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

Current version: April 2022

Mina Ameri[†] Elisabeth Honka[‡] Ying Xie[§]

Abstract

The authors study how strangers become friends within an evolving online social network. By modeling the co-evolution of individual users' friendship tie formations (when and with whom) and their concurrent online activities, the authors uncover important drivers underlying individuals' friendship decisions and, at the same time, quantify the resulting peer effects on individuals' actions. They estimate their model using a novel data set capturing the continuous development of a network and users' entire action histories within the network. The results reveal that similarity (homophily) with a potential friend, the properties of a potential friend's network, and her domain expertise all play a role in friendship formation. Via prediction exercises, the authors find that stimulating anime watching is the most effective site-wide intervention which leads to the highest overall site traffic and the largest number of active users and that recommending a friend of a friend as a potential friend is the most effective strategy in stimulating friendship tie formation. In contrast to the common finding for static networks, the results indicate that seeding to users with the most friends is *not* the most effective strategy to increase users' activity levels in an evolving network.

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

JEL Classification: D83, L82, M31

.

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

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

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

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

Online social networks have become an indispensable part of many individuals' everyday lives. By 2020, there were over 3.6 billion social network users worldwide with the average user having more than eight accounts across different platforms.[1] The growing connectedness among users as well as the increasing number of interactions and activities individuals engage in in online social networks provide ample opportunities for social network companies to grow revenues and profits. In 2020, despite the COVID-19 pandemic, these companies made an estimated revenue of 41.5 billion from social media advertising alone.[2]

Executives at social network companies utilize a variety of content intervention, friend recommendation, and seeding strategies to grow network connectedness and increase user engagement within the network.[3] For example, Facebook suggests content and potential friends to users under "Pages You May Like," "Suggested For You" posts in news feeds, "People You May Know," or "Groups You Should Join;" Instagram's suggestions can be found in "Instagram Explore," "Accounts You May Like," and "IGTV Discover." According to both companies' websites, these personalized recommendations are generated by proprietary algorithms based on content users have expressed interest in the past and actions users have taken on the social media platform.[4]

Despite wide industry practice, due to the proprietary nature of the recommendation algorithms, little is known about their effectiveness and academics have resorted to using simulations (based on model estimates) to evaluate the effectiveness of these strategies. Previous literature has primarily studied these strategies in the empirical context of mature (static) social networks where the number of friendship ties remains stable (e.g., Trusov, Bodapati, and Bucklin 2010; Hinz et al. 2011; Aral, Muchnik, and Sundararajan 2013; Lanz et al. 2019). Their findings may or may not hold for evolving networks in which an intervention is likely to trigger the formation of new friendships and therefore cause a cascading effect on other ac-

[1] https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/; https://www.statista.com/statistics/788084/number-of-social-media-accounts/

[2] https://www.iab.com/news/iab-internet-advertising-revenue/

[3] 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.

[4] More details can be found at https://www.facebook.com/help/1257205004624246 and https://help.instagram.com/313829416281232?fbclid=IwAR2YAPkXuscmrWPT2a5fnYY6bUXHEYLTYBtUIL11axt3LXLlPhBlNDdBAwg.

tivities due to the increased network connectedness. This issue becomes especially concerning in fast growing networks where users rapidly expand their friend connections. For example, Instagram users saw, on average, a growth of $13 - 16\%$ in the number of their followers during the first half of 2019.[5] As a result, to accurately evaluate the effectiveness of recommendation and seeding strategies in evolving networks, one needs to take the endogenous formation of new friendship ties into account, i.e., friendship ties are not made randomly between pairs of users, but are rather the result of users' deliberate decisions considering potential benefits and cost of each new friendship.

To fill this gap, we develop a framework for the co-evolution of individual users' friendship tie formations and their concurrent activities within an online social network and explicitly model the endogenous formation of new friendship ties. An intriguing aspect of making friends in many online social networks (including the one we study in this paper) is that people often do not know each other's identities in real life and their interactions are mostly confined to the online network, which is vastly different from the typical friendmaking scenarios in the real life. Thus one might ask how strangers become friends in these online networks characterized by anonymity in the first place.[6] While extant research in the strategic network formation literature including Christakis et al. 2010, Snijders, Koskinen, and Schweinberger 2010 tend to explain the formation of friendship ties in real life using only static individual characteristics such as age or gender, it is very likely that an individual's behaviors and opinions (such as product adoptions and content generation) as observed by other people in the online environment are also significant drivers of her friendship tie formation decision. By taking these factors into account when modeling tie formations in an online social network, our paper complements and enriches the strategic network formation literature.

We obtain our data from a special interest online community for animes (Japanese cartoons) called MyAnimeList.net. This website provides a gathering place for anime fans from all

[5] https://www.statista.com/statistics/307026/growth-of-instagram-usage-worldwide/ (users with more than 1K followers).

[6] We define "online networks characterized by anonymity" as online networks in which users are not required to provide personally identifying information to register an account and/or to participate in activities; the creation of a username and password suffices. Other examples of such online social networks are goodreads.com, boardgamegeek.com, dpadd.com, and last.fm. Even without personally identifying information, individuals may gain familiarity with each other and establish meaningful relationships in these networks over time.

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 becoming friends online and that the actions they observe on the website are the main drivers of their friendship decisions — making this platform an ideal environment for our research inquiry. We take advantage of this novel data set that documents both the continuous development of the network, i.e., which individuals become friends with each other and when that happens, and all users' entire activity (i.e., anime watching and posting of user-generated content (UGC)) histories on the platform. Access to these data enables us to overcome several modeling constraints present in recent network co-evolution models (e.g., Snijders, Steglich, and Schweinberger 2007) and in the strategic network formation literature (e.g., Christakis et al. 2010).[7] Most importantly, it allows us to model the friendship network development without the need to simulate the state of the network at each point in time and, as a result, to quantify the effects of users' time-varying activities on the probability that two individuals become friends. We model the endogenous formation of a social network and the occurrence of two types of online activities, namely, product adoptions and content generation, over time. More specifically, using a utility-maximizing framework, we model an individual user's daily decisions in three areas: (i) *with whom* to become friends, (ii) whether to watch any anime, and (iii) whether to publish a UGC post. We model friendship tie formation between two individuals as non-cooperative decisions. Drawing from the network formation and social psychology literature, we consider three types of drivers that impact individuals' friendship formation utility: characteristics of the potential friend's current network state ("network properties" for short in the following), similarity (homophily) between the pair, and the domain expertise of the potential friend. Each individual maximizes her own friendship formation utility and a friendship is formed if both users agree to it. A user's utilities of engaging in either product adoptions or content generation are functions of her past online activities and her friends' ac-

[7] We provide a detailed discussion of our versus prior modeling approaches in the Relevant Literature section.

tivities which can affect her actions through a direct peer effect (on the same type of activity) and a spillover effect (on the other type of activity). Finally, the three utility functions are connected in two ways: through observed variables and through correlated error terms.

Our results for friendship tie formation reveal that all three drivers, i.e., network properties, similarity, and expertise, affect a user's friendship formation decisions. A focal user is more likely to become friends with similar users, i.e., users who watch the same animes and are similar in terms of demographic traits. Although this has been shown in previous research studying more traditional friendship settings such as a friendship network among students, our finding reveals that even in (anonymous) online networks, users who share similar (though unverifiable) demographics are more likely to become friends. Among the three demographics we study, age is the most important driver with the largest marginal effect. We also find significant positive effects of both network property variables: a focal user is more likely to form a friendship tie with users who are popular and have many friends in common with the focal user. This result stands in contrast to Christakis et al. (2010) who find that students are less likely to become friends with popular students. We suspect that the disparate finding is due to our unique empirical context of an (anonymous) online social network in which popular users are more likely to be "known" by other users, and due to the lower cost of initiating and maintaining many connections in such an environment. Furthermore, we find a significant effect for domain expertise, i.e., a focal user is more likely to become friends with a user who publishes many posts. This finding is consistent with the notion that information sharing is the underlying mechanism that drives the expertise effect. Sharing comments and providing feedback about animes through UGC publication is the key to attract friendships using one's domain expertise.

Our results for in-site activities, i.e., product adoptions and UGC production, reveal significant positive peer effects on the focal user: while having more friends does not necessarily make a user more active, having more active friends does increase a user's activity level. Comparing the magnitude of the direct peer effect versus the spill-over effect, we find that the direct peer effect is the larger driver of a user's in-site activity levels.

Using predictive exercises, we uncover several novel findings regarding content/activity intervention, friend recommendation, and seeding strategies for evolving social networks. First, the winning friend recommendation strategy is context-specific. In other words, it depends on the objectives: while recommending a friend of a friend as a potential friend is the most effective strategy in stimulating friendship tie formation in an evolving network, recommending an active UGC creator as a potential friend works best in increasing anime watching, UGC publishing, and the overall level of in-site activities via friend recommendations. Second, seeding to users who make many UGC posts is, on average, more effective in increasing online activity levels in an evolving network than seeding to well-connected users. This result stands in contrast to the common finding for static networks that indicates that seeding to well-connected users is the most effective strategy in such an environment (see, e.g., Trusov, Bodapati, and Bucklin 2010; Hinz et al. 2011; Aral, Muchnik, and Sundararajan 2013). And lastly, shutting down the endogenous network formation in an evolving network when assessing the effectiveness of seeding strategies leads, on average, to an underestimation of seeding effectiveness by 41%.

The contribution of this paper is two-fold. First, our paper is the first systematic investigation in marketing that theorizes and quantifies the importance of various drivers behind friendship formation decisions among strangers in an online environment. Our finding that all three friendship drivers, i.e., network properties, similarity, and expertise, matter provides strong support for the popular practice of recommending people with common friends, similar traits, and/or domain expertise as potential connections by large social networks including Facebook and LinkedIn (see also Sun and Taylor 2020). In particular, by quantifying the effects of individuals' time-varying activities, such as UGC production, on friendship formations through expertise variables, our paper enriches the network co-evolution and strategic network formation literature. This richer specification is much needed when describing the network development in an online environment where many people encounter each other through the Internet and become friends even though they might never meet in real life.

Second, to the best of our knowledge, this paper is among the first ones to assess the effectiveness of various content intervention, friend recommendation, and seeding strategies in an evolving online social network. By modeling the interdependent dynamics between network formation and time-varying online activities, i.e., how tie formations, product adoptions, and UGC productions influence each other, we are able to account for the endogenous formation of friendship ties and, more importantly, the cascading effects of increased network connectedness on other in-site activities. We show that not accounting for the endogenous network formation leads to a severe underestimation of seeding effectiveness. Furthermore, we demonstrate that seeding to most connected users, the most effective seeding strategy in static networks, is no longer the winning strategy in evolving networks. Therefore, it is crucial for managers to consider the interdependence of network formation and in-site activities when devising optimal stimulation and seeding strategies for an evolving network.

The remainder of this paper is organized as follows: Next, we discuss the relevant literature and introduce our data. Then, we introduce our model, estimation approach, and identification strategy. Subsequently, we present our estimation approachand discuss the empirical results. Finally, we perform prediction exercises and conclude with a discussion of limitations and suggestions for future research.

RELEVANT LITERATURE

In this section, we review four relevant streams of literature on network co-evolution, drivers of friendship formation, the identification of peer effects, and seeding strategies in social networks and delineate our research vis-à-vis the findings from extant research.

