After the second round, it is clear that there is enough information on the whiteboard to reconstruct every graph, regardless of its topology. Although the correctness of the algorithm is apparent, proving that it is efficient for a given family of graphs can be non-trivial.

Remark 1 *The length of the message written by any node in the first round is $\mathcal{O}(\log n)$.*

2.1 Local popularity

Given a graph G and a vertex v of G, we write $\theta(v)$ to denote the number of neighbors of v that have at least as many connections as v. More precisely, $\theta(v)$ denotes the number of neighbors u of v such that $d_G(v) \leq d_G(u)$.

Let G_n be a random graph of size n generated by some random mechanism. We say that the class of graphs associated to such mechanism is *locally popular*

if for every node $v \in G_n$ the expectation of $\theta(v)$ is bounded above by some constant independent of n. If k is an upper bound then we say that the class of graphs is *k-locally popular*. The following proposition is obvious.

Proposition 2 *If we apply* Robust-Reconstruction *to a k-locally popular graph then, for every node v, the expected length of both messages is logarithmic. Moreover, the expected length of the second message written by v is bounded above by $k \log n$.*

2.2 Barabási-Albert networks

Barabási-Albert networks are scale-free graphs which represent many real-world situations ranging from the genome to the Internet [8]. They are generated by a stochastic process that uses a preferential attachment rule [3]. The well-known Barabási-Albert stochastic process proceeds in discrete time steps. The state of the process at each time step $n \geq 0$ is a connected graph G_n. At the beginning, G_0 is a clique of $m + 1$ nodes. At each time step $n \geq 1$ the graph G_{n-1} is augmented with a new node v_n that is connected to m already existing nodes (i.e., nodes of G_{n-1}). It is easy to deduce from this rule that the degeneracy of G_n is m. The m nodes are chosen in G_{n-1} following a preferential attachment rule, which means that the new node v_n is connected to node $w \in V(G_{n-1})$ with a probability proportional to the degree of w in G_{n-1}.

2.3 Our results

We conjecture that Robust-Reconstruction generates short, logarithmic expected size messages when it is applied to Barabási-Albert networks. In other words, we conjecture that Barabási-Albert networks are locally-popular. In Section 3 we provide results from computational experiments which strongly suggest this.

Contrasting the simulation-based approach, in Section 4 and Section 5 we provide an analytic result for the restricted case of Barabási-Albert trees ($m = 1$). More precisely, we prove that these trees are $\frac{31}{20}$-locally popular. Our proof does not scale naturally to cases using $m > 1$, as the nice recursive structure of Barabási-Albert trees is missing

In Section 6 we study random recursive trees [22]. These are trees where nodes also arrive one by one (and therefore older nodes have higher degree in

expectation), but each arriving node is attached to a node which is chosen *uniformly* among the existing ones. We prove that these trees are not locally popular. More precisely, we prove that the local popularity of the root is $\Omega(\sqrt{\log n})$. This result stresses the importance, at least in the setting of the ROBUST-RECONSTRUCTION protocol, of the preferential attachment mechanism, which is the defining characteristic of the Barabási-Albert networks.

2.4 Open problems

It should be pointed out that *if we fix the degeneracy* $m \in \mathbb{N}$ of the graph, then there is a trivial two-round robust protocol for which the message size is bounded above by $O(\log n)$ when G is m-degenerate. To see this, consider the following protocol: in the first round apply the protocol (appeared in [6]) that we have already described, which reconstructs G with message size upper bounded by $\mathcal{O}(\log n)$ if G is m-degenerate and answers "no, G is not m-degenerate" otherwise; if the answer is negative then, in the second round, use the protocol that reconstructs G using the indicator functions, which are long messages of size n. The problem with this protocol is that, in contrast to ROBUST-RECONSTRUCTION, the parameter m must be known *a priori* by the nodes.

Since ROBUST-RECONSTRUCTION produces short messages for a subclass of degenerate graphs (the Barabási-Albert ones), a natural open question arises: Is there a two-round protocol that reconstructs every network G such that the message size is upper bounded by $O(\kappa_{deg(G)} \log n)$, where $\kappa_{deg(G)}$ denotes a parameter that depends exclusively on $deg(G)$, the degeneracy of G?

In this work we prove that, for Barabási-Albert trees, $\max_k \mathbb{E}(\theta(v_k)) \leq \frac{31}{20}$. In the future, besides finding a formal proof for general Barabasi-Albert networks, it seems natural to study the value $\mathbb{E}(\max_k \theta(v_k))$, which is a much harder problem. Based on numerical simulations and by analogy with the case of the maximum of n exponentially distributed random variables, we speculate that $\mathbb{E}(\max_k(\theta(v_k)))$ should be roughly proportional to $\log n$. This is still interesting if we consider the length of the messages in bits. It would mean that the "worst case" message, in the expected sense, is about $m \log^2 n$ bits long, which is still short.

2.5 Related work

2.5.1 Broadcast congested clique

Drucker, Kuhn and Oshman [11] gave an upper bound to the round complexity of the subgraph detection problem. They made the following remark: the degeneracy of H-free graphs can be bounded above in terms of the Turán number $ex(n, H)$, which is the maximal number of edges of an n-node graph which does not contain a subgraph isomorphic to H. Plugging this into the reconstruction protocol introduced by Becker *et al.* [6], they designed a randomized protocol that solves the H detection problem in $\mathcal{O}(ex(n, H)\log^2 n/(nb)+\log^3 n/b)$ rounds with high probability (where b is the number of bits each node can broadcast in each round).

Kari et al. [18] tackled the problem of detecting *induced* subgraphs. They provided a one-round, randomized logarithmic message size protocol for detecting an induced P_4 (a path of length 4) in the input graph G. Ahn, Guha and McGregor [1, 2, 13] introduced a powerful technique that allows one to decide in one round whether G is connected using messages of size $\mathcal{O}(\log^3 n)$, with high probability.

Some negative results have also been obtained. For instance, deciding deterministically in one round whether a graph has a triangle requires messages of size $\Theta(n)$ [6]. On the other hand, if instead of bounding the number of rounds we bound the message size b, then the best known result is the following: detecting deterministically a triangle requires $\Omega(n/(e^{\mathcal{O}(\sqrt{\log n})}b)$ rounds [11].

