# A Model of Network Dynamics: Tie Formation, Product Adoption, and Content Generation

### PRELIMINARY AND INCOMPLETE. CURRENT VERSION: MARCH 2018

Mina Ameri<sup>†</sup> Elisabeth Honka<sup>‡</sup> Ying Xie<sup>§</sup>

#### Abstract

We develop a model for the co-evolution of individual users' friendship tie formation and their concurrent online activities (production adoptions and production of user-generated content) within an online social network. By explicitly modeling the endogenous formation of the network and accounting for the interdependence between decisions in friendship tie formation and in concurrent online activities, we are able to discover important drivers underlying individuals' friendship decisions while, at the same time, provide a clean identification of 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 history within the network. 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 this potential friend has with the focal user are the most important drivers of friendship formation. Further, while having more friends does not make a user more active, having more active friends does increase a user's activity levels in terms of both product adoptions and content generation through peer effects. We conduct several counterfactual exercises to assess the effectiveness of various seeding and stimulation strategies in increasing website traffic while taking the endogenous network formation into account. We find that seeding to users with the most friends is not always the best strategy to increase users' activity levels on the website.

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

JEL Classification: D83, L82, M31

<sup>\*</sup>We would like to thank Randy Bucklin, Brett Hollenbeck, Sylvia Hristakeva, Dmitri Kuksov, Xiaolin Li, B.P.S. Murthi, Ram Rao, Brian Ratchford, Upender Subramanian, and participants at Texas Marketing Faculty Research Colloquium. Mina Ameri and Elisabeth Honka gratefully acknowledge support by the Morrison Center for Marketing and Data Analysis. 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. All errors are our own.

<sup>&</sup>lt;sup>†</sup>University of Texas at Dallas, mina.ameri@utdallas.edu.

<sup>&</sup>lt;sup>‡</sup>University of California Los Angeles, elisabeth.honka@anderson.ucla.edu.

<sup>&</sup>lt;sup>§</sup>University of Texas at Dallas, ying.xie@utdallas.edu.

# 1 Introduction

(Online) social networks have become an indispensable part of many consumers' everyday lives. By the end of 2016, there were 2.3 billion active social network users worldwide<sup>1</sup> with the average user having more than 5 accounts across different platforms.<sup>2</sup> Despite these astonishing numbers, new (online) social networks are still being started every month.<sup>3</sup> The goal of social networks is to connect people. The number of connections users have can grow very fast especially in the online environment. For example, Instagram users saw, on average, a monthly growth of 16% in the number of their followers in 2017.<sup>4</sup> As the size and connectedness of social networks grow, people are increasingly sharing information and communicating with each other through these networks. For example, every minute, Facebook users shared 1.3 million pieces of content and Twitter users sent 0.35 million tweets in 2016.<sup>5</sup>

Research has shown that social networks greatly facilitate information dissemination through social learning (e.g., Duan et al. 2009; Katona et al. 2011; Christakis and Fowler 2013). Social networks provide a natural platform for users to produce and publish content (e.g., posts in discussions forums, product reviews, re-shares) related to various activities they engage in (e.g., Toubia and Stephen 2013) and to disseminate these contents to others beyond their own personal network (e.g., Moe and Trusov 2011). Moreover, these opinions and experiences shared by friends or other users through social networks have a significant effect on network users' decisions in a wide range of product markets (e.g., Bramoullé et al. 2009; Carrell et al. 2009; Aral and Walker 2011; Ameri et al. 2016). However, with the exception of Toubia and Stephen (2013), all these results were found for mature social networks. It is an open empirical question whether they also hold for platforms in which the

<sup>&</sup>lt;sup>1</sup>https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/

<sup>&</sup>lt;sup>2</sup>https://insight.globalwebindex.net/social-q4-2014

<sup>&</sup>lt;sup>3</sup>For example, Mastodon, a micro-blogging service, was founded in October 2016 and had 81,000 users by July 2017 (https://mastodon.social/about/more).

<sup>&</sup>lt;sup>4</sup>https://socialpilot.co/blog/151-amazing-social-media-statistics-know-2017/

<sup>&</sup>lt;sup>5</sup>https://socialpilot.co/blog/125-amazing-social-media-statistics-know-2016/

social network is still evolving and changing.

In this paper, we study the endogenous formation of an online social network in which users can become friends, adopt products, and produce user-generated content (UGC). Despite their significant role in today's society, very little is known about how (online) social networks develop and evolve and, in particular, how and with whom people become friends. An intriguing aspect of making friends in many online social networks is that people do not know each other's real identities. As a result, an individual's behaviors and opinions as observed by other people in the online environment are the main factors influencing friendship formation decisions in these types of networks. If online behaviors and opinions are likely drivers of friendship formations, studying online networks provides valuable insights into important factors that drive people's friend making decisions.

For platform owners – since online social networks mostly rely on advertising revenue for profitability and thus for platform survival – it is crucial to understand how social networks evolve and what drives users' friendship formations and activity levels (e.g. production of UGC). This knowledge would allow them to devise effective stimulation and seeding strategies that leverage friends' social influence to promote (more) friendships and other activities on the platform. 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 diffusion of a behavior (e.g., Hinz et al. 2011; Aral et al. 2013). However, these studies investigated mature (static) social networks. Thus their finding may or may not hold true in evolving networks in which a stimulation intervention is very likely to also change individuals' friendship networks and increase connectedness. As a result, to accurately evaluate the effectiveness of seeding strategies in evolving networks, one also needs to take the endogenous formation of new ties into account. We do so in a counterfactual in this paper.

The extant empirical studies on network formation typically examine changes in the network structure at a macro level, but are insufficient in providing a micro-foundation for network tie formations from an individual actor's perspective (see Jackson 2008 and Toivonen et al. 2009 for a review). A notable exception is a small set of studies that are known as strategic network formation models (Christakis et al. 2010; Snijders, Koskinen and Schweinberger 2010). In both papers, Snijders, Koskinen and Schweinberger (2010) and Christakis et al. (2010), the authors model future states of a network based on characteristics of the existing network state (current ties). However, the models proposed in these two papers are not appropriate for the study of network formation in an online environment for the following reason: Both papers have data on at most a few snapshots of the network and not the continuous development of the network. As a result, the authors resort to simulating network states between observed snapshots. Due to the simulated network states, besides changes in network structure, the authors are not able to capture the effects of timevarying variables such as users' activities on tie formation. While Snijders, Koskinen and Schweinberger (2010) and Christakis et al. (2010) explain the formation of ties via static user characteristics such as age or gender, it is possible that users' time-varying behaviors and actions are also significant drivers of their friendship formation decisions. This problem is aggravated in online environments where an individual's observed online activities and opinions user such as product adoptions or UGC may even be the more important drivers of friendship formation decisions since personal characteristics such as age or gender are either unavailable or non-verifiable.

Furthermore, while people's activities and opinions influence the friendships they form, their future activities and opinions are also subject to the influence of their friends. The latter is often termed social influence, peer or network effects in the literature (e.g., Sacerdote 2001; Katona et al. 2011; Iyengar et al. 2011). However, the endogeneity of network formation makes it a challenging task to correctly identify network effects (Manski 1993). Previous literature has suggested several approaches to deal with this challenge, e.g. instrumental variables (e.g., Bramoullé et al. 2009; De Giorgi et al. 2010), correlated group effects (e.g., Lee 2007; Lee et al. 2010; Ma et al. 2014), randomness/exogenous shocks (Sacerdote 2001; Tucker 2008), experiments (e.g., Aral and Walker 2011), individual-specific unobserved preferences (e.g., Nair et al. 2010; Trusov et al. 2010), and co-evolution models (e.g., Snijders et al. 2007). In this paper, we build on and extend the last two approaches: We account for individualspecific 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.

Our data come from a special interest online community website 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, and that the actions they observe on the website are the main drivers of the 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 histories on the platform, i.e. anime watching and UGC posts. Users report anime watching through personal watch lists, i.e., lists of animes that they have watched together with the dates of doing so, and exchange information and opinions, i.e. produce UGC, through publishing discussion forum posts, anime reviews, and recommendations on the platform. We observe the dates of all UGC posts as well. Access to these data containing the complete network evolution and all of users' actions on the platform allows us, unlike Christakis et al. (2010) and Snijders, Koskinen and Schweinberger (2010), to model the friendship network development without the need to simulate the state of the network at each point in time and thus we are able to measure the effects of time-varying variables such as anime watching or UGC posts 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. 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 make a UGC post. All decisions are made with 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 attractiveness of the potential friend and the similarity between the pair. A friendship tie is formed if and only if both users agree to it. A user's utilities of participating in either product adoptions or content generation are functions of past online activities of the same type and of the user's friendship network which can affect her actions through peer effects.

Furthermore, to capture any common shocks unobserved by the researcher which might result in simultaneity, we include time fixed effects and also allow the error terms in the utility functions (for friendship formations, product adoption, and content generation) to be correlated within a day. To tease apart homophily from peer effects, we estimate individualspecific intrinsic propensities to watch animes and to post UGC. To estimate these individualspecific unobserved preferences in the utility functions for these two online activities, we use observations before and after a user has made friends. We further incorporate individualspecific propensities to make friends in the friendship formation utility to capture any benefits and costs that users associate with making friends. The three utility functions are connected in three ways: through observed variables, through correlated error terms, and through correlated individual-specific unobserved preferences. By explicitly accounting for the interdependences between network formation and individuals' online activities in the modeling and by using data from both before and after individuals make friends to estimate individual-specific observed preferences, we are able to provide a clean identification of peer effects.

Our results for friendship formation reveal the relative importance of users' activities versus users' friendship networks in friend making decisions. We find that the friendship networks of other users with whom a focal user is not friends (yet), both in terms of how many friends other users have and how many common friends other users have with the focal user, are more important than other users' activities. In addition, even in (anonymous) online networks similar demographics matter. We further find significant effects of a user's friends on the user's in-site activities, i.e., product adoptions and the production of UGC. However, we do not find any spill-over effects of one type of activity of friends on the other type of activity of a focal users, i.e. friends' product adoptions (content generation) do not influence the focal user's content generation (product adoptions). Lastly, while having more friends does not make a user more active, having active friends does increase a user's activity level due to the positive social influence.