Network Co-evolution

Recent literature on network co-evolution typically adopts a continuous-time framework in which network users decide whether to alter their network ties or to perform other actions at random instants in time (Snijders, Steglich, and Schweinberger 2007, Steglich, Snijders, and Pearson 2010, Lewis, Gonzalez, and Kaufman 2012, Greenan 2015). For example, Snijders, Steglich, and Schweinberger (2007) develop a continuous time first-order Markov model that describes both the formation of a network from an individual's perspective and the incidences of individuals' other action(s) within the network. In the model, both the network structure and individuals' actions evolve in a dynamic fashion: individuals are selected at random rates and each selected individual decides whether to make a change in her friendship ties, to perform an action, or to do neither. The randomness in the timing of friendship and activity decisions is necessary in this class of models due to the data limitation of only observing the network at discrete moments (i.e., a few snapshots of the network). As a result, the timing and sequence of tie formations and other actions of users that take place between two observed discrete moments are unavailable to researchers. In an empirical context more closely related to ours, Bhattacharya et al. 2019 extend Snijders, Steglich, and Schweinberger (2007) and apply the model to study the co-evolution of users' social network structure and content posting behaviors using monthly data from a major social networking site.

There are several limitations to the modeling approach introduced in Snijders, Steglich, and Schweinberger (2007) and adopted in a number of later studies (e.g., Steglich, Snijders, and Pearson 2010, Lewis, Gonzalez, and Kaufman 2012, Greenan 2015, Bhattacharya et al. 2019). First, since individuals cannot both change their ties and perform other actions at the same point in time, simultaneous incidences of tie formation and other actions cannot be accommodated in this framework. Second, due to the randomness in the decision timing within a time period (between two sequential snapshots of the network observed by researchers), the effects of time-varying behaviors that happen continuously within the time period cannot be identified and, as a result, their effects cannot be properly assessed. Third, although Snijders, Steglich, and Schweinberger (2007) capture homophily by accounting for observed similarities among users when modeling tie formations, latent homophily (arising from the similarity among friends in their unobserved intrinsic preferences) remains a confounding factor that may bias the effect of friends' influence.

In this paper, we overcome these three limitations by proposing a co-evolution model of individuals' concurrent decisions to both form friendship ties and to perform online activities (such as product adoptions and content generation) at each point in time. We are able to do so because we observe the continuous development of tie formations and all concurrent online activities in our novel data. In other words, our data records the occurrence and timing of all tie formations and other online activities performed within the network. This allows us to capture the effects of any time-varying behaviors while studying simultaneous incidences of these decisions. Furthermore, we are able to account for the latent homophily by explicitly estimating individuals' unobserved intrinsic preferences for actions absent of their friends' influence and therefore provide a cleaner identification of peer effects. We are able to do so because we observe users' actions both before and after they make friends in our data.

Friendship Formation

More broadly, this paper is related to the network formation literature. Researchers have studied network formation using three main modeling approaches: nodal attribute models, exponential graph models, and strategic network formation models.[8] The first category of models explains the existence of ties and the resulting network structure via similarities among pairs of individuals (e.g., Hoff, Raftery, and Handcock 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. Exponential graph models explain the network development based on structural patterns such as triangular connections or transitivity, but do not provide insights into the mechanisms that drive individuals' tie formation decisions (e.g., Katona and Móri 2006; Mele 2017). These models are well-suited for predictions, but not for causal inferences and therefore not suited for counterfactual analysis. To overcome these shortcomings, strategic network formation models, the most recently developed modeling approach among the three, have taken the perspective of an individual actor's utility maximization when explaining the evolution of a network, and allow it to depend on the existing state of the network (e.g,.Hanaki et al. 2007).[9] This paper falls into the last category of modeling approaches. It is important to note that unlike network co-evolution models discussed in the previous section, strategic network formation models only describe friendship tie formations and *not* incidences of any other activities.

[8] We refer the interested reader to Jackson (2008) and Toivonen et al. (2009) for a detailed comparison of the three categories of models.

[9] 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.

Two notable papers in this stream of literature are Christakis et al. (2010) and Snijders, Koskinen, and Schweinberger (2010). In both papers, the authors model future states of a network based on characteristics of the existing network state (current friendship ties) and similarity in terms of demographic traits. However, the models proposed in these two papers are not able to capture the effects of users' changing behaviors and activities on tie formation outcomes due to the simulated network states between (a limited number of) observed snapshots of the network. Christakis et al. (2010) apply their model to study a network of 669 student 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 of the same gender and age. Snijders, Koskinen, and Schweinberger (2010) use a data set that contains six snapshots of a network of 32 students from the same class. They find that network characteristics such as transitivity matter in friendship formation. In terms of demographic characteristics, they find that males are more popular as friends, but having the same gender does not increase one's chance of making a friend.

Studies in the social psychology literature have found that similarity (homophily) is an important driver in individuals' friending decisions in social networks. In this literature, similarity is defined more broadly as a match between two persons' interests, experiences, backgrounds or personalities, and not limited to similarity in demographic traits. Similarity between two individuals increases the chance of them becoming friends (e.g., Berscheid and Reis 1998, McPherson, Smith-Lovin, and Cook 2001, Sun and Taylor 2020). To put it differently, "birds of a feather flock together." In addition, the expertise or knowledge of a person in the subject domain increases her desirability as a potential friend. The increased desirability is due to the potential gain from information sharing and learning from a friend who is knowledgable about the subject domain (Watson and Johnson 1972; Brandtzæg and Heim 2009).

Building on the extant research in the strategic network formation and social psychology literature, we identify three main drivers of users' friendship decisions in an online social network: (1) characteristics of the current network state or "network properties" such as out-degrees (i.e., number of friendship ties) and transitivity (i.e., number of common ties); (2) homophily between two users including similarity in their demographic traits as well as similarity in their interests such as animes both users have watched; (3) expertise of a user, which is reflected in the current level of activities and captures her domain expertise such as the number of anime related posts she published. While the first two types of drivers have been studied in previous research, to the best of our knowledge, no studies in the extant strategic network formation literature are able to quantify the impact of the last driver on network formation. This constraint is largely due to the data limitation that only one or a few snapshots of the network are available to researchers; therefore it is almost impossible to capture the effects of time-varying expertise variables.

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 and more importantly, we observe each individual's entire activity history, enabling us to explicitly account for the effects of individuals' time-varying behaviors in their friendship formation decisions. The latter improvement is especially important since we study network formation in an online environment. In such an environment, individual users' behaviors and opinions as observed by other individuals are likely to be among the most important drivers of their friending decisions.

Peer Effects

This paper is also related to the literature on the identification of peer effects. Making a casual inference of friends' influence is a challenging task (Manski 1993). Multiple social phenomena can confound the identification of social influence (see Hartmann et al. 2008; Shalizi and Thomas 2011, and discussion in the Identification section). Among these phenomena, homophily is probably the most challenging one to be accounted for. Due to similarity, friends may exhibit the same behavior without one necessarily influencing the other. Different approaches have been proposed to account for homophily including the use of instrumental variables (e.g., Bramoullé, Djebbari, and Fortin 2009; De Giorgi, Pellizzari, and Redaelli 2010; Claussen, Engelstätter, and Ward 2014), the inclusion of correlated group effects (e.g., Lee 2007; Lee, Liu, and Lin 2010; Ma, Krishnan, and Montgomery 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), the incorporation of individual-specific unobserved tastes/preferences (e.g., Nair, Manchanda, and Bhatia 2010; Trusov, Bodapati, and Bucklin 2010; Ameri, Honka, and Xie 2019), modeling the co-incidence of network and outcome formation (Snijders, Steglich, and Schweinberger 2007; Badev 2021), and explicitly accounting for the selection bias that arises from endogenous network formation (Hsieh and Lee 2016). Our approach is similar to Hsieh and Lee (2016) in which the authors simultaneously model the network formation and social interaction processes. By allowing the unobserved components in both processes to be correlated, they explicitly account for the selection bias that arises from endogenous network formation, and thereby correct biases on estimated peer effects in the social interaction model. In our paper, we take a similar approach by allowing the stochastic error term in our network formation equation to be correlated with the error terms in the other two online activity equations in which peer effects are estimated. However, the employed models differ. In particular, Hsieh and Lee (2016) model uni-directional tie formation which results in a simpler model than ours since mutual agreement is not necessary. Furthermore, in contrast to Hsieh and Lee (2016), we are able to account for latent homophily by explicitly estimating individuals' unobserved intrinsic preferences for actions absent of their friends' influence. Therefore, we provide a cleaner identification of peer effects. We can do so because we observe users' actions both before and after they make friends.

Seeding

And lastly, this paper is also related to the literature on 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"[10] or "betweenness centrality."[11] For example, Hinz et al. (2011) compare three seeding strategies — stimulating

[10]In-degree and out-degree centralities refer to the number of incoming and outgoing ties, respectively, of an individual.

[11]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.

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, Muchnik, and Sundararajan (2013) 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. 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. In a more recent empirical study of a music network, Lanz et al. (2019) also find that the majority of music creators does not benefit from seeding to influencers.

Many of the previous studies have focused on (one-time) 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, Bodapati, and Bucklin (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. They find that, on average, only 20% of a focal user's friends influence the focal user and suggest these friends for seeding.

Similar to Trusov, Bodapati, and Bucklin (2010), we add to the existing knowledge on the effectiveness of seeding strategies in increasing user engagement on online social platforms. Going beyond Trusov, Bodapati, and Bucklin (2010), we examine different seeding strategies that are not only based on network metrics, as is common in this literature, but can also depend on users' activity levels 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 development of future friendship ties and into future actions of users.

DATA

Our data come from MyAnimeList.net. This website is a consumption-related online community where online interactions are based upon shared enthusiasm for a specific consumption activity (Kozinets 1999). MyAnimeList.net was created to allow anime 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).[12] Users can write reviews of and recommendations for animes. The website also hosts a discussion forum where users can share information and exchange opinions about animes. We classify any post made by a user in the reviews, recommendations or discussion forum sections of the website as UGC. In addition, users have the option to become friends, which makes it easier for them to access their friends' pages and to be notified about their friends' activities, similar to bookmarking and RSS functions in web browsers.[13] Note that the platform does not have a friend recommendation system.

Anime fandom is a special interest and not very common. As a result, fans use special interest communities such as MyAnimeList.net to find and connect with other fans. This implies that most users of MyAnimeList.net meet other users for the first time on the website and their interactions happen within the website. Furthermore, this website is a worldwide community and attracts users from many different 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.[14] This observation further validates our assumption that meetings and interactions among users are mostly confined to the platform.

Estimation Sample

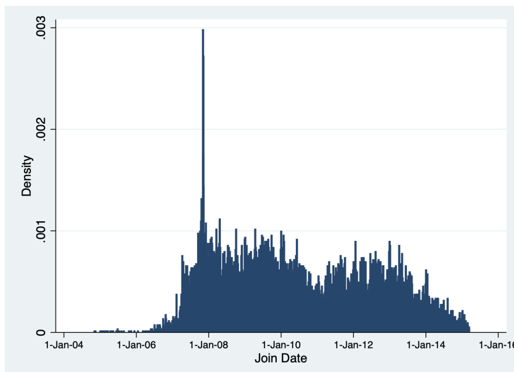
[12]Our anime watching data are self-reported. Thus accuracy in the reporting of watched animes is a potential concern. Note that, in contrast to incentivized surveys, there are no incentives for users on MyAnimeList.net to falsely report their true anime watching behavior. Furthermore, in the similar setting of TV shows, Lovett and Staelin (2016) compare survey panelists' self-reported viewing data and the actual streaming data and find that people tend to correctly report their actual watching behavior. However, even without explicit incentives, users may misrepresent their watching behavior due to social desirability concerns or the need to appear knowledgeable on the platform. This represents a limitation of our data.

