Link Formation in Networks Using Case-Based Decision-Making: A Laboratory Experiment*

Theodore L. Turocy
Department of Economics and Finance
John Carroll University
tturocy@jcu.edu

Elizabeth Watson
Department of Economics
Trinity University
elizabeth.watson2@trinity.edu

November 9, 2009

Abstract

We report results of laboratory experiments in which subjects make decisions about forming links with potential neighbors in a network context. In the field, decisions about forming links in social or professional networks are often made with local, incomplete information. This information often comes via the experiences of other agents with whom one is already a neighbor. We investigate the application of case-based decision theory (Gilboa and Schmeidler 1995) as a way to organize how subjects make their link formation decisions. We find that CBDT organizes choices better than several other competing heuristics, including choosing randomly and choosing the lowest-cost link available.

1 Introduction

Networks have long been the substrate on which much economic activity takes place. The buildout of the so-called "Web 2.0" has seen the creation of many platforms in which the abstract concept

*This is a preliminary version circulated for comments and suggestions. Please do not cite without checking first with the authors!

of network has been made concrete and explicit. Facebook helps individuals maintain lists of people whom they know in various facets of life, and one can keep track of status updates on Twitter, ranging from the sublime to the ridiculous, from friends and celebrities. Job search, in which networking has long been recognized as an important factor in the process of matching employers to applicants, has been transformed by the existence of sites like CareerBuilder.com and Monster.com.

A salient feature of all these networks is their dynamism. Network-themed websites have seen explosive growth in recent years. In a dynamic environment, the question of what links to form and maintain is central. Yet, precisely because the environment is so dynamic, those choices are necessarily made with imperfect information. Consider for example an employer wanting to post a job listing online. He can make use of various job networking sites, however, posting jobs and searching posted resumes is costly. How does the employer decide whether to post his job listing on CareerBuilder.com's website or to post his job listing on Monster.com's website? There is no way for him to know with precision which "link" will yield him the best employee (the most benefit). Perhaps he should form "links" with both websites, or perhaps neither.

Understanding how agents decide which networks to join and which links to form is an open research question. Over the past decade, much research in economics has been devoted to the formation of social networks (see Jackson 2005 for an overview). The equilibrium approach used in Jackson and Wolinsky (1996) makes assumptions about the information available to agents, and their strategic sophistication, which are implausible on the scales of real-world economies. As noted in Callander and Plott (2005), "[n]aturally occuring networks take place in a variety of institutional and informational environments." In many network environments, agents do not know the costs and benefits to themselves of forming various links. A reasonable response to such sparse information is to look to the experiences their "friends" (or even they themselves) have had in similar networks in the past. The employer trying to decide which job site to link with will know if he has had prior experience with a particular job site and will use that information to help him formulate an expected benefit from linking with that site again. If it is the case that he has not had direct experience with a given jobsite, he can compare the experiences of other employers with whom he has contact to compare their experiences. The employer can read the testimonials of others who have opted to link with a particular jobsite. Once the employer sees how the jobsite has benefitted others in a situation similar to his own he can make an educated judgment about how beneficial that particular link will be to him.

Gilles et al (2007) propose to use case-based decision theory (Gilboa and Schmeidler 1995, abbreviated CBDT hereafter) as a model of how agents choose to form and maintain links in a network environment in which the agent has limited information. CBDT is a decision making paradigm in which decision makers (DMs) use the past experiences of both themselves and others to help inform their evaluation of their current choices. A case-based DM knows the benefit others have received in the past from making a particular choice. The DM then judges how similar that past situation is to the one he is facing now, and weights the resulting benefit using that similarity. The DM opts for the choice with the highest similarity weighted benefit.

In this paper, we report on a series of network experiments in which agents have limited information on the benefit to them of forming any given link. We find that CBDT organizes subject choices better than several other heurstics that would be possible in such a limited information environment. In particular we find that CBDT does a better job organizing individual choices than does random choice or choosing the link with the lowest cost.

The paper is organized as follows. Section 2outlines the network model we consider, and the application of CBDT to a decision-maker's link formation problem in that context. Section 3 describes the implementation of the experimental design, and section 4 presents and analyzes the results. Section 5 then concludes with a discussion.

