Individual Motivations and Network Effects: A Multilevel Analysis of the Structure of Online Social Relationships

By
BROOKE FOUCAULT WELLES
and
NOSHIR CONTRACTOR

This article explores the relative influence of individual and network-level effects on the emergence of online social relationships. Using network modeling and data drawn from logs of social behavior inside the virtual world *Second Life*, we combine individual- and network-level theories into an integrated model of online social relationship formation. Results reveal that time spent online and the network pressure toward balance (individuals tending to form relationships with others who have relationships in common) predict the emergence of online relationship ties, while gender, age, proximity, homophily (the tendency of individuals to form relationships among people with similar traits), and preferential attachment are not significant predictors within the observed networks. We discuss these results in light of existing research on online social relationships and describe how digital data and network analytics enable novel insights about the emergence of online social relationships.

Keywords: network science; computational social science; online friendship; p*/ERGM; multilevel multitheoretical modeling; virtual worlds

Studies of online relationship formation date back to the earliest days of Internet research, mapping why and how individuals choose to form social relationships on the web. Historically, these studies have focused on individuals as the unit of analysis; investigating, for example, how age, gender, and Internet use influence the likelihood that an individual will form personal relationships online (Parks and Floyd 1996).

Brooke Foucault Welles is an assistant professor of communication studies and a faculty affiliate of the Network Science Institute at Northeastern University.

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

DOI: 10.1177/0002716214565755

More recently, thanks to the availability of electronic data, researchers have begun to use whole networks as the unit of analysis, mapping the structure and evolution of online personal relationship networks (Kossinets and Watts 2006). This study bridges these two streams of research, applying multilevel, multitheoretical network analysis (Monge and Contractor 2003) to test the relative explanatory power of individual and network effects on the emergence of online social relationships. Specifically, we apply p*/Exponential Random Graph Modeling (p*/ERGM; hereafter ERGM) to examine the relative influence of two individual-level theories (homophily and proximity) and two network-level theories (balance and preferential attachment) on the emergence of online social relationships.

This multilevel research is enabled by the availability of online social relationship data. Until recently, it was difficult to obtain data about social relationship networks to answer questions about the relative role of individual- and network-level factors on relationship formation. However, multilevel data about individuals and their social networks are increasingly collected unobtrusively and automatically within online games, virtual worlds, and social networking sites. In this article, we leverage data from the virtual world *Second Life* to examine a multilevel model of factors that influence the emergence of online social relationships. As online social interaction grows in both reach and frequency, it becomes increasingly important to understand the factors that drive the formation, maintenance, and dissolution of online relationships. Moreover, these relationships, and the trace data they leave behind, may serve as a lens for understanding the complex processes that drive social relationships more generally in the online and offline worlds (Williams 2010).

Theories of Online Relationship Formation

There are a number of theoretical models of online realtionship formation (e.g., Kossinets and Watts 2006; Peter, Valkenburg, and Schouten 2005); however, most are limited to a single unit of analysis—either individuals or networks, but not both. Increasingly, there is evidence that neither focus is suffficient for explaining the complexities of emergent social behavior; the choice of social partners is a product of an individual's motivations and the constraints of his/her social context (Kossinets and Watts 2006). In this article, we propose a multitheoretical, multilevel model of online social relationship formation that seeks to explain the emergence of online social relationship ties by performing analyses across analytic levels (Monge and Contractor 2003). Our model is based on a multitheoretical, multilevel model of relationship formation in the offline world (Espelage, Green, and Wasserman 2007) and examines two individual-level

NOTE: We would like to thank Linden Labs for providing the data used in this analysis. This research was funded, in part, by the National Science Foundation (IIS-0841583, OCI-0904356, and CNS-1010904) and the U.S. Army Research Laboratory (W911NF-09-2-0053). We gratefully acknowledge all our sponsors. The findings presented here do not in any way represent, either directly or through implication, the policies of these organizations.

effects (proximity and homophily) and two network-level effects (balance and preferential attachment).