In [7], the authors consider three variants of the broadcast congested clique model: randomized protocols with public coins, randomized protocols with private coins and deterministic protocols. They showed that this choice affects the message size complexity of some problems. More precisely, they introduced a problem called Translated-Twins. They proved that if only one round is allowed then the message size complexity is $\Theta(n)$ in the deterministic case and $\mathcal{O}(\log n)$ in the randomized, public coin case. For the private coins setting the message size complexity is bounded below by $\Omega(\sqrt{n})$ and bounded above by $\mathcal{O}(\sqrt{n}\log n)$.

2.5.2 Congested clique

No lower bounds are known for the general model, where nodes may send different messages to each of its neighbors. Drucker, Kuhn and Oshman [11] gave a possible explanation for such difficulty. In fact, they proved that in this case it is possible to simulate powerful classes of bounded-depth circuits (and therefore lower bounds in the congested clique would yield lower bounds in circuit complexity).

The intrinsic power of the model has allowed some authors to provide extremely fast protocols for solving some natural problems: $\mathcal{O}(1)$-round protocols for routing and sorting [19, 23], a $\mathcal{O}(n^{(d-2)/d}/\log n)$-round protocol for finding a particular d-vertex subgraph [10], a $\mathcal{O}(\log\log\log n)$-round protocol for finding a 3-ruling set [16], $\mathcal{O}(n^{0.158})$-round protocols for counting triangles, for counting 4-cycles and for computing the girth [9], a $\mathcal{O}(1)$-round protocol for detecting a 4-cycle [9], and a $\mathcal{O}(\log\log\log n)$-round protocol for constructing a minimum spanning tree [14]. Dolev, Lenzen and Peled [10] describe a protocol for reconstructing deterministically any graph in $\mathcal{O}(|E|/n)$ rounds. This result is interesting for sparse graphs. In particular, this means that graphs with bounded degeneracy m can be reconstructed in $\mathcal{O}(m)$ rounds. This protocol relies heavily on the possibility given by the general model to perform a load balancing procedure efficiently.

3 Local popularity of Barabási-Albert networks

Our experiment is as follows: fix values of N (the total number of nodes) and m. Generate 2000 random graphs using these parameters. For each $k \leq N$ we compute $\theta(v_k)$, where v_k is the vertex attached at time k. We estimate $\mathbb{E}(\theta(v_k))$ as the mean of the 2000 experiments. Call these estimators $\bar{\theta}_k$. To approximate $\max_k \mathbb{E}(\theta(v_k))$, it could in principle be inaccurate to simply use $\max_k \bar{\theta}_k$. This biased estimator would overestimate the real result as the statistical noise tends to drive the value up because of the *max* function. Therefore, we use a localized linear polynomial smoothing technique to reduce the noise first, and then we compute the *max*.

There are qualitative similarities in all the results, regardless of the choice of N and m. Consider for example Figure 1 (top and bottom). They represent the values of $\bar{\theta}_k$ computed from the 2000 graphs generated with $N = 1000$ and

$m = 1$. The first two nodes (the ones in the complete graph that seeds the process) have relatively low values of $\bar{\theta}$. There is a sharp increase of $\bar{\theta}$ when k is slightly bigger than two. Then the graph peaks, and then decreases "smoothly" (ignoring the local noise) until the value of θ reaches 1. This decay is monotonic and $\theta(k)$ looks convex, if we ignore the first few points.

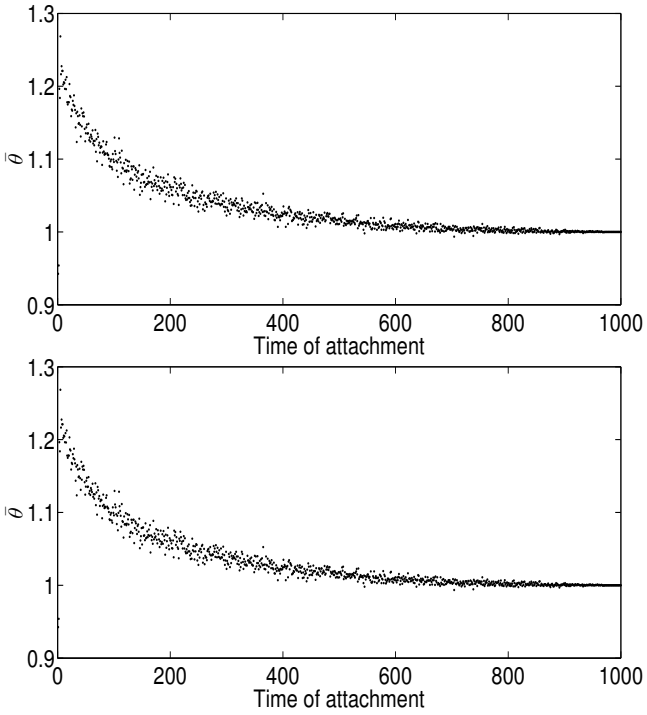

Figure 1: Top: values of $\bar{\theta}$ for $N = 1000$ and $m = 1$. Bottom: detail of the first 100 nodes.

In Figure 2 (top) we consider the simulations of graphs generated using $m = 64$ and $N = 1000$. If we increase N, while keeping m fixed, the change in the behavior is small. For instance, in Figure 2 (bottom), we used $m = 64$ but now $N = 10000$ (ten times larger than before). If we neglect the first 100 nodes or so, and adjust the horizontal scales, the plots are essentially the same. We see the same features, except that the decay after the peak is almost linear.

Finally, Table 1 shows how the numerical estimations (using smoothing) of $\max_k \mathbb{E}(\theta(v_k))$ change with m and N. Because of the scaling property, the heights of the peaks do not change much if one only modifies N, and therefore

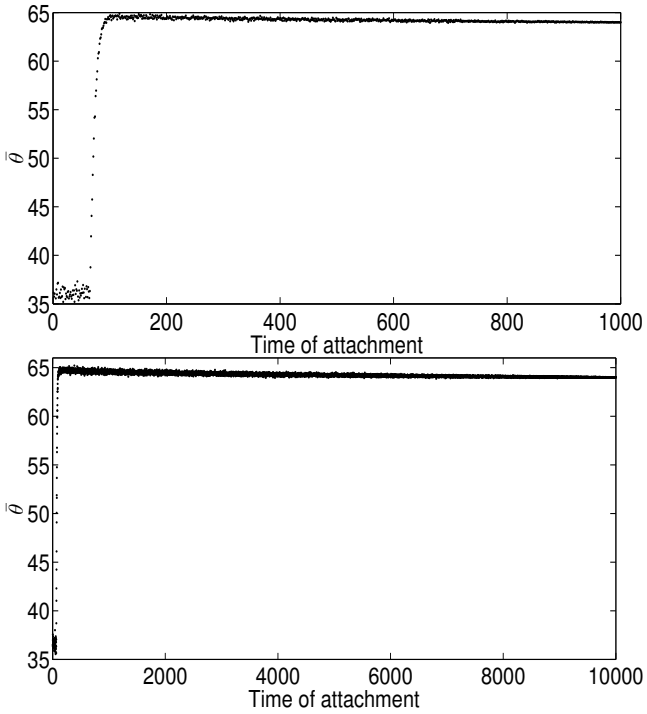

Figure 2: Values of $\bar{\theta}$. Top: $N = 1000$ and $m = 64$. Bottom: $N = 10000$ and $m = 64$.

