



Non-Searchability of Random Scale-Free Graphs

Philippe Duchon, Nicole Eggemann, Nicolas Hanusse

► To cite this version:

Philippe Duchon, Nicole Eggemann, Nicolas Hanusse. Non-Searchability of Random Scale-Free Graphs. 9ème Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications, May 2007, Ile d'Oléron, France. pp.27-30. inria-00176940

HAL Id: inria-00176940

<https://inria.hal.science/inria-00176940>

Submitted on 4 Oct 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Non-Searchability of Random Scale-Free Graphs

Philippe Duchon¹, Nicole Eggemann² and Nicolas Hanusse¹

¹ LaBRI, CNRS, Univ. Bordeaux, 351, cours de la Libération, 33405 Talence, France.

² Dept. of Mathematics, Brunel University, UK.

Nous nous intéressons à la complexité de recherche d'un nœud donné dans un graphe sans-échelle, en utilisant uniquement des informations locales. Pour les modèles généralisant les modèles de Barabási-Albert et de Cooper-Frieze, nous prouvons une borne inférieure de $\Omega(n^{1/2})$ sur le nombre de sommets à visiter pour trouver le n -ième sommet inséré dans un modèle de connaissance locale faible.

In this paper, we investigate the efficiency of local search algorithms for some relatively unstructured networks. By local search, we mean that each vertex has a knowledge of its very close neighborhood, typically the identities, degrees and possibly the contents. Assuming there is no preprocessing step, we are interested in seeking a vertex of a given identity starting from any vertex. The time complexity of search is expressed as the number of vertices to explore before reaching the target or a neighbor of the target. For an arbitrary graph and a worst-case labelling of vertices, time complexity can be of the same order as the size of the graph itself. However, there is some hope that for large classes of graphs, the size of the exploration sequence would drastically decrease.

Recent literature exhibits non trivial properties shared by the topologies of numerous real-world distributed networks: the combination of a low diameter, a specific degree distribution of the vertices and a sparse network are typically representative of a *small-world* topology.

On the other hand, in a seminal paper, Kleinberg [Kle00] focuses on the ability of a user or a mobile agent to route along short paths within a class of random small-world graphs, the *navigable small-worlds*[†] using a greedy distributed algorithm (distances within a spanning subgraph are known or can be computed locally). Under some conditions, Kleinberg also proposes polynomial lower bounds valid for *any* distributed algorithm using the same knowledge of the network even if the diameter is polylogarithmic.

However, Kleinberg's model lacks an important property observed in real networks which are often *scale-free*. The degree distribution of the vertices tends to follow a power-law distribution, that is the number of vertices of degree δ is proportional to $n \left(\frac{1}{\delta}\right)^k$ for an n -vertex graph and k a constant strictly greater than one. Moreover, it has been observed that for real-life networks, k tends to belong to the range $[2, 3]$. For graphs built using Kleinberg's model, the degree distribution is very different (close to a Poisson distribution).

All local[‡] distributed search algorithms presented in the literature work as follows: given the set of visited vertices, one should choose the next vertex to explore. The process stops as soon as the target is reached. In this paper, the underlying network is a random scale-free labeled graph whose node identities belong to the range $[1, n]$. Each vertex knows its own identity and its degree and we assume two models of local knowledge: some information about the neighbors can be known (*strong model*) or not (*weak model*). The first model is the more realistic one but our starting point is the second model. Note that Kleinberg's model assumes more information than our strong model.

Up to now, it is still an open problem to know if scale-free graphs are *navigable* in the polylogarithmic sense. In this paper, we answer negatively to this question for some random scale-free graphs built using a

[†] Roughly speaking, a labeled small-world graph is *navigable* if a greedy distributed algorithm, using labels of the vertices, is able to route a message for any source and any target along a path of $O(\log^c n)$ vertices for c constant.

[‡] A distributed algorithm is *local* if vertices only know information on a close neighborhood

combination of uniform and preferential attachment processes, by providing a polynomial lower bound of $\Omega(n^{1/2})$ expected number of visited vertices before reaching the target for *any local distributed algorithm*.

