NETWORK TOURISTS

August, 2016 © Ronald S. Burt,
University of Chicago Booth School of Business, Chicago, IL 60637
Tel: 312-953-4089, ron.burt@chicagobooth.edu

ACKNOWLEDGEMENT: I am grateful to the University of Chicago Booth School of Business for financial support during the work reported here. I am grateful to Thomas Burt for editorial comments and to Linden Lab, especially Cory Ondrejka, for providing a slice of Second Life data, and grateful to Noshir Contractor's Sonic Lab for archiving the data in an accessible form. This manuscript is available at http://faculty.chicagobooth.edu/ronald.burt/research.

Network tourists are people present as temporary observers. Tourists are likely under two conditions: low barriers to entering a population, and widespread curiosity about the population. Curiosity attracts tourists and low barriers to entry make it easy for them to satisfy their curiosity. Failure to control for the limited engagement of tourists could distort social-capital effects on trust and achievement. The suspicion is discussed and illustrated with data on the virtual world, Second Life. In keeping with the suspicion, there are a great many tourists in the virtual world (about half of the residents), and tourists have the expected characteristics of low achievement, low social-capital-network scores, and low trust. However, suspicion of tourists is unwarranted, at least in Second Life. The strong empirical evidence of trust higher in closed networks and achievement higher in open networks is unaffected by controls for tourists.

The scenes displayed in Figure 1 are from Second Life, a virtual world of more than 20 million registered accounts and something like a million regular users.[1] People interact with one another through animated characters discussed as residents. After its official release in 2003, Second Life quickly became home to meeting and event facilities run by real-world businesses, churches, foundations, government agencies, universities, and entertainment of high and low sorts. Residents retain copyright to the inworld content they produce, so there are active markets for diverse services including architecture and construction, events and education, as well as more generic goods such as clothing and painless plastic surgery giving one's avatar an attractive new "skin." Figure 1 contains some screen captures to provide a sense of the virtual world. After registering, residents can attend a meeting or class, shop for inworld goods, wander the world, or meet, engage, and exchange with folks. There are directories, maps, and people to guide residents to places and activities. Residents can teleport between locations inworld, and fly over a location to quickly get a sense of what is going on. Look at the "shop" screen in Figure 1. The resident avatar is hovering at a height several billboards above the boardwalk. He was in flying mode at the time.

---

[1] Linden Lab, which owns and operates Second Life, posted a graphic celebrating the virtual world's tenth anniversary (http://www.lindenlab.com/releases/infographic-10-years-of-second-life): 36 million accounts had been created, 400,000 new accounts were created monthly, and more than a million visits occurred monthly. Many of the accounts were tourists who exited shortly after creating their accounts. Wagner James Au, a continuing Second Life participant and reporter (e.g., Au, 2008), said in his April 19, 2013 blog that he "had it on good authority that Second Life's actual active userbase is about 600,000."

I am the pictured resident in the white shirt. I saw great potential for network analysis in virtual worlds, and had been given access to network data on Second Life. I wanted to get a sense of the place. I took the Figure 1 screen shots during my first day inworld. I did not return until a year later when a meeting of our research team was scheduled inworld.

During my time inworld, I was what I shall call a “tourist,” a person present as a temporary observer. In a sense, tourist is a label for what was earlier termed a lurker. The negative label “lurker” emerged in the mid-1980s to refer to people who tie up a modem line reading bulletin boards without ever contributing, thereby degrading access for active contributors. Lurkers are numerous. Observing a selection of email discussion lists (DL) in the late 1990s, Nonnecke and Preece (2000) report that half of the people on health DLs were lurkers (in the sense of not posting during the three-month observation period) and four-fifths of the people on software DLs were lurkers. Lurkers do not create the access problems today that they did in the mid-1980s, so one can talk today about the virtues of having lurkers around to give a community variety and scale.

——— Figure 1 About Here ———

As I travelled the virtual world, however, I wondered about the effect that a large number of tourists like me could have on the social capital hypotheses I would be studying. Suppose you have considerable experience in Second Life and you meet someone like me at an event on my first day inworld. It turns out that most people inworld for their first day are tourists, here today and gone tomorrow, never to return. As an experienced resident, you know that first-day visitors are often tourists. Why would you strike up a friendship with me? Wouldn’t it be a better use of your time to invest in someone likely to be inworld again tomorrow? Or suppose you’re recruiting for a group you’re excited about. Why waste time recruiting me? You’d do better to recruit someone whose membership you earn today will continue to be a membership tomorrow. These questions, and others like them, lead me to expect tourists to have small or nonexistent networks, because imminent departure makes the tourist a poor investment. At the same time, the positive effect of network closure on trust and cooperation should be especially positive for tourists, precisely because they are a poor investment. If the new person is a friend of a friend, you are likely to help the new person get comfortable inworld. Tourists also have implications for the achievement associated with networks rich in opportunities to broker connections between disconnected others. Tourists have small or nonexistent networks and achieve little — which is consistent with the brokerage-achievement hypothesis. But the consistency has nothing to do with the hypothesis. Tourists are not limited in their achievement by a lack of access to brokerage opportunities. They are present to see the sights, not change them.

My concern about the impact network tourists on evidence of social capital effects led to this note. The next section introduces the research setting. The subsequent section uses the probability of exit to measure of the probability that a person is a tourist. The third section shows that the tourist indicator has the expected associations with trust, networks, and achievement, but does not eliminate the strong empirical evidence of social capital effects on trust and achievement — at least not in Second Life.

STUDY POPULATION

The Second Life data to be analyzed describe all six million residents at the time the data were downloaded. Employees at Linden Lab selected the download day (September 6, 2007). The virtual world had been running commercially for about five years when the data were downloaded. Residents are only identified in the data by Linden Lab's code numbers. Personal identities remain confidential. The downloaded data include a few self-reported attributes such as gender and age, along with some data on each resident's history in Second Life, a roster of each resident's friends in Second Life, and a roster of Second Life groups the resident founded or with which the resident was affiliated.

A Site Prone to Tourists

Tourists are likely under two conditions: low barriers to entering a population, and widespread curiosity about the population. Curiosity attracts tourists and low barriers to entry make it easy for them to satisfy their curiosity. By these criteria, Second Life was tourist-prone at the time of the data download. Media stories attracted people curious to see what all the excitement is about. In the first half of 2006, a new arrival had the company of 8,062 other new users arriving in the same week (three days before through three days after ego's arrival). In the second half of 2006, the number increased to 63,244 per week. Through the first nine months of 2007, the number increased to 138,758 new users per week. Many of the new entrants were no more than tourists who lasted for a day, but many stayed to become residents.

Second, barriers to entry were minimal. The boundary around an organization is obvious by employment. The boundary around a subscription service like World of Warcraft, or EverQuest II, or Weibo is somewhat similar because people have to buy a subscription to use the service. In contrast, Second Life at the time of the data download was easy to enter and required neither payment nor credit identity. A new user went to the secondlife.com website, downloaded the free viewer software, launched the software, and entered (White, 2008, is a general guide; Boellstorff, 2008, describes the anthropology of entry and enculturation). Once in, the user selected a name and initial character that would be their resident avatar, and provided a few bits of demographic data. The porous boundary around Second Life made it easy for new people to show up, and existing people to disappear, which could made it difficult to trust people to meet obligations to which they'd agreed, or coordinate sustained group activity. Holding people to coordination commitments is a generic difficulty for people leading project work in the real world, and defining a boundary around any network can be problematic (Laumann, Marsden, and Prensky, 1989), but a porous boundary, such as the one around Second Life, exacerbates the generic difficulties.

