From Strangers to Friends: Tie Formations and Online Activities in an Evolving Social Network *

Current version: September 2019

Mina Ameri† Elisabeth Honka‡ Ying Xie§

Abstract

We study how strangers become friends within an evolving online social network. By modeling the co-evolution of individual users' friendship tie formations and their concurrent online activities (product adoptions and content generation), we are able to discover important drivers underlying individuals' friendship decisions and, at the same time, to quantify the resulting peer effects on individuals' actions. We estimate our model using a novel data set capturing the continuous development of a network and users' entire action histories within the network. Our results reveal that knowledgeability of and similarity (homophily) with a potential friend are the most important drivers of friendship formation. Next, through prediction exercises, we assess how to most effectively increase website traffic in an evolving network via seeding and stimulation strategies. And lastly, in contrast to results for static networks, we find that seeding to users with the most friends is *not* the most effective strategy to increase users' activity levels in an evolving network.

Keywords: Social Network Formation, Product Adoption, User-Generated Content, Peer Effects

JEL Classification: D83, L82, M31

*We would like to thank Bryan Bollinger, Randy Bucklin, Brett Hollenbeck, Sylvia Hristakeva, Dmitri Kuksov, Xiaolin Li, Xiao Liu, B.P.S. Murthi, Ram Rao, Brian Ratchford, Peter Rossi, Upender Subramanian, seminar participants at Cornell University, Duke University, Johns Hopkins University, LBS, NYU, University of Connecticut, University of Pittsburgh, UT Austin, Tsinghua University, Virginia Tech, University of North Carolina at Chapel Hill, University of Rochester, and Jinan University and participants at the 2017 Texas Marketing Faculty Research Colloquium, the 2018 INFORMS/Marketing Science conference, 2018 CMU Conference on Digital Marketing and Machine Learning, 2019 International Forum of Marketing Science and Applications, and 2019 Marketing Dynamics Conference. The authors thank Paul Rodriguez for his assistance with optimization of our code which was made possible through the XSEDE Extended Collaborative Support Service (ECSS) program (Gitler and Klapp 2015).The authors acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing HPC resources that have contributed to the research results reported in this paper. The authors acknowledge the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562. The authors used the Bridges system, which is supported by NSF award number ACI-1445606, at the Pittsburgh Supercomputing Center (PSC). All errors are our own.

†University of Pittsburgh, mina.ameri@pitt.edu.

‡University of California Los Angeles, elisabeth.honka@anderson.ucla.edu.

§University of Texas at Dallas, ying.xie@utdallas.edu.

1 Introduction

Online social networks have become an indispensable part of many individuals' everyday lives. By the end of 2016, there were 2.3 billion active social network users worldwide with the average user having more than 5 accounts across different platforms.[1] The goal of social networks is to increase the connectedness among people, not only by facilitating more interactions among existing acquaintances and friends, but also by enabling the meeting and making of new friends in the virtual world. As a result, the number of connections a user has can grow very fast in an online social network. For example, Instagram users saw, on average, a monthly growth of 16% in the number of their followers in 2017.[2] As the size and connectedness of online social networks grow, network users are also increasingly sharing information and interacting with each other through these networks.[3]

Despite a long list of studies documenting the significant role of online social networks in facilitating information transimission, social learning, and in shaping network users' purchase and consumption related decisions,[4] very little is known about how (online) social networks develop and evolve and, in particular, how and with whom people become friends. In this paper, we study the endogenous formation of an online social network. Specifically, we model the co-evolution of individual users' friendship tie formations and their concurrent activities, i.e., product adoption and content generation, within an online social network. An intriguing aspect of making friends in many online social networks (including the one we study in this paper) is that people often do not know each other's identities in real life.[5] Thus one might ask how these strangers become friends in the first place. In these situations characterized

[1] https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/; https://insight.globalwebindex.net/social-q4-2014

[2] https://socialpilot.co/blog/151-amazing-social-media-statistics-know-2017/

[3] For example, in 2016, Facebook users shared 1.3 million pieces of content every minute and Twitter users sent 0.35 million tweets every 60 seconds (https://socialpilot.co/blog/125-amazing-social-media-statistics-know-2016/).

[4] For example: Bramoullé, Djebbari, and Fortin (2009), Duan, Gu, and Whinston (2009), Aral and Walker (2011), Katona, Zubcsek, and Sarvary (2011), Moe and Trusov (2011), Christakis and Fowler (2013), Toubia and Stephen (2013), Ameri, Honka, and Xie (2019).

[5] Other examples of such online social networks are goodreads.com, boardgamegeek.com, dpadd.com and last.fm.

by anonymity, an individual's behaviors and opinions (such as product adoptions and content generation) as observed by other people in the online environment may be among the main factors influencing friendship tie formation decisions. Therefore, it is important to take them into account when modeling tie formations in an online social network.

We draw on the social psychology literature to identify the main drivers that may impact individuals' friending decisions in a social network. The first friendship determinant is (physical) proximity. Proximity describes the phenomenon that the more people, by chance, see and interact with each other, the more likely it is that they become friends (Berscheid and Reis 1998).[6] A second determinant of friendship formation is similarity (homophily), i.e., a match between two persons' interests, experiences, backgrounds or personalities increases the chance for them to become friends (Berscheid and Reis 1998, McPherson, Smith-Lovin, and Cook 2001). To put it differently, "birds of a feather flock together" and similar people have been shown to be more likely to become friends with each other than dissimilar people (e.g., Sun and Taylor 2019). And lastly, the expertise or knowledge of a person in the subject domain increases her desirability as a potential friend. The increased desirability is due to the expected benefit from becoming friends with a person who knows many other people (Langlois et al. 2000; Tong et al. 2008) and the potential gain from information sharing and learning from a friend who is knowledgable about the subject domain (Watson and Johnson 1972; Brandtzæg and Heim 2009).

Since online social networks mostly rely on advertising revenue for profitability and thus for platform survival, it is crucial for platform owners to understand how online social networks evolve and what drives users' friendship tie formations and their concurrent online activities. This knowledge would guide them to devise effective stimulation and seeding strategies that leverage friends' social influence to increase network connectedness and the amount of in-site activities on the platform.[7] For example, previous literature has shown that seeding to more connected users, i.e., users with more friends, is the most effective strategy to encourage the

[6]The proximity effect is also sometimes called the propinquity effect (Berscheid and Reis 1998).

[7]Seeding refers to the determination of whom to target for motivational stimulation with the goal of triggering large information cascades, adoptions, or other types of actions.

diffusion of a behavior (e.g., Trusov, Bodapati, and Bucklin 2010; Hinz et al. 2011; Aral, Muchnik, and Sundararajan 2013). However, these studies investigate mature (static) social networks where the number of friendship ties remains stable. Thus their finding may or may not hold for evolving networks in which an intervention is very likely to also cause individual users to form new friendship ties and therefore increase their connectedness. As a result, to accurately evaluate the effectiveness of stimulation and seeding strategies in evolving networks, one also needs to take the endogenous formation of new friendship ties into account. Our model considers this aspect when assessing the effectiveness of such interventions.

Our paper builds on and extends Snijders, Steglich, and Schweinberger (2007). The authors develop a network co-evolution model, i.e., a stochastic model that describes both the formation of a network from an individual's perspective and the incidences of individuals' other action(s) within the network. In the model, both the network structure and individuals' actions evolve in a dynamic fashion: individuals are selected at random rates and each selected individual decides whether to make a change in her friendship ties, to perform an action, or to do neither. There are several limitations to Snijders, Steglich, and Schweinberger (2007). First, since individuals cannot change both their ties and their actions at the same point in time, simultaneous incidences of tie formation and other actions are not accounted for. Second, due to the randomness in the decision timing, the effects of time-varying behaviors (such as online activities) cannot be identified and, as a result, their effects cannot be assessed. And third, although Snijders, Steglich, and Schweinberger (2007) capture homophily by accounting for observed similarities among users when modeling tie formations, latent homophily (arising from the similarity among friends in their unobserved intrinsic preferences) remains a confounding factor that may bias the effect of friends' influence.

In this paper, we overcome these limitations by proposing a co-evolution model of individuals' concurrent decisions to both form friendship ties and to perform online activities (such as product adoptions and content generation) at each point in time. This allows us to capture the effects of any time-varying behavior while controlling for simultaneous incidences of these decisions. Furthermore, we are able to account for the latent homophily by explicitly estimat-

ing individuals' unobserved intrinsic preferences for actions absent of their friends' influence and therefore provide a cleaner identification of peer effects. We are able to do so because we observe users' actions both before and after they make friends in an online social network in our data.

More broadly, this paper is also related to work on strategic network formation in the economics literature (see e.g., Christakis et al. 2010, Snijders, Koskinen, and Schweinberger 2010).[8] Strategic network formation models take the perspective of individual actors' utility maximizations and apply game-theoretic methods to explain the evolution of a network and to allow individuals' friendship tie formation decisions to depend on the existing state of the network (e.g., Hanaki et al. 2007). It is important to note that strategic network formation models only describe friendship tie formations and *not* incidences of any other activities. While we do not use the same model and methods as the strategic network formation literature, we borrow insights from this stream literature when developing the friendship formation part of our model and we also compare our findings on the drivers of friendship formation with this stream of literature.

Furthermore, while people's activities and opinions influence the friendship ties they form, their future activities and opinions are also subject to the influence of their friends. The latter is often termed as social influence, network effects or peer effects in the literature (e.g., Sacerdote 2001; Iyengar, den Bulte, and Valente 2011; Katona, Zubcsek, and Sarvary 2011). However, the endogeneity of network formation makes it a challenging task to correctly identify peer effects (Manski 1993). Previous literature has suggested several approaches to deal with this challenge by using, e.g., instrumental variables (e.g., Bramoullé, Djebbari, and Fortin 2009; De Giorgi, Pellizzari, and Redaelli 2010), correlated group effects (e.g., Lee 2007; Lee, Liu, and Lin 2010; Ma, Krishnan, and Montgomery 2014), randomness/exogenous shocks (e.g., Sacerdote 2001; Tucker 2008), experiments (e.g., Aral and Walker 2011), individual-specific unobserved preferences (e.g., Nair, Manchanda, and Bhatia 2010; Trusov, Bodapati, and Bucklin 2010;

[8]Strategic network formation models are also known as network evolution models (Toivonen et al. 2009) or actor based models (Snijders, van de Bunt, and Steglich 2010) in the economics literature.

Ameri, Honka, and Xie 2019), and co-evolution models (e.g., Snijders, Steglich, and Schweinberger 2007; Badev 2013). In this paper, we combine the last two approaches: we account for individual-specific unobserved preferences of performing an action and for the interdependence among actions, while explicitly modeling the evolution of the network to which an individual belongs.

We obtain our data from a special interest online community for animes (Japanese cartoons) called MyAnimeList.net. This website provides a gathering place for anime fans from all over the world to interact with each other and to form friendships. Since anime fandom is a special interest and anime fans are scattered around the world, the online channel naturally becomes the main venue through which anime fans interact with each other. This implies that most users of MyAnimeList.net do not know each other before forming their friendship ties online and that the actions they observe on the website are the main drivers of their friendship decisions — making this platform an ideal environment for our research inquiry. We take advantage of this novel data set that documents both the continuous development of the network, i.e., which individuals become friends with each other and when that happens, and all users' entire activity (i.e., anime watching and posting of user-generated content (UGC)) histories on the platform. Access to these data allows us to model the friendship network development without the need to simulate the state of the network at each point in time and, as a result, to quantify the effects of users' time-varying activities on the probability that two individuals become friends.

We model the endogenous formation of a social network and the occurrence of two types of online activities, namely, product adoptions and content generation, over time within the website. More specifically, each day, a user makes three types of decisions: (i) *with whom* to become friends, (ii) whether to watch any anime, and (iii) whether to publish a UGC post. All decisions are modeled within a utility-maximizing framework. We model friendship tie formation between two individuals as non-cooperative decisions. Each individual maximizes her own friendship formation utility which depends on the proximity and similarity between the pair, and the knowledgeability of the potential friend. A friendship tie is formed if and only if both users agree to it. A user's utilities of engaging in either product adoptions or content

generation are functions of her past online activities and her friendship network which can affect her actions through peer effects.

Furthermore, to capture any common shocks unobserved by the researcher which might result in correlated unobservables, we include time fixed effects and also allow the error terms in the utility functions associated with each type of decision to be correlated with each other within a day. To tease apart homophily from peer effects, we incorporate random effect to capture individual-specific intrinsic propensities to watch animes and to post UGC. We further incorporate individual-specific propensities to make friends in the friendship formation utility again through random effect to capture any inherent cost that users incur when making friends. To recap, the three utility functions are connected in three ways: through observed variables, through correlated error terms, and through correlated individual-specific unobserved preferences.

Our results for friendship tie formation reveal that a focal user is more likely to become friends with knowledgable and similar users, i.e., users who are experienced and active (watch many animes and publish a lot of posts), have many friends, and many friends in common with the focal user. Comparing the marginal effects of a potential friend's friendship network and her in-site activity levels in driving friendship formation, we find the former to be more important than the latter. In addition, even in (anonymous) online networks having similar demographic characteristics and close geographical proximity matters, i.e., users with the same (self-reported) age, gender, and location are more likely to become friends. Our results for in-site activities, i.e., product adoptions and the production of UGC, reveal significant positive peer and spill-over effects on the focal user: while having more friends does not necessarily make a user more active, having more active friends does increase a user's activity level due to the positive social influence.

We use our results to conduct a number of predictive exercises to assess the effectiveness of various seeding and stimulation strategies in increasing users' in-site activity levels in an evolving online social network. We find that stimulating UGC production is the most effective site-wide intervention which leads to the highest overall site traffic and the largest number

of active users. When evaluating the effectiveness of various seeding strategies, we find that seeding to the most active users outperforms seeding strategies targeting the least active users. Our results also shows that contrary to the findings in previous studies investigating static networks (e.g., Trusov, Bodapati, and Bucklin 2010; Hinz et al. 2011; Aral, Muchnik, and Sundararajan 2013), seeding to well-connected users, i.e. users with many friends, is not the most effective strategy to increase users' activities on the platform in evolving networks; seeding to users who watch many animes or make many UGC posts is more effective. And lastly, we find that not accounting for the endogenous network formation in an evolving network when assessing the effectiveness of seeding strategies leads, on average, to an underestimation of seeding effectiveness by 30%.