Related works

Scale-free models In the following we summarize two families of scale-free graphs models which are related to our work: *pure random graphs* and *evolving graphs*.

In the pure random graph models, a random graph is built from a fixed degree distribution [MR95]. In such a model, it is important to notice that the degrees of neighbors are *independant*. However, the pure random graph models do not explain how a graph comes to have a power law degree sequence. The evolving models attempt to explain this emergence of the power law distribution.

In these models, one starts from a fixed small graph (a single vertex, loop, edge) and at each time step, a new vertex and an out-going edge are added. The target of the out-edge is chosen randomly either uniformly (*uniform attachment*) or with probability proportional to the total degree of the target (*preferential attachment*). Variants with higher out-degree can be obtained either by adding more edges per time step, or, say, by building an nm -vertex graph and merging every m consecutive vertices into one. In these models, the degree and age of a vertex are positively correlated. In particular, the degrees of neighbors are *not independant*, and mean-field analysis of the models tends to give incorrect results. This will make a real difference with the pure random graph models whenever we aim at analysing a search process. Most of these models based on preferential attachment and their properties can also be found in the work of Barabási and Albert [BA99] or Bollobás and Riordan [BR03].

Extensions to more general models combining uniform and preferential attachment have been studied in [Mór05, CF03]. In these papers, each new vertex chooses with respective probabilities p and $1 - p$ whether to use uniform or preferential attachment. Depending on the value of parameter p , the power law distribution can be observed and is proved by Cooper and Frieze [CF03] or by Móri [Mór05].

Distributed search in scale-free graphs As an input, Adamic *et al.* [ALPH01] takes random graphs in the random power law model whose exponent k is strictly between 2 and 3. They propose two strategies: a pure random walk and a search process based on high degree vertices. This last distributed algorithm works as follows: at each step, the next visited vertex is the highest degree neighbor of the set of visited vertices. Using a mean-field analysis of this greedy algorithm, the authors prove that the target is reached on average in $O(n^{2(1-2/k)})$ steps whereas for a pure random walk, the time complexity becomes $O(n^{3(1-2/k)})$ steps. Simulations on random graphs and sampling of the peer-to-peer network Gnutella confirms the tendency of the theoretical analysis. For a P2P context, in the same graph model, Sarshar *et al.* [SBR04] proposed a searching protocol based upon the contents replication along short random walks and the epidemic broadcast of the query: the query is spread to the neighbors of the current vertex with a given probability. They get a sublinear lookup protocol assuming that each content is duplicated a polynomial number of times. Other extensions in the last direction have been proposed in [FGL05] for clustered scale-free networks.

1 Models and Contribution

In this section, we concentrate on random graph models for which rigorous degree distribution results are known, and which are connected by construction, since we want our searching processes to be able to terminate with probability 1. All graphs are constructed as oriented graphs, but searching always takes place in the corresponding unoriented graph.

Graph models The *Móri model* G_t of random trees starts, at time $t = 2$, with two vertices 1, 2 and a single edge between them; then, at each later time, a new vertex t is added, together with a single outgoing edge to an older vertex u , selected with probability p preferentially (based on indegree), or, with probability $1 - p$, uniformly, that is u is chosen with probability proportional to $pd_t(u) + (1 - p)$, $d_t(u)$ being the indegree of u at time t . To get the m -out Móri graph of size n , $G_t^{(m)}$, take the Móri tree of size nm and, for each $1 \leq i \leq n$, merge vertices $m(i - 1) + 1$ to mi into a new vertex i .

The *Cooper-Frieze model* is a very general model of evolving graphs, which is defined by several real parameters $\alpha, \beta, \gamma, \delta$ and two discrete probability distributions p, q . We refer to the paper [CF03] for a

precise definition, and simply give a short informal description of the evolution rules: at each time step, one randomly chooses whether to apply procedure **New** (with probability α) or procedure **Old** (with probability $1 - \alpha$); procedure **New** will add a new vertex and a random number (governed by distribution q) of outgoing edges, while procedure **Old** will add a random number (governed by distribution p) of new outgoing edges to a randomly selected existing vertex. Parameters β , γ and δ control probabilities that additional choices of vertices and endpoints are done preferentially or uniformly.

