Assessing and Quantifying Network Effects in an Online Dating Market

Gordon Burtch* Jui Ramaprasad†

Abstract

We empirically examine and quantify network effects on a large online dating platform in Brazil. We exploit a natural experiment, wherein a focal platform acquired its Brazilian competitor and subsequently imported the competitor's base of 150,000+ users over a 3-day period; a large exogenous shock to the composition of the purchasing platform. Our study context and the natural experiment provide two unique sources of identification: i) accounts purchased from the competitor were almost exclusively heterosexual users, even though the purchasing platform also played host to homosexual users, and ii) the treatment varied across cities, in that the "value" of new users to the existing user base differed. because purchased users differed from existing users in terms of their average characteristics (e.g., location within the city). We leverage the former to estimate a difference-in-differences specification, treating homosexual enrollment and exit rates as a plausible control for those of the heterosexual population, whereas the latter provides us with an opportunity to explore the importance of local market structure in the manifestation of network effects. We find that the treatment increased both rates of enrollment and rates of exit, amongst both genders, with a net positive effect that translated to a 17% increase in short-term revenue for the platform. We find clear evidence of local network effects; in cities where the average spatial distance between new users and existing users was larger, the treatment effect was significantly weaker. We discuss the implications for the literature and practice, and we suggest a number of avenues for future work in this space.

*Assistant Professor, Information Systems Department, Carlson School of Management, University of Minnesota. gburtch@umn.edu.

†Associate Professor, Information Systems Department, Desautels Faculty of Management, McGill University. jui.ramaprasad@mcgill.ca.

1 Introduction

Two-sided markets (or multi-sided markets more generally) are roughly defined as platforms that enable transaction and interaction amongst end-users of different types [Rochet & Tirole, 2006]. Examples of these markets are now quite prevalent, e.g., Hulu matches video content providers with end-consumers and advertisers, eBay connects buyers and sellers, and AirBNB links renters and hosts. Perhaps the oldest example, however, is the market for dating, which we consider here.

The success of two-sided markets hinges on network effects. Most commonly, individuals experience greater utility when the number of potential transaction partners is large (cross-side positive effects). At the same time, individuals experience lower utility when the number of competitors is large (same-side negative effects, or congestion). Platform operators employ a variety of techniques to stimulate and take advantage of network effects. In this work, we offer a case study of an online dating platform, which implemented one such technique: a seeding strategy. We do so with an eye toward identifying and quantifying network effects in this setting. Moreover, we seek to understand the importance of local network structure in the market (i.e., the fit between the seeding action and the preferences and interests of targeted users), and how this can determine the efficacy of the seeding strategy. Empirically understanding the nature of network effects is becoming increasingly important as the number of platforms competing in the online space grows. This is particularly true in the dating space, which has grown extremely competitive in recent years.[1]

Similar to more traditional markets of sale and purchase (e.g., eBay), which have received the bulk of consideration in the network effects literature to date, we might expect to observe positive cross-side and negative same-side network effects. However, we will argue that the story is somewhat more nuanced when it comes to online dating, for a number of reasons. First, broadly speaking, these nuances and differences derive from the fact that these markets facilitate social transactions and not economic ones, and thus individuals preferences for a transaction partner are likely to be highly heterogeneous. Second, by construction, participants in these markets generally have no direct visibility into levels of congestion or competition, because users do not typically receive information about other users who would not provide a viable relationship match.[2] Third, the

[1] Online dating magazine estimates that there are 2,500 online dating sites in the US alone: http://www.forbes.com/sites/martinzwilling/2013/03/01/how-many-more-online-dating-sites-do-we-need/

[2] Admittedly, this observation applies only to heterosexual users; homosexual users would of course serve, simultaneously, as possible matches and as competition for one another

online dating context, despite digital intermediation, nonetheless continues to depend heavily on offline word-of-mouth. Together, these differences highlight that the role and impacts of network effects may be highly contextual and should be considered in a case-by-case basis.

We seek to address the following three research questions: ***1) To what degree does seeding new users stimulate network effects in online dating markets; 2) What is the impact of this seeding strategy on enrollment of new users and on exits of existing users; and 3) How does this impact vary with differences between the characteristics of new (seeded) and existing users (i.e., matching potential)?*** To do so, we draw upon anonymized and aggregated data from the Ashley Madison leak of 2015. Following other researchers who have used this dataset for academic research purposes, we asked an external party to anonymize the data [Griffin, Kruger, & Maturana, 2016; Grieser, Kapadia, Li, & Simonov, 2016]. Our interest in this dataset comes from one key event that enables us to study seeding in online dating: Ashley Madison purchased a Brazilian competitor, Ohhtel, and thereby acquired its user base, an action which resulted in the sudden injection of more than 150,000 new user accounts into the platform.

We exploit two features of this study context to identify the effects deriving from the shock on subsequent patterns of enrollment and exit (attrition) on the dating site. First, whereas the competitor only allowed heterosexual matching, the focal site allowed both heterosexual and homosexual matching, Hence the exogenous shock to the number of user accounts on our focal site was to heterosexual users, which allows us to use homosexual users as a control group. Second, the treatment varied across cities, in that the characteristics of purchased user accounts (e.g., physical location within the city) differed to varying degrees from those of the existing base of users. We leverage the first feature to obtain a plausible control group that was not subject to the exogenous shock, to estimate a difference-in-differences (DD) specification, contrasting the enrollment and exit rates between heterosexual and homosexual users. We leverage the second feature to explore the notional of *local* network effects; the notion that the strength of network effects will vary based on the degree to which the new entrants can serve as a possible match for other users in the market.

We find, first, that the seeding action did deliver a positive return for enrollments, amongst both genders. Noting that site design precludes the manifestation of direct or indirect network effects on enrollment (because new users are incapable of viewing potential matches before they create an account), we attribute this finding to a mechanism of social proof, manifest via offline word-of-mouth. Second, we find that the treatment also increased the rate of exit amongst both genders, suggesting a countervailing result. Our calculations suggest an overall net benefit to the platform equivalent to an approximate 17% increase in the rate of acquisition of male users, and thus revenue. Third, exploring local structure in the market, we observe that the treatment effect was indeed heterogeneous between cities. In particular, we find that as the physical distance between purchased users and existing users grows in a given city, the effect of the treatment attenuates.