[13]We do not observe private communication between users. Two types of private communication are available on the platform: writing private comments on a user's wall and private messaging. This represents a limitation of our data.

[14]See Table A-1 in Web Appendix A for an overview of countries of origin for users.

The website was first started in 2004, however, as a private domain. In April 2006, it was moved to a public domain and began to take its current shape. At that point in time, the website had about 300 users. After its transfer to a public domain, the number of members started to grow quickly (see Figure 1). About a year later, on March 16, 2007, the function of forming friendships was added. At that point in time, the website had about 450 members and this number grew rapidly to 2,700 at the beginning of July 2007 and to about 10,000 by the end of 2007.

Figure 1: Dates Users Joined MyAnimeList.Net

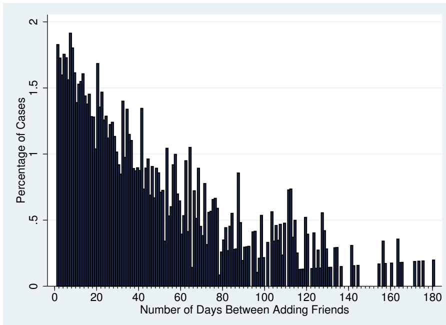


We focus on users who joined the website in the second half of 2007, mainly for two reasons. First, users who joined the website before March 2007 are likely to add each other as friends based on past interactions. To put it differently, had they had the option of adding friends before, they would have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

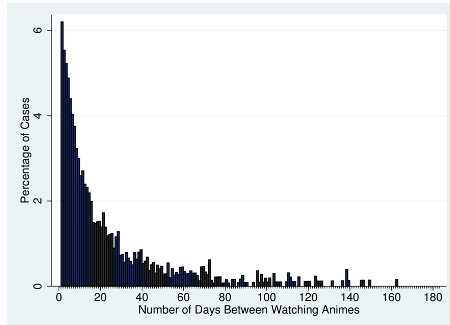
Studying daily friendship formation among *all* users who joined between July and December 2007 is, however, computationally impossible since the data set would include over 7 billion pair-day observations. One potential solution is to shorten the observation period. However, this approach would result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

Figure 2: Number of Days Between Activities

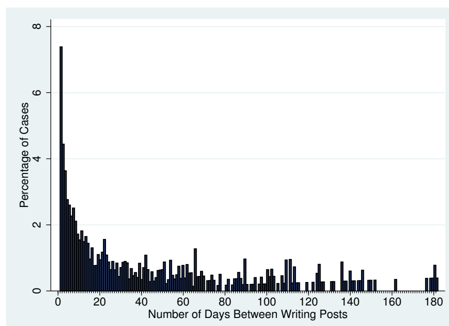
(a) Friend Addition



(b) Anime Watching



(c) Content Generation



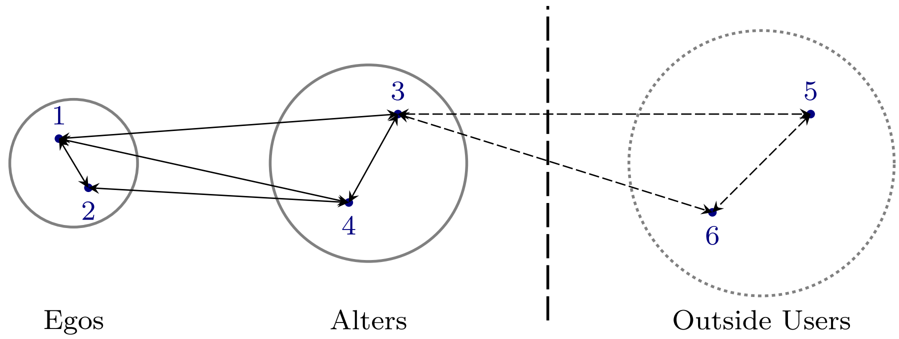
Our solution is to sample from the network.[15] We implement this solution via egocentric sampling:[16] we first draw a random sample of 400 users ("egos") out of 7,419 users who joined the website in the second half of 2007. We then include all of the egos' friends in our estimation sample. We term the users who are *not* egos themselves, but friends with an ego "alters" (986 users). Thus, our estimation sample contains 1,386 users (400 egos and 986 alters). Note that egos' friends can either be egos or alters and that *all* friends of all egos are included in the estimation sample. Figure 3 visualizes our sampling strategy. As an example, user 2, who is an ego, has three friends: he is friends with user 1, who is another ego, and with users 3

[15]A second potential solution would be to aggregate the observations to the weekly level. However, aggregation leads to information loss in both the dependent and independent variables. On top of that, we observe that users make more than one friend in a week in more than 20% of the cases. Anime watching and content generation also happen more than once a week in about 25% and 10% of the cases, respectively. Consequently, aggregating data to the weekly level would force us to model the sequence of users' actions within a week. As a result, similar to previous research (e.g., Snijders, Steglich, and Schweinberger 2007), the degree to which we could capture the effects of time-varying activities on friendship formation would be restricted.

[16]See Perry, Pescosolido, and Borgatti (2018) for a methodology overview and Trusov, Bodapati, and Bucklin (2010), Gargiulo and Benassi (2000), Banerjee et al. (2013), and Jackson, Rodriguez-Barraquer, and Tan (2012) for applications in marketing, organizational science, developmental economics/finance, and economics, respectively.

and 4, who are alters.

Figure 3: User Sampling Strategy



In the estimation, we model all anime adoptions and all UGC production activities for both egos and alters. For friendship formation, we model all potential ties among egos (e.g., between users 1 and 2 in Figure 3), all potential ties among egos and alters (e.g., between users 1 and 3 in Figure 3), and all potential ties among alters (e.g., between users 3 and 4 in Figure 3). We do not model the friendship formation decisions between alters and users outside our estimation sample, but we do take their friendships into account when creating related explanatory variables.

Given that activities of users in the three areas can be correlated, missing a portion of the network formation for the 986 alters could lead to bias in our estimates related to friendship formation decisions. To alleviate this concern, we estimate separate coefficients for the 400 egos (for whom we model all their tie formations) and for the 986 alters (for whom we do not model the portion of the friendship network that includes users outside of our sample) in all three utility functions, i.e., friendship formation, anime adoption, and UGC production. We will only use the coefficient estimates for egos to make inferences.[17]

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

[17] Snowball sampling is a common sampling method that has been used extensively in the network literature. However, a concern with this sampling method is the oversampling of active users. In our case, this concern is alleviated by drawing a random sample of 400 egos and controlling for unobserved heterogeneity among these users. For a focal ego, the other 399 randomly drawn egos and their friends who are not friends with that focal ego are a random representative sample of the whole network. Furthermore, we estimate separate coefficients for egos and alters in all three utility functions. As a result of this randomness and the inclusion of separate parameters for egos and alters, we believe that our estimates are unbiased for the 400 egos.

Representativeness

To evaluate whether our random sample of 400 egos is representative of the entire network of 7,419 users who joined the community in the second half of 2007, we calculate summary statistics of four network structure measures (degree centrality, betweenness centrality, eigenvalue centrality, and transitivity/clustering coefficient) for both groups. We then test whether the average network structure measures for the 400 egos and the 7,419 users are significantly different. The summary statistics and t-values of the mean comparisons are presented in Table 1. We find none of the differences to be significant at $p < 0.05$.[18]

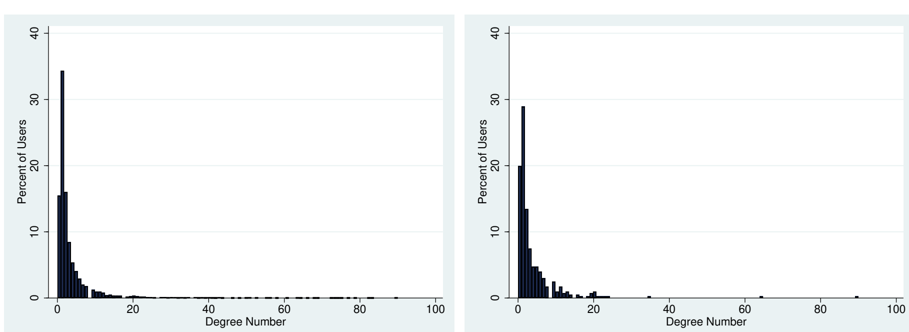
Table 1: Network Representativeness

	(i) Complete Network	(ii) Estimation Sample	(iii) t-value
Average Degree	4.37	3.81	1.12
Average Betweenness Centrality	8,986.13	5,234.28	1.50
Average Eigen Centrality	.0073	.0068	.94
Transitivity	.1986	.1779	1.32

To further validate the representativeness of our estimation sample, we examine the degree distributions among the 400 egos and compare it with the degree distribution in the entire network of 7,419 users who joined the community in the second half of 2007. The distribution graphs are shown in Figure 4. The degree distributions for both groups follow a similar power of law distribution with the majority of users having fewer than 20 friends.

Figure 4: Degree Distributions

(a) Complete Network **(b) Estimation Sample**



[18]Note that the sampled group of 400 egos can form ties with the 7,419 users and both groups can form ties with users who joined before July 2007. As a result, the networks of 400 and 7,419 individuals cannot be fully distinguished to calculate measures such as network diameter.

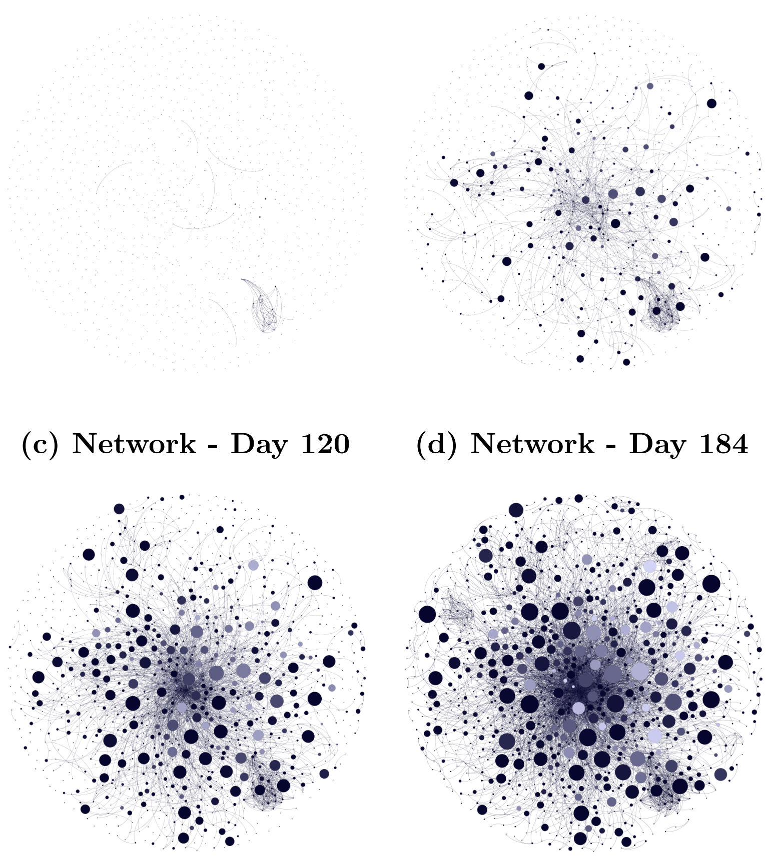
Data Description

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 5 shows the states of the network for snapshots taken 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 and the size of the nodes increases as users watch more animes. As expected, the nodes become darker, bigger, and more connected over time.