2 Model

We implement a version of the "connections model" from Jackson and Wolinsky (1996). In this model of social communication, each person has some inherent value, which we will call "coolness," which others can enjoy if they have a link with that person. People gain benefits, then, from being closely associated with cool people. Suppose, for instance, John is a very cool person. Then, if Sue is friends with John, i.e., has a direct social connection with him, then she gets the full benefit of his coolness. Another person, Mark, is Sue's friend, but not John's. As a "friend of a friend" of a very cool person, Mark enjoys some benefit from the indirect association with John, but less than if he were himself John's friend directly.

More formally, there are N agents who are connected in a graph G. G is a set of pairs (i, j), $i, j \in N$. which are interpreted as direct connections between pairs of agents i and j. Each agent i has an intrinsic value w_{ij} to each other agent j in the network. This intrinsic value is attenuated by distance; if t_{ij} is the shortest distance between i and j in the graph, then the benefit agent i receives from his connection, direct or indirect, with agent j is $\delta^{t_{ij}} w_{ij}$, for some parameter $0 < \delta < 1$. The total benefit agent i receives is given by

$$v_i = \sum_j \delta^{t_{ij}} w_{ij}.$$

Links are endogenous in the sense that for a link to exist, both players must find it beneficial to form the link. Links are costly, with c_{ij} denoting the cost to agent i of maintaining a link to j.

Therefore, the utility of agent i is given by

$$u_i = \sum_j \delta^{t_{ij}} w_{ij} - \sum_{j:(i,j)\in G} c_{ij}.$$

The equilibrium analysis of this model by Jackson and Wolinsky assumes that agents have full information and understanding about the consequences of forming or dissolving links, including the full matrix (w_{ij}) and the attenuation factor δ. Following the suggestion of Gilles et al (2007), we replace this by the assumption that agents use CBDT to evaluate whether to form links in the network. The role of the memory in CBDT for a decision-maker (DM) is played by the experiences of other agents in the network, whom we will refer to as the "friends" of the DM. Our DM observes the other agents to whom each friend is connected, and the total benefit that friend receives from his connections, but does not observe specifically how that benefit is derived.

In our setting, we will impose that $w_{ik} = w_{jk}$ for all agents i, j, and k. If the DM believes this is the case, but does not directly observe the (w_{ij}) matrix, then CBDT may be a reasonable heuristic for making decisions, insofar as the experiences of one's friends do provide useful information about which links would be desirable to form. We also make some simplifications to make this model amenable for laboratory experimentation. While we are ultimately interested in CBDT as a model for dynamic network formation, to focus on the question of whether CBDT models subject decisions, we will use a static setting, in which a subject will be given one network scenario, and asked whether she wants to form a link.

CBDT assumes the DM has a well-defined notion of similarity. Similarity is inherently a subjective concept, and different DMs may have different notions of similarity. We choose our experimental design so a few possible similarity concepts are salient. Following Gilles et al, we define two similarity measures, *network degree* (ND) similiarity and *neighborhood structure* (NS) similarity. Network degree similarity measures how similar two agents are in terms of the number of friends they have. Let $f_i(G) = |\{k : (i,k) \in G|$ denote the number of friends agent i has in network G. Then, the network degree similarity between agents i and j is

$$\sigma_{ij}^{ND}(G) = 1 - \frac{|f_i(G) - f_j(G)|}{n-2}. \tag{1}$$

Neighborhood structure similarity measures the extent to which two agents have the same set of friends. Let $F_i(G) = \{k : (i,k) \in G\}$ denote the set of friends of agent i in network G. Then, the neighborhood structure similarity between agents i and j is

$$\sigma_{ij}^{NS}(G) = 1 - \frac{|F_i(G) \Delta F_j(G)|}{n}, \tag{2}$$

where Δ denotes the symmetric set difference. Since we are not directly interested in testing what a particular agent sees as similar, we choose parameters in our design in which the same choice is predicted whether a subject uses *ND* or *NS* similarity.