We use our results to simulate various counterfactual scenarios to assess the effectiveness of various seeding and stimulation strategies in increasing users' activities on the website. Contrary to previous studies investigating static networks (e.g., Hinz et al. 2011; Aral et al. 2013), our results for evolving networks reveal that seeding to well-connected users, i.e. users with many friends, is not always the best strategy to increase users' UGC activities on the platform. This finding might be due to well-connected users generally becoming friends with other well-connected users who are not necessarily the ones who publish the most posts or watch the most animes. Further, we find that not accounting for the endogenous network formation in evolving networks when assessing the effectiveness of seeding strategies leads, on average, to an underestimation of seeding effectiveness by 10%.

The contribution of this paper is three-fold. First, our paper contributes to the strategic network formation literature by quantifying the effects of individuals' time-varying actions on friendship formation and their relative importance compared to users' static characteristics such as age or gender. This richer specification is much needed when describing the network development in an online environment. Second, our paper contributes to the social learning and network effects literature by providing a novel approach to identify the influence of friends' activities. Specifically, we account for the latent homophily by explicitly modeling the choice of friends and by incorporating individuals' intrinsic preferences for performing actions

that are identified absent of their friends' influence. And lastly, to the best of our knowledge, this paper is one of the first papers to model the strategic co-evolution of a friendship network along with users' actions within the network. By understanding the interdependent dynamics of network formation and online activities, our model yields important insights regarding strategies for companies and network platform owners to effectively stimulate user engagement.

The remainder of this paper is organized as follows: In the next section, we discuss the relevant literature. In Section 3, we describe our data and in Sections 4 to 6, we introduce our model, estimation approach, and identification strategy. We present and discuss our estimation and simulation results in Sections 7 and 8. 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 10.

# 2 Relevant Literature

In this section, we review three relevant streams of literature on network formation, peer effects, and seeding strategies in social networks.

#### 2.1 Network Formation

Researchers have studied network formation using three main modeling approaches: nodal attribute models, exponential graph models, and strategic network formation models. The first category of models explains the existence of ties and the resulting network structure via similarities among pairs of individuals (e.g., Hoff et al. 2002; Boguñá et al. 2004). However, this modeling approach explains the status quo of a network, i.e., who is/is not friends with whom, rather than its evolution. Further, it fails to take the structure of network connections into account. Exponential graph models explain the network development based on structural patterns such as triangular connections or transitivity, but do not provide in-

sights into the mechanisms that drive individuals' tie formation decisions (e.g., Katona and Móri 2006; Mele 2017). These models are well suited for making predictions but not for causal inferences and therefore they are not conducive for counterfactual analyses. To overcome these shortcomings, strategic network formation models, the most recently developed modeling approach among the three, have taken the perspective of individual actors' utility maximizations when explaining the evolution of a network, and allow them to depend on the existing state of the network (e.g., Hanaki et al. 2007). 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.<sup>6</sup> This paper falls into the last category of modeling approaches.

Christakis et al. (2010) and Snijders, Koskinen and Schweinberger (2010) are two empirical papers in the strategic network formation literature that are most closely related to our study. Both papers develop dynamic models to explain network formation through the effects of the existing network state (current ties) on individuals' future friendship tie formation decisions in the network. Christakis et al. (2010) propose a structural model in which each person's utility from forming a friendship tie with another person is a function of the number of friends the other person has as well as the distance or overlap between their respective friendship networks. In their model, given the chance of meeting based on a distribution of random sequences, each pair of individuals decide whether to become friends. However, due to the data constraint of observing only one snapshot of the network, Christakis et al. (2010) only incorporate static user characteristics such as age and gender in the utility function besides the existing network state. They apply their model to a network of 669 students with 1,541 friendship ties and find that, while having common friends is important for friendship formation, people are less likely to become friends with popular individuals. They also find that people prefer to become friends with people who are similar to themselves in terms of characteristics such as age or gender.

 $<sup>^{6}\</sup>mathrm{We}$  refer the interested reader to Jackson (2008) and Toivonen et al. (2009) for a detailed comparison of the three categories of models.

Unlike Christakis et al. (2010), Snijders, Koskinen and Schweinberger (2010) use a continuous Markov process to model the formation of the network as individuals decide whether to consider another individual as a friend. They use a data set that contains a few snapshots of a network. Although having more observations of the network gives Snijders, Koskinen and Schweinberger (2010) more flexibility in modeling than Christakis et al. (2010), due to not knowing the continuous sequence of tie formations, they are still unable to incorporate the effects of individuals' time-varying actions on their utilities for forming ties. They test their model using an empirical setting of a classroom of 32 students and 6 snapshots of the network and find that individuals prefer bi-directional ties to un-reciprocated uni-directional ties. In addition, students tend towards networks with transitive connections that are not singled out from the rest of the network. In terms of demographic characteristics, they find that males are more popular as friends, but that similarity of genders is not important.

Our paper complements and enriches the strategic network formation literature in two important ways. First, by observing the continuous evolution of the network, we do not need to rely on assumptions to simulate the current network state. Second, we directly observe each individual's entire action history in the network, which enables us to explicitly account for the effects of individuals' time-varying actions in their friendship formation decisions. The latter improvement is especially important since we study network formation in an online environment. In such an environment, individuals' behaviors and opinions as observed by other individuals are likely to be among the most important drivers of friendship tie decisions.

#### 2.2 Peer Effects

Previous research has shown that social networks greatly facilitate information dissemination and product diffusion (e.g., Duan et al. 2009; Katona et al. 2011; Christakis and Fowler 2013). These networks provide a platform for consumers to produce and publish content, for example, to post in discussions forums, to write product reviews or to re-share content (e.g., Toubia and Stephen 2013). Further, past research has also shown that friends influence each other's product adoption decisions through reviews and ratings they post online (e.g., Aral and Walker 2011; Ma et al. 2014; Ameri et al. 2016).

Making a casual inference of friends' influence, however, is a challenging task (Manski 1993). Multiple social phenomena can confound the identification and inference of social outcomes (Shalizi and Thomas 2011; Hartmann et al. 2008). Among these phenomena, homophily is probably the most challenging one to be accounted for. Homophily refers to the observation that individuals tend to become friends with similar individuals. Due to the similarity among friends, they exhibit the same behavior without one necessarily influencing the other. Different approaches have been proposed by previous literature to account for these issues including the use of instrumental variables (e.g., Bramoullé et al. 2009; De Giorgi et al. 2010; Claussen et al. 2014), the incorporation of individual-specific unobserved tastes/preferences (e.g., Nair et al. 2010; Trusov et al. 2010), controlling for correlated group effects (e.g., Lee 2007; Lee et al. 2010; Ma et al. 2014), the use of exogenous shocks to peers (e.g., Tucker 2008) or exogenous randomness (e.g., Sacerdote 2001), experiments (e.g., Aral and Walker 2011), and network co-evolution models (Snijders et al. 2007; Badev 2013). In our paper, we develop a network co-evolution model that explicitly incorporates individual-specific unobserved preferences to control for homophily.

The most notable among the co-evolution models was proposed by Snijders et al. (2007). Snijders et al. (2007) propose a stochastic model in which both the network structure and individuals' actions evolve simultaneously in a dynamic process: individuals are selected at random rates and each selected individual decides whether to make a change in her friendship ties or whether to perform an action of interest or whether to do neither. There are several limitations to Snijders et al. (2007). First, since individuals cannot change both their ties and their actions at the same point in time, simultaneity between tie formation and other actions is not accounted for. Second, although Snijders et al. (2007) capture homophily using friendship selection and similarity indices, 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. Third, due to the randomness in the decision timing, the effects of exogenous time-varying factors cannot be identified and, as a result, simultaneity between actions of friends cannot be controlled for.

In this paper, we overcome these limitations by proposing a structural model of individuals' concurrent decisions to both form ties and to perform activities at each point in time. This allows us to capture the effects of any time-varying variable while controlling for the simultaneity across these decisions. Furthermore, we are able to account for the latent homophily by explicitly modeling the choice of friends. In addition, by observing users' actions before and after they make friends, we estimate individuals' intrinsic preferences for actions absent of their friends' influence and therefore provide a better identification of peer effects.

#### 2.3 Seeding

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. The most commonly studied seeding strategies are based on network metrics such as "in/out-degree centrality"<sup>7</sup> or "betweenness centrality."<sup>8</sup> For example, Hinz et al. (2011) compare three seeding strategies — stimulating high-degree, low-degree, and high betweenness individuals with random seeding — in terms of their effectiveness in increasing adoption. They find that seeding to well-connected individuals is the most successful strategy in increasing participation in viral marketing campaigns. Similarly, Aral et al. (2013) examine the effectiveness of different seeding strategies under varying levels of homophily (self-tendency to adopt) and under the influence of friends on adoptions. They examine the effectiveness of seeding to high-degree individuals, dense regions of the network, and hubs unlikely to adopt against the effectiveness of random seeding and find that high-degree and dense region targeting generally perform better. They further show that seeding more than 0.2% of the population

<sup>&</sup>lt;sup>7</sup>In-degree and out-degree centralities refer to the number of incoming and outgoing ties, respectively, of an individual in a network.

<sup>&</sup>lt;sup>8</sup>Betweenness centrality of an individual refers to the number of shortest chains of links that connect all pairs in a network and include that individual.

is wasteful because the gain from their adoptions is lower than the gain from their natural adoptions. Katona (2013) studies seeding strategies in a theoretical framework and shows that highly connected influencers are valuable only if they cover consumers who are not connected to many other influencers.