Figure 5: Network Co-Evolution Over Time

(a) Network - Day 1 **(b) Network - Day 60**

(c) Network - Day 120 **(d) Network - Day 184**



Lines Between Nodes Indicate Friendship Ties. Node Size Increases with More Animes Watched. Node Color Darkens with More Posts Written.

Table 2 summarizes key statistics of our data. In terms of demographics, 78% of users report their age and are, on average, 19 years old. 93% of users report their gender with 39% being female and 54% being male.

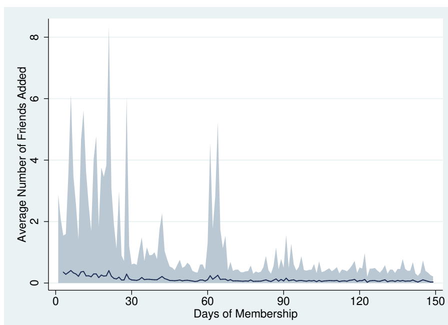
Table 2: Descriptive Statistics

	Mean	Std. Dev.	Min	1st Quartile	Median	3rd Quartile	Max	N
Age	19.3	5.2	12	17	18.5	23	78	1,088
Gender (% Females)	38.5							
Gender (% Males)	54.0							
Gender (% Not Specified)	7.5							
Egos:								
Number of Active Days	12.62	15.43	0	3	7	17	101	400
Number of Friend Adding Days	2.55	4.69	0	0	1	3	59	400
Number of Anime Watching Days	9.19	10.37	0	2	5	13	68	400
Number of Content Generating Days	2.47	9.17	0	0	0	1	95	400
Percentage of Active Days	17.76	17.07	0	5.04	12.90	24.24	100	400
Percentage of Friend Adding Days	3.80	6.19	0	0	1.63	4.60	40	400
Percentage of Anime Watching Days	13.55	14.11	0	3.23	9.70	19.44	85.71	400
Percentage of Content Generating Days	2.56	7.72	0	0	0	0.82	61.69	400
Friend Adding Interval in Days	44.26	36.05	1	15	34	64	156	12,414
Anime Adding Interval in Days	25.64	28.20	1	5	14	37	138	20,680
Post Adding Interval in Days	32.07	29.33	1	7	25	48	109	6,330
Alters:								
Number of Active Days	24.06	24.42	1	7	16	34	181	986
Number of Friend Adding Days	5.93	6.40	1	2	4	7	48	986
Number of Anime Watching Days	15.55	16.65	0	3	10	22	121	986
Number of Content Generating Days	6.31	17.57	0	0	0	3	181	986
Percentage of Friend Adding Days	22.49	18.36	.54	9.26	17.95	30.43	100	986
Percentage of Anime Watching Days	6.85	8.31	.54	2.17	4.23	8.15	80	986
Percentage of Content Generating Days	14.46	13.38	0	4.35	10.87	21.20	82.22	986
Percentage of Content Generating Days	4.72	12.01	0	0	0	2.78	98.37	986
Friend Adding Interval in Days	48.41	40.36	1	16	37	71	181	78,996
Anime Adding Interval in Days	24.82	29.86	1	5	13	32	163	84,133
Post Adding Interval in Days	47.98	48.96	1	7	28	78	182	40,287
Egos with Friends:								
Number of Friends on Last Day	4.72	7.83	1	1	2	6	94	320
Number of Animes on Last Day	76.49	83.02	0	20	50.5	100.5	522	320
Number of Posts on Last Day	6.13	22.50	0	0	0	2	228	320
Egos without Friends:								
Number of Friends on Last Day	0							
Number of Animes on Last Day	46.34	57.22	0	8	24	70.5	313	80
Number of Posts on Last Day	.68	3.19	0	0	0	0	23	80

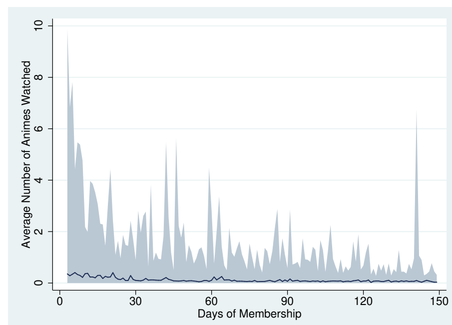
Figure 6 demonstrates how average activity levels of users who joined the website in the second half of 2007 change over time from the day they joined the website. Figure 6a shows a decreasing trend in making new friendship ties. Since one of the benefits of having friends are reduced costs associated with learning about the website and new animes, users are more likely to add friends shortly after joining the website. Motivated by this data pattern, we assume that users' intrinsic propensity to make friends contains two components: a constant intrinsic propensity component and a second component that decreases with membership length. Figures 6b and 6c show the activity trends for anime watching and content generation. Both graphs reveal a rather constant trend over time.[19] These two data patterns prompt us to assume that the intrinsic propensities in anime watching and content generation are constant over time.

Figure 6: Average Activity Levels Over Time Since Joining (New Users)

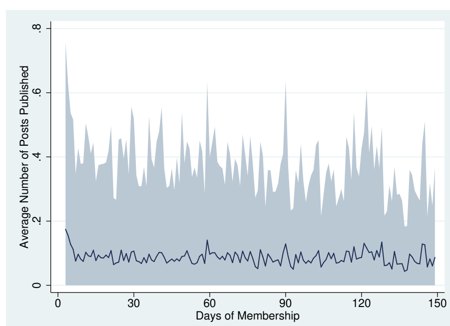
(a) Friend Addition



(b) Anime Watching



(c) Content Generation



Grey Areas Show +/- One Standard Deviation Bounded at 0 from Below.

Users can engage in multiple types of activities simultaneously. On average, egos have 2.6 active days in terms of friend adding, 9.2 active days in terms of anime watching, and 2.5 active days in terms of post writing (see Table 2). In total, they have 12.6 days in which they participate in at least one type of activity. To put it differently, on average, egos are active

[19] 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.

on about 18% of the days during the study period. We observe a similar pattern for alters albeit with higher average levels of all three activities. As shown on the bottom of Table 2, 20% of egos have no friends at the end of our study period. Egos with friends are more active in both anime watching and UGC production than egos without friends, suggesting the presence of peer influence from friends. Figure 7 shows a Venn diagram of the joint probabilities of each type of activity conditional on engaging in at least one type of activity. Users are active in only one area in 84.16% of the cases. Users are active in two and three of the areas of interest in 14.79% and 1.05% of cases, respectively. Figure 8 plots the percentages of egos engaging in 0, 1, 2 or 3 activities every calendar day and reveals constant levels during our study period.[20]

Figure 7: Percentage of Observations with Certain Activities Conditional on Performing at Least One Activity

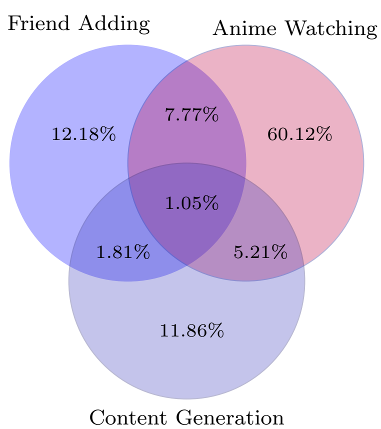
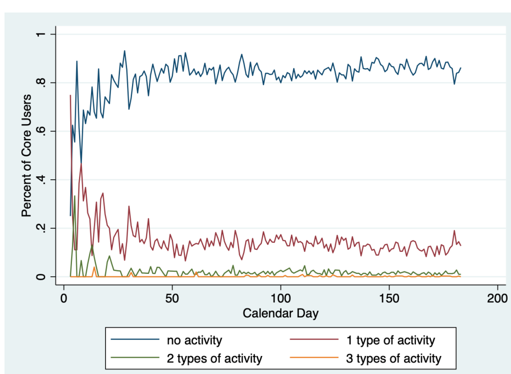


Figure 8: Percentages of Egos Engaging in 0, 1, 2 or 3 Activities Per Day

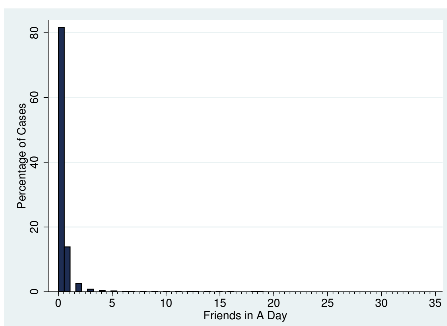


[20]The distribution of all permutations of activity types that can co-occur in a day exhibits a similar pattern and is available from the authors upon request.

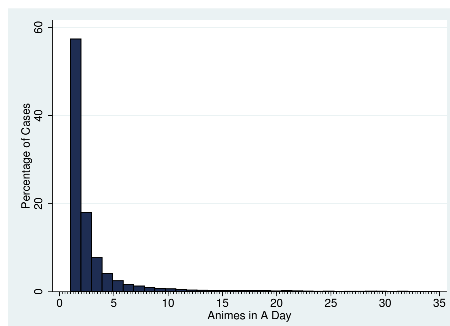
Lastly, Figure 9 shows histograms of individual users' daily activity intensities conditional on them being active. In more than 80% of the cases, users add only one friend on an active day. Similarly, in about 60% of the cases, users watch only one anime per active day. However, the content generation intensity is higher: users publish one post per active day in about half of the cases and publish 2 or 3 posts per active day in about 20% and 10% of the cases, respectively. Based on these data patterns, we make the simplifying assumption to model anime watching and content generation as binary indicator variables, i.e., we model whether a user watches an anime or publishes a post, but not the number of animes watched or posts published by a user, in a given day.[21]

Figure 9: Number of Activities in Each Area Per Day

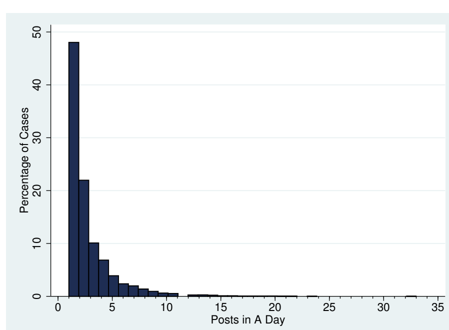
(a) Number of Friends Added in a Day (Truncated at 100)



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



(c) Number of Posts Written in a Day



MODEL

[21] 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.

In this section, we describe how we jointly model a user's decisions to form friendship ties, adopt animes, and generate content.

Tie Formation

We start by describing how we model tie formations among users over time. In each time period (day), a user makes decisions whether to become friends with any other user with whom she is not friends yet.[22] Since there are usually many users with whom the individual is not friends yet, at each point in time, a user can become friends with multiple users. Note that we model a user's tie formation decisions for each possible friendship pair in each time period and not whether a user makes a friend in a time period.

