Will You Be My Friend? An Exploration of Adolescent Friendship Formation Online in Teen Second Life

Brooke Foucault, Mengxiao Zhu, Yun Huang, Zeina Atrash & Noshir Contractor
Northwestern University
[b-foucault, mzhu, yun, nosh]@northwestern.edu, zeina@u.northwestern.edu

Abstract: *Although there have been many studies of the effect that online friendships have on adolescents, fewer have investigated how young people form friendships online. Those that have studied online friendship have found that it is generally different than real-world friendship in form and function. This study applies Monge & Contractor's (2003) Multi-theoretical Multilevel model to study the emergence of adolescent friendships in the online world. This study presents the results of a large scale social network analysis based on computer logs of friendships in Teen Second Life, an online social world for adolescents age 13-17. Results suggest that friendship formation follows several predictable patterns found in the real world, but not found in all previous studies of online friendship formation among adolescents, including tendencies towards balance and a preference for friends of high status and friends who are geographically, temporally, and physically proximate. These findings indicate that online friendship patterns in Teen Second Life mirror those found in real life.*

1. Introduction

The story is all too familiar. An adolescent boy, "Eddie," sits at his computer browsing his favorite website when a chat window appears, "Will you be my friend?" Thoughtlessly, Eddie agrees, "Sure." Naive to the dangers lurking on the Internet, he unwittingly puts himself at risk of sexual assault by befriending a cunning, and probably much older, Internet predator. Stories like this one have become a staple in primetime crime dramas, public service advertisements, and the evening news. Although recent evidence has suggested that Internet predation is strikingly rare (Wolak, Finkelhor, Mitchell, & Ybarra, 2008), the plot is nonetheless so recognizable and so prevalent that it is easy to believe that many, if not most, adolescents are making risky and haphazard decisions about who to befriend online.

However, a second look reveals just how unlikely this scenario is. When we accept, at face value, that children and teens befriend any stranger they meet online, we assume that adolescents' online friendship formation is random, un-thoughtful, and unpredictable. Not

Acknowledgements: *The authors would like to thank Linden Labs, especially Everett Harper and Bert Chi, for providing the data used in this analysis. The research reported here was supported by grants from the National Science Foundation (#NSF-IIS 0729505) and from the Army Research Institute (#W91WAW-08-C-0106).*

surprisingly, the literature on adolescent friendship formation suggests otherwise. Though adolescent friendships may appear fickle, a deeper look reveals that they have highly regular and predictable structures.

Interestingly, though, previous studies of adolescent friendships online have not always found the patterns characteristic of friendships in the real world. As a result, some have concluded that online friendship is somewhat more random (and, as a result more risky) than real-world friendship. However, in this paper, we challenge that assumption, and suggest that adolescent friendship formation online is deliberate and thoughtful, with predictable structural signatures. To do so, we present the results of a social network analysis of the friendship networks in Teen Second Life, an online social world for adolescents ages 13-17. Although there are many studies on the affects of adolescent social networks on health, behavior, and well-being, and on the extent to which certain behaviors, such as smoking, predict friendship formation and change over time (Espelage, Holt, & Henkel, 2003; Moody, 2001), few network studies have focused broadly on friendship formation, and fewer still consider how friendships form online. As young people increasingly form and maintain friendships online it is becoming critical for us to understand how and why these friendships form.

Here, we explore those questions in detail. First, we briefly review the literature on the form and function of adolescent friendships, highlighting the importance of these relationships in young people's lives. Then, we examine various theories of adolescent friendship formation, explaining what factors predict the adoption and termination of friendships for children and teens in the real world. Next, we review the literature on online friendship formation and posit several hypotheses and research questions about the nature of online friendship formation. Finally, we test our hypotheses using the computer logs of the friendship network data from the Teen Grid of

the Second Life virtual world. Results suggest that online friendship formation follows predictable patterns that are similar to those found in the offline world.

2. Theories of Adolescent Friendship Formation in the Offline World

Friendships are among the most important relationships in an adolescent's life. Perhaps second only to parents in influence, peers occupy a large portion of every adolescent's waking hours, and they are often the source of children's greatest pleasures and deepest frustrations (Rubin, 1980). During adolescence, as children's reliance on parents decreases, peers begin to occupy an increasing amount of an adolescent's time and attention (Brown, 2004). Moreover, as adolescence progresses, peers take on ever-increasing importance, and become the locus of significant social, emotional, and functional support (Savin-Williams & Berndt, 1990).

In their broadest sense, "friends" are non-familial relations that serve important functions in the lives of children and adults. However, the ways that people understand friendship changes throughout childhood and adolescence, and into adulthood (Rubin, 1980). Though children begin to play with others in their first and second year, and refer to others as "friends" as young as age three, it is not until early adolescence that friendship is recognized as anything more than a one-sided relationship of convenience (Rubin, 1980). Beginning around 11-13 years old, children come to recognize the reciprocal value of friends, and can identify characteristics of desirable friends, beyond shared interests. Later, beginning around 15-16 years of age, many of the friendship requirements from early adolescence remain, however additional characteristics gain importance. Older adolescents tend to want friends who are loyal and trustworthy, and who can provide support in an emotional or physical emergency (Bigelow & La Gaipa, 1980). Interestingly, overall stability in their relationships (Mannarino, 1980).

Interestingly, until late in adolescence, friendships are relatively unstable. Bigelow and La Gaipa note that, "it is not until the child is 12 years old that the notion of friendship implies an expectation of some kind of permanence," and "stability is not very characteristic of children's friendships until about 16 years of age," (Bigelow & La Gaipa, 1980, p. 38). Given this churn, it might be easy to believe that these decisions are random, but Cairns and Cairns (1994) note that, "Adolescent friendships are dynamic and predictably unreliable [but], the shifts are not willy-nilly, and they occur within a framework which is itself predictable" (p. 99). In this section we investigate factors and theories that predict adolescent friendship formation and dissolution. The remainder of this section reviews theories that explain the formation of friendship networks in the offline world using Monge and Contractor's (2003) Multi-theoretical Multilevel (MTML) models for the emergence of networks.