Modeling the searching process We aim at proving lower bounds on the complexity of searching for a given vertex in a random graph, using only local information for two models: *weak* model, and a *strong* model. In both models, the searching process has access to a list of already discovered vertices (initially reduced to a single vertex), each with its degree and a list of incident edges. At each time step, the searching process can try to discover a new vertex by making a request. In the *weak* model, a request is in the form of a pair (u, e) , where u is an already discovered vertex, and e is an edge incident to u . The answer to the request is the identity v of the other endpoint of edge e , together with the list of all edges incident to v . In the *strong* model, a request is in the form of a vertex u that is adjacent to an already discovered vertex, and the answer consists of the list of vertices adjacent to u , together with their respective lists of incident edges. Our measure of performance is the *number of requests* made prior to stopping.

We are also not making any assumptions on the computing power or memory requirements of the searching process. Since we deal with lower bounds, any results are valid under more restrictive assumptions.

Our contribution We concentrate on two models of evolving graphs mixing preferential and uniform attachment: the Móri model and the Cooper-Frieze model, which we rephrase in Section 1 to use *indegree* of vertices instead of total degree (this makes it possible to explore a wider range of parameters) Our main results are as follows and we will only sketch the probabilistic tools we use for Theorem 1.

Theorem 1. *For any $m \geq 1$, in the merged Móri tree with parameters m and p ($0 < p \leq 1$), no searching algorithm operating in the weak model can find a path to vertex n with an expected number of requests less than $\Omega(n^{1/2})$.*

For $p < 1/2$, no searching algorithm operating in the strong model can find a path to vertex n with an expected number of requests less than $\Omega(n^{1/2-p-\epsilon})$, for any $\epsilon > 0$.

Theorem 2. *In all Cooper-Frieze models with $0 < \alpha < 1$, no searching algorithm can find a path to vertex n , using only local information in the weak model, with an expected number of requests less than $\Omega(n^{1/2})$.*

2 Vertex equivalence and application to the Móri graph

Our proofs of lower bounds are based on a probabilistic notion of vertex equivalence. Intuitively, vertices are equivalent if their identities can be exchanged without modifying the probability distribution on graphs.

Definition 1. *Let G be some graph on the vertex set $[1, n] = \{1, \dots, n\}$, and $\sigma \in S_n$ some permutation of $[1, n]$. We write $\sigma(G)$ for the graph on the same vertex set that is obtained by applying permutation σ on the vertices of G .*

Definition 2. *Let G be some random graph model on the vertex set $[1, n]$, and $V \subset [1, n]$.*

We say that the vertices in V are (probabilistically) equivalent if, for any $\sigma \in S_V$, the random graphs G and $\sigma(G)$ have the same probability distribution.

If \mathcal{E} is some event, we say that the vertices in V are equivalent conditional on \mathcal{E} if, for all $\sigma \in S_V$, G and $\sigma(G)$ have the same probability distribution, conditional on \mathcal{E} .

We focus on finding, in a suitable (faithful) model for the construction of our random graph, an event \mathcal{E} with a large enough probability such that our equivalence property holds conditional on \mathcal{E} . Lemma 1 is the key to our lower bounds whereas Lemma 2 defines the conditional equivalence $\mathcal{E}_{a,b}$ and Lemma 3 approximates the probability of $\mathcal{E}_{a,b}$.