Proximity

Research suggests that proximity is the basic element in nearly every relationship (Priest and Sawyer 1967). Physical proximity (being physically close) increases the likelihood of communication, which in turn increases the likelihood that social relationships will develop (Monge et al. 1985). Somewhat unexpectedly, a number of researchers have demonstrated that offline physical proximity affects online social relationship choices, where people are far more likely to associate with others who are geographically proximate, even if they do not know them in the offline world (Huang et al. 2009; Mesch and Talmud 2007).

Monge and Contractor (2003) propose an alternate proximity measure for online environments called *digital proximity*. Just as geographic proximity increases the likelihood of interaction because two people are physically co-present, digital proximity refers to how "reachable" an individual is online, either because she or he spends a lot of time online, communicates frequently, or both. Several studies have demonstrated that digital proximity relates to online relationship formation and tie strength, whether measured as total time spent online (DiMaggio et al. 2001), or time since joining an online community, regardless of total time spent online (Parks and Floyd 1996; Peter, Valkenburg, and Schouten 2005). We therefore hypothesize:

Hypothesis 1a (H1a): Offline physical proximity is positively associated with online social relationship formation.

H1b: Time spent online is positively associated with online social relationship formation.

H1c: Time since joining an online community is positively associated with online social relationship formation.

Homophily

Proximity does not guarantee that social relationships will form; most people do not form social relationships with the majority of people they encounter every day. When proximity is controlled for, people tend to cluster around common traits, a phenomenon called *homophily* (McPherson, Smith-Lovin, and Cook 2001). Although lack of visual clues and a general disinclination to discuss offline personal characteristics may dampen the effects of homophily online, researchers have noted both age and gender homophily in online social relationships (Mazur and Richards 2011). We therefore hypothesize:

H2a: Individuals of similar age are more likely to form online social relationships than individuals of different ages.

H2b: Individuals of the same gender are more likely to form online social relationships than individuals of different genders.

Balance

Balance theory predicts that individuals will tend to form relationships with others who already have relational partners in common (Heider 1958). Put more simply, people tend to be friends with the friends of their friends. In networks, balance theory manifests as closed triads. In social settings, people almost always close open triads, an effect that has been observed in online and offline relationships alike (Brown and Miller 2000). Therefore, we hypothesize:

H3: Closed triads will appear in online social relationship networks more often than would be predicted by chance.

Preferential attachment

Given a choice, people disproportionately choose to socialize with high-status individuals in their networks (Epstein 1983). Formally, this phenomenon is described by the theory of preferential attachment, or a tendency for already-popular individuals to disproportionately attract more connections than their less popular peers (Gould 2002; Magee and Galinsky 2008). A number of researchers have noted that individuals with many social ties online tend to form more social ties over time, as the theory of preferential attachment would predict (Huffaker et al. 2009; Peter, Valkenburg, and Schouten 2005). Therefore, we hypothesize:

H4: The likelihood that an individual will form an online social relationship is positively related to the number of online social relationships that individual already has.

Methods

Data

To test the proposed hypotheses, we used data sampled from a large proprietary dataset provided to the research team by Linden Labs, the company that owns the virtual world *Second Life. Second Life* is an immersive virtual world where users interact via animated characters (avatars) in three-dimensional virtual space. It is similar to other online virtual worlds, including massively multiplayer online games (MMOGs), except that in *Second Life* there is no overall storyline or goal. Most users join *Second Life* to socialize, with a sizable minority of residents there to run and participate in professional or educational activities (de Nood and Attema 2006).

Network analytics

Unlike other types of data collected from individuals, relational data typically show strong interdependencies. These interdependencies make it inappropriate to analyze network data, including the data used in this analysis, using traditional statistical analyses. Instead, we used a social network analysis approach called p*/ERGM (Robins et al. 2007). ERGM is similar to traditional regression modeling, but it can be used to examine the interdependent mechanisms that drive the creation, maintenance, and dissolution of network ties (Contractor, Wasserman, and Faust 2006).