A Virtual World in Which Social Capital Matters

Second Life is also a fruitful research site for the purposes of this note because there is clear evidence of social capital effects among residents. When residents agree to a friendship, each defines a level of privileges granted to the other person (which can change over time). The strength of the relation from one resident to another, ego and alter respectively, can be scaled from the depth of privileges granted. Details are provided elsewhere in a technical report (Burt, 2011), but the four distinct levels of friendship in Second Life are 0.00 for no friendship, .28 for a friendship without privileges or limited to allowing alter to know when ego is inworld, .63 for ego allowing alter to know where ego is currently located inworld, and 1.00 for ego giving alter complete access to ego's resources ("modify objects" relationship).

The left graph in Figure 2 shows evidence of the closure-trust hypothesis: Ego grants a higher level of privileges to alter when the relationship between ego and alter is embedded in a more closed network of mutual friends.

——— Figure 2 About Here ———

The right graph shows evidence of the brokerage-achievement hypothesis: The residents who reach higher levels of achievement have more opportunities to broker connections across otherwise disconnected residents (see Burt, Kilduff, and Tasselli, 2013, on network constraint and other network measures of access to brokerage opportunities). Level of achievement is measured in Figure 2 by a z-score based on a resident's success in creating groups, that attract many members, and survive their initial founding (see Burt, 2011 for details).

OPERATIONALIZING TOURIST

Controlling for tourists is difficult because there is no universal indicator. Tourists in Second Life are unlikely to make the commitment of owning property, but many people active inworld do not own property. Tourists are unlike to have social attachments inworld, but some tourists enter with attachments because friends drag them inworld, or they enter with colleagues to fulfill a job or class obligation. Tourists have little experience inworld, and people new to Second Life are easily identified by their "newbie" face, hair, clothing, walk, and awkward fit to social situation (Boostrom, 2008). On the other hand, every experienced resident was once new to the virtual world.

I distinguish tourists by their probability of exit. The closer a person is to exit, the more likely they will behave and be treated like a tourist — no commitments, no engagement, just passing through. There is no rite of passage indicating exit from

Second Life. All I know is the first and last day on which a resident did some activity inworld (for 99.7% of the residents).

First Few Days Are Critical

I assume that days between first and last day were continuous affiliation, and the last day was an exit. The two assumptions are less problematic here than they could be elsewhere.

The problem with assuming continuous affiliation is that the person could have exited at any time then checked back for something. Time-stamped data would allow exit to be defined by the end of continuous activity. I do not have time-stamped activity data from Second Life, but I am not worried about continuous affiliation because exit is so concentrated around entry. The concentration is illustrated in Figure 3. Of the residents for whom I know first and last day, 46.35% made their first day their last. That is the exit rate indicated in Figure 3 by the bold line over day one. Of the residents who continued past their first day into the second, 16.43% made day two their last day, which defines the exit rate indicated by the bold line over day two. Exit quickly becomes less likely with further experience, 9.68% exit on the third day, 5.75% for the rest of the first week, down to 2.68% through the next two weeks, and down to 1.47% through the subsequent two weeks.

——— Figure 3 About Here ———

I measure exit probability with an event-history model. The first and last activity dates generate 62,803,432 person-day observations for a model of each person's first 35 days inworld. A person was at risk of exit each day she was affiliated with Second Life. She is counted as an exit, and thereafter dropped from the data, on her last day inworld.

A problem with treating last day as an exit is that active users have last days close to the day the data were downloaded. A person whose last day was yesterday could very well be back tomorrow. There is a high risk that her last day was not an exit. The risk decreases with the length of time between last day and the data download. I would be confident that last day does indicate exit for someone whose last day was two years ago.

I am not worried about treating last day as an exit for two reasons: Again, exit is concentrated in the time just after entry, so most exits are quickly detected. Second, to minimize the error of coding continuing people as exits when their last day was close to the data download, I censor the event-history data five weeks before the data download, which Figure 3 shows is well after most exits happened.[2]

Table 1 contains two event-history models predicting exit. The first is computed from the complete data. The second is computed from the data censored by ignoring observations during the last 35 days before the data download. Predictors are the same in the two models, and coefficients are very similar except during the bandwagon period (when exit was most ambiguous and therefore censored). The table is based on population data, so routine test statistics are not useful in the usual way, but they are reported as a familiar guide to highlight strong associations.

——— Table 1 About Here ———

The data provide controls for self-reported gender, age, and geographic region. These user attributes are recorded when a person first enters Second Life, so there are some assumptions to using them as time-varying correlates. Gender is assumed time invariant. Age is incremented for time passed since first entry (person 26 years old at entry is 28 years old two years later). Without information on changes in residence, I assume that people remained in the same broad geographic region in which they registered (e.g., Asia, North America, Western Europe). There are missing data on age, gender, geographic region, and time spent inworld on a user's first day, so the number of person-day observations is reduced in Table 1.[3]

---

[2]Two kinds of observations are censored: All 2,337,921 person-day observations are censored on the 530,815 people who first entered Second Life after August 2, 2007. Second, for people who first entered Second Life within 35 days of August 2, all person-days after the 2nd are censored. For example, a person who first entered on July 31 and last entered on August 14 would contribute to the event-history data three person-days of being at risk of exit (July 31, August 1, and August 2). His subsequent days inworld would be censored (August 3 through August 14). The 62,803,432 person-days through the data download reduce to 58,666,034 in the censored data. Table 1 contains results including and excluding the censored observations.

[3]I do not have exit probabilities for people missing data on the predictors in Table 1. Of the initial 6,391,823 registered users analyzed in the text, 82,562 are missing one or more of the self-report data on age, gender, or geographic region. Another 234,370 are missing the minutes of time they spent inworld on the first day they entered the virtual world. Although data are only missing on five percent of the population, five percent is lot of people. To test for selection bias, I imputed the missing data where I could from friends of friends. Homophily is more likely between closer friends, so I began with closest

The first rows of Table 1 show the effects of increasing experience. Exit is less likely with each additional day spent inworld (-.878 logit coefficient). The models contain a continuous association with time, adjusted for the initial days inworld. I wanted to identify the day after which the early tendency for exit declined to a level consistent with a continuous function of time inworld. The second day seemed to be it. Controlling for the strong monotonic association with days inworld, exit is especially high during the first day, and to a lesser extent during the second day. Exit rates in the third and four days are about what would be expected from the monotonic association with days inworld.

A crude control for tourists would be to put aside observations in the first two days of a person's time inworld. However, at the same time that new entrants are disproportionately tourists, there are future long-term residents among those new arrivals, so I want to measure tourist in terms of more than time alone.

Kind of Person (Individual Tastes)

The middle section of Table 1 distinguishes kinds of people prone to exit. Exit rates were uniform across broad geographic regions, but younger people were more likely to exit, especially males. The probability of exit decreases as age increases, up to the mid-40s. Exit rates increase slightly in the retirement years, but rates bounce up and down from year to year for more senior people and age is self-reported by people known only by an avatar, so I do not read the age data too closely. An age association with exit is most evident from age 18 to the mid-40s, so that is where I focus. In Table 1, age enters as "years younger than 46" (e.g., a person age 30 is 16

---

friends, then moved to weaker friendships as needed. For example, the correlation between a person's age and the average age of his friends is .53 for modify-objects friends, .33 for locate friends, and .27 for weaker friendships. If a person's age was missing, I set his age to the average age of his modify-objects friends. If none of their ages were known, or if the person did not have modify-objects friends, I turned to progressively more distant friends. Imputation provided exit probabilities for an additional 20,008 residents and 162,511 additional friendships. Without imputation, ego-exit-probability is correlated -.024 and -.108 respectively with the level of ego's trust in alter and the number of mutual friends ego has with alter (left graph in Figure 2). With imputation, the correlations are -.026 and -.108. Without imputation, resident exit-probability is correlated -.126 and .203 respectively with network constraint and achievement and network constraint (right graph in Figure 2). With imputation, the correlations are -.129 and .205. Given the similarity of the respective correlations, I ignore in the text the additional observations available from imputing values for the missing data.

years younger than 46). The variable equals zero for ages over 46. I get the same results if I replace the criterion age 46 with ages that are one or two years younger or older. I leave the criterion at 46 because that age is the upper boundary of a "middle-age" status for Americans (Burt, 1991:19).

