Decision Support Systems 55 (2013) 969–977

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

Distance matters: Exploring proximity and homophily in virtual world networks

Yun Huang [a,*], Cuihua Shen [b], Noshir S. Contractor [a]

[a] Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL, USA
[b] Emerging Media & Communication Program, School of Arts & Humanities, University of Texas at Dallas, Richardson, TX, USA

ARTICLE INFO

Available online 9 January 2013

Keywords:
Proximity
Homophily
Distance
Virtual worlds
Online games
Virtual collaboration

ABSTRACT

Distance has long been a powerful force that influences the ways in which we organize our personal relationships and collaborations. In the past two decades, globalization and the advent of Information and Communication Technologies (ICTs) have brought profound changes to the formation and maintenance of communication networks in contemporary work and social settings by eliminating the constraints of physical distance on human interaction. To study the role of distance in ICT-enabled virtual worlds, we propose three dimensions of proximity—space, time, and homophily—and analyze the impact of distance, time zones, gender, age, and game age on building collaborative relations in online games. The results show that spatial proximity, temporal proximity, and homophily in age and game age still have a strong impact on players' behavior in creating online relations in virtual worlds, however, there is no evidence of gender homophily. This study suggests that offline proximity plays an important role in bringing people together in virtual worlds.

1. Introduction

Distance has long been a powerful force that influences the ways in which we organize our personal relationships and collaborations. Yet in the past century, the limiting effect of distance has seemed to gradually fade as the early advances of transportation and telecommunication technologies such as railroads, cars, planes, and the telephone, have facilitated connections among spatially dispersed individuals [28]. In the past two decades, globalization and the advent of Information and Communication Technologies (ICTs), especially the Internet and mobile phones, have brought profound changes to the formation of communication networks in contemporary work and social settings. As ICTs largely eliminate the constraints of physical distance on human interaction, distributed work has become a common form of collaboration within and beyond organizational boundaries [14,18,33]. Meanwhile, in the social sphere, computer-supported social networks that tie friends and relatives close and afar have gradually replaced traditional place-bound neighborhood communities [41,42]. These notable societal trends have led scholars and observers alike to claim "the death of the distance" [7] and "the world is flat" [13].

Despite the development of advanced information and communication technologies that have dramatically increased our capability of interacting and collaborating with people across greater distances, many argue that distance "…would persist as an important element in human experience" [34] because of the unique characteristics of face-to-face interactions that ICTs are unable to emulate. The majority of these studies focused primarily on existing ties of a collaborative nature and within established organizational settings [34], but few studies examined the more free-form connections that emerged during the collaboration on virtual tasks and immersive digital worlds enabled by technology [46]. In today's digitally-networked environments, to what extent does distance still shape the structure of interaction?

Instead of focusing on existing ties in the physical world, this study investigates the role of proximity in establishing online collaborative relations in virtual worlds created by ICTs. The emergence of virtual worlds such as Second Life and online games as significant, new social phenomena has implications for the way we work and play. Among the few studies that have examined proximity and digital ties, the focus has been on informational ties such as connections formed on the micro-blogging site Twitter [37]. How proximity would affect more virtual interactions such as online collaborative ties that enable people to more narrational relations in a virtual world remains relatively unclear [45]. This question becomes increasingly important as the scale and complexity of human collaboration in virtual worlds offer an excellent opportunity to study the emergence and effectiveness of groups in complex social settings. In addition, virtual worlds provide unprecedented data traces of individual interactions and behavior that were often unavailable in earlier scholarly endeavors. Hence virtual worlds are promising potential test beds—conceptually and empirically—for examining how people form online

* Corresponding author at: Department of Industrial Engineering and Management Sciences, Northwestern University, 2145 Sheridan Rd, TECH C210, Evanston, IL 60208, USA. Tel.: +1 847 467 2563.
E-mail addresses: yun@northwestern.edu (Y. Huang), cuihua.shen@utdallas.edu (C. Shen), nosh@northwestern.edu (N.S. Contractor).

0167-9236/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.dss.2013.01.006

relations. More importantly, online relations are indeed dynamic and complex: some stem from previous personal ties and the rest are solely based on activities in virtual worlds. This study examines the role of proximity in online collaboration.

The following section begins with a review of the multidimensional conceptualizations of proximity. Based on this review, we hypothesize the influence of three dimensions of proximity on the formation of collaboration relations in a virtual world: space, time, and homophily.

2. Theory and hypotheses: proximity and online relationships

The effects of proximity on social and collaborative dynamics have been researched for decades [2,12,18,21,23]. To date, research on proximity and network formation has produced mixed findings that could be partially attributed to at least two factors: one conceptual and one methodological. With a few exceptions most existing studies tend to conceptualize proximity in terms of geographic distance. Yet, there is a growing recognition that proximity is better understood as a multidimensional concept. Scholars using concepts such as "proximity," "virtuality," and "geographic dispersion" tend to offer different operationalizations across research contexts [31,33,34]. A recent study unpacked "virtuality" into four distinct components: geographic dispersion, electronic dependence, dynamic structure, and national diversity [14]. In their recent essay, O'Leary and Cummings [33] further suggested a multidimensional conceptualization of geographic dispersion, including spatial, temporal, and configurational characteristics. Similarly, Armstrong and Cole [3] posited that objective measures of distance include geographic distance, time distance (time zones), organizational distance, and cultural differences.

Although some or all of these dimensions are implicitly present in many empirical studies, spatial proximity remains the dominant factor in many studies [21,24,34]. Clearly, proximity is a multi-faceted concept that encompasses related but distinct dimensions. For example, two people located 1000 miles apart may also span different time zones and have different individual characteristics, all of which may contribute to their unique interaction patterns. In many cases, two or more dimensions of proximity co-exist and exert influence on network dynamics together, but their mechanisms and effects cannot be disentangled unless each of those dimensions is carefully considered and examined in conjunction with others.