Suppose the website has $i = 1, \ldots, N$ users and these users can become friends with other users during $t = 1, \ldots, T$ time periods.[23] Let M denote the adjacency matrix of the network which shows the stati of ties between each pair of individuals i and j with $j = 1, \ldots, N$ and $i \neq j$. m_{ijt} equals 1 if i and j are friends at time t and 0 otherwise. Ties are bi-directional and symmetric, i.e., $m_{ijt} = m_{jit}$. Furthermore, both users have to agree to become friends. In our data, we do not observe users' requests for friendship with other users, only the formation of ties upon mutual agreement. Therefore our model describes the decision of both users to become friends regardless of who first initiated the friendship (see, e.g., Christakis et al. 2010).[24] In our estimation sample, $N = 1,386$ (400 egos and 986 alters), $T = 184$, and $M = 1,386 \times 1,386$. Note that the elements of M change over time as ties are formed.

The decision of two users to become friends depends on the utilities both individuals derive from becoming friends (see, e.g., Christakis et al. 2010). User i's utility of becoming friends with individual j in time period t, U^m_{ijt},[25] is given by

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

[22] A user's set of potential friends changes over time for two reasons: (i) with the focal user making more friends, the number of existing users with whom she is not friends yet decreases; and (ii) with new users joining the platform on a daily basis, her set of potential friends increases. Both events happen simultaneously and on a daily basis so the focal user's set of potential friends changes over time: it can both increase or decrease every day.

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

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

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

where $\mathcal{N}^m$ describes the characteristics of the current network state of user j, $\mathcal{R}^m$ captures the similarity (homophily) between user i and user j, $\mathcal{K}^m$ describes the expertise of user j, and C^m is a set of control variables. ϵ^m captures the part of the utility of user i at time t that is observed by the user but not by the researcher.

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

We model the tie formation between two users as a non-cooperative decision (e.g., Christakis et al. 2010, Boucher 2016), i.e., each pair's decision to become friends only depends on their own friendship utilities and is conditionally independent from friendship decisions of other pairs of users. A tie between user i and user j is formed if and only if both parties decide that it is beneficial for them to become friends, i.e.,

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

Recall that the platform does not have a friend recommendation system. A focal user's consideration set of potential friends consists of non-friend users the focal user naturally encounters on the platform. We do not explicitly model individuals' consideration sets of potential friends. This represents a limitation of our model. The random sampling is a means to approximate the consideration set in a scalable way and we control for individuals' visibility on the platform through observables in the friendship formation utility (see also Footnote 29).

Product Adoption and Content Generation

Next, we describe how we model a user's activities on the website. We study the incidence of users' activities in two broad areas, namely, product adoption and content generation. Let

U_{it}^{pa} denote user i's utility from watching an anime at time t and U_{it}^{cg} denote user i's utility from producing content at time t. Then U_{it}^{pa} and U_{it}^{cg} are given by

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

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

Integrating All Actions

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

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

Some variables, unobserved by the researcher, might influence more than one type of decision an individual user makes.[26] For example, a user might be traveling and, as a result, not spending any time on the website, i.e., be inactive in all three areas. To accommodate the simultaneous co-occurrence of activities user i makes at time t, we allow the three error terms in Equation (4) to be correlated, i.e.,

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

Utility Specifications

In this section, we present detailed utility specifications for the specific context of our data.

[26] 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.

We model the utility user i receives from forming a tie with user j as

$$U_{ijt}^m = \tilde{\alpha}_{it}^m + \delta^m \mathcal{N}_{j,t-1}^m + \mathcal{R}_{ij,t-1}^m + \gamma^m \mathcal{K}_{j,t-1}^m + \lambda^m C_{ijt}^m + \epsilon_{it}^m. \tag{6}$$

$\tilde{\alpha}_{it}^m$ captures user i's intrinsic preference for making friends at time t, i.e., the net of user i's benefit and cost of making friends at that point in time. As revealed in Figure 6, users newly joining the website are more likely to add friends compared to users who have already been members of the website for a longer time. This is likely due to having friends in the beginning reducing learning costs associated with navigating the website. Note that $\tilde{\alpha}_{it}^m$ does not depend on j, i.e., is identical across all potential friends. We model $\tilde{\alpha}_{it}^m$ as follows:

$$\tilde{\alpha}_{it}^m = \alpha_i^m + \kappa_1^m W_{it}$$

where α_i^m is user i's time-invariant tendency to have few or many friends and follows a normal distribution with mean $\bar{\alpha}^m$ and standard deviation σ_{α^m}. W_{it} denotes the length of time (in days) user i has been a member of the website and κ_1^m captures how the net of benefit and cost of forming friendship ties changes with membership length.

$\mathcal{N}_{j,t-1}^m$ captures the properties of the network concerning user j including the out-degree, operationalized as the cumulative number of friends user j has, and transitivity, operationalized as the proportion of friends user j and user i have in common.[27]

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

$$\mathcal{R}_{ij,t-1}^m = \kappa_2^m \mathsf{R}_{ij,t-1}^m + \zeta_{ij}^m$$

where $\mathsf{R}_{ij,t-1}^m$ captures observed similarity and includes the proportion of adopted animes users i and j have in common, and demographic similarity between user i and user j in terms of age, gender, and country.[28] Providing such demographic information is optional in the

[27] The proportion of friends users j and i have in common is defined as the number of friends user i and user j have in common divided by user $i's$ number of friends. Furthermore, given the bi-directional nature of friendships in our data and the operationalization of the out-degree variable as j's number of friends, the direction of the effect is not identified in our model, i.e., it is possible that either (i) users with more friends are more attractive as potential friends or (ii) users with more friends seek to form more friendships. The same issue also arises for the expertise variable defined as j's number of posts. We thank an anonymous reviewer for pointing this out. We discuss this issue in detail in the Results section and provide institutional details and descriptive data patterns that support our interpretation of the estimated coefficient in Web Appendix G.

[28] The proportion of adopted animes users i and j have in common is defined as the number of adopted animes users i and j have in common divided by user $i's$ number of adopted animes.

network we study. Thus we also control for whether age, gender, and country information of both individual i and individual j are available using three dummy variables. ζ_{ij}^m captures the *unobserved* similarity between users i and j. Note that $\zeta_{ij}^m = \zeta_{ji}^m$. The unobserved similarity ζ_{ij}^m follows a normal distribution with mean 0 and standard deviation σ_{ζ^m}.

$\mathcal{K}_{j,t-1}^m$ describes user j's expertise and only depends on j's attributes. We operationalize $\mathcal{K}_{j,t-1}^m$ as the cumulative number of UGC posts user j has published. This variable captures the utility gained from information sharing and learning from friends who are knowledgeable about animes (Watson and Johnson 1972; Brandtzæg and Heim 2009).

C_{ijt}^m contains several variables whose effects we control for. First, we include a weekend dummy. Second, we control for the cumulative number of animes user j has adopted. Third, we also include a dummy variable indicating whether user j was active on the platform during the previous week. This variable captures user j's visibility.[29] Fourth, as mentioned before, we include dummy variables indicating whether user i provided demographic information. Fifth, we include time fixed effects to address common shocks that might result in correlated unobservables affecting friending decisions across users. We operationalize these time fixed effects as week dummies.[30] And sixth, we also include a dummy variable indicating whether user i joined the website before July 2007. Lastly, we assume that ϵ_{it}^m follows a normal distribution with a correlation matrix as specified in Equation (5).

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

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

where α_i^{pa} represents user i's intrinsic tendency to watch animes and is assumed to follow a normal distribution with mean $\bar{\alpha}^{pa}$ and standard deviation $\sigma_{\alpha^{pa}}$. $\mathcal{F}_{i,t-1}^{pa}$ captures the effects of user i's friendship network on user i's actions. It includes user i's total number of friends by time $t-1$, the number of animes watched by all of user i's friends in time $t-1$, and the

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

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

number of posts written by all of i's friends in time $t-1$.[31] Note that we use lagged versions of the variables in $\mathcal{F}_{i,t-1}^{pa}$ to address simultaneity concerns in identifying peer effects (see also Trusov, Bodapati, and Bucklin 2010, Ameri, Honka, and Xie 2019, Bollinger, Burkhardt, and Gillingham 2020). Previous literature has shown that having more friends might directly affect the amount of social activities of network users (Shriver, Nair, and Hofstetter 2013; Toubia and Stephen 2013).[32] In addition, the number of animes watched by all of user i's friends captures the direct influence of friends' activities on user i's product adoptions, while the number of posts written by all of user i's friends reflects the spill-over effect of friends' activities in post publishing on user i's activity in anime watching.

$\mathcal{A}_{i,t-1}^{pa}$ captures state dependence in anime watching and is operationalized as a dummy variable which equals 1 if user i watched an anime at $t-1$ and 0 otherwise.[33] Furthermore, C_t^{pa} contains several controls including a weekend dummy and week fixed effects. And lastly, we assume that ϵ_{it}^{pa} is normally distributed with a correlation matrix as specified in Equation (5). Similarly, user i's utility from writing a post, U_{it}^{cg}, is given by

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

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

[31] Ideally, we would like to estimate separate coefficients for each friend. However, that is infeasible because of the large number of coefficients that would have to be estimated. So instead, we implicitly assume that "all friends are equal" and that their activities exert the same influence (per unit) on the focal user. In a robustness check, we conducted a median split of our three variables of making friends, watching animes, and posting UGC and estimate separate coefficients for popular/unpopular friends and active/inactive friends. The results are presented in Web Appendix D. We conducted t-tests and were not able to reject the null hypotheses that the difference in the coefficients is zero for each of the coefficient pairs at $p < 0.05$. We believe that this finding is driven by our specific empirical application: in the nascent stage of this evolving network, egos had a median of 2 friends (mean of 3.8 friends) at the end of the observation period. Therefore, it is likely that all friends and all friends' actions for sufficient attention from the focal user.

[32] 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 derive from posting on social media, the image these activities create for users also motivates them to perform these activities. They also find that image-related utility is more dominant for users with more friends.

[33] Heckman (1981) differentiates between *structural* and *spurious* state dependence. Structural state dependence arises when past choices (e.g., past purchases) affect current choices (e.g., current purchases). Spurious state dependence arises due to unobserved, time-persistent consumer differences. Both structural and spurious state dependence results in inert behavior. Following the literature in both marketing and economics (e.g., Keane 1997, Seetharaman, Ainslie, and Chintagunta 1999, Shum 2004), we include normally distributed individual-specific random effects to control for unobserved consumer heterogeneity in each of the three equations describing friendship formation, anime watching, and UGC production. As pointed out by Dubé, Hitsch, and Rossi (2010), a mis-specification of the distributional assumption on unobserved heterogeneity will likely result in upward biased state dependence estimates. Thus, the effect of the past behavior dummy can be interpreted as structural state dependence if our assumption of normally distributed individual-specific random effects is correct.

user i's total number of friends by time $t-1$, the number of animes watched by all of i's friends in time $t-1$, and the number of posts written by all of user i's friends in time $t-1$.