Theories of Proximity. Theories of proximity predict that physical proximity increases the likelihood of interaction, which, in turn, predicts the likelihood of friendship ties forming (Monge & Contractor, 2003). Research shows that proximity is the basic element in nearly every adolescent's friendships. As noted above, many researchers have identified proximity as a key factor influencing the likelihood of forming friendships, and children are more likely to become friends with others in their same classroom, school, or activity group (Hardy, Bukowski, & Sippola, 2002; Neckerman, 1996). Similarly, children show *temporal proximity* preferences for friends, almost always forming friendships with others from their cohort at school and in extracurricular organizations (Hardy, Bukowski, & Sippola, 2002). So, for most young people, proximity is a necessary condition for friendship. However, it is not a sufficient condition. As Epstein (1983) notes, many adolescents who are in frequent contact do not become friends.

Theories of Homophily. When proximity is controlled for, adolescents still do not form friendships at random. Even within classrooms, schools, etc. children tend to cluster around

predictable traits, as theories of homophily would predict (Monge & Contractor, 2003). Most notably, throughout adolescence, young people tend to cluster in relatively homogenous and demographically similar groups. For instance, children tend to be friends with others of the same gender. Although this pattern is strongest among very young children, it is still dominant throughout middle and high school, and even into college (Cairns & Cairns, 1994). Similarly, young people also tend to cluster by age. Even within grade levels, where ages are relatively constricted, students tended to be friends with others who were more similar in age.

Theories of Reciprocity and Balance. At their simplest, theories of balance predict that adolescent friendships will be reciprocal. Indeed, Hartup (1993) found that reciprocated friendships are more likely to be stable and emotionally satisfying over time. Similarly, Epstein surmises that unbalanced friendships may be stressful for adolescents, and therefore adolescents not only tend to reciprocate friendships within dyads, but they also tend to form transitive triads where they are friends with their friends' friends. Further, in her longitudinal study of middle and high school students, Epstein found that the tendency to balance friendship is strong for up to three "best friends" and their friends, and then declines as additional "best friends" are added (Epstein, 1983). In other words, adolescents may have a tendency to create balanced friendships, so long as their networks are not too large.

In their study of friendships among 150 sixth grade students, Gest and colleagues (2005) confirmed many findings from previous work. They found that the friendship networks of the children in their study were dense and tended to be reciprocal and transitive. That is, students in their study tended to have many friends, friendships tended to be mutual, and students tended to be friends with the friends of their friends.

Theories of Status. Theories of status are also relevant for explaining adolescent friendship formation. As Epstein notes, friendship may be one primary way for adolescents to

gain access to status, which will afford them items or opportunities that they desire (Epstein, 1983). She posits a *Status Based Initiation Response* model to explain friendship choices. This theory assumes that adolescents are capable of evaluating the personal attitudes, beliefs, and abilities that they possess, as well as those of others. They use these evaluations to make decisions about who to become friends with. According to Epstein, "They may 'pick up' in status on some criteria to improve a social position or to anticipate an expected social position. They may "pick equal" in status because of the expected cohesiveness in relationships with others similar to themselves. Or they may "pick down" on status characteristics to obtain recognition of superiority or to accept a friend who may be low on one quality but high on others," (Epstein, 1983, p. 41). So, adolescents may not have a tendency to choose one status level over another, but may instead use a variety of strategies when it comes to status selection, each of which serves a different practical or emotional need.

Other status-based theories indicate a tendency towards preferential attachment, where "the rich get richer." Hardy and colleagues (2002), for example, found that children tend to demonstrate a preference for befriending sociometrically popular students, where popular is defined by high peer-assessed likeability. Moreover, they found that students are very unlikely to form friendships with unpopular peers, and that unpopular children tend to become identified as such very quickly after new peer groups are established. This perspective illustrate network theories of contagion (Monge & Contractor, 2003), which posits that individuals are "infected" with the idea of seeking links with those who are already well linked to their peers and avoiding links to those who are not already well linked.

Theories of Exchange. Social Exchange theory predicts that adolescents will select and reject friends based on the rewards and costs that may result from their association (Homans, 1974). Although it is certainly possible to imagine that young people are capable of

understanding the rewards of productive friendships and the pain of unproductive ones, Epstein notes that they may be developmentally incapable of judging the costs and benefits of friendships at a global level (Epstein, 1983). Work by Inhelder and Piaget (1958) suggests that the ability to synthesize and predict the potential costs and benefits of a relationship may not emerge until late adolescence, around 16-18 years of age, when children learn to manipulate past, present, and future orientations together.

Instead, Epstein suggests that adolescents may focus on extant structural and social factors that may influence the value of a particular friendship. For instance, there are often pressures within a classroom to form certain friendships or avoid others. Students may be wary of being friends with the class clown, for example, but eager to partner with someone who has the resources to excel on an upcoming project. Even more likely, adolescents may find it prudent to be friendly with others who have been assigned to work on a project together, deemphasizing previous friendships with former partners who have moved to different groups (Epstein, 1983). Similarly, Kandel (1978) found that young people are likely to make adaptive friendship changes as their priorities and interests change over time. Adolescents are more likely to select friends who have the resources to help them achieve their present goals than to keep friends who had been helpful in the past or to make new friends who may be helpful in the future.

Finally, the tendency to employ social exchange strategies and choose friendships that maximize the likelihood of achieving goals and minimize the cost of doing so may become stronger throughout adolescence as children develop a better sense of what their own goals and interests are. Epstein (1983) found that, throughout adolescence, young people become increasingly aware of their own goals and interests, and it is not until the 12th grade that the majority can identify stable long-term goals. Thus, while exchange theories may predict state-based friendships in early adolescence, they may go on to predict friendship selection more

globally as adolescents grow up. While this section has reviewed theory and research on friendship formation in the offline world, the following section reviews recent efforts to investigate friendship formation online.