From a methodological standpoint, with a few exceptions [28], proximity is often operationalized as a dichotomous variable: dyads or teams are either collocated or dispersed [34]. This operationalization can be problematic because it does not recognize that in real life situations, proximity should be measured in terms of degree. For example, a distributed team involving members from two adjacent cities is different from a transnational team with members located in two or more countries. More refined measures are essential to reveal how proximity influences peoples' interaction patterns.

The literature on propinquity found that both physical and psychological similarities between people lead to interpersonal attraction [12]. For players' physical proximity in online games, distance constraints have been illuminated by the virtual world but time differences are still a cost of coordinating with other players. In this study, we examine the effects of these three aspects on the formation of online relations: spatial proximity (physical distance), temporal proximity (time zone difference), and homophily (social distance based on individual characteristics). In Sections 2.1 to 2.3, drawing upon the research on offline relations, we present a series of hypotheses on the formation of online relations based on three exogenous factors: spatial proximity, temporal proximity, and homophily. While these hypotheses examine the associations between online relations and offline proximity and homophily, they do not address causal and temporal sequencing of the mutual influence of online and offline networks.

2.1. Spatial proximity

There is a long tradition of examining the impact of spatial proximity on communication. Since the 1930s, researchers studied how spatial proximity affects constructs such as friendship, romantic relationships, and variables like the amount of communication [12]. When the spatial distance between workers reached 30 m and beyond, their frequency of spontaneous communications dropped drastically [2].

With the advent of communication technology, scholars have investigated whether spatial proximity is still relevant with the surge in computer-mediated communication. This line of research generally looks at teams spread across geographic distances and how computers mediate this interaction [16,23,34]. Scholars have focused on how individuals that know each other offline communicate online. There is limited research examining how individuals that do not know each other offline communicate online. In general, this research shows that individuals who are located closer to each other are more likely to communicate than individuals who are located farther away from each other regardless of whether or not communication is face-to-face [9]. Even taking into account organizational proximity and similarity in interests, researchers have found that individuals located closer together were more likely to collaborate with each other both face-to-face and by phone and email [23]. One explanation is that computer-mediated communication is more useful in augmenting, rather than replacing, face-to-face communication between people located closer to each other than those located farther apart [8].

Past research has found that spatial proximity exerts a significant and positive impact on individuals' interactions in both face-to-face and computer-meditated contexts. Even as a rich immersive social environment and new virtual meeting tool, virtual worlds may still inherit the impact of spatial proximity that augments existing relations [8]. This study extends this research by examining the impact of spatial proximity in virtual worlds where individual players distributed across geographical distances engage in various activities such as communication, economic transaction, and group collaboration. Hence, we propose the following hypothesis:

Hypothesis 1. (Spatial proximity) Individuals who are proximate in geographical distance are more likely to interact with one another than those who are not proximate.

When questioning the effect of proximity, researchers examine to what extent the impact is linear. Most of the early studies looked at rather short distances such as the location of offices in the same building [2,31]. The mechanism is intuitive: if two people's offices are located on the same floor, they are more likely to strike a hallway conversation than those who have offices on different floors or even in different buildings. Verbrugge [39] finds that residential proximity is the single best predictor of how often friends get together to socialize, however, the same mechanism may not become as intuitive as distance grows larger. To what extent can we argue that people are more likely to interact when they live within a shorter distance? In that case, the cost of communication and commute becomes a more useful organizing principle than that of the pure chance encounter. The travel time and expenses greatly reduce the likelihood of interaction when the mode of transportation changes, e.g., from walking distance to driving distance. The changes in travel and communication costs are not linear, but tend to increase sharply when the distance stretches beyond a short range. Hence, if the distances between individuals are large enough, the impact will be similar because the costs would remain largely at the same level.

Hypothesis 2. Short distance Individuals who are in close proximity are substantially more likely to interact with one another than those who are at medium or low proximity.

2.2. Temporal proximity

The previous hypotheses describe the spatial dimension of proximity characterized by geographical separation. A related perspective focuses on how space becomes less of an organizing principle than the costs associated with getting between places or communicating between places [10]. For example, it is more "costly" for people without a car to travel to another city located 100 miles away than those who have access to a car. With widespread connectivity to networked computers and the prevalence of global teams, the communication cost perspective has led some scholars to shift the attention from spatial distance to temporal distance, referred to as temporal proximity in the current study, or namely, the overlap of waking hours [33]. Temporal proximity is measured in terms of time zone difference. It is more costly for people who are in different time zones and hence share fewer waking hours to interact synchronously, compared with those who share more waking hours. Therefore, temporal proximity also affects the likelihood of communication, especially synchronous communication.

Even though spatial proximity and temporal proximity have distinct influences on the networks, most empirical studies do not simultaneously consider the effects of these two forms of proximity. Taking both effects as a whole, O'Leary and Cummings [33] provided a typology of different combinations of spatial and temporal dispersion in teams. While some teams are only spatially (but not temporally) dispersed, most global teams are confronted with both distances. Therefore, it is important to consider both spatial proximity and temporal proximity together when studying their impacts.

Hypothesis 3. Temporal Individuals who are proximate in time zone are more likely to interact with one another than those who have larger time zone differences.

2.3. Homophily

Finally, individuals can have similar social characteristics such as gender and age otherwise known as homophily, which implies that individuals are proximate in the social space [27]. Homophily facilitates communication, which is illustrated by the old saying "birds of a feather flock together." The word "homophily" was first coined by Lazarsfeld and Merton [25] when referring to a tendency for people to be attracted to others who have similar attitudes, beliefs, and personal characteristics. Monge and Contractor [30] summarized two lines of theoretical underpinnings of homophily: the similarity-attraction hypothesis and the theory of self-categorization. The similarity-attraction hypothesis postulates that people are more likely to interact with those who have similar traits [6]. The self-categorization theory argues that people tend to self-categorize with regard to race, gender, socio-economic status, etc., and they differentiate between similar and dissimilar people based on such attributes [1,38].

Homophily has received strong support from empirical research, in terms of gender [20], race [29], and status [26]. Homophily, especially with regard to gender, ethnicity, and occupation, has been found as a critical factor of relationship formation in entrepreneurial teams [36], work team composition [19], and generally in the formation of social networks [27].

