Published as: (2015). Building social networks out of cognitive blocks: factors of interest in agent-based socio-cognitive simulations. Computational and Mathematical Organization Theory, 21(2), 115-149.

What matters in social networks: Defining Factors of Interest for Agent-based Socio-cognitive Simulations

Changkun Zhao, Ryan Kaulakis, Jonathan H. Morgan, Jeremiah W. Hiam, Joseph P. Sanford, Frank E. Ritter
The College of Information Sciences and Technology
The Pennsylvania State University

Geoffrey P. Morgan
School of Computer Science
Carnegie Mellon University

Draft15_2012/11/06 Changkun Zhao

Abstract

This paper examines dyadic social tie formation, and how cognitive and environmental factors influence the formation of dyadic ties. Using agent models instantiated in ACT-R that interact in a large-scale social simulation, we capture the affect of environmental factors including population size, running time, and map configurations, as well as memory constraints. To explore and test these relationships, we ran simulations using a factorial design and used the simulation data to generate a large corpus of networks. Our analyses suggest five interesting conclusions: first, that the three environmental factors all influence both network density and some aspects of network structures; second, that agent memory strongly and decisively alters the network's density and structure; third, the growth pattern of these networks approximates a power law distribution; fourth, that the environment structure influences the networks' generation speed; and finally, certain map configurations tend to have more asymmetric activation patters. These findings are interesting in that they imply that the size of a social network primarily depends on internal cognitive factors rather than environmental factors, providing support for and deepening our understanding of Dunbar's(1998) number. These findings also suggest that future simulations examining generative social networks should account for and carefully report these factors.

1. Introduction

This paper examines the relationships between cognitive and environmental factors as they pertain to the formation of social networks by modeling how these factors affect the formation of dyads. This work is motivated by a desire to better understand how socio-cognitive processes influence the development of persistent patterns of relations, represented in this paper as network topologies. By socio-cognitive processes, we refer to both those cognitive resources and mechanisms necessary to create and sustain social ties, as well as those group-level factors known to moderate human decision-making (G. P. Morgan & Carley, 2011; J. H. Morgan, Morgan, & Ritter, 2010; Silverman, 2004; Yen et al., 2001). We focus on spatial reasoning and retention in this paper because these two processes seem foundational to understanding the emergence of social networks in a variety of contexts.

To explore this relationship, we introduce a set of agent-based models and experiments that test the influence of: population size, run time, map configuration, and agent navigation strategies. The outputs of this model are interaction networks (whole networks representing the total number of agent interactions that occurred within a single run) and ego-nets (declarative representations of the agents friends network). For any one run, there is one interaction network and as many ego-nets (networks from the agent's egocentric point of view) as there are agents in the experiment. We also merged the individual ego-nets to examine how each factor influence the generation process of the merged ego-nets.

Our model is unusual in that we model social processes using a cognitive architecture (ACT-R) that is primarily associated with cognitive science. To our knowledge, Carley (1991, 1992) and Newell (1994; Newell, Rosenbloom, & Laird, 1989) were the first to study organizations using a model based on a cognitive architecture (Plural-Soar). More recently, Gonzalez and Lebiere (2005), Lebiere, Gonzales, Dutt, and Warwick (2009), Reitter and Lebiere (Reitter & Lebiere, 2010), and Juvina, Lebiere, Martin, and Gonzalez (2011) have used cognitive architectures to model human decision making in collaborative tasks. Barrett, Eubank, and Marathe (2006) have developed a large simulation with millions of light, non-cognitive agents to model the influence paths in crisis situations. While our work builds upon these efforts, our interest in network formation poses some unique challenges. We review these challenge in light of the current literature and our solutions to them in the next three sections, and then discuss the results and implications in the final two sections.

2. Computational Social Models

Researchers have developed agent-based models to explore a variety of questions. We briefly examine two major modeling approaches: cognitive and social modeling. These approaches are not necessarily mutually exclusive; however, combined socio-cognitive models are relatively rare because they are generally expensive to create and run.

Social simulation models frequently, but not always, use bounded rational agents. Bounded agents are limited both cognitively and socially (Simon, 1991). These agents generally engage in some kind of goal-driven time-constrained decision cycle dependent on local information. In addition, agents are usually adaptive, though options are generally conditioned upon previous actions. Often, these simulations demonstrate aggregate behaviors that are emergent. These complex system-level behaviors arise out of the agents' discrete interactions, but cannot be explained entirely in reference to them. Further, at different levels of observation, different kinds of emergent behavior can be seen. It is often these kinds of traits that simulations are uniquely able to capture.

We treat the topology of interaction networks and their associated characteristics as an emergent property of the social system, defined in terms of interaction opportunities and memory constraints. Kraut et al. (2002) found that actor proximity fundamentally influences the evolution of network topologies by determining the interaction frequencies of actors across the network, while Allen (1977) demonstrated that the probability of two people communicating in an environment could be represented with a decreasing hyperbolic function of their distance. After a certain distance, the probability that two people will communicate decreases rapidly, making tie formation unlikely (Carley, 1991, 1992). We, thus, chose to focus on factors that are known to directly affect agent proximity or inter-agent distance: population size, run duration, and map configuration.

Cognitive models have historically focused on modeling human cognition at the symbolic and sub-symbolic level (Anderson et al., 2004; Newell, 1994; eg). Providing models of perception and memory, we can use cognitive architectures like ACT-R to simulate the formation of social ties in declarative memory. We believe that, as time increases, memory constraints fundamentally influence a social network's topology and capabilities by constraining the network's ability to process information, identify important changes in state, and respond to those changes.

Here, we look at the processing of social information by exploring the concept of nodal carrying capacity, the number of agents an agent can retain in memory. To that end, we examine how environmental factors contribute to the consolidation and retention of social ties in memory.

This concept is similar to Dunbar's (1998), where he argues that limitations associated with the neocortex limit the number bi-directional ties any one person can retain in memory to approximately 150 people. Dunbar argues that maintaining stable relationships requires repeated memory activations to identify not only one-on-one relationships but also third party relationships (i.e., the knowledge that my friend is also friends with other actors who I monitor). Further, he claims that the cognitive load associated with maintaining this set of relationships in memory rises exponentially as group size increases (Dunbar, 1998, p. 63). Based on retrospective empirical studies, Dunbar (Dunbar, 1998, pp. 68-75) argues that this ratio between cognitive load and group size underlies the small-world effect observed by Milgram(Milgram, 1967), and others McCarty, Killworth, Bernard, Johnsen, and Shelley (2001) propose a far larger number (n=291) of a person's social ties. In part, this discrepancy is rooted in a difference in definitions. McCarty et al.'s (2001) definition of a social tie requires mutual identification as opposed to Dunbar's stricter definition of mutual identification and placement in the network. In this work, the social ties we used are closer to McCarty et al.'s definition because mutual identification could be naturally implemented in the declarative memory of ACT-R. Nevertheless, we believe that our simulation is still relevant to the effect of Dunbar's number because both definitions bridge between cognitive limitation and number of social ties.

In the experiment, we expect that larger populations acting over longer time periods in fully connected environments will result in the most connected declarative network structures. We also expect that less connected environment layouts will result in interaction networks that consist of more fragmented networks, leading to smaller ego-nets. We also expect that map configurations characterized by nexus points (locations which act as hubs by virtue of being centrally located) will exhibit behaviors similar to the water-cooler effect (DiFonzo, 2008). We, however, are less certain where we might see thresholds in network formation, where for instance population growth no longer has an effect or run time is no longer relevant.