Many of the previous studies have focused on adoption behavior. However, for the growing number of online social platforms, a continuous engagement of its users with the website may be more important than one-time adoption behavior. Trusov et al. (2010) focus on the activity levels of users within an online platform instead of adoption behavior. The authors develop an approach to determine which of a focal user's friends have a significant influence on the focal user's activity level. Trusov et al. (2010) use their results to examine the effect of any change in the focal user's behavior on the behavior of those linked to him and find that, on average, only 20% of a focal user's friends are influencing the focal user.

Following Trusov et al. (2010), we add to the existing knowledge on the effectiveness of seeding strategies on user engagement in online social platforms. Going beyond Trusov et al. (2010), we examine different seeding strategies that are not only based on network metrics, as is common in the literature, but can also depend on users' actions on the website. Furthermore, by modeling the co-evolution of friendship formations and users' actions over time, we not only capture the immediate effect of friends' influence on each other, but also capture how that effect propagates into the further development of the network and into future actions of users.

### 3 Data

Our data come from MyAnimeList.net. This website is a consumption-related online community (Kozinets 1999) where online interactions are based upon shared enthusiasm for a specific consumption activity. 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 to 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 connect with other fans. This implies that most users in our data meet their friends for the first time on the website under study 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.

#### 3.1 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 started 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 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 2006 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 friendship feature.

Studying daily friendship formation among *all* the 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 in 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 in the corresponding area. 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 of observations leads to information loss on the dependent variables and the sequence of actions. On top of that, we observe that in more than 20% of the cases, users make more than one friend in a week. 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 studies on strategic network formation (Christakis et al. 2010, Snijders, Koskinen and Schweinberger 2010), 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 using snowball sampling, which is the common sampling method in the network literature. Figure 3 visualizes our sampling strategy. First, we draw a random sample of 400 users ("core users") out of about 8,800 users who joined the website in the second half of 2007, and then include all of their friends in our estimation sample. Note that the friends of the 400 core users can also be core users or they can be users not in the core. 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.<sup>9</sup> 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 noncore 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 noncore 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.

<sup>&</sup>lt;sup>9</sup>The set of non-core users includes 732 users who joined after July 2007 and the 254 users joined before July 2007.

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 friendship decisions. Note that since we model all actions of core and non-core users and incorporate the effects of friendships with outside users on non-core users' anime adoptions and UGC production, this bias is mainly a concern for the estimation of the friendship 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).

A concern with snowball sampling is the oversampling of active users. This concern is alleviated by controlling for unobserved heterogeneity among 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 incorporation of separate parameters, we believe that our estimates are unbiased for the core 400 users.

The observation period is 184 days between July 1 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.

#### 3.2 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 the interdependence between users' friendship formation and other activities.

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 and 94% of users report their gender with 39% being female and 53% being male.

Insert Table 1 about here

\_\_\_\_\_

\_\_\_\_\_

Figure 5 demonstrates how activities of users who joined the website in 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 is a cost reduction in 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.<sup>10</sup>

\_\_\_\_\_

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 active days in terms of anime watching,

<sup>&</sup>lt;sup>10</sup>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.

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 activity. 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 85.81% of the cases. Users are active in two and three of the areas of interest in 13.36% and 0.83% of cases, respectively. We observe similar a pattern for non-core users albeit with generally higher average activity amounts (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 this data pattern, we make the simplifying assumption of modeling the anime watching and content generation as binary indicator variables, i.e. we model whether a user watches an anime or makes a post, but not the number of watched animes or posts.<sup>11</sup>

\_\_\_\_\_

Insert Figure 7 about here

<sup>11</sup>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.

### 4 Model

In this section, we describe how we jointly model a user's decisions to form ties, adopt products, and generate UGC. For friendship formations, we model whether and with whom users become friends, while for the activities in the other two areas, namely, anime adoptions and content generation, we only model users' decisions to participate in an activity or not.

#### 4.1 Tie Formation

We start by describing how we model tie formation 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 or not in a time period.

Suppose the website contains i = 1, ..., N users and these users can become friends with other users during t = 1, ..., T time periods.<sup>12</sup> 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, ..., N and  $i \neq j$ .  $m_{ij,t}$  equals 1 if i and j are friends at time t and 0 otherwise. Ties in the network are bi-directional, i.e.,  $m_{ij,t} = m_{ji,t}$ . 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.<sup>13</sup>

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). Let us define  $U_{i,t}^m$  as individual *i*'s utility of becoming friends with individual *j* in time period *t* and  $U_{j,t}^m$  as individual *j*'s

 $<sup>^{12}\</sup>mathrm{Note}$  that t is the calendar day and not the day since a user joined the website.

<sup>&</sup>lt;sup>13</sup>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 do not believe unfriending is a common action among users and given that the observation period is 184 days, we view not modeling friendship dissolution as a minor limitation.

utility of becoming friends with individual i in time period t. User i's utility function is given by

$$U_{i,t}^m = f(\mathcal{A}^m, \mathcal{R}^m, \epsilon^m). \tag{1}$$

The utility individual *i* derives from forming a tie with individual *j* depends on the attractiveness of user *j* as judged by her past actions,  $\mathcal{A}^m$ , and the similarity between user *i* and user *j*,  $\mathcal{R}^m$ .  $\epsilon^m$  captures the part of the utility of user *i* at time *t* that is observed by the user but not by the researcher.<sup>14</sup>

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. 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. This ensures that the network formation will have a unique equilibrium.<sup>15</sup>

We model the tie formation between two users as a non-cooperative decision, i.e., each pair's decision to become friends only depends on the observed pair-specific variables and decisions of different pairs of users are conditionally independent. A tie between i and j is formed if and only if both parties decide that it is beneficial for them to become friends, i.e.,

$$m_{ij,t} = 1$$
 iff  $U_{i,t}^m > 0$  and  $U_{j,t}^m > 0.$  (2)

<sup>&</sup>lt;sup>14</sup>Throughout the paper, we use the superscript m to refer to variables associated with friendship utility. <sup>15</sup>We refer the interested reader to Jackson (2008) for an extensive discussion of equilibria in networks.

#### 4.2 Product Adoption and Content Generation

Next, we describe how we model a user's activities on the website. We study users' activities in two broad areas, namely, product adoption and content generation. Let  $A_{i,t}^k, k \in \{\text{Product} A \text{doption } (pa), \text{Content Generation } (cg)\}$  define user *i*'s activity at time *t*. If user *i* adopts an anime at time *t*, then  $A_{i,t}^{pa}$  equals 1 and 0 otherwise. If user *i* posts on the website at time *t*, then  $A_{i,t}^{cg}$  equals 1 and 0 otherwise.

User *i*'s utilities from watching animes,  $U_{i,t}^{pa}$ , and producing content,  $U_{i,t}^{cg}$ , are given by

$$U_{i,t}^{pa} = g(\mathcal{A}^{pa}, F^{pa}, \epsilon^{pa})$$

$$U_{i,t}^{cg} = h(\mathcal{A}^{cg}, F^{cg}, \epsilon^{cg}),$$
(3)

where the utilities depend on a user's past actions,  $\mathcal{A}^{pa}$  and  $\mathcal{A}^{cg}$ , as well as on the influence of a user's friendship network,  $F^{pa}$  and  $F^{cg}$ .  $\epsilon^{pa}$  and  $\epsilon^{cg}$  describe the part of the utility that is observed by the user but not by the researcher.

#### 4.3 Integrating All Actions

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

$$U_{i,t}^{m} = f(\mathcal{A}^{m}, \mathcal{R}^{m}, \epsilon^{m}) \qquad \forall j = 1 \dots N, \ i \neq j$$

$$U_{i,t}^{pa} = g(\mathcal{A}^{pa}, \mathcal{F}^{pa}, \epsilon^{pa}) \qquad (4)$$

$$U_{i,t}^{cg} = h(\mathcal{A}^{cg}, \mathcal{F}^{cg}, \epsilon^{cg}).$$

Some variables, unobserved by the researcher, might influence all three types of decisions a user makes.<sup>16</sup> For example, having watched an anime, a user might be excited to discuss the anime with other members in the forum section of the website. To accommodate simultaneity among these three decisions, we allow the three error terms in Equation (4) to be correlated,

<sup>&</sup>lt;sup>16</sup>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.

i.e.,

$$\Sigma = \begin{bmatrix} \sigma_m^2 & \rho_{m,pa} & \rho_{m,cg} \\ \rho_{m,pa} & \sigma_{pa}^2 & \rho_{pa,cg} \\ \rho_{m,cg} & \rho_{pa,cg} & \sigma_{cg}^2 \end{bmatrix}.$$
(5)

As is well known, we need to set one element in the covariance matrix to 1 for identification reasons (McCulloch and Rossi 1994). Thus, we normalize  $\Sigma_{11}$ , i.e., the variance of the error term in the friendship utility, to 1.

#### 4.4 Utility Specifications

Next, we present details of the utility specifications for the specific context of our data. We model the utility of user i forming a tie with user j as

$$U_{i,t}^{m_{ij}} = \alpha_i^m + \beta^m \mathcal{A}_{j,t-1}^m + \gamma^m \mathcal{R}_{ij,t-1}^m + \lambda^m C_{it}^m + \epsilon_{i,t}^m,$$
  
where  
$$\mathcal{A}_{j,t-1}^m = \sum m_{ij,t-1} + \sum^{t-1} A_{j,\tau}^{pa} + \sum^{t-1} A_{j,\tau}^{cg}$$
(6)

$$\begin{aligned} \mathcal{A}_{j,t-1}^{m} &= \sum_{\substack{i=1\\i \neq j}} m_{ij,t-1} + \sum_{\tau=0} A_{j,\tau}^{pa} + \sum_{\tau=0} A_{j,\tau}^{cg} \\ \mathcal{R}_{ij,t-1}^{m} &= R_{ij,t-1}^{m} + R_{ij,t-1}^{pa} + R_{ij}^{D}. \end{aligned}$$

 $\alpha_i^m$  captures user *i*'s intrinsic preference for making friends and follows a normal distribution with mean  $\bar{\alpha}^m$  and standard deviation  $\sigma_{\alpha^m}$ . The variable  $\mathcal{A}_{j,t-1}^m$  describes user *j*'s attractiveness as a potential friend. It only depends on *j*'s attributes and reflects how popular and/or knowledgeable user *j* is. We operationalize  $\mathcal{A}_{j,t-1}^m$  as the number of user *j*'s actions in the three areas, i.e., *j*'s number of friends,  $\sum_{i=1, i\neq j} m_{ij,t-1}$ , number of adopted animes,  $\sum_{\tau=0}^{t-1} \mathcal{A}_{j,\tau}^{pa}$ , and number of posts in the UGC part of the website,  $\sum_{\tau=0}^{t-1} \mathcal{A}_{j,\tau}^{cg}$ . The first variable describes the utility gained from becoming friends with popular users (Tong et al. 2008; Langlois et al. 2000), while the latter two represent the utility gained from information sharing and learning from friends who are knowledgeable about animes (Watson and Johnson 1972; Brandtzæg and Heim 2009).