This paper extends the current work on network effects in online platforms. Our main contribution lies in identifying and quantifying network effects in the context of online dating, in contrast to the current literature that mainly focuses on platforms that facilitate economic exchanges. Importantly, we look at how one seeding strategy (seeding the platform with additional users) can stimulate network effects, addressing the critical "chicken-and-egg" problem faced by two-sided platforms. From the practical perspective, we are able provide a measure of the economic effects of such a seeding strategy by estimating the proportional revenue returns. In exploring these questions, we consider that the nature of network effects can vary across contexts and market types. One significant reason that network effects may vary, as we demonstrate here, is based on local structure in the market, and how that structure aligns with the preferences and interests of users. In this setting, physical location is an important determinant of matching because, ultimately, matches must take place offline. In other settings, where transactions are fully mediated by the online platform, geography would be unlikely to play a significant role.

The remainder of this paper is structured as follows. We review the literature on network effects, two-sided markets, and online dating in the following section. We then describe our study context, the natural experiment and the data at our disposal, before detailing our empirical approach. We then report our results, discuss the implications and limitations of our work, and propose a number of avenues for future work.

2 Related Literature

In the literature review below, we highlight literature relevant to our paper, focusing on work that looks at network effects in two-sided platforms as well as the existing literature examining online dating platforms.

2.1 Network Effects

Two-sided markets (or multi-sided markets more generally) are roughly defined as platforms that enable transaction and interaction amongst end-users of different types [Parker & Van Alstyne, 2005; Rochet & Tirole, 2006]. Examples of these markets are now quite prevalent, e.g., Hulu matches video content providers with end-consumers and advertisers, eBay connects buyers and sellers, and AirBNB links renters and hosts. Perhaps the oldest example, however, is the market for dating, which we consider here.

Any two-sided network is generally faced with the challenge of how best to attract and retain participants on each side of the market, to achieve and maintain critical mass, and ultimately turn a profit [Rochet & Tirole, 2006]. Achieving and maintaining participation on either side is a difficult proposition because platforms face a cold-start problem; they must attract membership, yet members have a significantly reduced interest in joining if no other members exist to transact with. This notion, which speaks to cross-side network effects [Katz & Shapiro, 1985], implies that the platform often requires a specific strategy to achieve growth. That strategy often hinges on pricing. The platform operator may choose to subsidize participants on one side of the market, offering free or discounted access and features [Parker & Van Alstyne, 2005]. Indeed, this has long been a standard practice in dating markets, wherein free access is often granted to female users, as noted in the introduction. Long before the emergence of online dating, bars were using promotions such as "Ladies Night," letting ladies in for free in order to attract men (in the case of heterosexual pairs), i.e. to stimulate cross-side network effects.

Online, platforms have experimented with a variety of strategies in order to stimulate cross-side network effects. One of the most common techniques is the offer of subsidies. However, alternative, non-price strategies have also been shown to be quite effective. For example, if the platform can institute a perception of adoption by members on one side of the market, this can have a profound on enrollment on the opposite side. Tucker and Zhang [2010] report on a field experiment in which they demonstrate the benefit of advertising existing user volumes to potential users. They found that informing potential sellers of the volume of existing sellers in a B2B marketplace had a positive effect on the probability of sign-up. However, informing potential sellers of existing sellers and buyers, in tandem, eliminates the effect. The authors take this as evidence of an indirect network effect, wherein the absence of information about buyer volumes causes the potential seller to infer the number based on known seller volumes.

Anecdotal evidence also speaks to the success of platforms that have initially "faked" membership and usage (an example of a seeding strategy).[3] In 2012, Steve Huffman, co-founder of Reddit, revealed that the site stimulated early growth in its user base via the creation of fake accounts. Similar practices have been reported at PayPal, which reportedly grew its early base of eBay sellers by creating a bot that would automatically purchase goods and then insist that the payment be completed via PayPal. This strategy caused eBay sellers to flock to the Paypal platform, under the perception that it was heavily populated by consumers [Jackson, 2004].

Just as positive cross-side effects manifest with a growth in potential transaction partners, negative same-side effects can arise with the enrollment of competitors. Sun and Tse [2009] explain this phenomenon as follows: "Even if there are infinite potential participants of the network, the exponential growth of the network may be limited by other factors such as competition among participants. This is called a congestion effect in the two-sided market literature [Rysman, 2004] and it represents negative network externalities among the participants on the same side of the market. Sun [2007] finds that when such negative externalities exist, the growth of a two-sided network has a finite limit."

Network effects can manifest in a heterogeneous fashion in many settings, when individuals prefer to transact with specific subsets of the population. Sundararajan [2007] formalizes this idea, presenting a model in which users prefer adoption of a product by their social network neighbors. Other work has frequently acknowledged the importance of local network structure in the manifestation of network effects, considering, for example, local network density [Katana, Zubcsek, & Sarvary, 2011] and strong tie subsets [Suarez, 2005].

Given the highly personal nature of relationships, we would argue that a similar pattern of localized preference can play out in ways that need not depend strictly on social network structure. That is, we propose a broader definition of what is considered "local". In particular, we note that individuals may prefer adoption by a specific subset of the population for a variety of other reasons. Individuals may prefer adoption by people who speak the same language [Meer & Rigbi, 2013], who reside nearby, or who come from a similar cultural background [Burtch, Ghose, & Wattal, 2014].

[3] http://venturebeat.com/2012/06/22/reddit-fake-users/

2.2 Online Dating

Online dating itself is a fairly new phenomenon that has just recently achieved social acceptability in the mainstream. The related online dating literature is nascent but growing. While the bulk of the of the work in online dating has been from the psychology literature, with a focus on individual self-presentation [Ellison, Heino, & Gibbs, 2006], deception [Toma, Hancock, & Ellison, 2008], and image, the economics literature has also examined online dating, looking specifically at mate preferences and sorting in matching [Hitsch, Hortacsu, & Ariely, 2010].

Most relevant to our work is the literature on platform design of online dating sites. Recent literature in Information Systems has looked at the impact of different features (e.g. anonymous browsing) on matching outcomes [Bapna, Ramaprasad, Shmueli, & Umyarov, 2016]. Other work has examined the design of recommendation systems for human-to-human matching, which is different from human-to-product matching that we see on sites like Amazon.com [Pizzato, Rej, Chung, Koprinska, & Kay, 2010], and the inclusion of search features to narrow down the choices of potential matches [Fiore & Donath, 2004]. Though critically important to the success of an online dating site, few papers have looked at network effects in the online dating context, and in particular little work has examined impacts on enrollment of new users and attrition of existing users–critical factors in the success of such a site. The objective of this project is to address this gap.