3. Nodal Carrying Capacity: The Effect of Agents' Memory and Space

Having summarized our model's environmental and cognitive factors, we provide both a definition and a prediction as to how each factor will influence network formation.

3.1 Interaction Frequency Factors

We model three factors that influence interaction frequency: population size, run time, and map configuration.

Population density: We predict population density will have the greatest impact on social interaction frequency. Here, we model shifts in population density by changing the population size, and holding the environment size constant.

Length of simulation (run time): We predict longer run times will lead to more ties and denser networks. Consequently, determining the run time lengths necessary for a network to reach a stable state under a given set of conditions (e.g., memory decay of ties) is important for accurately representing the formation of a group of interest.

Environment configuration: We predict the configuration of the environment will influence the structure of the simulated social network. We measure the relative connectivity of our three map configurations by defining its *grid ratio*. The *grid ratio* is the ratio of the number of edges over the total number of edges possible for a rectangular grid containing the same number of rooms.

We tested three map configurations (Figure 1). The first configuration (1a) is a full 5x5 grid with *grid ratio* 1.0. We expect this environment will result in relatively high connectivity. The second configuration (1b) has a central area with *grid ratio* 0.75. We believe this central meeting point will lead to network densities and clustering that are less pronounced than those associated with the 5x5 map but more than those associated with the hallway map. The third configuration (1c) is a two-hallway configuration with *grid ratio* 0.6. This configuration should lead to low connectivity due to the large distances between agents.

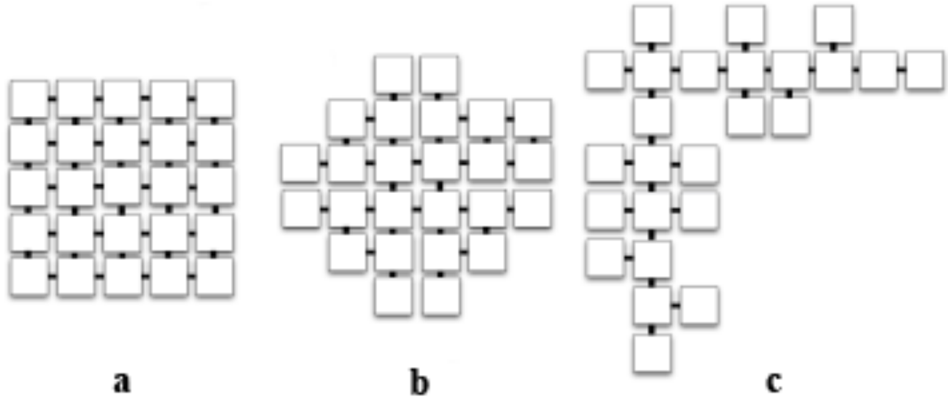

Figure 1. (a) 5by5 grid map, (b) Central Area map, (c) and Hallway grid map

3.2 Memory and Nodal Carrying Capacity Factors

Because we are interested in nodal carrying capacity, we needed a theory of memory and its decay. We use Anderson's activation theory (2004) as implemented in ACT-R, as our theoretical framework. Each agent may have one or more memory chunks which represent other agents (from an ego perspective, these perceived entities are called alters[1]). The number of active memory chunks is influenced by several factors, including initial memory activation, retrieval threshold, memory decay rate, time of retrievals, and practice time. We use the activation of alters (agents in a dyad with the target agent), which is directly related to the likelihood of recalling these alters, to determine whether a tie exists in each network. Thus, as the threshold for determining if a network exists, the semantics for identifying another alter as an acquaintance become more and more stringent.

Figure 2 offers an example of how a single merged ego-network can reveal the strong ties in a network as the criteria for inclusion in that network becomes stricter. By stricter, we mean that the activation threshold necessary to be considered a tie increases. While the three networks shown in Figure 4 arise from the same set of declarative representations, these networks reflect activation thresholds of -3.5, 0.0, and 1.0 (ACT-R's declarative memory activation ranges from negative numbers to small positive numbers), corresponding to networks a, b, and c respectively. In essence, network c identifies the agent's core relationships, those of which the agent has the strongest memory.

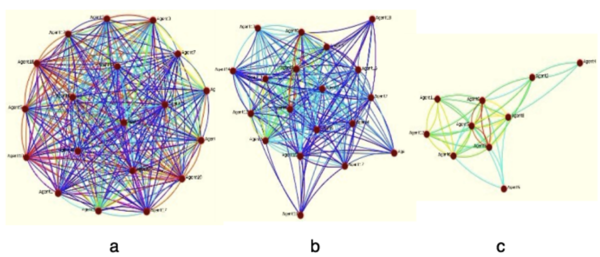

Figure 2. The same ego network at various memory thresholds, (a) -3.5, (b) 0.0, and (c) 1.0.

[1] In network science, the term 'alter' is often used to indicate that this is the ego's perception of other agents.

Other cognitive factors may be important, but we have reserved those for future work. We are particularly interested in considering agent movement patterns, and how that may affect social networks. For example, a policeman walking a beat will have more acquaintances (alters) than a person who spends most of their time at home, because the policeman has more opportunities to meet people.

4. Experiment Environment

To model multi-agent social behavior using cognitive architectures, we constructed a simulation environment, VIPER. All of our experiments were conducted on a 2GHz eight-core server with 8GB of RAM. The server runs Linux 2.6.31 under Ubuntu 11.04, with SBCL 1.0.52 Lisp, and ACT-R 6 (Anderson et al., 2004).

4.1 The VIPER Server

VIPER models the constraints associated with embodiment on social networks. It supports multi-agent simulations to study network science. It is lightweight in that it is text-based, but is extendable and records agent behaviors. VIPER is designed to be a part of a distributed model that resolves events in either real or accelerated time. The network's speed and frequency of communication are determined by its component agents, with no queue of events being enforced. VIPER is designed so variations in performance originate from the agents participating in the environment, versus being a function of the environment. The VIPER server is based on NakedMUD, an open-source MUD environment. It communicates with client programs using the Telnet Protocol.

Within the environment provided by the server, agents or human subjects are situated on maps of interconnected rooms. The agents can see and communicate within each room. Agents can walk between the rooms, and can interact with objects in the rooms.

To connect ACT-R to VIPER, we implemented the Telnet Agent Wrapper for ACT-R (TAWA) in Common Lisp. It supports logging in, waiting for synchronization, logging, halting, and writing results to CSV files. It also exports a number of functions that provide ways to examine the environment, speak, listen, move, and otherwise control a virtual body in VIPER.

When an ACT-R model is wrapped by TAWA, executions of model code are delayed until a privileged administrator agent signals for synchronization. Error conditions are also caught by TAWA and standard UNIX error codes are returned instead of dropping into the more standard debugger. For example, a successful run returns 0 to the parent process, while any error (e.g., network errors like the server being unreachable) returns a non-0 value. Returning error codes like this allows automated error checking in large-scale experiments.