Lemma 1. *If a set V of vertices is equivalent conditional on some event \mathcal{E} , then any search process operating in the weak model, starting from any vertex u and searching for any vertex $v \in V$ ($v \neq u$), has an expected complexity of at least $|V| \mathbb{P}(\mathcal{E})/2$.*

Lemma 2. For any integers a, b, t with $a \leq b \leq t$, in G_t , the vertex set $V = [[a+1, b]]$ is equivalent conditional on the event

$$\mathcal{E}_{a,b} = \bigcap_{a < k \leq b} \{N_k \leq a\},$$

where N_k denotes the father of vertex k (that is, the destination of its outgoing edge).

Lemma 3. Let $b = a + \lfloor (a-1)^{1/2} \rfloor$. Then $\mathbb{P}(\mathcal{E}_{a,b}) \geq e^{-(1-p)}$.

We are now able to sketch the proof of Theorem 1 for $m = 1$ (for $m > 1$, the extension is straightforward). Combining Lemmas 2 and 3, we see that, in G_t (for $t \geq n + \sqrt{n-1}$), the vertices in $[[n, n + \sqrt{n-1}]]$ are equivalent conditional on an event of probability bounded away from 0. Lemma 1 then concludes the proof for the weak model. The strong model case then follows from Móri's result [Mór05] that the maximum degree in graph G_t is of order t^p , which implies that, with high probability, this maximum degree is less than $t^{p+\varepsilon}$. Any algorithm operating in the strong model can be simulated in the weak model by replacing each request about vertex u with requests about all edges incident to u , which gives a slowdown factor of at most the maximum degree. \square

3 Conclusion and open problems

We have proved, in the weak model of local information, a polynomially high lower bound on the time required to find a given vertex in the two considered models of scale-free graphs. This is in contrast with the logarithmic diameter of such graphs, proved in expectation and with high probability. Thus, these graph models do not have the easy searchability “small world” property as exhibited by several authors in different models of random graphs. The technique we used seems broad enough to be adapted to other models of growing random graphs (Due to space limitations, we are not able to present the more complex setup of Cooper Frieze graphs but the starting point is still the existence of a set of $\Theta(\sqrt{n})$ equivalent vertices).

Our results carry over somehow weakly to the strong model of local information, whenever the maximum degree can be proved to be significantly smaller than $n^{1/2}$. Unfortunately, most rigorous results concerning the maximum degree of scale-free graphs (notably the ubiquitous Barabási-Albert model and its generalizations based on the preferential attachment with total degree) yield a maximum degree that is larger than this limit, making our upper bound trivial.

References

- [ALPH01] L. Adamic, R. Lukose, A. Puniyani, and B. Huberman. Search in power law networks. *Physical Reviews E*, 64(046135):1–8, 2001.
- [BA99] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.
- [BR03] Béla Bollobás and Oliver Riordan. Mathematical results on scale-free random graphs. In *Handbook of graphs and networks*, pages 1–34. Wiley-VCH, Berlin, 2003.
- [CF03] Colin Cooper and Alan M. Frieze. A general model of web graphs. *Random Struct. Algorithms*, 22(3):311–335, 2003.
- [FGL05] Pierre Fraigniaud, Philippe Gauron, and Matthieu Latapy. Combining the use of clustering and scale-free nature of user exchanges into a simple and efficient p2p system. In *Euro-Par*, volume 3648 of *Lecture Notes in Computer Science*, pages 1163–1172, 2005.
- [Kle00] Jon Kleinberg. The small-world phenomenon: An algorithmic perspective. In *Proceedings of the ACM Symposium on Theory of Computing (STOC)*, pages 163–170, 2000.
- [MR95] Michael Molloy and Bruce Reed. A critical point for random graphs with a given degree sequence. *Random Structures and Algorithms*, 6(2-3):161–179, 1995.
- [Mór05] Tamás F. Móri. The maximum degree of the Barabási-Albert random tree. *Combinatorics, Probability and Computing*, 14:339–348, 2005.
- [SBR04] Nima Sarshar, Oscar Boykin, and Vwani P. Roychowdhury. Percolation search in power law networks: Making unstructured peer-to-peer networks scalable. In *Proc. of the IEEE International Conference on Peer-to-Peer Computing (P2P'04)*, pages 2–9, 2004.