Similar to more traditional markets of sale and purchase (e.g., eBay), which have received most of the consideration in the network effects literature to date, we might expect to observe positive cross-side and negative same-side network effects. However, we will argue that the story is quite different in the online dating context, for a number of reasons, most notably because these markets do not facilitate economic transactions; instead, they facilitate social transactions. Moreover, by construction, despite advertisements that tout the "millions of possibilities" available on online dating sites [Heino, Ellison, & Gibbs, 2010], participants in these markets generally have no direct visibility into levels of congestion or competition, because users do not typically receive information about other users who would not provide a viable relationship match. Finally, a successful transaction on an online dating platform may result in the loss of two participants, thus requiring the platform to continually seek new enrollments in order to sustain itself. Together, these differences highlight that the role and impacts of network effects may be highly contextual and should be considered on a case-by-case basis.

Online dating markets provide perhaps the best opportunity to explore the importance of local preferences as a determinant of network effects. In online dating, individuals vary quite a bit in their statements about the characteristics they deem to be desirable in a relationship partner (i.e., "turn-ons" vs. "turn-offs"). However, recent work suggests that, by and large, despite what individuals claim, they generally prefer, in practice, to date others who are extremely similar to themselves.[4] Accordingly, an ideal seeding strategy for an online dating market would likely focus on seeding new users who most closely resemble the existing user base.

Of course, when it comes to platform growth strategies in the general sense, it likely behooves platform operators to explore the preference distributions of users in a given market, and to then optimize based on what is found. That is, firms may wish to incentivize participation amongst *specific* subsets of the population or they may wish to seed specific types of content, to maximize the resultant demand.

3 Methodology

3.1 The Natural Experiment

On April 10$^{\text{th}}$ of 2012, Ashley Madison's acquisition of Brazilian competitor Ohhtel was announced, for an undisclosed dollar amount[5]. With the announcement, the media reported that the Ohhtel user data was transferred directly to Ashley Madison's servers. Ohhtel users, upon attempting to login to their user accounts after that transfer, would be redirected to the Ashley Madison website and asked to accept a set of terms and conditions before proceeding. This purchase and user transfer reportedly resulted in the addition of approximately 150,000 new users to Ashley Madison's Brazilian website.

The user account transfer described by the media is readily apparent in our sample. Figure 1 depicts the volume of Brazilian user account creations over time in our sample, in the two months surrounding April 10$^{\text{th}}$. Here, we clearly see the creation of approximately 150,000 new user accounts

[4] Five Thirty Eight: http://fivethirtyeight.com/features/in-the-end-people-may-really-just-want-to-date-themselves/

[5] Globo News: http://g1.globo.com/economia/negocios/noticia/2012/04/site-de-traicao-ashley-madison-compra-base-de-usuarios-do-ohhtel.html

between April 9$^{\text{th}}$ and April 11$^{\text{th}}$. When we look more closely at these accounts, we observe that they were all created from the same IP address. A *whois* lookup on that IP reveals that it is maintained by OnX Enterprise Solutions (US) Ltd., a hosting services provider that also hosts AvidLifeMedia.com (ALM is the umbrella company that owns Ashley Madison). These observations are consistent with the occurrence of a large database import of user accounts having taken place at the time the acquisition was announced in the media.

A notable feature of these purchased accounts is that they were almost exclusively heterosexual; just 78 of the 150,000+ users indicated an interest in same-sex partners. This implies that the natural experiment presents an exogenous shock to the volume of heterosexual user accounts on Ashley Madison's Brazilian platform, yet virtually no shock to the volume of homosexual user accounts.

[Figure 1 about here.]

3.2 Data

We employ aggregate, anonymized data from the Ashley Madison data leak of 2015. In anonymizing the data, we follow the approach of other recent academic research that has drawn on the same data set [Griffin et al., 2016; Grieser et al., 2016]. The natural experiment takes place from April 9$^{\text{th}}$ to April 11$^{\text{th}}$ of 2012, and our sample spans a 2-month window around these dates (March through May of 2012). We focus on the platform's 120 largest Brazilian markets (cities). Figure 2 depicts the geographic distribution of users in our sample who had registered prior to the treatment date. As can be seen, user volumes follow a power distribution; São Paulo and Rio de Janeiro jointly account for more than 30% of registered users, or approximately 80,000 individuals, while the smallest city in our sample, Rio Grande, hosts slightly more than 400 users.

[Figure 2 about here.]

We aggregate user accounts to construct daily counts enrollment and exit, i.e., last date of profile modification. We are relatively confident that the last date of account activity in our data is a reliable indicator of exit, because although we do not study data post-2012, we draw upon all data through 2015 to identify a user's last action. Thus, a user whose last date of activity was observed in 2012 was inactive for at least 3 years afterward. Accordingly, right censoring is not a serious concern in our data. We construct four panels for each city in our data, for a total of 480 panels. These panels reflect the cross-product of gender and sexual orientation; one panel for heterosexual males, one for heterosexual females, one for homosexual males and one for homosexual females.

In Table 1, we report descriptive statistics related to daily enrollments and exits in each of the cities over over the month prior to the natural experiment, for each combination of gender and sexual orientation.

[Table 1 about here.]

3.3 Econometric Specification

All purchased user accounts were heterosexual. As such, homosexual enrollment serves as our control. Because our dependent variable represents a strictly positive count measure and exhibits a power (right skewed) distribution, we draw upon the Conditional Fixed Effect Poisson Quasi-Maximum Likelihood (PQML) estimator proposed by Wooldridge [1999], and we subsequently evaluate the robustness of our primary results to the use of a Log-OLS specification.

Bertrand et al. [2004] advocate a number of strategies to avoid issues of false significance deriving from the use of a difference-in-differences estimator on data with an outcome variable that exhibits serial correlation. One such strategy, easily implemented, is to ignore time-series data and average observations across the pre period, as well as observations across the post period, collapsing the panel into pairs of observations for each unit. We take that approach here. Thus, we begin by constructing a panel of observation pairs, reflecting enrollment volumes for each gender and sexual orientation, in each city, in the pre and post periods. Our panel is thus short ($T = 2$), yet relatively wide. In our estimations, we consider two alternative time windows; 2 weeks before and after the treatment (i.e., 1-month), and 4 weeks before and after the treatment (i.e., 2-months).