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

ESTIMATION

We show details of the log likelihood derivation in Web Appendix B. The log likelihood of the model is given by

$$\begin{aligned} LL = \log \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \prod_{t=1}^{T} \prod_{i=1}^{N} & \left[\Pr\left(A^{pa}_{it}=1\right)\right]^{A^{pa}_{it}} \left[1-\Pr\left(A^{pa}_{it}=1\right)\right]^{1-A^{pa}_{it}} \\ & \cdot \left[\Pr\left(A^{cg}_{it}=1\right)\right]^{A^{cg}_{it}} \left[1-\Pr\left(A^{cg}_{it}=1\right)\right]^{1-A^{cg}_{it}} \\ & \cdot \prod_{j=i+1}^{N} \left[\left[\Pr\left(m_{ijt}=1\right)\right]^{m_{ijt}} \left[1-\Pr\left(m_{ijt}=1\right)\right]^{1-m_{ijt}}\right]^{1-m_{ij,t-1}} d\epsilon\, d\alpha\, d\zeta \end{aligned} \tag{9}$$

where A^{pa}_{it} and A^{cg}_{it} indicate the incidence of an activity – anime watching and UGC production, respectively – of user i in time period t. We estimate our model using Simulated Maximum Likelihood (SMLE). To calculate standard errors of the parameter estimates, we use the BHHH estimator, i.e., the outer product of the gradient, instead of the numerical Hessian (Berndt et al. 1974). Here, we would like to make two comments on the estimation.

First, for computational reasons, the conventional approach of estimating a model via MLE and SMLE involves taking the logarithm of the model likelihood to convert an extremely-small-in-value product of probabilities to a sum of the logarithms of these probabilities. This approach cannot be applied to the likelihood of our model for several reasons (see Web Appendix B for details), i.e., we *cannot* convert the product of the probabilities into a sum of the logarithms of these probabilities (see Equation (9)). 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.[34] To make the likelihood estimation computationally tractable, we use a transformation of the logarithm of a sum of variables to a function of the logarithm of those variables. Details on the transformation and our estimation approach are presented in Web Appendix B.

And second, to speed up the estimation, we use tensorization of large matrices and parallel computing methods to estimate the model. Due to the large size of the data and parallelization, we cannot run the estimation code on conventional computing systems.[35] We utilize several large memory super-computing servers including the Texas Advanced Computing Center (TACC), the Pittsburgh Supercomputer Center (PSC) (Towns et al. 2014; Nystrom et al. 2013), and Jetstream cloud-computing (Stewart et al. 2015; Towns et al. 2014).[36] Details on estimation using supercomputers are presented in Web Appendix C.

IDENTIFICATION

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

The mean intrinsic propensities, $\bar{\alpha}^m$, $\bar{\alpha}^{pa}$, and $\bar{\alpha}^{cg}$, are identified by average user behavior in each of the three areas across users and across time (see Figure 6). The diagonal covariance matrix of the user random effects, Σ^α, is identified by variation in average activity levels across users (see Figure 6). These individual-specific random effects capture unobserved, time-invariant behavior such as a user's tendency to accept friendships or to watch animes. The correlation matrix of the error terms, G, is identified by variation in the simultaneous co-occurrence of activities in a day (see Figure 8). The term σ_ζ^m captures the standard devi-

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

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

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

ation of the unobserved similarity between user i and user j and is identified by variation in friendship formation across different pairs of users. And lastly, conditional on G, the three utilities are separately identified since each action is modeled as a function of other actions in the previous time period.

Identifying influence in social settings is a challenging task (Manski 1993). To be able to identify peer effects, one has to address three factors that can also result in correlated behavior: correlated unobservables, simultaneity, and endogenous group formation. We address correlated unobservables by including week fixed effects and simultaneity by including lagged versions of variables capturing friends' activities. We address the issue of endogenous group formation by explicitly modeling how users become friends and by accounting for observed and unobserved homophily.

Recall that homophily refers to friends behaving in a similar manner due to their similar preferences and not because of one influencing the other. Similarity in unobserved preferences, if unaccounted for, can lead to correlated errors which, in turn, lead to upward biased estimates of friends' influence. We address this issue by incorporating unobserved time-invariant components, α_i^m, α_i^{pa}, and α_i^{cg}, in a user's decisions to form ties, to adopt animes, and to generate content (similar approach as in Nair, Manchanda, and Bhatia 2010, Trusov, Bodapati, and Bucklin 2010, and Ameri, Honka, and Xie 2019). Since we model the incidence of users' actions and not the specific taste for *which* product to adopt or *what* type of content to generate, homophily only plays a role in the frequency level of users' actions, i.e., whether they perform an action on each day. For example, two friends are similar to each other if both tend to publish a lot of posts. In our model, this unobserved heterogeneity in the propensity to perform each of the three actions is captured by α_i^m, α_i^{pa}, and α_i^{cg}. Motivated by the data patterns shown in Figure 6, α_i^m, α_i^{pa}, and α_i^{cg} are assumed to be time-invariant.[37] Moreover, since many of the users are new to the network, the latent propensities are identified not only by the variation in behavior after any friendship formation, but also by behavior before any

[37]Note that we account for the declining trend in friendship formation by including membership length as a covariate (see discussion in the Data section).

ties are formed, i.e., when friends' influence is absent.

RESULTS

We present the estimation results for egos in Table 3. As discussed in the Estimation section, we infer separate coefficients for egos and alters in all three utility functions. The complete sets of results for both egos and alters are available in Web Appendix F. The majority of the estimates for egos and alters have the same sign and significance level, indicating that, overall, the explanatory variables have qualitatively similar effects on both groups of users. In the following, we focus on discussing the results for the egos. Column (i) in Table 3 contains the results for a model in which the decisions about the three types of actions of making friends, watching animes, and publishing posts are made independently of each other. Column (ii) presents the parameter estimates for a model in which we allow these three decisions to be correlated, but there is no unobserved heterogeneity among users. And lastly, column (iii) depicts the results for our full model in which we allow for both correlated errors and unobserved heterogeneity among users.

The results across the three different specifications are overall consistent. We therefore focus on evaluating the results for our main model in column (iii). Potential simultaneous incidences of the three types of actions a user might engage in each day are captured through correlations of the error terms. We find a significant positive correlation between friendship formation and product adoption. We also find a significant coefficient for the standard deviation of the individual-specific random effect in the friendship formation utility suggesting the presence of unobserved heterogeneity in the intrinsic propensity of making friends. And lastly, as expected, user i's net benefit from making friends declines with the length of her membership on the platform. This result suggests that having friends helps to reduce the learning costs associated with navigating the website.

Friendship Formation

We start by discussing the parameter estimates for the friendship formation utility. As expected, network properties matter. The effect of the number of friends user j has is positive and significant (marginal effect of .0003). Given the bi-directional nature of friendships in our data and the operationalization of the out-degree variable as user j's number of friends, the direction of the effect is not identified in our model, i.e., it is possible that either (i) users

Table 3: Model Estimation Results for Egos

	(i) Independent	(ii) Homogenous	(iii) Main
Friendship Formation			
Network Properties			
j's Number of Friends by $t-1$[a]	.4434***	.4601***	.4512***
	(.0071)	(.0067)	(.0082)
Ratio of Number of Friends in Common with j	.0727***	.0768***	.1147***
to $i's$ Number of Friends by $t-1$[a]	(.0100)	(.0106)	(.0106)
Similarity			
Ratio of Number of Animes in Common with j	.1842***	.1828***	.1652***
to $i's$ Number of Animes by $t-1$[a]	(.0051)	(.0084)	(.0311)
Dummy for Whether i and j Are	.2030***	.1904***	.2526***
Within 5 Years of Age	(.0191)	(.0185)	(.0216)
Dummy for Whether i and j	.1394***	.1405***	.1263***
Have the Same Gender	(.0056)	(.0225)	(.0349)
Dummy for Whether i and j Are	.2240***	.0034	.2338***
from Same Country	(.0058)	(.0107)	(.0327)
Standard Deviation of Pair-Specific	.0001	.0069	.0004
Random Effect	(.0043)	(.0040)	(.0044)
Expertise			
j's Number of Written Posts by $t-1$[a]	.0413**	.0500***	.0540***
	(.0127)	(.0117)	(.0131)
Control Variables			
Number of Membership Days by t[a]	−.5192***	−.5396***	−.5384***
	(.0027)	(.0024)	(.0028)
j's Number of Watched Animes by $t-1$[a]	−.0198***	−.0154***	−.0202***
	(.0051)	(.0043)	(.0049)
Dummy for Whether t Is a Weekend	−.3105***	−.3038***	−.5239***
	(.0072)	(.0055)	(.0076)
Dummy for Whether j was Active from	.0801***	.0824***	.1126***
$t-7$ to $t-1$	(.0188)	(.0176)	(.0194)
Dummy for Whether Both i and j	−.2631***	−.2756***	−.2007***
Indicate Their Gender	(.0047)	(.0101)	(.0341)
Dummy for Whether Both i and j	−.3187***	−.3199***	−.3907***
Indicate Their Age	(.0051)	(.0113)	(.0336)
Dummy for Whether Both i and j	−.3731***	−.3504***	−.7357***
Indicate Their Country	(.0069)	(.0154)	(.0260)
Constant	−1.6586***	−1.6650***	−1.5335***
	(.0071)	(.0136)	(.0209)
Dummy for i Having Joined Before	.1364***	.1404***	.1538***
July 2007	(.0018)	(.0017)	(.0017)
Standard Deviation of Individual-Specific	.0303***		.290***
Random Effect	(.0004)		(.0004)
Week Dummies	yes	yes	yes

Standard errors in parentheses.

* $p < .05$, ** $p < .01$, *** $p < .001$

[a] Measured on logarithmic scale.

Table 3: Model Estimation Results for Egos (Continued.)

	(i) Independent	(ii) Homogenous	(iii) Main
Anime Watching			
Number of Friends by $t-1$[a]	$-.0193^{**}$	$-.0170$	$-.0088$
	(.0072)	(.0134)	(.0229)
Number of Animes Watched by Friends in $t-1$[a]	$.1477^{***}$	$.1396^{***}$	$.1506^{***}$
	(.0064)	(.0206)	(.0361)
Number of Posts Published by Friends in $t-1$[a]	$-.0394^{***}$	$-.0398^{**}$	$-.0388$
	(.0026)	(.0143)	(.0375)
Dummy for Whether i Watched an Anime in $t-1$	1.1275^{***}	1.1236^{***}	1.1169^{***}
	(.0020)	(.0045)	(.0197)
Dummy for Whether t Is a Weekend	$.1465^{***}$	$.1479^{***}$	$.1570^{***}$
	(.0039)	(.0097)	(.0477)
Constant	-1.8350^{***}	-1.8094^{***}	-1.9329^{***}
	(.0046)	(.0178)	(.0219)
Standard Deviation of Individual-Specific Random Effect	$.0099^{***}$		.0099
	(.0016)		(.0336)
Week Dummies	yes	yes	yes
Content Generation			
Number of Friends by $t-1$[a]	$.1701^{***}$	$.1723^{***}$	$.2038^{***}$
	(.0107)	(.0105)	(.0175)
Number of Animes Watched by Friends in $t-1$[a]	$.1278^{***}$	$.1329^{***}$	$.1101^{***}$
	(.0134)	(.0154)	(.0248)
Number of Posts Published by Friends in $t-1$[a]	$.1848^{***}$	$.1872^{***}$	$.1666^{***}$
	(.0152)	(.0146)	(.0237)
Dummy for Whether i Published a Post in $t-1$	1.9848^{***}	1.9013^{***}	2.2277^{***}
	(.0005)	(.0017)	(.0064)
Number of Animes Watched by $t-1$[a]	$-.0526^{**}$	$-.0276$	$-.0169$
	(.0175)	(.0185)	(.0379)
Dummy for Whether t Is a Weekend	$-.0049^{***}$	$-.0008$	$.0765^{***}$
	(.0011)	(.0040)	(.0094)
Constant	-2.7046^{***}	-2.6870^{***}	-3.0452^{***}
	(.0024)	(.0027)	(.0110)
Standard Deviation of Individual-Specific Random Effect	$.0142^{***}$		.0114
	(.0019)		(.0125)
Week Dummies	yes	yes	yes
Error Correlation Matrix			
Correlation between Friendship and Adoption		$.3245^{***}$	$.3095^{***}$
		(.0028)	(.0032)
Correlation between Friendship and UGC		$-.0243^{**}$	$-.0247$
		(.0086)	(.0210)
Correlation between Adoption and UGC		$-.0036$	.0007
		(.0192)	(.0472)
Model Summary Statistics			
Number of Observations	69,020,774	69,020,774	69,020,774
AIC	359,515.60	359,501.20	359,180.20
BIC	361,698.39	361,683.99	361,411.14
LogLikelihood	-179,621.80	-179,614.60	-179,451.10

Standard errors in parentheses.