The contribution of this paper is two-fold. First, our paper is the first systematic investigation in marketing that theorizes and quantifies the importance of various drivers behind friendship formation decisions among strangers in an online environment. Our finding, that all three friendship drivers, i.e., proximity, similarity, and knowledgeability, matter, provides strong support for the popular practice of recommending people with similar traits as potential connections by large social networks such as Facebook or LinkedIn (see also Sun and Taylor 2019). And second, to the best of our knowledge, the current paper is among the first ones to model the interdependent dynamics between network formation and time-varying online activities, i.e., how tie formations, product adoptions, and UGC productions interact and influence each other, within an evolving online social network. This knowledge is crucial for network platform managers when developing effective seeding and stimulation strategies to increase user engagement – and thus the probability of platform survival – in an evolving online social network.

The remainder of this paper is organized as follows: In the next section, we describe our data. In Sections 3 to 5, we introduce our model, estimation approach, and identification strategy. We present and discuss our estimation and simulation results in Sections 6 and 7. In the following section, we examine limitations of the current work and opportunities for future research. Finally, we conclude by summarizing our findings in Section 9.

2 Data

Our data come from MyAnimeList.net. This website is a consumption-related online community where online interactions are based upon shared enthusiasm for a specific consumption activity (Kozinets 1999). MyAnimeList.net was created to allow anime (Japanese cartoons) fans to gather and to share their excitement and opinions about animes. The website has developed into one of the most popular platforms for anime fans over the years. Users of the website create a profile page when they join the website. On their profile page, users can share some information about themselves (e.g., age, gender, or location) and create a list of the animes they have watched or are watching (which we refer to as "watch list" throughout this paper). The website also provides a forum where users can share information and exchange opinions about animes with each other. In addition, users have the option to become friends, which makes it easier for them to access their friends' pages and to be notified about their friends' activities, similar to bookmarking and RSS functions in web browsers.

Anime fandom is a special interest and not very common. As a result, fans use special interest communities such as MyAnimeList.net to find and connect with other fans. This implies that most users of MyAnimeList.net meet other users for the first time on the website and their interactions are happening within the website. Furthermore, this website is a worldwide community and attracts users from different cities and countries around the globe. About half of the users reveal their locations on their profile pages. We can see that users frequently become friends with other users from different countries. This observation further validates our assumption that meetings and interactions among the users are mostly confined to the platform.

2.1 Estimation Sample

The website was first started in 2004, however, as a private domain. On April 6, 2006, it was moved to a public domain and began to take its current shape. At that point in time, the website had about 300 users. After its transfer to a public domain, the number of members

started to grow quickly (see Figure 1). About a year later, on March 16, 2007, the function of forming friendships was added. At that point in time, the website had about 450 members and this number grew rapidly to 2,700 at the beginning of July 2007 and to 11,500 by the end of 2007.

==========

Insert Figure 1 about here

==========

We focus on users who joined the website in the second half of 2007, mainly for two reasons. First, users who joined the website before March 2007 are likely to add each other as friends based on past interactions. To put it differently, had they had the option of adding friends before, they would have done so. And second, it might have taken existing members some time to learn about this new feature. Therefore, we start our study period about three months after the introduction of the friending function.

Studying daily friendship formation among *all* users who joined between July and December 2007 is, however, computationally impossible since the data set would include over 7 billion pair-day observations. One potential solution is to shorten the observation period. However, this approach would result in insufficient variation in the dependent variables. Figure 2 shows the distributions of the number of days between activities of each type. In about 50% of the cases, users add a friend and publish a post more than a month after their last action of the same kind. In 40% of the cases, users watch an anime more than a month after the last watched anime.

==========

Insert Figure 2 about here

==========

A second potential solution is to aggregate the observations to the weekly level. However, aggregation leads to information loss in both the dependent and independent variables. On top of that, we observe that users make more than one friend in a week in more than 20% of the cases. Anime watching and content generation also happen more than once a week in

about 25% and 10% of the cases, respectively. Consequently, aggregating data to the weekly level would force us to model the sequence of users' actions within a week. As a result, similar to previous research (e.g., Snijders, Steglich, and Schweinberger 2007), the degree to which we could capture the effects of time-varying activities of users on their friendship making would be restricted.

A third potential solution is to sample from the network. We implement this solution by drawing a random sample of 400 users ("core users") out of about 8,800 users who joined the website in the second half of 2007. We then include all of the core users' friends in our estimation sample. Note that friends of the 400 core users can also be core users or they can be users not in the core. Figure 3 visualizes our sampling strategy. For example, in Figure 3, users 1 and 2 are both core users. User 2 is friends with user 1 who is another core user and with users 3 and 4 who are not core users. We term all users who are *not* core users themselves, but friends with a core user "non-core users." This second set of users includes 986 users.[9] Thus, our estimation sample contains 1,386 users (400 core users and 986 non-core users).

==========

Insert Figure 3 about here

==========

In the estimation, we model all anime adoptions and all UGC production activities for both core and non-core users. For friendship formation, we model all potential ties among core users (e.g., between users 1 and 2 in Figure 3), all potential ties among core and non-core users (e.g., between users 1 and 3 in Figure 3), and all potential ties among non-core users (e.g., between users 3 and 4 in Figure 3). However, we assume that non-core users' friendship formations with outside users, i.e., users who are neither core nor non-core users, are exogenous. This exogeneity assumption means that we do not model them becoming friends, but we do take their friendship into account when creating friendship-related independent variables.

Given that activities of users in the three areas can be correlated, missing a portion of the network formation for these 986 non-core users can lead to bias in our estimates of the

[9]The set of non-core users includes 732 users who joined after July 2007 and 254 users who joined before July 2007.

friendship decisions. Note that since we model all actions of core and non-core users and incorporate the effects of friendship ties with outside users on non-core users' anime adoptions and content generation, this bias is mainly a concern for the estimation of the friending decision. To alleviate this concern, we estimate separate coefficients for the 400 core users (for whom we have their complete tie formations) and for the 986 non-core users (for whom we do not model the portion of the friendship network that includes users outside of our sample).[10]

The observation period is 184 days between July 1, 2007, and December 31, 2007. However, we have fewer observations for users who joined after July 1, 2007. On average, we observe each user for 140 days.

2.2 Data Description

Within our sample of 1,386 users, we observe 5,038 ties out of 947,155 possible ties being formed during the observation period and about 68 million daily observations of possible pairs. Figure 4 shows the states of the network for snapshots at days 1, 60, 120, and 184. The nodes represent individual users in the network and the links between nodes represent friendships ties. Furthermore, the color of a node reflects the quantity of a user's UGC production and the size of a node reflects the number of animes a user watched. The color of the nodes becomes darker as users publish more posts on the website and the size of the nodes increases as users watch more animes. As expected, the nodes become darker, bigger, and more connected over time. The larger and darker nodes are also associated with more links, suggesting interdependence between users' friendship formation and other activities.

[10]Snowball sampling is a common sampling method that has been used extensively in the network literature. However, a concern with this sampling method is the oversampling of active users. In our case, this concern is alleviated by drawing a random sample of 400 core users and controlling for the unobserved heterogeneity among these users. Furthermore, we estimate separate coefficients for the core and non-core users in the friendship utility. And lastly, since we draw a random sample of users and include the friendship network of those users in our sample, for a focal user, the other 399 randomly drawn users and their friends who are not friends with that focal user are a random representative sample of the whole network. As a result of this randomness and the inclusion of separate parameters for core and non-core users, we believe that our estimates are unbiased for the 400 core users.

==========

Insert Figure 4 about here

==========

Table 1 summarizes key statistics of our data. In terms of demographics, 78% of users report their age and are, on average, 19 years old. 93% of users report their gender with 39% being female and 54% being male.

==========

Insert Table 1 about here

==========

Figure 5 demonstrates how average activity levels of users who joined the website in the second half of 2007 change over time from the day they joined the website. Figures 5a shows a decreasing trend in making new friendship ties. Since one of the benefits of having friends are reduced costs associated with learning about the website and new animes, users are more likely to add friends shortly after they join the website. Figures 5b and 5c show the activity trends for anime watching and content generating. Both graphs reveal a rather constant trend over time.[11]

==========

Insert Figure 5 about here

==========

Users can engage in multiple types of activities simultaneously. On average, our core users have 2.6 active days in terms of friend adding, 9.2 active days in terms of anime watching, and 2.5 active days in terms of post writing (see Table 1). In total, they have 12.6 days in which they participate in at least one type of the three activities. To put it differently, on average, core users are active on about 18% of the days during the study period. Figure 6 shows a Venn diagram of the joint probabilities of each type of activity conditional on engaging in at least one type of activity. Users are active in only one area in 84.16% of the cases. Users are active

[11]Note that the high number of animes shortly after joining is mainly due to users adding animes that they watched before joining the website to their watch lists.

in two and three of the areas of interest in 14.79% and 1.05% of cases, respectively. We observe a similar pattern for non-core users albeit with higher average activity levels in all three types of activities (see lower half of Table 1).

==========

Insert Figure 6 about here

==========

Lastly, Figure 7 shows histograms of individual users' daily activity intensities conditional on them being active. In more than 80% of the cases, users add only one friend on an active day. Similarly, in about 60% of the cases, users watch only one anime per active day. However, the content generation intensity is higher: users publish one post per active day in about half of the cases and publish 2 or 3 posts per active day in about 20% and 10% of the cases, respectively. Based on these data patterns, we make the simplifying assumption to model anime watching and content generation as binary indicator variables, i.e., we model whether a user watches an anime or publishes a post, but not the number of animes watched or posts published by a user, in a given day.[12]

==========

Insert Figure 7 about here

==========

3 Model

In this section, we describe how we jointly model a user's decisions to form friendship ties, adopt animes, and generate content. For friendship tie formation, we model whether and with whom users become friends, while for anime adoption and content generation, we only model users' decision whether to participate in the activity.

[12]Since we model the decision of a user to become friends with each of the other users as separate independent decisions, even if users make more than one friend in a day, our model captures that.

3.1 Tie Formation

We start by describing how we model tie formations among users over time. In each time period (day), a user makes decisions whether to become friends with any other user with whom she is not friends yet. Since there are usually many users with whom the individual is not friends yet, at each point in time, a user can become friends with multiple users. Note that we model a user's tie formation decisions for each possible friendship pair in each time period and not whether a user makes a friend in a time period.

Suppose the website contains $i = 1, \ldots, N$ users and these users can become friends with other users during $t = 1, \ldots, T$ time periods.[13] Let M denote the adjacency matrix of the network which shows the stati of ties between each pair of individuals i and j with $j = 1, \ldots, N$ and $i \neq j$. m_{ijt} equals 1 if i and j are friends at time t and 0 otherwise. Ties in the network are bi-directional and symmetric, i.e., $m_{ijt} = m_{jit}$. Furthermore, both users have to agree to become friends. In our data, we do not observe users' requests for friendship with other users, only the formation of ties upon mutual agreement. Therefore our model describes the decision of both users to become friends regardless of who first requested the friendship.[14]

The decision of two users to become friends depends on the utilities both individuals derive from becoming friends (see e.g., Christakis et al. 2010). User i's utility of becoming friends with individual j in time period t, U^m_{ijt},[15] is given by

$$U^m_{ijt} = f(\mathcal{P}^m, \mathcal{R}^m, \mathcal{K}^m, C^m, \epsilon^m) \tag{1}$$

where $\mathcal{P}^m$ describes the proximity between user i and user j, $\mathcal{R}^m$ captures the similarity (homophily) between user i and user j, $\mathcal{K}^m$ describes the knowledgeability of user j, and C^m is a set of control variables. ϵ^m captures the part of the utility of user i at time t that is observed

[13] Note that t is the calendar day and not the day since a user joined the website.

[14] We do not model the dissolution of friendship ties, i.e., once users become friends, they stay friends. This is due to a limitation of our data: if two users "unfriend" each other, they appear as non-friends. We believe unfriending is not a common action among users. Therefore, we view not modeling friendship dissolution as a minor limitation.

[15] In the following, we use the superscript m to refer to variables associated with friendship utility. We suppress the subscripts for m for readability.

by the user but not by the researcher.

In characterizing user i's friendship utility, we build on the social psychology literature. The first friendship determinant is (physical) proximity $\mathcal{P}^m$, which refers to the phenomenon that the more people see and interact with each other by chance, the more likely it is that they become friends (Berscheid and Reis 1998). Proximity increases the chance of forming a friendship due to familiarity or the mere exposure effect: the more we see certain people, the more familiar they become and the more likely friendship is to bloom (Bornstein 1989, Griffin and Sparks 1990, Moreland and Beach 1992). A second determinant of friendship formation is similarity (homophily) $\mathcal{R}^m$, i.e., a match between two potential friends' interests, experiences, backgrounds or personality (Berscheid and Reis 1998, McPherson, Smith-Lovin, and Cook 2001). To put it differently, "birds of a feather flock together," people with similar traits are also more likely to become friends. And lastly, in our empirical context of an online social network for anime fans, the expertise and knowledge $\mathcal{K}^m$ of a user (about animes in general and about the platform in particular) increases the desirability of forming a potential friendship with the user. It encompasses the utility gained from becoming friends with a user who knows many other users (Langlois et al. 2000; Tong et al. 2008) and the utility gained from information sharing and learning from a friend who is knowledgable about animes (Watson and Johnson 1972; Brandtzæg and Heim 2009).

We assume that individuals have myopic utilities, i.e., individuals do not anticipate future states of the network and only care about the current state of the network when deciding to form a tie (Snijders, Steglich, and Schweinberger 2007, Christakis et al. 2010). They do not take future links of themselves or the other party into consideration when making the decision to become friends. The assumption of myopic utility is appropriate for large networks in which individuals can meet numerous other individuals at each point in time and the number of future states of the network increases exponentially. Furthermore, users are not limited in the number of ties they can make in an online friendship network. As a result, users independently and non-strategically form ties if the utility of such ties is positive.

