Preferential Acquisition of Scale-Free and Social Network Graphs

Jason Jeffrey Jones (jasonj@ucsd.edu)
Department of Psychology, 9500 Gilman Drive #0109
La Jolla, CA 92037 USA

Abstract

Human subjects were tasked with learning the structure of formal graphs. Deep structure and surface features were varied independently. It was found that subjects were faster to acquire knowledge of scale-free graphs as opposed to random graphs. Subjects also learned the structure of a graph more quickly when it was described as a social network as opposed to a computer network.

Keywords: learning; social networks

Introduction

Formal graphs are flexible, abstract structures that can be used to represent phenomena as diverse as the physical form of organic molecules and the sexual history of Swiss survey-takers. The more quickly people can acquire knowledge about the structure of a novel formal graph, the more efficient they will be at asking and answering questions about the phenomenon it represents. It would thus be useful to know what conditions lead to the most efficient acquisition of graph structure.

How people learn the structure of graphs in general is too broad a research question for one paper, however. Therefore, in this paper I use the concept of a social network to generate and test hypotheses concerning graph learning.

While popular attention to social networks is new (Fincher, 2010), the study of acquisition of social network structure is not. Many researchers have examined the process by which social tie information is learned (e.g. DeSoto, 1960; Krackhardt & Kilduff, 1999; Zajonc & Burnstein, 1965).

Two examples of reliable findings in this field are (1) the expectation that friendships are reciprocal and (2) the expectation that friendships are transitive. People expect the fact that A is friends with B to imply that B is friends with A. They also expect that if A is friends with B and B is friends with C, that A and C are also friends. These expectations have both the salutary effect of causing networks that conform to these beliefs to be easily learned and the harmful effect of causing people to incorrectly infer that such connections exist when they do not (DeSoto, 1960; DeSoto, Henley & London, 1968).

The experimental research to date has focused on the training of small networks – typically in the range of 3-5 individuals. Networks of this size allow for both tight control of the variables of interest (e.g. transitivity) and easy implementation of a training regimen (often carried out by the researcher interacting with the subject directly or through a set of flash cards).

In the present study, I take advantage of advances in technology to train subjects on larger and more complex networks. I use computer programming to generate and analyze networks that would be unwieldy to work with by hand. I deliver the experiment online to reach a large number of subjects efficiently. With these advantages, I am able to approach the study of social network structure acquisition from a different perspective than past work.

Hypotheses

I test two hypotheses concerning the learning of graph structure. The first is that the rate of learning depends upon the type of graph to be learned: *Human subjects will acquire a network structure more quickly if it resembles a true human social network rather than an arbitrary network.* To operationalize this, I will measure learning curves as subjects learn the structure of scale-free or random graphs.

The second hypothesis is that the rate of learning depends upon the surface description given to a network: *Human subjects will acquire a network's structure more quickly if it is framed as a social network* as opposed to the same network framed in some other manner (e.g. a computer or transport network).

To test these hypotheses, I developed an experiment in which subjects were tasked with learning the structure of a graph. The deep structure of the graph and its surface features were manipulated independently.

Graph Structure Manipulation. Subjects learned the structure of a Random Graph or a Scale-Free graph.

The Random Graph was generated using the model of Erdős and Rényi (1960). To form links, two nodes are chosen at random and a link formed between them.

The Scale-Free Graph was generated using the model of Barabási and Albert (1999). In this process, one node at a time is added to the network, and links are formed to other nodes based on the number of links those existing nodes already have. Nodes with more exisiting links are preferred over nodes with fewer existing links. This preferential attachment results in the power-law degree distribution characteristic of a scale-free graph.

Surface Feature Manipulation. Subjects learned the structure of a Social Network or a Computer Network.

The Social Network was described as a class of 20 students. Subjects were asked to learn "Who is friends with whom?"

The Computer Network was described as a collection of 20 computer servers. Subjects were asked to learn "Which computers are connected to each other by fiber optic wires?"

Method

68 UCSD undergraduates participated in the experiment.

Subjects were trained and tested on four separate networks, spending about twelve minutes on each. The experiment was administered online through a PHP/mySQL/Flash Web application.

The design was a 2 (Graph Structure, within) x 2 (Surface Description, between) design.

In the Social Network condition, subjects were told they were to learn the friendship ties in a group of 20 students. In the Computer Network condition, subjects were told they were to learn which computer servers in a collection of 20 were connected by fiber optic wires.

All subjects were trained on two Random graphs and two Scale-Free graphs.