Equation 1 reflects the log-linear specification, where i indexes cities, j indexes sexual orientation and t indexes time. Our outcome of interest, Y is the count of enrollments or exits, which we model as a function of *Orientation*, an indicator that equals 1 if the panel pertains to heterosexual users, *Post*, an indicator of whether the observation takes place after the natural experiment date, and the interaction between the two. Additionally, we incorporate a city fixed effect, α. Here, we are particularly interested in the coefficient β_3, the difference-in-differences (DD) estimate. In performing these estimations, we exclude the purchased user accounts from our enrollment and exit measures. Moreover, in the specific case of our attrition (exit) estimations, we exclude any users who registered on or after the treatment date. This is necessary in order to rule out the possibility that users who arrive due to the treatment may be systematically different than users who arrived "organically." For example, it may be the case that treatment-induced enrollments are systematically less interested or committed to the platform, and thus that they may exit more rapidly. Accordingly, were we to include post treatment entrants in our calculation of exit rates over time, we might observe an uninformative increase in the exit rate following treatment.

$$\log(Y_{ijt}) = \alpha_i + \beta_1 \cdot Orientation_j + \beta_2 \cdot Post_t + \beta_3 \cdot Orientation_j * Post_t + \varepsilon_{ijt} \quad (1)$$

Before progressing to our primary estimations, it is also useful to first explore the validity of our assumption that the homosexual population can serve as a valid control for the heterosexual population. To determine this, we first estimate two relative time difference-in-differences models [Autor, 2003; Angrist & Pischke, 2009], pooling the gender panels, separately considering enrollments and exits. We estimate the specification presented in Equation 2, treating April 8$^{\text{th}}$ as the reference period (the day prior to the natural experiment). If the homosexual population is a valid control group, we would expect to find no statistically significant coefficients associated with pre-treatment relative time dummies, yet significant coefficients associated with post-treatment relative time dummies. We estimate this model on our un-collapsed panels for the 28 days around April 9$^{\text{th}}$. We plot our estimates of γ from each regression in Figures 3 and 4. As can be seen in the two figures, there are no significant differences in enrollment or exit between the heterosexual and homosexual populations before the natural experiment, yet we begin to see differences emerge shortly after. The lack of significant differences prior to the experiment suggests that we have no reason to believe that the homosexual population cannot serve as a valid control in our setting.

$$\log(Y_{ijt}) = \alpha_i + Orientation_j + \sum_q \tau_{qt} + \gamma \cdot \sum_q \tau_{qt} * Orientation_j + \varepsilon_{ijt} \quad (2)$$

[Figure 3 about here.]

[Figure 4 about here.]

4 Results

We next proceed with our primary analyses. Table 2 reports our findings for new user enrollments, by gender, for both the 4-week and 2-week pre/post windows. Table 3 reports the results using a Log-OLS specification. We observe consistent results in both cases. Referring to Table 3, we see that female enrollments increased by approximately 21.1% in the two weeks after treatment, whereas male enrollments rose by approximately 23.8%. The latter result is of particular note, given that only male users pay to use the Ashley Madison website.

[Table 2 about here.]

[Table 3 about here.]

Next, we consider the natural experiment's impact on exit rates. We observe significant positive effects across all estimations. Table 4 reports our findings for user exits, by gender, for both the 4-week and 2-week pre/post windows, while Table 5 reports the results using a Log-OLS specification. Referring to Table 5, we see that female exits increased by approximately 24.8% in the two weeks after treatment, whereas male exits rose by approximately 56.3%. The latter result is once again of particular note, because only male users pay to use the Ashley Madison website.

[Table 4 about here.]

[Table 5 about here.]

Having obtained estimates of the impact of the treatment on both entry and exit rates, we can now assess the net effect on revenue by calculating the shift in the net acquisition of paying (male) users, accounting for both entry and exit. Again considering the two-week pre-post window, where the effects were strongest, the elasticity of enrollment, equal to 24%, equates to an approximate increase of 1.66 new male users per day in the average city, or 72 new male users per day in the largest city, São Paulo. The elasticity of exit, equal to 56%, equates to an approximately increase of 0.66 additional male user exits per day in the average city, or 8 additional male exits per day in

São Paulo. Accordingly, in the average city, we see an approximate net daily acquisition of 1 new male (paying) user per day in the average city, over and above the baseline net acquisition rate of approximately 6 new male (paying) users, or an estimated 17% increase in short term revenue.

4.1 Local Network Effects: The Importance of Co-Location

To explore the role of local network effects in this marketplace, i.e., the degree to which our results depend on the distribution of preference in the existing user base, in a given city, we consider geographic separation. Geographic distance has been highlighted as a particularly important factor in users selection of online dating partners [Couch & Liamputtong, 2008]. This is because physical distance translates to increased travel time. As such, intuitively, if our results are indeed driven by network effects, we would expect to observe weaker effects in cities where the average location of pre-existing users deviates from the average location of purchased users (e.g., analogous to the notion of cluster separation) . To determine average location, we draw upon latitude and longitude data recorded for each user account at the time of creation, which are likely based on the user's IP address at registration, or a self-reported address. For each city, we calculate the average geographic coordinates of pre-existing heterosexual users, and another set of average coordinates for newly entered (purchased) users. Based on these coordinate pairs, we then calculate our measure of geographic separation, *Geo*, as the Euclidean distance between the two (e.g., analogous to a centroid distance). This measure then enters our baseline specification as a moderator, along with all lower-order interactions, as reflected in Equation 3. In this model, we are thus particularly interested in β_6 and β_7, our difference in differences estimate and its moderation by geographic distance. In this setup, we would expect to once again observe a positive coefficient for β_6 and a negative coefficient on the moderator, β_7.

$$\begin{aligned}\log(Y_{ijt}) = \alpha_i + \beta_1 \cdot Orientation_j + \beta_2 \cdot Post_t + \beta_3 \cdot Geo_i + \\ \beta_4 \cdot Orientation_j * Geo_i + \beta_5 \cdot Post_t * Geo_i + \beta_6 \cdot Orientation_j * Post_t + \\ \beta_7 \cdot Orientation_j * Post_t * Geo_i + \varepsilon_{ijt}\end{aligned} \tag{3}$$

It should be noted that our measure of geographic separation is not strictly accurate, because degrees of latitude and longitude vary in their relationship to traditional measures of physical distance (e.g., miles), depending upon the one's location on the earth. Fortunately, however, Brazil is situated on the equator, where a degree of latitude and a degree of longitude both translate to roughly 68 miles or 110km. Accordingly, a simple Euclidean distance measure can provide a rough approximation of actual physical distance in a localized area (e.g., within a city). We incorporate our distance measure as a moderator on the interaction term in each of our 2-week, pre-post regressions, for both enrollments and exits.

The results of the enrollment regressions are reported in Table 6, and those for the exit regressions are reported in Table 7. Broadly speaking, as we expected, we find that a greater average geographic distance between pre-existing users and purchased users results in weaker treatment effects. The one exception to this is that we observe no significant moderating effect when it comes to male exits.