4.2 Synchronization

Because memory decay and networks are strongly temporal, we paid special attention to time. To synchronize the agents, the administrator agent (which does not take part in the experiment) waits for all of the TAWA wrapped agents to finish loading and logging in. It then signals to TAWA to begin the simulation. Because TAWA delays the evaluation of the model code until synchronization, no agent experiences time before the synchronization signal. Further, all ACT-R models are set to run in real-time and for the same amount of "real time", so they all halt after the same perceived period. Thus, the total time experienced is the same for all agents.

4.3 Scalability

Early benchmarks showed that ACT-R processes took up about 80MB per process. We would only have been able to run about 100 processes on a single 8GB machine before swapping would occur. To reduce the per-process footprint, a number of optimizations were implemented. Basic space reductions were achieved by using the DECLARE Lisp construct, as well as by pre-compiling the components, removing the debugger, and saving the whole system (sans the ACT-R agent model) as a system image. This reduced our per-process memory footprint somewhat, but they were not the biggest contributions towards memory usage reduction.

In SBCL Lisp 1.0.52, the "--merge-core-pages" flag was recently added. This flag enables Kernel SamePage Merging under recent versions of Linux (Arcangeli, Eidus, & Wright, 2009). This optimization flags shared areas of memory as being able to be merged unless modified. Because a significant percentage of our agents were replicated, we found that we could reduce the per-process memory footprint as low as 8MB per process (with one shared copy of the merged pages excepted). Thus, the only activities that increase the size of this footprint are changes within individual agent models. Sharing and merging copies increases the number of agents capable, whether on single processors or HPC. Together, these optimizations permit orders of magnitude more agents to be run in a single experiment than many previous efforts, enabling larger-scale analysis than has been previously done. Such large-scale work is planned for future research.

4.4 ACT-R Agents

We built an ACT-R model to conduct "random walks" in VIPER. The model contains two declarative memory types: a "goal" type containing the agent's current location, remaining steps, total friends counts; and a "friend" type used to store friend names. The model consists of four basic components with 9 productions. First, the agent's "walking" component selects an available direction randomly, and sends a *moving* message to VIPER. Second, the "waiting" component utilizes ACT-R's temporal buffer to wait 16 real-time seconds before allowing the agent to enter a new room to simulate transition time. Third, the agent's "checking" component, consisting of 3 productions, checks if the current room is empty. Fourth and finally, the "memorizing" component, consisting of 3 productions, first checks declarative memory to see if the agent has previously encountered the agent it has just met. If not, the model creates a new friend chunk, using the imaginal buffer to store the new agent name.

5. Virtual Experiments and Results

In this section, we consider three important questions. Do these environment topologies influence the agent's perceived social network? Are cognitive limitations important to understand how environments influence an agent's social network? How much exposure to a novel environment is necessary for agents to reach activation equilibrium?

To answer these questions, we run two sets of experiments to examine influential factors and activation equilibrium separately .

5.1 Experiment 1

In the first experiment, we created a dataset of 810 runs to test three environmental factors: population density, running time, and environment configuration, shown in Table 1.

Table 1. Experiment 1 parameters.

Variable	Values	#
Population	20, 40, 60	3
Run time	125, 250, 500 s	3
Environment	Full-Grid, Central, Hall (100%, 75%, 60%)	3
Total possible combinations		27
Replications per combination		30
Total Runs		810

The VIPER system generates both simulation logs, which can be used to evaluate movement patterns, and also ego files for each agent, showing the activation of each alter chunk, which identifies each alter. Each agent represents actors it has met as a working memory element (WME) chunk. The activation of each WME can be used to derive the probability of retrieving a chunk and the amount of time a human would require to recall the actor. The semantics of each dyadic tie is important in interpreting the network. An activation value of '-3' indicates that the actor will need as much as 200 ms to recall the chunk (if they can retrieve it at all), whereas an activation of '3' indicates that the actor will need less than 5 ms to recall the chunk (but longer to report it because of additional processes including planning an utterance and speaking). Figure 3 shows an example ego network.

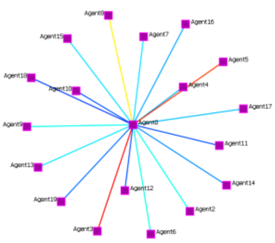

Figure 3. An example ego network, links are colored by activation value.

These ego files thus can be considered analogous to network survey data, with a survey instrument used to gather each individual's view on their social network. We merge these ego files to create merged ego networks. We have two distinct strategies for merging these ego files to create networks, symmetric-necessary and directed-sufficient.

To merge our ego-nets, we considered two rules for determining whether the tie should exist: symmetric-necessary (where threshold, t, must be met by both A_{ij} and A_{ji}) and directed-sufficient (where threshold, t, must be met by either A_{ij} or A_{ji}). In either case, a binary two-way link was formed if the necessary conditions were met. We chose to form binary two-way links to ensure that the network measures being used could be compared directly. The directed-sufficient network merge method is less conservative; typically creating more links, differences between the networks generated by each method may also offer interesting insights on the ramifications of environmental topology. Dunbar's definition of relationships more closely follows the semantics implied by the symmetric-necessary merge method (Dunbar, 1998).

Given that our alter chunks were inherently weighted by activation value, it was tempting to simply use the activation values to inform a weighted network measure analysis. We chose not to do that for three reasons. Firstly, we consider that changes to the threshold value change the meaning of the network, which makes such a network analysis problematic. Secondly, ACT-R's activation values range from negative to positive values, spanning zero (0). Zero is usually reserved in network analysis for the absence of a link. Transforming the data is possible, for example, by converting the activation value into the time to recall or probability of recall,, but adds an additional processing step. Finally, network measures are sensitive to the edge weighting and may produce non-intuitive results—many network analysis tools automatically binaries networks before calculating measures.

For each simulation run, we created both a symmetric-necessary and directional-sufficient network using thresholds between 5 and -5, with intervals of .1. Thus, we created two networks per run for each threshold value of -5.0, -4.9, -4.8, -4.7, and so on to 5.0. This process created a corpus of (810 runs * 2 merge methods * 100 threshold values) 160,000 merged ego-networks. We included isolates, egos with no connections, in our networks.

5.1.1 Network Measures

We use these generated networks to evaluate our questions of interest. We focus on five network-level measures to inform this analysis: density, average distance, clustering coefficient, average eigenvector, and average betweenness. We define those terms now.

Density is the number of ties present in the network compared to the maximum number of possible ties, excluding self-loops. Network density is a fundamental measure of a network, with implications for the interpretation of many other measures. Networks tend to be less dense as the population grows, but should be denser, other factors held equal, as the simulation run-time increases (at least until activation equilibrium is achieved for networks with memory decay).

Average distance is the number of jumps required, on average, for each agent to reach every other agent. If every agent is connected to every other agent (a density of 1) then average distance will also be 1. Thus, dense networks tend to have average distance near 1. Isolated nodes report a distance of 0 to every other agent, which can complicate the interpretation of this measure. Thus, networks with isolated nodes will have their average brought down towards 0.

Clustering coefficient is a transformation, over all agents, of each ego's local network density (how many alters is that ego connected to of all possible alters). A high clustering coefficient in a social network suggests a decentralized information structure, and one where actors tend to have an accurate understanding of the state of their local work-group. This measure is highly correlated with density.

Average eigenvector is a measure of the second-order connections an agent has. For example, a person who knows group leaders, but not the group members, has a high eigenvector. In very dense networks, average eigenvector is very low, typically near 0. In very sparse networks with isolated nodes, eigenvector will again be near 0. In medium density networks, the eigenvector may be large. Thus, eigenvector values tend to make an inverse U-shaped curve as network connectivity increases.