3. Theories of Adolescent Friendship Formation Online

Returning now to our story of Eddie, the online teen from the introduction to this paper, the wealth of evidence about the strategies and techniques that adolescents have for choosing, evaluating, and ending friendships renders it highly unlikely that Eddie would choose his online friends so thoughtlessly. However, we still know little about how he will make decisions about who to befriend online. There is one thing we do know for sure – it is very likely that Eddie has online friends, and that he will make more online friends in the future. A recent studies from the Pew Internet and American Life Project show that 93% of American youth, ages 12-17, are online and, of these, 55% have profiles on one or more social networking sites where they keep in touch with current friends and make new friends online (Lehnhart & Madden, 2007; Macgill, 2007). Similarly, Williamson (2007) estimates that 24% of adolescents, ages 13-17, currently use virtual worlds such as World of Warcraft and Second Life to socialize and meet new people, and she predicts this number will jump to 53% by 2011. Thus, the trend appears to be strong and growing – adolescents are making friends online and they will continue to make friends online in the future.

Given this trend, and the abundance of research focused on offline youth friendships, it is puzzling that there is so little written about teens' online friendship formation. Moreover, in the literature that has been published, findings have been mixed regarding whether or not the patterns of online friendships are similar to those used in the real world.

Generally speaking, the literature on computer-mediated communication (CMC) supports the notion that offline and online strategies for choosing friends may be comparable. Although some may argue that adolescents simply do not have enough information about potential friends online to make decisions as they would in the real world (Sproull & Kiesler, 1986), others posit that the critical difference between offline communication and online communication is not the richness of information available, but the rate at which it is transmitted (Walther, 1995). That is, given sufficient time, individuals communicating online will discover many of the same traits and attributes they can identify in face-to-face communication.

This suggests that many of the same strategies that adolescents may use in the real world to form friendships will apply online as well, however they may take longer. Indeed, in their study of friendships in the real world and in online newsgroups, Chan and Cheng (2004) found that although adolescents talk about online and offline friendships differently at first, these descriptions become more similar after friendships carry on for six months or more. So, there appears to be reason to believe that at least some of the structures and attributes that influence offline friendship may also influence online friendship, even if they take longer to manifest.

In the literature specifically dedicated to online youth friendships, some have found that adolescents use strategies similar to those they use in the real world for form friends. For instance, Peter and colleagues (2005) studied friendship formation in chat rooms. Their work suggests that theories of preferential attachment and status apply online just as they do offline. In their study of 800 adolescents, ages 9-18, they found that extroverted children tended to disclose more personal information online, leading them to make online friends quickly, which in turn led their popularity to increase in a mechanism very similar to offline preferential attachment. Interestingly, however, introverts eventually became popular online as well, although through an entirely new social attachment mechanism. To compensate for their lack of popularity, introverts

tended to communicate online more frequently, a strategy that eventually led to increased social status and more friendship ties (Peter et al., 2005).

Peter and colleagues' model further suggests that certain proximity mechanisms that predict offline friendship may predict online friendship as well. They found that time spent online and frequency of communication are significantly correlated with online friendship, where more time online and more frequent communication is associated with more friendships (Peter et al., 2005). Though these behaviors are not proximity in the physical sense, they can be understood as *digital proximity* (Monge & Contractor, 2003). Adolescents who are more reachable online, either because they spend a lot of time online or because they communicate frequently online, may be considered to be more digitally proximate, and this proximity predicts friendship formation in much the same way it does in the real world.

Mesch and Talmud (2007) also found that *geographic proximity* is an important predictor of online friendship, as is similarity in traits such as age and gender. Among a representative sample of Israeli adolescents, they found that "when a friend was met at school the likelihood of similarity in age, gender and place of residence was higher than when contact was made online" (p. 455). However, considering only friends made online, those who were more similar in age, gender, and (real-life) place of residence had stronger social ties than those who were different on these dimensions (Mesch & Talmud, 2007). This suggests that while geographic proximity and trait homophily may be less important online, they are still strong predictors of online friendship formation.

However, other studies of online friendships have concluded that offline friendship patterns do not apply equally online. For instance, in their newsgroup study Chan and Cheng (2004) found that, while gender homophily seems to be a significant predictor of offline friendships, it does not predict online friendships for the same people. The students in their

sample were just as likely to have cross-gender friendships as same-gender friendships online, while they were far more likely to have same-gender friendships offline. Based on these findings, the authors concluded that the absence of offline structural and social constraints in the virtual world enables certain types of emergent friendships online that may be discouraged offline, and they suggest that future work should investigate the likelihood forming online friendships across a variety of dimensions such as age, race, social class, and culture that generally render real-world friendships unlikely (Chan & Cheng, 2004).

4. Hypotheses & Research Questions

The literature reviewed above indicates there is some evidence in support and some against the similarity of mechanisms that explain offline and online friendship formation. Adolescents may use many of their offline strategies to make friends online, but, on the other hand, removing some of the social and structural constraints of the real world may enable new social structures to emerge online. In order to resolve some of the ambiguity in the literature, and to begin to understand the mechanisms behind adolescent friendship formation online, we propose the following hypotheses about online friendship formation to be tested using data collected from the Teen Grid of the online social world Second Life.