The theory of homophily posits that people of the same attributes tend to interact with each other because of their common background and shared interests. Although in virtual worlds players' demographic attributes such as gender and age may not be visible, other players may still be able to detect behavioral patterns related to these attributes. For example, male players are more likely to choose male characters and engage more in combat activities. Furthermore, through online interactions and social activities, players may exchange some personal information and become familiar with each other. Therefore, in virtual worlds, demographic attributes such as gender and age are expected to display similar effects as in other social contexts.

Hypothesis 4. Gender Individuals of the same gender are more likely to interact with one another than those of opposite genders.

Hypothesis 5. Age Individuals who have smaller age differences are more likely to interact with one another than those who have bigger differences.

We expect that players tend to interact with others who joined the virtual world at similar time periods. Individuals with similar game age, i.e. in the same cohort in a game, usually have similar levels of knowledge, status, tenure, and the same goals in the virtual world. Some virtual worlds provide a common starting place for new players to get familiar with the environment. New players can design their characters, learn basic functions in the world, and meet other fellow players. In addition, players joining at the same time period are presented with the same "version" of the world and the same player population, which may further enhance their perception of similarity.

Hypothesis 6. Game age Players who enter a virtual world during the same time period are more likely to interact with one another than those who have bigger differences in starting time.

To summarize, spatial proximity, temporal proximity, and homophily all demonstrate similar effects on creating and facilitating interactions between individuals. In general, when people are physically close to each other, located in similar time zones, and have similar social and demographic attributes, they are more likely to initiate interactions and have more frequent communication.

3. Methods

This study examines the impact of spatial proximity, temporal proximity, and homophily in a large virtual world—EverQuest II, a massively multiplayer online role-playing game (MMORPG) launched in November of 2004. Both EverQuest II and its predecessor EverQuest maintained a large share of the North American market until the launch of World of Warcraft (WoW), the current US MMOGs leader. Despite losing its market lead, the EverQuest franchise continues to expand and still attracts hundreds of thousands of players. As two of the most popular (and largest) of MMOGs, EverQuest and EverQuest II have inherited features from text-based MUD (Multi-User Dungeon) games and implemented an immersive 3D environment like some precedents such as Meridian 59 and The Realm Online. Sharing many similar designs and concepts, EverQuest II and WoW, two of the best games in the role playing world, have been used by many studies as representatives of MMOGs.

EverQuest II operates virtual game worlds based on individual game servers. For most activities, players are only allowed to interact with those who are on the same server. Transferring characters from one server to another is not encouraged and requires a charge so therefore a server can be considered as an independent virtual world with a stable population of players. This study examines player interactions on the Guk server. In the game, each player can create one or more characters (or avatars) and each character needs to develop combat skills to kill monsters. In order to advance levels faster, game characters can collaborate with others to fight monsters and finish game quests in the virtual world.

Through analyzing the data from EverQuest II, we identify players' online relations in the game based on their collaboration activities and construct precise measures of spatial proximity, temporal proximity, and homophily. We then examine the effects of proximity and homophily on virtual world networks through cross-sectional network analyses.

3.1. Data samples—partnership in EverQuest II

A detailed server log from May 1, 2006 to September 11, 2006 provided by Sony describes players' demographic information and collaborative activities that occurred in the Guk server: a typical player-versus-environment (PvE) server focusing on killing monsters, or non-player characters (NPC)s. In the world of EverQuest II players are allowed to form *groups* among themselves in order to play together and fight monsters more efficiently. Similar to friendship offline, two players can be considered as partners if they group together and earn experience points in combat collaboration activities. The partnership relation is especially important because this form of collaboration represents a stronger dyadic relation between two players than the interactions in bigger groups. In this study, *partnership* is defined as a binary and undirected tie between two players if they use their characters to achieve more than two outcomes together. A player could establish partnership relations with others through different game characters.

To test the overall impacts of proximity and homophily as well as the changes of the relations, data samples are extracted based on the log data to test the overall effect of proximity and homophily among all players in a given time period. Because the Guk server is designed to host players in North America, only players in the United States and Canada are included in the samples. From September 5, 2006 to September 11, 2006, 1525 unique players who partnered with others (1478 in the United States and 47 in Canada) are identified and their partnership network is illustrated in Fig. 1.

3.2. Measurements

Detailed activity records in EverQuest II provide information to build relation networks, represented by an adjacency matrix with 1525 nodes and 1179 edges (network density 0.5%). This partnership network is used as the dependent variable in ERG models.

Upon their first entry in EverQuest II, players need to report their demographic information such as gender, date of birth, and address zip code. The players' zip codes and country codes are used to estimate their offline location and to construct distance and time zone differences between players. The individual attributes such as gender, age, and game age (based on the dates of registration) are used to construct the homophily measures between players.

3.2.1. Distance

In order to calculate distance, the players' zip codes and country codes are first mapped to latitude/longitude using ZIPList5 and Canada Geocode databases from ZipInfo.com. The shortest distance between any two players is calculated using their latitude and longitude coordinates according to Spherical law of cosines (Eq. (1)):

$$d_{ij} = a\cos\left(\sin(lat_i) \times \sin\left(lat_j\right) + \cos(lat_i) \times \cos(lat_j) \times \cos\left(long_j - long_i\right)\right) \times 6371\,km \quad (1)$$

The marginal impact of distance on interactions between two locales declines significantly as the distance between them increases. For example, the impact of distance between 800 km away and 850 km away is very small and the difference is much bigger between 50 km and 100 km. Therefore, four categorical measures and one continuous measure are developed based on the raw distance to capture the non-linear impact of distance. The categorical measures include four binary matrices: *Same_zip_code*, *Short_distance*, *Medium_distance*, and *Long_distance*. If player i and j are located in the same zip code, $Same\_zip\_code_{ij}$ equals one. Based on the study on human travel by Brockmann et al. [4], we use $Short\_distance_{ij}$, $Medium\_distance_{ij}$, and $Long\_distance_{ij}$ to indicate whether the distance between two players

