Comput Math Organ Theory (2011) 17:318–343
DOI 10.1007/s10588-011-9088-4

Leaving us in tiers: can homophily be used to generate tiering effects?

Brian R. Hirshman · Jesse St. Charles · Kathleen M. Carley

Published online: 10 May 2011

Abstract Substantial evidence indicates that our social networks are divided into tiers in which people have a few very close social support group, a larger set of friends, and a much larger number of relatively distant acquaintances. Because homophily—the principle that like seeks like—has been suggested as a mechanism by which people interact, it may also provide a mechanism that generates such frequencies and distributions. However, our multi-agent simulation tool, Construct, suggests that a slight supplement to a knowledge homophily model—the inclusion of several highly salient personal facts that are infrequently shared—can more successfully lead to the tiering behavior often observed in human networks than a simplistic homophily model. Our findings imply that homophily on both general and personal facts is necessary in order to achieve realistic frequencies of interaction and distributions of interaction partners. Implications of the model are discussed, and recommendations are provided for simulation designers seeking to use homophily models to explain human interaction patterns.

Keywords Homophily · Social group size · Tiering · Agent-based simulation · Dynamic network analysis

B.R. Hirshman (✉) · J. St. Charles · K.M. Carley
Computation, Organization and Society Program, School of Computer Science, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA
e-mail: hirshman@cs.cmu.edu

J. St. Charles
e-mail: jstcharl@cs.cmu.edu

K.M. Carley
e-mail: kathleen.carley@cs.cmu.edu

1 Introduction

Homophily—the principle that like seeks like—has been suggested as a fundamental mechanism which can drive interaction (Lazarsfeld and Merton 1954; McPherson et al. 2001). Homophily can lead to increased interaction between pairs or groups of individuals and has been related to outcomes at the dyad, small group, and network level. However, can a simplistic homophily model generate the overall macro patterns observed in human ego networks? It is widely accepted that our social networks are in some fashion tiered, containing both strong and weak ties (Granovetter 1973; Marsden 1987). Can homophily be used as a mechanism that both generates the observed human network distribution of strong and weak ties and predicts the relative frequency with which an ego interacts with them?

Human social networks are not uniform in terms of tie strength, and indeed we are known to have few very strong ties and many weak ties (Granovetter 1973; Hill and Dunbar 2003; Zhou et al. 2005). While aggregations of these different ties have been called many names in the literature, we choose to call such aggregations 'tiers' in homage to the 'tiered grouping' concept identified by Zhou et al. (2005). It is widely accepted that a number of factors can lead to inequality in the frequency and strengths of relationships. Factors such as geography (Butts 2002), family (Agneessens et al. 2006), education (Marsden 1987), and employment (McPherson et al. 2001) have been shown to have important effects on the composition of human social networks. The role of a more general homophily mechanism, however, has often been overlooked. Indeed, many empirical studies, in trying to stress the importance of one or more particular factors, minimize the role of homophily in the formation of such tiers. Thus, this paper asks a slightly different question than has commonly been asked: is the principle of homophily a sufficient condition for generating the patterns of tie strength that match the empirical literature? Is it possible to observe empirical distribution of tiering patterns without explicitly modeling geographic and other factors? To the best of our knowledge, previous work has not attempted to address this question.

The question of the whether homophily can lead to tiering behavior has important practical considerations, especially for simulations. For instance, the accurate modeling of networks and tiering behaviors may have profound effects on attempts to model large-scale social processes such as models of disease spread (Mniszewski et al. 2008), information diffusion (Cowan and Jonard 2004; Rogers 1995; Valente 1995), belief propagation (Friedkin and Johnsen 1999), web services (Leskovec and Horvitz 2008), meme propagation (i.e. Dawkins 1976), and others. Such models, in turn, can be critical for policy makers, system designers, managers, and executives seeking to understand how a network or system will evolve. Additionally, an improved understanding of interaction mechanisms can be useful for designing and applying successful interventions to modify dynamical social systems (Carley 2003). Without accurate models of interaction frequency, and without an understanding of why such interaction patterns occur, simulation models may suggest inaccurate or ineffective interventions.

We use Construct (Carley 1991; Carley et al. 2009), a multi-agent dynamic-network simulation, to understand whether and how homophily can lead to the formation of tiers.[1] Specifically, we simulate the emergence of social ties using Construct's implementation of homophily in order to understand whether it can lead to the tiering behavior observed by anthropologists and sociologists. We then examine the strength of these interactions by examining simulation results by tier. Our primary finding is that a simple knowledge homophily model is insufficient to explain the tiering behavior frequently observed in human societies. We therefore introduce and verify a plausible modification to the homophily model that produces networks and interaction frequencies that align with well-documented social patterns.

The remainder of this paper is organized as follows. Section 2 describes past social and computational work investigating the tiered nature of social systems. Section 3 presents our tool, Construct, while Sect. 4 describes the virtual experiment used in our investigation and Sect. 5 presents our analysis methodology. Section 6 discusses our results, and Sect. 7 presents the broader socio-cognitive and simulation implications that can be drawn from our work.

2 Background

The tiered nature of human social networks is an area of considerable active research with a rich history that can be traced back decades (Dunbar 1993, 1998; Granovetter 1973; Marsden 1987; Zhou et al. 2005). While active research in this area is still ongoing, and though culture and context are known to play important roles in shaping a social network, several general trends have been observed. For instance, the majority of members in most human societies tend to have a small group of close friends or strong ties (Granovetter 1973) from whom they draw core support (Marsden 1987), a larger group of friends whom they stay in touch with regularly, a collection of weaker ties with whom they interact less often, and acquaintances with whom they interact on an infrequent basis. This tiered structure can be described using a network perspective by saying that each ego agent in the network is strongly connected to a small number of others, less strongly connected to quite a few, and weakly connected to many more (Dunbar and Spoors 1995). In other words, tie strength is inversely proportional to the number of alter agents that have similar tie strength. In this context a tier can be conceptualized as a bin of alters in a histogram of tie strength. Distant ties or weak ties are important sources of information for people and can serve as the underpinnings of a successful community (Granovetter 1973). Large social networks, however, are well known to be sparse (Leskovec et al. 2005) and full of structural holes (Burt 1992). In both large and small networks these quantities can be important for determining power, influence, and types of action (Wasserman and Faust 1994). General social network structure, both at the local level as well as global level, has been correlated with health, happiness, satisfaction, and many other attributes of network members (Christakis and Fowler 2007; Freeman 1979;

[1] We use Construct 3.9, a completely refactored version of Construct that utilizes new agent technology and is composed of more robust and validated mechanisms. Additional details and technical literature about the tool can be found on the project website, http://www.casos.cs.cmu.edu/projects/construct/.

Wellman and Wortley 1985). However, the number of tiers and the frequency of interaction with members of each of these tiers is a matter of rich discussion (Zhou et al. 2005).

Human social networks are known to have very different network properties as compared to other non-human networks such as neuronal networks or power grids (Newman and Park 2003). While strict structuralists contend that social network behavior can be largely explained by social position, there is increasing evidence which demonstrates that individual and cognitive properties also play an important role in determining interaction patterns within networks. Analyses of tiered networks from the small (Kilduff and Krackhardt 1994; Levine and Moreland 1998; Stiller and Dunbar 2007) to the very large (Bandura 2001; Leskovec and Horvitz 2008) increasingly take into account the individual attributes of networked actors. To operationalize the presence of these attributes one can take a meta-network perspective (Carley 2003; Krackhardt and Carley 1998). The meta-network is an ontology that supplements a standard social network with information about individual knowledge, beliefs, and attributes. Such additional factors can help explain network change and evolution over time (Carley 2003).

Cognitive limits also play an important role in determining the size and structure social network tiers. Anthropologists such as Dunbar postulated a physiological basis for social group size in the early 1990s—Dunbar's "social brain hypothesis" suggested that humans' social behavior was at least partially constrained by biological factors (Dunbar 1993, 1998; Hill and Dunbar 2003). Specifically, Dunbar (1993) ran a regression of neocortex and group sizes for different primates species and found evidence that humans should have a mean social network size of 150 people with a 95% confidence interval between 100 and 230. Initial support for this assertion was found in tribal and military structures, but several other types of corroborating evidence have subsequently been observed. Mathematicians and others have seized upon these results in order to demonstrate the relationship between cognitive limitations and general social structures. Zhou et al. (2005) extended Dunbar's work to provide mathematical support for a concept that sociologists had long posited: subgroup sizes among people were likely to be tiered by emotional distance. Their analysis suggested that most humans had 3–5 alters in the "core discussion group" portion of their network, 15–20 in the "sympathy group", 30–50 in the "band", and 150 in the clan or regional group, and 500 and circa 2000 in increasing shells around the ego like babushka dolls, though it is important to observe that each larger network is not always identical in structure or organization to the smaller one (Dunbar and Spoors 1995). Members of the physics community have used such findings, in conjunction with Milgram's (1967) small world results, in order to build networks to investigate diffusion social processes (e.g. Lopez and Sanjuan 2002). Researchers from multiple disciplines have seized upon Dunbar's result in order to generate plausible social network structures. For instance, work with Dunbar's cognitive limits has been used to inform and support generators for building plausible communities such as those found in large social and information networks like citation networks or peer-to-peer file sharing networks (i.e. Leskovec et al. 2008).