Average betweenness is a measure of a node's prominence on the paths between other agents. We expect map topology to strongly affect this measure. In thresholded networks, we would expect that betweenness will be higher in environments with the hallway configuration. Betweenness is low on highly dense networks, and also tends towards 0 in highly sparse networks with isolated nodes. Betweenness, thus, also tends have an inverse U-Shaped curve.

5.1.2 Simulation results

In this section, we examine the results of the simulation with respect to the questions we discussed at the introduction of this section.

5.1.2.1 Simple data description

Before entering data analysis section, we discuss some preliminary findings in this section. We plot 10 figures to display the changes of five network measures along activation threshold. In each figure, show the distribution of a network measures instead of showing average measures. To provide a better understanding and to compare each plots, we create Figure 4 to combine 10 plots together.

Figure 4 contains 10 sub figures to show the distribution of five measures when we increase threshold value from -5 to 5. It is obvious that all measures change significantly, especially when the threshold value is over -1. Comparing the Directed network and Symmetric network analyses, we also find that Symmetric network is more sensitive to cognitive limitations because the measures of the Directed network change earlier than Symmetric network when we increase the threshold values.

We could also find that the threshold has positive influences on Average Distance, Eigenverctor, and Betweenness. It has negative influences on Density, and Clustering Co-efficiency. This result implies that when increasing activation threshold, the overall size of the network will shrink but the network will split into several smaller and tighter groups with more second-order connections.

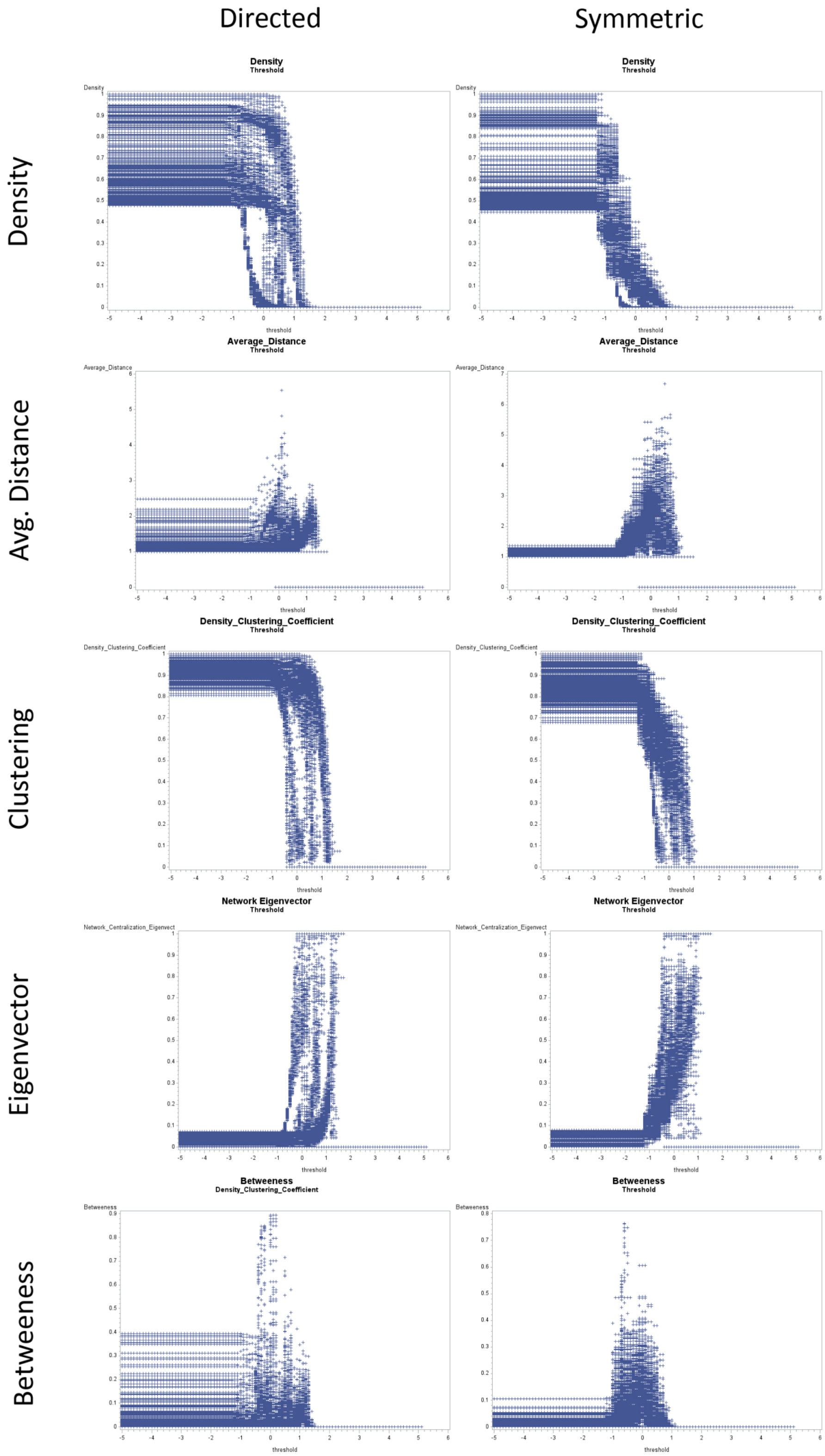

Figure 4. The distribution of five network measures along threshold

5.1.2.2 Environment Factors influence network formation

To examine the impact of environmental topology, we used a model selection approach, comparing the impact of various terms and whether their inclusion offered sufficient benefit, as calculated by Bayesian Information Criterion (BIC)(Schwarz, 1978) and Adjusted R-Square, to justify their presence. Environmental Topology, Running Time, and Agent Count were coded as ordinal variables, while threshold was, for this analysis, a quantitative independent value. Each model was applied to each dependent variable of interest in turn: density, average distance, clustering coefficient, average eigenvector, and average betweenness. In addition to four basic variables that we intended to examine, we also added two supplemental variables: a Squared Thresholding variable to allow for matching non-linear relations in the data; and a Spline variable to represent the spline difference between the range -3 to 3 and the range -5 to5. We present six regression models as:

Model 1: The thresholding value alone.
Model 2: Model 1 + Running Time
Model 3: Model 2 + Agent Count
Model 4: Model 3 + Environmental Topology
Model 5: Model 4 + Squared Thresholding (to allow for non-linearity)
Model 6: Model5+ Spline variable (to represent difference between (-3,3) and (-5,5))

We assert that if Model 4 provides a sufficient improvement over Model 3, then the environmental topology term is useful for understanding network formation. Typically, a decrease of more than 2 in the BIC score indicates that the new variable provides useful information. The Model 1 column shows the actual BIC score, whereas later columns show the relative improvement from the previous model. Table 2 shows our model selection results.

Table 2. Model selection results using a BIC analysis(Schwarz, 1978). Directed refers to networks generated with the directed-sufficient method. Symmetric refers to networks generated with the symmetric-necessary method.