we see little change within each column. These heights do change with m, but in a predictable way. We note that our estimation for $\max_k \mathbb{E}(\theta(v_k))$ is never larger than $m + 1$ and this bound seems to get tighter as m and N increase.

		m				
		1	4	16	64	128
N	1000	1.23	4.42	16.54	64.56	128.57
	10000	1.31	4.53	16.68	64.76	128.79
	100000	1.36	4.58	16.74	64.88	128.93

Table 1: $\max_k \ \bar{\theta}_k$, for different values of N and m.

We summarize our conjectures:

(i) Given a family of Barabási-Albert graphs generated using parameters N and

m, for all $k \leq N$, $\mathbb{E}(\theta(v_k)) \leq m + 1 - o(1)$. Furthermore, the bound is tight if $N, m \to \infty$ and $m \ll N$.

(ii) $\mathbb{E}(\theta(v_k))$ is scale independent. Informally, $\mathbb{E}(\theta(v_k))$ is mostly determined by k/N. This is suggested by comparing experiments that use the same m but different N, such as those described in Figure 2.

4 Local popularity of Barabási-Albert trees

Contrasting the simulation-based approach, we provide an analytic result for the restricted case of Barabási-Albert trees, which are simply Barabási-Albert graphs generated choosing $m = 1$. Recall that in particular this means that we start with a clique of size 2, so at the beginning we have the tree $T_0 = (V_0, E_0)$ where $V_0 = \{v_0, v_0'\}$ and $E_0 = \{v_0 v_0'\}$. Note that the degree of v_0 (and v_0') at the beginning is 1. Let $d_T(v)$ denote the degree of the vertex v in the tree T. The process evolves in discrete time steps $n = 1, 2, \ldots$ as follows:

1. We choose a unique $w_n \in V_{n-1}$ according to the following probability distribution:

$$\forall v \in V^{n-1}, \quad \mathbb{P}(w_n = v | T_{n-1}) = \frac{d_{T_{n-1}}(v)}{\sum_{u \in V_{n-1}} d_{T_{n-1}}(u)}$$

2. $V^n = V^{n-1} \cup \{v_n\}$ and $E_n = E_{n-1} \cup \{w_n v_n\}$.

We state now the main result of the paper:

Proposition 3 *The class of Barabási-Albert trees is* $\frac{31}{20}$*-locally popular.*

As will become clear in the proof, this result can be slightly sharpened to say that the class of Barabási-Albert trees is "asymptotically" $\frac{3}{2}$-locally popular, in the sense that a Barabási-Albert tree of size n is $(\frac{3}{2} + o(1))$-locally popular. As a direct consequence of Propositions 2 and 3 we obtain a bound on the maximal expected length among all messages written when applying Robust-Reconstruction to Barabási-Albert trees.

Corollary 1 *If we apply* Robust-Reconstruction *to Barabási-Albert trees then the expected length of the messages written by every node is bounded above by* $\frac{31}{20}\log n$.

Before turning to the proof of Proposition 3 let us provide the intuition behind it. We first study how the degree of the root v_0 compares to the degree of its children as the tree grows (by symmetry the same will apply to v_0'). Informally, we want to bound $\mathbb{E}(\theta(v_0))$. This is done by studying a process similar to the conventional Barabási-Albert tree construction algorithm but considering only attachments involving the root or its children. To bound $\mathbb{E}(\theta(v_0))$ we study the "contest for higher degree" that occurs between the root and each child *individually*. This competition process is described using a Pólya-Eggenberger urn model. After we obtain a bound for $\mathbb{E}(\theta(v_0))$, the recursive structure of Barabási-Albert trees allows us to extend the result to all the remaining nodes.

We can partition the Barabási-Albert tree T_n into two subtrees:

- $T_n^{v_0}$, the v_0-subtree rooted at v_0.
- $T_n^{v_0'}$, the v_0'-subtree rooted at v_0'.

These two subtrees are joined by the edge v_0v_0'. More precisely: $V(T_n) = V(T_n^{v_0}) \cup V(T_n^{v_0'})$ and, in terms of edges, $E(T_n) = E(T_n^{v_0}) \cup \{v_0v_0'\} \cup E(T_n^{v_0'})$.

In order to estimate $\mathbb{E}(\theta(v_0))$ we study the subtree $T_n^{v_0}$ rooted at v_0. For that purpose, we only need to focus on what happens in the first 3 layers of $T_n^{v_0}$. More precisely, we will study how the degree of the root v_0 changes compared to the degrees of its children. The process in this case is exactly the same as the general one except that, when adding a new node, we only consider as possible neighbors those nodes that are at distance at most 1 from v_0 (including v_0 itself, which is at distance 0).

We start the process with $n = 0$ and we stop it with $n = N$. Note that the generated tree, that we denote by $\mathcal{T}_N$, will have 3 layers. This is equivalent to considering the process in the complete v_0-subtree up to the time when the number of nodes at distance at most 2 from the root v_0 first equals N.

Definition 2 *Let us consider the situation after the n-th node is added to the 3-layer tree* $\mathcal{T}_n$. *More precisely, let us define the following random variables.*

- d_n^0 = *the degree of node* v_0 *in step* n. *Recall that* $d_0^0 = 1$ *and therefore* $d_1^0 = 2$.

- d_n^k = *the degree of the* k*-th neighbor of* v_0 *(*v *is the* k*-th neighbor of* v_0 *if* $k-1$ *neighbors of* v_0 *arrived before* v*; when* v_0 *has less than* k *neighbors we impose* $d_n^k = 0$*).*

- $S_n^k = \begin{cases} 1 & \text{if } d_n^k \geq d_n^0, \\ 0 & \text{otherwise.} \end{cases}$

- $M_n = \sum_{k=1}^{n} S_n^k$. *In other words,* M_n *is the number of neighbors of* v_0 *having degree greater than or equal to the one of* v_0.