[Table 6 about here.]

[Table 7 about here.]

Considering the point we had made above, that a single unit change in our average geographic separation measure equates to approximately 68 miles, with the fact that a number of the largest cities in our sample are quite large, in some cases spanning hundreds of miles, we observe what appears to be a relatively significant attenuation of the treatment effect from locational separation amongst users.

5 Discussion & Conclusion

This work offers a first look at network effects in online dating. We consider a case study in which a Brazilian online dating platform purchased its primary competitor, and subsequently imported the entirety of the competitor's heterosexual user base over a three day period. This action, i.e., seeding new users in the market, imposed an exogenous shock on the volume of heterosexual users in the market, with little to no effect on the homosexual population. We demonstrate through a series of difference-in-difference estimations (treating the homosexual population as a control) that this strategy was indeed effective, increasing the rate of new user entry as well as the rate of attrition, with a net positive effect that translated to an approximate 17% increase in short term revenue. Further, we demonstrate that the strength of the treatment, and thus the resultant network effects, was moderated by the degree of co-location amongst new and existing users, such that greater geographic separation between the two groups resulted in a weaker effect.

We thus contribute to the literature by identifying and quantifying network effects in the context of online dating, in contrast to the current literature that mainly focuses on platforms that facilitate economic exchanges. In doing this, we suggest that given the differences between the different types of platforms, such as in the visibility of network size, the length of the transaction between the two parties, the importance of offline interactions, and importantly, the individual preferences of the users, the nature of network effects can vary across contexts and market types. Moreover, network effects can manifest to varying degrees within a market, depending on local preferences or network structure. We demonstrate here how one feature of users, geography, influences the manifestation of network effects in an online dating market.

It is readily apparent that relevant factors will vary by transaction-type and context. As noted previously, there are numerous online platforms in which the geography of users would be unlikely to play an important role in determining network effects, namely those platforms which fully mediate matching and transaction (e.g., video game consoles and online multiplayer games). At the same time, other factors that bear no relevance to the dating context may play a particularly important role in other markets. Returning to the example of online games, players may, for example, exhibit a preference for opponents of a comparable skill level, in order to ensure a "challenge."

Interestingly, these ideas suggest that more traditional launch strategies for highly innovative products, such as targeted launch [Lee & O'Connor, 2003], remain somewhat relevant for products and services characterized by network effects. Although the volume of adopting users may be of paramount importance in the presence of network effects, our findings suggest that, under a resource constraint, it may be to the platform operator's advantage to allocate it's resources toward the acquisition of specific individuals, namely those who are likely to be of greatest interest to others, and thus most likely to stimulate subsequent adoption.

Thus, the results of this work can help inform strategies for online dating platforms, whose number and popularity are growing. As most platform operators are now aware of the importance of network effects, it is important for firms not only to achieve a large install base, but to do so as quickly as possible. By optimizing platform launch and growth strategies, our findings suggest that firms can accelerate their acquisition of new users and increase their likelihood of succeeding in a winner-take-all scenario. At the same time, platforms need to remain cognizant of congestion and search costs. As the install base of an online dating platform grows large, users will face greater difficulty sorting through suggested matches to identify those with real potential. By limiting growth to users who are quite likely to match with others in the market, the platform can maintain efficiency and ensure an ideal user experience. Interestingly, some relatively new online dating applications appear to have struck upon this same conclusion. Newer online dating services like *Coffee Meets Bagel*[6] place a greater focus on relevant matches, providing users with just one recommendation each day, based on user-specified filters. In taking this approach, the platform operators seek to reduce noise and congestion while simultaneously ensuring the relevance of potential matches.

Our work is of course subject to a number of limitations. First, we are unable to identify the exact mechanism of the observed treatment effect on exit rates. That being said, as noted above, the immediacy of the response and the fact that heterosexual users are unable to directly observe their competition implies that the increase in exits is due to successful matches. That being said, future work might look to build on our findings employing more granular, nuanced data, enabling clearer identification of this mechanism.

Second, our data is limited to a single geography (Brazil), which could limit the generalizability of our results in other countries, given that dating norms vary across cultures [Hatfield & Rapson, 1996]. Future work could examine online dating seeding strategies in other countries. Third, and last, we note that we are only able to identify the short-term impact of this influx of new users to the focal website on enrollments and exits. While indirect network effects could take some time to develop, there are other factors that also confound drivers of enrollments and exits as the time window increases making it difficult to isolate the impact of this seeding strategy.

To further pursue this line of work, it would be particularly useful for researchers to explore the features and characteristics of individuals and markets that may drive matching outcomes in various settings characterized by network effects. We have demonstrated an example of one feature that impacts matching in this particular context. Going forward, researchers might seek to develop a theoretical typology [Bailey, 1994] of features that can result in local network effects, based on transaction and market characteristics.

[6] https://coffeemeetsbagel.com

References

Angrist, J. & Pischke, J.-S. [2009]. *Mostly harmless econometrics: an empiricists companion*. Princeton, NJ: Princeton University Press.

Autor, D. H. [2003]. Outsourcing at will: the contribution of unjust dismissal doctrine to the growth of employment outsourcing. *Journal of Labor Economics*, *21*[1], 1–42.

Bailey, K. [1994]. *Typologies and taxonomies: an introduction to classification techniques*. Thousand Oaks, CA: SAGE Publications.

Bapna, R., Ramaprasad, J., Shmueli, G., & Umyarov, A. [2016]. One way mirrors in online dating: a randomized field experiment. *Management Science*, *Articles in Advance*.

Bertrand, M., Duflo, E., & Mullainathan, S. [2004]. How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, *119*[1], 249–275.

Burtch, G., Ghose, A., & Wattal, S. [2014]. Cultural differences and geography as determinants of online prosocial lending. *MIS Quarterly*, *38*[3], 773–794.

Couch, D. & Liamputtong, P. [2008]. Online dating and mating: the use of the internet to meet sexual partners. *Qualitative Health Research*, *18*[2], 268–279.

Ellison, N., Heino, R., & Gibbs, J. [2006]. Managing impressions online: self-presentation processes in the online dating environment. *Journal of Computer-Mediated Communication*, *11*[2], 415–441.

Fiore, A. & Donath, J. [2004]. Online personals: an overview. *Chi'04 extended abstracts on human factors in computing systems*, ACM, 1395–1398.

Grieser, W. D., Kapadia, N., Li, R., & Simonov, A. [2016]. Fifty shades of corporate culture. *Available at SSRN*, *http://ssrn.com/abstract=2741049*.