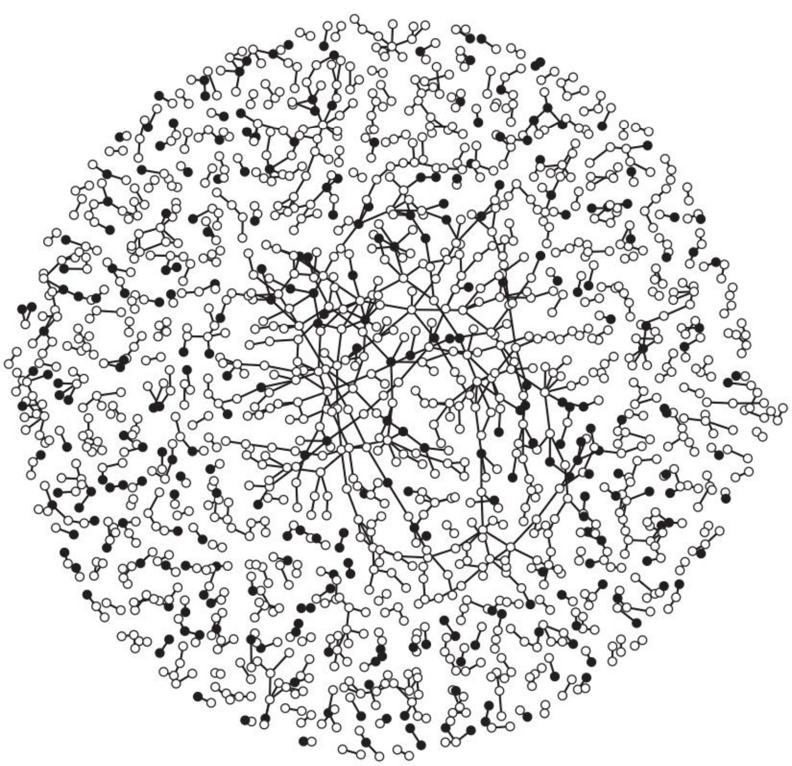


Fig. 1. Partnership among 1525 players in Guk in the sample week (white: male; black: female).

is smaller than 50 km (but not in the same zip code), between 50 and 800 km, or more than 800 km. Short distance represents a range reachable for daily activities and long distance represents places out of a geographical region. The geographic distribution of players in the sample is consistent with the U.S. Census data and most of them are located far away from each other: only 0.05% pairs of them are located in the same zip code, 0.6% are in short distance, 15% are in medium distance, and the rest are in long distance. However, players in partner relations tend to be much closer to each other: in 20.9% of partner relations two players live in the same zip code and in 9.7% of relations players live in a short distance.

Kleinberg [22] suggested that the tie probability should be a negative exponential function of distance. The continuous measure is constructed using an exponential decay function to measure the decreasing effect of distance (Eq. (2)). If two players are located in the same time zone, the impact of distance between them equals one, which indicates the maximal impact between the two; if the distance is d_0, the distance impact is reduced to 0.368. In this study, we use 50 km as d_0.

$$Distance_{impact}t_{ij} = \mathrm{EXP}\left(-d_{ij}/d_0\right). \quad (2)$$

3.2.2. Timezone_difference

Similarly the players' zip codes can be mapped into time zones. The time zone difference between two players is calculated as the hour difference between their time zones in absolute value.

Three individual attribute variables include *Gender*, *Age*, and *Game_age*. *Gender* is a binary variable indicating whether a player is female. *Age* is calculated at the time point September 11, 2006. *Game_age* is the number of months each individual played since they registered their first characters. Table 1 shows the descriptive statistics of individual locations and attributes in the sample set. 19% of the players are female with an average age of 33.63. This is consistent with the survey conducted in the same game [44].

Based on *Gender*, *Age*, and *Game_age* variables, the similarity of the attributes is measured in three matrices: *Same_gender*, indicating whether two players have the same gender, *Age_difference*, indicating the age difference between players, and *Game_Age_differernce*, indicating the game age difference between players.

3.3. Analyses

Many statistical models assume that relational ties among a given set of nodes are independent even though some relations are affected by others. For example, certain popular users have many ties and friends of friends tend to become friends themselves. Exponential Random Graph Models (p*/ERGM) [35,40] explicitly incorporate the dependence of the relations within a network by considering the observed network as one realization of a network generation process and estimate how likely the observed structure emerges out of all possible configurations generated in the process. Therefore, we can draw inferences of multiple factors that could be associated with the likelihood of a partnership tie existing between two players: attributes of players (e.g., gender, age and game age), joint attributes of two players (e.g., gender homophily), and relational attributes between two players (e.g., distance).

In our analysis, *Same_zip_code*, *Short_distance*, *Medium_distance*, *Long_distance*, and *Distance_impact* are used as measures of spatial proximity to test Hypotheses 1 and 2. *Timezone_difference* is used as a measure of temporal proximity to test Hypothesis 3. *Same_gender*, *Age_difference*, and *Game_Age_differernce* are used as homophily measures to test Hypotheses 4, 5, and 6. *Gender(Female)*, *Age*, and *Game_Age* are used as baseline control variables. The estimated coefficients can be explained similar to those in logistic regressions, positive coefficients indicate that the corresponding attributes or structures are more likely to occur than by random chance, and negatives indicate less likely. The effect size of one additional unit is measured by the odds ratio (OR), which equals to the exponential function of the corresponding coefficient, e.g. e^{β}.

