Network Discovery and Verification[⋆]

Zuzana Beerliova[1], Felix Eberhard[2], Thomas Erlebach[2], Alexander Hall[1], Michael Hoffmann[2], Matúš Mihaľák[2] and L. Shankar Ram[1]

[1] Department of Computer Science, ETH Zürich
{bzuzana,mhall,lshankar}@inf.ethz.ch
[2] Department of Computer Science, University of Leicester
{te17,mh55,mm215}@mcs.le.ac.uk

Abstract. Consider the problem of discovering (or verifying) the edges and non-edges of a network, modelled as a connected undirected graph, using a minimum number of queries. A query at a vertex v discovers (or verifies) all edges and non-edges whose endpoints have different distance from v. In the network discovery problem, the edges and non-edges are initially unknown, and the algorithm must select the next query based only on the results of previous queries. We study the problem using competitive analysis and give a randomized on-line algorithm with competitive ratio $O(\sqrt{n \log n})$ for graphs with n vertices. We also show that no deterministic algorithm can have competitive ratio better than 3. In the network verification problem, the graph is known in advance and the goal is to compute a minimum number of queries that verify all edges and non-edges. This problem has previously been studied as the problem of placing landmarks in graphs or determining the metric dimension of a graph. We show that there is no approximation algorithm for this problem with ratio $o(\log n)$ unless $\mathcal{P} = \mathcal{NP}$. Furthermore, we prove that the optimal number of queries for d-dimensional hypercubes is $\Theta(d/\log d)$.

Keywords. Metric dimension, landmarks in graphs, set cover, on-line algorithm

1 Introduction

In recent years, there has been an increasing interest in the study of networks whose structure has not been imposed by a central authority but arisen from local and distributed processes. Prime examples of such networks are the Internet and unstructured peer-to-peer networks such as Gnutella. For these networks, it is very difficult and costly to obtain a "map" providing an accurate representation of all nodes and the links between them. Such maps would be useful for many purposes, e.g., for studying routing aspects or robustness properties of these networks.

In order to create maps of the Internet, a commonly used technique is to obtain local views of the network from various locations (vantage points) and combine them into a map that is hopefully a good approximation of the real network [1,2]. More generally, one can view this technique as an approach for discovering the topology of

[⋆] Research supported by European Commission Integrated Project DELIS (Dynamically Evolving Large Scale Information Systems), IST-001907.

Dagstuhl Seminar Proceedings 05031
Algorithms for Optimization with Incomplete Information
http://drops.dagstuhl.de/opus/volltexte/2005/59

an unknown network by using a certain type of queries—a query corresponds to asking for the local view of the network from one specific vantage point. In this paper, we formalize *network discovery* as a combinatorial optimization problem whose goal is to minimize the number of queries required to discover all edges and non-edges of the network. We study the problem as an on-line problem using competitive analysis. Initially, the network is unknown to the algorithm. To decide the next query to ask, the algorithm can only use the knowledge about the network it has gained from the answers of previously asked queries. In the end, the number of queries asked by the algorithm is compared to the optimal number of queries sufficient to discover the network. In this paper, we consider a query model in which the answer to a query at a vertex v consists of all edges and non-edges whose endpoints have different (graph-theoretic) distance from v.

In the off-line version of the network discovery problem, the network is known to the algorithm from the beginning. The goal is to compute a minimum number of queries that suffice to discover the network. Although an algorithm for this off-line problem would not be useful for network discovery (if the network is known in advance, there is no need to discover it), it could be employed for network verification, i.e., for checking whether a given map is accurate. Therefore, we refer to the off-line version of network discovery as *network verification*. For network verification, we are interested in polynomial-time optimal or approximation algorithms.

1.1 Motivation

As mentioned above, the motivation for our research comes from the problem of discovering information about the topology of communication networks such as the Internet or peer-to-peer networks. The query model that we study is motivated by the following considerations. First, notice that our query model can be interpreted in the following way: A query at v yields the shortest-path subgraph rooted at v, i.e., the set of all edges on shortest paths between v and any other vertex. To see that this is equivalent to our definition (where a query yields all edges and non-edges between vertices of different distance from v), note that an edge connects vertices of different distance from v if and only if it lies on a shortest path between v and one of these two vertices. Furthermore, the shortest-path subgraph rooted at v implicitly confirms the absence of all edges between vertices of different distance from v that are not part of the shortest-path subgraph.

It is clear that our model of network discovery is a simplification of reality. In real networks, routing is not necessarily along shortest paths, but may be affected by routing policies, link qualities, or link capacities. Furthermore, routing tables or traceroute experiments will often reveal only a single path (or at most a few different paths) to each destination, but not the whole shortest-path subgraph. Nevertheless, we believe that our model is a good starting point for a theoretical investigation of fundamental issues arising in network discovery.