Gender is not associated with exit as clearly as age, but there is a discernable gender difference. Young males are about six percentage points more likely than females to exit. I tested for a gender-age interaction association with exit, but it is negligible relative to the direct associations with gender and age (2.47 z-score test statistic for the interaction of male and "years younger than 46," versus the 85.79 and 387.19 test statistics in Table 1 for direct associations).

Beyond age, gender, and region, there are unobserved characteristics of users that can be held constant with another time variable recorded by Linden Lab. People predisposed to activities inworld can be expected to spend more time when they first enter the world, and first-day-hours-inworld is strongly associated with exit in Table 1. People differed widely on this time variable. The average first day was two hours inworld. At the extremes are people who spent less than a minute, and others who spent more than eight hours.

Spending little time inworld on the first day is in some part a result of frustrations in figuring out what to do. Linden Lab made an effort to minimize entry frustrations, and so lengthen first time inworld, with features such as an "orientation island" set aside for new entrants, and designated "greeters" to ease new entry into areas of the virtual world. This frustration interpretation of short time on the first day is visible in the convergence of the solid and dashed thin lines in Figure 3. The lines show an exit-rate difference between people who spent little time inworld on their first day (dashed line) versus the people who spent a lot of time (solid line). The gap between the two lines is large for the first day (exit rates of 64.3% versus 6.2% respectively), much smaller on the second day (16.4% versus 10.1%), then disappears as people at either extreme become subject to the same inworld retention factors. The results in Table 1 support the frustration interpretation by the concentration of the first-day-hours-inworld exit effect in the first day. First-day-hours-inworld strongly inhibits exit during the first day

(-.761 logit coefficient, -513.92 z-score test statistic). By the second day, the exit association with first-day time is much weaker (-21.75 z-score).

At the same time, spending little time inworld on the first day reveals a personal preference. People with a taste for virtual world activity should be less likely to exit at any time. Table 1 shows a continuous exit association with first-day-hours-inworld: the more time spent inworld on the first day, the less likely a person will exit in subsequent days (-.115 logit coefficient, -287.68 z-score). Although the solid and dashed lines in Figure 3 converge with more time inworld, they never cross: The dashed line for residents who spent less than 30 minutes their first day is consistently higher than the solid line for residents who spent more than three hours. In other words, people who had a taste for Second Life found the virtual world interesting upon entry and so spent more time inworld on their first day. In this revealed-preference interpretation, the higher exit rates for people who spent little first-day time inworld resulted from a selection effect that removed people who did not enjoy Second Life, thereby clearing the virtual world for people who did enjoy it.

Entry Conditions (Operations and Crowding)

Exit is associated with the state of the virtual world when a person entered it. Second Life matured over time. Operations became more reliable and capable as the developers acquired experience. Crowding is also relevant. The thousands of people who registered with Second Life before 2004, became twice as many in 2004, three times that in 2005, and ten times that in 2006. More people make the world more interesting, but they also put more demand on hardware and support services, which slows response time, increasing the odds of exit. Popularity also brings a shift in the kind of people entering. The people first aware of a virtual world are likely to be people familiar with such things. Au (2008:41) quotes a person active in the early days of Second Life: “I think most of us were sci-fi fans, [and] we saw what looked like the beginning of the metaverse . . . we were the eternal Beta testers tilling the virtual land for future Residents to come in.” As a virtual world becomes more popular and discussed in the media, it attracts a more general population of people curious to see what the excitement is about. Whatever the proportion tourist in the early population, the proportion must be higher in the later populations of uninformed curious, who could be expected to be especially sensitive to newbie frustrations.

The bottom section of Table 1 shows the effects of operations and crowding. First, the probability of exit decreased as Second Life matured (-.711 logit coefficient for "Second Life age," which is the log of the years for which Second Life had been running when a person first entered).

Second, crowding varied across three periods in the history of Second Life. The periods are illustrated in Figure 4. The bold line shows the cumulative number of residents. The thin line shows the percent of person-days that ended in exit at each point in time, and the histogram below shows the number of person-day observations at each point in time. Three periods are distinguished. The first period begins with the beta launch of Second Life as "Linden World" in March, 2002 and ends with the official launch of Second Life on June 23, 2003. The number of users increased through this period from a few thousand to many thousands. The density of person-days is difficult to see in the graph at the bottom of Figure 4 because the millions who entered later dwarf the thousands of early users. What is visible in Figure 4 is the increasing likelihood of exit through the first period. As the number of early users increased, the probability of their exit increased.

——— Figure 4 About Here ———

Then a second period began with the official launch of Second Life, a period in which the user base grew steadily to hundreds of thousands, and the probability of exit decreased quickly to a stable, low level. An accumulation of person-days during the second period is visible in the graph at the bottom of Figure 4. In the graph at the top of Figure 4, the decreasing tendency for exit during the second period resembles a learning curve, implying collaborative effort at Linden Lab improving operations.

A third period began on June 1, 2006. It was a period of exponential growth that put a strain on operations, increasing the probability of exit. Two events marked the third period. First, Second Life was on the cover of the May 1, 2006 issue of Business Week. The cover article waxed enthusiastic about the possibilities of Second Life and people who had prospered from its opportunities. Second, Linden Lab made a policy change to open registration on June 28, 2006. Before the change, new entrants were required to give a credit card or other payment information when they first entered Second Life, even if they did not plan to make purchases inworld. The request was a deterrent to people merely curious about the virtual world. After the change, people just had to give a name and a few demographic characteristics. The national media attention and eased entry were coincident with a dramatic increase in the rate of new entrants to Second Life. Before June 2006, an average of 5,698 new people entered per week. After June, new entrants averaged 59,639 per week, increasing in the first nine months of 2007 to a peak of 190,583 during one week. The graph at the bottom of Figure 4 shows that most observations in the event-history data come from this third period (92.9%).

Judging from the probability of exit, tourists reached their largest proportion in the population during two intervals in the spread of Second Life — just after the official release, and during the bandwagon period. Figure 4 shows a spike in exits just after the official release in 2003. People were attracted to Second Life, but many of the new residents wandered around without being engaged, so they left. From a high of 29% just after the official release, the percentage of residents exiting decreased to a stable 5% through 2004 and 2005. The probability of exit never reaches 29% again, but the number of tourists increases. The 29% exit is based on thousands of residents. The 7% exit rate during the bandwagon period is based on millions of residents.

Table 1 contains level and slope adjustments for the periods in Second Life's development: (1) a pre-launch period, (2) a steady growth period following the official launch of Second Life in the middle of 2003, and (3) an exponential growth period following national media attention and relaxed entry requirements in the middle of 2006. The level adjustments drop exit to a low level at the beginning of the first and third periods.

The slope adjustments correct by period the learning-curve tendency for exit to become less likely with time. The learning-curve effect is concentrated in the second period of Second Life, a period of relatively stable growth (illustrated in Figure 4). Before and after the second period, exit becomes more likely with time, as Second Life became more crowded. In the pre-launch period, Linden Lab worked with beta users to prepare Second Life for official launch. The growing number of users put a strain on operations, increasing the odds of exit with time (combine the 3.414 slope adjustment in Table 1 for period one with the -.711 direct association with time to get a positive 2.703 period-one association with time). Similarly, exponential growth during the third period is associated with increasing exit (combine the 2.437 slope adjustment for period three with the -.711 direct association with time to get a positive 1.726 period-three association with time). The slope adjustment for period three is stronger in the uncensored data because last days near the data download are erroneously coded as exits for continuing users.[4,5]

---

[4]In fact, the variance explained by the mixture of level and slope adjustments for virtual-world-age in Table 1 can be obtained more simply by predicting exit from age and age squared. A negative exit association with age (-1.222) shows the learning-curve effect and a positive exit association with age squared (.164) shows the effect of crowding in period three. The problem with the age and age-squared model is that it is data fitting without explanation. The learning curve in the second period is detected. Crowding in the third period is detected. Crowding in the first period is dismissed because it is based on such a small percentage of the observations (though looking forward from 2002, those several thousand users must have seemed considerable). The model in Table 1 is algebraically more complicated than an age and age-squared model, but the model's substantive meaning is more obvious, facilitating replication in other virtual worlds.