First, given the wealth of evidence that adolescent friendships, online and offline, are influenced by a variety of structural and personal constraints it seems highly unlikely that online friendships are random. Therefore,

H1: *Adolescents' online friendship ties are not random.*

Next, given the evidence that proximity is a key predictor of friendship formation in the real world, posit that there are three kinds of proximity that affect online friendship formation as well:

H2: *Real-world geographic proximity is positively associated with online friendship formation (geographic proximity).*

H3: *Time spent online is positively associated with friendship formation (digital proximity).*

H4: *Adolescents who join the virtual world at similar times are more likely to form friendships than adolescents who join at very different times (temporal proximity).*

Next, several studies have shown that trait homophily predicts adolescent friendship formation, therefore:

H5: *Adolescents of similar age are more likely to form friendships than adolescents who are not.*

Furthermore, based on balance theories, we might expect certain structural signatures to emerge within networks of adolescent friends. Although ties in our data are necessarily reciprocal because of system design, we also expect to find balance among larger groups:

H6: *Adolescents' online friendships tend to be balanced.*

Finally, based on previous findings, we would expect status to play a role in friendship formation. Specifically:

H7: *Higher-status individuals, such as those with more experience in-world and premium accounts (where premium users incur a monthly fee in exchange for enhanced privileges) are more likely to form friendships than lower-status individuals.*

5. Data and Model

Data. In order to test the hypotheses and explore the research questions listed above, we analyzed friendship and attribute data gathered from computer logs of the users of the online

social world Teen Second Life (TSL). TSL is an age- and content-restricted portion of the Second Life virtual world, reserved for teenagers ages 13-17. Members of TSL are required to verify their age using a mobile phone or Paypal account and are prohibited from posting or participating in mature content. The only adults (18+) in TSL are either paid employees of Linden Labs, the company that owns TSL, or screened volunteer guides who offer assistance and advice to the teenage users. Otherwise, TSL is much like the main Second Life virtual world, where users interact with one another via avatars and can socialize, join groups, own land, and build a wide range of objects, from virtual tennis shoes to virtual night clubs.

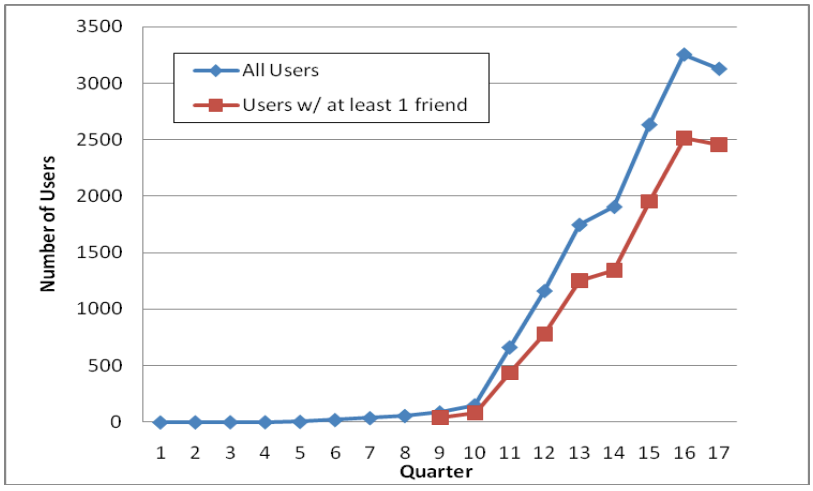
In total, there are 69,119 registered users in TSL. At registration, each of these users provided a variety of information, including date of birth, home city and country, and gender. In addition, a variety of data were generated about the users as they used TSL. For instance, for each user we have an account create date and a date of last login. We also have information about lifetime usage minutes, lifetime revenue or in-world money earned, and current account type (basic vs. premium). Finally, for each user, we have data about friendship. Friendship ties are formed in TSL when one user extends friendship to another, and the other user accepts. By design, friendships are necessarily reciprocal – friendship must be extended and accepted to exist. Therefore, our friendship network data are non-directed. TSL friends can see when one another log in and log out of the world, and they can also see where one another are while they are online.

Of the 69,119 total users, 28,992 users have at least one friend. Since this study focuses on how friendships are formed online, users without at least one friend are excluded from this analysis, and from hereon these users are referred to as "active users." The active users are organized into 458 components, where a component is defined as a group where all members have direct or indirect ties to all other members. 457 of these components are size 9 or less, with

the majority (414) being of size 2. The last component is a giant component consisting of the vast majority of active TSL users, or 28,023 users.

Typically, social network analyses of data sets of this size focus on one or several sampled components within the network data, however because the largest component comprises nearly all of the users in this network, we opted to further subdivide the data into meaningful, yet manageable parts. Using data on when accounts were created and when users last logged into their accounts, we divided the active users by calendar-year quarter, where a user was included in a quarter if his/her account was created before the quarter started, and still in use at the end of the quarter. Using this technique, a user may be included in multiple quarters (all the quarters where the account was in existence and in use), but would not be included in quarters before their accounts were created, or after their accounts were no longer being used. The result was 17 network "slices" representing the 17 full quarters for which we have data, ranging in size from one user to 3,256 total users, and 40 to 2,456 active users with at least one friend (see Figure 1).

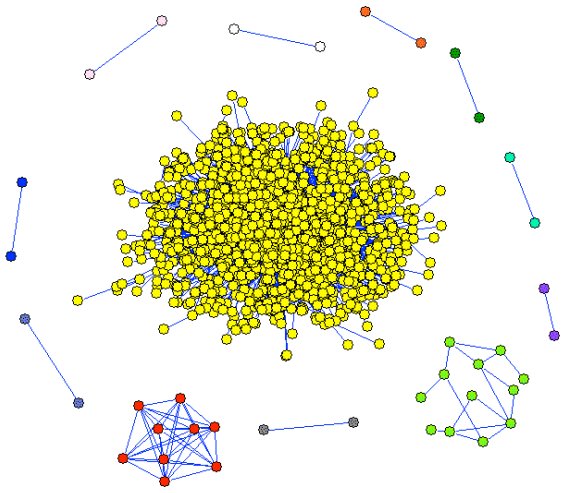
Figure 1. Number of Active Users in Each Quarter



In this paper, we perform network analysis on the most recent quarter, the 17th quarter (March-May 2007), which has 2,456 active users with 42,464 friendship relations. The

friendship network is well-connected: the largest component includes 2,417 active users, which is 98.41% of all active users; the rest active users cluster into 9 dyads, one group of 9, and another group of 12. Figure 2 shows all 12 components indicated by different colors.