Training. In a training trial, the subject saw the name of one student (or computer) at the top of the screen, the words "is friends with" (or "is connected to") just below, and below that a list of the connected nodes (i.e. friends or connected servers). This information was displayed for three seconds. A blank white screen followed for 500 milliseconds and then a test trial began.

Test. In a test trial, two student (or computer) names were presented above the question "Friends?" (or "Connected?"). Below this, two equally-sized buttons were labeled YES and NO. The experiment waited for a response before moving on to the next training trial.

Feedback was provided after the subject made a response. The button for the correct answer was highlighted, and a 300 millisecond sound clip of a bell was played for a correct response or a buzzer for an incorrect response.

Training and Test trials were interleaved. Subjects completed 80 test trials for each of the four networks they learned.

For each network, a matching number of valid (existing) and invalid (non-existent) edges were generated. A random number on each test trial determined if the subject would be tested with a valid or invalid edge. These edges were sampled without replacement until all had been shown. At that point, all of the constructed test edges became available again for re-testing. This procedure was followed so that responses should be distributed evenly between YES and NO, even though most edges in the graph were invalid.

Results

The simplest measure to use to characterize the performance of subjects is trials-to-criterion. I use a criterion of 40 correct trials. An agent with perfect knowledge of the network could reach the criterion in 40 trials. An agent with no knowledge and no ability to learn would simply guess (with a 50% chance of being correct on each trial) and would be expected to reach the criterion in the neighborhood of 80 trials. Of course, subjects reached criterion somewhere within this range. (In the few cases in which a subject did not reach the criterion within the 80 test trials, the subject was assigned the maximum score of 80.)

The trials to criterion measure for each of the 4 cells in the design is plotted in Figure 1. A mixed-design ANOVA with Greenhouse-Geisser correction confirmed a significant main effect for Graph Structure $F(1,66) = 17.94$, $p < .001$, a significant main effect for Surface Description $F(1,66) = 13.42$, $p < .001$, and a marginally significant interaction $F(1,66) = 2.98$, $p < .10$.

Plotted in Figure 2 is the percentage of incorrect responses as a function of trial for Random and Scale-Free Graphs. The best-fit power-law curve is fit to each set of points separately. The main effect of Graph Structure is evident.

Plotted in Figure 3 is the percentage of incorrect responses as a function of trial for Social and Computer Networks. The best-fit power-law curve is fit to each set of points separately. The main effect of Surface Description is evident.

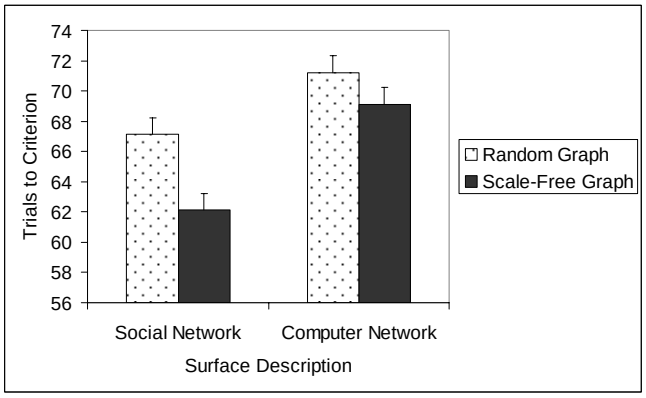

Figure 1: Trials-to-Criterion by Graph Type and Surface Description. (Lower numbers indicate faster learning.)

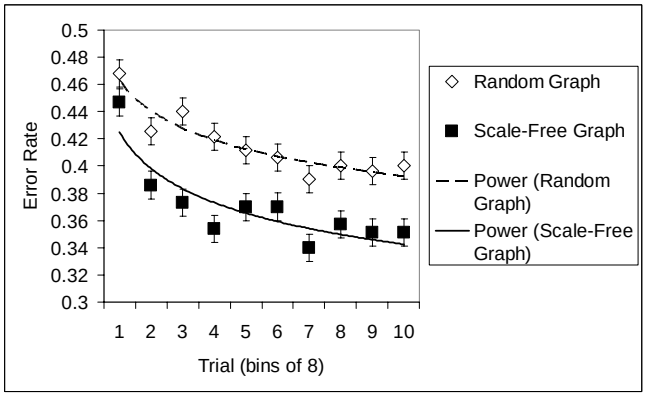

Figure 2: Error Rate as a Function of Trial for Random and Scale Free Graph Structure Acquisition.