In ERGM, a network is a set of nodes connected by relationships. In any given network, the particular pattern of nodes and relationships is called the *network structure*. ERGM examines the statistical likelihood that certain network structures will be observed. As a baseline, ERGM assumes that all relationships in a network are formed at random, and therefore, no one network structure is any more or less likely to emerge than any other. Of course, in practice, social network structures are rarely random. Instead, the structure of an observed network represents the culmination of social processes where individuals form relationships based on their goals and motivations. Each of those goals and motivations can be mapped to a theory of networked behavior and, ultimately, back to a particular network structure. ERGM tests for the statistical likelihood of theoretically hypothesized network structures, compared to random chance alone. For additional details about ERGM, including its application to social science, we refer readers to Robins et al. (2007) or Shumate and Palazzolo (2010).

Weighted least squares aggregation

One of the chief limitations of ERGM is that computational limitations present challenges for the analysis of very large online networks, including the network analyzed here. However, recent advances in network meta-analysis offer a work around. Instead of modeling an entire network, researchers can get a valid picture of network properties by sampling bounded subnetworks from the overall network, analyzing their properties and aggregating the results via meta-analysis (Handcock and Gile 2007). Although aggregation of ERGM results is still under development, the currently accepted standard is a weighted least squares (WLS) estimation procedure proposed by Snijders and Baerveldt (2003). They suggest that sampled networks modeled using the same model and parameter terms can be accurately combined via WLS meta-analysis as if they were individual experiments. This technique reweighs individual parameter estimates by the inverse of their effect on the variance of each model's error such that individual terms with large errors are less influential in summary estimates than terms with small errors, controlling for non-normal estimate distributions. WLS meta-analysis has been used successfully to break down large networks, compare across multiple similar networks, and deal with missing data in network analyses (Handcock and Gile 2007; Snijders and Baerveldt 2003).

Data sampling

There are a number of different sampling techniques for large networks, including sampling connected components, snowball sampling, and sampling based on exogenous characteristics such as membership in a particular organization (Handcock and Gile 2007). Recall that one of the requirements for accurate WLS aggregation is sampling bounded subnetworks from the larger network. To accommodate that requirement, we sampled our data by an exogenous characteristic, in this case, examining the social relationships among *Second Life* group members. Groups in *Second Life* are formally designated organizations with clear membership rosters that facilitate subnetwork samples with clear boundaries. Most *Second Life* groups are free to join, and although members do not have to form social relationships with one another, group membership may afford opportunities to do so.

Our random sample of 20 groups included a total of 1,254 users (1,210 unique). Group sizes ranged from 13 members to 402 members, with an average of 32.4 members per group. This average is consistent with the overall average group size for all groups included in the sample frame, 32.2 members per group.

Variables

The ERGM used in this analysis was developed in StatNet (Goodreau et al. 2008). The dependent variable for all analyses was the presence of a social relationship tie.

Independent variables included the following:

Geographic proximity. Offline geographic location was measured by country (indicated in user registration data). More than forty countries were represented in the data, with a majority from the United States (81 percent), followed by other English-speaking countries, including the United Kingdom (11 percent) and Canada (6 percent). Our model included a variable to estimate the likelihood that relationships will emerge between users from the same country (H1a).

Usage minutes. Total usage minutes were automatically logged for each user (μ = 52,950.9 minutes, *SD* = 67,184.7), and our model included a variable to test the relationship between the total amount of time a user spends in *Second Life* and his or her likelihood of forming social relationships (H1b).

Online tenure. Logs of user behavior also included information about first and last login. To test H1c, we created a variable to test the relationship between duration of time between first and last login (μ = 264.2 days, *SD* = 217.8 days) and the likelihood of social relationship formation.

Age homophily. When registering in *Second Life*, users report their dates of birth. Using this information, we calculated ages (μ = 33.59 years, *SD* = 10.84 years) and constructed a variable to test for the effect of age similarity on social relationship formation (H2a).

Gender homophily. When users create accounts in *Second Life*, they are asked to report their offline-world gender. At the time of data collection, users had to choose between two gender options (male [44 percent] and female [56 percent]). Using this information, we created a variable to examine the effect of gender matching on social relationship formation (H2b).