Griffin, J. M., Kruger, S. A., & Maturana, G. [2016]. Do personal ethics influence corporate ethics. *Available at SSRN*, *http://ssrn.com/abstract=2745062*.

Hatfield, E. & Rapson, R. L. [1996]. *Love and sex: cross-cultural perspectives*. Needham Heights, MA: Allyn & Bacon.

Heino, R. D., Ellison, N. B., & Gibbs, J. L. [2010]. Relationshopping: investigating the market metaphor in online dating. *Journal of Social and Personal Relationships*, *27*[4], 427–447.

Hitsch, G. J., Hortacsu, A., & Ariely, D. [2010]. Matching and sorting in online dating. *The American Economic Review*, *100*[1], 130–163.

Jackson, E. [2004]. *The paypal wars: battles with ebay, the media, the mafia, and the rest of planet earth*. Washington, D.C.: World Ahead Publishing.

Katana, Z., Zubcsek, P., & Sarvary, M. [2011]. Network effects and personal influences: the diffusion of an online social network. *Journal of Marketing Research*, *48*[3], 425–443.

Katz, M. & Shapiro, C. [1985]. Network externalities, competition, and compatibility. *American Economic Review*, *75*[3], 424–440.

Lee, Y. & O'Connor, G. C. [2003]. New product launch strategy for network effects products. *Journal of the Academy of Marketing Science*, *31*[3], 241–255.

Meer, J. & Rigbi, O. [2013]. The effects of transaction costs and social distance: evidence from a field experiment. *The B.E. Journal of Economic Analysis & Policy*, *13*[1].

Parker, G. & Van Alstyne, M. [2005]. Two-sided network effects: a theory of information product design. *Management Science*, *51*[10], 1494–1504.

Pizzato, L., Rej, T., Chung, T., Koprinska, I., & Kay, J. [2010]. Recon: a reciprocal recommender for online dating. *Proceedings of the fourth acm conference on recommender systems*, ACM, 207–214.

Rochet, J.-C. & Tirole, J. [2006]. Two-sided markets: a progress report. *The RAND Journal of Economics*, *37*[3], 645–667.

Rysman, M. [2004]. Competition between networks: a study of the market for yellow pages. *The Review of Economic Studies*, *71*[2], 483–512.

Suarez, F. [2005]. Network effects revisited: the role of strong ties in technology selection. *Academy of Management Journal*, *48*[4], 710–720.

Sun, M. [2007]. *Weaving a two-sided network: winning strategies for network platform competition.* VDM Publishing.

Sun, M. & Tse, E. [2009]. The resource-based view of competitive advantage in two-sided markets. *Journal of Management Studies*, *46*[1], 45–64.

Sundararajan, A. [2007]. Local network effects and complex network structure. *The BE Journal of Theoretical Economics*, *7*[1], Article 46.

Toma, C. L., Hancock, J. T., & Ellison, N. B. [2008]. Separating fact from fiction: an examination of deceptive self-presentation in online dating profiles. *Personal and Social Psychology Bulletin*, *34*[8], 1023–1036.

Tucker, C. & Zhang, J. [2010]. Growing two-sided networks by advertising the user base: a field experiment. *Marketing Science*, *29*[5], 805–814.

Wooldridge, J. [1999]. Distribution-free estimation of some nonlinear panel data models. *The Journal of Econometrics*, *90*[1], 77–97.