The principle of homophily presents a way to relate cognitive abilities with larger social patterns and thus can serve as the glue between micro behavior and macro structure. Homophily, as initially defined by Lazarsfeld and Merton (1954), was specified as "a tendency for friendships to form between those who are alike in some designated respect." While this definition has been extended in various ways by different authors, for instance to generalize the definition to encompass general interactions in addition to friendship, most authors continue to note that homophily is often contextual and impacted by one or more salient dimensions (i.e. McPherson et al. 2001). This "designated respect" can be contextualized in many ways and can be socio-demographic (Borgatti and Foster 2003; Harrison et al. 1998; McPherson et al. 2001), geographic (Butts 2002; McPherson et al. 2001; Wellman 1996), due to shared interests (Carley 1986; McPherson et al. 2001), related to common knowledge (Borgatti and Foster 2003; Carley 1986; Rogers 1995), a function of shared beliefs (Harrison et al. 1998; Lazarsfeld and Merton 1954), and/or related to common behavioral patterns (McPherson et al. 2001). Many of these factors are essential for building both close and distant relationships, and the greater the overlap on multiple dimensions the better (McPherson et al. 2001). However, homophily is neither monolithic nor static. For instance, the degree of similarity on a single dimension can change, which can affect interaction between different individuals, or more salient dimensions may become applicable and lead individuals to interact with different alters. Nevertheless, if individuals can modify their position on the salient dimension, frequent interaction may lead them to become even more similar as they interact and exchange information. Such a process of increasing similarity is at the heart of the constructuralism, a theory which suggests that social networks and knowledge networks co-evolve and changes in the who-knows-what network induce changes in who-knows-who network, and vice versa (Carley 1986, 1991).

If those who are similar to each other become increasingly so—via a preferential-attachment-like mechanism where the similar become even more similar (e.g. Barabási and Reka 1999)—then homophily may be one way of growing a society where individuals have the pattern of friendships observed by anthropologists, sociologists, and others studying the cognitive abilities of human beings. This would be an example of generative social science (Axtell et al. 1996; Epstein and Axtell 1999; Goldstein 1999). If a generative social mechanism based on homophily is unable to generate realistic social networks, it would be an indication that omitted factors are perhaps more important in determining the shape and structure of human behavior. While homophily or even contextualized homophily may not be the dominant factor actually driving human tiering behavior, a set of virtual experiments could demonstrate whether or not such a mechanism could yield realistic results.

A computational model is ideal for examining this behavior since it would be difficult to envision a set of laboratory or field experiments that could assess only the effects of homophily on subject populations. Since human social networks are quite large, it would be practically impossible to grow a realistic social network with hundreds of people in a laboratory setting. More importantly, it can take a substantial amount of time for a social network to grow to fruition and for individuals to begin to find the types of deep homophily that are needed for sustained relationships (Harrison et al. 1998). Observational studies of real world behavior suffer from other limitations. Accurate data collection would be challenging for very

Table 1 Comparison of simulations

Type of model	Cognitive model	Social model	Tiering	Examples
Cognitive models	often good	often non-existent	usually restricted to a single agent and particular task	ACT-R (Anderson 1983) SOAR (Laird and Congdon 2006)
Small group or social models	often adequate	often good	generally take social space as given and evolve patterns there	Friedkin belief models (Friedkin and Johnsen 1999)
Societal or swarm intelligence models	often simplistic	often simplistic	little cognitive representation, often lack a meta-network approach	Sugarscape (Epstein and Axtell 1999) Flocking models (i.e., Olfati-Saber 2006)

large network for reasons including cost, privacy, and methodology. Since the collection of data can limited by subject fatigue, and because observers cannot follow a particular person everywhere to record all of their interactions, it would be extremely difficult to track all of a person's acquaintances; even if such problems could be overcome, there is considerable debate as to what constitutes a link between two people in a social network, how robust this link is to measurement, and how individual differences in self-reporting can skew results (e.g. Borgatti et al. 2006; Wasserman and Faust 1994). However, even with perfect data, it likely would be impossible to demonstrate how homophily alone could produce the total pattern of interactions of an individual. Not only would it would be necessary to understand the dimensions on which the individual and partner were similar, but it would also be necessary to rule out competing processes such as geographic proximity, economic pressure, and other factors. In the face of such difficulties, we believe a generative computer model is appropriate for addressing our research question.

The process of growing artificial societies in order to evaluate social hypotheses is not new and is becoming increasingly popular in a variety of disciplines (Carley 1995; Epstein and Axtell 1999; Ilgen and Hulin 2000). Nevertheless, using generative social science to examine the tiered nature of agent interactions has not, to our knowledge, been extensively explored. In framing our search for comparable models, we divided the modeling literature into three general areas—the individual, as seen in Table 1—that fall very close to the divisions that Alan Newell proposed in his hierarchy of cognition (Newell 1994, Fig. 3-3). At all three of these levels, models that grow realistic societies based on homophily have been underexplored. While powerful cognitive architectures such as ACT-R or SOAR have the ability to represent critical attributes and to model human behavior, such models have focused more on the modeling of individuals and have only recently begun to explore the rich social world. Though some cognitive models have attempted to examine social phenomena, most research is limited to a handful of agents—a quantity insufficient to explore tiering behavior in detail. Models that focus on small groups of individuals occasionally use homophily to help determine who should interact with whom, as well as to determine the salience of particular factors. However, most of this work is usually domain-specific. Other models that focus on slightly broader societal phenomena such as information or belief diffusion generally assume a lattice or other fixed structure on which agents interact; such structures prevent realistic tiering behavior from emerging since agents are often confined in their choice of neighbors. Though the most complex of such models examine more realistic structures such as small-world or hierarchical networks, even these models tend to leave the network fixed throughout the course of the simulation. Lastly, while population-level simulations tend to explore very large phenomena such as flocking or polarization, the agents in these models are usually assumed to interact with a fixed set of neighbors. More importantly, there are usually only a very small number of salient attributes that drive interaction. This does not allow for a more complex study of general homophily. While we recognize that our survey in Table 1 captures only a small fraction of the literature on agent-based socio-technical simulation, our reading of the agent-based simulation literature suggests that tiering has been underexplored and that simulations that tend to employ homophily do so once the pattern of interactions has largely been defined.

It is worth mentioning, however, that a number of researchers have approached the problem of generating or identifying realistic social networks. Techniques such as the ERGM family of models have been used to test a variety of hypotheses about empirical networks, including homophily based on one or more attributes (Anderson et al. 1999b). Models such as SIENA can be used to evaluate the effects of homophily over time and to attempt to determine whether homophily is a cause or an effect of network structure (Steglich et al. 2006). The generation of plausible social networks has also been undertaken from a statistical perspective using Bayes Nets and other techniques (e.g. Thiriot and Kant 2008). Other models have used spatial locations to generate realistic social networks (Wong et al. 2006). While the approach taken in this paper is a generative one based on homophily, the question of building and identifying realistic interaction patterns continues to be a rich and active research area.

3 Construct

To examine the tiering effects of social ties, we employed version 3.9 of Construct. Construct is a dynamic network simulator that combines sociological, organizational, and cognitive theories to simulate the co-evolution of the social networks (who knows whom), knowledge networks (who knows what), and belief networks (who believes what). Construct was the first multi-agent simulation system that attempted to include realistic social networks at its core, and while the code has evolved and the system redesigned, the social network realism remains of paramount importance (Carley 1991). Construct is the embodiment of constructuralism, a mega-theory which states that the socio-cultural environment is continually being constructed and reconstructed through individual cycles of action, adaptation and motivation (Carley 1986). Past simulation work with Construct simulations has investigated the formation and changes to small groups (Carley 1991), the effect of non-human agents on organizational performance (Carley 1999; Carley et al. 2009), the effect of cognitive and

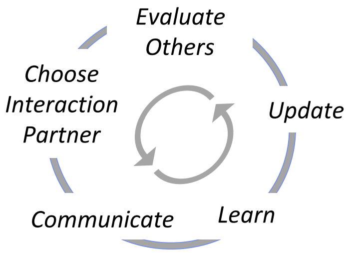

Fig. 1 The Construct interaction cycle

information access restrictions on information diffusion patterns (Carley et al. 2009; Hirshman and Carley 2008; Hirshman et al. 2008b; Hirshman and St. Charles 2009).