The variable  $\mathcal{R}_{ij,t-1}^m$  is tie-specific and captures the similarity between individual i and individual j.  $\mathcal{R}_{ij,t-1}^m$  includes the number of common friends,  $R_{ij,t-1}^m$ , the number of common animes,  $R_{ij,t-1}^{pa}$ , at time t - 1, and demographic similarity between user i and user j in terms of age, gender, and country of origin,  $R_{ij}^D$ . 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 et al. 2007).  $R_{ij}^D$  additionally includes three dummy variables that indicate whether age, gender, or location information of both individual i and individual j are available.<sup>17</sup>

 $C_{it}^m$  contains several variables whose effects we control for. The first variable is the length of time (in days) since user *i* joined the website. As revealed in Figure 5, new users are more likely to add friends compared to experienced users since having friends in the beginning helps to reduce learning costs associated with navigating the website. Second, we also include weekend dummies. Third, we also include a dummy variable indicating whether use *j* was active on the platform during the previous week. This variable captures user *j*'s visibility.<sup>18</sup> Fourth, we include time fixed effects to address common shocks that might result in simultaneity of friendship actions across users. We operationalize these time fixed effects as week dummies.<sup>19</sup> And lastly, we assume that  $\epsilon_{i,t}^m$  follows normal distribution.

<sup>&</sup>lt;sup>17</sup>To address the potential bias in the estimation results due to not modeling the formation of the full network of non-core users, we further estimate separate coefficients of  $\mathcal{A}_{j,t-1}^m$ ,  $\mathcal{R}_{ij,t-1}^m$ , and  $R_{ij}^D$  for core and non-core users.

<sup>&</sup>lt;sup>18</sup>In addition to user *i*'s preference for friendship with user *j*, both  $\mathcal{A}_{j,t-1}^m$  and  $\mathcal{R}_{ij,t-1}^m$  also capture the degree to which user *j* is visible to user *i*. Unlike previous studies in the strategic network formation literature (e.g., Christakis et al. 2010; Snijders, Koskinen and Schweinberger 2010) which model the meeting and the decision to become friends separately, we follow the conventional approach in the choice model literature and model the combined effect of visibility and preference in the utility.

<sup>&</sup>lt;sup>19</sup>While it would be desirable to include daily dummy variables, for computational reasons (see Section 5), 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 many months to converge.

User *i*'s utility from watching an anime,  $U_{i,t}^{pa}$ , is given by

$$U_{i,t}^{pa} = \alpha_i^{pa} + \beta^{pa} \mathcal{A}_{i,t-1}^{pa} + \gamma^{pa} \mathcal{F}_{i,t-1}^{pa} + \lambda^{pa} C_{it}^{pa} + \epsilon_{i,t}^{pa},$$
  
where  
$$\mathcal{A}_{i,t-1}^{pa} = A_{i,t-1}^{pa}$$
  
$$\mathcal{F}_{i,t-1}^{pa} = \sum_{\substack{j=1\\j\neq i}} m_{ij,t-1} + \sum_{\substack{j\in\{m_{ij,t-1}=1\}}} A_{j,t-1}^{pa} + \sum_{\substack{j\in\{m_{ij,t-1}=1\}}} A_{j,t-1}^{cg},$$
  
(7)

where  $\alpha_i^{pa}$  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  $\sigma_{\alpha^{pa}}$ .  $\mathcal{A}_{i,t-1}^{pa}$  captures the state dependence of anime watching and is operationalized as a dummy variable which equals 1 if user *i* watched an anime at t-1 and 0 otherwise. Next,  $\mathcal{F}_{i,t-1}^{pa}$  captures the effects of user *i*'s friendship network on user *i*'s actions. It includes user *i*'s number of friends,  $\sum_{j=1, j\neq i} m_{ij,t-1}$ , the number of animes watched by all of user *i*'s friends at t-1,  $\sum_{j\in\{m_{ij,t-1}=1\}} \mathcal{A}_{j,t-1}^{pa}$ , and the number of posts written by all of *i*'s friends at t-1,  $\sum_{j\in\{m_{ij,t-1}=1\}} \mathcal{A}_{j,t-1}^{cg}$ . Previous literature has shown that having more friends might directly increase the level of social activities of network users (Toubia and Stephen 2013; Shriver et al. 2013).<sup>20</sup> 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 over user *i*'s activity in anime watching.

Furthermore,  $C_{it}^{pa}$  contains several variables whose effects we control for. It includes a weekend dummy and week fixed effects. And lastly, we assume that  $\epsilon_{i,t}^{pa}$  is normally distributed.

 $<sup>^{20}</sup>$ 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.

Similarly, user *i*'s utility from writing a post,  $U_{i,t}^{cg}$ , is given by

$$U_{i,t}^{cg} = \alpha_i^{cg} + \beta^{cg} \mathcal{A}_{i,t-1}^{cg} + \gamma^{cg} F_{i,t-1}^{cg} + \lambda^{cg} C_{it}^{cg} + \epsilon_{i,t}^{cg},$$
  
where  
$$\mathcal{A}_{i,t-1}^{cg} = A_{i,t-1}^{cg} + \sum_{\tau=0}^{t-1} A_{i,\tau}^{pa}$$
  
$$\mathcal{F}_{i,t-1}^{cg} = \sum_{\substack{j=1\\ j\neq i}} m_{ij,t-1} + \sum_{j\in\{m_{ij,t-1}=1\}} A_{j,t-1}^{pa} + \sum_{j\in\{m_{ij,t-1}=1\}} A_{j,t-1}^{cg},$$
  
(8)

where  $\alpha_i^{pa}$  is user *i*'s intrinsic tendency to produce content and follows a normal distribution with mean  $\bar{\alpha}^{cg}$  and standard deviation  $\sigma_{\alpha^{cg}}$ .  $\mathcal{A}_{i,t-1}^{cg}$  represents user *i*'s past actions and contains two variables. The first variable,  $A_{i,t-1}^{cg}$ , captures state dependence and is operationalized as a dummy variable indicating whether user *i* wrote a post at t-1. In addition, user *i*'s past anime watching behavior may also influence her posting decision because a user who watches more animes may have more things to talk about. Therefore, we also include the cumulative number of animes watched by user i,  $\Sigma_{\tau=0}^{t-1} A_{i,\tau-1}^{pa}$ , as a covariate in Equation 8. The variable  $\mathcal{F}_{i,t-1}^{cg}$  describes the effects of user *i*'s friendship network on user *i*'s actions and is defined in a similar manner as  $\mathcal{F}_{i,t-1}^{pa}$  in Equation (7), i.e., it includes user *i*'s number of friends,  $\sum_{\substack{j=1\\ j\neq i}} m_{ij,t-1}$ , the number of animes watched by all of *i*'s friends at t-1,  $\sum_{j\in\{m_{ij,t-1}=1\}} A_{j,t-1}^{pa}$ the number of posts written by all of user *i*'s friends at t-1,  $\sum_{j\in\{m_{ij,t-1}=1\}} A_{j,t-1}^{pa}$ . As in the previous  $j\in\{m_{ij,t-1}=1\}$  contains a weekend dummy and week fixed effects. And lastly,  $\epsilon_{i,t}^{cg}$  follows a normal distribution.<sup>21</sup>

### 5 Estimation

Given the conditional independence assumption of user i's decision to become friends with each user j (as discussed in Section 4.1) and given the need for mutual agreement to become

<sup>&</sup>lt;sup>21</sup>All continuous variables in the three utility functions are incorporated in the form of natural logarithms.

friends, the probability of a tie forming between individual i and individual j is given by

$$P(m_{ij,t} = 1) = P(U_{i,t}^m > 0) \cdot P(U_{j,t}^m > 0).$$
(9)

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

$$L^{m}_{ij,t|\alpha_{i},\alpha_{j},\epsilon_{i},\epsilon_{j}} = \left[ \left[ P(m_{ij,t}=1) \right]^{m_{ij,t}} \left[ 1 - P(m_{ij,t}=1) \right]^{1-m_{ij,t}} \right]^{1-m_{ij,t-1}},$$
(10)

where  $\alpha_i = \{\alpha_i^m, \alpha_i^{pa}, \alpha_i^{cg}\}$  and  $\alpha_j$  is defined similarly. Note that  $L^m_{ij,t|\alpha_i,\alpha_j,\epsilon_i,\epsilon_j}$  conditions on the two users not being friends before time t through the exponent  $1 - m_{ij,t-1}$ . Given the above equation, the likelihood of all friendship formations of user i in time period t is denoted by

$$L_{i,t|\alpha,\epsilon}^{m} = \prod_{j=1}^{N} \left[ \left[ P(m_{ij,t}=1) \right]^{m_{ij,t}} \left[ 1 - P(m_{ij,t}=1) \right]^{1-m_{ij,t}} \right]^{1-m_{ij,t-1}} \qquad i \neq j$$
(11)

where  $\alpha = \{\alpha_1, \ldots, \alpha_N\}$  and  $\epsilon = \{\epsilon_1, \ldots, \epsilon_N\}.$ 