Since $\mathbb{E}(M_n) = \sum_{k=1}^{n} \mathbb{E}(S_n^k)$, we will bound $\mathbb{E}(S_n^k)$ for every fixed k. In other words, we only need to worry about the dynamic competition between two nodes: v_0 and its k-th neighbor v_k. This dynamic competition corresponds exactly to the Pólya-Eggenberger urn model [12]. In this model, the urn starts with r red balls and b black balls; one ball is drawn randomly from the urn and its color observed; it is then replaced in the urn, and an additional ball of the same color is added. The process is repeated.

In our case, the competition between v_0 and v_k starts as soon as the degree of v_0 becomes equal to $k+1$ and the degree of v_k becomes 1 (i.e., as soon as v_k is connected to v_0). This is equivalent to starting the urn process with $k+1$ red balls and 1 black ball. Define J_n^k as the fraction of black balls in step n.

Remark 2 *It is known that* $J_n^k \xrightarrow[a.s.]{n\to\infty} J_\infty^k$, *where* $J_\infty^k \sim \beta(1, k+1)$ *and* $\beta(1, k+1)$ *denotes the Beta distribution with parameters* 1 *and* $k+1$ *(for a comprehensive treatment of the subject see for instance [21]).*

The corresponding density function is given as follows:

$$f_{\beta(1,k+1)} = \frac{\Gamma(k+2)}{\Gamma(k+1)\Gamma(1)}(1-x)^k = (k+1)(1-x)^k.$$

Proposition 4 *Let* $k \in \mathbb{N}$, $k \geq 2$. *Then, for all* $n \in \mathbb{N}$,

$$\mathbb{P}\left(J_n^k \geq \frac{1}{2}\right) \leq \mathbb{P}\left(J_\infty^k \geq \frac{1}{2}\right).$$

On the other hand, for the case $k = 1$ we have

$$\mathbb{P}\left(J_n^1 \geq \frac{1}{2}\right) = \frac{1}{4} + o(1) \text{ and } \mathbb{P}\left(J_n^1 \geq \frac{1}{2}\right) \leq \frac{3}{10}.$$

Proof See next section. □

As a consequence we obtain the following result.

Corollary 2 *$\mathbb{E}(M_n) \leq \frac{1}{2} + o(1)$ and, for all $n \in \mathbb{N}$, we have $\mathbb{E}(M_n) \leq \frac{11}{20}$.*

Proof For $k \geq 2$

$$\begin{aligned}\mathbb{P}\left(J_\infty^k \geq \frac{1}{2}\right) &= \int_{\frac{1}{2}}^1 f_{\beta(1,k+1)}\, dx = (k+1)\int_{\frac{1}{2}}^1 (1-x)^k\, dx \\ &= \frac{1}{2^{k+1}}.\end{aligned}$$

It follows that

$$\begin{aligned}\mathbb{E}\left(M_n\right) &= \sum_{k=1}^n \mathbb{E}(S_n^k) = \sum_{k=1}^n \mathbb{P}\left(J_n^k \geq \frac{1}{2}\right) \\ &\leq \mathbb{P}\left(J_n^1 \geq \frac{1}{2}\right) + \sum_{k=2}^n \mathbb{P}\left(J_\infty^k \geq \frac{1}{2}\right) \\ &\leq \frac{3}{10} + \sum_{k=2}^n \frac{1}{2^{k+1}} = \frac{3}{10} + \frac{1}{4} = \frac{11}{20}.\end{aligned}$$

□

Proof of Proposition 3 Consider now the general Barabási-Albert tree. Since T^i, the subtree rooted at any node v_i, is also a Barabási-Albert tree, it follows that at any time of the process, the expected number of children of v_i having degree greater than or equal to $d_{T^i}(v_i)$ is bounded by $\frac{11}{20}$. Nevertheless, the parent of v_i could eventually have more neighbors than v_i. Therefore, $\frac{11}{20}+1 = \frac{31}{20}$ is an upper bound for $\mathbb{E}(\theta(v_i))$ and Proposition 3 follows. □

Remark 3 *Proposition 3 gives a bound for the local popularity of every node in the tree. For the two particular sibling nodes v_0 and v_0' from which the Barabási-Albert process starts, this bound can be improved. In fact, the expected number of children of v_0 (resp. v_0') having degree larger than or equal to the degree of v_0 (resp. v_0') is bounded above by $\frac{1}{2} + o(1)$ thanks to Proposition 4. On the other hand, by symmetry, the probability that the degree of v_0 is larger or equal to than the degree of v_0' is also $\frac{1}{2} + o(1)$ (this is because the probability that both have the same degree goes to 0 as $n \to \infty$). Therefore, the local popularity of these two nodes is bounded above by $\frac{1}{2} + \frac{1}{2} + o(1) = 1 + o(1)$. This particularity can be seen in the first two points of Figure 1 (bottom).*

5 Proof of Proposition 4

Definition 3 *Let $\tilde{B}_n^k$ be the number of black ball draws in the Pólya-Eggenberger urn after n draws, starting the process with 1 black ball and $k+1$ red balls.*

Lemma 1 *Let $n, k, i \in \mathbb{N}$, $i + 1 \leq n$. The following holds:*

$$\mathbb{P}(\tilde{B}_n^k = i) = \mathbb{P}(\tilde{B}_n^k = i+1) \cdot \frac{k+n-i}{n-i}.$$

Proof The probability that the first i draws correspond to black balls and that the next $n-i$ draws to red balls is given by

$$\frac{1 \cdot 2 \cdots i \cdot (k+1) \cdot (k+2) \cdots (k+n-i)}{(k+2) \cdot (k+3) \cdots (k+n+1)}.$$

Any other order in which exactly i black balls are drawn corresponds to a permutation of the terms in the numerator of previous expression (with the same denominator). Therefore, the probability of drawing exactly i black balls is

$$\frac{1 \cdot 2 \cdots i \cdot (k+1) \cdot (k+2) \cdots (k+n-i)}{(k+2) \cdot (k+3) \cdots (k+n+1)} \binom{n}{i}. \tag{1}$$

Using this we get

$$\begin{aligned}
\mathbb{P}(\tilde{B}_n^k = i) &= \frac{i!(k+1)(k+2)\cdots(k+n-i)}{(k+2)(k+3)\cdots(k+n+1)} \binom{n}{i} \\
&= \frac{(k+1)(k+2)\cdots(k+n-i-1)}{(k+2)(k+3)\cdots(k+n+1)} \frac{n!}{(n-i-1)!} \frac{k+n-i}{n-i} \\
&= \mathbb{P}(\tilde{B}_n^k = i+1) \frac{k+n-i}{n-i}.
\end{aligned}$$