As Construct is an agent based model, it consists of multiple interacting yet independent agents. At the beginning of the experiment, the simulation designer initializes the agents, sets a number of important weights, and lets the simulation run for a number of time periods. During each time period the following sub-processes occur:

(1) agents compute a probability of interaction with available agents
(2) agents use these probabilities to select an available agent with which to interact
(3) agents communicate with one of their interaction partners, sending a subset of facts as a message
(4) agents learn new information from this communication
(5) agents update their knowledge to reflect what they have just communicated, and the cycle repeats.

This process is diagrammed in Fig. 1.

It is worthwhile to describe this cycle in slightly more detail in order to understand the operation of the simulation, though readers are also encouraged to examine the technical literature for additional details.

At the beginning of each time period, agents first employ their transactive memory to rank their possible interaction partners. Transactive memory represents an agent's perception of who knows what and can potentially be incomplete or even incorrect. As agents evaluate possible interaction partners, they use various criteria to create a relative score which reflects their preference for interacting with each other agent. Construct allows multiple factors to affect this probability of interaction: relative similarity (knowledge homophily), which occurs when an ego agent perceives that an alter knows a fact the ego already knows; knowledge expertise, which occurs when an ego perceives that the alter holds new information; socio-demographic (or surface-level) similarity, which is operationalized as similarity based on easily-observable properties (Harrison et al. 1998; Hirshman et al. 2008a; McPherson et al. 2001), physical proximity, which is similarity due to similar positions in the physical environment (i.e. Barnlund and Harland 1963; Butts 2002; Wellman 1996), and social proximity, a catch-all for other types of similarity not explicitly modeled.

In this experiment, we employ only relative similarity and expertise in order to drive agent interactions. Relative similarity and knowledge expertise, diagramed in Fig. 2, are fundamental to Construct's operation (Hirshman and Carley 2007a). The

Fig. 2 Similarity and expertise in Construct

	Ego perceives that alter does not know	Ego perceives that alter knows
Ego does not know	no effect	increases expertise
Ego knows	no effect	increases similarity

relative similarity of i and j, from i's perspective, is characterized as

$$relative\ similarity_{ij} = \frac{\sum_{k<K}(AK_{ik} * TM_{ijk} * W_{ik})}{\sum_{j<1}\sum_{k<K}(AK_{ik} * TM_{ijk} * W_{ik})}$$

where individual i's relative similarity to j, is determined in terms of similarity in the agent-to-knowledge matrix AK. The more facts that agent i knows (AK_{ik}) and perceives that j knows using its transactive memory about agent j (TM_{ijk}), the greater the similarity, though this similarity is tempered by the weight that agent i places on each fact k (W_{ik}). Since relative similarity is calculated in contrast to the rest of the agent population, it is necessary to normalize this similarity score using the score calculated for the remainder of the population in order to calculate a relative value.

Relative expertise is a search-based mechanism and derives from the idea that individuals are more likely to interact if one has information that the other wants. The relative expertise of agent j as judged by i is characterized as

$$relative\ expertise_{ij} \frac{\sum_{k<K}(X_{jk} * W_{ik})}{\sum_{j<1}\sum_{k<K}(X_{jk} * W_{ik})} \quad \text{where } X_{jk} = \begin{cases} TM_{ijk}, & AK_{ik} = 0 \\ 0, & \text{otherwise} \end{cases}$$

where individual i's relative similarity to j is determined in terms of differences in the agent-to-knowledge matrix AK (Schreiber et al. 2004). The more facts that agent i lacks ($AK_{ik} = 0$) but perceives that agent j knows when checking its transactive memory about agent j (TM_{jk}), the greater the expertise, though agent i also weights this expertise on a per-fact basis (W_{ik}). Note that relative expertise, like relative similarity, is also normalized per agent.

The relative interaction probability is then a weighted average of relative similarity and relative expertise. Other factors, such as socio-demographic homophily, can be included in the weighted term $static\ factors_{ij}$.

$$\begin{aligned} relative\ probability_{ijt} &= \alpha_t * relative\ similarity_{ijt} + \beta_t * relative\ similarity_{ijt} \\ &\quad + \gamma_t * static\ factors_{ij} \end{aligned}$$

where weights α, β, and γ must sum to one and can vary from simulation time period to simulation time period. In this set of simulations, however, no static factors were modeled ($\gamma = 0$ for all time periods) and similarity was assumed to be the driving force ($\alpha = 0.9, \beta = 0.1$). Thus, while our simulation was largely driven by homophily, we included a slight amount of information seeking behavior in order to allow agents to consider interesting interaction partners who had slightly different knowledge.

Once each ego agent has evaluated its potential alters and has calculated all relative probabilities, it then selects an alter with which to interact. This selection is based on the probabilities just computed. Alters that have higher relative probabilities of interaction will have higher chances of being selected as interaction partners, but such selections are not guaranteed. The order in which ego agents choose their alters is randomized each time period, so agents may be first to choose (and thus have a large pool of potential alters from which to select a partner) while others will choose later and may not have as many options. If some alters are not available for interaction because they are interacting with third parties, then the ego will only choose from those available. Note that egos will always consider themselves as a possible alter with whom to interact, a factor which will greatly decrease the selection of extremely dissimilar interaction partners. While agents are likely to exhibit strong preferences towards agents with which they are similar—since the relative probability of interaction with such agents will be high—the stochastic nature of Construct guarantees that agents will have the potential for occasional interactions with those who are not necessarily highly similar.

After all agents have chosen their appropriate interaction partners, each ego and each alter prepares a message to send to the other. While Construct contains several cognitive filters which may affect what agents choose to send in their messages (Hirshman and Carley 2008), such filters were turned off for this experiment. The prepared message usually consists of a subset of the agent's knowledge, but it can also contain transactive memory information about third parties—a factor that allows agents to learn about other agents who may have new and interesting knowledge. The exact nature of the message sent can be controlled by the simulation designer. Certain facts may be weighted more heavily and thus are more likely to be chosen in the message. Other facts may be weighted less heavily and thus will be less likely to be shared. This weighting allows certain facts to be considered more salient when agents are choosing interaction partners but less relevant when agents are choosing information to communicate.

Once agents have sent their messages, they are able to learn new information sent to them by others. Construct can apply several cognitive filters that affect how agents learn this information (Hirshman and Carley 2008), but such filters were also disabled in this experiment. All information sent by the alter agent is learned by the ego. During this learning phase, each agent also updates its transactive memory to reflect its new knowledge as well as its knowledge of its social environment. These changes in both knowledge and transactive memory may lead to modifications when relative similarity and relative expertise are recalculated in the next time period.

It should be noted, however, that new knowledge or transactive memory learned in one period may not lead to increased similarity or expertise in the next. If an ego agent learns a fact from an alter, the receiver will have increased similarity with the sender on an *absolute* scale; however, the same fact may cause the ego agent to become more similar to many other agents, thus leading to a decrease in *relative* similarity. For this reason, the results of agent behavior in Construct are highly non-linear—just like those of human behavior. Interaction between each pair of agents is as dependent on the similarity of the dyad as it is on the differences between each agent in the dyad and its other potential interaction partners. As Construct iterates over the course

Table 2 Design of virtual experiment

Condition	Values	Number of categories
Number of experiment parameters	3	
Experimental parameters		
Number of agents	250, 500, 750, 1000, 1250	5
Number of personal facts per agent	0, 1, 5, 10	4
Transmission weight per general fact	0, 1, 5, 10, 50	5
Number of experimental conditions	$5 \times 4 \times 5$	100
Total replications per condition	10	
Total number of simulation runs	1000	

of the virtual experiment, such non-linearities may have important effects on final outcomes, a feature that makes Construct a chaotic system.