[5]The number of people participating in Second Life grew so persistently that direct measures of crowding are too correlated with virtual-world age to enhance the prediction in Table 1. I computed three measures of crowding around ego at entry: the number of new entrants during the day that ego entered, the number of new entrants that week (people entering three days before ego through three days after ego), and the number of new entrants that month (fifteen days before or after ego). The below table gives the crowding correlations with exit and the virtual-world age variable in Table 1 (period three correlations above the diagonal, periods one and two below). The three measures of crowding are highly correlated with one another and with Second Life age, and none is strongly correlated with exit, so it is difficult to separate learning from crowding effects. The .330 pseudo R^2 in Table 1 does not increase if I add any of the three crowding measures to the prediction, and remains .330 if I replace Second Life age with the number of people who entered during the week that ego entered. Second Life age is slightly more associated with exit in the predictions, so age is the predictor included in Table 1. I also looked at exit rates by the day of the week and week of the year in which a person first entered the virtual world. Exit was no more or less likely for people entering on major holidays (in the US or elsewhere). Exit was slightly more likely among people who entered early in the week. Averaged across the person-day observations, exit is 8.30% among people who first entered on Monday versus 7.52% among people who first entered on Friday. The slight difference is negligible relative to the predictors in Table 1 (26.18 z-score test statistic when entry-day-of-the-week is added to Table 1).

Exit	——	.050	.019	.020	.018
Second Life Age (log)	-.032	——	.927	.726	.767
Number Same Day	.002	.671	——	.899	.852
Number Same Week	-.003	.720	.950	——	.954
Number Same Month	-.005	.752	.909	.972	——

Tourist Operationalized

I use the censored-data model in Table 1 to compute a probability of exit for each resident at the time they were observed for the data download. Days inworld longer than 35 days are coded as 35 (maximum days in Table 1 is 35). For a resident inworld for the first time on the data download day, the "First Day" dummy variable in Table 1 would be 1 and the dummy variables for days two, three, and four would be zero. For a resident inworld for longer than four days, all four dummy variables for days one, two, three, and four would be zero.

Figure 5 is a quantitative and qualitative display of tourists in Second Life. The exit probability most characteristic of the data — the mode — is near zero, as indicated by the concentration of scores to the left in the histogram at the top of Figure 5. Rows in Figure 5 show how the distribution is different for residents who spent different numbers of days inworld. For example, residents inworld for their first day are very likely tourists. Their exit probabilities vary, but around a high mean of .56. Residents inworld for a second day are much less likely to be tourists, but they too vary in their probabilities of exit. Exit becomes more concentrated around low exit probabilities for residents who have been inworld for more than four days. Exit probabilities have a bi-modal distribution with very few observations in a gap between the two modes. Within the gap, the exit probability with the fewest observations is .31. If everyone above a .31 cut-off is deemed a tourist, then the column to the right in Figure 5 shows that 43% of all residents are tourists, with 94% of the first-day residents tourists, 1% of the second day, and less than a percentage point after that.

——— Figure 5 About Here ———

TOURISTS IN THE SOCIAL CAPITAL PREDICTIONS

Tourists play their expected role in social capital, but they do not much affect the evidence of social capital effects in Second Life. Table 2 shows that, as expected, tourists are low on achievement, low on social-capital network variables, and low on trust. Reading down the rows of the table, tourists have friendships less-embedded in closed networks, tourists are not much different from non-tourists in terms of trusting friends, but they are much less likely to be trusted, and tourists have less access to brokerage opportunities and lower levels of achievement.

——— Table 2 About Here ———

Nevertheless, Table 3 shows that controlling for tourists has little effect on the association between closure and trust in Second Life. The first column contains a regression model predicting ego's level of trust in alter, where ego and alter are two residents connected by a friendship. The primary predictor is the number of mutual friends that ego and alter share. As illustrated in the graph to the left in Figure 2, the more mutual friends ego and alter share, the higher ego's level of trust in alter (106 t-test). Also as illustrated in Figure 2, the association is strongest for the initial mutual friends. The increase in trust associated with adding another mutual friend decreases at higher numbers of mutual friends (-57 t-test).

The model in the second column of Table 3 adds controls for ego and alter being tourists. As expected, tourists are unlikely to be trusted (-55 t-test), but the other test statistics are negligible: tourists are no less likely to trust (7 t-test), and the closure association with trust is little affected by ego or alter being a tourist (-1 and 5 t-tests respectively). More, adding controls for ego and alter being tourists has no effect on the overall prediction of trust (R^2 is identical for the two models), and has almost no effect on the coefficients for the network variables. The results in the third column of Table 3 use the continuous exit-probability measure of tourist. The column-three results support the same conclusions reached with the column-two binary distinction between tourist versus not tourist.

——— Table 3 and Table 4 About Here ———

Table 4 shows that controlling for tourists has little effect on the association between brokerage and achievement in Second Life. The model in the first column of the table shows that achievement is associated with the number of hours a resident has spent inworld (1113 t-test), and with having a friendship network rich in structural holes which gives the resident numerous opportunities to broker connections between otherwise disconnected friends (476 t-test). There is relatively little achievement association with the other predictors. Adding level and slope adjustments for tourists has the expected associations with achievement — tourists achieve less (-29 t-test) and their achievement is unaffected by access to brokerage opportunities (.015 coefficient for non-tourists, adjusted by -.015 for tourists, yields a coefficient of .000 for the tourist association between brokerage and achievement). As with the closure-trust association, adding controls for tourist residents does not improve the achievement prediction (R^2 is .32 in the first and second columns of Table 4), does not change the regression coefficient predicting achievement from access to brokerage opportunities (.015 in first row of first and second columns), and similar results are obtained when the binary distinction between tourist and not tourist is replaced in the third column of Table 4 by the continuous exit-probability measure of tourist.

CONCLUSIONS

Network tourists are people present as temporary observers. I suspected that failure to control for the limited engagement of tourists could distort social-capital effects on trust and achievement. The suspicion is discussed and illustrated with data on the virtual world, Second Life. In keeping with the suspicion, there are a great many tourists in the virtual world (about half of the residents), and tourists have the expected characteristics of low achievement, low social-capital-network scores, and low trust. However, my suspicion of tourists seems unwarranted, at least in Second Life. The strong empirical evidence of trust higher in closed networks and achievement higher in open networks is unaffected by controls for tourists.

I walk away with a conclusion and a caution. I conclude that it would be wise to test for tourist distortion of social capital evidence if the study population has low barriers to entry and there is wide-spread curiosity about the population. Tourists can exist in large numbers and they have their expected achievement, network, and trust characteristics. However, until future research reveals significant distortion, one need not worry about tourists distorting the basic social capital evidence of trust more likely within closed networks or achievement more likely with more access to brokerage opportunities.

The caution is that a key predictor is missing from the definition of tourists here. Social attachments are missing from Table 1. This is not to say that tourists can be identified by their lack of social attachments. Some tourists form social attachments. Some long-term Second Life residents do not. On average, however, exit is correlated with social attachments in that users more active socially in other virtual worlds are less likely to exit (Kawale, Pal, and Srivastava, 2009). In Second Life too, social attachments are correlated with exit. If I add number of friends to the predictors in Table 1, I get a -.245 logit coefficient for number of friends (-344.41 test statistic), which shows that exit is less likely among people with friends. If I replace friends with number of group affiliations, I get a -.659 coefficient and -407.09 test statistic, which shows that exit is less likely among people affiliated with groups.