□

Definition 4 *Let R_n^k and B_n^k be the number of red and black balls respectively in the Pólya-Eggenberger urn, after n draws, starting the process with 1 black ball and $k+1$ red balls. (Note that $B_n^k = \tilde{B}_n^k + 1$).*

Remark 4 *Note that $\mathbb{P}\left(J_n^k \geq \frac{1}{2}\right) = \mathbb{P}(B_n^k \geq R_n^k)$. Hence, by Remark 2, it follows that $\left(\mathbb{P}(B_n^k \geq R_n^k)\right)_{n\in\mathbb{N}}$ converges. Therefore, in order to prove Proposition 4 (for $k \geq 2$), it would be enough to show that sequence $\left(\mathbb{P}(B_n^k \geq R_n^k)\right)_{n\in\mathbb{N}}$ grows monotonically. Nevertheless, as we will see below, this is not true. Instead, we will use the fact that this monotonicity holds if we focus separately on the two subsequences for n even and n odd (which is enough for our purposes). To this end, we need to distinguish the cases k even and k odd.*

5.1 The case k even

Proposition 5 *Let $k, n \in \mathbb{N}$ with $k \geq 2$ even. It follows that*

$$\mathbb{P}(B_{2n+2}^k \geq R_{2n+2}^k) \geq \mathbb{P}(B_{2n+1}^k \geq R_{2n+1}^k).$$

Proof After $2n+1$ draws, we will have a total of $(k+2n+3)$ balls in the urn (an odd number). Then,

$$\begin{aligned}
\mathbb{P}(B_{2n+2}^k \geq R_{2n+2}^k) &= \mathbb{P}\left(B_{2n+2}^k \geq R_{2n+2}^k | B_{2n+1}^k = \frac{k+2n+2}{2}\right) \\
&\quad \cdot \mathbb{P}\left(B_{2n+1}^k = \frac{k+2n+2}{2}\right) \\
&\quad + 1 \cdot \mathbb{P}\left(B_{2n+1}^k \geq \frac{k+2n+2}{2} + 1\right) \\
&\quad + 0 \cdot \mathbb{P}\left(B_{2n+1}^k \leq \frac{k+2n+2}{2} - 1\right) \\
&\geq \mathbb{P}\left(B_{2n+1}^k \geq \frac{k+2n+2}{2} + 1\right) \\
&= \mathbb{P}(B_{2n+1}^k \geq R_{2n+1}^k).
\end{aligned}$$

□

In view of this result, which tells us that $\mathbb{P}(J^k_{2n+1} \leq \frac{1}{2}) \leq \mathbb{P}(J^k_{2n+2} \leq \frac{1}{2})$ for each n, it is enough to verify that the sequence $\left(\mathbb{P}(B^k_{2n} \geq R^k_{2n})\right)_{n\in\mathbb{N}}$ is increasing.

Proposition 6 *Let $k, n \in \mathbb{N}$ with $2n \geq k \geq 2$, k even. It follows that*

$$\mathbb{P}(B^k_{2n+2} \geq R^k_{2n+2}) \geq \mathbb{P}(B^k_{2n} \geq R^k_{2n}).$$

Proof Proceeding similarly to the previous proof,

$$\begin{aligned}
\mathbb{P}(B^k_{2n+2} \geq R^k_{2n+2}) = \quad & \mathbb{P}\left(B^k_{2n+2} \geq R^k_{2n+2} | B^k_{2n} = \frac{k+2n+2}{2}\right) \\
& \cdot \mathbb{P}\left(B^k_{2n} = \frac{k+2n+2}{2}\right) \\
& + \mathbb{P}\left(B^k_{2n+2} \geq R^k_{2n+2} | B^k_{2n} = \frac{k+2n+2}{2} - 1\right) \\
& \cdot \mathbb{P}\left(B^k_{2n} = \frac{k+2n+2}{2} - 1\right) \\
& + 1 \cdot \mathbb{P}\left(B^k_{2n} \geq \frac{k+2n+2}{2} + 1\right).
\end{aligned}$$

The following identity holds because the only favorable scenario for the event inside is to draw black balls in the last 2 draws:

$$\begin{aligned}
\mathbb{P}\left(B^k_{2n+2} \geq R^k_{2n+2} \,\middle|\, B^k_{2n} = \frac{k+2n+2}{2} - 1\right) &= \frac{\frac{k+2n+2}{2} - 1}{k+2n+2} \cdot \frac{\frac{k+2n+2}{2}}{k+2n+3} \\
&= \frac{1}{4} \cdot \frac{k+2n}{k+2n+3}
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
\mathbb{P}\left(B^k_{2n+2} \geq R^k_{2n+2} \,\middle|\, B^k_{2n} = \frac{k+2n+2}{2}\right) &= 1 - \frac{\frac{k+2n+2}{2}}{k+2n+2} \cdot \frac{\frac{k+2n+2}{2} + 1}{k+2n+3} \\
&= 1 - \frac{1}{4} \cdot \frac{k+2n+4}{k+2n+3},
\end{aligned}$$

since now the only unfavorable scenario is to draw red balls in the last 2 draws. From Lemma 1 (taking $i = \frac{k+2n+2}{2} - 2$) we get

$$\begin{aligned}
\mathbb{P}\left(B^k_{2n} = \frac{k+2n+2}{2} - 1\right) &= \mathbb{P}\left(\tilde{B}^k_{2n} = \frac{k+2n+2}{2} - 2\right) \\
&= \mathbb{P}\left(\tilde{B}^k_{2n} = \frac{k+2n+2}{2} - 1\right) \cdot \left(\frac{k+2n+2}{2n-k+2}\right) \\
&= \mathbb{P}\left(B^k_{2n} = \frac{k+2n+2}{2}\right) \cdot \left(\frac{k+2n+2}{2n-k+2}\right).
\end{aligned}$$

Putting all this together we deduce that

$$\begin{aligned}\mathbb{P}(B^k_{2n+2} \geq R^k_{2n+2}) = \mathbb{P}&\left(B^k_{2n} \geq \frac{k+2n+2}{2}+1\right)\\ &+\left(1-\frac{1}{4}\cdot\frac{k+2n+4}{k+2n+3}\right)\mathbb{P}\left(B^k_{2n}=\frac{k+2n+2}{2}\right)\\ &+\left(\frac{1}{4}\frac{k+2n}{k+2n+3}\right)\left(\frac{k+2n+2}{2n-k+2}\right)\\ &\qquad\cdot\mathbb{P}\left(B^k_{2n}=\frac{k+2n+2}{2}\right).\end{aligned}$$