During the update phase, the Construct simulation also records relevant outcome measures. For this experiment, only one type of measurement was made: the number of times each agent interacted with each particular alter. This data, recorded as an agent-by-agent network, was printed at the end of the simulation. While Construct can print a variety of other outputs, our focus on tiering behavior led us to use only this one output feature. Additional details on Construct outputs, as well as further information on the operation of the simulation, can be found elsewhere (e.g. Hirshman and Carley 2007a, 2007b).

4 Experiment design

In this experiment, we vary three simulation parameters: the size of the agent population, the number of personal facts unique to each agent that no other agent initially knows, and the likelihood that each agent will want to share general facts versus personal ones when communicating. Each of these parameters takes on multiple values, as can be seen in Table 2. Each will be discussed in turn.

The number of agents in the simulation was varied between 250 and 1250 to examine the effect of network size on simulation results.[2] The agent population was closed, meaning that agents did not enter, leave, or otherwise turn over once the simulation was started. Agents also could not exogenously learn new facts in the simulation as performed; agents could only learn new information from other

[2]Note that each Construct had to keep track of transactive memory for each other agent in the simulation, meaning that in the largest possible simulation all 1250 agents had to keep track of any knowledge known by each of the 1250 agents in the simulation. While Construct uses sparse memory storage and other optimizations, the nature of this experiment meant that each agent had to keep track of a very, very large amount of knowledge about potential interaction partners. While an obvious memory-saving strategy might be to exogenously prevent agents from storing information about some alters, such a modification would go against the purpose of this simulation: to see whether tiered structures could arise from homophily alone without any exogenous pruning. Construct is capable of using larger numbers of agents, but agents in such simulations will not store transactive memory about all others.

agents. The range of agents used was chosen for mathematical, social, and pragmatic reasons. Simple combinatorics suggests that the greater the number of agents, the greater the number of potential interaction partners among which each agent could choose. Modifying the number of agents in the simulation would allow us to investigate whether interaction partner availability substantially affects tier sizes and distributions. Increasing the number of agents in the population could also minimize the effect of random chance: in the smaller simulations, there may be fewer agents with which to interact and random chance interactions could have greater impacts on outcomes. Additionally, social network analysis has suggested that important social network properties are highly dependent on the size of the network and modifications to this parameter allow for examinations of such effects (Anderson et al. 1999a; Wasserman and Faust 1994). If the group is too small, or the number of potential alters not large enough, certain effects cannot be observed. However, such size considerations had to be balanced against the space and running-time considerations of Construct, which is $O(\text{agent}^2 \times \text{knowledge} \times \text{time})$ when running under the conditions of this experiment (Hirshman and Carley 2007a). This meant it could take tens of hours to run a large simulation. The values chosen for this experiment were selected to balance these competing pressures.

The number of personal facts served as a proxy for the importance of personal information in the simulation. Personal factors such as core values, important experiences, deep similarity, and close friendship are known to strongly influence individual behavior in social settings, and homophily on such factors is known to be especially strong (Harrison et al. 1998; Marsden 1987). As modeled in Construct, the personal facts were distinct for each agent, meaning that the set of personal facts for agent 1 was completely separate from those for agent 2. At the beginning of the simulation, only the owner of each personal fact would know it; during subsequent interactions, however, the owner could choose to share the fact with others. These personal facts had several properties that differentiated them from the general facts of the simulation. First, agents would have a very low chance of transmitting these personal facts in a given interaction, a factor which made them both more realistic as well as more valuable to dyads that shared them. Second, the number of personal facts was varied in the experiments while the number of general facts remained constant. While the number of personal facts per agent varied between zero and ten as can be seen in Table 2, the number of general facts remained at two thousand, as seen in Table 3. Third, personal facts had a very high interaction weight relative to that of general facts, meaning that egos were more likely to interact with alters with whom they had shared personal facts, and alters were more likely to contact egos about whom they had learned personal facts. Last, alters who learned personal facts about an ego could not relay such facts to third parties. While agents would always have a small probability of sharing these personal facts with any interaction partner, the rarity of such sharing meant that very few agents would learn any of an ego's personal facts by the end of the simulation.

In contrast to personal facts, which were rarely shared and would impact relatively few pairs of agents, general facts usually dominated the relative similarity and relative expertise calculations for most agent-agent pairs. While the number of general facts

Table 3 Parameters held constant

Parameter	Value
Number of time periods	2000
Max interactions per period	10 (5 initiations, 5 receptions)
Number of "general" simulation facts	2000
Transmission weight for one's own personal fact	1
Transmission weight for another agent's personal fact	0
Percent "general" simulation facts known per agent	1%
Percent transactive memory known per agent	50% (of 1%)

was kept constant during each simulation,[3] a parameter sweep was performed over the transmission weight on general facts. This had the effect of changing how often agents were to communicate general facts with each other versus their likelihood of communicating one's own personal facts. Changes to the transmission weight served as changes in the propensity to disclose personal information: lower transmission weights for general facts meant that personal facts were more likely to be shared, while higher transmission weights for general facts implied that individuals were more reluctant to share information about a highly salient dimension. The transmission weight values that we used, specified in Table 2, were weights per fact. Thus, the chance of sending any general fact was proportional to the number of general facts that the agent knew (the transmission weight per one's own personal fact, also a weight per fact, was a constant value 1 as seen in Table 3, while the transmission weight for relaying the personal fact of another agent was set to 0). As agents learned general facts via interaction with others, the chance of sending one of their personal facts decreased since the number of personal facts unique to each agent was established at the start of the simulation and did not change once the simulation was underway. Note that the transmission weight only affected the composition of messages communicated between agents and did not directly affect how agents calculated relative similarity or expertise (Hirshman and Carley 2007b). However, increasing the transmission weight of the general facts made it less likely that agents would share their personal facts—both with those that they interact with frequently, and with those that they do not—and thus could affect later similarity and expertise calculations since fewer agents would know those facts in later time periods.

In addition to the parameters varied during the simulation, several factors were held constant to facilitate analysis. These parameters are described in Table 3. A total of two thousand time periods were simulated per experiment replication. This was

[3] A parameter sweep—not reported here—suggested that changing the number of facts did not greatly influence the qualitative pattern of results observed using the techniques described in Sect. 6. Increasing the number of knowledge facts prevented analysis of larger societies. While transactive memory is stored in a sparse format, each Construct agent still must have a transactive memory of all other potential alters. Doubling the number of agents led to a quadrupling of the memory requirement for simulation, assuming that agents keep track of a roughly equal amount of each alter's knowledge. The number of general knowledge facts used in this simulation (2000 facts) was about the largest possible number of bits that could be used in the largest population size (1250 agents) without leading to to out-of-memory errors with Construct.

found to be sufficient for agents to interact with others, exchange information, and develop stronger preferences for some agents over others. Each agent was allowed to initiate up to five interactions (Hirshman and Carley 2007a) and receive up to five interactions from others (Hirshman and Carley 2007a) during each simulated time period. This meant that an agent could potentially interact up to twenty thousand times in the course of the simulation, though the total number of interaction partners was found to be less than three thousand for the entire simulation, once self-interactions and repeated interactions between agent-pairs per time period were dropped. By allowing an agent to interact multiple times per time period, it was possible to re-use the expensive-to-calculate similarity scores between agents in order to better simulate human activity. For all experiments performed, the number of general facts was held constant at two thousand, as this was a large enough value to ensure that agents did not learn all the general facts (reach quiescence). At simulation start, each agent had a 1% chance of knowing each of the general facts, meaning that each agent knew a mean of twenty random facts. Furthermore, each agent had a 50% chance of knowing whether or not every other alter agent had knowledge a particular fact, meaning that agents had about ten transactive memory facts per alter at the start of the simulation. A parameter sweep, whose results are not described in this paper, was performed to examine the effects of some of these parameters; moderate changes to these parameters did not greatly change the observed results.

The full factorial design of ten thousand replications, as described in Table 2, was performed on a heterogeneous cluster of computers, containing eighty processors distributed among fifteen machines. About half the cluster consisted of 8-processor, 64-bit, 2.6-GHz machines with 64 GB of RAM; the other half consisted primarily of 32-bit workstations with about 2 GHz processors and less than 2-GB of memory. Each simulation took between 3 hours and 96 hours to complete, depending on the machine type and the number of agents in the simulation. The entire suite of simulations took three weeks running in parallel to perform. The resulting interaction matrices required 16 GB of storage space.

5 Analysis

In analyzing the data, we had two primary goals: first, to understand whether any tiering behavior occurred, and second, to explore how changes to a naïve homophily model could affect such results.