We model the tie formation between two users as a non-cooperative decision (e.g., Christakis

et al. 2010), i.e., each pair's decision to become friends only depends on their own friendship utilities and is conditionally independent from friendship decisions of other pairs of users. A tie between user i and user j is formed if and only if both parties decide that it is beneficial for them to become friends, i.e.,

$$m_{ijt} = 1 \qquad \text{iff} \qquad U^m_{ijt} > 0 \;\text{ and }\; U^m_{jit} > 0. \tag{2}$$

3.2 Product Adoption and Content Generation

Next, we describe how we model a user's activities on the website. We study the incidence of users' activities in two broad areas, namely, product adoption and content generation. Let U^{pa}_{it} denote user i's utility from watching an anime at time t and U^{cg}_{it} denote user i's utility from producing content at time t. Then U^{pa}_{it} and U^{cg}_{it} are given by

$$\begin{aligned} U^{pa}_{it} &= g(\mathcal{F}^{pa}, \mathcal{A}^{pa}, C^{pa}, \epsilon^{pa}), \\ U^{cg}_{it} &= h(\mathcal{F}^{cg}, \mathcal{A}^{cg}, C^{cg}, \epsilon^{cg}), \end{aligned} \tag{3}$$

respectively. Both utilities depend on a user's friendship network denoted by $\mathcal{F}^{pa}$ and $\mathcal{F}^{cg}$, respectively, to capture peer effects; a user's past actions denoted by $\mathcal{A}^{pa}$ and $\mathcal{A}^{cg}$, respectively, to capture state dependence; and a set of control variables denoted by C^{pa} and C^{cg}, respectively. ϵ^{pa} and ϵ^{cg} denote the parts of the utilities that are observed by the user but not by the researcher.

3.3 Integrating All Actions

We now present the full model integrating user i's actions in all three areas:

$$\begin{aligned} U^m_{ijt} &= f(\mathcal{P}^m, \mathcal{R}^m, \mathcal{K}^m, C^m, \epsilon^m) \qquad \forall j = 1 \ldots N,\; i \neq j \\ U^{pa}_{it} &= g(\mathcal{A}^{pa}, \mathcal{F}^{pa}, C^{pa}, \epsilon^{pa}) \\ U^{cg}_{it} &= h(\mathcal{A}^{cg}, \mathcal{F}^{cg}, C^{cg}, \epsilon^{cg}). \end{aligned} \tag{4}$$

Some variables, unobserved by the researcher, might influence more than one type of decision an individual user makes.[16] For example, a user might be traveling and, as a result, not spending any time on the website, i.e., be inactive in all three areas. To accommodate the simultaneous co-occurrence of activities user i makes at time t, we allow the three error terms in Equation (4) to be correlated, i.e.,

$$G = \begin{bmatrix} 1 & \rho_{m,pa} & \rho_{m,cg} \\ \rho_{pa,m} & 1 & \rho_{pa,cg} \\ \rho_{cg,m} & \rho_{cg,pa} & 1 \end{bmatrix}. \tag{5}$$

3.4 Utility Specifications

In this section, we present detailed utility specifications for the specific context of our data. We model the utility user i receives from forming a tie with user j as

$$U^m_{ijt} = \tilde{\alpha}^m_{it} + \mathcal{R}^m_{ij,t-1} + \delta^m \mathcal{P}^m_{ij,t-1} + \gamma^m \mathcal{K}^m_{j,t-1} + \lambda^m C^m_{ijt} + \epsilon^m_{it}. \tag{6}$$

$\tilde{\alpha}^m_{it}$ captures user i's intrinsic preference for making friends at time t, i.e., the net of user i's benefit and cost of making friends at that point in time. As revealed in Figure 5, users newly joining the website are more likely to add friends compared to users who have already been members of the website for a longer time. This is likely due to having friends in the beginning reducing learning costs associated with navigating the website. Note that $\tilde{\alpha}^m_{it}$ does not depend on j, i.e., is identical across all potential friends. We model $\tilde{\alpha}^m_{it}$ as follows:

$$\tilde{\alpha}^m_{it} = \alpha^m_i + \kappa^m_1 W_{it}$$

where α^m_i is user i's time-invariant tendency to have few or many friends and follows a normal distribution with mean $\bar{\alpha}^m$ and standard deviation σ_{α^m}. W_{it} denotes the length of time (in days) user i has been a member of the website and κ^m_1 captures how the net of benefit and cost of forming friendship ties changes with membership length.

[16] If we were to assume that the decision a user makes regarding one action is independent of the user's decision regarding actions in the other areas, each of the decisions in the three areas could be estimated separately.

$\mathcal{R}^m_{ij,t-1}$ is tie-specific and captures the similarity (homophily) between individual i and individual j. We model $\mathcal{R}^m_{ij,t-1}$ as follows:

$$\mathcal{R}^m_{ij,t-1} = \kappa^m_2 \mathsf{R}^m_{ij,t-1} + \zeta^m_{ij}$$

where $\mathsf{R}^m_{ij,t-1}$ captures observed similarity (homophily) and includes the number of common friends, the number of common animes, and demographic similarity between user i and user j in terms of age and gender. Providing such demographic information is optional for users. However, the presence of such information may signal honesty and thus increase the credibility and perceived utility gained from forming a friendship (Lampe, Ellison, and Steinfield 2007). Thus $\mathsf{R}^m_{ij,t-1}$ also includes two dummy variables that indicate whether age and gender information of both individual i and individual j are available. ζ^m_{ij} captures the *unobserved* similarity (latent homophily) between users i and j. Note that $\zeta^m_{ij} = \zeta^m_{ji}$. The unobserved similarity (latent homophily) ζ^m_{ij} follows a normal distribution with mean 0 and standard deviation σ_{ζ^m}.

$\mathcal{P}^m_{ij,t-1}$ captures (physical) proximity and is operationalized as a dummy variable indicating whether user i and user j live in the same country.[17] Previous research (e.g., Mazur and Richards 2011, Amichai-Hamburger, Kingsbury, and Schneider 2013) has shown that – even in online social networks – proximity matters.

$\mathcal{K}^m_{j,t-1}$ describes user j's knowledgeability as a potential friend and only depends on j's attributes. We operationalize $\mathcal{K}^m_{j,t-1}$ as the (cumulative) number of friends user j has, the (cumulative) number of animes user j has adopted, and the (cumulative) number of posts user j has published in the UGC part of the website, by time $t-1$. The first variable describes the utility gained from becoming friends with a popular user, i.e. a user who knows many other users (Langlois et al. 2000; Tong et al. 2008), while the latter two variables represent the utility gained from information sharing and learning from friends who are knowledgeable about animes (Watson and Johnson 1972; Brandtzæg and Heim 2009).

C^m_{ijt} contains several variables whose effects we control for. First, we include a weekend

[17] We also include a dummy variable indicating if the geographic location information of both users was not available.

dummy. Second, we also include a dummy variable indicating whether user j was active on the platform during the previous week. This variable captures user j's visibility.[18] Third, we include time fixed effects to address common shocks that might result in correlated unobservables affecting friending decisions actions across users. We operationalize these time fixed effects as week dummies.[19] And fourth, we also include a dummy variable indicating whether user i is a non-core user and a dummy variable indicating whether user i joined the website before July 2007. Lastly, we assume that ϵ_{it}^{m} follows a normal distribution with a correlation matrix as specified in Equation (5).

User i's utility from watching an anime, U_{it}^{pa}, is given by

$$U_{it}^{pa} = \alpha_i^{pa} + \beta^{pa}\mathcal{F}_{i,t-1}^{pa} + \gamma^{pa}\mathcal{A}_{i,t-1}^{pa} + \lambda^{pa}C_t^{pa} + \epsilon_{it}^{pa} \tag{7}$$

where α_i^{pa} represents user i's intrinsic tendency to watch animes and is assumed to follow a normal distribution with mean $\bar{\alpha}^{pa}$ and standard deviation $\sigma_{\alpha^{pa}}$. $\mathcal{F}_{i,t-1}^{pa}$ captures the effects of user i's friendship network on user i's actions. It includes user i's total number of friends by time $t-1$, the number of animes watched by all of user i's friends in time $t-1$, and the number of posts written by all of i's friends in time $t-1$. Previous literature has shown that having more friends might directly affect the level of social activities of network users (Toubia and Stephen 2013; Shriver, Nair, and Hofstetter 2013).[20] In addition, the number of animes watched by all of user i's friends captures the direct influence of friends' activities on user i's product adoptions, while the number of posts written by all of user i's friends reflects the spill-over effect of friends' activities in post publishing on user i's activity in anime watching.

$\mathcal{A}_{i,t-1}^{pa}$ captures state dependence in anime watching and is operationalized as a dummy

[18] In addition to user i's preference for friendship with user j, both $\mathcal{K}_{j,t-1}^{m}$ and $\mathcal{R}_{ij,t-1}^{m}$ also capture the degree to which user j is visible to user i, i.e., we follow the conventional approach in the choice model literature and model the combined effect of visibility and preference in the utility.

[19] While it would be desirable to include daily dummy variables, for computational reasons (see Section 4), we are not able to do so as the number of additional parameters to be estimated ($552 = 184$ days $\times 3$ activities) would be too large and the estimation would take a long time (i.e., several months) to converge.

[20] A potential explanation for this effect is the image or prestige utility users gain from performing social activities within a network. Toubia and Stephen (2013) find that, aside from the intrinsic utility users derived from posting on social media, the image these activities create for users also motivated them to perform these activities. They also found that image-related utility was more dominant for users with more friends.

variable which equals 1 if user i watched an anime at $t-1$ and 0 otherwise. Furthermore, C_t^{pa} contains several variables whose effects we control for. It includes a weekend dummy and week fixed effects. And lastly, we assume that ϵ_{it}^{pa} is normally distributed with a correlation matrix as specified in Equation (5).

Similarly, user i's utility from writing a post, U_{it}^{cg}, is given by

$$U_{it}^{cg} = \alpha_i^{cg} + \beta^{cg}\mathcal{F}_{i,t-1}^{cg} + \gamma^{cg}\mathcal{A}_{i,t-1}^{cg} + \lambda^{cg}C_t^{cg} + \epsilon_{it}^{cg} \tag{8}$$

where α_i^{pa} is user i's intrinsic tendency to produce content and follows a normal distribution with mean $\bar{\alpha}^{cg}$ and standard deviation $\sigma_{\alpha^{cg}}$. $\mathcal{F}_{i,t-1}^{cg}$ captures the effects of user i's friendship network on user i's actions, and is defined in a similar manner as in Equation (7): it includes user i's total number of friends by time $t-1$, the number of animes watched by all of i's friends in time $t-1$, and the number of posts written by all of user i's friends in time $t-1$.

$\mathcal{A}_{i,t-1}^{cg}$ represents user i's past activities and contains two variables: a dummy variable capturing state dependence in UGC posting behavior and the (cumulative) number of animes watched by user i by time $t-1$. User i's past anime watching behavior may influence her posting decisions because a user who watches more animes is likely to have more things to write about. As in the previous equation, C_t^{cg} contains a weekend dummy and week fixed effects. And lastly, ϵ_{it}^{cg} follows a normal distribution with a correlation matrix as specified in Equation (5).[21]

4 Estimation

Given the conditional independence assumption of user i's decision to become friends with each user j (as discussed in Section 3.1) and given the need for mutual agreement to become friends, the probability of a tie forming between individual i and individual j is given by

$$\Pr\left(m_{ijt} = 1\right) = \Pr\left(U_{ijt}^{m} > 0\right) \cdot \Pr\left(U_{jit}^{m} > 0\right). \tag{9}$$

[21] All continuous variables in the three utility functions are incorporated in the form of natural logarithms.

Then the likelihood of user i becoming friends with user j at time t is given by

$$L^{m}_{ijt\,|\,\alpha_i,\alpha_j,\zeta_{ij},\epsilon_{it},\epsilon_{jt}} = \left[\left[\Pr\left(m_{ijt}=1\right)\right]^{m_{ijt}}\left[1-\Pr\left(m_{ijt}=1\right)\right]^{1-m_{ijt}}\right]^{1-m_{ij,t-1}}, \tag{10}$$

where $\zeta_{ij} = \zeta^{m}_{ij}$, $\alpha_i = \{\alpha^{m}_i, \alpha^{pa}_i, \alpha^{cg}_i\}$ and α_j is defined similarly.[22] Note that $L^{m}_{ijt\,|\,\alpha_i,\alpha_j,\zeta_{ij},\epsilon_{it},\epsilon_{jt}}$ conditions on the two users not being friends before time t through the exponent $1 - m_{ij,t-1}$.

The likelihoods for the other two types of activities, i.e., product adoption and content generation, at time t are given by

$$\begin{aligned} L^{pa}_{it\,|\,\alpha_i,\epsilon_{it}} &= \left[\Pr\left(A^{pa}_{it}=1\right)\right]^{A^{pa}_{it}}\left[1-\Pr\left(A^{pa}_{it}=1\right)\right]^{1-A^{pa}_{it}} \\ L^{cg}_{it\,|\,\alpha_i,\epsilon_{it}} &= \left[\Pr\left(A^{cg}_{it}=1\right)\right]^{A^{cg}_{it}}\left[1-\Pr\left(A^{cg}_{it}=1\right)\right]^{1-A^{cg}_{it}} \end{aligned} \tag{11}$$

where A^{pa}_{it} and A^{cg}_{it} indicate the incidence of an acitivity – anime watching and UGC production, respectively – of user i in time period t.