To fix design parameters, there are a total of $N = 9$ agents in each network. The DM has four immediate neighbors. In each network, the DM observes the set of direct links each of his four neighbors has, and the total benefit the agent receives, but does not know exactly how the benefit is calculated. The DM does know that the benefit calculation is symmetric in that, if he had the same set of neighbors as another agent, he would receive the same total benefit. The DM is then offered the opportunity to link with one of the remaining four agents, or to make no link at all. The DM knows the cost of forming a link with each of those agents. The DM may form at most one link, or may choose not to form a link.

Under CBDT, given a similarity function σ, the DM assesses the benefit of connecting to a non-neighbor agent k using the formula

$$u_{ik} = \sum_{j:(i,j)\in G} \sigma_{ij}(G) \frac{v_j}{f_i(G)+1}.$$

CBDT then predicts that the DM will connect to the agent k which maximizes this $u_{ik} - c_k$, or will make no connections if all $u_{ik} - c_k < 0$.

3 Experimental Design

We report the results from three experimental sessions conducted in June 2009 at the Economic Research Laboratory at Texas A&M University. A total of 29 students recruited from the undergraduate student body at Texas A&M participated. Each session lasted approximately one hour, including instructions and distribution of earnings to the subjects.

Each session began with reading the instructions aloud.[1] The instructions were distributed in hardcopy to each student for reference, and were also projected on a screen at the front of the lab. The instructions presented the decision environment using a network frame, using the term "neighbor" to denote another simulated agent with whom the subject had a pre-existing link. The instructions indicated that there would be 20 networks in which the subject would make a decision, and that decisions in each of these networks would be independent. Furthermore, the other agents in each network were "simulated," and did not represent other participants in the experimental session. Finally, subjects made all 20 decisions prior to finding out the results of their decisions.

The subject decision interface is shown in Figure 1. The subject's links, and the links of his four neighbors, were presented as rows in a matrix. Each column corresponded to one of the nine agents

[1] The instructions are available from the authors on request.

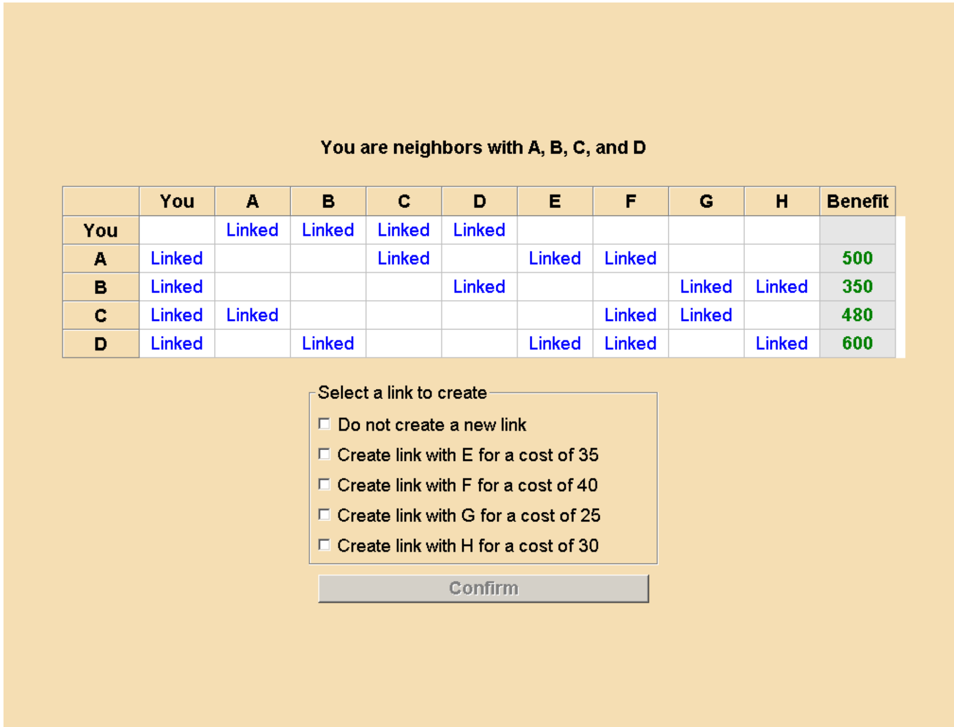

Figure 1: Screenshot of subject decision interface.

in the network. Reading down a column identifies all the agents in the subject's neighborhood who had direct links with that column's agent; therefore, similarity could be assessed by comparing the similarity of two rows in the matrix. In presenting the networks to the subjects, the assignment of letters to nodes was randomized, as was the ordering of listing the nodes in the subject interface, to minimize the potential for anchoring on letters or placement on the screen.