In our analysis, we focused our search for four tiers, as inspired by the work of Zhou et al. (2005) described earlier. Due to the computational limitations discussed in Sect. 4, we were not able to thoroughly investigate the larger groups of the hierarchy—for instance, the groups with five hundred or more agents. Thus, our analysis focuses primarily on the four sections of Zhou's hierarchy closest to the ego: the tiers containing 3–5 agents, 15–20 agents, 30–50 agents, and 150 agents. These correspond to the "core discussion network" (Dunbar and Spoors 1995; Marsden 1987; Zhou et al. 2005), the "sympathy group" (Dunbar and Spoors 1995; Stiller and Dunbar 2007; Zhou et al. 2005), the "band" (Dunbar and Spoors 1995; Zhou et al. 2005), and the "clan" containing most general acquaintances (Dunbar

1993; Dunbar and Spoors 1995; Zhou et al. 2005) in Zhou's taxonomy. For simplicity, we will refer to these tiers as T1 (the tier that is closest to the ego and therefore interacts most frequently with it), T2, T3, and T4 (the tier that is most distant, and therefore interacts least frequently).

Our approach was to take the interaction counts for each agent-agent dyad as our baseline unit of analysis. Our Construct simulation output was an agent-by-agent matrix of interaction counts; we removed any self-interactions in order to better understand how frequently agents interacted with others. Because we cared about how frequently agents communicated and were not primarily concerned with which agent initiated the interaction, we marked a one for both agents if either agent contacted the other. We analyzed the data by simulation time period, so agents who interacted with each other multiple times during one time period were marked as having interacted only once. Thus, our variable of interest was a count that varied from zero (if a pair agents never interacted) to two thousand (if a pair agents interacted during each simulated time period).

As agents could have different numbers of total interactions but similar trends in behavior, it was necessary to normalize the raw interaction counts before further examining the data. We did so by taking the network of interaction counts network and treating it as a collection of vectors, one vector per agent. Each element in each row vector corresponded to an alter who interacted with a particular ego agent. Thus, we normalized each vector relative to its largest component so all values in the vector were in the range [0, 1]. We expressed these values as the percent of interaction that ego had with each alter relative to the time spent with the ego's most frequent interaction partner.

Once this normalization had been performed, we compared agents. We reordered the vectors so that agents ranked their normalized interaction frequency from largest to smallest. This of course destroyed the meaning of the indices in each vector, but as we were most concerned with the distribution of interaction frequencies for each agent, this transformation helped us better address our research question. This sorting procedure gave a slightly revised meaning to each index: that of *agent position* in terms of interaction frequency. This allowed us to rank alters by the time spent with the ego; for the agent at position zero, this value would have a normalized value of 1 since that agent had the maximum number of interaction, while the agent at a high position, the value would be close to 0. We could then compare different agents in the same simulation, as well as compare agents across different simulations by looking at trends in these normalized and resorted vectors of agent position.

After some initial investigation, it became clear that our data was taking on a long-tailed distribution as seen in Fig. 3, which plots agent position versus mean percent frequency of interaction. While Fig. 3 illustrates the case with five hundred agents and a general fact transmission weight of one, a similar drop-off was observed for other experimental conditions. Results confirmed that agents tended to communicate with a handful of alters many times more frequently than they communicated with the majority of others, a result that is consistent with literature values (Wellman 1996; Zhou et al. 2005). While this was in accordance with the general power-law behavior found by Zhou et al. (2005), the drop-off was much too steep to show conclusive evidence of tiering behavior. Thus, to better analyze the resulting frequencies of interaction, we decided to take the $\log_2$ transform of the original interaction

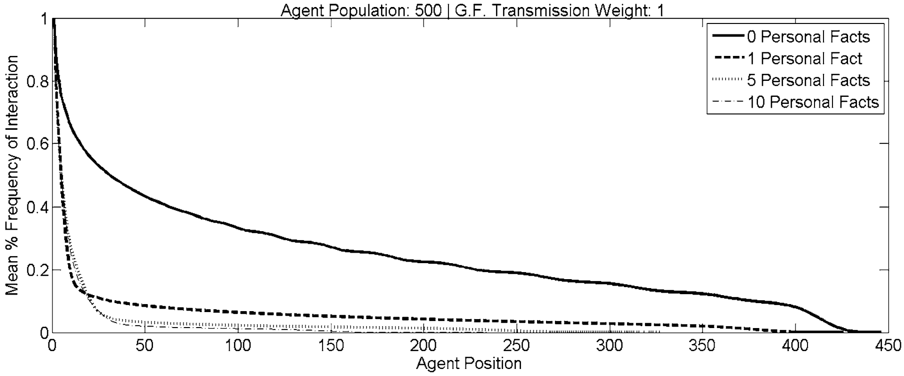

Fig. 3 Distribution of interaction frequency

count matrix prior to normalization. This transformation provided further differentiation between the tiers: alters in successively higher tiers would have significantly more frequent interaction with the ego, and the increase in interaction would be non-linear.

Also in this preliminary analysis, we noticed that a nontrivial proportion of all interactions were interactions which occurred only once or twice out of two thousand possible time periods. Such interactions contributed to an extremely long tail. Exploring this trend, we became concerned about what a single interaction meant—a measurement question that has been extensively discussed in the sociological literature but not often tackled in the modeling community (Marsden 1990; Wasserman and Faust 1994). Clearly, hundreds of interactions suggest an elevated relationship. However, since one agent's probability of interactions with another will be non-zero if it has transactive memory about another agent (see Fig. 2), some interactions may be due more to chance than to legitimate manifestations of substantial homophily. Upon further consideration, we decided that interaction frequencies that fell beneath a threshold of three interactions per two thousand time periods were not significant enough to merit consideration in the analysis that we were performing. One interaction could be due to chance, a second interaction coincidence, but three or more contacts between ego and alter should indicate an elevated relationship. For this reason, we then eliminated any agent-agent pairs with less than three interactions also prior to normalization.

As noted previously, we were searching for the presence of four tiers. We thus divided the normalized log frequency of interaction into four groups, corresponding to 1.00–0.75 (T1), 0.75–0.50 (T2), 0.5–0.25 (T3), and 0.25 to 0.00 (T4). For each agent in each simulation, we then counted the number of interaction partners within each range to calculate the number of agents in each tier. We then computed the average size of T1, T2, T3, and T4 by averaging each tier over the total number of agents in each experimental condition. Figure 4 provides a representative example of one of these experimental conditions—the 500 agent, general fact transmission weight 1 condition—after the truncation, transformation, and reordering have been performed.

Fig. 4 Frequency of interaction by agent position

It is worthwhile to mention variance measurements that were taken in both the preliminary and final simulation results. Within each simulation instance, the variance on agent position was relatively low. After the interaction partner counts for each agent were truncated, transformed, and reordered, the standard deviation around the mean log frequency of interaction (the coefficient of variance) was less than 5%. This suggested that there was not substantial variation between the ratio of interactions between ego's most common interaction partner and nth most common partner across multiple agents in the same simulation—though noting that the nth most frequently contacted agent would very likely differ when the ego changed. However, when tiers were created from the agent position vectors, the variance in the size of the tiers was consistently found to be about 20–40% of the mean for each tier size (coefficient of variance was 0.2–0.4). Variation of a similar magnitude was found when results for multiple agents was collapsed to get an overall measure of variation per simulation replication, and when multiple replications were collapsed to achieve an overall result for the experimental condition. Such ranges in tier size were not unexpected—even Dunbar's number and Zhou's tiers have very wide confidence intervals (Dunbar 1993; Zhou et al. 2005). For purposes of presentation clarity, however, we have omitted these variances from our tables and figures.

6 Results

We calculate the number of agents in each tier by following the procedure outlined in Sect. 5. The results, presented in Table 4, have been averaged over all ten replications performed for each experimental condition. Several trends are observable from this data.