Then the joint likelihood of user i's actions at time t is denoted by

$$\begin{aligned} L_{it\,|\,\alpha_i,\alpha_j,\zeta_{ij},\epsilon_{it},\epsilon_{jt}} = &\left[\Pr\left(A^{pa}_{it}=1\right)\right]^{A^{pa}_{it}}\left[1-\Pr\left(A^{pa}_{it}=1\right)\right]^{1-A^{pa}_{it}} \\ &\cdot\left[\Pr\left(A^{cg}_{it}=1\right)\right]^{A^{cg}_{it}}\left[1-\Pr\left(A^{cg}_{it}=1\right)\right]^{1-A^{cg}_{it}} \\ &\cdot\prod_{j=i+1}^{N}\left[\left(\Pr\left(m_{ijt}=1\right)\right]^{m_{ijt}}\left[1-\Pr\left(m_{ijt}=1\right)\right]^{1-m_{ijt}}\right]^{1-m_{ij,t-1}} \qquad i\neq j. \end{aligned} \tag{12}$$

Note that α_j and ϵ_{jt} also enter the above equation for user i because they are part of the probability of user i and user j becoming friends at time t (see Equation (9)).

The full likelihood can be written as

$$\begin{aligned} L = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\prod_{t=1}^{T}\prod_{i=1}^{N} &\left[\Pr\left(A^{pa}_{it}=1\right)\right]^{A^{pa}_{it}}\left[1-\Pr\left(A^{pa}_{it}=1\right)\right]^{1-A^{pa}_{it}} \\ &\cdot\left[\Pr\left(A^{cg}_{it}=1\right)\right]^{A^{cg}_{it}}\left[1-\Pr\left(A^{cg}_{it}=1\right)\right]^{1-A^{cg}_{it}} \\ &\cdot\prod_{j=i+1}^{N}\left[\left[\Pr\left(m_{ijt}=1\right)\right]^{m_{ijt}}\left[1-\Pr\left(m_{ijt}=1\right)\right]^{1-m_{ijt}}\right]^{1-m_{ij,t-1}} d\epsilon d\alpha d\zeta \end{aligned} \tag{13}$$

[22] Note that we allow the individual-specific random effects α_i to be correlated across the three activities, i.e., we estimate a full covariance matrix Σ^{α} for the individual-specific intrinsic propensities.

and the log likelihood of the model is given by

$$LL = \log \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \prod_{t=1}^{T} \prod_{i=1}^{N} \left[\Pr\left(A_{it}^{pa} = 1\right)\right]^{A_{it}^{pa}} \left[1 - \Pr\left(A_{it}^{pa} = 1\right)\right]^{1-A_{it}^{pa}}$$
$$\cdot \left[\Pr\left(A_{it}^{cg} = 1\right)\right]^{A_{it}^{cg}} \left[1 - \Pr\left(A_{it}^{cg} = 1\right)\right]^{1-A_{it}^{cg}}$$
$$\cdot \prod_{j=i+1}^{N} \left[\left[\Pr\left(m_{ijt} = 1\right)\right]^{m_{ijt}} \left[1 - \Pr\left(m_{ijt} = 1\right)\right]^{1-m_{ijt}}\right]^{1-m_{ij,t-1}} d\epsilon d\alpha d\zeta. \tag{14}$$

We estimate our model using Simulated Maximum Likelihood (SMLE). To estimate the full covariance matrix of user random effects, Σ^{α}, we take 30 random draws from a standard normal distribution for each user and each activity, estimate the parameters of the Cholesky decomposition of the covariance matrix, and convert the Cholesky decomposition into a covariance matrix. To estimate the standard deviation of the pair-specific random term, σ_{ζ}^{m}, we take 30 random draws from a standard normal distribution for each pair of users, and directly estimate the parameter for the logarithm of the standard deviation. To estimate the correlation matrix of the three error terms, G, we take 30 random draws from a standard normal distribution for each user and each activity in each time period, estimate the off-diagonal elements of a Cholesky decomposition after setting the diagonal elements to 1, convert the Cholesky decomposition to a covariance matrix, and, lastly, the covariance matrix into a correlation matrix. This procedure allows us to estimate the elements of the error correlation matrix without putting restrictions on specific parameters (see Li, Honka, and Chintagunta 2018 for details). To calculate standard errors of the parameter estimates, we use the BHHH estimator, i.e. the outer product of the gradient, instead of the numerical Hessian (Berndt et al. 1974).

For computational reasons, the conventional approach of estimating a model via MLE and SMLE involves taking the logarithm of the model likelihood to convert an extremely-small-in-value product of probabilities to a sum of the logarithms of these probabilities. This approach cannot be applied to the likelihood of our model for three reasons. First, recall that, at any time t, the error terms in the three utility functions are correlated (see Equation (5)).[23] Therefore

[23] Another reason is that the individual-specific intrinsic propensities for each type of activity, α_i^m, α_i^{pa}, and α_i^{cg}, are correlated as well.

the integral taken over $f(\epsilon)$ has to include user i's likelihood of all three activities at time t. Second, recall that the probability of a friendship formation depends on both user i's and user j's utilities for the tie formation, i.e. a friendship is only formed if both users derive positive utilities from doing so (see Equation (2)). Since at each time t, all users can become friends with any other user with whom they are not friends yet, all friendship formation decisions of all users at time t are connected through the pair-specific random effect ζ_{ij} and the error terms in users' friendship formation utilities. In other words, due to the second reason, the integrals over $f(\epsilon)$ and $f(\zeta)$ have to to include all friendship formation probabilities of all users at time t. Combining the first and second reason, it is evident that the integrals over $f(\epsilon)$ and $f(\zeta)$ have to include the probabilities of all actions of all users at time t.

Third, recall that our model includes time-invariant individual-specific intrinsic propensities for each type of activity, i.e., α_i^m, α_i^{pa}, and α_i^{cg}, with a full covariance matrix Σ^α and a time-invariant pair-specific random effect ζ_{ij}. Therefore, for each user and each type of activity, the integrals over $f(\alpha)$ and $f(\zeta)$ have to include all activities of that type over all time periods. Given that the first two reasons necessitate that the integrals over $f(\epsilon)$ and $f(\zeta)$ contain the probabilities of all actions of all users at each time t and given that the third reason necessitates that the integrals over $f(\alpha)$ and $f(\zeta)$ contain all probabilities over all time period for a specific type of activity and a specific user, the integrals over $f(\epsilon)$, $f(\alpha)$, and $f(\zeta)$ have to contain the probabilities of all actions of all users over all time periods (see Equation (13)). As a result of these three issues, when we take the logarithm of the model likelihood, we *cannot* convert the product of the probabilities into a sum of the logarithms of these probabilities (see Equation (14)). This poses a problem for common computing technologies since the likelihood is the product of a very large number of probabilities and too small in magnitude to be detected.[24] To make the likelihood estimation computationally tractable, we use a transformation of the logarithm of a sum of variables to a function of the logarithm of those variables. Details on the transformation and our estimation approach are presented in Web Appendix A.

To speed up the estimation, we use OpenBLAS as the system BLAS (Basic Linear Algebra

[24]For the interested reader, the likelihood is given by the product of over 136,000,000 probabilities.

Subprograms), tensorization of large matrices, and parallel computing methods to estimate the model. Due to the large size of the data and parallelization, we cannot run the estimation code on conventional computing systems.[25] We utilize several large memory super-computing servers including the Texas Advanced Computing Center (TACC), the Pittsburgh Supercomputer Center (PSC) (Towns et al. 2014; Nystrom et al. 2013), and Jetstream cloud-computing (Stewart et al. 2015; Towns et al. 2014).[26]

5 Identification

The set of parameters to be estimated is given by $\{\bar{\alpha}^m, \bar{\alpha}^{pa}, \bar{\alpha}^{cg}, \Sigma^\alpha, \sigma_\zeta^m, \kappa_1^m, \kappa_2^m, \delta^m, \beta^{pa}, \beta^{cg}, \lambda^m, \lambda^{pa}, \lambda^{cg}, \gamma^m, \gamma^{pa}, \gamma^{cg}, G\}$. The identification of $\{\kappa_1^m, \kappa_2^m, \delta^m, \beta^{pa}, \beta^{cg}, \lambda^m, \lambda^{pa}, \lambda^{cg}, \gamma^m, \gamma^{pa}, \gamma^{cg}\}$ is standard. In the following, we first informally discuss the identification of $\{\bar{\alpha}^m, \bar{\alpha}^{pa}, \bar{\alpha}^{cg}, \Sigma^\alpha, \sigma_\zeta^m, G\}$ and then present our identification approach for peer effects.

The mean intrinsic propensities, $\bar{\alpha}^m$, $\bar{\alpha}^{pa}$, and $\bar{\alpha}^{cg}$, are identified by the average user behavior in each of the three areas across users and across time. The covariance matrix of the user random effects, Σ^α, is identified by variation and covariation in average activity levels across users. In contrast, the correlation matrix of the error terms, G, is identified by the variation in the simultaneous co-occurance of activities in a day. The term σ_ζ^m captures the standard deviation of the unobserved similarity (latent homophily) between users i and user j and is identified by the variation in friendship formation across different pairs of users. And lastly, conditional on G, the three utilities are separately identified since each action is modeled as a function of other actions in the previous time period.

Separating homophily from influence is a challenging task (Manski 1993). Recall that homophily refers to friends behaving in a similar manner due to their similar preferences and not because of one influencing the other. Similarity in unobserved preferences, if unaccounted for, can lead to correlated errors which, in turn, lead to upward biased estimates of friends'

[25] The fully parallelized estimation code requires at least 350GB of RAM.

[26] It takes more than 3 weeks to estimate a model with 131 parameters on a super computer utilizing 32 CPU cores using our data containing 68 million observations.

influence. We address this issue by incorporating unobserved time-invariant components, α_i^m, α_i^{pa}, and α_i^{cg}, in a user's decisions to form ties, to adopt animes, and to generate content (similar approach as in Nair, Manchanda, and Bhatia 2010, Trusov, Bodapati, and Bucklin 2010, and Ameri, Honka, and Xie 2019). Since we model the incidence of users' actions and not the specific taste for *which* product to adopt or *what* type of content to generate, homophily only plays a role in the frequency level of users' actions, i.e., whether they perform an action on each day. For example, two friends are similar to each other if both tend to publish a lot of posts. In our model, this unobserved heterogeneity in the propensity to perform each of the three actions is captured by α_i^m, α_i^{pa}, and α_i^{cg}. Furthermore, α_i^m, α_i^{pa}, and α_i^{cg} are assumed to be time-invariant since levels of homophily are unlikely to change during the relatively short time span of our observation period. Moreover, since many of the users are new to the network, the latent propensities are identified not only by the variation in behavior after any friendship formation, but also by behavior before any ties are formed, i.e., when friends' influence is absent. And lastly, to capture correlations among a user's intrinsic propensities to perform the three types of activities (i.e., make friends, watch animes, and produce UGC), we allow for correlations among α_i^m, α_i^{pa}, and α_i^{cg}.

6 Results

We present the estimation results in Table 2. As discussed in Section 2.1, we estimate separate coefficients for core and non-core users in the friendship formation utility. The estimation results for core and non-core users are presented in separate columns in the first half of Table 2. In the following, we focus on discussing the results for the core users. Column (i) in Table 2 contains the results for a model in which the decisions about the three types of actions of making friends, watching animes, and publishing posts are made independently of each other. Column (ii) presents the parameter estimates for a model in which we allow these three decisions to be correlated, but there is no unobserved heterogeneity among users. And lastly, column (iii) depicts the results for our full model in which we allow for both correlated errors and

unobserved heterogeneity among consumers.

==========

Insert Table 2 about here

==========

The results across the three different specifications are overall consistent. Potential simultaneous incidences of the three types of actions a user might engage in each day are captured through correlations of the error terms. We find significant correlations among all three actions. We also find significant coefficients for the standard deviations of the individual-specific random effects suggesting the presence of unobserved heterogeneity in intrinsic propensities across users. Further, the correlations among these random effects are all significant. And lastly, as expected, user i's net benefit from making friends declines with the length of her membership on the platform since, especially shortly after joining. This result suggests having friends helps to reduce the learning costs associated with navigating the website. In the following, we focus on evaluating the results for our main model in column (iii).

6.1 Friendship Formation

We start by discussing the parameter estimates for the friendship formation utility for the 400 core users in our sample. Common friends and common animes capture the similarity between two users. As expected, we find a significant positive effect for common friends implying that users are more likely to connect with friends of friends. Having more friends in common serves as a signal for similar interests of individuals i and j. This finding is in line with results in the previous literature (e.g., Aral, Muchnik, and Sundararajan 2009; Shalizi and Thomas 2011). The coefficient for the number of common animes is significant albeit with a negative sign. A potential explanation is that users are less inclined to become friends with users who watched many of the same animes since they can learn less about not-yet-watched animes that might be of interest to them from these users.

In terms of demographic similarities, we find positive and significant coefficients for user i and user j being close in age and having the same gender if both individuals reveal this

information. However, the coefficients for dummies indicating whether both users provide the information are negative. In other words, knowing demographic information about each other only increases the chance of forming a friendship tie if both users are similar in those characteristics. Otherwise, it actually hurts the chance of forming a friendship tie. Lastly, the coefficient for the standard deviation of the pair-specific random effect capturing latent homophily between individual i and individual j is positive but insignificant. This finding suggests that our flexible modeling approach together with the observed variables included in our model capture the similarity between two individuals well.

Next, we examine the results for proximity. Consistent with prior research (e.g., Mazur and Richards 2011, Amichai-Hamburger, Kingsbury, and Schneider 2013), we find that (physical) proximity matters even in an online social network: the coefficient for the dummy variable indicating whether both users are from the same country is positive and significant. Similar to the results for demographic similarity, we find a significant negative coefficient for the dummy indicating whether both users provide this information.

User j's number of friends, j's number of watched animes, and j's number of written posts represent the knowledgeability of user j as a potential friend for user i. We find a significant positive effect of the number of friends user j has indicating that users gain utility from becoming friends with well-connected users. This finding stands in contrast to the findings in Christakis et al. (2010) who find that students are less likely to become friends with popular students. A potential explanation for this result might be the unique context of the online environment. Next, we find that user j's (cumulative) number of watched animes and her (cumulative) number of posts have significant positive effects on friendship formation. This result is consistent with the notion that domain knowledge and information sharing are the primary incentives for friendship tie formations in our specific empirical context.