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

$$L_{i,t|,\alpha_{i},\epsilon_{i}}^{pa} = [P(A_{i,t}^{pa}=1)]^{A_{i,t}^{pa}} [1 - P(A_{i,t}^{pa}=1)]^{1 - A_{i,t}^{pa}}$$

$$L_{i,t|,\alpha_{i},\epsilon_{i}}^{cg} = [P(A_{i,t}^{cg}=1)]^{A_{i,t}^{cg}} [1 - P(A_{i,t}^{cg}=1)]^{1 - A_{i,t}^{cg}}.$$
(12)

Combining equations (11) and (12), the joint likelihood of user i's actions at time t is

denoted by

$$L_{i,t|\alpha} = \int_{-\infty}^{+\infty} \left[ P(A_{i,t}^{pa} = 1) \right]^{A_{i,t}^{pa}} \left[ 1 - P(A_{i,t}^{pa} = 1) \right]^{1 - A_{i,t}^{pa}} \cdot \left[ P(A_{i,t}^{cg} = 1) \right]^{A_{i,t}^{cg}} \left[ 1 - P(A_{i,t}^{cg} = 1) \right]^{1 - A_{i,t}^{cg}} \cdot \prod_{j=1}^{N} \left[ (P(m_{ij,t} = 1))^{m_{ij,t}} \left[ 1 - P(m_{ij,t} = 1) \right]^{1 - m_{ij,t}} \right]^{1 - m_{ij,t-1}} d\epsilon \qquad i \neq j.$$

$$(13)$$

The full likelihood can be written as

$$L = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \prod_{t=1}^{T} \prod_{i=1}^{N} \left[ P(A_{i,t}^{cg} = 1) \right]^{A_{i,t}^{pa}} \left[ 1 - P(A_{i,t}^{pa} = 1) \right]^{1 - A_{i,t}^{pa}} \\ \cdot \left[ P(A_{i,t}^{cg} = 1) \right]^{A_{i,t}^{cg}} \left[ 1 - P(A_{i,t}^{cg} = 1) \right]^{1 - A_{i,t}^{cg}} \\ \cdot \prod_{j=i+1}^{N} \left[ \left[ P(m_{ij,t} = 1) \right]^{m_{ij,t}} \left[ 1 - P(m_{ij,t} = 1) \right]^{1 - m_{ij,t}} \right]^{1 - m_{ij,t-1}} d\epsilon d\alpha$$
(14)

and the log likelihood of the model given by

$$LL = \log \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \prod_{t=1}^{T} \prod_{i=1}^{N} \left[ P(A_{i,t}^{cg} = 1) \right]^{A_{i,t}^{pa}} \left[ 1 - P(A_{i,t}^{pa} = 1) \right]^{1 - A_{i,t}^{pa}} \\ \cdot \left[ P(A_{i,t}^{cg} = 1) \right]^{A_{i,t}^{cg}} \left[ 1 - P(A_{i,t}^{cg} = 1) \right]^{1 - A_{i,t}^{cg}} \\ \cdot \prod_{j=i+1}^{N} \left[ \left[ P(m_{ij,t} = 1) \right]^{m_{ij,t}} \left[ 1 - P(m_{ij,t} = 1) \right]^{1 - m_{ij,t}} \right]^{1 - m_{ij,t-1}} d\epsilon d\alpha.$$
(15)

We estimate our model using Simulated Maximum Likelihood (SMLE). To estimate the full covariance matrix of user random effects, we take 30 draws from a standard normal distribution for each user and each activity and use the Cholesky decomposition of the covariance matrix. Similarly, to estimate the full covariance matrix of the three error terms, we take 30 draws from a standard normal distribution for each user and each activity in each time period and use the Cholesky decomposition of the error covariance matrix in the estimation.

For computational reasons, the conventional approach to estimate a model via MLE and SMLE involves taking the logarithm of the model likelihood in order to convert an extremelysmall-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)).<sup>22</sup> Therefore the integral taken over  $\Sigma$  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 error terms in users' friendship formation utilities. In other words, due to the second reason, the integral over  $\Sigma$ has to to include all friendship formation probabilities of all users at time t. Combining the first and second reason, it is evident that the integral over  $\Sigma$  has to include the probabilities of all actions of all users at time t.

Third, recall that our model includes time-invariant individual-specific preferences for each type of activity, i.e.  $\alpha_i^m$ ,  $\alpha_i^{pa}$ , and  $\alpha_i^{cg}$ . Therefore, for each user and each type of activity, the integral over  $\alpha$  has to include all activities of that type over all time periods. Given that the first two reasons necessitate that the integral over  $\Sigma$  contains the probabilities of all actions of all users at each time t and given that the third reason necessitates that the integral over  $\alpha$  contains all probabilities over all time period for a specific type of activity and a specific user, the integrals over both  $\Sigma$  and  $\alpha$  have to contain the probabilities of all actions of all users over all time periods (see Equation (14)). As a result of these three issues, when we take the logarithm of the model likelihood, we *cannot* convert the product

<sup>&</sup>lt;sup>22</sup>Another reason is that the individual-specific intrinsic propensities for each type of activity,  $\alpha_i^m$ ,  $\alpha_i^{pa}$ , and  $\alpha_i^{cg}$ , are correlated as well.

of the probabilities into a sum of the logarithms of these probabilities (see Equation (15)). 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.<sup>23</sup> To make the likelihood estimation computationally tractable, we use a transformation of the logarithm of sum of variables to a function of the logarithm of those variables. Details of the transformation are presented in Appendix A.

To speed up the estimation, we use OpenBLAS as the system BLAS (Basic Linear Algebra Subprograms), tensorization of large matrices and parallel computing methods to run the estimation program. Due to the large size of the data and parallelization, we cannot run the estimation code on conventional computing systems.<sup>24</sup> We utilize several large memory super-computing servers including the Texas Advanced Computing Center (TACC) and Jetstream cloud-computing (Stewart and Gaffney 2015; John Towns 2014).<sup>25</sup>

### 6 Identification

The set of parameters to be estimated is given by  $\{\bar{\alpha}^m, \bar{\alpha}^{pa}, \bar{\alpha}^{cg}, \Sigma^{\alpha}, \beta^m, \beta^{pa}, \beta^{cg}, \lambda^m, \lambda^{pa}, \lambda^{cg}, \gamma^m, \gamma^{pa}, \gamma^{cg}, \Sigma\}$ . The identification of  $\{\beta^m, \beta^{pa}, \beta^{cg}, \lambda^m, \lambda^{pa}, \lambda^{cg}\}$  is standard.

The mean intrinsic propensities,  $\{\bar{\alpha}^m, \bar{\alpha}^{pa}, \bar{\alpha}^{cg}\}$ , are identified by the average user behavior across users and across times. The covariance matrix of the user random effects,  $\Sigma^{\alpha}$ , is identified by the variation in average activity levels across users. In contrast, the covariance matrix of the error terms,  $\Sigma$ , is identified by the variation in the co-occurance of activities on the same day.

The parameter  $\gamma^m$  captures the effects of common factors for each pair of users and is identified by the variation in the percentage of common friends and the percentage of common animes among different pairs. The parameters { $\gamma^{pa}, \gamma^{cg}$ } capture friends' influence

 $<sup>^{23}</sup>$ For the interested reader, the likelihood is given by the product of over 136,000,000 probabilities.

<sup>&</sup>lt;sup>24</sup>The estimation code requires at least 350GB of RAM.

 $<sup>^{25}\</sup>mathrm{It}$  takes more than 3 weeks to estimate a model with 49 parameters on a super computer utilizing 32 CPU cores using our data containing 68 million observations.

on a user's actions and are identified by changes in average behavior of friends. Lastly, conditional on  $\Sigma$ , 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. The similarity in unobserved preferences, if unaccounted for, can lead to correlated errors which, in turn, lead to upward biased estimates of friends' influence. We address this issue by incorporating unobserved timeinvariant components,  $\alpha_i^m$ ,  $\alpha_i^{pa}$ , and  $\alpha_i^{cg}$ , in a user's decisions to form ties, to adopt animes, and to generate content (similar approach as in Nair et al. 2010 and Trusov et al. 2010). 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  $\alpha_i^m$ ,  $\alpha_i^{pa}$ , and  $\alpha_i^{cg}$ . Furthermore,  $\alpha_i^m$ ,  $\alpha_i^{pa}$ , and  $\alpha_i^{cg}$  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 propensity is 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 correlation 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  $\alpha_i^m$ ,  $\alpha_i^{pa}$ , and  $\alpha_i^{cg}$ .

## 7 Results

We present the estimation results in Table 2. As discussed in Section 3.1, we estimate separate coefficients for core and non-core users. The estimation results for non-core users are also presented in a separate column in Table 2. In the following, we focus on discussing the results for the core users. Column (i) in the 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. Lastly, column (iii) depicts the results for our full model in which we allow for both correlated errors and unobserved heterogeneity among consumers. Overall, the results across the three different specifications are consistent. Potential simultaneity among the three types of actions a user might engage in each day is captured through the correlated errors. We find significant correlations among two of the three actions.<sup>26</sup> We also find significant coefficients for the std. deviations of the individual-specific random effects suggesting the presence of heterogeneity in intrinsic propensities across users. However, the correlations among the random effects are insignificant.

Insert Table 2 about here

\_\_\_\_\_

\_\_\_\_\_

We first discuss the parameter estimates for the friendship formation utility for the 400 core users in our sample. User j's number of friends, j's number of watched animes, and j's number of written posts represent the attractiveness 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

 $<sup>^{26}</sup>$ The correlation between product adoption and UGC production is insignificant.

unique context of the online environment. Next, we find that user j's cumulative number of watched animes and his 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.

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 the unobserved match between individual i and individual j. This finding is in line with results in the previous literature (e.g., Aral et al. 2009; Shalizi and Thomas 2011) suggesting that "birds of a feather flock together." The coefficient for the number of common animes, however, is insignificant. In terms of demographic similarities, we find positive coefficients for user i and user j being from the same country, being close in age, and having the same gender if both individuals reveal this information. However, the coefficients for dummies indicating whether both users provided the information are negative. In other words, knowing demographic information about each other only increases the chance of a friendship if both users are similar in those characteristics. Otherwise, it actually hurts the chance of forming a friendship tie.