Balance. Hypothesis 3 predicts social relationships in *Second Life* will have the tendency to be balanced. To test this hypothesis, we used a variable called *Geometrically Weighted Edge Shared Partners* (*GWESP*) in the model (Hunter 2007). This statistic counts the number of edgewise shared partners in the network (where users who are "friends" have a third "friend" in common), applies a decay rate to account for diminishing returns on additional shared partners, and combines them into one single variable by giving counts different weights (Hunter 2007).

Preferential attachment. To test for evidence of preferential attachment in the network (H4) we included the variable *Alternating K-Star* (*altkstar*) in the model. This variable checks for a particular pattern of network relationships where certain popular individuals (with a large number of relationship ties) are surrounded by other, less popular individuals (with a small number of relationship ties), suggesting a preference for forming ties with already-popular individuals, with a certain decay rate to account for diminishing returns as new relationships are added (Snijders et al. 2006).

Controls. We also included terms in the model to control for the overall density of the network (edges) and for the nodal effects of gender, age, and country of origin. These variables account for baseline differences in relationship formation tendencies among various groups, resulting in more accurate estimates of the matching (homophily and proximity) effects.

Results

The network results reported below are based on twenty converged models that were a good fit for the observed data (Hunter et al. 2008). Hypotheses were tested using a WLS meta-analysis of the results of ERGM models fitted individually to each of the twenty sampled groups. The results in Table 1 give the weighted average effects, summarized across all twenty network models.

This table is organized by hypotheses and their corresponding structural effects. *N* denotes the total number of parameter estimates included in each meta-analysis, and μ_{WLS} is the estimated average effect size, weighted and summarized across all twenty samples, with its standard error (*SE*). Significant effects are noted with asterisks (*). T^2 is a statistic that tests whether the total effect of each parameter is zero. This value can also be used to determine the relative influence of each structural effect in the model, where structural effects with larger T^2 values have larger effects in the model (Snijders and Baerveldt 2003).

TABLE 1
Aggregate Weighted Average Effects for Twenty Sampled Networks

Hypothesis	Structural Effect	*N*	μ_{WLS}	*SE*	T^2
H1a	Physical proximity	20	–0.7269	0.7918	34,362.61
H1b	Digital proximity	20	3.9637*	1.2987	258,941.70
H1c	Online tenure	20	0.0038	0.0049	335,968.70
H2a	Age homophily	20	–0.0029	0.0016	402.48
H2b	Gender homophily	20	–1.1239	0.8177	17,074.06
H3	Balance	20	0.5722*	0.2181	132.66
H4	Preferential attachment	20	–1.0488	0.6762	114,906.80
control	Edges	20	–14.7444	23.2577	633,805.40
control	Country (U.S.)	20	11,660.52	–1.2178	0.8391
control	Age	20	–0.0069	0.0055	61,668.79
control	Gender (male)	20	–1.7976	0.9320	16,015.90

*p < .05.

As seen in Table 1, there were two statistically significant effects in the model. The first was for digital proximity (total time spent online, H1b). As H1b predicted, people who spend a great deal of time in *Second Life* are more likely to form social relationships than those who do not, regardless of how long they have had an account. Time spent online was the largest significant effect in the model, accounting for more variance in the emergence of online social relationships than any other factor.

The other significant effect was for balance. As predicted in H3, the *balance* variable in the model is significant and positive, suggesting that social relationships tend to cluster in closed triads more often than would be predicted by chance.

The remaining hypotheses on geographic proximity (H1a), online tenure (H1c), age and gender homophily (H2a and H2b), and preferential attachment (H4) were not supported.

Discussion and Conclusion

This study examined individual and network-level factors contributing to the formation of online social relationships. Unlike prior research that has focused on either individuals or networks as the unit of analysis, we leveraged the availability of online data and multilevel network modeling to test an integrated model that examines factors influencing relationship formation across analytic units. The model revealed two significant effects: time spent online and network balance.