Figure 2. Network of Active Users in the 17th Quarter



Methods. Unlike other types of data collected from individuals, relational data usually show strong interdependencies. For example, the existence of one friendship relation may depend on the total number of friends the focal node has, what other friends the focal node has and/or the relations between the existing friends and the potential friends. The interdependencies among relational data make it inappropriate to analyze network data using traditional statistical analysis approaches, such as regression analysis or correlation tests.

To test the hypotheses about the network without losing information about interdependencies among relations, we adopt a recently developed social network analysis approach Exponential Random Graph Models (ERGM). ERGM, also known as p^* models, can

be used to examine both endogenous and exogenous mechanisms that drive the creation, maintaining and dissolving of network ties (Contractor, Wasserman & Faust, 2006). The theoretical foundation and early work of ERGMs can be traced back to Frank and Strauss' work on Markov Graphs (Frank & Strauss, 1986). More recently, research by Wasserman, Pattison and Robins (Pattison & Wasserman, 1999; Robins, Pattison, & Wasserman, 1999; Wasserman & Pattison, 1996) made it an influential methodology in social network analysis.

Dependence assumptions of network data can be incorporated into ERGMs using different network statistics. The simplest dependence assumption is that all edges in the network are homogenous and independent of each other. The model under this assumption is also known as Bernoulli Model and the Erdős–Rényi Model (Erdős & Rényi, 1959). In this simplest case, ERGM only include the measure of edges as network statistics. When other network statistics are includes, such as reciprocal ties, triangles, node attributes or even high-order statistics (Robins, Snijders, Wang, Handcock, & Pattison, 2007) , more complex dependencies can be included in the model.

Software and programs, such as PNet, SIENA and Statnet, have been developed to assist analysis using ERGMs. In this study, StatNet is used for network data analysis, which is developed by the research group at University of Washington (Goodreau, Handcock, Hunter, Butts, & Morris, 2008). The reason to choose StatNet is that this suite of R packages enables us to analyze large scale network dataset with 2,000+ nodes. In this analysis, we used StatNet to estimate and test the following model for our hypotheses.

Measures. In order to test our hypotheses using StatNet, we needed to develop a model for our network. We built our model based on a similar analysis that was done on the data from the National Longitudinal Study of Adolescent Health, more commonly known as the Add Health study of Faux Magnolia High. The data in that sample was collected from over 90,000 US

students to study relationships, and contained measures similar to those we are interested in here, including age attributes and measures of balance in friendships (Goodreau et al., 2008). The specific nodal attributes, dyadic covariates, and network effects we include in our model were:

- *Edges*: We use the network statistic edges, which equals to the number of edges in the network, to control the density of the network and as a basic test of Hypothesis 1. That is, friendships are not created at random.
- *Same_city*: To test Hypothesis 2 on *geographic proximity*, we add into the model one independent variable "Same City" which counts the number of edges (i, j) for which users i and j come from the same city. A positive coefficient of this variable predicts the tendency of users from the same city creating ties with each other, and a negative coefficient predicts otherwise.
- *Usage_category*: To test *digital proximity* in Hypothesis 3, we use the lifetime usage of users to study the impact of user's online time on friendship relations. We classify users into five categories according to the logarithm of their usage minutes to the base 10. For example, the users in Category i has a total usage time between 10^i and 10^{i+1} -1 minutes in TSL. Hypothesis 3 predicts that users with a higher Usage_category are more likely to make friends.
- *Different_registration_date*: Variable "Different_registration_date" is included into the model to test Hypothesis 4 on *temporal proximity*. Different_registration_dateij equals to the absolute difference of user i and user j's registration time (measured by day). The variable was taken as an edge covariance factor in the model. Positive coefficient of this variable means higher tendency of making friends between users with large difference in registration time. Negative coefficient means higher tendency of observing friendship ties between users with small difference in registration time, i.e. users who registered at similar times.

- *Different_age*: When registered in TSL, users were asked to report their date of birth. Based on this information, we construct a dyadic variable Different_age$_{ij}$, which equals to the absolute year difference between users i and j's ages. This measure is used to test the *age homophily* in Hypothesis 5.
- *GWESP*: Hypothesis 6 predicts balanced structure in the friendship network. To test this hypothesis, one curved linear statistics "gwesp" (Hunter, 2007) is included in the model. GWESP stands for geometrically weighted edge shared partners. This statistics counts the number of edgewise shared partners in the network and combine them into one single variable by giving counts different weights (Hunter, 2007). Positive coefficient of this variable indicates that the tendency of triangle formation in the network is higher than expectated at random. Negative coefficient indicates lower tendency than what would be expected at random. .
- *Account_type*: We use account types to measure *user status* in Hypothesis 7. There are four types of accounts in the 17th quarter: basic, premium, non-revenue, and unknown (coded as 1, 2, 8, and 0 respectively). Basic accounts are free but cannot own land in TSL; Subscription-based premium accounts have more features such as monthly stipends, free land quota, and building objects on private lands; non-revenue accounts are for teachers and Linden Liaisons, Linden Lab's staff who are there to make Teen Second Life a safe and pleasant place to be. Users with premium and non-revenue accounts have more resources and higher status in TSL. Based on Hypothesis 7, they should be more likely to establish friendship with others.

6. Results

Descriptive Statistics. Among 2,456 active users in the 17th quarter, 28.9% users reported to be female upon registration and the rest 71.1% reported male. Four types of accounts appeared

in the 17th quarter: 92.3% basic accounts, 3.87% premium, 0.45% non-revenue accounts, and 3.78% unknown (coded as 1, 2, 8, and 0 respectively).

Even though the age of users in the 17th quarter ranges from 13 to 60, 95.15% users are less than 18 years old. The rest 4.85% are 119 adults (18+) who are either Linden Lab employees, or other adults who volunteered to assist in TSL. All of the active accounts in the 17th quarter were created between May 8, 2004 (second month in the 6th quarter) and March 20, 2007 (first day of 17th quarter). 85.06% of the accounts were created in the past year (13th to 16th quarter).