However, some proportion of the correlated social activity arose after tourists had exited. My social attachment data are cumulative to the date of the data download. People who stayed inworld longer had more opportunities to make friends and join groups. Consistent with attachments accumulating over time, the number of days a person spent in Second Life is correlated with the cumulative number of her friends (.34) and the cumulative number of groups with which she was affiliated (.45). People who had spent less than a week inworld averaged less than one friend and no group affiliations (means of .34 friends, .07 groups). People inworld for longer than five weeks averaged eleven friends and three group affiliations (means of 10.6 and 2.8 respectively).

Further, social attachments are most associated with exit near the time a person first entered Second Life. A zero-order logit coefficient of -.427 predicting exit from number of friends is a stronger -.715 when the coefficient is estimated from event-history data on the first two days of time inworld, -.22 if estimated from days 3 through 7, and -.13 if estimated from exit rates in the second through the fifth weeks inworld. An exit association with group affiliations is similarly concentrated around entry: a -1.083 logit coefficient predicting exit from number of group affiliations is -1.564 for days 1-2, -.676 for days 3-7, and -.427 for days 8-35.

It would be illogical to explain early tourist exits by the later social activity of residents who continued inworld, so I interpret the above exit associations with social attachments as exit-prevents-network effects rather than network-inhibits-exit effects. To be interpreted as a network-inhibits-exit effect, the associations between exit and social attachments need to be estimated from time-stamped data with which social attachments can be measured as they precede exit. I do not have such data on Second Life. So, to keep the tourist control variables separate from my network predictors in Tables 3 and 4, I exclude social attachments from the predictors in Table 1. Future research needs to estimate tourist effects including time-stamped network data that measure a person's social attachments on the day of exit. The negligible distortion effect of tourists on the social capital evidence in Tables 3 and 4 might be stronger if the social attachments of tourists were included in the Table 1 definition of a tourist.

REFERENCES

Au, Wagner James. 2008. The Making of Second Life. New York: HarperCollins.

Boellstorff, Tom. 2008. Coming of Age in Second Life. Princeton, NJ: Princeton University Press.

Boostrom, Robert. 2008. "The social construction of virtual reality and the stigmatized identity of the newbie." Journal of Virtual Worlds Research 1(2), http://jvwresearch.org/index.php?_cms=default,4,1.

Burt, Ronald S. 1991. "Measuring age as a structural concept." Social Networks 13:1-34.

Burt, Ronald S. 2011. "Structural holes in a virtual world." Working Paper, University of Chicago.

Burt, Ronald S., Martin Kilduff, and Stefano Tasselli. 2013. "Social network analysis, through a focus on network advantage." Annual Review of Psychology 64: 527-547.

Kawale, Jaya, A. Pal, and Jaideep Srivastava. 2009. "Churn prediction in MMORPGs: a social influence based approach." Paper presented at the SIN09 Workshop, in conjunction with the IEEE Social Computing Conference, Vancouver, BC.

Laumann, Edward O., Peter V. Marsden, and David Prensky. 1989. "The Boundary Specification Problem in Network Analysis." Pp. 61-87 in Research Methods in Social Network Analysis, edited by Linton C. Freeman, Douglas R. White, and A. Kimball Romney. Fairfax, VA: George Mason University Press.

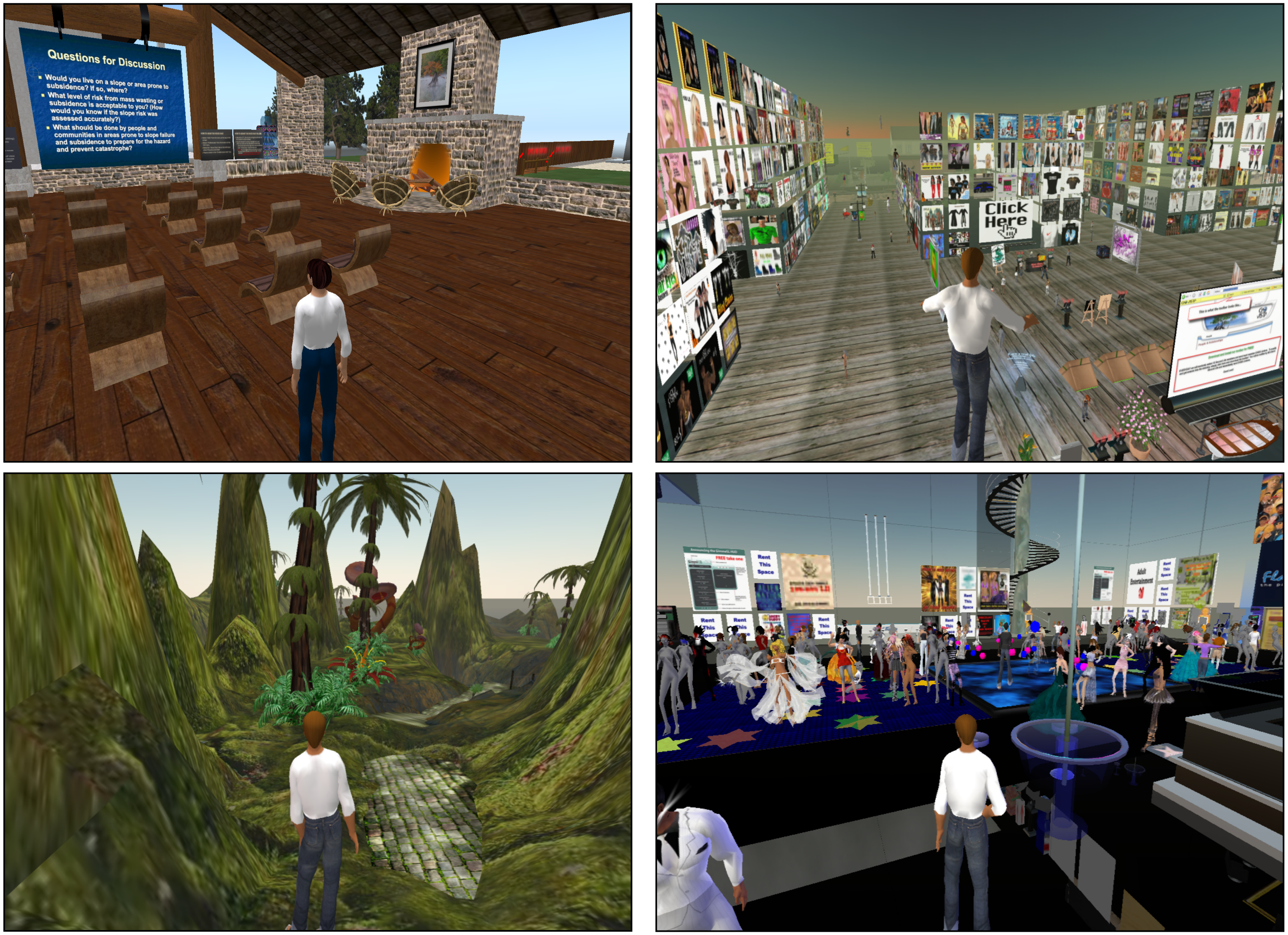
Nonnecke, Blair and Jenny Preece. 2000. “Lurker demographics: counting the silent.” Proceedings of CHI 2000 (Conference on Computer-Human Interaction), Association for Computing Machinery (ACM). The Hague, Netherlands.

White, Brian A. 2008. Second Life. Indianapolis, IN: Que.