Consistent with prior research, our results suggest that time spent online significantly predicts online relationship formation (Parks and Floyd 1996). Specifically, the more time someone spends logged into *Second Life*, the more likely she or he is to form social relationships within *Second Life*. This evokes findings from the sociological literature on community participation, where value is generated and derived from "showing up" to community events (Putnam 1995). Time spent online and social relationship formation may be mutually reinforcing, where spending time online increases the chances of being available for social interaction, and having social relationships makes it more enjoyable to spend time online.

Results also revealed a significant effect for network balance; individuals tended to form closed triads in their online social relationship networks. Consistent with the pattern of "showing up," the significant effect on balance suggests a tendency to form online social relationship networks characterized by supportive clusters of like-minded others, yielding opportunities to access social capital and support (Krackhardt 1992).

In addition to time spent online, balance was the only significant predictor of online social relationship formation in our model. This suggests that balance may be a better predictor of online social relationship formation than individual-level effects, such as proximity or homophily. Recent research has noted that individual effects tend to be conflated with network effects in social theoretical models (Rivera, Soderstrom, and Uzzi 2010; Steglich, Snijders, and Pearson 2010). The results of this study suggest the same conflation could be true for single-level models of online social relationship formation as well; prior studies may have overestimated the effect of individual effects because they did not explore underlying network effects accounting for patterns of online social behavior. The results of this research underscore the importance of looking across analytic units to understand online social behavior.

Of course, there are limitations to note. Because of data and algorithmic limitations, process-based theories of online relational formation and measures of personality are absent from the current analysis. Additionally, we know little about the precise nature of the social relationships examined, only that they occurred in groups, which could include social and professional organizations, interest groups, and educational groups, to name just a few. Although *Second Life* is similar to other online communities and virtual worlds, to be able to make claims about online relationships more broadly, researchers should conduct similar investigations with different datasets. Notwithstanding these limitations, this study demonstrates how using online data and network analytics enables more detailed understandings of the processes that drive online relationship formation, understandings that were not possible just a few years ago.

References

Brown, Thomas M., and Charles E. Miller. 2000. Communication networks in task-performing groups: Effects of task complexity, time pressure, and interpersonal dominance. *Small Group Research* 31 (2): 131–57.

Contractor, Noshir S., Stanley Wasserman, and Katherine Faust. 2006. Testing multi-theoretical, multilevel hypotheses about organizational networks: An analytic framework and empirical example. *Academy of Management Review* 31:681–703.

de Nood, David, and Jelle Attema. 2006. *Second life, the second life of virtual reality*. The Hague: EPN - Electronic Highway Platform.

DiMaggio, Paul, Eszter Hargittai, W. Russell Neuman, and John P. Robinson. 2001. Social implications of the Internet. *Annual Review of Sociology* 27:307–36.

Epstein, Joyce L. 1983. Examining theories of adolescent friendships. In *Friends in school: Patterns of selection and influence in secondary schools*, ed. Joyce L. Epstein, 39–61. New York, NY: Academic Press.

Espelage, Dorothy L., Harold D. Green Jr., and Stanley Wasserman. 2007. Statistical analysis of friendship patterns and bullying behavior among youth. *New Directions for Child and Adolescent Development* 118:61–76.

Goodreau, Steven M., Mark S. Handcock, David R. Hunter, Carter T. Butts, and Martina Morris. 2008. A statnet tutorial. *Journal of Statistical Software* 24 (9): 1–27.

Gould, Roger V. 2002. The origins of status hierarchies: A formal theory and empirical test. *American Journal of Sociology* 107 (5): 1143–78.

Handcock, Mark S., and Krista Gile. 2007. *Modeling social networks with sampled data*. Seattle, WA: University of Washington.

Heider, Fritz 1958. *The psychology of interpersonal relations*. New York, NY: John Wiley.

Huang, Yun, Cuihua Shen, Dmitri Williams, and Noshir S. Contractor. 2009. Virtually there: Exploring proximity and homophily in a virtual world. Paper presented at the Symposium on Social Intelligence and Networking, Vancouver, Canada.