Now

$$1-\frac{1}{4}\cdot\frac{k+2n+4}{k+2n+3}+\frac{1}{4}\cdot\frac{k+2n}{k+2n+3}\cdot\frac{k+2n+2}{2n-k+2}\geq 1 \tag{2}$$

because the inequality reduces to $\frac{1}{4}\cdot\frac{k+2n}{k+2n+3}\cdot\frac{k+2n+2}{2n-k+2}\geq\frac{1}{4}\frac{k+2n+4}{k+2n+3}$, which is equivalent to $\frac{k+2n+2}{2n-k+2}\geq\frac{k+2n+4}{k+2n}$ and holds whenever $2n\geq k$ and $k\geq 2$. Therefore,

$$\begin{aligned}\mathbb{P}(B^k_{2n+2}\geq R^k_{2n+2}) &\geq \mathbb{P}\left(B^k_{2n}\geq\frac{k+2n+2}{2}+1\right)+\mathbb{P}\left(B^k_{2n}=\frac{k+2n+2}{2}\right)\\ &=\mathbb{P}\left(B^k_{2n}\geq\frac{k+2n+2}{2}\right)\\ &=\mathbb{P}(B^k_{2n}\geq R^k_{2n}).\end{aligned}$$

□

5.2 The case k odd

The following two propositions are analogous to Propositions 5 and 6. The proof of the first one is analogous to the one for the case k even, so we omit it. The second one is also proved similarly, but we need to deal with the case $k=1$ separately.

Proposition 7 *Let $k,n\in\mathbb{N}$ with k odd. Then*

$$\mathbb{P}(B^k_{2n+1}\geq R^k_{2n+1})\geq\mathbb{P}(B^k_{2n}\geq R^k_{2n}).$$

Proposition 8 *Let $k,n\in\mathbb{N}$ with $2n+1>k$ and $k\geq 3$ odd. Then*

$$\mathbb{P}(B^k_{2n+3}\geq R^k_{2n+3})\geq\mathbb{P}(B^k_{2n+1}\geq R^k_{2n+1}).$$

On the other hand, for the case $k = 1$ we have

$$\mathbb{P}(B^1_{2n+3} \geq R^1_{2n+3}) \leq \mathbb{P}(B^1_{2n+1} \geq R^1_{2n+1}).$$

Proof Proceeding analogously to the case k even leads to the identity

$$\begin{aligned}\mathbb{P}(B^k_{2n+3} \geq R^k_{2n+3}) = \mathbb{P}&\left(B^k_{2n+1} \geq \frac{k+2n+3}{2} + 1\right) \\ &+ \left(1 - \frac{1}{4} \cdot \frac{k+2n+5}{k+2n+4}\right) \mathbb{P}\left(B^k_{2n+1} = \frac{k+2n+3}{2}\right) \\ &+ \left(\frac{1}{4}\frac{k+2n+1}{k+2n+4}\right)\left(\frac{k+2n+3}{2n-k+3}\right) \\ &\qquad \cdot \mathbb{P}\left(B^k_{2n+1} = \frac{k+2n+3}{2}\right)\end{aligned}$$

(we omit the details). The statement for the case $k \geq 3$ now follows from the inequality

$$1 - \frac{1}{4} \cdot \frac{k+2n+5}{k+2n+4} + \frac{1}{4} \cdot \frac{k+2n+1}{k+2n+4} \cdot \frac{k+2n+3}{2n-k+3} \geq 1,$$

which is analogous to (2) and holds for $k \geq 3$. However, for $k = 1$ the opposite inequality holds, which implies that the sequence $\left(\mathbb{P}(B^1_{2n+1} \geq R^1_{2n+1})\right)_{n\geq 1}$ is decreasing as claimed. □

Proof of Proposition 4 The case $k \geq 2$ follows from Propositions 5, 6, 7 and 8 (together with Remark 4). To see this, consider the case k even. By Proposition 6 we have that the sequence $\left(\mathbb{P}(J^k_{2n} \geq \frac{1}{2})\right)_{n\in\mathbb{N}}$ is non-decreasing, and thus

$$\mathbb{P}(J^k_{2n} \geq \tfrac{1}{2}) \leq \mathbb{P}(J^k_\infty \geq \tfrac{1}{2}) \tag{3}$$

for all $n \in \mathbb{N}$. By Proposition 5 we have

$$\mathbb{P}(J^k_{2n+1} \geq \tfrac{1}{2}) \leq \mathbb{P}(J^k_{2n+2} \geq \tfrac{1}{2}),$$

which means that the bound (3) also holds along odd times. The case k odd, $k \geq 3$ is analogous.

For $k = 1$, Proposition 8 gives

$$\mathbb{P}(J^1_{2n+1} \geq \tfrac{1}{2}) \leq \mathbb{P}(J^1_3 \geq \tfrac{1}{2}) = \tfrac{3}{10}$$

for all $n \geq 1$, where the equality follows from (1) by a simple calculation. Since $\mathbb{P}(J^1_{2n} \geq \frac{1}{2}) \leq \mathbb{P}(J^1_{2n+1} \geq \frac{1}{2})$ we deduce that $\mathbb{P}(J^1_n \geq \frac{1}{2})$ is bounded by $\frac{3}{10}$ for all $n \geq 2$.

We are left showing that $\mathbb{P}(J_n^1 \geq \frac{1}{2}) = \frac{1}{4} + o(1)$. For this, take $n \geq 3$ and write

$$\mathbb{P}\left(J_n^1 \geq \frac{1}{2}\right) = \frac{2}{3}\mathbb{P}\left(J_n^2 \geq \frac{1}{2}\right) + \frac{1}{3}\mathbb{P}\left(\tilde{J}_{n-1}^{2,2} \geq \frac{1}{2}\right),$$

where $\tilde{J}_n^{2,2}$ denotes the same quantity as J_n^k but with the urn starting with two balls of each color. The first probability is bounded by $\frac{1}{8}$ by the case k even, so

$$\mathbb{P}\left(J_n^1 \geq \frac{1}{2}\right) \leq \frac{1}{12} + \frac{1}{3}\mathbb{P}\left(\tilde{J}_{n-1}^{2,2} > \frac{1}{2}\right) + \frac{1}{3}\mathbb{P}\left(\tilde{J}_{n-1}^{2,2} = \frac{1}{2}\right).$$