First, the models of interaction that are based on a simplistic version of homophily—i.e., those models that include just general facts without personal facts—do not fit the Zhou distribution. These models, found in the left-hand block of Table 4, tend to have very large first, second, and (generally) third tiers relative to those Zhou found for human social networks. The sharp decrease in the fourth tier was partially an artifact of the analysis technique, since agents must have at least three interactions with the ego in order to be placed in that tier. This effect was also partially due to a

Table 4 Counts of agents in each tier of % frequency of interaction (log transformed)

Pop 250	# of Personal Facts															
	0				1				5				10			
Gen. Trans. Weight	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
0	117	128	4	0	58	30	11	34	32	19	11	25	23	15	11	23
1	49	127	55	16	6	43	150	41	8	10	81	113	8	9	30	92
5	49	127	55	16	6	106	109	25	5	28	148	55	6	9	116	89
10	49	127	55	16	13	123	89	21	4	51	142	42	5	19	137	64
50	48	128	55	17	38	128	62	19	13	111	91	26	4	74	114	41
Pop 500	# of Personal Facts															
	0				1				5				10			
Gen. Trans. Weight	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
0	97	319	66	15	74	42	16	38	40	22	12	28	29	17	12	26
1	52	209	140	45	7	38	268	118	10	13	121	183	10	14	46	165
5	53	208	140	48	4	120	244	70	6	26	251	126	7	11	186	157
10	52	209	141	46	10	165	209	58	4	51	260	105	5	21	228	131
50	52	209	141	43	38	201	155	52	9	148	206	64	3	90	226	84
Pop 750	# of Personal Facts															
	0				1				5				10			
Gen. Trans. Weight	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
0	83	354	206	36	84	54	20	48	46	25	15	38	34	19	14	35
1	52	262	197	62	7	35	334	161	11	16	149	228	10	20	59	183
5	52	262	197	65	4	120	351	95	6	26	310	173	8	12	230	197
10	52	263	196	61	8	172	310	75	4	48	333	153	6	21	281	176
50	52	264	196	63	36	218	247	64	7	153	301	80	3	91	313	95
Pop 1000	# of Personal Facts															
	0				1				5				10			
Gen. Trans. Weight	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
0	70	450	198	46	91	64	23	55	51	29	16	38	37	23	15	36
1	49	258	265	67	7	33	383	189	11	18	166	248	11	23	71	195
5	49	258	265	71	3	121	411	94	6	25	343	203	9	12	252	236
10	49	259	264	73	7	188	355	89	4	45	412	127	6	21	325	184
50	49	258	266	70	33	258	270	74	7	167	343	96	3	91	367	99
Pop 1250	# of Personal Facts															
	0				1				5				10			
Gen. Trans. Weight	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
0	70	403	260	53	96	74	27	58	54	34	18	40	40	26	17	43
1	47	239	314	69	7	30	445	155	12	19	173	254	11	27	79	201
5	46	239	314	77	3	103	457	96	6	24	416	153	9	13	260	237
10	47	240	312	65	7	161	409	82	4	44	446	127	6	21	381	152
50	47	239	313	66	32	234	322	75	6	145	392	89	3	105	379	98

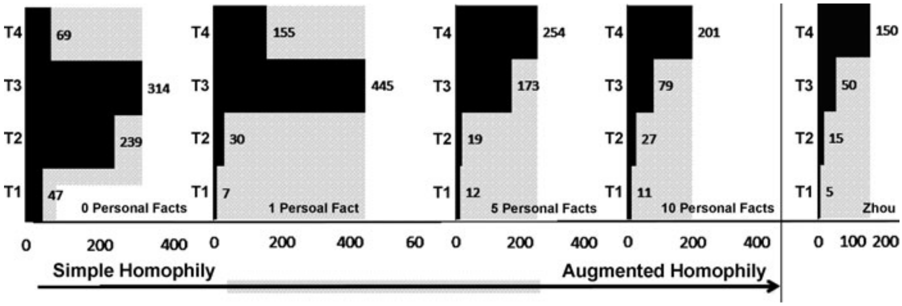

Fig. 5 Number of interaction partners by tier

saturation effect since such a large percentage of the society occupied the first three tiers that there were few agents left to fill the remaining tier. Regardless of the cause of the smaller last tier, it is clear that the simplistic homophily model does not lead to the expected interaction pattern.

This suggested that some sort of correction to the simplistic mechanism would be needed in order to generate realistic tiering behavior using homophily. Our correction, the inclusion of highly-valued personal facts, helped to generate realistic results. But how many of such facts are necessary? When there are too few personal facts, the model fit is generally poor as seen in the leftmost and left-center blocks of Table 4 while the fits of the rightmost blocks are much better. However, the transmission of personal facts, if included, becomes important. If personal facts are used, and the general facts are given a low transmission weight and thus not transmitted, unrealistic results are obtained. On the first row of every block of tables are the results occurring when the general knowledge transmission weight is zero, or what happens when agents only are able to send personal facts when communicating. In this setting, agents only (infrequently) share personal facts when communicating and otherwise do not share information. In this situation, a more or less uniform pattern is achieved across the tiers, a result which also does not match the literature. Increasing the general fact transmission weight is necessary.

The fact that extreme cases lead to implausible results suggests that personal facts and a non-trivial transmission weights are needed for homophily to lead to accurate tiering behavior. When a moderate number of personal facts and a moderate general fact weight are used, the ensuing pattern appears much more consistent with the literature values. The shaded cases in Table 4 represent distributions which have the same general shape as observed in the real data, though the numbers of agents in each tier do not always correspond to Zhou et al. (2005) results. The majority of these cases contain a first tier that is much smaller than the second, a second which is smaller than the third, and a third that is in turn smaller than the fourth. Such distributions occur under the same simulation conditions regardless of the simulation sample size.

The tiered nature of the data in Table 4 can be clearer if illustrated graphically. For instance, Fig. 5 displays the number of interaction partners in each tier when there are 1250 agents and a general fact transmission weight of one. This is the case occurring in the fourth-from-bottom row of Table 4. A simple homophily model using only the general facts leads to a normal distribution of tier sizes while gradually increasing

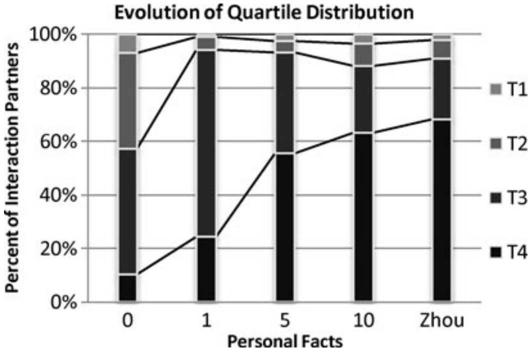

Fig. 6 Percent of interaction partners by tier

the number of personal facts generates the more expected skew in the distribution. With ten personal facts, the number of communication partners in the first tier is close to a third of the number in the second tier, which in turn is about a third of the results in the third tier, which in turn is about a third of the number in the fourth tier—results which have the exponential coefficient remarkably similar to that of Zhou et al. (2005). While the raw counts in each tier are slightly inflated from the values found by most anthropologists and sociologists, the Zhou results are not outside standard confidence intervals for some of our simulation conditions. More importantly, though, the results of Table 4 indicates that such a pattern of interactions was consistently found regardless of simulation size, though the size of the tier increased slightly with greater numbers of agents.

The relative number of agents in each tier can be seen most clearly in Fig. 6, which plots the raw values of Fig. 5 as percentages of the total number of interaction partners in all four tiers. The distribution of partners imputed from the Zhou data is plotted on the right-hand side for comparison. The simplistic homophily model fits the data poorly, as the first, second, and third tiers are slightly larger and the fourth tier dramatically smaller than the human data. Including even a few personal facts will lead to a substantial improvement, though the fourth tier remains small relative to what would be expected for human social networks. Increasing the number of personal facts further will modify the simulation by increasing both the frequency of personal fact transmission as well as increase the maximum relative homophily of agents sharing multiple personal facts. Such changes lead to more frequent interaction with an agent's most common interaction partners and ultimately generate the more skewed (and realistic) interaction frequencies.

7 Discussion

Our findings suggest that homophily can be used to make simulated agents follow patterns of interaction that are analogous to those observed for humans. While there is no firm agreement on what constitutes a sociological tier, a nomenclature has been developed to separate small numbers of close friends from large numbers of relatively distant acquaintances (e.g. Dunbar and Spoors 1995; Zhou et al. 2005). We find that when we augment the homophily mechanism with highly valued personal facts, we are better able to generate tiering behavior than when a simple homophily model is used in which all facts are equal.