The lifetime usage of the TSL users ranges from the minimum of 36 minutes (0.60 hour) to the maximum of 168,727 minutes (2812.12 hours), with mean 13,700 minutes and standard deviation 29,915 minutes. The distribution of total usage minutes is highly skewed to the left. It indicates that there are a large amount of users who played in TSL for only a short period of time and a small amount of users who stayed in TSL for a long time.

The physical locations of TSL users in the 17th quarter are distributed across 48 countries, among which the United States has 79.8% of TSL users followed by United Kingdom (6.8%), Canada (4.8%) and Australia (1.4%). The rest 7.2% come from 44 other countries. 76.6% users provided city information and they came from 1143 unique cities. The city with the largest number of users is Los Angeles (23) followed by Chicago (15) and Miami (13). The majority of cities (72.4%) only have one player in TSL.

Network Analysis Results. In the network analysis, we include two network variables, two nodal variables, and three dyadic variables in an ERGM model. For *Account_type*, we take account types 0 and 1 as the base, which have the majority of users and test the relative tendencies of high status account types compared to the baseline population. Similarly for *Usage_category*, we use usage categories 1 and 2 as the base and study more active users

compared to users who stay less than one thousand minutes in TSL. The results of the estimation are reported in Table 1.

Table 1. ERGM estimation results for TSL Friendship Network

Variables	Friendship network (N=2456)	
	coefficients	(S.E.)
Network effects		
Edges *(H1)*	-8.454***	(.017)
GWESP(0.2) *(H6)*	1.058***	(.013)
Main effects of Nodal Variables		
Account_type=2 *(H7)*	0.071***	(.010)
Account_type=8 *(H7)*	0.657***	(.042)
Usage_category=3 *(H3)*	1.177***	(.005)
Usage_category=4 *(H3)*	2.094***	(.012)
Usage_category=5 *(H3)*	2.518***	(.015)
Dyadic variables		
Same_city *(H2)*	2.259***	(.060)
Different_age *(H5)*	0.0006	(.001)
Different_registration_date *(H4)*	-0.005***	(.0001)

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

The strongest effects were observed for all variables except *Different_age*. The negative coefficient of *Edges* supports Hypothesis 1 and indicates the friendship network is very sparse and adolescents are not likely to establish friendship randomly. The strong positive effect of GWESP suggests that two users are more likely to become friends if they have common friends. Therefore, the balance theory posited in Hypothesis 6 is supported.

Compared to basic accounts and a few unknown accounts, both premium and non-revenue accounts have higher tendencies to form friendship. This is in accord with Hypothesis 7, which proposed that higher-status individuals, such as paid subscribers and Linden Liaisons, are more likely to form friendships than lower-status individuals with basic account. However the results also indicate that non-revenue accounts have much higher likelihood than premium accounts. That is, non-revenue accounts, typically Linden lab liaisons, were considered of even higher status than premium accounts

The coefficients of *Usage_category* are positive and larger for higher categories. Users with thousands, tens of thousands, and hundreds of thousands online minutes are $e^{1.177}=3$, $e^{2.094}=8$, and $e^{2.518}=12$ times more likely to form friendships than the users who have less than one thousand minutes online time. This supports Hypothesis 3 on digital proximity which suggests users with more usage will make more friends.

The strong positive effect of *Same_city* suggests that a friend relation is almost 10 times more likely to be observed in a pair of users from the same city than in a pair of users from different cities. Consistent with the temporal proximity argument in Hypothesis 4, the negative coefficient of *Different_registration_date* suggests users registered during similar time points are more likely to become friends. However, the magnitude of the impact is very small. Finally, the effect of age homophily (Hypothesis 5) is not significant. That is users of the same age were no more likely to form friendships after controlling for the remaining hypotheses in the model

Goodness of Fit. Figures 3 and 4 show the goodness-of-fit for the degree distribution, the distribution of edgewise shared partners, and the geodesic distance. The degree distribution generated from the model fits well with the observed network. The model captured the basic shape of the edgewise shared partner distribution of the observed network. However the model overestimates the number of edges with less than four partners and underestimates the number of edges with four or more partners. Geodesic distance distribution is also captured by the model, with the distances larger than four slightly underestimated. Given geodesic distance is a global measurement and was not directly included in the model, we consider the model to be a good fit of the observed data. To sum it up, our model well captured the structures of the observed network.