While examining the impacts of the exogenous factors on the formation of online relations, ERG models control for the endogenous factors that enable and constrain the formation. Interactions between individuals can be explained endogenously by the presence and pattern of other ties within the same network. In a similar scenario, research on relation networks of offline friendship [5,11] indicates that endogenous structural tendencies such as sparsity, popularity, and transitivity influence the structure of virtual world networks. Based on the theory of self-interest, individuals establish ties if there are premiums associated and therefore they would not randomly create ties with others. Individual degrees in relational networks usually follow the power law distribution, i.e., a few individuals have many ties and become hubs while most individuals only have a few ties. Individuals tend to connect to others who already have many ties because of preferential attachment [32] and therefore the popular individuals will get even more ties. As a consequence of balance theory [17], researchers have found that two individuals who are both tied to a third person are also tied to each other and this is called a tendency toward transitivity. In ERG models, to control for sparsity, popularity, and transitivity, we included three network statistics: the number of edges (Edges) that indicate the network density; the geometrically weighted degree distribution (GWDegree) that summarize the degree distribution in a network; and geometrically weighted edgewise shared partners (GWESP) that measure the number of players connecting two other players in a network [20,33].

Four ERG models are estimated to reveal the marginal contributions of the explanatory variables. Model 0, as a baseline, includes six control variables; Model 1 estimates the impacts of temporal proximity as well as homophily; Model 2 combines *Distance_impact* and *Timezone_difference* to estimate the impact of spatial proximity and temporal proximity at the same time; Model 3 uses *Same_zip_code* and *Short_distance* (taking *Medium_distance* and *Long_distance* as the base category) to study the detailed impact of spatial proximity. Statnet [15] v2.2-3 with R-2.10.1 is used for estimation.

4. Results

Table 2 shows the results of the ERG models, which assess the relative importance of different ranges of offline distance and the impact of homophily and network structures in partnership networks.

Three structural control variables in Model 0 are all significant and reveal inherent structures in the partnership network. The edge parameter controls the density of a network. The significant and negative coefficients suggest that partnership relations in EverQuest II are sparse and individuals are not likely to engage in interaction randomly. As an anti-preferential attachment measure, the positive impact of geometrically weighted degree distribution (GWDegree) shows that popular players with many relations are less likely to engage in partnership with others. The degrees of the players, i.e., numbers of partners, are

Table 1
Descriptive statistics of player locations and attributes.

	Statistics (1525 players)			
	Mean	S.D.	Min.	Max.
Latitude	38.73	6.05	21.28	64.86
Longitude	−95.61	18.02	−158.0	−64.85
Time zone	−6.16	1.23	−10	−4
Gender (female = 1)	0.19	n.a.	0	1
Age	33.63	9.92	6	76
Game age	14.98	7.60	0	22

more evenly distributed compared to a random network with the same density and over effects modeled. The positive coefficients of geometrically weighted edgewise shared partners (GWESP) indicate that partnership in the game is transitive, which suggests that if two players have common partners they are more likely to become partners with each other.

Model 1 shows the impact of time difference: the odds ratio of becoming partners for two players within a one-hour difference is 61% of the odds ratio of players in the same time zone. Hypothesis 3 on temporal proximity is supported.

Model 2 examines the impact of time zone differences controlling for distance. Given the impact of distance, the negative and significant coefficient of time zone differences shows that time zone differences reduce the likelihood of interaction. In Model 2, the positive coefficient of distance impact supports the same conclusion. If two players are located closer offline, i.e., distance impact between them is larger, they have a greater chance to be partners online.

The proximity part of Model 3 shows positive and significant impacts of *Same_zip_code* and *Short_distance* taking *Medium_distance* and *Long_distance* as the baselines. This suggests that two players in the same zip code are more likely to form relations than two players within a short distance (less than 50 km but not in the same zip code). Similarly short distance relations are more likely to form than medium distance relations, which are more likely than long distance relations. For example, in Model 3, the odds ratio of forming partnership relations between players living in the same zip code is 721 times more compared to the odds ratios of players living more than 50 km apart. Players living within 50 km but not in the same zip code are 21.5 times more likely to be partners than the ones farther apart. The results show that players who are proximate in geographical distance are more likely to engage in interaction than those who are not proximate. Moreover, close proximity has a substantially bigger impact than medium or low proximity. Hypotheses 1 and 2 are supported.

In all models, the negative coefficients of *Age_difference* and *Game_age_difference* support Hypotheses 5 and 6. Individuals tend to play with people who have similar age and join the game at the similar time i.e., have the similar level of online experience. Among the two, the game age difference, i.e., the difference in registration dates, has a bigger impact than age. On the other hand, the impact of gender homophily (Hypothesis 4) is not significant. There is no evidence supporting that individuals with the same gender are more likely to become partners.

The goodness-of-fit diagnostics for Model 3 is illustrated in Fig. 2. The left-hand graph plots the degree distribution predicted by the Model 1b (gray lines) and the observed degree distribution (the solid line). The model captures the overall shape of the observed degree distribution, although our model overestimates the number of isolates and underestimates the number of high degree nodes. This is because no isolates are included in the sample. The right-hand graph plots the edgewise shared partner distribution and the GWESP statistic generated by the estimated model captures the correct amount of clustering in the observed partnership network.

5. Discussion

5.1. Distance still matters in EverQuest II

Similar to previous research on social structures in the offline world [28], this study shows that spatial proximity has a strong impact on the structure of online relation networks. Even though in virtual worlds players are not immediately aware of others' offline locations, shorter distance between players' offline locations significantly increases the likelihood of online collaboration. The impact of distance is not linear: short distance plays a dominant role in structuring online relations. When two players are located in the same zip code, they have a greater chance to be partners online. Such an impact decreases dramatically as distance between players exceeds city boundaries (ranging from 2 to 86 km according to the U.S. Census 2000). Model 2 suggests that the distances between 50 and 800 km only have about 20% more impact than the distance beyond 800 km. This effect suggests that the online relations are connected to the ability of making face-to-face contact.

Temporal proximity, measured by time zone difference, also exerts some influence on player interactions. People located in different time zones are less likely to interact and play with each other. However, after controlling for spatial proximity, the magnitude of the impact of temporal proximity is very small. This might be the result of the small variance in time zones in our sample. Since all players are in North America and the standard deviation of time zones is just 1.2 h, players are less affected by temporal proximity when coordinating group play across a smaller time difference, especially in the evening or during weekends.