While the most realistic results of the simulation are perhaps the most useful, it is of interest to examine the cases that did not produce results that follow the social and anthropological literature for general human societies. For instance, when no general facts were present (the leftmost block of results in Table 4), the results were clearly skewed towards the middle tiers. This suggests that at least some personal facts are necessary in order to achieve the results observed in human behavior. When only the two thousand general facts were present, agents would have insufficient reason to prefer one agent to another. Agents tended to interact with others frequently, and would not build up close relationships with favorite agents. Though systematic preferences did arise in these simulations, the degree of preference was insufficient to mimic real-world models. Such results provide strong support for some critiques of simple homophily or social categorization theories: homophily is an idea that works, except when it does not (Levine and Moreland 1998; McPherson et al. 2001; Turner et al. 1994). By including salient dimensions (McPherson et al. 2001) on which agents should be homopholous in order to interact—in our approach, the personal facts—we both strengthen and address this criticism. When we include such personal facts, we see that they contribute greatly to homophily between a few agent-agent pairs and so contribute to realistic social structures. Those seeking to use homophily to drive interaction in similar simulations might consider such results when building their own models. Those seeking to use homophily as an explanatory mechanism for phenomena in the real world may wish to consider the importance of dimension salience and to acknowledge that some forms of homophily may be more important than others.

When the transmission weight for general facts was very large or very small, the distribution of interactions was also different from that observed in the sociological data. When the general fact transmission weight was zero, agents were bound by the expertise and similarity knowledge present at simulation initialization. When interacting with other agents, egos would always attempt to share personal facts with every alter, a behavior which greatly increased the likelihood of continued contact between ego and alter. Without the exchange of general facts, exploration and the search for agents with similar general facts did not occur. As can be seen in Table 4, this resulted in a smaller number of overall interaction partners than that found in human societies, as well as a distribution which contained first and second tiers that were much larger than expected. Nevertheless, such interaction patterns are not entirely unrealistic and may be present in some venues, such as among students arriving at college (e.g. Hays and Oxley 1986; Newcomb 1961). However, such relationships are likely not as stable as our simulation would suggest; inevitably, as other processes affect the individuals and change what is salient, new interaction patterns begin to emerge and the initial friendship patterns may dissolve.

On the other hand, an excessive emphasis on general fact transmission also generated skewed results. If agents were insufficiently willing to share personal facts, they would share these facts with very few others. Such a behavior would lead to the formation of only a small number of close relationships. Such close ties with personal information would also be heavily-used relative to the vast majority of other ties. If these ties become excessively strong, a scenario which occurred to some extent in Granovetter's Boston neighborhoods, information may diffuse slowly and small clusters of agents will be isolated from each other (Granovetter 1973; Marsden 1987). Such societies may be vulnerable to changes in conditions or health of the members and may be more likely to fall apart.

While the most realistic results presented in Table 4—the shaded values—follow a pattern that may be classified as an exponential distribution, we are reluctant to fit an exact model to the data. First, we are not fully convinced that all real-world networks share the binning properties that Zhou et al. (2005) summarized. Secondly, given the fact that Dunbar's number (Dunbar 1993) has a very wide confidence interval and that other ranges on Zhou's scale can vary greatly, we feel that trying to fit a set of our values to Zhou's exact values would run the risk of overfitting the simulation model. More importantly, though, our primary goal in this paper has been to present our simulation technique an exploratory study of a mechanism and parameter space, and not as an optimally tuned model. While we find that certain parameters have important effects on the overall distribution of simulation results, our primary focus has been on the trends observed with the modified homophily mechanism. From the trends that we have observed, simulations with larger numbers of personal facts and relatively small general fact transmission weights are best for generating the types of distributions that Zhou observed.

Even without an exact fit to the Zhou data, the results that we present in Table 4 lead to an interesting and perhaps startling conclusion. We have posited a society in which agents are likely to interact with those who are similar to them. Under certain circumstances, such a mechanism is sufficient to generate tiering behavior. If all facts are treated equally, we tend to see that agents have no real preference for other interaction partners: their maximum frequencies of interaction are small and tiering behavior does not occur. However, if we allow agents to have increasing numbers of highly-valued personal facts, agents will become more selective in their choice of interaction partners and will begin to form structures that resemble those seen in human networks. If the transmission weight for general facts falls in a relatively broad band—the boundaries of which we explore but do not fully define—then agents will naturally assume a frequency of interaction which is found in the general anthropology, sociology, and groups literature. Using only transactive memory to guide their preferences, as well as a principle of seeking out others who are either similar or who are known to have interesting knowledge, agents settle into interaction patterns that resemble those observed in human populations.

We recognize there are a number of limitations to this work. First, we have deliberately fixed a number of simulation parameters, such as the number of facts, in order to explore other effects which we feel to be valuable and informative. This is good simulation practice, but may lead to questions regarding the roles that such parameters play in producing our results (Epstein and Axtell 1999). Though some of these factors were indirectly analyzed—for instance, we partially examined the effect of changing the number of facts by changing the transmission weight for the general facts—it was felt that such factors would be of secondary interest in understanding patterns of interaction. Second, we have deliberately omitted factors such as sociodemographic attributes (McPherson et al. 2001), geospatial location (Butts 2002), and affect (Lawler 1999), which have been known to play important roles in determining who interacts with whom and for how long. While these attributes are known to have an important impact on the development of human social networks, our research sought to demonstrate that minor tweaks to the homophily and transactive memory assumptions are strong enough to lead to the development of a plausible social network structure. Lastly, we acknowledge the effect of timescale. We recognize that human social networks are built up over the course of a lifetime and are the result of a number of complex social, psychological, and even environmental processes that cannot be easily modeled. Here, we have used an arbitrary representation of time, even as we acknowledge that past research has documented that social networks are highly dependent on time-dependent factors such as age and location (Hill and Dunbar 2003; Marsden 1987). Additional research may be needed to understand how homophily—and personal information—may affect the growth, maintenance, and decay of social networks and social tiers over time.

In our work, we wish to suggest that Construct is a plausible model for human interactions, not the definitive one. The homophily assumption in Construct, in conjunction with agent-specific yet shareable facts, represents a plausible way by which human networks may adopt a tiered structure. Other models may use slightly different core assumptions to explain human interaction; as previously mentioned, the empirical literature has suggested that geography, family, schooling, and work environment play important roles. Nevertheless, our results suggest it is possible to simulate the emergence of multi-tiered social ties using a homophily model and that these tiers do correspond with patterns observed in real-world human networks. Specifically, we believe our results demonstrate that if one wishes to use a homophily model to simulate realistic processes on human social networks, then one could include personal facts (or analogous attributes) in order to ensure that agents have both the proper number of interaction partners and correct frequency of interactions with them.

Thus, our findings suggest important considerations for other simulation designers, especially those seeking to model realistic patterns of human interaction. Simulations that use homophily to drive interaction can include highly salient or personal facts in their model to help ensure that interactions with alter agents follow literature-supported distributions of interaction frequencies. The power of these facts is both in their high salience and in their relatively low probability of being transmitted to a favored other. If this probability of transmission is either too low or too high results will be unrealistic. A homophily model that incorporates such personal facts, and tempers their transmission appropriately, will lead to results which will more closely match those observed in the real world. Such models can then be used more pragmatically to address questions of information diffusion, disease transmission, or a wide variety of other areas in which human agents must interact with realistic and meaningful frequencies.

References

Agneessens F, Waege H, Lievens J (2006) Diversity in social support by role relations: a typology. Soc Netw 28:427–41

Anderson JR (1983) The architecture of cognition. Harvard University Press, Cambridge

Anderson B, Butts C, Carley K (1999a) The interaction of size and density with graph-level indices. Soc Netw 21:239–67