	Threshold	Time	Count	Env. Top.	Non-Linearity	Spline
Directed Density	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Adjusted R^2	.71	.72	.76	.76	.76	0.84
Δ BIC (from prev)	-249865	-5079	-10354	-62	-993	-31290
Symmetric Density	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Adjusted R^2	.7248	.7265	.7603	.7603	.7777	0.8517
Δ BIC (from prev)	-265634.25	-506.19	-10874.5	-12.85	-6221.33	-33465.49
Directed AVG Distance	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Adjusted R^2	.5448	.5479	.5545	.5546	.6085	0.6489
Δ BIC (from prev)	-152230.62	-564.22	-1208.89	-12.1	-10660	-8995.77
Symmetric AVG Distance	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Adjusted R^2	.3734	.3935	.3948	.3948	.4549	0.4993
Δ BIC (from prev)	-91075.07	-2696.45	-170.947	-0.697	-8643.49	-7017.06
Directed Clustering	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Adjusted R^2	.7588	.7698	.7719	.7720	.7781	0.8581
Δ BIC (from prev)	-247427.27	-3833.92	-768.43	-8.82	-2242.15	-36971.49
Symmetric Clustering	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Adjusted R^2	.8022	.8048	.8092	.8092	.8138	0.9003
Δ BIC (from prev)	-273429.14	-1101.3	-1858.32	0.34	-2013.1	-51584.97
Directed Eigenvector	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Adjusted R^2	.0031	.0084	.0236	.0236	.0798	0.0819
Δ BIC (from prev)	-335699.46	-441.1	-1269.84	-3.52	-4889.12	-185.21

Symmetric Eigenvector	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Adjusted R^2	.0066	.0134	.0156	.0156	.133	0.1331
Δ BIC (from prev)	-303507.63	-567.78	-178.77	-3.72	-10481.4	-8.78

Directed Betweenness	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Adjusted R^2	.0111	.0128	.0265	.0272	.0419	0.091
Δ BIC (from prev)	-475378.96	-140.53	-1154.65	-53.87	-1256.62	-53519.95
Symmetric Betweenness	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Adjusted R^2	.0053	.0143	.0159	.0159	.0873	0.091
Δ BIC (from prev)	-524066.57	-750.7	-133.14	0.1	-6223.09	-331.18

From this analysis, we can see that environment topology provide useful information to all of these measures except for Symmetric Average Distance, Symmetric Clustering, and Symmetric Betweenness. The improvement in information is relatively slight, compared to other factors, but the difference is there.

The BIC score criterion show another pattern – room topology is a more useful term if you consider networks where directed ties are sufficient to allow a connection. This suggests, in turn, that some of the environmental topologies used in this analysis fostered agents with mismatched activation values, while some did not.

5.1.2.3 Cognitive limitations influence network structure

From the previous analysis, we can see strong evidence already that cognitive limitations influence network structure. We, however, did an additional analysis comparing cognitive limited agents at a specific threshold value (0) to agents that are able to recall all previous contacts using the symmetric-necessary networks. Those results are in Table 3. MOVE PROSE TO HERE

Table 3. network measures for agents with no memory limits and those limited by a cog arch

Measure		**95% CI +**	**Mean**	**95% CI -**	**F**	**Sig.**
Density	No Limit	0.781	0.769	0.757	9136.43	< .005
	Limited Recall	0.112	0.106	0.100		
AVG Distance	No Limit	1.033	1.030	1.025	347.86	< .005
	Limited Recall	1.889	1.808	1.726		
Clustering	No Limit	0.909	0.905	0.900	720.59	< .005
	Limited Recall	0.355	0.338	0.320		
Eigenvector	No Limit	0.034	0.033	0.031	3822.25	< .005
	Limited Recall	0.340	0.322	0.304		

Betweenness	No Limit	0.006	0.005	0.004	1002.96	< .005
	Limited Recall	0.103	0.097	0.090		

These results tell us, strongly and unambiguously, that cognitive limitations are important for understanding these networks. The density results also suggest that networks of co-occurrence are often likely to vastly over-exaggerate the number of meaningful connections between human actors. Although two people may have been in the same room (or even the same elevator) at the same time, they may not have any memory of each other.

5.1.3 Influential models

As we showed in Table 3, the Model 6 shows the best fit of our simulation data because it has the best R-Square outcomes, all R-Square values of Model6 are high. Consequently, we believe that Model 6 could provide a reasonable influential model to explain the relations between network measures and dependent influential variables.

Directed

Variables	**Intercept**	**Threshold**	**Population**	**Running length**	**Room Map**	**Square threshold**	**Spline**
Density	0.28217	-0.09378	-0.08592	0.01921	0.0031	0.00625	-0.08938
Average distance	0.89594	-0.12322	0.03171	0.12651	-0.00398	-0.02339	-0.13189
Clustering co-efficiency	0.39878	-0.11005	-0.0347	0.02696	0.00101	0.00346	-0.10868
Eigenvector	0.1314	-0.00343	0.00911	0.01616	-0.00158	-0.00706	-0.00146
Betweenness	0.0249	-0.00046956	-0.00207	0.00489	0.00023542	-0.00146	-0.00219

Symmetric

Variables	**Intercept**	**Threshold**	**Population**	**Running length**	**Room Map**	**Square threshold**	**Spline**
Density	0.43969	-0.09717	-0.08986	0.06593	0.00686	-0.0031	-0.09632
Average distance	0.82885	-0.12635	0.05859	0.04028	-0.00625	-0.01794	-0.10208
Clustering co-efficiency	0.54157	-0.11284	-0.02579	0.05836	0.00304	-0.0049	-0.11085
Eigenvector	0.08018	-0.00305	0.01979	-0.01177	-0.00108	-0.00399	0.00514
Betweenness	0.0249	-0.00046956	-0.00207	0.00489	0.00023542	-0.00146	-0.00219

Table4 shows the regression result of Model6. Each number represents the contribution of each dependent variable to a specific measure. Based on this table, we could give an influence equation for every measure. In Eq1, we show an example equation of symmetric density.

$$SymmetricDensity = 0.28217 - 0.09378 \times Threshold - 0.08592 \times Population + 0.0121 \times RunningTime + 0.0031 \times MapConnectiveness + 0.00625 \times ThresholdSquare - 0.08938 \times Spline \quad \text{(Eq1)}$$

This equation clearly shows that activation threshold, population, and spline between -3 and 3 has negative influence on the symmetric density. This result aligns with our earlier result(Zhao, Kaulakis, Morgan, Hiam, & Ritter, 2012) suggesting that density decrease when increase threshold and population. The parameter of Spline variable is relatively large suggests that there is a significant difference between the range (-5,5) and range (-3,3). This also matches the Figure 4, which shows the measures changes mostly between -2 and 2.

5.2 Experiment2

In this section, we show the effects of the models three parameters on the rate of tie formation. Based on these figures, we will discuss how memory thresholds, map configurations, and navigation strategies influence the formation rates of simulated networks. This section focuses on showing the influences of navigation strategies and map configurations.

5.2.1 Experiment parameters

This experiment is a preliminary test, and we only run one set of parameters rather than a full set of conditions like Experiment 1. Because drawing the growth curves of activation networks needs to take numerous sample points on a curve, which requires nearly thousands of running hours.