As predicted by Hypotheses 5 and 6, age and game age homophily exists in virtual worlds. Individuals are more likely to interact with players with similar ages. And when they join the virtual world at a similar time, they are more likely to interact with each other. Gender homophily, on the other hand, is not supported as the same gender has a negative impact on partnership relations. A detailed analysis

Table 2
ERGM estimation results for proximity and homophily.

	Model 0	Model 1	Model 2	Model 3	Hypotheses
Proximity					
Same zip code				6.58^{***} (.18)	H1, 2: supported
Short distance				3.07^{***} (.13)	
Distance impact			5.99^{***} (.14)		H1, 2: supported
Time zone difference		-0.49^{***} (.04)	−0.01 (.005)	-0.11^{**} (.04)	H3: supported
Homophily					
Same gender		−0.09 (.29)	−0.15 (.35)	−0.17 (.34)	H4: not supported
Age difference		-0.04^{***} (.004)	-0.03^{***} (.005)	-0.03^{***} (.005)	H5: supported
Game age difference		-0.05^{***} (.008)	-0.04^{***} (.008)	-0.04^{***} (.01)	H6: supported
Control					
Gender—female	−0.08 (.09)	−0.14 (.25)	−0.21 (.28)	−0.27 (.29)	
Age	0.004 (.006)	0.01 (.02)	0.009 (.03)	0.01 (.02)	
Game age	-0.01^{*} (.005)	-0.02^{*} (.009)	-0.02^{+} (.01)	-0.02^{*} (.01)	
Edges	-8.93^{***} (.53)	-7.84^{***} (2.14)	-8.69^{**} (2.84)	-8.50^{***} (2.55)	
Weighted degree	3.80^{***} (.22)	3.90^{***} (.30)	3.93^{***} (.56)	3.88^{***} (.42)	
Shared partners	1.86^{***} (.08)	1.82^{***} (.10)	1.03^{***} (.13)	0.85^{***} (.13)	
AIC	17,678	17,151	14,600	14,511	
BIC	17,750	17,271	14,732	14,654	
Log likelihood	−8833 (df=6)	−8566 (df=10)	−7289 (df=11)	−7243 (df=12)	

Signif. codes: $0 < ^{***} < 0.001 < ^{**} < 0.01 < ^{*} < 0.05 < ^{+} < 0.1$.

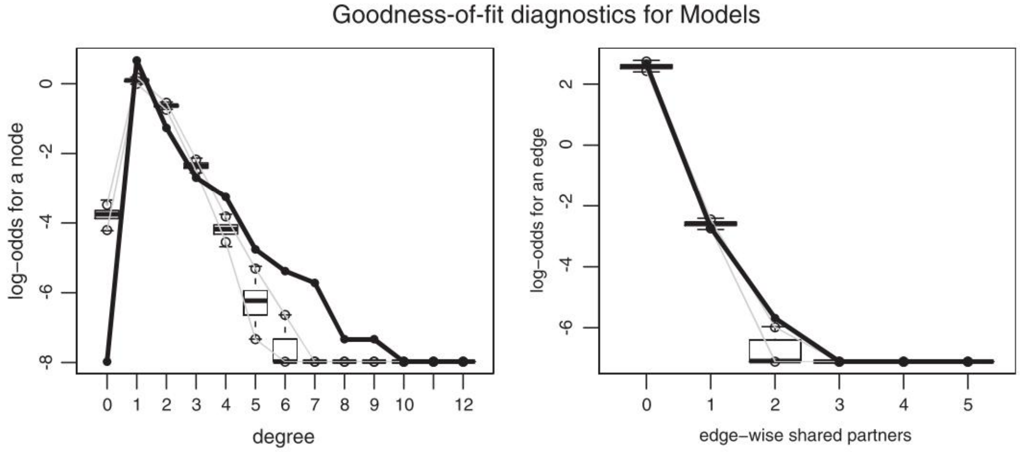


Fig. 2. Goodness-of-fit diagnostics for Models.

indicates that the negative impact is mostly from female–female matching, i.e., female players are very unlikely to partner with other female players. Although it is not consistent with the research findings on offline friendship, this still leads to a special play pattern connecting online relations with offline attributes: as Williams et al. [44] reported, 32% of people play with a romantic partner, e.g., spouse, fiancé, or boyfriend/girlfriend, which brings many male–female ties into the game.

5.2. Migration from offline to online

Our results generally support the association between proximity and interactions in virtual worlds. In addition, the detailed data analysis suggests that the migration of offline relations into online worlds plays an important role in bringing offline proximity and homophily into virtual worlds. When people decide to collaborate online, players prefer others with short distance. This leads to an assumption that the players seem to know where other players are located offline prior to their online interactions. A possible explanation is that individuals may bring their offline relations into virtual worlds and play with offline friends in the game. The box plot diagram in Fig. 3 shows the distance distribution of partnership relations established during the first 20 days players registered in EverQuest II. For the 1525 players included in the sample (i.e., the egos in Fig. 3), the distances of their new collaboration relations are calculated and grouped by their game ages (in days). Among the players with one day of game age, almost one quarter of their partners were located in the same zip code, i.e., the distance of zero. This implies that many players started the game with their offline friends. For the players registered more than one day, the patterns are similar: the distances have a bigger variance and there are still some ties with zero distance, i.e., players find new partners in the game and may connect with a few offline friends later on.