The parameters for the networks were chosen using a 5×4 design matrix. There were five different network structures, shown in Figure 2. The dashed box in each figure includes the agents who were in the subject's neighborhood. Therefore, network links originating inside the box were visible to the subjects, but network links entirely outside the box were not known to the subjects. For each network structure, four different cost functions were used, to vary systematically the choice which CBDT predicts. The quantitative parameters were chosen to satisfy two criteria. First, for all network-cost function pairs, the prediction made by CBDT is the same under either the network degree (1) or neighborhood structure (2) formulations of similarity. Second, for each network, the prediction made by CBDT differs from the choice which would maximize earnings if the subject had full information about the decision problem.

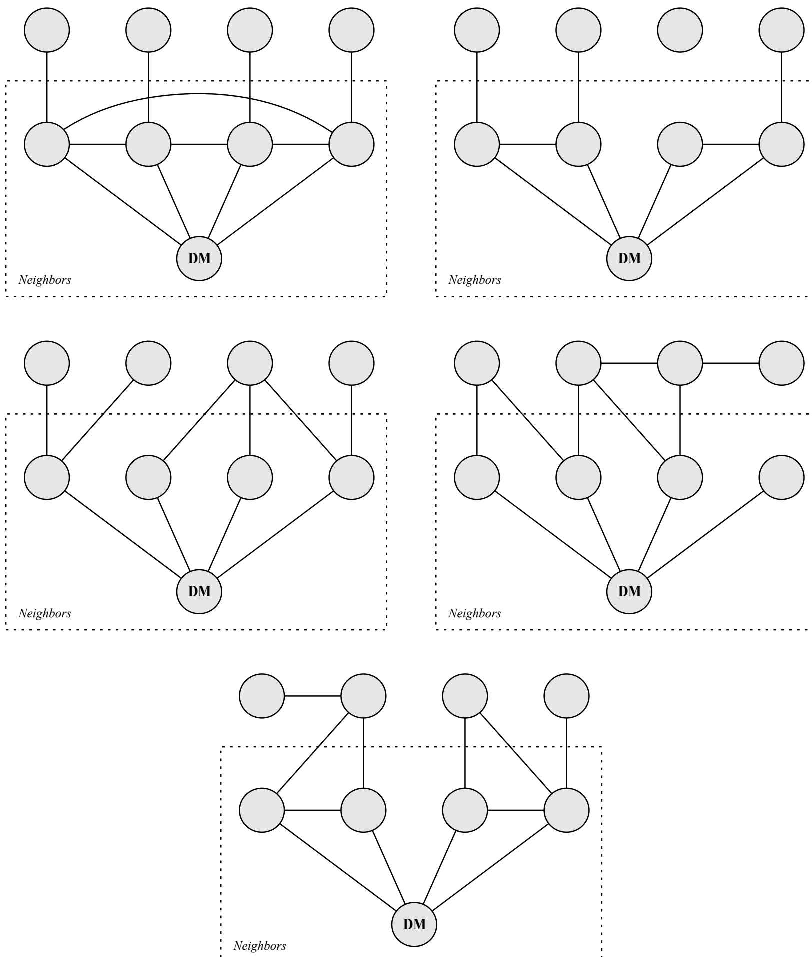

Figure 2: The five networks.