* $p < .05$, ** $p < .01$, *** $p < .001$

[a] Measured on logarithmic scale.

with more friends are more attractive as potential friends or (ii) users with more friends seek to form more friendships. However, institutional details and descriptive data patterns, which we discuss in detail in Web Appendix G, suggest that (i) is the more appropriate interpretation, i.e., that well-connected users are desirable candidates for friendship, in our empirical context. The most important pieces of evidence presented in Web Appendix G are that users had no monetary incentives to add more friends (such as influencers have nowadays) and that users mostly form friendships during the first 12 months of platform membership. Older users, even those with many friends, only add few friends after the first year.

The result that users with many friends are more attractive as potential friends stands in contrast to the negative estimate for the number of friends a potential friend has found in Christakis et al. (2010). We suspect that this result is due to the unique context of the online social network environment in which users come from many different countries and do not know one another in real life. In such an environment, popular users are more likely to be "known" by other users, i.e., visible to other users, which is a prerequisite for tie formation. In comparison, students attending the same school usually know each other so popular students do not have an advantage in this regard. In addition, initiating and maintaining many connections in the virtual world may be less costly than in the real life.

Next, the effect of the proportion of friends user i and user j have in common is significant and positive (marginal effects of .0008), implying that users are more likely to connect with friends of friends. Two average users who have 50% of friends in common are 44% more likely to become friends than two average users who have no friends in common. This finding is in line with results in the previous literature (e.g., Aral, Muchnik, and Sundararajan 2009; Shalizi and Thomas 2011; Bhattacharya et al. 2019).

In terms of similarity, the proportion of common animes has a significant positive effect with a marginal effect of .0008, suggesting that common interests increase the probability of forming a friendship. Further, similarity in the three demographic traits, i.e., age, gender, and country, also increases the likelihood of friendship formation, consistent with the finding in Christakis et al. (2010). Comparing the magnitudes of the effects of the three demographics (marginal effects in parentheses), we find that having similar age (.0019) is the most important trait, followed by coming from the same country (.0014), and having the same gender (.0009) being the least important trait. Put together, two users who share one demographic trait are, on average, three times more likely to become friends as two users who do not share any demographic traits. Lastly, the coefficient for the standard deviation of the pair-specific random effect capturing latent homophily between individual i and individual j is positive but insignificant. This finding suggests that our flexible modeling approach together with the observed variables included in our model captures the similarity between two individuals well. User j's number of written posts represents her domain expertise and has a significant positive effect on friendship formation, albeit a small one (marginal effect of .0000). UGC publishing is conducive to the formation of a social relationship.[38]

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 that user j's activities increased her visibility and awareness among other platform users. User j's (cumulative) number of watched animes has a significant negative effect on her forming a friendship tie with user i, albeit the effect is very small.[39] Further, the weekend dummy has a significant negative coefficient, suggesting that users are less likely to form friendships on weekends.

Online Activities

Next, we discuss our results related to a user's anime watching decisions. The number of animes watched by friends has a significant positive effect on a user's anime watching behavior. For example, if a focal user's friends watch two animes the previous day, the probability that

[38]The same directionality issue discussed in context with user j's number of friends also pertains to user j's number of posts, i.e., given the bi-directional nature of friendships in our data and the operationalization of the expertise construct, the direction of the effect is not identified in our model. It is possible that either (i) users with more posts are more attractive as potential friends or (ii) users who post more seek to form more friendships. We discuss institutional details and descriptive data patterns in Web Appendix G that support our interpretation of the estimated coefficient. The most important piece of evidence is that users who post more do not form more new friendships.

[39]Although both the number of animes user j watched and the number of posts user j wrote capture aspects of a user's experience and interest in anime watching, there is a key difference between these two activities: users tend to watch animes alone (i.e., not with their friends within the network), but they write posts to interact with each other within the online community. In other words, UGC posting is a "social" or "interactive" activity but anime watching is not. As a result, users who watch a lot of animes may have little time for other activities such as interacting with other users and making new friends, while users who post a lot of UGC interact with other users which create opportunities to meet and make new friends.

the focal user watches animes the next day increases by 15%. However, the number of friends a user has does not have a significant influence. These results indicate that having more active friends increases a user's activity level due to the presence of direct peer effects. We do not find a spill-over effect of friends' posting behavior on a user's anime watching: a user is not more likely to watch an anime if her friend made a post the previous day. Further, our results reveal positive state dependence in anime watching: a user is more likely to watch an anime if she did so 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.

We now describe our results related to content generation. Similar to our findings for anime watching, we observe evidence of a direct peer effect on a user's content generation: the number of posts published by friends during the previous day has a significant and positive effect on a user's content generation decision on the following day. We also note a positive spill-over effect: the number of animes watched by friends increases a user's UGC production. In addition, the coefficient for the number of friends a user has is significant and positive, suggesting that having more friends makes a user more active in publishing content. Further, there is evidence of positive state dependence in content generation, i.e., we observe a significant positive effect of a user's posting on her posting behavior the following day. And lastly, the weekend dummy has a significant positive effect.

Comparing the magnitudes of the direct peer effect versus the spill-over effect using marginal effects, we observe the direct peer effect to have a larger impact than the spill-over effect in increasing activity levels for both anime watching and UGC publishing.[40] For example, if a focal user's friends write two more posts the previous day, the probability that the focal user also writes a post the next day increases by 26%. In contrast, if a focal user's friends watch two more animes, the probability that the focal user writes a post the next day only increases by 17%.

Model Fit

[40] The following are marginal effects on anime watching: number of animes watched by friends (.0053), number of posts published by friends (−.0028). The following are marginal effects on UGC publication: number of animes watched by friends (.0004), number of posts published by friends (.0012).

We evaluate the fit of our models using two criteria: AIC and BIC. Our main model outperforms the independent model and the homogeneous model on both criteria (see bottom of Table 3). We also compare the (in-sample) predictive performance of the three models using hit rates for three holdout periods: the last 30 days, the last 60 days, and the last 90 days of our sample period. The average hit rates across these three holdout periods for each of the three activities (i.e., friendship formation, anime watching, and UGC production) are presented in Table 4. The average hit rates are very similar across the three models with the main model performing slightly better than the other two models in predicting UGC posts.

Table 4: Hit Rates

	(i) Independent Model	(ii) Homogenous Model	(iii) Main Model
Friending Actions	99%	99%	99%
Adoption Actions	72%	72%	72%
UGC Actions	94%	95%	96%

PREDICTION EXERCISES

For companies operating social networks, advertising revenue represents their primary source of income. In 2020, the industry earned revenues of over 40 billion through advertisements.[41] 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.

Using our model estimates, we examine the effects of stimulating different types of users and different types of in-site activities through a series of prediction exercises. More precisely, we assume that the platform can trigger an increase in any of the three activities of making friends, watching animes, and generating content by, e.g., posting a recommendation list on a user's page: the platform can recommend to a user to become friends with some other users, to adopt some specific animes, or to participate in forum discussions that are active and related to the user's past adoptions or posts. Although we do not observe the login or page

[41] https://www.statista.com/statistics/736971/social-media-ad-spend-usa/

view activities of a user and, as a result, cannot directly translate the changes in activity levels to changes in ad viewership, as long as users are not spending *less* time on each activity compared to before the stimulation, an increase in the total activity level will also lead to an increase in the time spent on the website. Furthermore, an increase in the activity level is observable by other users and non-users of the website and therefore can lead to activity cascades and a growing user base.

Platform-Wide Stimulation

We compare the effects of different platform-wide stimulations on *overall* activity levels, i.e., the sum of activities in all three areas. The overall activity levels serve as a proxy for the total site traffic or total time spent on the site. The predictive exercises are implemented as follows: in each scenario, we increase one type of activity (friendship tie formation, product adoption or UGC generation) among all egos and all alters by the amount of that activity on a typical day for an average user. To put it differently, our stimulation doubles the amount of activity of a particular type on a given day for each ego and each alter. We do so on days 30, 90, and 150 and predict users' behavior going forward until day 184.[42] When presenting our findings, we focus on findings for the egos.

Table 5: Prediction from a Platform-Wide Stimulation

Stimulation Activity	Stimulation Day	(i) Change in Active USERS in %	(ii) Change in ACTIVITIES in %
Friendship Formation			
	30	2.76	1.97
	90	3.23	.43
	150	12.66	2.63
	Average	*6.22*	*1.68*
Anime Watching			
	30	1.80	12.66
	90	4.03	2.36
	150	17.14	6.47
	Average	*7.66*	*7.17*
UGC Production			
	30	1.74	1.69
	90	.43	.44
	150	13.90	3.21
	Average	*6.20*	*1.78*

[42] If a user joined the website after the stimulation day, her actions are simulated from her join date (without stimulation).

The results are presented in Table 5. Column (i) shows the changes in average number of active egos per day, i.e., the mean number of egos who perform at least one activity in a day, and column (ii) depicts the changes in the number of total activities performed by egos. Out of the three types of stimulations the platform can implement (i.e., to recommend friends, animes, or forum discussion topics), stimulating users to watch more animes is the most effective intervention resulting in the highest overall increase in the number of active users and in the level of in-site activities. This is because anime watching is the most frequent activity of the three in-site activities and because the peer effect of friends' anime watching is a stronger driver than the spill-over effect of users' anime watching. The effects of stimulating users to make more friends or to post more UGC are very similar in terms of magnitude.

Friendship Recommendations

Friendship recommendation systems are a popular tool used by online social networks such as Facebook or LinkedIn to increase network connectedness (Sun and Taylor 2020). Here, we ask *whom* to recommend as a potential friend in an evolving social network, i.e., what type of individual should a platform suggest for friendship.

We evaluate the effects of forming a friendship with five types of potential friends: (i) popular user, i.e., the user with the largest number of friends; (ii) friend of a friend, i.e., the user who has the most common friends with the focal user; (iii) active anime watcher, i.e., the user who watches the most animes; (iv) active UGC creator, i.e., the user who publishes the most posts; (v) user with similar interest, i.e., the user who watches the largest number of common animes with the focal user. Furthermore, this predictive exercise relies on the assumption that a friend recommendation strategy is successful in convincing all egos and alters to accept one recommended user as a new friend on day 150.[43] We then simulate users' behavior going forward until day 184, and compare the increases in each of the three activities as well as the overall activity level among all egos.