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. his number of watched animes and his 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 the increased visibility and awareness of user j that comes with activity. Next, the weekend dummy has a significant negative coefficient. Users are less likely to make friends on weekends. And lastly, as expected, user i's utility of making friends declines with the length of her membership on the platform since, especially shortly after joining, having friends helps to reduce the learning costs associated with navigating the website.

Next, we discuss our results for anime watching. We find a negative significant effect for the cumulative number of friends implying that having more friends does not make a user more active. The number of animes watched by friends reflects the influence friends have on a user's anime watching behavior. We find a significant positive coefficient for the effect of the number of animes watched by friends the previous day indicating the existence of peer effects. However, we do not find any spill-over effects of friends' posting behavior on a user's anime watching. Our results also reveal a 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 an insignificant coefficient for the cumulative number of friends indicating that users with more friends do not publish more content. We also find evidence for friends' influence on users' content generation behavior. The number of posts published by friends has a significant positive effect on a user's content generation decision. However, there is no spill-over effect. The number of animes watched by friends does not have a significant effect on a user's UGC production. 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. Additionally, as discussed in Section 4.4, we model a user's utility of UGC as a function of the number of animes watched by that user. We find a significant positive effect of the 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 do not find a significant effect for the weekend dummy.

To summarize, our results for friendship formation reveal that both the attractiveness of and similarity with a potential friend matter. Further, the number and overlap of users' friends are more important drivers of friendship formation than product adoption and content generation activities. And lastly, even in (anonymous) online networks having similar demographics matters. With regard to online activities of anime watching and content generation, we find evidence of significant peer effects. 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 result in more activity.

## 8 Counterfactual

For companies operating social networks, advertising revenue represents their primary source of income. In 2015, the industry had revenues of over \$25 billion through advertisements.<sup>27</sup> 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 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.

The existence of peer effects within social networks implies that an increase in a user's activity level can have a cascading effect on the user's friends and friends of friends and so on. Previous literature has also shown that seeding to more connected users is the most effective

<sup>&</sup>lt;sup>27</sup>https://www.statista.com/statistics/271406/advertising-revenue-of-social-networks-worldwide/

way of increasing the total number of product adoptions within a community (e.g., Hinz et al. 2011; Aral et al. 2013). However, the studies in this area assume a static network structure that does not evolve over time. While this result may hold true for mature networks where a static assumption applies, a stimulation intervention in evolving networks is very likely to also lead to a change in the structure of the network due to the possibility of newly formed 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. 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 using a recommendation system. For example, the platform can recommend 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 nonusers of the website and therefore can lead to activity cascades as well as a growing user base.

For the simulations, we use the state of the network for all users in our sample on day 150. Note that on day 150, only 1,194 out of the 1,386 users in our sample were members of the website. The remaining users joined the website sometime between day 151 and 184 and their actions are simulated from the time they join. Furthermore, as discussed in Section 3.1, we simulate the actions of all core users and all non-core users going forward until day 184. However, we take the actions of friends of non-core users as exogenous and adjust the relevant independent variables for non-core users. Furthermore, we only compare the changes in activity levels of core users for whom we estimated the model based on their full network.

# 8.1 How to Increase In-Site Activities through Platform-Wide Stimulation System?

In the first set of simulations, we examine and compare the overall activity level of all core users due to the implementation of different platform-wide stimulation systems where taking the evolving network structure due to the stimulation into account. Here we sum the three different activities to make the overall activity level as a proxy for the total site traffic or total time spent on the site. In each scenario, we increase one type of activity (friendship ties, product adoptions or UGC generation) among all core and all non-core users by one standard deviation on day 150 and simulate users' behavior going forward until day 184. Figure 8 plots daily overall activity levels and the number of active users among all core users. Out of the three types of stimulation strategies, stimulating users to watch more animes leads to the highest overall activity level among users, 35% higher than the result of the stimulation strategy to have users make more friends and two times higher than the result of the UGC generation stimulation. However, in terms of the number of users who engage in at least one activity, the stimulation strategy to encourage users to make more friends slightly outperforms product adoption stimulation strategy (by 1%).

These findings imply that out of the three recommendation systems the platform can implement (i.e., to recommend friends, animes, or forum discussion topics), the anime recommendation system that increases users' anime watching is the most effective in driving the overall site traffic while the friending recommendation system that increases the number of friendships is the most effective in producing more active users. Insert Figure 8 about here

\_\_\_\_\_

\_\_\_\_\_\_

### 8.2 How to Increase In-Site Activities through Seeding?

Previous literature has found that not all users have the same degree of influence on their friends (e.g., Manchanda et al. 2008; Iyengar et al. 2011). Further, users can also have varying degrees of activity in different areas. 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 the second set of simulations, we examine the effectiveness of different seeding strategies in increasing tie formations and UGC production within the network. For these simulations, we select the top/bottom 50 most/least active users (about 15% of core users) among the core users based on their activity levels in making friends, adopting animes, producing UGC or conducting any activity.<sup>28</sup> Next, we increase the activity level for these selected users in one of the three activities by one standard deviation in the beginning of day 150 and simulate the network evolution and users' activities until day 184.

We report the results in Table 3. The numbers in Table 3 represent the percentage changes in users' activities compared to the baseline scenario without any stimulation. First, we discuss the effectiveness of seeding strategies in increasing the number of friendships. The two best strategies are to target the bottom 50 users based on their anime watching and to stimulate their friend making. The number of formed friendships increases by 34% when this strategies is applied. In general, motivating users to make friends is the best strategy to increase friendship ties in a network. Furthermore, promoting UGC production is more effective than promoting adoptions to increase connectedness within the network.

 $<sup>^{28}{\</sup>rm For}$  the selection of users in the "any activity" category, we use the sum of the standardized activity levels in the three areas.

#### Insert Table 3 about here

\_\_\_\_\_

\_\_\_\_\_

The best two strategies for increasing UGC production on the website are selecting the top 50 users in terms of UGC production and the bottom 50 users in terms of anime watching and stimulating them to watch more animes. These two strategies result in a 16% and 18% increase in UGC production, respectively. Our simulation results show that, if platform owners want to increase UGC on their website, promoting post publishing behavior is far less effective that promoting product adoptions. Furthermore, our results reveal that selecting users based on their number of friends is not always the best seeding strategy for firms. This result is also in line with Katona et al. (2011)'s finding that average influential power of individuals decreases with their total number of contacts. Although in comparing the effectiveness of selecting the top 50 and the bottom 50 users in terms of their number of friends, selecting the better connected users generally results in more UGC, this strategy is still far less effective (by about 150%) than selecting the top 50 users based on their UGC production and stimulate them to watch more animes.

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 counterfactual scenarios 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 10%.

The best two strategies for increasing UGC production on the website are to select the top 50 users in UGC production and the bottom 50 users in anime watching and to have them watch more animes. These two strategies result in a 16% and 18% increase in UGC production, respectively. Our simulation results show that, if platform owners want to increase UGC on their website, promoting post publishing behavior is far less effective that promoting anime watching. Furthermore, our results reveal that targeting the most connected users is not always the best seeding strategy for firms. We find that well-connected users in the network tend to become friends with other well-connected users who are not necessarily the ones who would produce a lot of UGC or watch many animes. Because of this non-overlap of users with a lot of friends and users with a lot UGC, selection based on the metric of number of friends is not the best strategy. This result is in line with Katona et al. (2011) in which the authors find that average influential power of individuals decreases with their total number of contacts. Although seeding the more connected users generally leads to more UGC production than seeding the less connected users, this strategy is still far less effective (by about 150%) than targeting the top 50 users in UGC production and stimulating them to watch more animes.

Lastly, we examine to what extent the effectiveness of seeding strategies is underestimated when the endogenous network formation is not accounted for. To do so, we re-run the counterfactual scenarios discussed in this section, but do not allow users to form new friendship ties. We find that not accounting for the endogenous network formation leads, on average, to an underestimation of seeding effectiveness by 10%.

## 9 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 do not observe the first request for the friendship and 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 tendency to form friendships or due to an increase of attractiveness of that user 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 action 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 watch an anime and not where and how many posts users write or what particular anime they watch. Studying the details of each action can shed further light on the co-evolution process of users' friendship formations and concurrent actions and is left for future research. Fourth, we do not model platform growth in our relatively short observation period, i.e., we do not model users' joining behavior and assume it is exogenous. However, in the long run, popularity of a platform in terms of its user base and richness of its content can change the rate of users' joining the website. And lastly, in this paper, we do not consider the content of users' posts. Longer or more detailed posts may imply the writer is more knowledgeable and thus more attractive. However, studying the effects of characteristics of users' UGC is left for future research.

### 10 Conclusion

We develop a structural model for the co-evolution of individuals' friendship tie formations and their concurrent online activities (product adoptions and production of user-generated content) within a 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) provides a clean identification of peer effects and of important drivers of individuals' friendship decisions. 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, compared to a potential friend's product adoptions and content generation activities, the total number of friends and the number of common friends this potential friend has with the focal individual are the most important drivers of friendship formation. Further, while having more friends does not 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 effects. Via counterfactuals 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 et al. 2013), we find that seeding to users with the most friends is not always the best strategy to increase users' activity levels on the website.