Comparing the magnitudes of the effects of user j's friends versus user j's activities on the platform using marginal effects, we find the following: the two largest drivers of user i's utility of forming a friendship with user j are user j's number of friends and the number of friends user i and user j have in common. User j's online activities, i.e. her number of watched animes

and her number of written posts, only play a secondary role.

And lastly, we find significant effects for all our control variables. The coefficient for the dummy variable indicating whether user j showed any activity during the previous week is positive and significant. One likely reason is user j's activities increased her visibility and awareness among other platform users. Further, the weekend dummy has a significant negative coefficient, suggesting that users are less likely to make friends with other users on weekends.

To summarize, our results for friendship formation reveal that all three determinants, i.e., similarity with, proximity to, and knowledgeability of a potential friend, matter. Further, the number of friends and overlap with a potential friend's friends are more important drivers of friendship formation than a potential friend's product adoption and content generation activities. And lastly, even in (anonymous) online social networks having similar demographics and geographic proximity matter.

6.2 Online Activities

Next, we discuss our results related to a user's anime watching decisions. We find a negative and significant effect for the number of friends a user has implying that users with more friends are less likely to watch an anime. The coefficient associated with the number of animes watched by friends reflects the influence friends' behavior has on a user's anime watching decisions. We find a significant positive effect of the number of animes watched by friends the previous day indicating the existence of peer effects. We also find a spill-over effect of friends' posting behavior on a user's anime watching: a user is significantly more likely to watch an anime if her friend made a post the previous day. Further, our results reveal positive state dependence in anime watching: users are more likely to watch an anime if they did so on the previous day. And lastly, the coefficient for the weekend dummy is positive and significant implying that users are more likely to watch animes on weekends. A potential explanation is that users might have more free time during weekends.

We now describe our findings related to content generation. We find a significant positive

coefficient for the number of friends a user has indicating that having more friends makes a user more active in publishing content. We also find evidence for the direct influence of friends' UGC production on a user's content generation: the number of posts published by friends during the previous day has a significant and positive effect on a user's content generation decision on the following day. However, the number of animes watched by friends does not have a significant effect on a user's UGC production, i.e., there is no spill-over effect. In addition, there is evidence of positive state dependence in content generation, i.e., we find a significant positive effect of a user's posting on her posting behavior the following day. One likely reason is the conversation/discussion nature of content generation. Other users can post a reply or comment on a post published by the focal user and/or the user herself might respond by writing another post. Further, we find a significant positive effect of the cumulative number of animes a user has watched: the more animes a user watched, the more likely it is that the user publishes content (likely about the watched anime). This provides another reason for the existence of state dependence. When a user watches animes, she is likely to want to talk about them. This interest in talking might last for a few days and state dependence captures this effect. And lastly, we find a significant positive effect for the weekend dummy.

To summarize, we find evidence of significant peer and spill-over effects in online activities (anime watching and content generation). Having friends who watch many animes and post a lot makes a user more likely to do the same. However, simply having many friends does not necessarily result in more activity.

7 Prediction Exercises

For companies operating social networks, advertising revenue represents their primary source of income. In 2015, the industry earned revenues of over $25 billion through advertisements.[27] Advertising revenues depend on site traffic: the more active users are, the more ads can be shown to them. In addition, having more active users can increase the appeal of the website to

[27] https://www.statista.com/statistics/271406/advertising-revenue-of-social-networks-worldwide/

non-users and lead to continuous growth of the user base. Therefore, it is in platform owners' best interest to motivate users (or a subset of users if stimulating all users is not feasible) to increase their in-site activities.

Previous literature on mature networks has developed a set of results for static networks on how to most effectively increase in-site activity levels. For example, previous research has shown that seeding to more connected users is the most effective way of increasing the total number of product adoptions within a community through peer effects (e.g., Hinz et al. 2011; Aral, Muchnik, and Sundararajan 2013). While this result holds true for mature networks with a static network structure, a stimulation intervention in an *evolving* network is very likely to also lead to changes in the structure of the network due to the possibility of newly formed friendship ties. As a result, to understand diffusion patterns in evolving networks, one needs to take the evolving ties in the network into account as well. By modeling the co-evolving friendship network and users' actions under their friends' influence, we capture the cascading effects of stimulating users to conduct more activities of a specific type on future states of the network and users' future activity levels.

Using our estimation results, we examine the effects of stimulating different types of users and different types of in-site activities through a series of prediction exercises. More precisely, we assume that the platform can trigger an increase in any of the three activities of making friends, watching animes, and generating content by, for example, posting a recommendation list on a user's webpage: the platform can recommend to a user to become friends with some other users, to adopt some specific animes, or to participate in forum discussions that are active and related to the user's past adoptions or posts. Although we do not observe the login or page view activities of a user and, as a result, cannot directly translate the changes in activity levels to changes in ad viewership, as long as users are not spending *less* time on each activity compared to before the stimulation, an increase in the total activity level will also lead to an increase in the time spent on the website. Furthermore, an increase in the activity level is observable by other users and non-users of the website and therefore can lead to activity cascades as well as a growing user base.

7.1 Effects of a Platform-Wide Stimulation

In the first set of prediction exercises, we examine and compare the increases in *overall* activity levels, i.e., the sum of activities in all three areas, of all core users due to a platform-wide stimulation. The overall activity levels serve as a proxy for the total site traffic or total time spent on the site. The prediction exercises are implemented as follows: in each scenario, we increase one type of activity (friendship tie formations, product adoptions or UGC generation) among all core and all non-core users by the amount of that activity on a typical day for an average user. To put it differently, our stimulation doubles the amount of activity of a particular type on a given day for each core and each non-core user. We do so on days 30, 90, and 150 and predict users' behavior going forward until day 184.[28] When presenting our findings, we focus on findings for the core users.

The results are presented in Table 3. Column (i) shows the changes in average number of active core users per day, i.e., the number of core users who perform at least one activity in a day, and column (ii) depicts the changes in the number of total activities performed by core users. Out of the three types of stimulations the platform can implement (i.e., to recommend friends, animes, or forum discussion topics), stimulating users to post more UGC is the most effective intervention resulting in the highest overall increase in the number of active users and in the level of in-site activities. One reason might be the interactive nature of UGC: after making a post, users might return to the platform a few days later to check for comments and feedback to their original post, which might prompt them to respond by posting even more. Stimulating users to watch more animes is the second most effective strategy while stimulating users to make more friends is the least effective strategy among the three.

==========

Insert Table 3 about here

==========

[28] If users joined the website after the stimulation day, their actions are simulated from the time they join (without a stimulation).

7.2 Effects of Seeding

Previous literature has found that users often have varying degrees of activity in different areas (e.g., Manchanda, Xie, and Youn 2008; Iyengar, den Bulte, and Valente 2011). Our data confirm this pattern (see also Section 2.2 and Figure 4). For example, a user might make many friends or publish many posts but only watch few animes. Consequently, carefully choosing whom to target and which type of activity to stimulate are crucial for platform owners in achieving a desired outcome such as an increase in UGC production.

In the second set of prediction exercises, we examine the effectiveness of different seeding strategies in increasing tie formations, anime watching, and UGC production in an evolving social network. For these prediction exercises, we select the 15% most/least active core users among all core users based on their activity levels in making friends, adopting animes, or producing UGC ("selection activity") as our seeding targets.[29] As a benchmark, we also randomly select 15% of core users as the initial seeds. Next, we increase the activity level of these selected core users in one of the three areas of activity ("seeding activity") by the amount of that activity on a typical day for an average user. This is the same stimulation as in the previous set of prediction exercises. We do so on days 30, 90, and 150 ("stimulation date") and simulate users' behavior going forward until day 184. To recap, we perform a total of 63 prediction exercises, including $3 \cdot 3 \cdot 3 = 27$ prediction exercises for the 15% most active core users, $3 \cdot 3 \cdot 3 = 27$ prediction exercises for the the 15% least active core users, as well as $3 \cdot 3 = 9$ prediction exercises for our benchmark case of randomly selected users. And lastly, when presenting our findings, we again focus on findings for the core users.

Overall, we find that seeding to the 15% most active users – whether active in the area of friendship formation or anime watching or UGC production – is a more effective seeding strategy than seeding to the 15% least active users. Across all three selection activities, all three seeding activities, and all three stimulation dates, i.e., 27 prediction exercises, we find seeding to the 15% most active users to increase friendship formation, on average, by 13.08%, while seeding

[29] 15% of core users are equivalent to 8 core users on day 30, 26 core users on day 90, and 50 core users on day 150.

to the 15% least active users increases friendship formation, on average, by 2.65%. Similarly, seeding to the 15% most active users increases anime watching and UGC production, on average, by 0.25% and 0.35%, respectively, while seeding to the 15% least active users increases anime watching and UGC production, on average, by only 0.90% and 0.21%, respectively.[30] Seeding to 15% randomly selected users performs inbetween for friendship formation and anime watching with 15.45% and 0.14%, respectively, while it is the most effective seeding strategy for UGC production with an increase of 0.51%.

Next, we focus on the results for the 27 prediction exercises targeting the 15% most active users as initial seeds and evaluate whether seeding to the most connected users, i.e., users with the most friends, is the most effective seeding strategy in terms of the choice of selection activity. In other words, we wonder whether the most-connected users are indeed better candidates for seeding than the users who are most active in anime watching or UGC creation in an evolving social network. We find that, across the three seeding activities and the three stimulation days (9 prediction exercises), on average, seeding to the most connected users results in an average increase of 3.67% of overall activities, i.e., activity levels in all three areas of friendship formation, anime watching, and UGC production. In comparison, seeding to users who watch a lot of animes leads to an average increase of 5.21% of overall activities and seeing to users who post a lot of UGC leads to an average increase of 4.79% of overall activities.[31] Therefore, among the three selection activities, choosing seeds based on the number of friendship connections is the least effective seeding strategy.

Digging deeper, we next investigate the most effective seeding strategy (in terms of choosing both a selection and a seeding activity) for each desired outcome. Our results indicate that – for all three potential desired outcomes of increasing the number of friendships, increasing anime watching or increasing UGC posting – selecting the 15% of users who watch the most

[30]This finding also holds when comparing possible pairs of prediction exercises, e.g., when comparing results from a pair of prediction exercises in which either the 15% of users with the most or least friends were stimulated to watch more animes on day 30. Across the 27 pairs of prediction exercises, for 65% of pairs seeding to the 15% most active users is more effective than seeding to the 15% least active users.

[31]This findings also holds for the individual prediction exercises: in 25 out of the 27 prediction exercises, we find that seeding to the most connected users is less effective than either seeding to users watching the most animes or seeding to users posting the most UGC.

animes is the best strategy, i.e., anime watching should be used as the selection activity. If the goal is to increase the number of friendship, then choosing the 15% of users who watch the most animes and encouraging them to watch even more animes is the most effective strategy. If instead the goal is to increase anime watching or UGC production, then choosing the 15% of users who watch the most animes and encouraging them to post more UGC is the winning strategy.

Lastly, we examine by how much the effectiveness of seeding strategies is underestimated when the endogenous network formation is not accounted for. To do so, we re-run the prediction exercises discussed in this section, but do not allow users to form new friendships. We find that not accounting for the endogenous network formation leads, on average, to an underestimation of seeding effectiveness by 30%. This finding underscores the practical importance of modeling the co-evolution of individual users' friendship tie formations and their concurrent in-site activities. The choice of the most effective seeding strategy can be misguided due to the ignorance of newly formed ties and their influence on network users' activities in an evolving network.

To summarize, first, we find that seeding to the most active users – whether active in the area of friendship formation or anime watching or UGC production – is a more effective seeding strategy than seeding to the least active users. Second, among the most active users, our results show that seeding to well-connected users is less effective than seeding to users who watch a lot of animes or post a lot of UGC. Third, we find the most effective seeding strategies for specific desired outcomes to be as follows: if the desired outcome is an increase in friendships, then the platform should target users who watch a lot of animes and encourage them to watch even more animes. If the desired outcome is to increase anime watching or UGC production, then the 15% of users who watch the most animes should be encouraged to post more UGC. And lastly, our results show that not accounting for the endogenous network formation leads, on average, to an underestimation of seeding effectiveness by 30%.

8 Limitations and Future Research

There are several limitations to our research. First, we only observe a friendship if both users agree to become friends. In other words, we observe neither the friendship request nor the potential rejection of that request. This is a limitation of our data. As a result, we cannot separately identify whether an increase in a user's number of friends is due to that user's elevated preference to form friendships or due to her increased desirability as a potential friend to other users. Second, in our data, we do not observe tie dissolution and thus assume everlasting friendships. Although, due to the small cost of friendship ties for users, we do not believe unfriending is a frequent act in the network under study, it is still possible for users to break their friendship ties. This unfriending behavior in itself is interesting and can provide additional insights into network formation dynamics.

Third, in this paper, we model whether users post something on the website or whether they watch an anime, but not the topic or number of posts or which anime they watch. Studying the details of each action can shed further light on the co-evolution process of users' friendship formations and concurrent activities which we leave for future research. Fourth, we do not consider the length or content of users' posts. Longer or more detailed posts may imply the writer is more knowledgeable. Studying the effects of such UGC characteristics will be an interesting extension of our current research. And lastly, in this paper, we do not model platform growth in our relatively short observation period, i.e., we do not model users' joining behavior and assume that it is exogenous. However, in the long run, the popularity of a platform in terms of the size of its user base and volume and variety of its content can change the rate of users joining the website. We hope future research can relax the exogeneity assumption and provide further insights into this research question.

9 Conclusion

In this study, we develop a model for the co-evolution of individuals' friendship tie formations and their concurrent activities (product adoptions and UGC production) within an online social network. Explicitly modeling the endogenous formation of the network and accounting for the interdependence between decisions in these two areas (friendship formations and concurrent online activities) allows us to discover important drivers of individuals' friendship decisions and, at the same time, to identify the resulting peer effects on individuals' other online activities. We estimate our model using a novel data set capturing the continuous development of a network and users' entire action histories within the network.