In this preliminary experiment, we use samples' population size as 40 with a running time of 500s. To draw the growth curve, we collected network data at 18 time points, ranging from 10s to 500s. The sample points range from 10 s to 500s

5.2.2 Growth patterns of activation network

Figure 6 shows the growth curve of a network consisting of agents using the fixed-path navigation strategy within the Hallway map. The lower line represents the network formation rate of a network where no memory threshold was applied—if an agent met an agent, they formed a permanent tie (the directed-sufficient method). We find that the lower curve increase rapidly and then flattens when it reaches 1,336 ties (the maximum is 40*39, or 1,560, if the agents' paths completely overlap, which they do not). This flattening occurs once the network has achieved equilibrium and is fully connected.

In Figure 6, the top solid line represents the network formation rate of a network where an activation threshold of 0.0 was applied. According to the ACT-R theory, the activation threshold represents a memory limitation, meaning that memory chunks with an activation value lower than the threshold cannot be retrieved. The top curves more gradual progression illustrates the influence of memory on the formation rate, multiple exposures are required to remember another agent, while the difference in total number of links (800 versus 1,336) illustrates memory's effect on the network's density. In addition, this network never achieves a fully connected state, in the sense that the agent's declarative representation at no point includes the total set of possible interactions. In other words, these agents must continue to maintain their relationships because they continue to forget. Nevertheless, this network does eventually achieve equilibrium at 150 seconds with a network size of 800 links.

Comparing the two solid curves in the Figure 6, we noticed another difference, the time at which the rate of growth begins to increase. For the thresholded network, this time happens later than for the un-thresholded network. This is because the agents tie formation requires multiple exposures. Initially, agents are busy simply encountering other agents and building their friends list. As they, however, begin to meet more "old friends", the activation values of friendships start to increase.

The x-axis of the Figure 6 represents the simulation running time in real seconds. In our experiment, we set the travel interval between rooms at 16 seconds to make the effect of memory decay more prominent. Nevertheless, this interval is still not long enough to be realistic because people might take minutes or hours to find another person. As this work only focuses on the growth pattern of the social network, we would argue that the measurement of time is a secondary factor of our study because over 80 percent of the decay happens in the first 16 seconds according to ACT-R's decay function, with little additional decay occurring at greater time scales. Consequently, we believe total running time of 500 seconds and a short travel interval of 16 seconds are acceptable for initial explorations simulating the growth pattern.

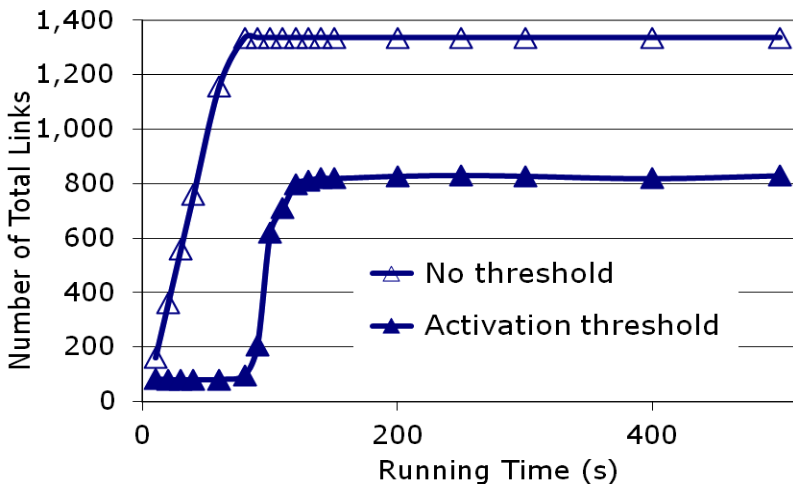

Figure 6: The effect of memory threshold on network formation over time for the fixed path navigation strategy in the hallway map (n=40).

Figure 7 shows the growth curve of a network of agents moving through the Hallway map using the random navigation strategy. Comparing Figure 6 with Figure 7, the non-threshold curves have the same growth pattern, but the threshold curves appear to be different. Memory appears to have different effects based on the setting in which the agents operate.

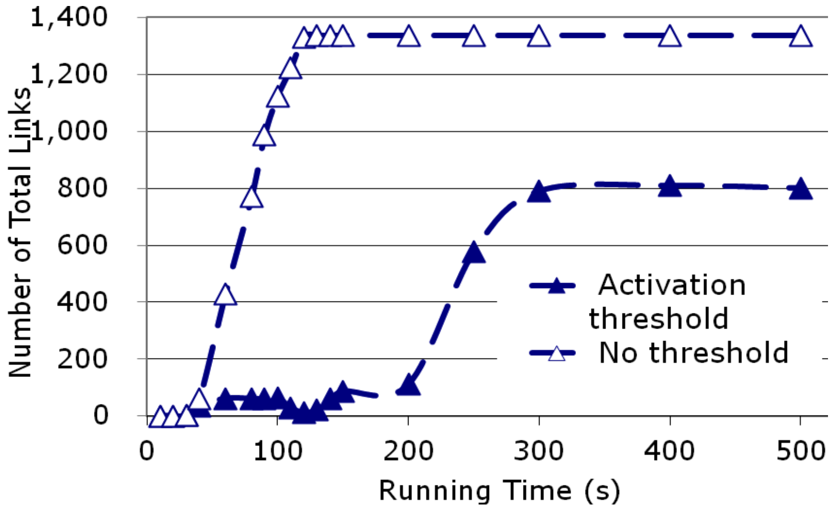

Figure 7: The effect of memory threshold on network formation over time for the random walk strategy in the hallway map (n=40).

Figure 8 compares the growth curves of two networks where a memory retrieval threshold of 0.0 was applied; these networks differ with respect to the navigation strategy used by their members. The fixed path-strategy (dash line) forms ties more quickly than the random-path strategy. We suspect that the fixed-path strategy achieves equilibrium sooner because it is more localized, and thus provides more chances for agents to meet their "old friends". On the other hand, both networks achieve equilibrium at about 800 links, suggesting that the navigation strategies in this simulation do not constrain the number of relations an agent can maintain in memory.

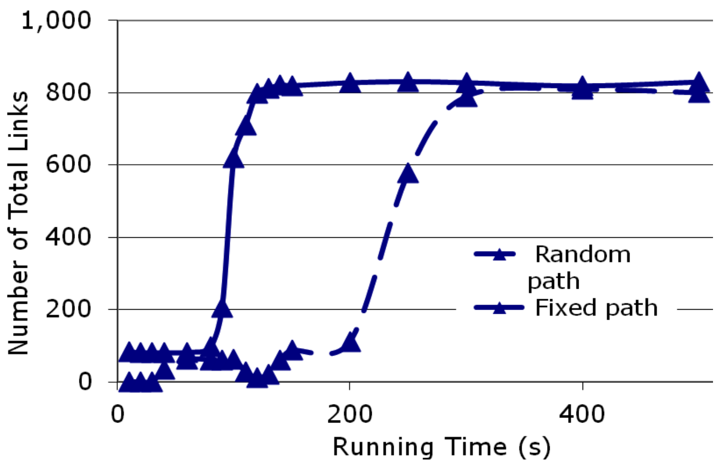

Figure 8: The effect of navigation strategy on network formation over time in the hallway map with threshold (n=40).

Figure 9 compares the network formation rates of networks occurring in each of the three map configurations (full grid, central, and hallway); all these networks consist of agents with a memory activation threshold of 0.0. We find that the map configurations have a similar influence on the networks' growth curves as the navigation strategies. Again, the map configurations influence the rate of formation but not the network's density at equilibrium.