## References

- Ameri, Mina, Elisabeth Honka and Ying Xie (2016), 'Word-of-mouth, observational learning, and product adoption: Evidence from an anime platform', *Working Paper*.
- 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.
- 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.
- Aral, Sinan, Lev Muchnik and Arun Sundararajan (2013), 'Engineering social contagions: Optimal network seeding in the presence of homophily', Network Science 1(02), 125–153.
- Badev, Anton (2013), 'Discrete games in endogenous networks: Theory and policy', *Working Paper*.
- Boguñá, Marián, Romualdo Pastor-Satorras, Albert Díaz-Guilera and Alex Arenas (2004), 'Models of social networks based on social distance attachment', *Physical Review E* **70**(5), 056122.
- Bramoullé, Yann, Habiba Djebbari and Bernard Fortin (2009), 'Identification of peer effects through social networks', *Journal of Econometrics* **150**(1), 41–55.
- Brandtzæg, Petter Bae and Jan Heim (2009), Why people use social networking sites, *in* 'International Conference on Online Communities and Social Computing', Springer, pp. 143– 152.
- Carrell, Scott, Richard Fullerton and James West (2009), 'Does your cohort matter? measuring peer effects in college achievement', *Journal of Labor Economics* **27**(3), 439–464.
- 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*.
- Claussen, Jörg, Benjamin Engelstätter and Michael Ward (2014), 'Susceptibility and influence in social media word-of-mouth', ZEW-Centre for European Economic Research Discussion Paper (14-129).

- 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.
- Hanaki, Nobuyuki, Alexander Peterhansl, Peter S Dodds and Duncan J Watts (2007), 'Cooperation in evolving social networks', *Management Science* **53**(7), 1036–1050.
- Hartmann, Wesley, Puneet Manchanda, Harikesh Nair, Matthew Bothner, Peter Dodds, David Godes, Kartik Hosanagar and Catherine Tucker (2008), 'Modeling social interactions: Identification, empirical methods and policy implications', *Marketing Letters* 19(3-4), 287–304.
- Hinz, Oliver, Bernd Skiera, Christian Barrot and Jan U Becker (2011), 'Seeding strategies for viral marketing: An empirical comparison', *Journal of Marketing* 75(6), 55–71.
- Hoff, Peter D, Adrian E Raftery and Mark S Handcock (2002), 'Latent space approaches to social network analysis', *Journal of the American Statistical Association* 97(460), 1090– 1098.
- Iyengar, Raghuram, Christophe Van den Bulte and Thomas W. Valente (2011), 'Opinion leadership and social contagion in new product diffusion', *Marketing Science* 30(2), 195– 212.
- Jackson, Matthew (2008), Social and Economic Networks, Princeton University Press.
- John Towns, Timothy Cockerill, Maytal Dahan Ian Foster Kelly Gaither Andrew Grimshaw Victor Hazlewood Scott Lathrop Dave Lifka Gregory D. Peterson Ralph Roskies J. Ray Scott Nancy Wilkins-Diehr (2014), 'Xsede: Accelerating scientific discovery', Computing in Science & Engineering 16(5), 62–74.
- Katona, Zsolt (2013), 'Competing for influencers in a social network', Working Paper.
- Katona, Zsolt, Peter Pal 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.
- Katona, Zsolt and Tamás F. Móri (2006), 'A new class of scale free random graphs', Statistics
   & Probability Letters 76(15), 1587–1593.

- Kozinets, Robert (1999), 'E-tribalized marketing?: The strategic implications of virtual communities of consumption', *European Management Journal* **17**(3), 252–264.
- Lampe, Cliff AC, 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', ACM, pp. 435–444.
- Langlois, Judith H, Lisa Kalakanis, Adam J 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.
- 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.
- McCulloch, Robert and Peter E Rossi (1994), 'An exact likelihood analysis of the multinomial probit model', *Journal of Econometrics* **64**(1), 207–240.
- Mele, Angelo (2017), 'A structural model of dense network formation', *Econometrica* **85**(3), 825–850.
- Moe, Wendy and Michael Trusov (2011), 'The value of social dynamics in online product ratings forums', *Journal of Marketing Research* **48**(3), 444–456.
- Nair, Harikesh S., 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.
- Sacerdote, Bruce (2001), 'Peer effects with random assignment: Results for dartmouth roommates', *The Quarterly Journal of Economics* **116**(2), 681–704.

- Shalizi, Cosma Rohilla and Andrew Thomas (2011), 'Homophily and contagion are generically confounded in observational social network studies', *Sociological Methods & Research* 40(2), 211–239.
- Shriver, Scott K, Harikesh S 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, Christian Steglich and Michael Schweinberger (2007), Modeling the coevolution of networks and behavior, Lawrence Erlbaum Associates Publishers, pp. 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.
- Snijders, Tom, Johan Koskinen and Michael Schweinberger (2010), 'Maximum likelihood estimation for social network dynamics', *The Annals of Applied Statistics* 4(2), 567–588.
- Stewart, C.A., Cockerill T.M. Foster I. Hancock D. Merchant N. Skidmore E. Stanzione D. Taylor J. Tuecke S. Turner G. Vaughn M. and N.I. Gaffney (2015), Jetstream: a selfprovisioned, scalable science and engineering cloud environment, in 'XSEDE Conference: Scientific Advancements Enabled by Enhanced Cyberinfrastructure'.
- 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 Tom, Brandon Van Der Heide, Lindsey Langwell and Joseph B 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.
- Trusov, Michael, Anand V Bodapati and Randolph E 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 W Johnson (1972), Social Psychology: Issues and insights, Lippincott Philadelphia.

# Figures and Tables



Figure 1: Dates Users Joined MyAnimeList.Net



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.



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







Figure 7: Number of Activities in Each Area Per Day

(a) Number of Activities Per Day (b) Number of Active Users Per Under Different Recommendation Day Under Different Recommen-Strategies dation Strategies



### Figure 8: Number of Activities and Active Users Per Day Under Different Recommendation Strategies

Colors Represent Seeding in Indicated Areas of Activities.

	Mean	Std. Dev.	Min	Median	Max	Ν
Age	19.29	5.20	12	18.5	78	1088
Gender (% Females)	38.5					
Gender ( $\%$ Males)	54					
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

	(;)		(;;`	\	(;;;)	)	
	(1) Independent		(11) Homographics		(III) Main Model		
	Core	Non-Core	Core	Non-Core	Core	Non-Core	
	0010	11011 0010	0010	1.011 0.010	0010	11011 0010	
Friendship Formation							
Attractiveness	0.400***	0.005***	0.400***	0.007***	0 40/2***	0.000***	
$j$ 's Number of Friends by $t - 1^{\alpha}$	$0.426^{****}$	0.097****	$0.426^{+++}$	0.097****	0.426***	0.098****	
	(0.010)	(0.010)	(0.010)	(0.010)	(0.009)	(0.009)	
$j$ 's Number of Watched Animes by $t - 1^{\circ}$	(0.008)	-0.036	0.007	-0.037	(0.007)	-0.037	
i's Number of Written Dogte by t 18	(0.010)	(0.010)	(0.011)	(0.011)	(0.001)	(.) 0.045***	
$j$ s Number of Written Posts by $i = 1^{-1}$	(0.000)	-0.040	(0.001)	-0.045	(0.002)	-0.045	
Similarity	(0.010)	(0.011)	(0.010)	(0.011)	(0.010)	(0.010)	
Number of Friends in Common with $i$	0 106***	-0.172***	0 106***	-0 173***	0.106***	-0.173***	
by $t = 1^{a}$	(0.100)	(0.026)	(0.100)	(0.026)	(0.100)	(0.026)	
Number of Animes in Common with $i$	-0.008	0.006	-0.008	0.008	(0.023)	0.008	
by $t = 1^{a}$	(0.000)	(0.014)	(0.003)	(0.000)	(0.000)	(0.000)	
Dummy for Whether Both $i$ and $i$	$-0.577^{***}$	0.516***	$-0.577^{***}$	0.517***	$-0.577^{***}$	$0.517^{***}$	
Indicate Their Country	(0.085)	(0.087)	(0.084)	(0.087)	(0.080)	(0.082)	
Dummy for Whether $i$ and $j$ Are	0.180***	-0.096	0.180***	-0.097	0.180***	-0.097	
from Same Country	(0.051)	(0.054)	(0.051)	(0.054)	(0.048)	(0.051)	
Dummy for Whether Both $i$ and $j$	$-0.327^{***}$	0.216***	$-0.327^{***}$	0.216***	$-0.327^{***}$	0.216***	
Indicate Their Age	(0.040)	(0.045)	(0.041)	(0.045)	(0.039)	(0.044)	
Dummy for Whether $i$ and $j$ Are	0.221***	-0.047	0.221***	-0.047	0.223***	-0.047	
Within 5 years of Age	(0.033)	(0.038)	(0.033)	(0.038)	(0.032)	(0.037)	
Dummy for Whether Both $i$ and $j$	-0.168***	0.085	$-0.167^{***}$	0.085	$-0.167^{***}$	0.086	
Indicate Their Gender	(0.047)	(0.051)	(0.048)	(0.051)	(0.046)	(0.050)	
Dummy for Whether $i$ and $j$	0.128***	$-0.077^{***}$	0.128***	$-0.077^{***}$	0.128***	$-0.077^{***}$	
Have the Same Gender	(0.026)	(0.029)	(0.026)	(0.029)	(0.025)	(0.028)	
Control Variables	0 100***	0 157***	0 101***	0 157***	0 101***	0 155***	
Dummy for Whether $j$ was Active from	(0.100)	$(0.157)^{(1)}$	$(0.101^{+1})$	(0.157)	(0.101)	$(0.15)^{(11)}$	
t - t to $t - 1$	(0.030)	(0.032)	(0.030)	(0.032)	(0.029)	(0.031)	
Dummy for whether $t$ is a weekend	-0.494	(0.075)	-0.495	(0.076)	-0.495	(0.072)	
Number of Membership Dave by ta	(0.075)	(0.073)	(0.075)	(0.070)	(0.073)	(0.073)	
Number of Membership Days by t	-0.5	02	-0.	002	(0.004)		
Dummy for <i>i</i> Boing Non Coro Usor	(0.0) _0.0_	/1***	) 	000) 040***	(0.004)		
Duminy for <i>i</i> being from-Core Oser	(0.0)	41 87)	-0	940 187)	-0.941 (0.076)		
Dummy for $i$ Having joined before	(0.087)		(0.0)	0.007		0.90/***	
July 2007	(0.203)		(0,000)		(0.008)		
Constant	(0.000)		(0.0	(0.009)		-1 575***	
Constant	(0.0)	81)	(0.0	185)	(0.0	(71)	
Standard Deviation of Friendship	0.025***		(0.000)		0.021***		
Bandom Effect	(0.0	(0.005)		(0.021)			
Week Dummies	ves		ves		ves		
	3.00		,00		3.00		
Madal Commencer Ctatistics							
Number of Observations	60 000 774		60 000 774		60 000 774		
Number of Observations	09,020,774	51	09,020,774	346	09,020,774	260	
	201,000.0	90 91	281,884.0	940 994	281,833.3	000 056	
DIU LogLikolihood	200, 000.0 142, 970.0	29 25	288,022.	924 000	288,702.0	390	
LogLikeimood	-143,879.0	20	-143,890.	020	-143,803.0	000	

Standard errors in parentheses. \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001

<sup>a</sup> Measured on logarithmic scale.