Our findings shed light on the important factors that drive strangers to become friends in an online environment. Specifically, our results reveal that, compared to a potential friend's product adoptions and content generation activities, the total number of friends and the number of common friends a potential friend has with the focal individual are the most important drivers of friendship formation. Further, while having more friends does not necessarily make a person more active, having more active friends does increase a user's activity levels in terms of both product adoptions and content generation through peer and spill-over effects. Via prediction exercises we assess the effectiveness of various seeding and stimulation strategies in increasing website traffic while taking the endogenous network formation into account. Contrary to previous studies (e.g., Hinz et al. 2011; Aral, Muchnik, and Sundararajan 2013), we find that seeding to most connected users is not the best strategy to increase users' activity levels on the website.

References

Ameri, Mina, Elisabeth Honka, and Ying Xie (2019), "Word-of-Mouth, Observed Adoptions, and Anime Watching Decisions: The Role of the Personal versus the Community Network," *Marketing Science*, 38 (4), 567–583.

Amichai-Hamburger, Y., M. Kingsbury, and B. Schneider (2013), "Friendship: An Old Concept with a New Meaning?" *Computers in Human Behavior*, 29 (1), 33–39.

Aral, Sinan, Lev Muchnik, and Arun Sundararajan (2009), "Distinguishing Influence-Based Contagion from Homophily-Driven Diffusion in Dynamic Networks," *Proceedings of the National Academy of Sciences*, 106 (51), 21544–21549.

— (2013), "Engineering Social Contagions: Optimal Network Seeding in the Presence of Homophily," *Network Science*, 1 (2), 125–153.

Aral, Sinan and Dylan Walker (2011), "Creating Social Contagion Through Viral Product Design: A Randomized Trial of Peer Influence in Networks," *Management Science*, 57 (9), 1623–1639.

Badev, Anton (2013), "Discrete Games in Endogenous Networks: Theory and Policy," *Working Paper*.

Berndt, E., B. Hall, R. Hall, and J. Hausman (1974), "Estimation and Inference in Nonlinear Structural Models," *Annals of Economic and Social Measurement*, 3/4 (1), 653–665.

Berscheid, Ellen and Harry Reis (1998), "Attraction and Close Relationships," in D. Gilbert, S. Fiske, and G. Lindzey (Editors), "The Handbook of Social Psychology," 193–281.

Bornstein, Robert (1989), "Exposure and Affect: Overview and Meta-Analysis of Research, 1968-1987," *Psychological Bulletin*, 106 (2), 265–289.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin (2009), "Identification of Peer Effects Through Social Networks," *Journal of Econometrics*, 150 (1), 41–55.

Brandtzæg, Petter and Jan Heim (2009), "Why People Use Social Networking Sites," in "International Conference on Online Communities and Social Computing," 143–152.

Christakis, Nicholas and James Fowler (2013), "Social Contagion Theory: Examining Dynamic Social Networks and Human Behavior," *Statistics in Medicine*, 32 (4), 556–577.

Christakis, Nicholas, James Fowler, Guido Imbens, and Karthik Kalyanaraman (2010), "An Empirical Model for Strategic Network Formation," *Working Paper*.

De Giorgi, Giacomo, Michele Pellizzari, and Silvia Redaelli (2010), "Identification of Social Interactions Through Partially Overlapping Peer Groups," *American Economic Journal: Applied Economics*, 2 (2), 241–275.

Duan, Wenjing, Bin Gu, and Andrew Whinston (2009), "Informational Cascades and Software Adoption on the Internet: An Empirical Investigation," *MIS Quarterly*, 33 (1), 23–48.

Gitler, Isidoro and Jaime Klapp (Editors) (2015), *High Performance Computer Applications - 6th International Conference, ISUM 2015, Revised Selected Papers*, Mexico City, Mexico: Springer International Publishing.

Griffin, Em and Glenn Sparks (1990), "Friends Forever: A Longitudinal Exploration of Intimacy in Same-Sex Pairs and Platonic Pairs," *Journal of Social and Personal Relationships*, 7, 29–46.

Hanaki, Nobuyuki, Alexander Peterhansl, Peter Dodds, and Duncan Watts (2007), "Cooperation in Evolving Social Networks," *Management Science*, 53 (7), 1036–1050.

Hinz, Oliver, Bernd Skiera, Christian Barrot, and Jan Becker (2011), "Seeding Strategies for Viral Marketing: An Empirical Comparison," *Journal of Marketing*, 75 (6), 55–71.

Iyengar, Raghuram, Christophe Van den Bulte, and Thomas Valente (2011), "Opinion Leadership and Social Contagion in New Product Diffusion," *Marketing Science*, 30 (2), 195–212.

Katona, Zsolt, Peter Zubcsek, and Miklos Sarvary (2011), "Network Effects and Personal Influences: The Diffusion of an Online Social Network," *Journal of Marketing Research*, 48 (3), 425–443.

Kozinets, Robert (1999), "E-Tribalized Marketing?: The Strategic Implications of Virtual Communities of Consumption," *European Management Journal*, 17 (3), 252–264.

Lampe, Cliff, Nicole Ellison, and Charles Steinfield (2007), "A Familiar Face (Book): Profile Elements as Signals in an Online Social Network," in "Proceedings of the SIGCHI Conference on Human Factors in Computing Systems," 435–444.

Langlois, Judith, Lisa Kalakanis, Adam Rubenstein, Andrea Larson, Monica Hallam, and Monica Smoot (2000), "Maxims or Myths of Beauty? A Meta-Analytic and Theoretical Review," *Psychological Bulletin*, 126 (3), 390–423.

Lee, Lung-fei (2007), "Identification and Estimation of Econometric Models with Group Interactions, Contextual Factors and Fixed Effects," *Journal of Econometrics*, 140 (2), 333–374.

Lee, Lung-fei, Xiaodong Liu, and Xu Lin (2010), "Specification and Estimation of Social Interaction Models with Network Structures," *The Econometrics Journal*, 13 (2), 145–176.

Li, Xiaolin, Elisabeth Honka, and Pradeep Chintagunta (2018), "Technology Adoption and Depression Diagnosis: Evidence from the Adoption of Multiple Technologies," *Working Paper*.

Ma, Liye, Ramayya Krishnan, and Alan Montgomery (2014), "Latent Homophily or Social Influence? An Empirical Analysis of Purchase Within a Social Network," *Management Science*, 61 (2), 454–473.

Manchanda, Puneet, Ying Xie, and Nara Youn (2008), "The Role of Targeted Communication and Contagion in Product Adoption," *Marketing Science*, 27 (6), 961–976.

Manski, Charles (1993), "Identification of Endogenous Social Effects: The Reflection Problem," *The Review of Economic Studies*, 60 (3), 531–542.

Mazur, E. and L. Richards (2011), "Adolescents' and Emerging Adults' Social Networking Online: Homophily or Diversity?" *Journal of Applied Developmental Psychology*, 32 (4), 180–188.

McPherson, M., L. Smith-Lovin, and J. Cook (2001), "Birds of a Feather Flock Together: Homophily in Social Networks," *Annual Review of Sociology*, 27, 415–444.

Moe, Wendy and Michael Trusov (2011), "The Value of Social Dynamics in Online Product Ratings Forums," *Journal of Marketing Research*, 48 (3), 444–456.

Moreland, Richard and Scott Beach (1992), "Exposure Effects in the Classroom: The Development of Affinity Among Students," *Journal of Experimental Social Psychology*, 28 (3), 255–276.

Nair, Harikesh, Puneet Manchanda, and Tulikaa Bhatia (2010), "Asymmetric Social Interactions in Physician Prescription Behavior: The Role of Opinion Leaders," *Journal of Marketing Research*, 47 (5), 883–895.

Nystrom, Nick, Joel Welling, Phil Blood, and Eng Lim Goh (2013), "Blacklight: Coherent Shared Memory for Enabling Science," in "Contemporary High Performance Computing," Chapman and Hall/CRC, 421–439.

Sacerdote, Bruce (2001), "Peer Effects with Random Assignment: Results for Dartmouth Roommates," *The Quarterly Journal of Economics*, 116 (2), 681–704.

Shalizi, Cosma and Andrew Thomas (2011), "Homophily and Contagion are Generically Confounded in Observational Social Network Studies," *Sociological Methods & Research*, 40 (2), 211–239.

Shriver, Scott, Harikesh Nair, and Reto Hofstetter (2013), "Social Ties and User-Generated Content: Evidence from an Online Social Network," *Management Science*, 59 (6), 1425–1443.

Snijders, Tom, Johan Koskinen, and Michael Schweinberger (2010), "Maximum Likelihood Estimation for Social Network Dynamics," *The Annals of Applied Statistics*, 4 (2), 567–588.

Snijders, Tom, Christian Steglich, and Michael Schweinberger (2007), "Modeling the Coevolution of Networks and Behavior," in Kees van Montfort, Johan Oud, and Albert Satorra (Editors), "Longitudinal Models in the Behavioral and Related Sciences," 41–71.

Snijders, Tom, Gerhard van de Bunt, and Christian Steglich (2010), "Introduction to Stochastic Actor-Based Models for Network Dynamics," *Social Networks*, 32 (1), 44–60.

Stewart, C., Timothy Cockerill, Ian Foster, D. Hancock, N. Merchant, E. Skidmore, D. Stanzione, J. Taylor, S. Tuecke, G. Turner, M. Vaughn, and N. Gaffney (2015), "Jetstream: A Self-Provisioned, Scalable Science and Engineering Cloud Environment," in "XSEDE Conference: Scientific Advancements Enabled by Enhanced Cyberinfrastructure," .

Sun, Tianshu and Sean Taylor (2019), "Displaying Things in Common to Encourage Friendship Formation: A Large Randomized Field Experiment," *Quantitative Marketing and Economics*, forthcoming.

Toivonen, Riitta, Lauri Kovanen, Mikko Kivelä, Jukka-Pekka Onnela, Jari Saramäki, and Kimmo Kaski (2009), "A Comparative Study of Social Network Models: Network Evolution Models and Nodal Attribute Models," *Social Networks*, 31 (4), 240–254.

Tong, Stephanie, Brandon Van Der Heide, Lindsey Langwell, and Joseph Walther (2008), "Too Much of a Good Thing? The Relationship Between Number of Friends and Interpersonal Impressions on Facebook," *Journal of Computer-Mediated Communication*, 13 (3), 531–549.

Toubia, Olivier and Andrew Stephen (2013), "Intrinsic vs. Image-Related Utility in Social Media: Why Do People Contribute Content to Twitter?" *Marketing Science*, 32 (3), 368–392.

Towns, John, Timothy Cockerill, Maytal Dahan, Ian Foster, Kelly Gaither, Andrew Grimshaw, Victor Hazlewood, Scott Lathrop, Dave Lifka, Gregory Peterson, Ralph Roskies, Ray Scott, and Nancy Wilkins-Diehr (2014), "XSEDE: Accelerating Scientific Discovery," *Computing in Science & Engineering*, 16 (5), 62–74.

Trusov, Michael, Anand Bodapati, and Randolph Bucklin (2010), "Determining Influential Users in Internet Social Networks," *Journal of Marketing Research*, 47 (4), 643–658.

Tucker, Catherine (2008), "Identifying Formal and Informal Influence in Technology Adoption with Network Externalities," *Management Science*, 54 (12), 2024–2038.

Watson, Goodwin Barbour and David Johnson (1972), *Social Psychology: Issues and Insights*, Lippincott Philadelphia.

Figures and Tables

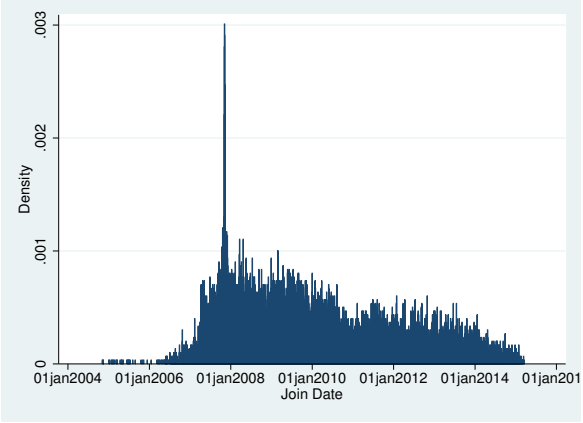
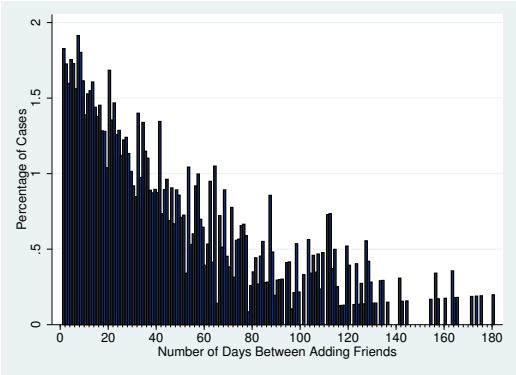
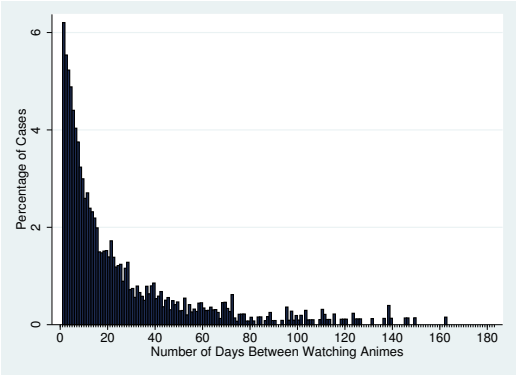