1.2 Related Work

We are not aware of any theoretical work on network discovery problems using query models similar to the one we consider in this paper. Graph discovery problems have been studied in distributed settings where one or several agents move along the edges of the graph (see, e.g., [3]); the problems arising in such settings appear to require very different techniques from the ones in our setting.

It turns out, however, that the network verification problem has previously been considered as the problem of placing landmarks in graphs. Here, the motivation is to place landmarks in as few vertices of the graph as possible in such a way that each vertex of the graph is uniquely identified by the vector of its distances to the landmarks. This problem has been studied by Khuller et al. in [4]. They prove that the problem is $\mathcal{NP}$-hard in general (by reduction from 3-SAT) and give an $O(\log n)$-approximation algorithm for graphs with n vertices, based on SetCover. For trees, they show that the problem can be solved optimally in polynomial time. For d-dimensional grids, $d \geq 2$, they show that d landmarks suffice and claim that d landmarks are also necessary; we will show that this lower bound is incorrect in the case of d-dimensional hypercubes (grids with side length 2). Furthermore, they prove that one landmark is sufficient if and only if G is a path, and discuss properties of graphs for which 2 landmarks suffice. They also show that if k landmarks suffice for a graph with n vertices and diameter D, we must have $n \leq D^k + k$. The smallest number of landmarks that are required for a given graph G is also called the *metric dimension* of G [5]. For further known results about this concept, we refer to the survey [6]. Results for the problem variant where extra constraints are imposed on the set of landmarks (e.g., connectedness or independence) are surveyed in [7].

1.3 Our Results

For network discovery, we give a lower bound showing that no deterministic on-line algorithm can have competitive ratio better than 3, and we present a randomized on-line algorithm with competitive ratio $O(\sqrt{n \log n})$ for networks with n nodes. For the network verification problem, we prove that it cannot be approximated within a factor of $o(\log n)$ unless $\mathcal{P} = \mathcal{NP}$, thus showing that the approximation algorithm from [4] is best possible (up to constant factors). We also give a useful lower bound formula for the optimal number of queries of a given graph, and we show that the optimal number of queries for d-dimensional hypercubes is $\Theta(d/\log d)$.

2 Directions for Future Work

For the network discovery problem, a significant gap remains between our randomized upper bound of $O(\sqrt{n \log n})$ and the small constant lower bounds. Thus, the major problem left open by our work is to close this gap. Another interesting question is finding a deterministic construction of query sets of size $O(d/\log d)$ for hypercubes.

The subject of our study can be seen as one example of a family of problem settings in which the goal is to discover or verify information about a graph using certain kinds of queries. Different problems are obtained if the query model is varied, or if the objective is changed. Other natural query models are, for example, that a query at v returns only the distances from v to all other vertices of the graph; that a query is specified by two vertices u and v, and returns the set of all edges on shortest paths (or just one shortest path) between u and v; or that a query returns an arbitrary shortest-path tree rooted at v. Concerning the objective, the goal could be to discover or verify a certain graph parameter such as the diameter, the average path length, or the independence number. One could also relax the requirement and only ask for an approximate answer. For example, one could pose the problem of minimizing the number of queries required to approximate the average path length of the graph within a factor of $1 + \varepsilon$.

We believe that the study of such problems could be a fruitful area of research with applications in the monitoring and analysis of communication networks such as the Internet or peer-to-peer networks.

References

1. Barford, P., Bestavros, A., Byers, J., Crovella, M.: On the marginal utility of deploying measurement infrastructure. In: Proceedings of the ACM SIGCOMM Internet Measurement Workshop 2001. (2001)
2. Subramanian, L., Agarwal, S., Rexford, J., Katz, R.: Characterizing the Internet hierarchy from multiple vantage points. In: Proceedings of INFOCOM'02. (2002)
3. Bender, M.A., Slonim, D.K.: The power of team exploration: Two robots can learn unlabeled directed graphs. In: Proceedings of the 35th Annual IEEE Symposium on Foundations of Computer Science (FOCS'94). (1994) 75–85
4. Khuller, S., Raghavachari, B., Rosenfeld, A.: Landmarks in graphs. Discrete Applied Mathematics **70** (1996) 217–229
5. Harary, F., Melter, R.: The metric dimension of a graph. Ars Combinatorica (1976) 191–195
6. Chartrand, G., Zhang, P.: The theory and applications of resolvability in graphs: A survey. Congressus Numerantium **160** (2003) 47–68
7. Saenpholphat, V., Zhang, P.: Conditional resolvability in graphs: A survey. Inernational Journal of Mathematics and Mathematical Sciences **38** (2004) 1997–2017