Figure 3. Goodness-of-fit Diagnostics for Degree Distribution

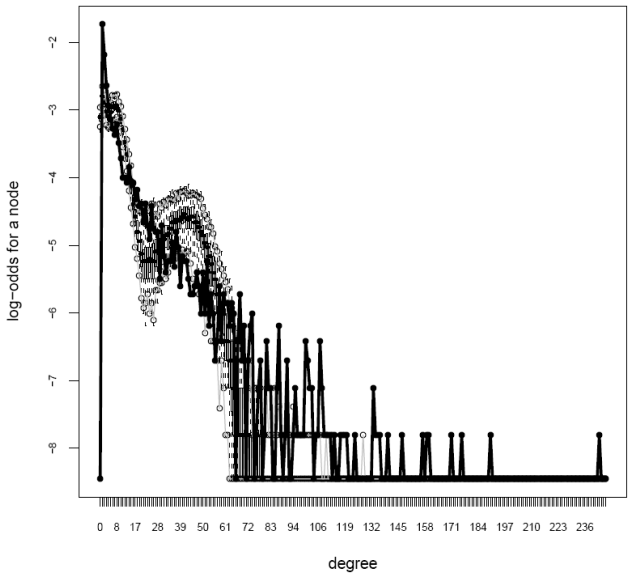
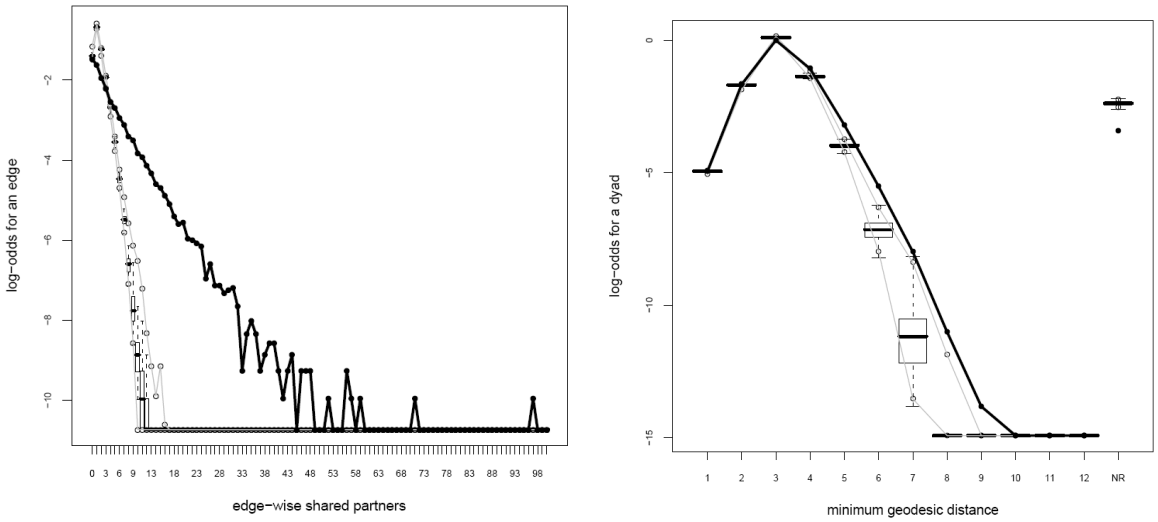


Figure 4. Goodness-of-fit Diagnostics for edge-wise Shared Partners and Distance



7. Discussion and Conclusion

The popular media might have us believe that teens make haphazard decisions about who to befriend online. However, our data demonstrate that this is not the case. Instead, adolescents appear to be following predictable patterns as they form online friendships. Much like the offline world, teens appear to be making thoughtful and predictable decisions about who to befriend online. The results reported here indicate that friendships are driven by three dimensions of proximity: geographic, digital and temporal. Teens are significantly more likely to forge friendship ties with others who are physically collocated (geographically proximate); they are more likely to seek friendship with others who spend a large amount of time online (digital proximity); and they are more likely to forge ties with others who joined TSL at the same time as them (temporal proximity). In addition, the results reported here indicate that teens sought friendship ties with others of higher status and with friends of their friends. After controlling for these factors, there was no additional evidence that they sought friends of the same age.

Taken together, these results indicate that the use of virtual worlds such as TSL, appear to augment and reinforce existing patterns of friendship formation. These findings challenge the received view in the popular press that new virtual world technologies disrupt (often deleteriously) patterns of friendship formation in the offline world.

In addition to the substantive insights, this paper also showcases significant contributions in instrumentation and analytics. It is one of the first studies in the communication discipline that used computer logs to test models of friendship formation in large scale virtual worlds. Previous studies in both the offline and online world have relied heavily on self-report data on friendships. As is often acknowledged in this literature, these data are especially error prone when they are solicited from children. From an instrumentation standpoint, the ability to capture friendship

networks from computer generated logs is both a more reliable and a more scalable approach, allowing for the investigation of networks involving millions of potential friendship ties among 1000s of individuals.

Methodologically, the study also showcases one of the first efforts in the field of communication research to use advanced network analytic techniques to test the hypothesized structures leading to friendship formation. The vast majority of network analytic research tends to be descriptive – providing insights into characteristics such as the density and centralization of the network. Until recently, there was no statistically defensible approach to conduct confirmatory network analysis of the type illustrated in this study. This is because network data violate assumptions of independence that form the bedrock of traditional statistical techniques. Recently, the emergence of p^* or exponential random graph modeling techniques have ushered in the possibility of testing for the presence of hypothesized structures in networks using techniques that explicitly model the dependencies among the ties. This is one of the first studies that have applied those techniques to large-scale networks in the field of communication.

Despite, and perhaps because of its groundbreaking contributions, the findings of the study must be interpreted with caution and calls for additional research. Teen Second Life is just one online world, with a specialized user base and a specific set of rules and social customs. Clearly, generalizing to all virtual worlds is not warranted. In order to be able to make knowledge claims on online friendship patterns more broadly, it will be important to conduct similar investigations in other virtual worlds with different rules and mechanisms , including those with much larger user bases such as World of Warcraft, Club Penguin, or Habbo Hotel. Indeed a future area of study is to hypothesize differences in the mechanisms of friendship formation that can be derived from variation in the contexts of these virtual worlds. For instance, the rules driving friendship formation in TSL where bonding is a primary goal, are likely to be

different from Massively Multiplayer Onliner (MMOs) games where banding together to go on quests is a primary goal. The difference between bonding (in the case of TSL) and banding (in the case of MMOs) suggest different theoretical mechanisms for friendship formation.

There are several theoretical insights to be derived from this study and the ensuing research program. First it addresses a perennial debate associated with the introduction of new technologies in society: how are these similar to, or different from, prior socio-technical configurations? Second, it serves to advance our basic understanding of fundamental social processes of friendship formation by considering specific technological features of the virtual world as experimental and design manipulations that provide more insight about baseline mechanisms such as proximity.

Likewise, the study also has important practical implications for those involved in *developing* these virtual worlds (what features in the virtual worlds facilitate or inhibit friendship formation), *marketing* these virtual worlds (what friendship structures are more likely to retain users within the world or what friendship structures signal a "problem" user), and policy makers who must decide what regulatory policies are justified based on empirical evidence and what are the consequences of potential policy changes.