Figure 1: Dates Users Joined MyAnimeList.Net

(a) Friend Addition



(b) Anime Watching



(c) Content Generation

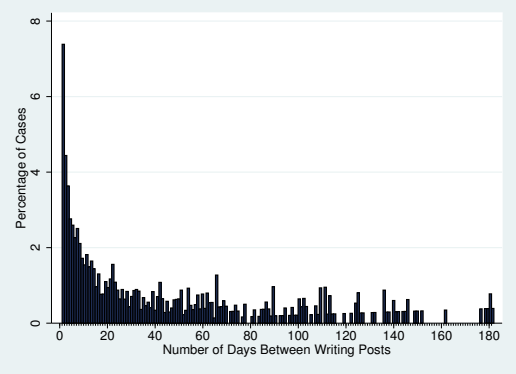


Figure 2: Number of Days Between Activities

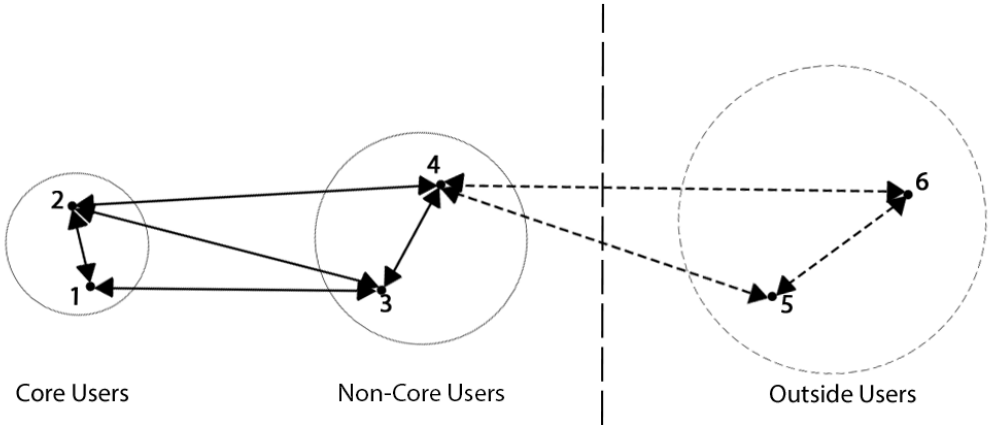


Figure 3: User Sampling Strategy

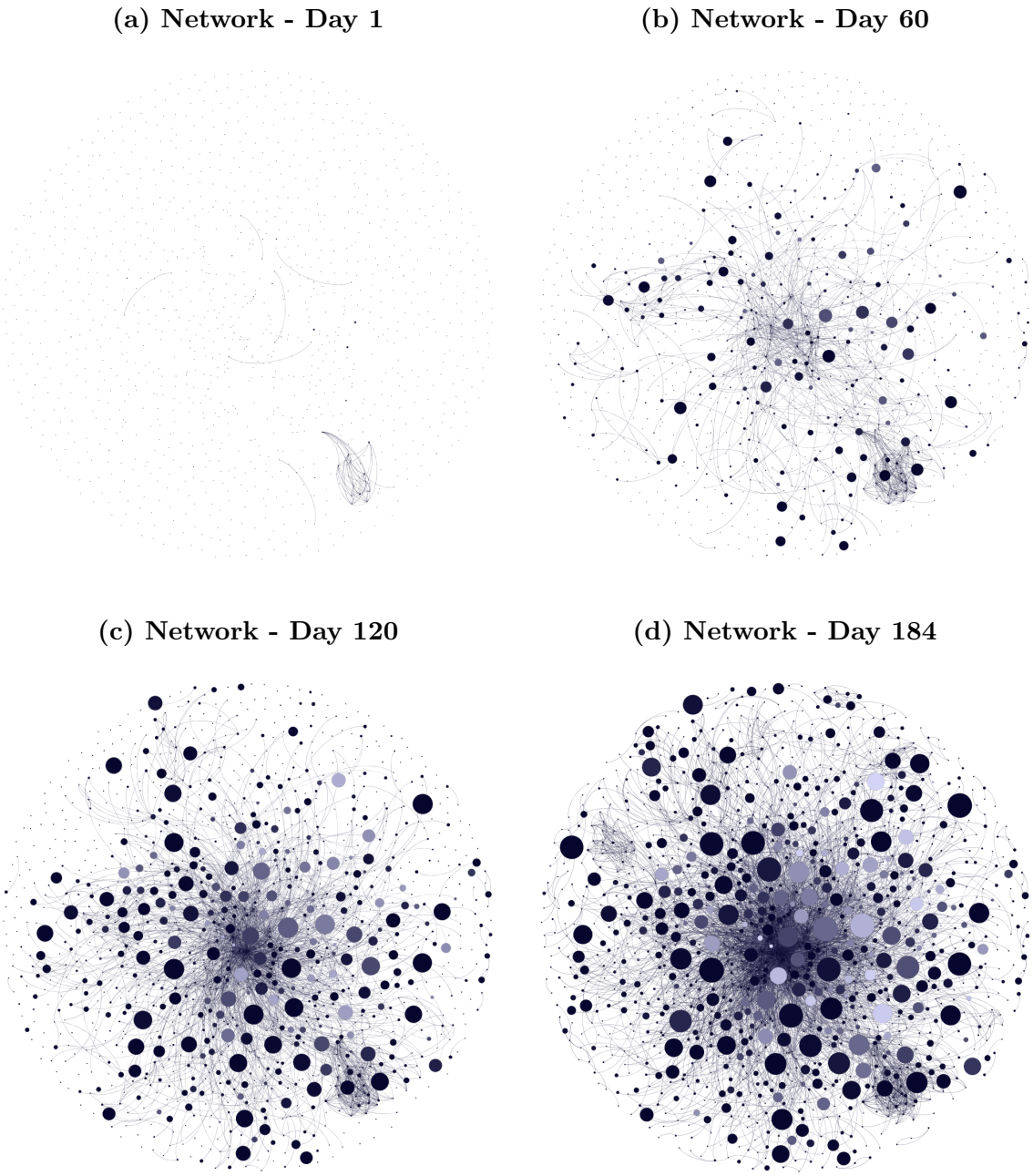
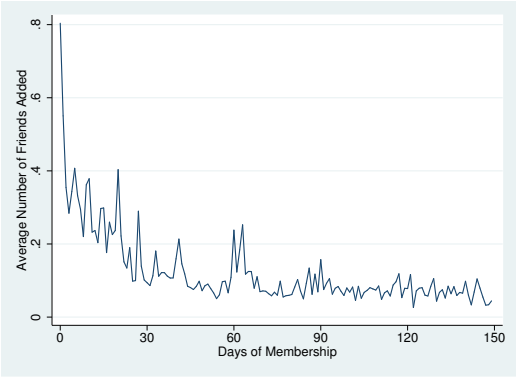
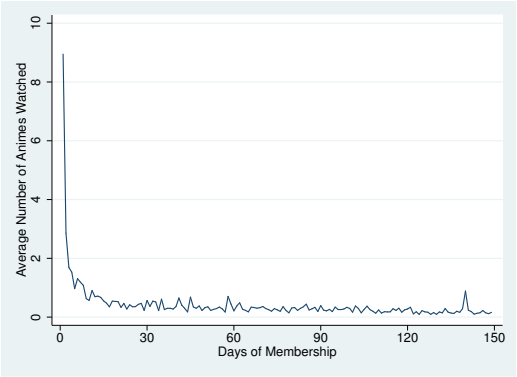


Figure 4: Network Co-Evolution Over Time
Lines Between Nodes Indicate Friendship Ties. Node Size Increases with More Animes Watched. Node Color Darkens with More Posts Written.

(a) Friend Addition



(b) Anime Watching



(c) Content Generation

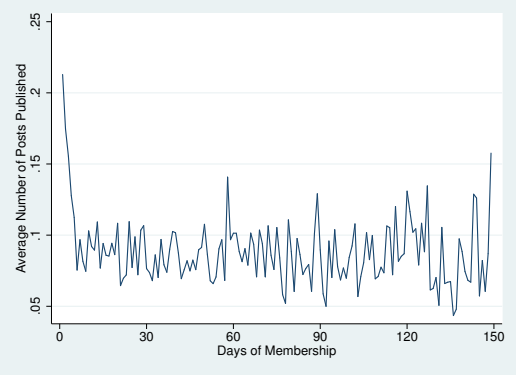


Figure 5: Average Activity Levels Over Time Since Joining (New Users)

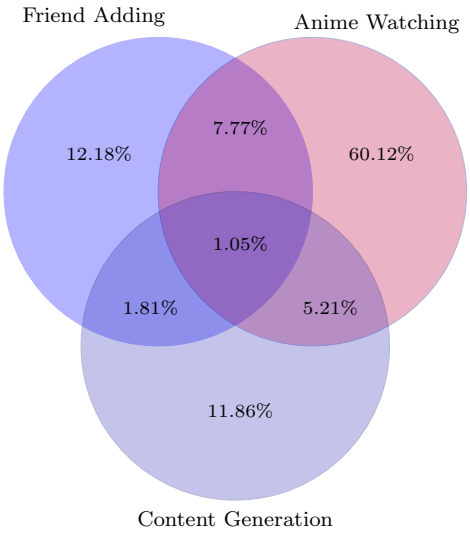
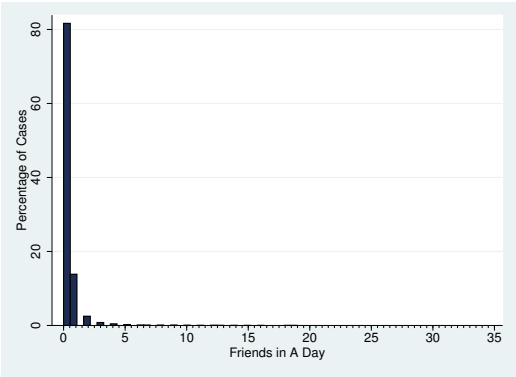
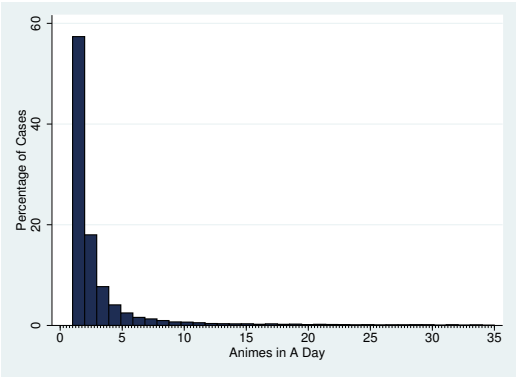


Figure 6: Percentage of Observations with Certain Activities Conditional on Performing at Least One Activity

(a) Number of Friends Added in a Day (Truncated at 100)



(b) Number of Animes Watched in a Day (Truncated at 100)



(c) Number of Posts Written in a Day

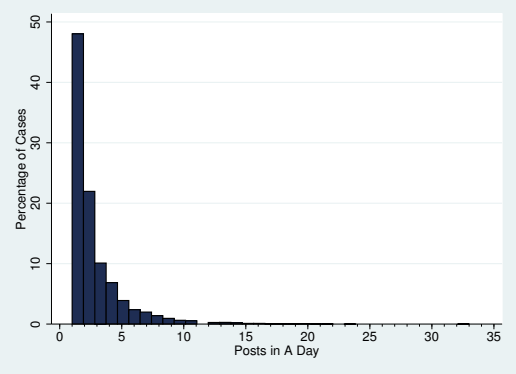


Figure 7: Number of Activities in Each Area Per Day

	Mean	Std. Dev.	Min	Median	Max	N
Age	19.3	5.2	12	18.5	78	1,088
Gender (% Females)	38.5					
Gender (% Males)	54.0					
Gender (% Not Specified)	7.5					
Core Users:						
Number of Active Days	12.62	15.43	0	7	101	400
Number of Friend Adding Days	2.55	4.69	0	1	59	400
Number of Anime Watching Days	9.19	10.37	0	5	68	400
Number of Content Generating Days	2.47	9.17	0	0	95	400
Percentage of Active Days	17.76	17.07	0	12.90	100	400
Percentage of Friend Adding Days	3.80	6.19	0	1.63	40	400
Percentage of Anime Watching Days	13.55	14.11	0	9.70	85.71	400
Percentage of Content Generating Days	2.56	7.72	0	0	61.69	400
Friend Adding Interval in Days	44.26	36.05	1	34	156	12,414
Anime Adding Interval in Days	25.64	28.20	1	14	138	20,680
Post Adding Interval in Days	32.07	29.33	1	25	109	6,330
Non-Core Users:						
Number of Active Days	24.06	24.42	1	16	181	986
Number of Friend Adding Days	5.93	6.40	1	4	48	986
Number of Anime Watching Days	15.55	16.65	0	10	121	986
Number of Content Generating Days	6.31	17.57	0	0	181	986
Percentage of Friend Adding Days	22.49	18.36	.54	17.95	100	986
Percentage of Anime Watching Days	6.85	8.31	.54	4.23	80	986
Percentage of Content Generating Days	14.46	13.38	0	10.87	82.22	986
Percentage of Content Generating Days	4.72	12.01	0	0	98.37	986
Friend Adding Interval in Days	48.41	40.36	1	37	181	78,996
Anime Adding Interval in Days	24.82	29.86	1	13	163	84,133
Post Adding Interval in Days	47.98	48.96	1	28	182	40,287