Figure 1. Scenes in Second Life

(February 2008, just after the data download)

Attend a Meeting or Class | **Shop**



Wander the Earth | **Meet, Engage, and Exchange with Folks**

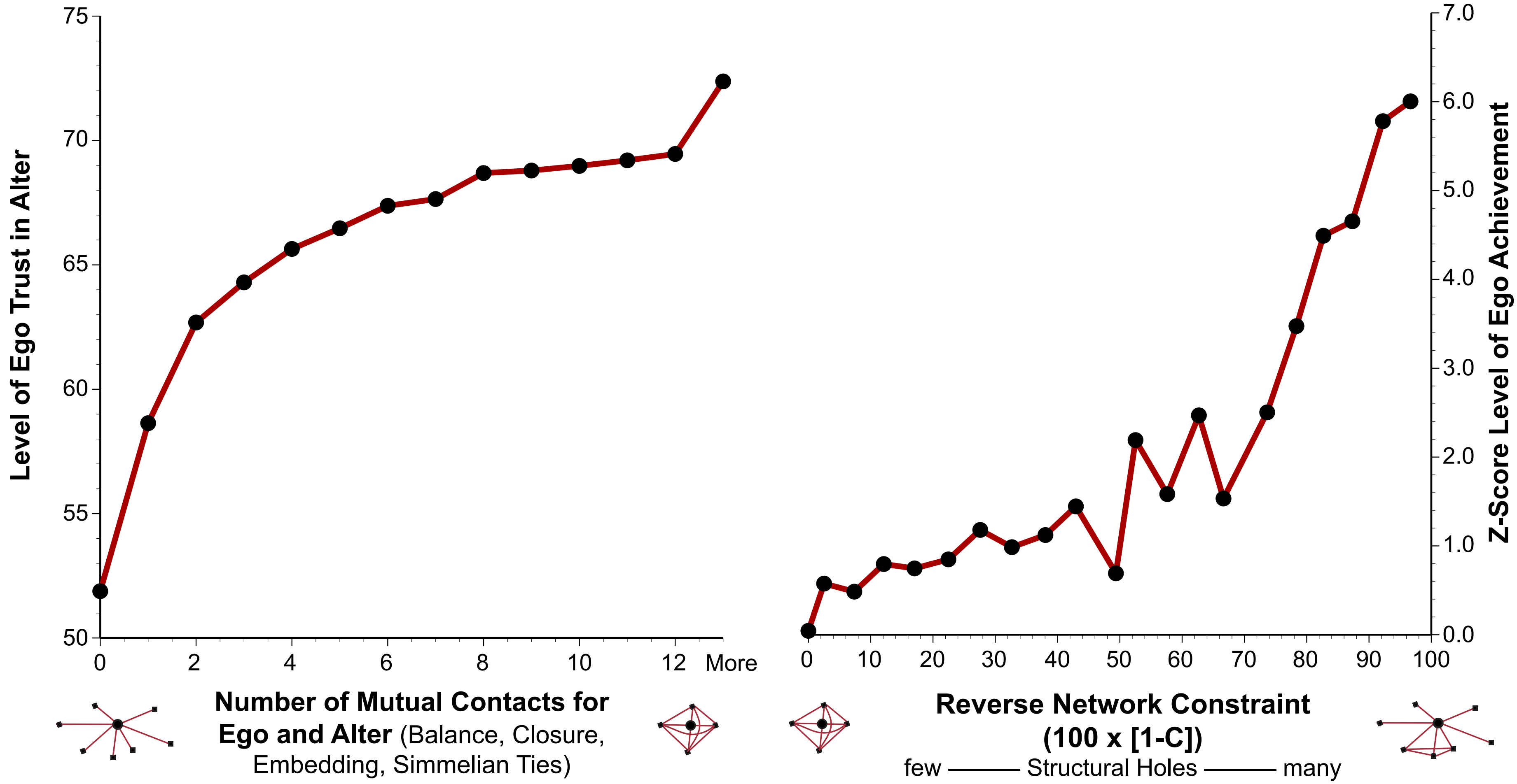


Figure 2. Basic Social Capital Results in Second Life

NOTE — Dots are average Y scores within intervals of X. Closure-Trust is graphed to the left. Trust is friendship rights ego grants to alter. Brokerage-Achievement is graphed to the right. Achievement is a z-score based on a resident's success in creating groups, that attract many members, and survive their initial founding.

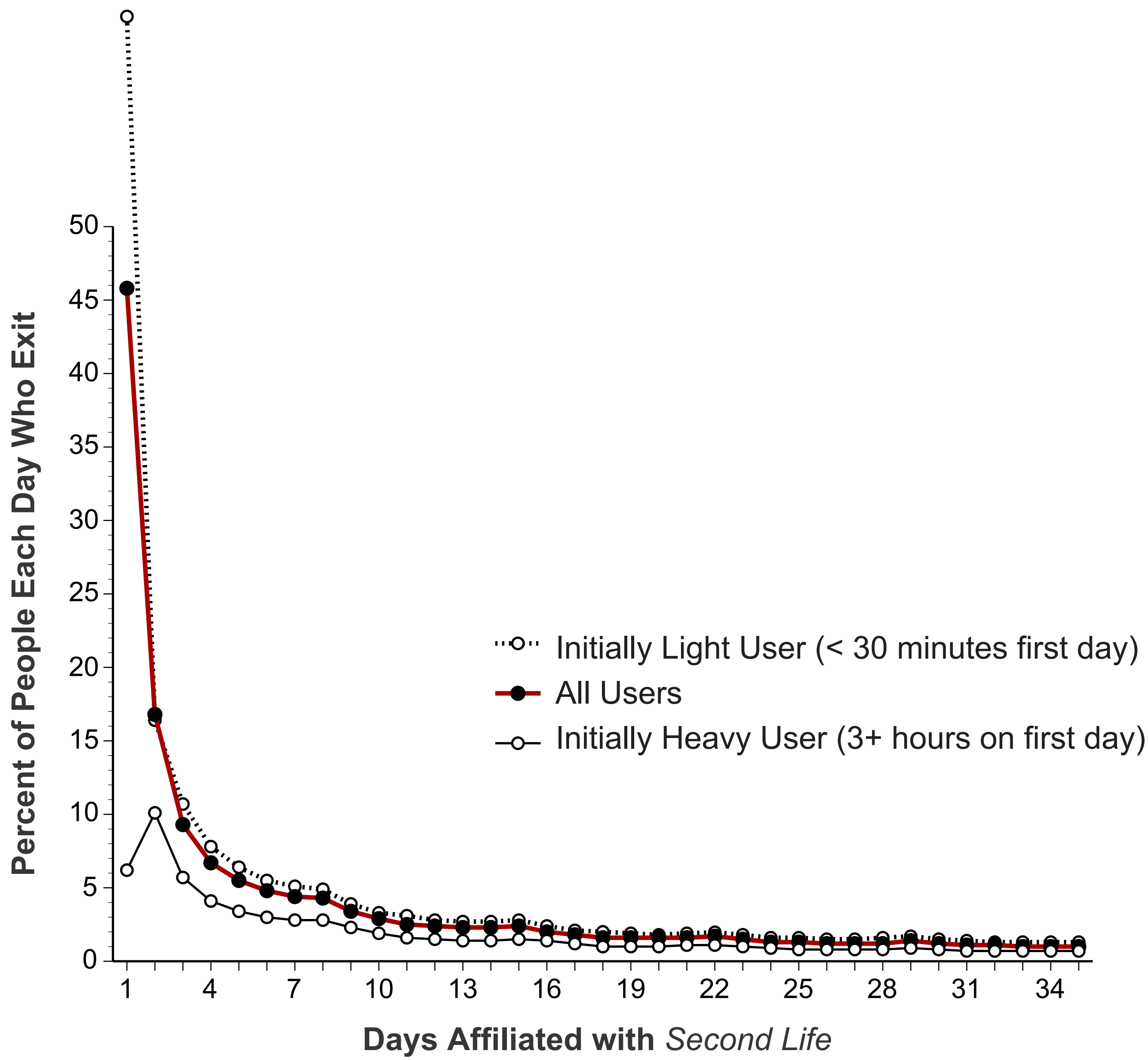


Figure 3. Exit Is Concentrated In The First Few Days

Figure 4.