References

Bigelow, B. J., & La Gaipa, J. J. (1980). The Development of Friendship Values and Choice. In H. C. Foot, A. J. Chapman & J. R. Smith (Eds.), *Friendship and Social Relations in Children*. New York, NY: John Wiley & Sons.

Brown, B. B. (2004). Adolescents' Relationships With Peers. In R. M. Lerner & L. Steinberg (Eds.), *Handbook of Adolescent Psychology* (Second ed.). Hoboken, New Jersey: John Wiley & Sons, Inc

Cairns, R. B., & Cairns, B. D. (1994). *Lifelines and Risks: Pathways of Youth in Our Time*. Cambridge, England: Cambridge University Press.

Chan, D. K., & Cheng, G. H. (2004). Comparison of Offline and Online Friendship Qualities at Different Stages of Relationship Development. *Journal of Social and Personal Relationships*, *21*(3), 305-320.

Contractor, N., Wasserman, S., & Faust, K. (2006). Testing multi-theoretical multilevel hypotheses about organizational networks: An analytic framework and empirical example. *Academy of Management Review*, 31, 681-703.

Epstein, J. L. (1983). Examining Theories of Adolescent Friendships. In J. L. Epsten (Ed.), *Friends in School: Patterns of Selection and Influence in Secondary Schools*. New York, NY: Academic Press.

Erdős, P., & Rényi, A. (1959). On Random Graphs. I. *Publicationes Mathematicae*, *6*, 290-297.

Espelage, D. L., Holt, M. L., & Henkel, R. R. (2003). Examination of Peer-Group Contextual Effects on Aggression During Early Adolescence. *Child Development*, *74*(1), 205-220.

Frank, O., & Strauss, D. (1986). Markov Graphs. *Journal of the American Statistical Association*, *81*(395), 832-842. http://sonic.northwestern.edu/digital-readings/frank&strauss_1986.pdf

Gest, S. D., Moody, J., & Rulison, K. L. (2005). Density or Distinction? The Roles of Data Structure and Group Detection Methods in Describing Adolescent Peer Goups. *Journal of Social Structure*, *6*. http://www.cmu.edu/joss/content/articles/volume8/GestMoody/

Goodreau, S., Handcock, M., Hunter, D., Butts, C., & Morris, M. (2008). A statnet Tutorial. *Journal of Statistical Software*, *24(8)*. http://www.jstatsoft.org/v24/i08/

Hardy, C. L., Bukowski, W. M., & Sippola, L. K. (2002). Stability and Change in Peer Relationships During the Transition to Middle-Level School. *Journal of Early Adolescence*, *22*, 117-142.

Hartup, W. W. (1993). Adolescents and Their Friends. *New Directions for Child Development*, *60*, 3-22.

Homans, G. C. (1974). *Social Behavior: Its Elementary Forms*. New York: Harcourt.

Hunter, D. R. (2007). Curved exponential family models for social networks. *Social Networks*, *29*(2), 216-230.

Inhelder, B., & Piaget, J. (1958). *The Growth of Logical Thinking*. New York: Basic Books.

Kandel, D. B. (1978). Homophily, Selection, and Socialization in Adolescent Friendships. *The American Journal of Sociology*, *84*(2), 427-436.

Lehnhart, A., & Madden, M. (2007). *Teens, Privacy & Online Social Networks.* Washington, D.C.: Pew Internet and American Life Project.

Macgill, A. R. (2007). *Parent and Teen Internet Use.* Washington, D. C. : Pew Internet and American Life Project.

Mannarino, A. P. (1980). The Development of Children's Friendships. In H. C. Foot, A. J. Chapman & J. R. Smith (Eds.), *Friendship and Social Relations in Children*. New York, NY: John Wiley & Sons.

Mesch, G. S., & Talmud, I. (2007). Similarity and the Quality of Online and Offline Social Relationships Among Adolescents in Israel. *Journal of Research on Adolescence*, *17*(2), 455-466.

Monge, P. S., & Contractor, N. S. (2003). *Theories of Communication Networks*. New York: Oxford University Press.

Moody, J. (2001). Peer Influence Groups: Identifying dense Clusters in Large Networks. *Social Networks*, *23*, 261-283.

Neckerman, H. J. (1996). The Stability of Social Groups in Childhood and Adolescence: The Role of the Classroom Social Environment. *Social Development*, *5*(2), 131-145.

Peter, J., Valkenburg, P. M., & Schouten, A. (2005). Developing a Model of Adolescent Friendship Formation on the Internet. *CyberPsychology and Behavior*, *8*(5), 423-430.

Robins, G., Snijders, T., Wang, P., Handcock, M., & Pattison, P. (2007). Recent developments in exponential random graph (p*) models for social networks. *Social Networks*, *29*(2), 192-215. http://www.sciencedirect.com/science/article/B6VD1-4M69JM0-1/2/c9edb0da2dcdc21156ba684d6b0df9fc

Rubin, Z. (1980). *Children's Friendships* (J. Bruner, M. Cole & B. Lloyd, Ed.). The Developing Child. Cambridge, Massachusetts: Harvard University Press.

Savin-Williams, R. C., & Berndt, T. J. (1990). Friendship and Peer Relations. In S. S. Feldman & G. R. Elliott (Eds.), *At the Threshold: The Developing Adolescent*. Cambridge, Massachusetts: Harvard University Press.

Sproull, L., & Kiesler, S. (1986). Reducing Social Context Cues: Electronic Main in Organizational Communication. *Management Science*, *32*, 1492-1512.
Walther, J. B. (1995). Relational Aspects of Computer-Mediated Communication: Experimental Observations Over Time. *Organizational Science*, *6*(2), 52-90.
Williamson, D. A. (2007). *Kids and Teens: Virtual Worlds Open New Universe.* New York, NY: eMarketer.
Wolak, J., Finkelhor, D., Mitchell, K. J., & Ybarra, M. L. (2008). Online "Predators" and Their Victims: Myths, Realities, and Implications for Prevention and Treatment. *American Psychologist*, *63*(2), 111-128.