Table 1: Descriptive Statistics

	(i) Independent		(ii) Homogenous		(iii) Main Model	
	Core	Non-Core	Core	Non-Core	Core	Non-Core
Friendship Formation						
Similarity						
Number of Friends in Common with j	0.0997^{***}	-0.0936^{***}	0.1824^{***}	-0.1792^{***}	0.1791^{***}	-0.1824^{***}
by $t-1^{a}$	(0.0054)	(0.0057)	(0.0069)	(0.0072)	(0.0069)	(0.0072)
Number of Animes in Common with j	-0.0201^{***}	0.0046	-0.0104^{***}	0.0063^{*}	-0.0130^{***}	0.0054
by $t-1^{a}$	(0.0034)	(0.0034)	(0.0029)	(0.0029)	(0.0029)	(0.0029)
Dummy for Whether i and j Are	0.3170^{***}	-0.0900^{***}	0.2669^{***}	-0.0773^{***}	0.2684^{***}	-0.0761^{***}
Within 5 Years of Age	(0.0023)	(0.0022)	(0.0020)	(0.0019)	(0.0018)	(0.0019)
Dummy for Whether i and j	0.2079^{***}	-0.0978^{***}	0.1626^{***}	-0.0960^{***}	0.1642^{***}	-0.0954^{***}
Have the Same Gender	(0.0034)	(0.0026)	(0.0025)	(0.0026)	(0.0025)	(0.0027)
Dummy for Whether Both i and j	-0.6105^{***}	0.5308^{***}	-0.4171^{***}	0.3128^{***}	-0.4123^{***}	0.3167^{***}
Indicate Their Age	(0.0024)	(0.0026)	(0.0024)	(0.0025)	(0.0024)	(0.0025)
Dummy for Whether Both i and j	-0.0839^{***}	-0.0877^{***}	-0.2382^{***}	0.1663^{***}	-0.2353^{***}	0.1687^{***}
Indicate Their Gender	(0.0023)	(0.0028)	(0.0021)	(0.0020)	(0.0020)	(0.0019)
Standard Deviation of Pair-Specific	0.2745^{***}		0.0008		0.0002	
Random Effect	(0.0038)		(0.0034)		(0.0034)	
Proximity						
Dummy for Whether i and j Are	0.2259^{***}	-0.0972^{***}	0.1351^{***}	-0.0101^{***}	0.1359^{***}	-0.0083^{***}
from Same Country	(0.0029)	(0.0026)	(0.0020)	(0.0020)	(0.0021)	(0.0019)
Dummy for Whether Both i and j	-0.8645^{***}	0.7751^{***}	-0.6250^{***}	0.5532^{***}	-0.4372^{***}	0.3437^{***}
Indicate Their Country	(0.0025)	(0.0030)	(0.0017)	(0.0021)	(0.0016)	(0.0020)
Knowledgeability						
j's Number of Friends by $t-1^{a}$	0.6971^{***}	-0.0371^{***}	0.5273^{***}	0.1404^{***}	0.5271^{***}	0.1401^{***}
	(0.0054)	(0.0056)	(0.0036)	(0.0038)	(0.0036)	(0.0038)
j's Number of Watched Animes by $t-1^{a}$	0.0426^{***}	-0.0824^{***}	0.0089^{*}	-0.0430^{***}	0.0089^{*}	-0.0429^{***}
	(0.0045)	(0.0047)	(0.0035)	(0.0037)	(0.0036)	(0.0037)
j's Number of Written Posts by $t-1^{a}$	0.1768^{***}	-0.1434^{***}	0.0259^{***}	-0.0226^{***}	0.0243^{***}	-0.0234^{***}
	(0.0070)	(0.0073)	(0.0054)	(0.0056)	(0.0053)	(0.0056)
Control Variables						
Dummy for Whether j was Active from	0.0636^{***}	0.2951^{***}	0.1513^{***}	0.1235^{***}	0.1513^{***}	0.1227^{***}
$t-7$ to $t-1$	(0.0007)	(0.0008)	(0.0025)	(0.0026)	(0.0030)	(0.0031)
Dummy for Whether t Is a Weekend	-0.4931^{***}	0.5181^{***}	-0.5530^{***}	0.6048^{***}	-0.5540^{***}	0.6041^{***}
	(0.0063)	(0.0064)	(0.0047)	(0.0048)	(0.0048)	(0.0049)
Dummy for i Being Non-Core User	-1.1456^{***}		-1.0651^{***}		-1.0645^{***}	
	(0.0030)		(0.0024)		(0.0024)	
Dummy for i Having joined before	0.2123^{***}		0.1321^{***}		0.1333^{***}	
July 2007	(0.0021)		(0.0013)		(0.0013)	
Constant	-1.6330^{***}		-2.0494^{***}		-2.0492^{***}	
	(0.0036)		(0.0030)		(0.0029)	
Standard Deviation of Individual-Specific	0.0225^{***}				0.0034^{***}	
Random Effect	(0.0004)				(0.0000)	
Number of Membership Days by t^{a}	-0.6922^{***}		-0.6246^{***}		-0.6263^{***}	
	(0.0003)		(0.0002)		(0.0002)	
Week Dummies	yes		yes		yes	
Model Summary Statistics						
Number of Observations	69,020,774		69,020,774		69,020,774	
AIC	404,210.40		358,885.80		358,582.20	
BIC	406,264.79		360,892.04		360,684.74	
LogLikelihood	−201,977.20		−179,317.90		−179,160.10	

Standard errors in parentheses.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

[a] Measured on logarithmic scale.

Table 2: Results

	(i) Independent Core & Non-Core	(ii) Homogenous Core & Non-Core	(iii) Main Model Core & Non-Core
Anime Watching			
Number of Friends by $t-1$[a]	-0.0249^{***}	-0.0267^{***}	-0.0270^{***}
	(0.0052)	(0.0040)	(0.0040)
Number of Animes Watched by Friends in $t-1$[a]	0.0811^{***}	0.0707^{***}	0.0714^{***}
	(0.0027)	(0.0018)	(0.0018)
Number of Posts Published by Friends in $t-1$[a]	0.0496^{***}	0.0355^{***}	0.0361^{***}
	(0.0021)	(0.0014)	(0.0014)
Dummy for Whether i Watched an Anime in $t-1$	1.1197^{***}	0.8931^{***}	0.8935^{***}
	(0.0006)	(0.0006)	(0.0006)
Dummy for Whether t Is a Weekend	0.0657^{***}	0.0532^{***}	0.0531^{***}
	(0.0009)	(0.0002)	(0.0002)
Constant	-3.5236^{***}	-1.5394^{***}	-1.5796^{***}
	(0.0024)	(0.0020)	(0.0020)
Standard Deviation of Individual-Specific Random Effect	0.2595^{***}		0.0001^{***}
	(0.0001)		(0.0000)
Week Dummies	yes	yes	yes
Content Generation			
Number of Friends by $t-1$[a]	0.0684^{***}	0.0496^{***}	0.0488^{***}
	(0.0061)	(0.0041)	(0.0041)
Number of Animes Watched by Friends in $t-1$[a]	-0.0105^{**}	0.0023	0.0013
	(0.0035)	(0.0022)	(0.0022)
Number of Posts Published by Friends in $t-1$[a]	0.2484^{***}	0.2519^{***}	0.2510^{***}
	(0.0031)	(0.0021)	(0.0021)
Dummy for Whether i Published a Post in $t-1$	2.1675^{***}	2.1375^{***}	2.1375^{***}
	(0.0006)	(0.0004)	(0.0004)
Number of Animes Watched by $t-1$[a]	0.0377^{***}	0.0461^{***}	0.0469^{***}
	(0.0100)	(0.0070)	(0.0070)
Dummy for Whether t Is a Weekend	0.0251^{***}	0.0244^{***}	0.0244^{***}
	(0.0009)	(0.0002)	(0.0002)
Constant	-5.6849^{***}	-2.9280^{***}	-2.9283^{***}
	(0.0025)	(0.0017)	(0.0017)
Standard Deviation of Individual-Specific Random Effect	0.1052^{***}		0.1160^{***}
	(0.0002)		(0.0000)
Week Dummies	yes	yes	yes
Error Correlation Matrix			
Correlation between Friendship and Adoption		-0.0542^{***}	-0.0540^{***}
		(0.0003)	(0.0000)
Correlation between Friendship and UGC		0.0189^{**}	0.0186^{***}
		(0.0059)	(0.0000)
Correlation between Adoption and UGC		-0.0002	0.0374^{***}
		(0.0035)	(0.0000)
Random Effects Correlation Matrix			
Correlation between Friendship and Adoption			0.0003^{***}
			(0.0000)
Correlation between Friendship and UGC			0.0000^{***}
			(0.0000)
Correlation between Adoption and UGC			0.0022^{***}
			(0.0000)

Standard errors in parentheses.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

[a] Measured on logarithmic scale.

Table 2: Results (Continued.)

Stimulation Activity	Stimulation Day	(i) Change in Active USERS in %	(ii) Change in ACTIVITIES in %
Friendship Formation			
	30	1.2	4.6
	90	0.4	0.3
	150	0.3	0.2
	Average	*0.6*	*1.7*
Anime Watching			
	30	1.5	6.0
	90	1.0	1.1
	150	1.7	2.3
	Average	*1.4*	*3.1*
UGC Production			
	30	2.7	10.0
	90	3.5	2.5
	150	6.4	4.7
	Average	*4.2*	*5.7*

Table 3: Platform-Wide Stimulation

Web Appendix A: Log-Likelihood Derivation

In this section, we explain the estimation techniques used to estimate the log-likelihood. Recall that the model log-likelihood is given by

$$
\begin{aligned}
LL = \log \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \prod_{t=1}^{T} \prod_{i=1}^{N} & \left(\Pr\left(A_{it}^{pa}=1\right)\right)^{A_{it}^{pa}} \left(1-\Pr\left(A_{it}^{pa}=1\right)\right)^{1-A_{it}^{pa}} \\
& \cdot \left(\Pr\left(A_{it}^{cg}=1\right)\right)^{A_{it}^{cg}} \left(1-\Pr\left(A_{it}^{cg}=1\right)\right)^{1-A_{it}^{cg}} \\
& \cdot \prod_{j=i+1}^{N} \left[\left(\Pr\left(m_{ijt}=1\right)\right)^{m_{ijt}} \left(1-\Pr\left(m_{ijt}=1\right)\right)^{1-m_{ijt}}\right]^{1-m_{ij,t-1}} d\epsilon d\alpha d\zeta.
\end{aligned}
\tag{A1}
$$

We use simulated maximum likelihood (SMLE) to calculate the log-likelihood. R denotes the number of random draws. Given this estimation method and the law of large numbers, the log-likelihood can be written as follows:

$$
\begin{aligned}
LL = & \log \frac{1}{R} \sum_{r=1}^{R} \Big[\prod_{t=1}^{T} \prod_{i=1}^{N} \left(\Pr\left(A_{it}^{pa}=1\right)\right)^{A_{it}^{pa}} \left(1-\Pr\left(A_{it}^{pa}=1\right)\right)^{1-A_{it}^{pa}} \\
& \cdot \left(\Pr\left(A_{it}^{cg}=1\right)\right)^{A_{it}^{cg}} \left(1-\Pr\left(A_{it}^{cg}=1\right)\right)^{1-A_{it}^{cg}} \\
& \cdot \prod_{j=i+1}^{N} \left[\left(\Pr\left(m_{ijt}=1\right)\right)^{m_{ijt}} \left(1-\Pr\left(m_{ijt}=1\right)\right)^{1-m_{ijt}}\right]^{1-m_{ij,t-1}} \Big] \Bigg|_{r} \\
= & -\log R + \log \sum_{r=1}^{R} Q_r
\end{aligned}
\tag{A2}
$$

with

$$
\begin{aligned}
Q_r = \prod_{t=1}^{T} \prod_{i=1}^{N} & \left(\Pr\left(A_{it}^{pa}=1\right)\right)^{A_{it}^{pa}} \left(1-\Pr\left(A_{it}^{pa}=1\right)\right)^{1-A_{it}^{pa}} \cdot \left(\Pr\left(A_{it}^{cg}=1\right)\right)^{A_{it}^{cg}} \left(1-\Pr\left(A_{it}^{cg}=1\right)\right)^{1-A_{it}^{cg}} \\
& \cdot \prod_{j=i+1}^{N} \left[\left(\Pr\left(m_{ijt}=1\right)\right)^{m_{ijt}} \left(1-\Pr\left(m_{ijt}=1\right)\right)^{1-m_{ijt}}\right]^{1-m_{ij,t-1}} \Bigg|_{r}
\end{aligned}
\tag{A3}
$$

Note that each of the probabilities in Q_r is a small number and the number of probabilities being multiplied to calculate Q_r is very large, i.e., $N \cdot N \cdot \frac{N(N-1)}{2}$. Thus Q_r is extremely small and most likely not processed properly by a computer. To bypass this issue, we use the following transformation:

$$\log \sum_{i=0}^{N} a_i = \log a_0 + \log \left(1 + \sum_{i=1}^{N} e^{(\log a_i - \log a_0)} \right). \tag{A4}$$

Thus we can write the model log-likelihood as

$$LL = -\log R + \log Q_1 + \log \big(1 + \sum_{r=2}^{R} e^{(\log Q_r - \log Q_1)}\big), \tag{A5}$$

with

$$\begin{aligned}
&\log Q_r = \\
&\sum_{t=1}^{T} \sum_{i=1}^{N} \Big[A_{it}^{pa} \cdot \log \left(\Pr \left(A_{it}^{pa} = 1 \right) \right) + (1 - A_{it}^{pa}) \cdot \log \left(1 - \Pr \left(A_{it}^{pa} = 1 \right) \right) \\
&\qquad + A_{it}^{cg} \cdot \log \left(\Pr \left(A_{it}^{cg} = 1 \right) \right) + (1 - A_{it}^{cg}) \cdot \log \left(1 - \Pr \left(A_{it}^{cg} = 1 \right) \right) \\
&\qquad + \sum_{j=i+1}^{N} \left[(1 - m_{ij,t-1}) \cdot \left[m_{ijt} \cdot \log \left(\Pr \left(m_{ijt} = 1 \right) \right) + (1 - m_{ijt}) \cdot \log \left(1 - \Pr \left(m_{ijt} = 1 \right) \right) \right] \right] \Big].
\end{aligned} \tag{A6}$$

We now explain the simulation procedure within SMLE. For each of the random effects, α_i^m, α_i^{pa}, α_i^{cg} and ζ_{ij}, and each of the error terms, ϵ_{it}^m, ϵ_{it}^{pa}, and ϵ_{it}^{cg}, we start by taking $R = 30$ random draws from a standard normal distribution. Let us define a "set of random draws" as a set of random draws containing one random draw for each of the four random effects and the three error terms, i.e., we have $R = 30$ sets of random draws. Note that we use the Cholesky decomposition for Σ^α, Σ, and σ^m. Next, for each set of random draws out of the R sets, we calculate $\log Q_r$ and, subsequently, LL.