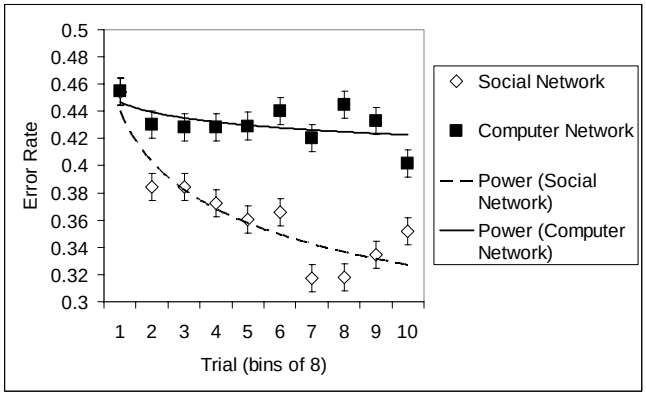

Figure 3: Error Rate as a Function of Trial for Social and Computer Network Graph Structure Acquisition.

Discussion

Clear effects of deep structure and surface description are evident. Scale-free graph structure is acquired more efficiently than random graph structure. Given similar graph structures, it is always beneficial to view the graph as a social network as opposed to a computer network. In addition, the factors interact such that the benefit of learning a scale-free graph is diminished when it is viewed as a computer network.

The meaning of all these effects is mostly a matter for speculation until further research is undertaken. Nonetheless, I offer some interpretation and suggestions as to the implications.

To some degree, it should not be surprising that subjects outperform themselves when learning scale-free versus random graphs. Learning one edge in a scale-free graph gives some information (in the Shannon sense) as to the probabilities of other edges, whereas this is not the case for an edge in a random graph. The whole of the advantage for scale-free graph acquisition may be in subjects' ability to learn which nodes are "popular" and which are not.

Secondly, scale-free graphs – due to the method of their construction – are more likely to conform to the assumptions subjects bring to social network learning. In this experiment, all friendships are reciprocal (A→B implies B→A and B→A implies A→B), so there is no advantage in that respect. Transitive relationships, however, are more likely to develop in scale-free graphs (with their core of highly-connected nodes) than in random graphs where a transitive relationship would have to form by the confluence of three independent chance events.

A third, intriguing explanation is that subjects have observed many scale-free human social networks in their lives and bring this assumption to this experiment. They know that some people are more desirable as friends than others and that this will cause the preferential attachment evident only in the scale-free networks in this experiment. If the subjects make this assumption specifically about people (and not about computer servers), that would explain the interaction between factors observed in the data.

Why the difference in rates of acquisition for social networks versus computer networks? Two explanations are possible here. The first has to do with the stimuli within this experiment. In the Social Network condition, nodes are labeled with names pulled from the Social Security Administration's list of the most popular baby names in the 1990s. In the Computer Network condition, the nodes are labeled with computer server names constructed by concatenating a uniquely identifying first letter (A, B, C … T) with a string of random numbers and a ".com" or ".net" suffix. It is possible these stimuli were simply more difficult to form associations with because they were not pronounceable or because they were less familiar. It should be noted, however, that due to the unique first letter, the server names were just as distinctly identifiable as the student names. This explanation could be tested by asking subjects to learn the structure of a computer network consisting of servers with human names.[1]

The second possibility is that people simply are better at learning graph structure when it is framed as a social network than otherwise. Humans have much experience learning social networks informally. Humans have great incentive to learn the social network around them in order to find jobs, friends and mates. It may be that we have developed a specialization for acquiring social network structure over and above the ability to acquire graph structure in general.

References

Barabási, Albert-László & Albert, Réka (1999). Emergence of scaling in random networks. *Science 286,* 509–512.

DeSoto, C. B. (1960). Learning a social structure. *Journal of Abnormal and Social Psychology, 60,* 417–421.

DeSoto, C. B., Henley, N. M., & London, M. (1968). Balance and the grouping schema. *Journal of Personality and Social Psychology, 8,* 1–7.

Erdős, P. & Rényi, A. (1959). On Random Graphs. I. *Publicationes Mathematicae, 6,* 290–297.

Fincher, D. (2010). *The Social Network* [motion picture]. USA: Columbia Pictures.

Krackhardt, D., & Kilduff, M. (1999). Whether close or far: Social distance effects on perceived balance in friendship networks. *Journal of Personality and Social Psychology, 76,* 770–782.

Zajonc, R. B., & Burnstein, E. (1965). The learning of balanced and unbalanced social structures. *Journal of Personality, 33,* 153–163.

[1] Network systems administrators may have already discovered the answer. It would explain why many computer servers are given the names of hobbits, wizards and Star Wars characters.