#### Table 2: Results

		(i)	(ii)		(iii)	
	Inde	ependent	Homogenous		Main Model	
	Core	Non-Core	Core	Non-Core	Core	Non-Core
Anime Watching						
Number of Friends by $t - 1^{a}$	-	-0.030***		-0.028***		-0.028***
		(0.005)		(0.005)		(0.005)
Number of Animes Watched by Friends in $t - 1^{a}$		0.082***		0.082***		0.082***
		(0.005)		(0.005)		(0.006)
Number of Posts Published by Friends in $t - 1^{a}$						0.003
						(0.006)
Dummy for Whether $i$ Watched an Anime		0.872***		0.872***		0.872***
$\operatorname{in} t - 1$		(0.011)		(0.011)		(0.011)
Dummy for Whether $t$ Is a Weekend		0.051***		0.051***		0.051***
		(0.009)		(0.009)		(0.009)
Constant	-	$-1.328^{***}$		$-1.328^{***}$		$-1.328^{***}$
		(0.008)		(0.008)		(0.008)
Standard Deviation of Anime		0.011***				$0.010^{*}$
Adoption Random Effect		(0.004)				(0.004)
Week Dummies	У	ves	3	yes		yes
Contant Commetica						
Content Generation		0.019		0.014		0.019
Number of Friends by $t - 1^{\circ}$		0.013		0.014		0.013
Number of Animar Wetched by Drive do in 4 18		(0.007)		(0.007)		(0.009)
Number of Animes Watched by Friends in $t - 1^{\circ}$						0.000
Number of Deste Dublished by Driveds in t 18		0.000***		0.000***		(0.008)
Number of Posts Published by Friends in $t - 1^{\circ}$		0.208		(0.208)		(0.208)
Durante for Whather i Dublished a Destin t 1		(0.008)		(0.008)		(0.008)
Dummy for whether i Published a Post in $t - 1$		$1.700^{-11}$		$1.700^{-1.1}$		$1.700^{-11}$
Number of Anima Wetched by t 18		(0.017)		(0.017)		(0.017)
Number of Animes watched by $t - 1^2$		$(0.032^{\circ})$		$(0.032^{\circ\circ})$		$(0.032^{\circ})$
Demonstra Whathan the Washend		(0.004)		(0.004)		(0.004)
Dummy for whether $t$ is a weekend		0.023		0.024		0.024
C		(0.014)		(0.014)		(0.014)
Constant	-	-2.201		-2.201		$-2.201^{\circ}$
Standard Deviation of UCC Draduction Dandars Effect		(0.019)		(0.019)		(0.019)
Standard Deviation of UGC Production Random Effect		(0.013)				$(0.010^{\circ})$
Week Duramies	_	(0.007)	_			(0.000)
week Dummes	y	es	3	yes		yes
Error Correlation Matrix						
Friendship Error Standard Deviation		1		1		1
		1		T		T
Adoption Error Standard Deviation		0.009***		0.014***		0.013***
		(0.004)		(0.004)		(0.004)
UGC Error Standard Deviation		0.005		0.012***		0.011***
		(0.006)		(0.005)		(0.004)
Correlation between Friendship and Adoption		(0.000)		0.729***		0.716***
correlation between ritenaship and rideption				(0.215)		(0.226)
Correlation between Friendship and UGC				$-0.874^{***}$		$-0.865^{***}$
ettersten seeneen triendship and e de				(0.219)		(0.237)
Correlation between Adoption and UGC				-0.449		-0.397
ettersten seeneen Haepvon and e de				(0.409)		(0.436)
Random Effects Correlation Matrix				(0.100)		(3.100)
Correlation between Friendship and Adoption						-0.426
						(0.290)
Correlation between Friendship and UGC						0.051
· · · · · · · · · · · · · · · · · · ·						(.)
Correlation between Adoption and UGC						-0.085
1						(0.241)

Standard errors in parentheses.

\* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001<sup>a</sup> Measured on logarithmic scale.

54 Table 2: Results (Continued.)

		Changes	in
SELECTION Activity	SEEDING Activity	Friendships in $\%$	UGC in %
Top 50 Selection			
-	Friend	4.91	0.87
Friendship	Anime	0.00	7.36
	Post	2.39	0.15
	Friend	6.97	1.56
Anime Watching	Anime	0.86	0.97
	Post	1.59	0.23
	Friend	5.11	2.22
UGC	Anime	0.33	15.92
	Post	1.26	0.00
Bottom 50 Selection			
	Friend	5.44	1.26
Friendship	Anime	1.20	1.88
	Post	1.79	0.71
	Friend	34.06	2.22
Anime Watching	Anime	0.47	17.53
	Post	7.04	0.23
	Friend	0.40	0.01
UGC	Anime	0.47	0.00
	Post	0.40	0.00

Table 3:	Counterfactual	Results
----------	----------------	---------

## **Appendix A: Likelihood Derivation**

In this section, we explain the estimation techniques used to estimate the likelihood. Recall that the full likelihood is

$$L = \int_{-\infty}^{+\infty} \prod_{t=1}^{T} \prod_{i=1}^{N} (P(A_{i,t}^{cg} = 1))^{A_{i,t}^{pa}} (1 - P(A_{i,t}^{pa} = 1))^{1 - A_{i,t}^{pa}} \cdot (P(A_{i,t}^{cg} = 1))^{A_{i,t}^{cg}} (1 - P(A_{i,t}^{cg} = 1))^{1 - A_{i,t}^{cg}} \cdot \prod_{j=i+1}^{N} \left[ (P(m_{ij,t} = 1))^{m_{ij,t}} (1 - P(m_{ij,t} = 1))^{1 - m_{ij,t}} \right]^{1 - m_{ij,t-1}} .d\Sigma.d\alpha.$$
(A1)

The log-likelihood is calculated by taking log of the likelihood

$$LL = \log \int_{-\infty}^{+\infty} \prod_{t=1}^{T} \int_{-\infty}^{+\infty} \prod_{i=1}^{N} (P(A_{i,t}^{cg} = 1))^{A_{i,t}^{pa}} (1 - P(A_{i,t}^{pa} = 1))^{1 - A_{i,t}^{pa}} \cdot (P(A_{i,t}^{cg} = 1))^{A_{i,t}^{cg}} (1 - P(A_{i,t}^{cg} = 1))^{1 - A_{i,t}^{cg}} \cdot \prod_{j=i+1}^{N} \left[ (P(m_{ij,t} = 1))^{m_{ij,t}} (1 - P(m_{ij,t} = 1))^{1 - m_{ij,t}} \right]^{1 - m_{ij,t-1}} .d\Sigma.d\alpha.$$
(A2)

First, we explain how we estimate the integral using simulation. First, for each user-time combination, we draw R random draws from standard normal distribution, which will be used as error terms in Cholesky decomposition of covariance of actions. Next, for each set of user-time draws out of the R draws, each probability is calculated at its drawn value and the final value of the expression inside the integral is calculates. Lastly, we take average over the R final values of each drawn set. Given the law of large numbers and the aforementioned

process, the log-likelihood is given by

$$LL = \log \frac{1}{R} \sum_{r=1}^{R} \left[ \prod_{t=1}^{T} \prod_{i=1}^{N} (P(A_{i,t}^{cg} = 1))^{A_{i,t}^{pa}} (1 - P(A_{i,t}^{pa} = 1))^{1 - A_{i,t}^{pa}} \\ \cdot (P(A_{i,t}^{cg} = 1))^{A_{i,t}^{cg}} (1 - P(A_{i,t}^{cg} = 1))^{1 - A_{i,t}^{cg}} \\ \cdot \prod_{j=i+1}^{N} \left[ (P(m_{ij,t} = 1))^{m_{ij,t}} (1 - P(m_{ij,t} = 1))^{1 - m_{ij,t}} \right]^{1 - m_{ij,t-1}} \right] \Big|_{r} \\ = -\log R + \log \sum_{r=1}^{R} Q_{r},$$
(A3)

where the calculated values of the expression inside integration for each  $r = 1 \dots R$  are denoted as  $Q_1 \dots Q_R$ . Since the number of probabilities being multiplied in each drawn set is large  $(N * N * \frac{N(N-1)}{2})$  and each probability is a small number, the final value calculated for each set will be extremely small, and most likely not processed properly by computer. In order to bypass this limitation, we use the following transformation

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

Thus we can transit the log into the sum, i.e.

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

where

$$\log Q = \sum_{t=1}^{T} \sum_{i=1}^{N} \left[ A_{i,t}^{pa} \cdot \log(P(A_{i,t}^{cg} = 1)) + (1 - A_{i,t}^{pa}) \cdot \log(1 - P(A_{i,t}^{pa} = 1)) + (A_{i,t}^{cg} \cdot \log(P(A_{i,t}^{cg} = 1)) + (1 - A_{i,t}^{cg}) \cdot \log(1 - P(A_{i,t}^{cg} = 1)) + \sum_{j=i+1}^{N} \left[ (1 - m_{ij,t-1}) \cdot [m_{ij,t} \cdot \log(P(m_{ij,t} = 1)) + (1 - m_{ij,t}) \cdot \log(1 - P(m_{ij,t} = 1))] \right] \right].$$
(A6)