Huffaker, David, Jing Wang, Jeffrey Treem, Muhammad A. Ahmad, Lindsay Fullerton, Dmitri Williams, Marshall Scott Poole, and Noshir S. Contractor. 2009. The social behaviors of experts in massive multiplayer online role-playing games. Paper presented at the 2009 International Conference on Computational Science and Engineering Vancouver, Canada.

Hunter, David R. 2007. Curved exponential family models for social networks. *Social Networks* 29 (2): 216–30.

Hunter, David R., Mark S. Handcock, Carter T. Butts, Steven M. Goodreau, and Martina Morris. 2008. ERGM: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of Statistical Software* 24 (3). Available from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2743438/.

Kossinets, Gueorgi, and Duncan J. Watts. 2006. Empirical analysis of an evolving social network. *Science* 311 (57): 88–90.

Krackhardt, David. 1992. The strength of strong ties: The importance of Philos in organizations. In *Networks and organizations: Structure, form and action*, eds. Nitin Nohrida and Robert G. Eccles, 216–39. Boston, MA: Harvard Business School Press.

Magee, Joe C., and Adam D. Galinsky. 2008. Social hierarchy: The self-reinforcing nature of power and status. *Academy of Management Annals* 2 (1): 351–98.

Mazur, Elizabeth, and Lacey Richards. 2011. Adolescents' and emerging adults' social networking online: Homophily or diversity? *Journal of Applied Developmental Psychology* 32 (4): 180–88.

McPherson, Miller, Lynn Smith-Lovin, and James M. Cook. 2001. Birds of a feather: Homophily in social networks. *Annual Review of Sociology* 27:415–44.

Mesch, Gustavo S., and Ilan Talmud. 2007. Similarity and the quality of online and offline social relationships among adolescents in Israel. *Journal of Research on Adolescence* 17 (2): 455–66.

Monge, Peter R., and Noshir S. Contractor. 2003. *Theories of communication networks*. New York, NY: Oxford University Press.

Monge, Peter R., Lynda W. Rothman, Eric M. Eisenberg, Katherine I. Miller, and Kenneth K. Kirste. 1985. The dynamics of organizational proximity. *Management Science* 31 (9): 1129–41.

Parks, Malcolm R., and Kory Floyd. 1996. Making friends in cyberspace. *Journal of Communication* 46 (1): 80–97.

Peter, Jochen, Patti M. Valkenburg, and Alexander Schouten. 2005. Developing a model of adolescent friendship formation on the Internet. *CyberPsychology and Behavior* 8 (5): 423–30.

Priest, Robert F., and Jack Sawyer. 1967. Proximity and peership: Bases of balance in interpersonal attraction. *American Journal of Sociology* 72 (6): 633–49.

Putnam, Robert D. 1995. Bowling alone: America's declining social capital. *Journal of Democracy* 6 (1): 65–78.

Rivera, Mark T., Sara B. Soderstrom, and Brian Uzzi. 2010. Dynamics of dyands in social networks: Assortative, relational and proximity mechanisms. *Annual Review of Sociology* 36:91–115.

Robins, Gary, Pip Pattison, Yuval Kalish, and Dean Lusher. 2007. An introduction to exponential random graph (p*) models for social networks. *Social Networks* 29 (2): 173–91.

Shumate, Michelle, and Edward T. Palazzolo. 2010. Exponential random graph (p*) models as a method for social network analysis in communication research. *Communication Methods and Measures* 4 (4): 341–71.

Snijders, Tom A. B., and Chris Baerveldt. 2003. A multilevel network study of the effects of delinquent behavior on friendship evolution. *Journal of Mathematical Sociology* 27 (2): 123–51.

Snijders, Tom A. B., Philippa E. Pattison, Garry L. Robins, and Mark S. Handcock. 2006. New specifications for exponential random graph models. *Sociological Methodology* 36 (1): 99–153.

Steglich, Christian, Tom A. B. Snijders, and Michael Pearson. 2010. Dynamic networks and behavior: Separating selection from influence. *Sociological Methodology* 40 (1): 329–93.

Williams, Dmitri. 2010. The mapping principle, and a research framework for virtual worlds. *Communication Theory* 20 (4): 451–70.