By symmetry, $\mathbb{P}\left(\tilde{J}_{n-1}^{2,2} > \frac{1}{2}\right) = \mathbb{P}\left(\tilde{J}_{n-1}^{2,2} < \frac{1}{2}\right)$, so they are both bounded by $\frac{1}{2}$. Since $= \mathbb{P}\left(\tilde{J}_{n-1}^{2,2} = \frac{1}{2}\right) \longrightarrow 0$ as $n \to \infty$, this means that

$$\mathbb{P}\left(J_n^1 \geq \frac{1}{2}\right) \leq \frac{1}{12} + \frac{1}{6} + o(1) = \frac{1}{4} + o(1).$$

□

6 Local popularity of random recursive trees

Random recursive trees were first studied in [22]. As in the Barabási-Albert case, nodes also arrive one by one; nevertheless, each arriving node is attached to a node which is chosen *uniformly* among the existing ones. Let θ_n denote the local popularity of the root at time $n-1$, that is, when the tree has n nodes.

Proposition 9 $\mathbb{E}(\theta_n) = \Omega(\sqrt{\log n})$.

Proof Let $f : \mathbb{N} \longrightarrow \mathbb{N}$ be given by $f(n) = \lfloor \sqrt{\log n} \rfloor$. Let $d_{f(n)}$ denote the degree of the root at time $f(n)$ and define the event

$$A_n = \{d_{f(n)} \geq \log f(n)\}.$$

Let ξ_k be 1 if the vertex added at time k is connected to the root and zero otherwise. Since the ξ_k's are independent, with $\mathbb{P}(\xi_k = 1) = \frac{1}{k}$, we have

$$\mathbb{E}(d_{f(n)}) = H_1(f(n)) := \sum_{k=1}^{f(n)} \frac{1}{k},$$

and

$$\mathrm{Var}(d_{f(n)}) = H_2(f(n)) := \textstyle\sum_{k=1}^{f(n)} \frac{k-1}{k^2}.$$

Thus, by the Lindeberg Central Limit Theorem and the fact that $|H_1(m) - \log m|$ is bounded in m,

$$\begin{aligned}\mathbb{P}(A_n) &= \mathbb{P}\left(\tfrac{d_{fn} - H_1(f(n))}{\sqrt{H_2(f(n))}} \geq \tfrac{\log f(n) - H_1(f(n))}{\sqrt{H_2(f(n))}}\right) \\ &\xrightarrow[n\to\infty]{} \mathbb{P}(Z \geq 0),\end{aligned}$$

where Z is a standard normal random variable. We deduce that there exists a constant $p_1 > 0$ such that $\mathbb{P}(A_n) \geq p_1$ for large enough n. Since $\mathbb{E}(\theta_n) \geq \mathbb{E}(\theta_n | A_n)\mathbb{P}(A_n)$, we deduce that it is enough to show that

$$\mathbb{E}(\theta_n | A_n) = \Omega(f(n)). \tag{4}$$

Define the event

$$\begin{aligned}B_n = \{&\text{at time } n \text{ the degree of the } \log f(n)\text{-th child} \\ &\text{of the root is} \geq \text{the degree of the root}\}.\end{aligned}$$

Then we have

$$\begin{aligned}\mathbb{E}(\theta_n | A_n) &\geq \sum_{i=1}^{\log f(n)} \mathbb{P}\big(\text{in step } n \text{ the degree of the } i\text{-th} \\ &\qquad\qquad \text{child is} \geq \text{the degree of the root} | A_n\big) \\ &\geq \log f(n) \mathbb{P}(B_n | A_n).\end{aligned} \tag{5}$$

Now let Δ_m be the difference between the degree of the root and the degree of the $\log f(n)$-th child of the root at time m (if the $\log f(n)$-th child has not appeared by time m let $\Delta_m = \infty$), so that $B_n = \{\Delta_n \leq 0\}$. Let K_n be the number of nodes which are attached to either the root or the $\log f(n)$-th child during the steps $f(n)+1, f(n)+2, \ldots, n$. Conditional on K_n and the event A_n, Δ_n has the distribution of a simple random walk started at $\Delta_{f(n)}$ after taking K_n steps, and so if X_n is a binomial random variable with parameters $(K_n, 1/2)$ then we have

$$\begin{aligned}\mathbb{P}(B_n | A_n, K_n) &= \mathbb{P}\big(X_n \geq \tfrac{1}{2}(K_n + \Delta_{f(n)}) | A_n, K_n\big) \\ &\geq \mathbb{P}\big(X_n \geq \tfrac{1}{2}(K_n + f(n)) | K_n\big),\end{aligned}$$

and where the inequality follows from the facts that $d_{f(n)} \leq f(n)$ and that, for the event A_n, the $\log f(n)$-th child of the root has already arrived by time $f(n)$.

Reasoning as above, we deduce from the definition of Δ_n and the Lindeberg CLT that $\mathbb{P}(K_n \geq \log n - \log f(n) | A_n) \geq p_2$ for some $p_2 > 0$ and large enough n. On the other hand it is not hard to see that there is a constant $p_3 > 0$ so that if $K_n \geq M_n$ for some constant M_n and Y_n is a binomial random variable with parameters $(M_n, 1/2)$, then

$$\mathbb{P}\big(X_n \geq \tfrac{1}{2}(K_n + f(n)))\big) \geq p_3\, \mathbb{P}\big(Y_n \geq \tfrac{1}{2}(M_n + f(n)))\big).$$

Therefore, letting $M_n = \log n - \log f(n)$ we deduce that

$$\begin{aligned}
&\mathbb{P}(B_n | A_n) \\
&\quad \geq \mathbb{P}(B_n | A_n \cap \{K_n \geq M_n\}) \mathbb{P}(\{K_n \geq M_n\} | A_n) \\
&\quad \geq p_2 p_3\, \mathbb{P}\big(Y_n \geq \tfrac{1}{2}(\log n - \log f(n) + \tfrac{1}{2} f(n)\big) \\
&\quad = p_2 p_3\, \mathbb{P}\left(\frac{Y_n - \frac{\log n - \log f(n)}{2}}{2\sqrt{\log n - \log f(n)}} \geq \frac{f(n)}{4\sqrt{\log n - \log f(n)}} \right).
\end{aligned}$$

By the Central Limit Theorem it follows that, for large enough values of n, the last probability is approximately $\mathbb{P}(Z \geq f(n)/(4\sqrt{\log n - \log f(n)}))$ with Z a standard normal random variable.