Diffusion of <u>Second Life</u>

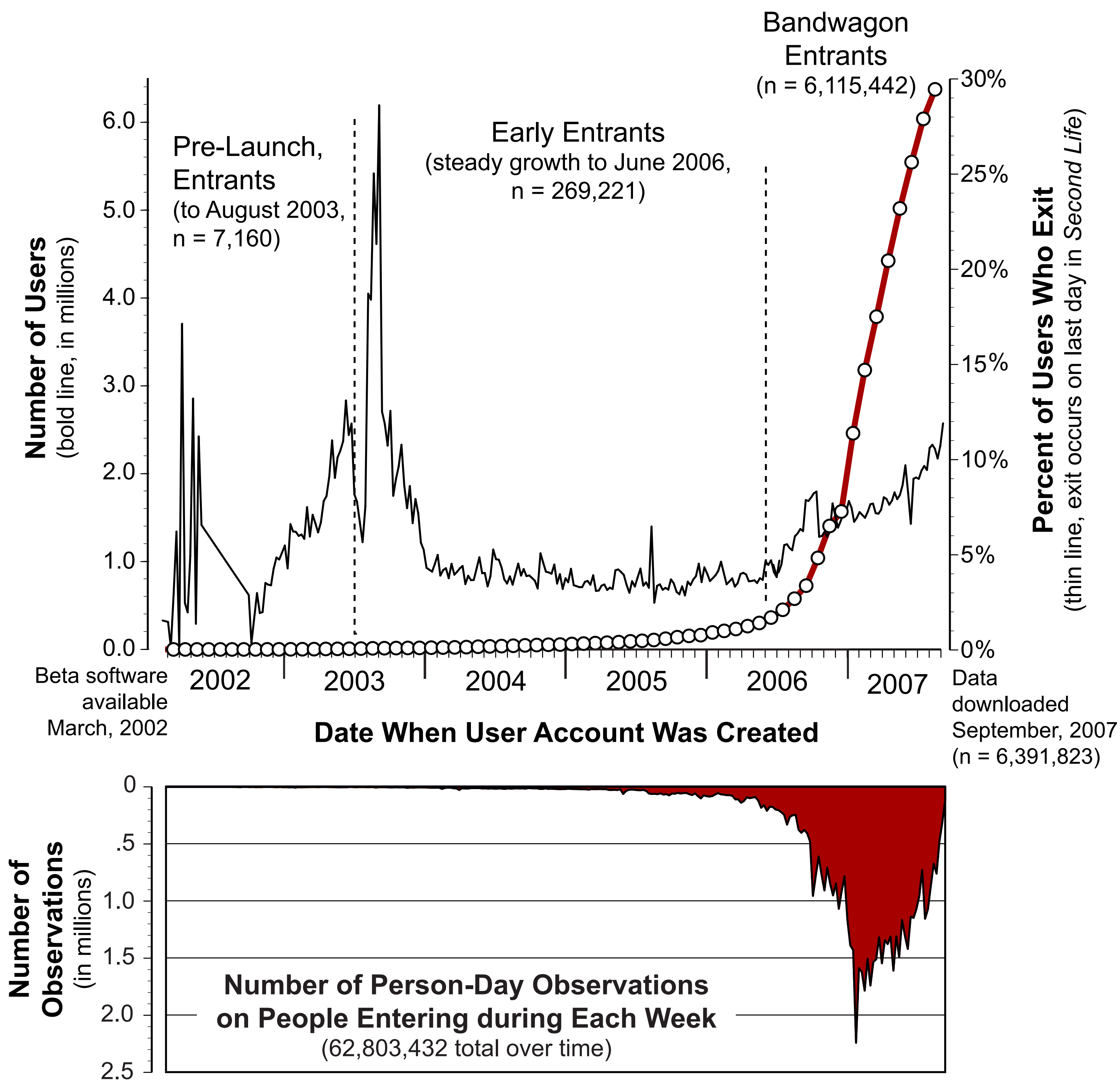
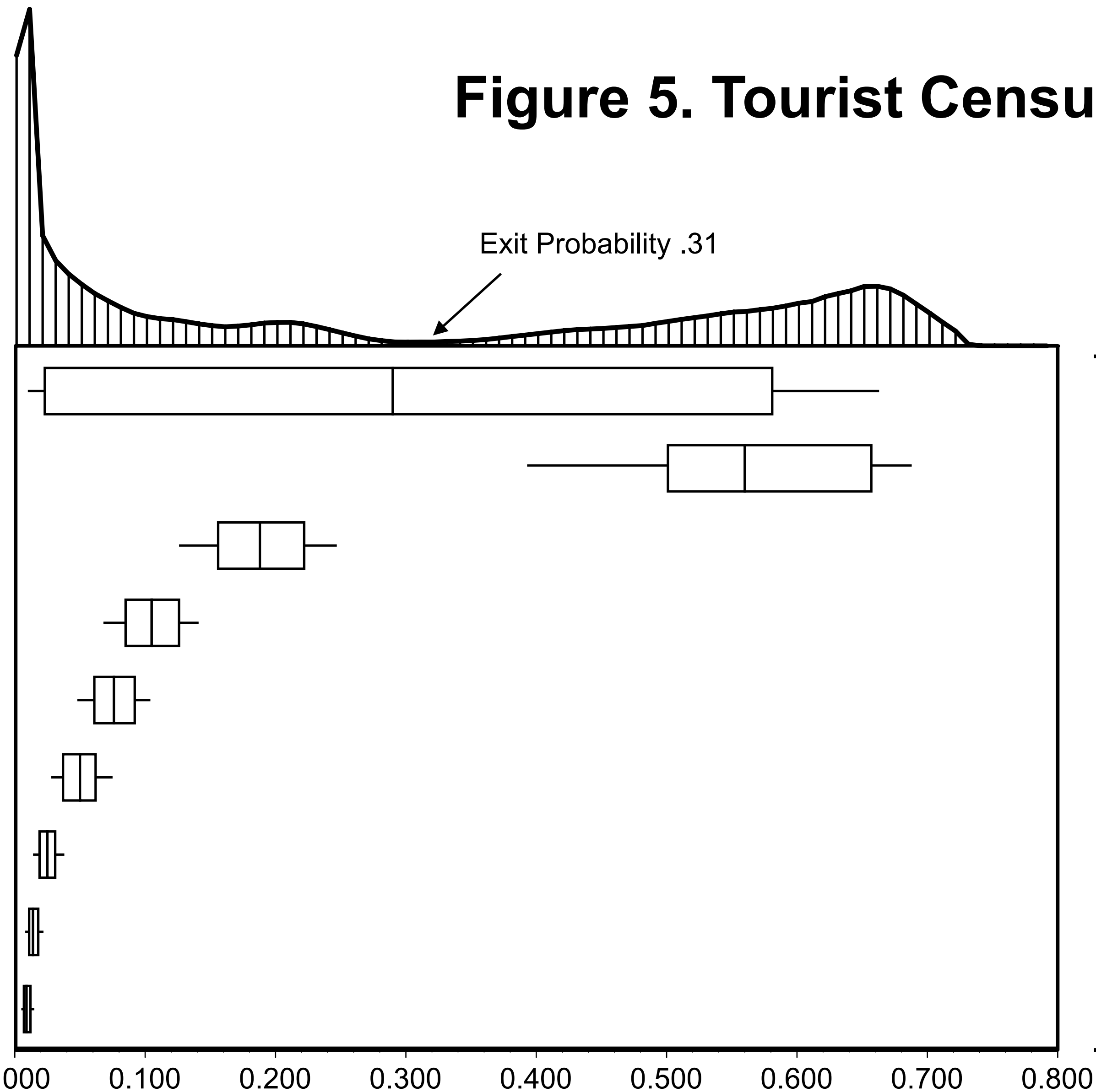


Figure 5. Tourist Census



Days Inworld	Residents	Percent Tourists
All	6,080,514	43%
One	2,954,836	94%
Two	577,399	1%
Three	266,840	0%
Four	175,867	0%
5 - 10	560,428	0%
11 - 20	363,069	0%
21 - 35	276,799	0%
More	1,075,171	0%

NOTE — These are the 6,080,514 residents for whom all characteristics are known in the Table 1 exit-prediction equation. Top histogram shows distribution of exit probabilities. Rows distinguish residents by their days in Second Life (last day active minus first day active). Box plots are defined by the 25th and 75th percentile, around the mean exit probability, with whiskers extending to the 10th and 90th percentiles. For the right column, residents above a .31 exit probability are deemed tourists.