List of Figures

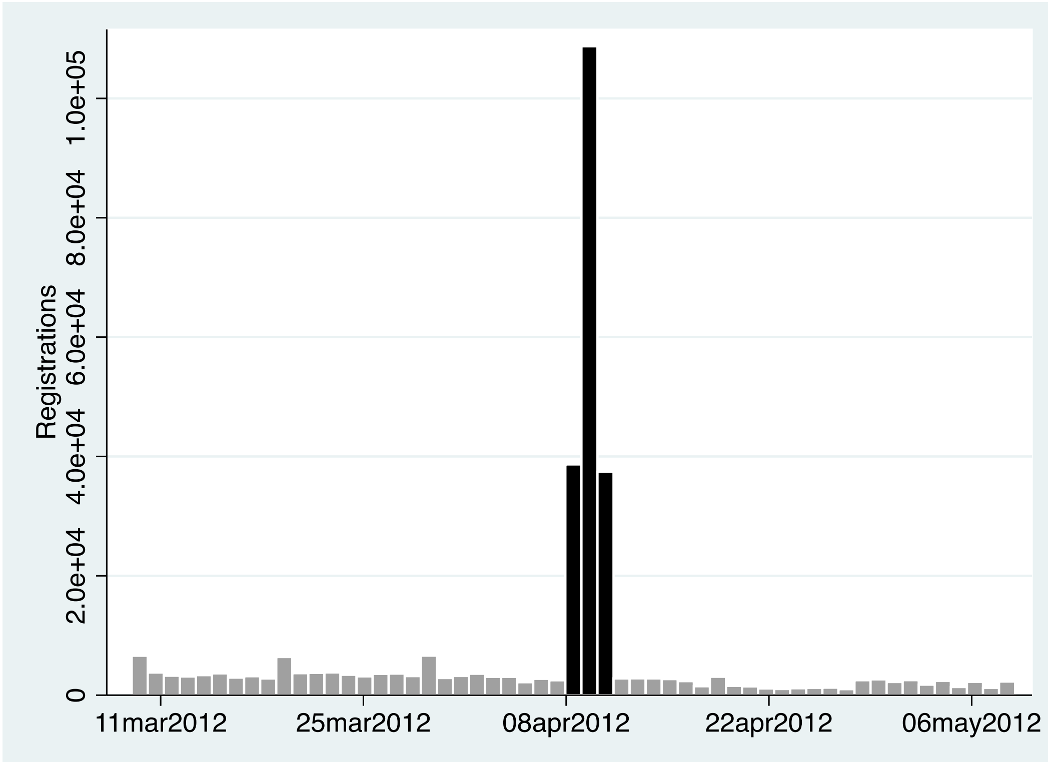

Figure 1: User Account Creation Over Time

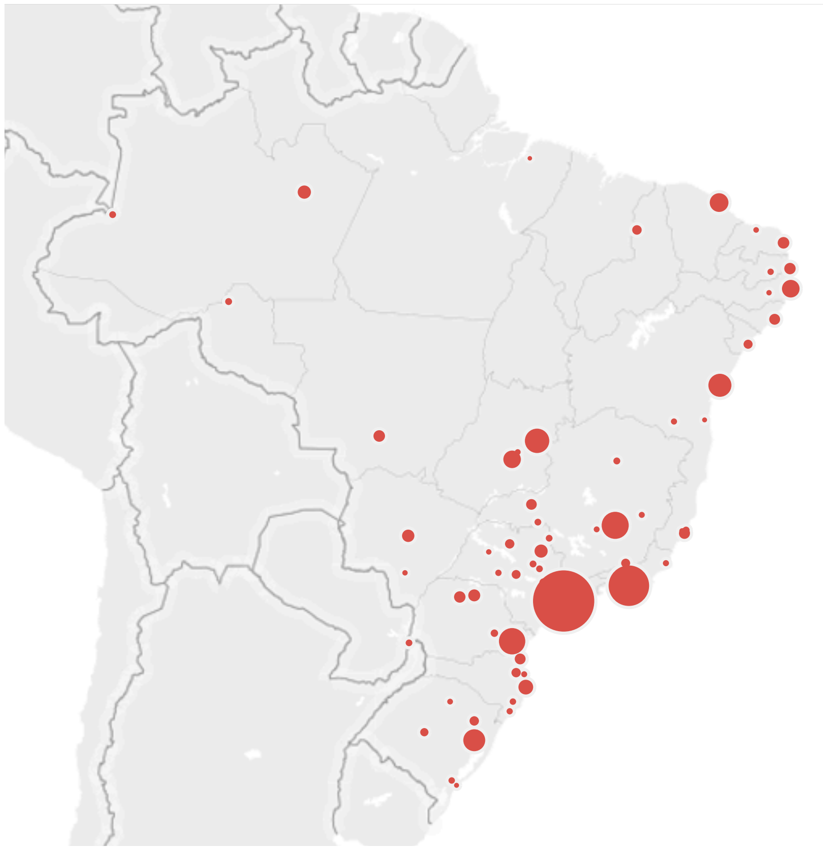

Figure 2: Geographic Distribution of Users

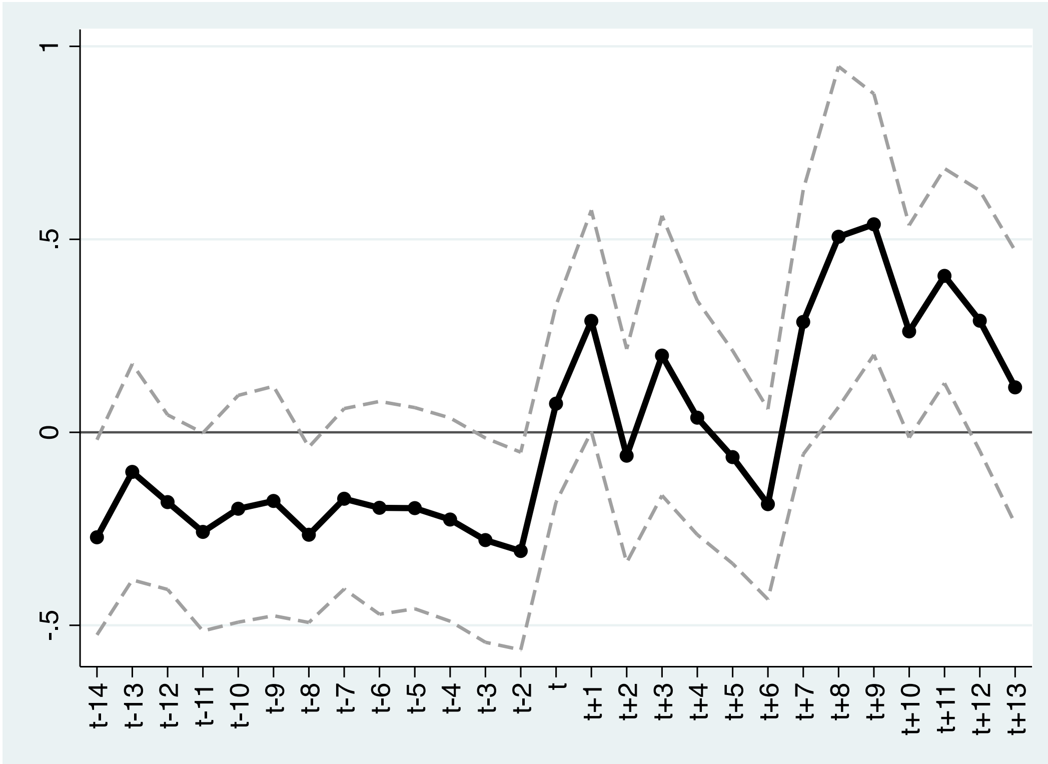

Figure 3: Poisson Relative Time Estimates Enrollments; t-1 Omitted

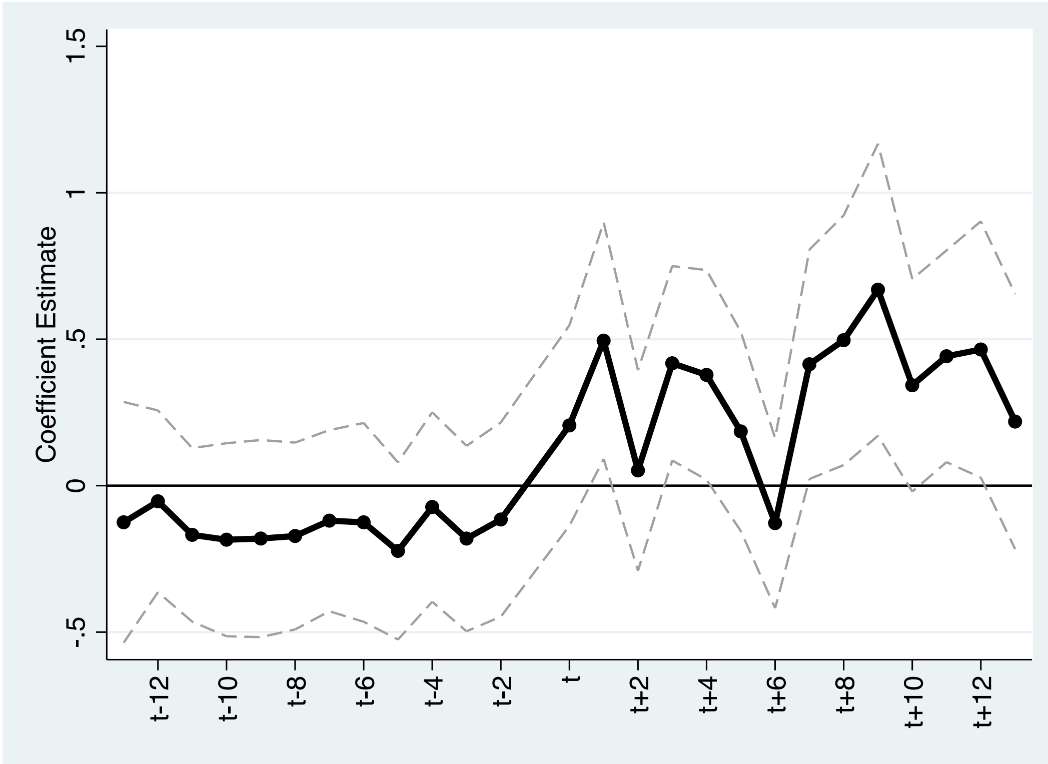

Figure 4: Poisson Relative Time Estimates for Exits; t-1 Omitted

List of Tables

Table 1: Average Daily Enrollment & Exit Over 4-Weeks Prior to Treatment

	Heterosexual						Homosexual					
		Male			Female			Male			Female	
Variable	Max	Mean	STDev	Max	Mean	STDev	Max	Mean	STDev	Max	Mean	STDev
Enrollment	302.071	13.476	31.886	36.893	1.463	3.830	14.893	0.674	1.550	13.393	0.583	1.423
Exit	7.750	0.493	0.821	34.929	1.375	3.600	14.393	0.645	1.500	12.964	0.556	1.368
Cities	120	120	120	120	120	120	120	120	120	120	120	120

Table 2:
Enrollments: Poisson FE

	Male		Female	
DV = Enrolls	4-Week	2-Week	4-Week	2-Week
Post	-0.799*** (0.035)	-0.970*** (0.043)	-0.697*** (0.042)	-0.795*** (0.050)
Orientation	2.995*** (0.032)	2.970*** (0.034)	0.920*** (0.034)	0.895*** (0.054)
Post ∗ Orientation	0.288*** (0.041)	0.467*** (0.048)	0.211*** (0.051)	0.309*** (0.050)
Observations	480	476	476	476
Cities	120	119	119	119
Wald χ^2	10,923.64 (3)	14,645.41 (3)	1,272.02 (3)	1,461.96 (3)

Notes: Robust standard errors in brackets; *** p<0.001
Some cities dropped because no variation was observed in the DV within the group

Table 3:
Enrollments: Log-OLS FE

	Male		Female	
DV = *Log*(*Enrolls*)	4-Week	2-Week	4-Week	2-Week
Post	-0.759*** (0.059)	-0.804*** (0.064)	-0.738*** (0.062)	-0.806*** (0.069)
Orientation	3.032*** (0.040)	3.018*** (0.051)	0.926*** (0.043)	0.834*** (0.057)
Post ∗ *Orientation*	0.207*** (0.062)	0.238*** (0.067)	0.169*** (0.075)	0.211*** (0.086)
Observations	469	454	463	445
Cities	120	119	119	119
R^2	0.967	0.960	0.735	0.664
F statistic	3,042.41 (3, 119)	2,103.36 (3, 118)	359.25 (3, 118)	239.78 (3, 118)

Notes: Robust standard errors in brackets; *** p<0.001
Some observations dropped when no enrollments are observed, i.e., Log of 0 is undefined

Table 4:
Exits: Poisson FE

	Male		Female	
DV = Exits	4-Week	2-Week	4-Week	2-Week
Post	-0.721*** (0.039)	-0.570*** (0.040)	-0.606*** (0.043)	-0.333*** (0.048)
Orientation	-0.268*** (0.093)	-0.238* (0.092)	0.906*** (0.033)	0.838*** (0.051)