Since $f(n)/\sqrt{\log n - \log f(n)} \longrightarrow 1$ as $n \to \infty$, there is a $p_4 > 0$ such that

$$\mathbb{P}(Z \geq f(n)/(4\sqrt{\log n - \log f(n)})) \geq p_4 \text{ for all } n.$$

Using this above gives (4) and thus the result. □

Acknowledgments

Partially supported by CONICYT via Basal in Applied Mathematics (I.R., D.R.), Núcleo Milenio Información y Coordinación en Redes ICM/FIC RC130003 (I.R.), Fondecyt 1130061 (I.R., J.U.), Fondecyt 1120309 (D.R.) and Núcleo Milenio Modelos Estocásticos de Sistemas Complejos y Desordenados NC130062 (D.R.).

References

[1] K.J. Ahn, S. Guha, and A. McGregor, Analyzing graph structure via linear measurements, Proc Twenty-Third Ann ACM-SIAM Symp Discrete Algorithms, Kyoto, Japan, 2012, pp. 459–467.

[2] K.J. Ahn, S. Guha, and A. McGregor, Graph sketches: Sparsification, spanners, and subgraphs, Proc 31st ACM SIGMOD-SIGACT-SIGART Symp Principles of Database Systems, Scottsdale, AZ, USA, 2012, pp. 5–14.

[3] A.L. Bárabasi and R. Albert, Emergence of scaling in random networks, Science 286 (1999), 509–512.

[4] F. Becker, A. Fernandez Anta, I. Rapaport, and E. Reémila, Brief announcement: A hierarchy of congested clique models, from broadcast to unicast, Proc 2015 ACM Symp Principles Distributed Computing, New York, NY, USA, 2015, pp. 167–169.

[5] F. Becker, A. Kosowski, N. Nisse, I. Rapaport, and K. Suchan, Allowing each node to communicate only once in a distributed system: Shared whiteboard models, Proc 24th ACM Symp Parallelism in Algorithms and Architectures, Pittsburgh, PA, USA, 2012, pp. 11–17.

[6] F. Becker, M. Matamala, N. Nisse, I. Rapaport, K. Suchan, and I. Todinca, Adding a referee to an interconnection network: What can(not) be computed in one round, Proc 25th IEEE Int Symp Parallel and Distributed Processing, Anchorage, Alaska, USA, 2011, pp. 508–514.

[7] F. Becker, P. Montealegre, I. Rapaport, and I. Todinca, The simultaneous number-in-hand communication model for networks: Private coins, public coins and determinism, Proc 21st Int Colloq, Structural Information and Communication Complexity, Takayama, Japan, 2014, pp. 83–95.

[8] S. Bornholdt and H.G. Schuster (Editors), Handbook of graphs and networks: From the genome to the internet, John Wiley & Sons, Inc., New York, NY, USA, 2003.

[9] K. Censor-Hillel, P. Kaski, J.H. Korhonen, C. Lenzen, A. Paz, and J. Suomela, Algebraic methods in the congested clique, Proc 2015 ACM

Symp Principles of Distributed Computing, Donostia-San Sebastián, Spain, 2015, pp. 143–152.

[10] D. Dolev, C. Lenzen, and S. Peled, "Tri, tri again": Finding triangles and small subgraphs in a distributed setting - (ext. abstract), Proc 26th Int Symp Distributed Computing, Salvador, Brazil, 2012, pp. 195–209.

[11] A. Drucker, F. Kuhn, and R. Oshman, On the power of the congested clique model, ACM Symp Principles of Distributed Computing, Paris, France, 2014, pp. 367–376.

[12] F. Eggenberger and G. Pólya, Über die statistik verketteter vorgänge, ZAMM-Zeitschrift für Angewandte Mathematik und Mech 3 (1923), 279–289.

[13] S. Guha, A. McGregor, and D. Tench, Vertex and hyperedge connectivity in dynamic graph streams, Proc 34th ACM Symp Principles of Database Systems, Melbourne, Victoria, Australia, 2015, pp. 241–247.

[14] J.W. Hegeman, G. Pandurangan, S.V. Pemmaraju, V.B. Sardeshmukh, and M. Scquizzato, Toward optimal bounds in the congested clique: Graph connectivity and MST, Proc 2015 ACM Symp Principles of Distributed Computing, New York, NY, USA, 2015, pp. 91–100.

[15] J.W. Hegeman and S.V. Pemmaraju, Lessons from the congested clique applied to MapReduce, Proc 21st Int Colloq Structural Information and Communication Complexity, Takayama, Japan, 2014, pp. 149–164.

[16] J.W. Hegeman, S.V. Pemmaraju, and V. Sardeshmukh, Near-constant-time distributed algorithms on a congested clique, Proc 28th Int Symp Distributed Computing, Austin, TX, USA, 2014, pp. 514–530.

[17] S. Holzer and N. Pinsker, Approximation of distances and shortest paths in the broadcast congest clique, arXiv preprint arXiv:1412.3445 (2014).

[18] J. Kari, M. Matamala, I. Rapaport, and V. Salo, Solving the induced subgraph problem in the randomized multiparty simultaneous messages model. Proc 22nd Int Colloq Structural Information and Communication Complexity, Montserrat, Spain, 2015, pp. 370–384.

[19] C. Lenzen, Optimal deterministic routing and sorting on the congested clique, Proc 2013 ACM Symp Principles of Distributed Computing, Montreal, QC, Canada, 2013, pp. 42–50.

[20] Z. Lotker, E. Pavlov, B. Patt-Shamir, and D. Peleg, MST construction in O(log log n) communication rounds, Proc 15th ACM Symp Parallelism in Algorithms and Architectures, San Diego, California, USA, 2003, pp. 94–100.

[21] H. Mahmoud, Polya urn models, Chapman & Hall/CRC Texts in Statistical Science, Taylor & Francis, Boca Ratón, FL, USA, 2008.

[22] H.S. Na and A. Rapoport, Distribution of nodes of a tree by degree, Math Biosciences 6 (1970), 313–329.

[23] B. Patt-Shamir and M. Teplitsky, The round complexity of distributed sorting: extended abstract, Proc 30th Ann ACM Symp Principles of Distributed Computing, San Jose, CA, USA, 2011, pp. 249–256.

[24] D. Peleg, Distributed computing: A locality-sensitive approach, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2000.

[25] V. Raghavan and J.P. Spinrad, Robust algorithms for restricted domains, Proc 12th Ann Symp Discrete Algorithms, Washington, D.C., 2001, pp. 460–467.