Adaptive Geographical Search in Networks

Andrea Wiggins

March 5, 2007

1 Introduction

Problems related to geographical search in networks arise in a variety of fields; searching a network can be a very inefficient task, and the goal of research into geographical search in networks is to identify better strategies for finding the shortest (geodesic) paths through a network. The varying properties of networks yield different structures that are more and less conducive to known search techniques and strategies. For example, the typical small world network has a higher link density that improves its searchability over a sparsely connected network. Nearly identical networks with different ratios of long-range and short-range links may cause a simple "greedy" algorithm to perform very poorly in some cases and very well in others.

While useful and interesting results have been achieved with studies into the performance of search algorithms on networks of varying topological structures, at most a handful of similar algorithms are tested and usually only on a relatively limited number of network data sets. By testing a much broader spectrum of network topologies with an adaptive algorithm, I hope to identify more efficient strategies for a much greater variety of network archetypes. This will be achieved through the creation of a large number of search strategy rules based on elements from well-known, successful algorithms combined with adaptive learning processes and rule discovery mechanisms.

2 Proposal

2.1 Experimental Environments

This research will be conducted as a series of experiments using computer simulation. The simulation requires the generation of environments and agents, where the environment is determined by the structure of the graph being explored by the agents. For the purposes of an initial study, the use of Erdös-Rényi random graphs will be the control group. Experimental treatments can be widely varied, and will begin with well-known graph structures such as small-world networks, for which successful search algorithms have been established. By beginning with these structures, I can validate the performance of the adaptive agents before moving on to less thoroughly investigated network topologies. In this simulation, graph structures will be laid out on a lattice upon which nodes and links are placed. The underlying lattice structure allows a uniform concept of link distance and direction across many network graph models.

After exploring the best-known models, such as Erdös-Rényi random graphs and small-world networks, I will study numerous graphs generated by the network growth models in the literature, and will examine these as thoroughly as possible by testing a wide variety of parameters for each model. The study's utility can be further extended by using the characteristic parameters from many different network growth models to create a network growth "gene pool" from which more network topologies can be generated for testing. Due to the combinatorial explosion of generated network models that could be created, the research must also determine a method of choosing which networks to test based on potential applicability to real-world problems.

2.2 Agents and Interactions

Agents in this simulation will be made up of combinations of graph traversal rules with different weights; each agent is essentially a new search algorithm made up of genes contributed by elements from graph theory and search algorithms that are currently in use. The agents' genetic structure determines their movement on the graph, and the diversity of the population of agents that can be tested on a series of graph environments in simulation takes full advantage of the parallel processing that allows a more efficient approach to a complex problem. The agents must also possess a memory of the nodes they have traversed, as a basic condition to enable backtracking and prevent infinite loops. There are multiple ways to achieve this; most approaches "color" visited nodes so that the agent is able to recognize a previously-visited position in the network based on the color of the nodes.

The agents are subject to rules of global interaction, which describes the actions of the agents in the experimental trials. The goal of each agent is to reach a target node, beginning from a starting node, and it must travel along established links between the nodes in the network. The agents begin their traversal from a randomly selected node in the network, and at all times they know the relative position of the goal node, expressed as a directional vector based on the current location on the graph. The agents' interactions are limited by the condition of limited information that is relevant to most search problems. That is to say, an agent cannot ask its neighbor nodes for the complete path to the goal node, which would be a trivial solution; these agents must operate in an environment in which they can only know how many neighbors exist for each node adjacent to its current position, and the difference in direction between the neighbor nodes and the target node.

2.3 Rules of Graph Traversal

The rules in this system are the agents' genetic makeup, dictating their choices of moves in a graph. Each agent is composed of a different rule set with different weights on the rules. At each time step, each agent will fully traverse a graph, adjusting the weights of its rules based on their contribution to the goal of reaching a target node.

To promote new rule discovery, agents duplicate and recombine after each traversal with a probability based on the length of their traversal with respect to the other agents traversals in the same time step. Agents that were able to swiftly traverse the graph will generate recombined offspring at a higher rate than agents who required many steps to find the target node. This strategy of recombination serves as an implicit fitness function, where fitness is determined within the context of the simulation by comparison to the performance of the other agents traversing the same graphs.

Each series of traversals that make up each experimental condition are performed on newly generated graphs for each traversal, where the graphs are stochastically similar or identical for key properties such as link density or ratio of short- and long-range links. This allows the rapid testing of many possible search algorithms without the risk of local optimization based on a single graph; instead, the agents will evolve search strategies based on traversing many graphs that are topologically similar but not identical.

3 Expected Outcomes

Geographic search in networks is a problem that is well-suited to an approach that emphasizes learning over optimization. The goal of this research is to generate search algorithms that function well for the network topology of a class of graphs, rather than a single algorithm optimized to a single graph. The problem of geographical search in networks can be approached like a checkersplayer game, where each agent is a search algorithm made up of weighted traversal strategies.

Through rule weighting and recombination, the population of agents are expected to converge to a few good algorithms for each graph tested. The rules and weights that make up successful algorithms are expected to vary between different graph types; it is also expected that currently popular search algorithms will be discovered as agent combinations, and quite probably surpassed.

Besides the many applications of the search algorithms to practical problems of information retrieval that can be generated through this simulation, future work will also explore which search strategy characteristics are successful for particular graphs, and why. I expect that the best-of-breed algorithms for groups of graph types will have many common characteristics, but will show wide diversity in weighting and exactly which characteristics are held in common. Finally, I expect to be surprised by emergent patterns; this research proposes to explore the process of exploration itself, and a lack of novelty would come as a surprise.