4 Results

To evaluate the performance of CBDT as an organizing model, we compare against several other plausible heuristics:

1. **Earnings maximization.** The experimental design is such that subjects do not have enough information to back out the earnings-maximizing choice. To the extent that it might be possible to draw partial inference about the earnings-maximizing choice, or to reverse-engineer the maximizing choices based upon some subtle patterns in our design parameterization, our parameters were chosen so that the CBDT prediction never coincides with the choice which would actually maximize earnings.

2. **Random choice.** In an environment with limited information and no feedback on choices, subjects might believe they have no way of distinguishing good choices from poor ones, and therefore choose at random.

3. **Choosing the cheapest link.** In our networks, there is ambiguity regarding the potential benefits of forming any given link; however, there is certainty as to the cost that would be incurred by choosing to form a link. Therefore, it is possible that subjects might focus on the (known) costs instead of the (unknown) benefits, and choose the link which is least costly, since cost minimization is a heuristic that can be implemented.

We organize the analysis of the results in two dimensions: performance of CBDT and alternative models for each network-cost function pair across subjects (longitudinal), and for each subject across network-cost function pairs (cross-sectional).

4.1 Longitudinal analysis

Table 1 summarizes the frequencies with which each of the models predicted the actual choice of each of the 29 subjects.

Result 1. *The frequency with which CBDT matches actual subject choices is significantly greater than random choice.*

Support. Assuming the choice in each network is made independently, and that each of the five choices is chosen with equal probability, there is a 3.2% chance of matching 8 or more choices. Of the 29 subjects, CBDT matches 8 or more choices made by the subject for 16 subjects (55%).

Subject	**CBDT**	**Min cost**	**Max earnings**	**Subject**	**CBDT**	**Min cost**	**Max earnings**
1	14	5	1	**16**	10	6	4
2	12	1	3	**17**	6	5	2
3	8	7	1	**18**	11	7	0
4	7	6	2	**19**	10	6	2
5	7	2	3	**20**	9	7	1
6	7	4	3	**21**	7	8	2
7	9	1	4	**22**	6	6	1
8	6	2	2	**23**	8	9	1
9	5	2	4	**24**	9	2	4
10	9	5	0	**25**	9	5	1
11	12	2	1	**26**	11	5	2
12	9	7	0	**27**	10	2	5
13	10	6	0	**28**	5	5	2
14	4	3	1	**29**	4	5	8
15	6	4	4				

Table 1: Performance of heuristics by subject (number of correct predictions, out of 20)

Result 2. *CBDT outperforms the minimum-cost heuristic for a significant majority of the subjects.*

Support. CBDT outperforms the minimum-cost heuristic on 24 of 29 subjects, with minimum-cost matching more often three times, and two ties. We reject the null hypothesis that the two models are equally likely to better predict a given subject's choices (a binomial random variable with $n = 29$, $\pi = 0.5$ has probability $\sim 10^{-5}$ of taking on values 24 or higher). The margin by which CBDT outperforms is also significant. The median number of correct predictions by CBDT is 9; on only one subject does mincost predict 9 choices correctly.

Taken together, Results 1 and 2 indicate a majority of subjects are responding to some degree to the characteristics of the decision environments, and that their decisions are not driven solely by the costs of link formation.

Result 3. *CBDT outperforms the max-earnings heuristic for a significant majority of the subjects.*

Support. CBDT outperforms the max-earnings heuristic on 27 of 29 subjects, with max-earnings matching more often once, and one tie. We therefore reject the null hypothesis that the two models are equally likely to better predict a given subject's choices (a binomial random variable with $n = 29$, $\pi = 0.5$ has probability $\sim 10^{-8}$ of taking on values 27 or higher). The margin by which CBDT outperforms is also significant. Max-earnings predicts 8 decisions correctly for

Network	CBDT	Min cost	Max earnings	Network	CBDT	Min cost	Max earnings
1	22	1	4	**11**	1	1	10
2	3	4	3	**12**	16	7	2
3	22	22	1	**13**	17	17	2
4	23	1	1	**14**	16	1	1
5	10	8	6	**15**	9	10	5
6	0	0	6	**16**	4	4	8
7	14	14	1	**17**	13	14	2
8	15	2	0	**18**	12	5	1
9	12	10	1	**19**	12	11	1
10	11	2	3	**20**	8	1	7

Table 2: Performance of heuristics by network (number of correct predictions, out of 29)

one subject, 5 for a second subject, and no more than 4 for any other subject. In fact, max-earnings generally predicts fewer choices correctly than the random choice model; the median number of correct predictions is only 2.