Williams [43] used the concept of "the mapping principle" to describe the phenomenon of a behavior in the virtual world matching that of the offline parallel. Our results support one test scenario in which dyad level behavior of human collaboration that occurred in virtual spaces will map to it in real spaces. The principle of proximity characterized by a range of social theories is still applicable online. Although the models do not explain the causal relationships between online and offline behaviors, a survey conducted in EverQuest II [44] reported that almost 70% of respondents played together with persons

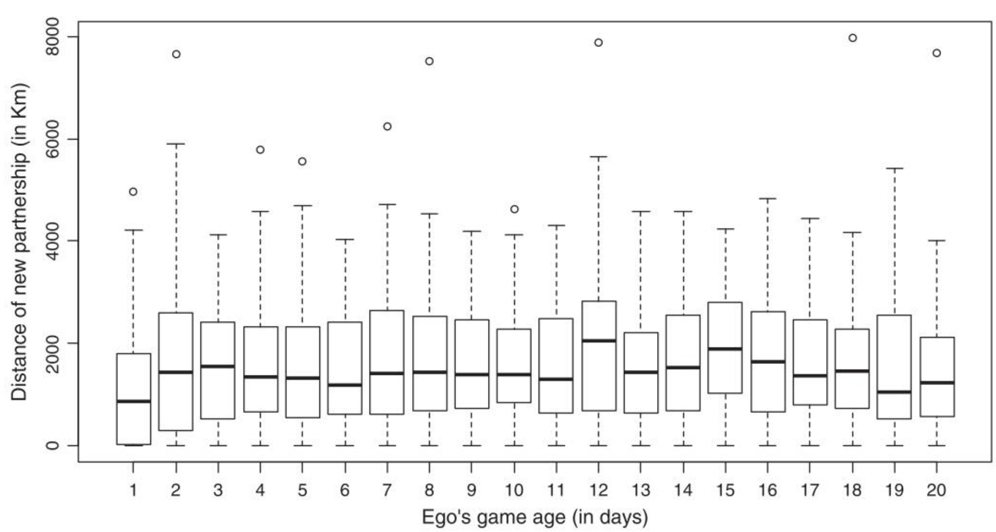


Fig. 3. Distances of new partnership during the first 20 days in the game.

they knew offline. People migrate their offline relations to the game and therefore the proximity associated with the relations is projected onto virtual worlds.

6. Conclusion

In this study, we analyze the impact of distance, time zones, players' gender, age, and game age on the process of relation building in virtual worlds given the endogenous network structures of online ties. To differentiate the effects on various human interactions online and offline among distributed individuals, this study suggests a more nuanced approach to examine the impact of proximity by decomposing the construct itself. The results show that spatial proximity, temporal proximity, as well as homophily in age and game age have a significant impact on players' behavior in creating online relations. There is no evidence, however, of gender homophily in EverQuest II. The theories of proximity and homophily are still valid in virtual worlds. The findings of this study illuminate the complex interplay between spatial proximity, temporal proximity, and homophily. This study suggests that migration of offline relations into online worlds plays an important role to bring offline proximity and homophily into virtual worlds. With our findings we hope to stimulate future research that will further explore the process through which offline proximity and homophily are introduced to online relation networks.

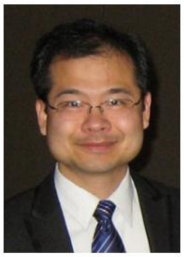
In this study, all partnership ties are considered the same and converted to binary relations. When studying the impact of proximity and homophily, partnership relations with a few activities together have the same weight as active relations. This setting could be a limitation when studying the filtering effects of online relation building. A player may still play with partners in different time zones, but less frequently than other partners in the same time zone. Furthermore, because these dimensions of proximity differ in their theoretical mechanisms, their effects also vary according to the type of interaction. For example, as O'Leary and Cummings [33] proposed, spatial proximity is more likely to influence the possibility of face-to-face communication, while temporal proximity has a stronger impact on real-time synchronous communication. Future studies will examine the impact of individual offline attributes on various types of online relations as well as their frequency and duration in virtual worlds.

Acknowledgments

This study was supported by the National Science Foundation (Grant Nos. CNS-1010904, OCI-0904356, & IIS-0841583).