Table 1.

Event-History Model Predicting Exit

Note — These logit models predict ego's exit based on person-day observations. Days Inworld is the 35 days on the horizontal axis of Figure 3. Years younger than 46 is zero for ages over 46. Hours inworld on the first day is the hours ego spent inworld on the day ego first entered Second Life. Second Life age is how much time had been spent developing the world (number of days between January 1, 2002 and ego's first session in Second Life, divided by 365; e.g., January 30, 2003 entry would be 395/365, or 1.082). The three periods are distinguished in the text and Figure 4. Standard errors are adjusted for repeated observation of the same person using the "cluster" option in STATA.

	All Data	Censored Data		
	Logit Coefficient	Logit Coefficient	Standard Error	Test Statistic
User Experience				
2. Days Inworld (log)	-.799	-.878	.0017	-511.68
3. First Day	1.800	1.674	.0045	371.13
4. Second Day	.493	.403	.0037	109.02
5. Third Day	.101	.065	.0033	19.66
6. Fourth Day	.000	-.024	.0034	-7.12
Kind of Person (individual tastes)				
7. Male	.124	.123	.0014	87.28
8. Years Younger than 46	.027	.028	.0001	387.17
9. North American	.057	.033	.0017	19.06
10. Western European	-.021	-.009	.0014	-6.07
11. Hours Inworld Day 1	-.098	-.115	.0004	-287.68
12. (Day 1 Hours)(Day 1)	-.778	-.761	.0015	-513.92
13. (Day 1 Hours)(Day 2)	-.039	-.018	.0008	-21.75
Entry Conditions (operations-crowding)				
14. Second Life Age (log)	-.701	-.711	.0128	-55.49
15. Entered in Period 1	-1.030	-1.047	.0361	-29.04
16. Entered in Period 3	-5.958	-3.441	.0263	-130.98
17. (Period 1)(SL Age)	3.402	3.414	.1240	27.53
18. (Period 3)(SL Age)	4.013	2.437	.0179	136.24
Intercept	-1.328	-1.140		
Observations	61,863,520	57,744,508		
Pseudo R^2	.320	.330		

Table 2. Tourists Are Low on Achievement, Network, and Trust

	Residents	Tourists	Test Statistic
Closure (ego-alter number of mutual contacts)	1.16	.05	-91
Trust from Ego (0-100 scale)	55.44	53.66	-17
Trust to Alter (0-100 scale)	56.00	47.94	-79
Brokerage (0-100, reversed network constraint)	4.97	.07	-490
Achievement (z-score)	.23	.00	-298

Note — Closure and trust are from the graph to the left in Figure 2 for the 1,755,736 friendships on which ego and alter attribute data and exit probabilities are available. Tourists are distinguished as in Figure 5 by an exit probability higher than the median .184. Routine t-test statistics are presented (based on standard errors adjusted for correlation between ego's relations using STATA's "cluster" option; 655,596 clusters). Brokerage and achievement are from the graph to the right in Figure 2 for the 5,631,917 residents on whom attribute data are available. Routine t-test statistics are reported.

Table 3. Closure-Trust Association in Second Life

Network Variables:						
Number of Mutual Contacts	5.25	(106)	5.19	(105)	5.16	(97)
Number of Mutual Contacts Squared	-.28	(-57)	-.28	(-56)	-.28	(-56)
One-Contact Network	2.42	(29)	2.62	(32)	2.79	(32)
Ego Network Size	-.13	(-25)	-.13	(-25)	-.13	(-25)
Alter Network Size	.03	(97)	.03	(94)	.03	(92)
Tourist Controls:						
Ego Tourist			.76	(7)	.48	(2)
Alter Tourist			-5.34	(-55)	-13.41	(-59)
Ego Tourist x Number of Mutual Contacts			-.08	(-1)	-3.77	(-2)
Alter Tourist x Number of Mutual Contacts			1.35	(5)	4.97	(6)
Homophily Controls:						
Ego Player Lives in North America	.66	(6)	.62	(6)	.60	(5)
Alter Player Lives in North America	.42	(6)	.39	(5)	.36	(5)
Same Geographic Region (five regions)	2.64	(41)	2.62	(40)	2.61	(40)
Ego Player’s Age (years)	.16	(36)	.16	(35)	.16	(34)
Alter Player’s Age (years)	.09	(31)	.08	(27)	.07	(25)
Age Difference (|ego – friend|)	-.20	(-58)	-.20	(-59)	-.20	(-58)
Ego Player Is Female	4.65	(49)	4.64	(49)	4.61	(48)
Alter Player Is Female	-2.13	(-31)	-2.20	(-33)	-2.24	(-33)
Both Female	1.44	(13)	1.46	(13)	1.46	(13)
Intercept [R^2]	43.55	[.12]	44.12	[.12]	44.59	[.12]

Note — These regression models predict the level of ego’s trust in alter in the 1,755,736 friendships on which ego and alter attribute data and exit probabilities are available (0 to 100 scale, vertical axis to the left in Figure 2). A resident is a tourist in the second column if his or her exit probability is greater than the median (see Figure 5). Tourist in the third column is a resident’s exit probability predicted by the censored-data logit model in Table 1. Routine t-test statistics are in parentheses (standard errors are adjusted for correlation between ego’s relations using STATA’s “cluster” option; 655,596 clusters).

Table 4.
Brokerage-Achievement Association in Second Life

Network Variables:						
Reversed Constraint	.015	(476)	.015	(468)	.016	(472)
Semi-Isolated	.082	(66)	.073	(57)	.073	(57)
Tourist Controls:						
Tourist Level Adjustment			-.021	(-29)	-.043	(-30)
Tourist Slope Adjustment			-.015	(-56)	-.068	(-108)
Player Attributes:						
North American	.084	(88)	.083	(87)	.082	(86)
Western European	.010	(12)	.010	(12)	.009	(11)
Age (decades)	.004	(9)	.001	(4)	.000	(0)
Female	-.032	(-38)	-.033	(-39)	-.034	(-40)
Experience (hours inworld)	.137	(1113)	.137	(1106)	.136	(1090)
Intercept [R^2]	-.023	[.32]	-.004	[.32]	.003	[.33]

Note — These regression models predict level of achievement for the residents on whom attribute data and exit probability are available. Achievement is a z-score based on resident success in creating groups, that attract many members, and survive their initial founding (vertical axis to the right in Figure 2). Reversed constraint is 100 times one minus ego's network constraint score (horizontal axis to the right in Figure 2). Semi-isolated is a resident who is a member of one or more groups, but has no friends. A resident is a tourist in the second column if his or her exit probability is greater than the median (see Figure 5). Tourist in the third column is a resident's exit probability predicted by the censored-data logit model in Table 1. Routine test statistics are reported in parentheses (n = 5,631,917).