Anderson C, Wasserman S, Crouch B (1999b) A P* Primer: logit models for social networks. Soc Netw 21:37–66
Axtell R, Axelrod R, Epstein J, Cohen M (1996) Aligning simulation models: a case study and results. Comput Math Organ Theory 1(2):123–142
Bandura A (2001) Social cognitive theory of mass communication. Media Psychol 3:265–99
Barabási A-L, Reka A (1999) Emergence of scaling in random networks. Science 286:509–12
Barnlund D, Harland C (1963) Propinquity and prestige as determinants of communication networks. Sociometry 26:467–79
Borgatti S, Foster P (2003) The network paradigm in organizational research: a review and typology. J Manag 29:991–1013
Borgatti S, Carley K, Krackhardt D (2006) On the robustness of centrality measures under conditions of imperfect data. Soc Netw 28:124–36
Burt R (1992) The social structure of competition. In: Structural holes. Harvard University Press, Cambridge, pp 57–89
Butts C (2002) Spatial models of large-scale interpersonal networks. Carnegie Mellon University, Pittsburgh
Carley K (1986) An approach for relating social structure to cognitive structure. J Math Sociol 12:137–89
Carley K (1991) A theory of group stability. Am Soc Rev 56:331–54
Carley K (1995) Computational organization theory. Comput Math Organ Theory 1:39–56
Carley K (1999) On the evolution of social and organizational networks. In: Special issue of research in the sociology of organizations on networks in and around organizations, pp 3–30
Carley K (2003) Dynamic network analysis. In: Dynamic social network modeling and analysis: workshop summary and papers, Washington, DC, pp 133–145
Carley K, Martin M, Hirshman B (2009) The etiology of social change. Top Cogn Sci 1(4):621–650
Christakis N, Fowler J (2007) The spread of obesity in a large social network over 32 years. N Engl J Med 357:370–9
Cowan R, Jonard N (2004) Network structure and the diffusion of knowledge. J Econ Dyn Control 28:1557–75
Dawkins R (1976) The selfish gene. Oxford University Press, London
Dunbar RIM (1993) Co-evolution of neocortex size, group size, and language in humans. Behav Brain Sci 16:681–735
Dunbar RIM (1998) The social brain hypothesis. Evol Anthropol 6:178–90
Dunbar RIM, Spoors M (1995) Social networks, support cliques, and kinship. Human Nat 6:273–90
Epstein J, Axtell R (1999) Growing artificial societies. MIT Press, Cambridge
Freeman L (1979) Centrality in social networks: conceptual clarification. Soc Netw 1:215–239
Friedkin N, Johnsen E (1999) Social influence networks and opinion change. Adv Group Process 16:1–29
Goldstein J (1999) Emergence as a construct: history and issues. Emergence 1:49–72
Granovetter MS (1973) The strength of weak ties. Am J Sociol 78:1360–1380
Harrison D, Price K, Bell M (1998) Beyond relational demography: time and the effects of surface- and deep-level diversity on group cohesion. Acad Manag J 41:96–107
Hays R, Oxley D (1986) Social network development and function during a life transition. J Pers Soc Psychol 50:305–13
Hill RA, Dunbar RIM (2003) Social network size in human. Human Nat 14:53–72
Hirshman B, Carley K (2007a) Specifying agents in construct. Carnegie Mellon University, School of Computer Science, Pittsburgh
Hirshman B, Carley K (2007b) Specifying networks in construct. Carnegie Mellon University, School of Computer Science, Pittsburgh
Hirshman B, Carley K (2008) Modeling information access in construct. Carnegie Mellon University, School of Computer Science, Pittsburgh
Hirshman B, St. Charles J (2009) Simulating emergent multi-tiered social ties. In: Proceedings of the 2009 human behavior and computational intelligence modeling conference, Oak Ridge National Laboratory
Hirshman B, Martin M, Bigrigg M, Carley K (2008a) The impact of educational interventions by socio-demographic attribute. Carnegie Mellon University, School of Computer Science, Pittsburgh
Hirshman B, Martin M, Birukou A, Bigrigg M, Carley K (2008b) The impact of educational interventions on real & stylized cities. Carnegie Mellon University, School of Computer Science, Pittsburgh
Ilgen D, Hulin C (2000) Computational modeling of behavior in organizations: the third scientific discipline. American Psychological Association, Washington

Kilduff M, Krackhardt D (1994) Bringing the individual back in: a structural analysis of the internal market for reputation in organizations. Acad Manag J 27:87–108
Krackhardt D, Carley K (1998) A PCANS model of structure in organization. In: Proceedings of the 1998 international symposium on command and control research and technology, Monterey, CA
Laird J, Congdon CB (2006) The soar user's manual, version 8.6.3. University of Michigan
Lawler E (1999) Bringing emotions into social exchange theory. Annu Rev Sociol 25:217–44
Lazarsfeld P, Merton R (1954) Friendship as social process: a substantive and methodological analysis. In: Berger M, Abel T, Page C (eds) Freedom and control in modern society. Van Nostrand, Princeton
Leskovec J, Horvitz E (2008) Planetary-scale view on a large instant-messaging network. In: Proceedings of the world wide web 2008, Beijing, China
Leskovec J, Kleinberg J, Faloutsos C (2005) Graphs over time: densification laws, shrinking diameters, and possible explanations. In: Proceedings of the 2005 conference on knowledge and data discovery, Chicago, IL
Leskovec J, Lang K, Dasgupta A, Mahoney M (2008) Statistical properties of community structure in large social and informational networks. In: Proceedings of the 17th international conference on the world wide web conference, Beijing, China
Levine J, Moreland R (1998) Small groups. In: Gilbert SFD, Lindzey G (eds) The handbook of social psychology. Oxford University Press, London
Lopez L, Sanjuan M (2002) Relation between structure and size in social networks. Phys Rev E 65:036107
Marsden P (1987) Core discussion networks of Americans. Am Sociol Rev 52:122–31
Marsden P (1990) Network data and measurement. Annu Rev Sociol 16:435–63
McPherson M, Smith-Lovin L, Cook J (2001) Birds of a feather: homophily in social networks. Annu Rev Sociol 27:415–44
Milgram S (1967) The small world problem. Psychol Today 2:60–7
Mniszewski S, Del Valle S, Stroud P, Riese J, Sydoriak S (2008) EpiSimS simulation of a multi-component strategy for pandemic influenza. In: Proceedings of the 2008 spring simulation multiconference, Ottawa, Canada. ACM, New York
Newcomb T (1961) The acquaintance process. Holt, Reinhart, and Winston, New York
Newell A (1994) Unified theories of cognition. Harvard University Press, Cambridge
Newman M, Park J (2003) Why social networks are different from other types of networks. Phys Rev E 68:036122
Olfati-Saber R (2006) Flocking for multi-agent dynamic systems: algorithms and theory. IEEE Trans Autom Control 51:401–20
Rogers E (1995) Diffusion of innovation. Free Press, New York
Steglich C, Snijders T, West P (2006) Applying Sienna: an illustrative analysis of the co-evolution of adolescents' friendship networks, taste in music, and alcohol consumption. Methodology 2:48–56
Schreiber C, Siddhartha S, Kathleen C (2004) Construct—a multi-agent network model for the co-evolution of agents and socio-cultural environments. Technical Report ID CMU-ISRI-04-109. Carnegie Mellon University School of Computer Science, Pittsburgh PA
Stiller J, Dunbar RIM (2007) Perspective taking and memory capacity predict social network size. Soc Netw 29:93–104
Thiriot S, Kant J-D (2008) Generate country-scale networks of interaction from scattered statistics. In: Proceedings of the fifth conference of the european social simulation association, Brescia, Italy
Turner J, Oakes P, Haslam SA, McGarty C (1994) Self and collective: cognition and social context. Pers Soc Psychol Bull 20:454–63
Valente T (1995) Network models of the diffusion of innovations. Hampton Press, Cresskill
Wasserman S, Faust K (1994) Social network analysis. Cambridge University Press, Cambridge
Wellman B (1996) Are personal communities local: a dumptarian reconsideration. Soc Netw 18:347–54
Wellman B, Wortley S (1985) Different strokes from different folks: community ties and social support. Health Educ Behav 12:5–22
Wong LH, Pattison P, Robins G (2006) A spatial model for social networks. Physica A 360:99–120
Zhou WX, Sornette D, Hill RA, Dunbar RIM (2005) Discrete hierarchical organization of social group sizes. Proc R Soc Lond B, Biol Sci 272:439–44

Brian R. Hirshman is a fourth year graduate student in the Computation, Organization, and Society program at Carnegie Mellon University. His interests include social network modeling and cognition, and will shortly be proposing a thesis to investigate how improved cognitive modeling will affect our understanding of information and belief diffusion in social networks.

Jesse St. Charles is a second year graduate student in the Computation, Organization, and Society program at Carnegie Mellon University.

Kathleen M. Carley is a Professor in the Institute for Software Research in the School of Computer Science at Carnegie Mellon University and the director of the center for Computational Analysis of Social and Organizational Systems (CASOS) (http://www.casos.cs.cmu.edu/). Dr. Carley did her undergraduate work at the Massachusetts Institute of Technology and received her Ph.D. from Harvard in Mathematical Sociology. She has published over 200 articles and co-authored 3 books related to dynamic network analysis, social networks, and multi-agent modeling of complex socio-technical systems. Her research combines cognitive science, dynamic social network science, social/organization science and computer science to address complex socio-technical issues from an interdisciplinary perspective. She has developed and deployed multiple technologies related to simulation and networks. These include: ORA, a statistical toolkit for analyzing and visualizing complex multilevel socio-cultural networks with geo-spatial and longitudinal aspects; AutoMap, a text-mining system for extracting network information and beliefs from unstructured texts; and Construct, a large scale multi-agent dynamic-network simulation for assessing information diffusion and belief change in a cyber environment.