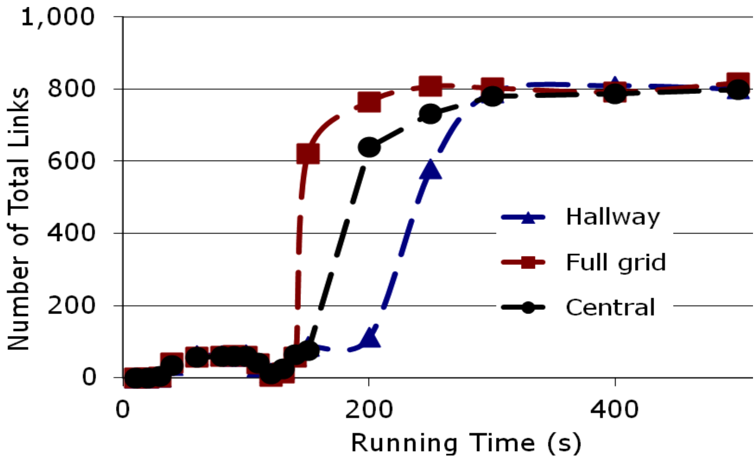

Figure 9: The effect of map configuration on network formation over time (n=40).

Comparing the three curves, we find the Hallway map (grid ratio=60%) is associated with the longest delay in network formation and the lowest rate of increase; the full grid map (grid ratio=100%) has the shortest delay and the fastest rate of link formation. These results show that delay in the network's growth rate is negatively correlated with map configuration (operationalized as grid ratio), while the network's growth rate during its growth spurt is positively correlated.

5.3.4 Interaction Density on Locations

Additionally, environment configurations can create loci of interaction or activity spaces (Brantingham & Brantingham, 1993). These locations are where the majority of all interactions occur. Brantingham and Brantingham use this concept to study crime densities, but this idea can be expanded to other activities, such as co-occurrence or socialization. When traveling to or between these spaces, people tend to take routine paths. Costanzo et al. (1986) demonstrated that people near one another tend to travel along the same paths to activity hotspots. Therefore, we expect that agents will also tend to take high frequency paths to common locations because they are constrained by the world's geometry.

These high activity spaces for one of our environments are shown in Figures 10a and 10b. Figure 10a shows a heat map of room activity, while Figure 10b shows the connectivity between all agents and the rooms in which they have interacted. Given the concentration and degree of these spaces, we show that agents who traveled between activity spaces tended to travel along the same path. This result is similar to the water-cooler effect (DiFonzo, 2008), which suggests that interaction happens naturally in shared public locations.

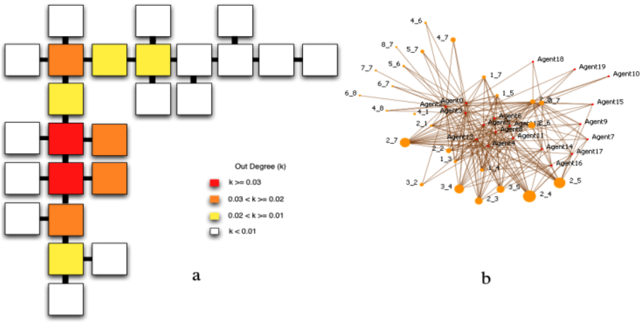

Figure 10. (a) Hallway map's heatmap; (b) Agents-by-location network.

6. Discussion and Future Work

In this study, we created a multi agent social network simulation that provides a flexible platform to examine several influential factors in social networks. Based on the existing literature (Brantingham & Brantingham, 1993; DiFonzo, 2008), we distinguish between environmental and cognitive factors. Our environmental factors included population size, run time, and map configuration. We focused on one cognitive factor, memory activation. Our experiments' results show the impact of these factors on the *network topologies* of our agents' interaction networks and on the *growth patterns* of these networks.

We first examined whether and how these factors influenced the *topologies* of our agent networks by using five independent variables, specifically threshold, threshold square, environmental topology, running time, and agent count to determine how much each of these factors could account for variation in five network measures of interest. The five network measures of interest were density, average distance, clustering, average eigenvector, and average betweenness. We used a model selection procedure to determine whether each independent term was useful in accounting for variation in each dependent term. We found that all three environmental factors were useful, although the environment topology term was the least useful. Our cognitive factor, memory thresholding, accounted for a large amount of the variation as well. We found that network that all three factors explained variation in these measures, and that these measures significantly differed with respect to cognitive threshold.

The usefulness of the environmental topology term differed depending on what semantic was used to generate the network. When one-way perceptions were sufficient to suggest a tie, environmental topology was a more useful term. This suggests that certain environmental configurations allow more asymmetry in the perceptions of the agents within them.

To test the influence of cognitive limit, we also did an additional analysis to compare the network without a cognitive limit and the network with an activation threshold 0. From Table 3, we can find that the comparison results are very significant ($p < 0.005$) in all five measures. This means the cognitive limitations are important for understanding these networks. It also suggests that networks of co-occurrence that do not represent the limits of memory are likely to vastly over-exaggerate the proliferation of connections between human actors.

We also generated the distribution plots of five measures when we increase the threshold value from -5 to 5. The plots align with the result we found in Table 3, and we also found Symmetric network is more sensitive to cognitive limitation. Another interesting finding is that five measures started to decrease/increase when the threshold is around -1. Based on ACT-R theory, we believe this turning point should be related to memory decay speed and other mental noise factors. Examining the relations between this turning point and other cognitive factors will be a important topic in our future work.

When examining the formation of our networks, the effect of running time was not as significant as we expected, and shows plateauing after 250s run for these configurations. The large running time also weakens the effect that map configurations have because it provides agents sufficient time to travel around the whole map. By examining interaction density on locations, we also found that the shared public locations have higher interaction densities, resembling Brantingham and Brantinham's (1993) focal loci.

Taking advantage of the ACT-R memory mechanism, we were able to create an egocentric view of our agents' interaction networks in memory by reconstructing the declarative representations used by our agents to recall past associations. We then merged these egonets to create a system wide representation of our agent's recalled interactions. We found the structure and density of these merged egocentric networks to depend heavily on the criteria for tie formation, *directed-sufficient* and *symmetric necessary*.

By examining the activation values between agents, we also found that our model's four factors influenced the activation values of agent ties. The agent activation logs show that the population size has a negative influence on the average activation (smaller groups have stronger ties); that running time has a positive influence on the average activation value; and that map configuration has some influence on the average activation but that this change does not correspond to changes in the grid ratio. This suggests that grid ratio is not a sufficient measure of map configuration at least with these maps, and we need to find a more accurate measure in the future. We also found that navigation strategies do influence activation values, with the Fixed Path strategy resulting in a neighborhood effect (strong localized ties). Further, we found that environmental factors did impact our network measures, finding running time to be the chief main effect explaining the variance in our density, average distance, clustering, and eigenvector across all the networks. For betweenness, population size contributed more information, but this is not surprising as betweenness is a path measure. Population size, more generally, was the most influential parameter for our thresholded networks (activation values of 0). This is as expected because thresholded agents must maintain ties in memory through repeated exposures.

Examining *network formation*, we found that navigation strategies and map configurations did influence network formation. Holding population size and run time constant, we examined to what degree cognitive limitations (represented by a memory activation threshold) influenced a network's generative process. The results suggest that cognitive limitations influence both the rate of network formation and the network's size at equilibrium. These findings roughly mirror empirical studies (Brantingham & Brantingham, 1993).