Post ∗ Orientation	0.688*** (0.051)	0.880*** (0.069)	0.212*** (0.046)	0.199*** (0.051)
Observations	476	476	476	476
Cities	119	119	119	119
Wald χ^2	395.76 (3)	251.87 (3)	994.72 (3)	695.95 (3)

Notes: Robust standard errors in brackets; *** p<0.001, * p<0.05
Some cities dropped because no variation was observed in the DV within the group

Table 5:
Exits: Log-OLS FE

	Male		Female	
DV = Log(Exits)	4-Week	2-Week	4-Week	2-Week
Post	-0.608*** (0.059)	-0.649*** (0.066)	-0.619*** (0.068)	-0.715*** (0.071)
Orientation	-0.183*** (0.089)	-0.001 (0.093)	0.929*** (0.044)	0.838*** (0.060)
$Post * Orientation$	0.679*** (0.084)	0.563*** (0.103)	0.148* (0.074)	0.248*** (0.079)
Observations	459	434	461	445
Cities	119	119	119	119
R^2	0.178	0.214	0.700	0.642
F statistic	39.21 (3, 118)	38.11 (3, 118)	306.79 (3, 118)	171.79 (3, 118)

Notes: Robust standard errors in brackets; *** p<0.001, * p<0.05
Some observations dropped when no enrollments are observed, i.e., Log of 0 is undefined

Table 6:
Geographic Distance Moderation of Enrollment: Poisson FE

DV = Enrolls	Male	Female
Post	-0.980*** (0.045)	-0.832*** (0.054)
Orientation	2.973*** (0.035)	0.887*** (0.056)
$Post * Geo$	0.019** (0.007)	0.060*** (0.017)
$Orientation * Geo$	0.001 (0.014)	0.027 (0.020)
$Post * Orientation$	0.479*** (0.049)	0.350*** (0.049)
$Post * Orientation * Geo$	-0.018* (0.008)	-0.060* (0.025)
Observations	456	456
Cities	114	114
Wald χ^2	23,172.01 (6)	1,400.89 (6)

Notes: Robust standard errors in brackets; *** p<0.001, ** p<0.01, * p<0.05
Some cities dropped because no variation was observed in the DV
within the group or because no geographic data was available for users

Table 7:
Geographic Distance Moderation of Exit: Poisson FE

DV = Exits	Male	Female
Post	-0.888*** (0.048)	-0.726*** (0.050)
Orientation	-0.242* (0.096)	0.869*** (0.053)
$Post * Geo$	0.021** (0.008)	0.059** (0.022)
$Orientation * Geo$	-0.013 (0.028)	0.025 (0.021)
$Post * Orientation$	0.875*** (0.067)	0.340*** (0.049)
$Post * Orientation * Geo$	-0.004 (0.010)	-0.044* (0.019)
Observations	456	456
Cities	114	114
$Wald\ \chi^2$	697.27 (6)	867.96 (6)

Notes: Robust standard errors in brackets; *** p<0.001, * p<0.05
Some cities dropped because no variation was observed in the DV within the group or because no geographic data was available for users