Again, subjects did not have enough information to determine the choice which would actually maximize their earnings. Result 3 confirms that the experimental design achieved this objective, and that the predictive power of CBDT (or any other model) cannot be due to coincidence with the actual earnings-maximizing choice.

4.2 Cross-sectional analysis

Table 2 summarizes the frequencies with which each of the models predicted the actual choice in each of the 20 network settings. We see that CBDT in general does well on most networks.

Result 4. *CBDT outperforms the minimum-cost heuristic for a majority of the networks.*

Support. On three networks (3, 7, and 13) the CBDT and mincost predictions coincided. On the remaining 17, for 11 CBDT predicts more subject choices than minimum-cost, with minimum-cost predicting more choices 3 times, and 3 ties. We can reject the null hypothesis that each model is equally likely to predict a plurality of subject choices across networks at the 10% level; there is a 7.2% chance a binomial random variable with $n = 17$, $\pi = 0.5$ takes on a value of 11 or greater.

4.3 Other results

An additional prediction of CBDT is that choices about which the subject has no information will not be taken. In our setting, in 8 networks there were agents who were neither neighbors of the subject, nor neighbors of the subject's neighbors.

Result 5. *Consistent with CBDT, subjects rarely chose to form links with agents who were not neighbors of neighbors.*

Support. Eight networks contained an agent who was not a neighbor of the subject's neighbors. In six of these networks (3, 4, 8, 13, 18, and 19), only one subject linked with that agent; in two of those networks (9 and 14), two subjects chose to link with that agent.

5 Discussion

In a network formation setting designed to suggest that case-based reasoning is a reasonable approach to choosing which links to initiate, we find that case-based decision theory predicts an appreciable number of subject choices. Most subjects make the choice consistent with CBDT more often than with any other heuristic, with CBDT predicting half or more of the subject's choices 13 out of 29 times. Similarly, for most of the network environments we study, CBDT predicts significantly better than other heuristics. The strength in performance of CBDT over the minimum cost and random heuristics indicates that subjects are attempting to use all the information, and not just focusing on cost, in making their decisions. The poor performance of the maximum "heuristic" indicates that the performance of CBDT cannot be attributed to it coinciding with the action which would maximize earnings.

We interpret the results as being favorable to CBDT. In practice, the very nature of the concept of a heuristic is such that different heuristics may suggest themselves in different decision settings. Therefore, it is too much to expect that any decision-maker would always choose in accordance with CBDT, even if he found CBDT compelling in some circumstances. The usefulness of a decision theory is to the extent it can organize enough of individual decision-making in practice to be a viable choice in constructing larger models. The evidence from our experiments indicates that CBDT is a credible choice for theoretical modeling some dynamic aspects of network formation.

Acknowledgments

We thank John Van Huyck for support in using the facilities of the Economic Research Laboratory at Texas A&M University to conduct the experiments we report.

References

[1] Callander, Steven and Charles R. Plott. 2005. Principles of Network Development and Evolution: An Experimental Study. Journal of Public Economics 89, 1469-1495.

[2] Gilboa, Itzhak and David Schmeidler. 1995. Case-Based Decision Theory. The Quarterly Journal of Economics 110(3), 605-639.

[3] Gilles, Robert, Tabitha James, Reza Barkhi, and Dimitrios Diamantaras. 2007. Simulating Social Network Formation: A Case-Based Decision Theoretic Model. Manuscript.

[4] Jackson, Matthew O. 2005. A Survey of Models of Network Formation: Stability and Efficiency. in Group Formation in Economics: Networks, Clubs, and Coaltions ed by G. Demange and M. Wooders. Cambridge University Press. Cambridge, UK.

[5] Jackson, Matthew O. and Asher Wolinsky. 1996. A Strategic Model of Social and Economic Networks. Journal of Economic Theory 71: 44-74.