We can view the progression of the curves in Figures 6-9 as corresponding to three stages in network formation, though at abbreviated time scales. Between 0 and 100 seconds, the size of the network does not grow significantly, and the average number of relations stays constant at 60. This represents the tendency of people to initially remain in localized relations with a few people. Between approximately 100 to 150 seconds, there is a rapid increase in ties as they become more familiar with a new activity space. Finally, between 150 to 500 seconds, the network stops growing because the agents have shifted from primarily establishing to maintaining their friends network. In the stable state, the number of total links stays around 800 in these networks, meaning that the average number of relations in these networks is about 20.

In the second analysis, we examined the influences of two navigation strategies. The results suggest that navigation strategies do not influence the ultimate densities of our networks over time, with non-thresholded networks reaching a higher density than thresholded networks. Navigation strategies do, however, change the growth pattern of both networks. Figure 6 showed that the network using the fixed-path strategy grows much faster. This is because the fixed-path strategy is a more focused strategy that provides more chances for people to meet their "old friends". In this case, people more easily form small groups associated with their starting location, such as people living on the same street or attending the same school. We see that the fixed-path strategy facilitates the rapid creation of smaller tighter groups than the random-walk strategy.

The third analysis focused on examining the influences of spatial configurations on generative networks. We operationalized map configurations using a constant (grid ratio). We defined grid ratio as the ratio of the number of edges over the total number of possible edges to quantify the connectivity of the map configurations. We find that delay increment is negatively correlated with grid ratio, while the formation rate during the growth phase is positively correlated with the grid ratio. This result validates our definition of grid ratio, because it shows the grid ratio does have influence on network formation; it also proves that lower grid ratio maps with more gaps and obstacles decrease the network's growth rate. We also found that our map configurations did not influence the final density of the network over time, but did influence its rate of growth.

Comparing the two rounds of analyses, we noticed an interesting conflict: the environmental topology showed relatively slight influence on network measures in the first round of experiment, but it exhibited significant influences on the growth pattern of networks. After examining Figure 9, we noticed that the significant differences between map configurations happened between 150 seconds and 250 seconds, however, the first round of experiments used 125 seconds and 250 seconds as running time, which do not include the influence of map configurations. Combining these two experiments, we may conclude that the influence of map configuration changes over time. It is slight at the beginning, but will increase at a point (approximately 150 seconds) and the influence will diminish when the networks reach an equilibrium. We will rerun the experiment with different running times to verify this conclusion in the future work.

Future avenues of work will build upon some of the more interesting issues. First, we would look at analysis of normalized thresholds to see if there are regularities in their effects on network topology. Second, we should run more agents and more runs (Ritter, Schoelles, Quigley, & Klein, 2011), because the system to demonstrate these effects was kept deliberately small. Finally, we would extend our analysis on the effects of cognition on network measures analogous to Dunbar's Number.

Acknowledgements

This work was supported by a grant from DTRA (HDTRA1-09-1-0054). We thank Jaeyhon Paik for

help on the agent design and Jeremy Lothian for assistance with VIPER.

References

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An intergrated theory of the mind. *Psychological Review, 111*(4), 1036-1060.

Arcangeli, A., Eidus, I., & Wright, C. (2009). Increasing memory density by using KSM. In *Proceedings of the Linux Symposium*, 19-28.

Barrett, C., Eubank, S., & Marathe, M. (2006). Modeling and simulation of large biological, information and socio-technical systems: An interaction based approach. In D. Goldin, S. A. Smoka & P. Wegner (Eds.), *Interactive computation: The new paradigm* (pp. 353-394). Berlin Heidelberg: Springer.

Brantingham, P. L., & Brantingham, P. J. (1993). Nodes, paths and edges: Considerations on the complexity of crime and the physical environment. *Journal of Environmental Psychology, 13*(1), 3-18.

Carley, K. M. (1991). A theory of group stability. *American Sociological Review, 56*(331-354).

Carley, K. M. (1992). Organizational learning and personnel turnover. *Organizational Science, 3*(1), 20-46.

Costanzo, C. M., Halperin, W. C., & Gale, N. (1986). Criminal mobility and the directional component in jour¬neys to crime. In R. Figlio, S. Hakim & G. Rengert (Eds.), *Metropolitan Crime Patterns* (pp. 73-96). Monsey, NY: Criminal Justice Press.

DiFonzo, N. (2008). *The watercooler effect: a psychologist explores the extraordinary power of rumors*. New York: Avery.

Dunbar, R. I. M. (1998). *Grooming, gossip, and the evolution of language*. Cambridge,MA: Harvard University Press.

Gonzales, C., & Lebiere, C. (2005). Instance-based cognitive models of decision making. Pitsburg: Carnegie Mellon University.

Juvina, I., Lebiere, C., Martin, J. M., & Gonzalez, C. (2011). Intergroup prisoner's dilemma with intragroup power dynamics. *Games and economic behavior, 2*(1), 21-51.

Lebiere, C., Gonzalez, C., Dutt, V., & Warwick, W. (2009). Predicting cognitive performance in open-ended dynamic tasks: A modeling comparison challenge. In *the 9th International Conference on Cognitive Modeling*. Manchester, UK.

McCarty, C., Killworth, P. D., Bernard, H. R., Johnsen, E. C., & Shelley, G. A. (2001). Comparing two methods for estimating network size. *Human Organization, 60*(1), 28-39.

Milgram, S. (1967). The small world problem. *Psychology Today, 1*(1), 61-67.

Morgan, G. P., & Carley, K. M. (2011). *Exploring the impact of a stochastic hiring function in dynamic organizations*. Paper presented at the Brims. Retrieved.

Morgan, J. H., Morgan, G. P., & Ritter, F. E. (2010). A preliminary model of participation for small groups. *Computational & Mathematical Organization Theory, 16*(3), 246-270.

Newell, A. (1994). *Unified theories of cognition*. Cambridge: Harvard University Press.

Newell, A., Rosenbloom, P., & Laird, J. (1989). Symbolic architectures for cognition. In M. Posner (Ed.), *Foundations of Cognitive Science*. Cambridge: MIT Press.

Reitter, D., & Lebiere, C. (2010). A cognitive model of spatial path planning. *Computational and Mathematical Organization Theory, 16*(3), 220-245.

Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*(2), 461-464.

Silverman, B. G. (2004). Human performance simulation. In J. W. Ness, D. R. Ritzer & V. Tepe (Eds.), *The science and simulation of human performance* (pp. 469-498). Amsterdam: Elsevie.

Simon, H. A. (1991). Bounded rationality and organizational learning. *Organizational Science, 2*(1).

Yen, J., Yin, J., Ioerger, T. R., Miller, M. S., Xu, D., & Volz, R. A. (2001). CAST: Collaborative agents for simulating teamwork. In *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence (IJCAI-0 1)*, 1135-1142. Los Altos, CA.

Zhao, C., Kaulakis, R., Morgan, J. H., Hiam, J. W., & Ritter, F. E. (2012). Socio-cognitive networks: modeling the effects of space and memory on generative social structures. In *the Proceedings of the 21st Conference on Behavior Representation in Modeling and Simulation. 12-BRIMS-007. 24-31. BRIMS Society*. Amelia Island, FL.