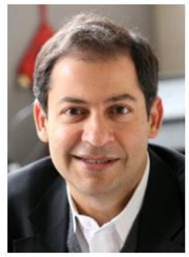
References

[1] D. Abrams, M.A. Hogg, Social Identity and Social Cognition, Blackwell, Oxford; Malden, Mass., 1999.
[2] T. Allen, Managing the Flow of Technology, MIT Press, Cambridge, MA, 1977.
[3] D.J. Armstrong, P. Cole, Managing distances and differences in geographically distributed work groups, in: P. Hinds, S. Kiesler (Eds.), Distributed Work, The MIT Press, Cambridge, MA, 2002, pp. 167–186.
[4] D. Brockmann, L. Hufnagel, T. Geisel, The scaling laws of human travel, Nature 439 (2006) 462–465.
[5] R.S. Burt, Models of network structure, Annual Review of Sociology 6 (1) (1980) 79–141.
[6] D.E. Byrne, The Attraction Paradigm, Academic Press, New York, 1971.
[7] F. Cairncross, The Death of Distance: How the Communications Revolution Will Change Our Lives, Harvard Business School Press, Boston, Mass., 1997.
[8] M.J. Culnan, M.L. Markus, Information technologies, in: F.M. Jablin, L. Putnam, K.H. Roberts, L.W. Porter (Eds.), Handbook of Organizational Communication: An Interdisciplinary Perespective, Sage, Newbury Park, CA, 1987, pp. 420–443.
[9] J.D. Eveland, T.K. Bikson, Evolving electronic communication networks: an empirical assessment, Proceedings of the 1986 ACM Conference on Computer-supported Cooperative Work, Austin, TX, 1986.
[10] T. Falk, R. Abler, Intercommunications, distance, and geographical theory, Geografiska Annaler 62 (2) (1980) 59–67.
[11] S.L. Feld, The focused organization of social ties, The American Journal of Sociology 86 (5) (1981) 1015–1035.
[12] L. Festinger, S. Schacter, K. Back, Social Pressures in Informal Groups, 1st ed. Stanford University Press, Palo Alto, CA, 1950.
[13] T.L. Friedman, The World is Flat: A Brief History of the Twenty-first Century. 1st further updated and expanded hardcover ed. Farrar, Straus and Giroux, New York, 2007.
[14] C.B. Gibson, J.L. Gibbs, Unpacking the concept of virtuality: the effects of geographic dispersion, electronic dependence, dynamic structure, and national diversity on team innovation, Administrative Science Quarterly 51 (3) (2006) 451–495.
[15] S. Goodreau, M. Handcock, D. Hunter, C. Butts, M. Morris, A statnet tutorial, Journal of Statistical Software 24 (8) (2008).
[16] K. Hampton, B. Wellman, Long distance community in the network society: contact and support beyond netville, American Behavioral Scientist 45 (3) (2001) 476–495.
[17] F. Heider, The Psychology of Interpersonal Relations, Wiley, New York, 1958.
[18] In: P. Hinds, S. Kiesler (Eds.), Distributed Work, MIT Press, Cambridge, MA, 2002.
[19] P.J. Hinds, K.M. Carley, D. Krackhardt, D. Wholey, Choosing work group members: balancing similarity, competence, and familiarity, Organizational Behavior and Human Decision Processes 81 (2) (2000) 226–251.
[20] H. Ibarra, Homophily and differential returns: sex differences in network structure and access in an advertising firm, Administrative Science Quarterly 37 (3) (1992) 422–447.
[21] J. Katz, Geographical proximity and scientific collaboration, Scientometrics 31 (1) (1994) 31–43.
[22] J. Kleinberg, The small-world phenomenon: an algorithmic perspective, Proceedings of 32nd ACM Symposium on Theory of Computing, 2000.
[23] R. Kraut, C. Egido, J. Galegher, Patterns of contact and communication in scientific research collaboration, Proceedings of 1988 ACM Conference on Computer-supported Cooperative Work, Portland, Oregon, United States, 1988.
[24] R.E. Kraut, S.R. Fussell, S. Brennan, J. Siegel, Understanding effects of proximity on collaboration: implications for technologies to support remote collaborative work, in: P.H.S. Kiesler (Ed.), Technology and Distributed Work, MIT Press, Cambridge, MA, 2002, pp. 137–162.
[25] P.F. Lazarsfeld, R.K. Merton, Friendship as a social process: a substantive and methodological analysis, in: M. Berger, T. Abel, C.H. Page (Eds.), Freedom and Control in Modern Society, Van Nostrand, New York, 1954.
[26] M. McPherson, L. Smith-Lovin, Homophily in voluntary organizations: status distance and the composition of face-to-face groups, American Sociological Review 52 (3) (1987) 370–379.
[27] M. McPherson, L. Smith-Lovin, J.M. Cook, Birds of a feather: homophily in social networks, Annual Review of Sociology 27 (2001) 415–444.
[28] D. Mok, B. Wellman, Did distance matter before the Internet?: interpersonal contact and support in the 1970s, Social Networks 29 (3) (2007) 430–461.
[29] K.A. Mollica, B. Gray, L.K. Trevino, Racial homophily and its persistence in newcomers' social networks, Organization Science 14 (2) (2003) 123–136.
[30] P.R. Monge, N.S. Contractor, Theories of Communication Networks, Oxford University Press, New York, 2003.
[31] P.R. Monge, L.W. Rothman, E.M. Eisenberg, K.I. Miller, K.K. Kirste, The dynamics of organizational proximity, Management Science 31 (9) (1985) 1129–1141.
[32] M.E.J. Newman, Clustering and preferential attachment in growing networks, Physical Review E 64 (2001) 025102.
[33] M.B. O'Leary, J.N. Cummings, The spatial, temporal, and configurational characteristics of geographic dispersion in teams, MIS Quarterly 31 (3) (2007) 433–452.
[34] G.M. Olson, J.S. Olson, Distance matters, Human Computer Interaction 15 (2000) 139–178.
[35] G. Robins, T. Snijders, P. Wang, M. Handcock, P. Pattison, Recent developments in exponential random graph (p*) models for social networks, Social Networks 29 (2) (2007) 192–215.
[36] M. Ruef, H.E. Aldrich, N.M. Carter, The structure of founding teams: homophily, strong ties, and isolation among U.S. entrepreneurs, American Sociological Review 68 (2) (2003) 195–222.
[37] Y. Takhteyeva, A. Gruzd, B. Wellman, Geography of Twitter networks, Social Networks 34 (2012) 73–81.
[38] J.C. Turner, Rediscovering the Social Group: Self-categorization Theory, Blackwell Pub., Oxford, UK, 1987.
[39] L.M. Verbrugge, A research note on adult friendship contact: a dyadic perspective, Social Forces 62 (1) (1983) 78–83.
[40] S. Wasserman, P.E. Pattison, Logic models and logistic regressions for social networks: I. An introduction to Markov graphs and p^*, Psychometrika 61 (1996) 401–425.
[41] B. Wellman, Computer networks as social networks, Science 293 (5537) (2001) 2031–2034.
[42] B. Wellman, J. Boase, W. Chen, The networked nature of community online and offline, IT & Society 1 (1) (2002).
[43] D. Williams, The mapping principle, and a research framework for virtual worlds, Communication Theory 20 (2010) 451–470.
[44] D. Williams, M. Consalvo, S. Caplan, N. Yee, Looking for gender (LFG): gender roles and behaviors among online gamers, Journal of Communication 59 (2009) 700–725.
[45] A. Wittel, Toward a network sociality, Theory, Culture & Society 18 (6) (2001) 51–76.
[46] N. Contractor, Some assembly required: leveraging web science to understand and enable team assembly, Philosophical Transactions of Royal Society A (2013).



Yun Huang is a research associate in the Science of Networks in Communities (SONIC) research group in the department of Industrial Engineering and Management Sciences at Northwestern University. His research explores the dynamic and evolution of individual behavior and interactions in digital-enabled environments such as scientific collaboration, online communities, and virtual worlds using data mining, social network analysis, and economics approaches.

Cuihua Shen (www.shencuihua.com) is an assistant professor at the Emerging Media and Communication Program, University of Texas at Dallas. Her research focuses on the evolution and effects of virtual environments such as online groups, MMOs, and social networking sites.



Noshir S. Contractor is the Jane S. & William J. White Professor of Behavioral Sciences in Department of Industrial Engineering & Management Sciences in the McCormick School of Engineering & Applied Science, Department of Communication Studies in the School of Communication and Department of Management & Organizations in the Kellogg School of Management at Northwestern University, USA. He is the Director of the Science of Networks in Communities (SONIC) Research Group at Northwestern University.