Network Discovery in Random Graphs*

Thomas Erlebach† Alexander Hall‡ Matúš Mihaľák†

Abstract

We study the number of queries at randomly selected nodes that are needed for approximate network discovery. For the approximate discovery of Erdős-Rényi random graphs $G_{n,p}$ in the layered graph query model, we show that a constant number of queries is sufficient if p is a constant, but $\Omega(n^{\alpha})$ queries are needed if $p = n^{\varepsilon-1}$ for arbitrarily small $\varepsilon > 0$, where $\alpha > 0$ is a constant depending only on ε.

A fundamental problem in the study of complex networks is how to obtain accurate information about the topology of a network using a limited number of measurements or observations. For example, attempts to map the Internet can be based on traceroute experiments [3] or on the analysis of BGP routing tables [5]. A simplified theoretical model of such network discovery settings, the so-called *layered graph query model*, has been introduced in [1]. The goal is to discover the edges and non-edges (for $u, v \in V$, we call $\{u, v\}$ a non-edge if it is not an edge of the graph) of an unknown graph or network $G = (V, E)$ using queries; a query at a node v reveals all edges and non-edges whose endpoints have different distance from v. Preliminary simulation experiments with (scale-free as well as Erdős-Rényi) random graphs reported in [4] indicate that the number of queries needed to discover all edges and non-edges typically grows with the size of the graph, as expected, but in some cases appears to be bounded by a small constant independent of the size of the graph if only a large fraction (say, 95%) of the edges and of the non-edges needs to be discovered. This shows that for the practically relevant goal of approximate network discovery, a surprisingly small number of queries is often sufficient. Now we study this phenomenon analytically for Erdős-Rényi random graphs $G_{n,p}$. These are graphs on n nodes in which each possible edge is present independently with probability p. We consider the simple query strategy that selects the query nodes uniformly at random. We say that a set of random queries approximately discovers $G_{n,p}$ in expectation, if the expected number of edges discovered by the queries is at least $\rho|E|$, and the analogous condition is satisfied for non-edges. Here, ρ is a constant such as 0.95. Surprisingly, we find that if p is a constant strictly between 0 and 1 (i.e., if we consider dense $G_{n,p}$ graphs), then a constant number of query nodes is sufficient to approximately discover $G_{n,p}$ in expectation, but if $p = n^{\varepsilon}/n$ for an arbitrarily small constant $\varepsilon > 0$, then $\Omega(n^{\alpha})$ queries are necessary, where $\alpha > 0$ is a constant depending on ε. Our results show that the number of random queries needed to approximately discover $G_{n,p}$ depends on the density of the graph, and in the query model considered it is actually easier to discover dense random graphs than relatively sparse ones.

Our analysis for constant p is based on the observation that the probability that a query at node q discovers an edge or non-edge $\{u, v\}$ is at least $2p(1 - p)$, the probability that q is adjacent to one of u, v but not the other. For the case of $p = n^{\varepsilon}/n$, we use bounds from [2] on the size of the i-neighborhood and on the size of the i-th breadth-first search layer of a node in $G_{n,p}$. These bounds allow us to show that for an edge or non-edge $\{u, v\}$, a query node q is very likely to have the same distance from u and v (and thus does not discover the edge or non-edge). It would be interesting to extend the analysis to scale-free random graphs.

References

[1] Z. Beerliová, F. Eberhard, T. Erlebach, A. Hall, M. Hoffmann, M. Mihaľák, and L. S. Ram. Network discovery and verification. In *Proceedings of the 31st International Workshop on Graph-Theoretic Concepts in Computer Science (WG'05)*, LNCS 3787, pages 127–138. Springer-Verlag, 2005.

[2] F. Chung and L. Lu. The diameter of random sparse graphs. *Advances in Applied Math*, 26:257–279, 2001.

[3] DIMES. Mapping the Internet with the help of a volunteer community. http:// www.netdimes.org/.

[4] F. Eberhard, T. Erlebach, and A. Hall. A simulation study of network discovery strategies. In *European Conference on Complex Systems 2005 (ECCS'05)*. Proceedings CD, 2005.

[5] Oregon RouteViews. University of Oregon RouteViews project. http://www.routeviews.org.

*Work partially supported by European Commission – Fet Open project DELIS IST-001907 Dynamically Evolving Large Scale Information Systems, for which funding in Switzerland is provided by SBF grant 03.0378-1.

†Department of Computer Science, University of Leicester, Leicester LE1 7RH, UK. {te17, mm215}@mcs.le.ac.uk

‡Institute for Theoretical Computer Science, ETH Zurich, CH-8092 Zurich, Switzerland. alex.hall@inf.ethz.com