[43] We could relax this assumption by lowering the acceptance rate from 100% to, e.g., 50%. In this case, the effects of the friendship recommendations would become smaller. However, as long as the acceptance rate of friendship recommendations is constant across the different recommendation strategies (recommending a user j with different characteristics), the rank order of the effectiveness of different recommendation strategies remains the same. That is why we focus on the effectiveness rank order when interpreting the results.

Among the five friendship recommendation strategies, recommending a friend of a friend is the most effective strategy in stimulating subsequent friendship formation within the evolving network.[44] This finding stands in contrast to the common result for static networks for which recommending a popular user has been found to be the most effective strategy.[45] A potential reason for our different result is that a popular user does not necessarily share interests with or is not necessarily similar to the focal user. Thus friends of the popular user are also less likely to be potential candidates for new friendships, thereby limiting the cascading effect. On the other hand, a friend of a friend is more likely to share interests with or be similar to the focal user. Thus friends of a friend of a friend are more likely to be potential candidates for new friendships, thereby enhancing the cascading effect. Our data and estimation results support this explanation: friends of a friend have, on average, 1.5 common friends with the focal user, while friends of a popular user have, on average, 0.3 common friends with the focal user. Friends of a friend also watch a larger number of the same animes than friends of a popular user (9.3 versus 8.4). And lastly, we find significant and positive effects of similarity variables in the estimation (see Results section).

Seeding

Previous literature has found that users often have varying degrees of activity in different areas (e.g., Manchanda, Xie, and Youn 2008; Iyengar, Van den Bulte, and Valente 2011). Our data confirm this pattern (see Figure 5). 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 decisions for platform owners in achieving a desired outcome such as an increase in UGC production.

In this set of prediction exercises, we examine the effectiveness of different seeding strategies in increasing tie formations, anime watching, and UGC production in an evolving social network. For these prediction exercises, we select the 15% most/least active egos among all

[44] Recommending an active UGC creator is the most effective strategy in increasing anime watching, UGC publishing, and the overall level of in-site activities.

[45] The strategy of recommending a popular user as a potential friends works equally well (or better) as the strategy of recommending a friend of a friend for users who joined the platform very early, who have relatively many friends, and who are very active (in terms of anime watching and UGC posting), i.e., resemble active users in more mature networks.

egos based on their activity levels in each of the three activities ("selection activity") as our seeding targets.[46] As a benchmark, we also randomly select 15% of egos as the initial seeds. Next, we increase the activity level of these selected egos in one of the three areas of activity ("seeding activity") by the amount of that activity on a typical day for an average user. This is the same stimulation as in the Platform-Wide Stimulation section. We do so on days 30, 90, and 150 ("stimulation date") and simulate users' behavior going forward until day 184. To recap, we perform a total of 63 prediction exercises, including $3 \cdot 3 \cdot 3 = 27$ prediction exercises for the 15% most active egos, $3 \cdot 3 \cdot 3 = 27$ prediction exercises for the 15% least active egos, as well as $3 \cdot 3 = 9$ prediction exercises for our benchmark case of randomly selected users. And lastly, when presenting our findings, we again focus on the results for egos.

We start by investigating the question whether platform owners should seed to the most or least active users. We find that seeding to the 15% most active users is a more effective strategy than seeding to the 15% least active users when the goal is an increase in anime watching or in UGC production. However, for friendship formation, seeding to the 15% least active users is a more effective strategy than seeding to the 15% most active users. Seeding to 15% randomly selected users performs the worst for all three outcomes.[47]

Next, inspired by results for static networks, we evaluate whether seeding to the most-connected users, i.e., users with the most friends, is also the most effective seeding strategy in terms of the choice of selection activity in evolving networks. In other words, we wonder whether the most-connected users are indeed better candidates for seeding than users who are most active in anime watching or UGC creation in an evolving social network. In answering this question, we focus on the 27 prediction exercises which use the most active users as initial seeds. Our results indicate that, on average, seeding to users who post the most UGC is the most effective seeding strategy, followed by seeding to most-connected users, and lastly by seeding to

[46] 15% of egos are equivalent to 8 egos on day 30, 26 egos on day 90, and 50 egos on day 150.

[47] Across all three selection activities, all three seeding activities, and all three stimulation dates, i.e., twenty-seven prediction exercises for *each* outcome activity, seeding to the 15% most active users increases anime watching and UGC production, on average, by .32% and 1.40%, respectively, while seeding to the 15% least active users increases anime watching and UGC production, on average, by only .15% and .96%, respectively. Seeding to the 15% least active users increases friendship formation, on average, by 4.20%, while seeding to the 15% most active users increases friendship formation, on average, by 3.88%. Random seeding increases friendship formation, anime watching, and UGC production by 3.48%, .14%, and .44%, respectively.

users watching the most animes.[48] A similar pattern also holds for the individual prediction exercises: in 25 out of the 27 prediction exercises, seeding to the most connected users is less effective than either seeding to users posting the most UGC or users watching the most animes. We conclude that, in contrast to the common finding for static networks, seeding to the most-connected users is not the most effective strategy in evolving networks.

While the discussion in the previous paragraph focused on the choice of the best selection activity with the goal of increasing overall activity levels, here, we investigate the most effective seeding strategy for each desired outcome in terms of choosing *both* a selection *and* a seeding activity. For all three potential desired outcomes of increasing the number of friendships, increasing anime watching or increasing UGC posting, the most effective seeding activity is the desired activity itself. However, the selection activity does not follow the same pattern. If the goal is to increase the number of friendship, then choosing the users who watch the fewest animes and encouraging them to make more friends is the most effective strategy. To increase anime watching, selecting the most-connected users and encouraging them to watch even more animes is the best strategy. And lastly, to increase UGC posting, the winning strategy is to select users who create the most content and encourage them to post more content.

Accounting for Network Evolution

First, we examine by how much the effectiveness of seeding strategies is underestimated when the endogenous network formation is shut down. To do so, we re-run 54 of the 63 prediction exercises discussed in the Seeding section, but do *not* allow users to form new friendships.[49] We find that not allowing for endogenous network formation leads, on average, to an underestimation of seeding effectiveness by 41%. This result underscores the practical importance of modeling the co-evolution of individual users' friendship tie formations and their concurrent in-site activities to correctly measure seeding effectiveness.

Second, we investigate whether we can replicate previous literature's common finding that

[48] Across the three seeding activities and the three stimulation days (9 prediction exercises for *each* selection activity), on average, seeding to the most connected users results in an average increase of 1.84% of activity levels across friendship formation, anime watching, and UGC production. In comparison, seeding to users who watch a lot of animes and post a lot of UGC leads to an average increase of 1.80% and 1.95%, respectively, of activity levels.

[49] We do *not* re-run the predictions for randomly sampled users.

seeding to the most-connected users is the most effective strategy to increase online activity levels (anime watching and UGC posting in our case) in static networks. To do so, we focus on the subset of 27 prediction exercises in which users are not allowed to form new friendships and which use the 15% most active users (in terms of friends, anime watching, and UGC posting) as initial seeds, i.e., we center on the selection strategy. Consistent with previous literature (e.g., Trusov, Bodapati, and Bucklin 2010; Hinz et al. 2011; Aral, Muchnik, and Sundararajan 2013), we find that, in static networks, seeding to the most-connected users is, on average as well as in all but two individual prediction exercises, the most effective strategy to increase online activity levels.

And lastly, we investigate whether our results for the case when *both* the selection *and* the seeding strategy can be chosen continue to hold when the endogenous network formation is shut down. For the two desired outcomes of increasing anime watching or increasing UGC posting, we indeed find the same two strategies to be most effective: to increase anime watching, selecting the most-connected users and encouraging them to watch even more animes is the best strategy. To increase UGC posting, the winning strategy is to select users who create the most content and encourage them to post more content.

CONCLUSION

In this paper, we study the co-evolution of individuals' friendship tie formations and their concurrent activities within an evolving online social network. Our findings shed light on the important factors that drive strangers to become friends in an online environment. Specifically, our results reveal that, all three drivers, i.e., network properties, similarity, and expertise, matter for a user's friendship formation decisions. Further, while having more friends does not necessarily make a person more active, having more active friends does increase a user's activity levels in terms of both product adoptions and content generation through peer and spill-over effects. Via prediction exercises, we find that stimulating product adoptions is the most effective site-wide intervention which leads to the highest overall site traffic and the largest number of active users and that recommending a friend of a friend as a potential friend is the most effective strategy in stimulating friendship tie formation. Contrary to previous studies investigating static networks, we find that seeding to most connected users is not the best strategy to increase users' overall activity levels in an evolving network.

We believe that our results are generalizable in two directions: first, MyAnimeList.com is only one example of an interest-based online community. There are many such online communities including standalone ones (e.g., goodreads.com, boardgamegeek.com, dpadd.com, etc.) and special interest groups on large social media platforms (e.g., subreddit on reddit.com, online groups on Facebook, etc.). And second, we believe that the insights about friending/recommendation/stimulation interventions (e.g., evolving versus static networks) are useful for social networks in general, especially for those which are in a fast growing phase.

This paper has several limitations and presents opportunities for future research. First, we only observe a friendship if both users agree to become friends. In other words, we observe neither the friendship request nor the potential rejection of that request. This is a limitation of our data. As a result, we cannot separately identify whether an increase in a user's number of friends is due to that user's elevated preference to form friendships or due to her increased desirability as a potential friend to other users. Second, our anime watching data are self-reported and accuracy in the reporting is a potential concern. While there are no explicit incentives for users of MyAnimeList.net to falsely report their true anime watching behavior, even without explicit incentives, users may misrepresent their watching behavior due to social desirability concerns or the need to appear knowledgeable on the platform. This is a limitation of our data. Third, we do not observe churn, i.e., users abondoning the platform. This is a limitation of our data. Understanding drivers of churn and whether it is contagious is important for online platforms which are commonly populated with a large number of inactive users.

Fourth, we do not explicitly model time-varying unobserved factors that may impact subsets of users. Although some institutional features of the data and model (e.g., users of the website are spread out around the world, we include random effects and time fixed effects to control for individual preference and some unobserved correlated factors) alleviate this concern, they do not completely eliminate it. Furthermore, a more robust solution of incorporating group-time fixed effects is computationally infeasible given the model structure and the large number of required fixed effects (see Lee 2007; Lee, Liu, and Lin 2010; Ma, Krishnan, and Montgomery 2014). Fifth, we do not allow for heterogeneity in peer influence in our main model. While we did not find significantly different effects of popular/unpopular friends and active/inactive friends in a robustness check, a more thorough investigation of this important question is left for future research.

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

References

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

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

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

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

Badev, Anton (2021), "Nash-Equilibria on (Un)Stable Networks," *Econometrica*, 89 (3), 1179–1206.

Banerjee, Abhijit, Arun Chandrasekhar, Esther Duflo, and Matthew Jackson (2013), "The Diffusion of Microfinance," *Science*, 341 (6144).

Berndt, Ernst, Bronwyn Hall, Robert Hall, and Jerry Hausman (1974), "Estimation and Inference in Nonlinear Structural Models," *Annals of Economic and Social Measurement*, 3/4 (1), 653–665.

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

Bhattacharya, Prasanta, Tuan Phan, Xue Bai, and Edoardo Airoldi (2019), "A Coevolution Model of Network Structure and User Behavior: The Case of Content Generation in Online Social Networks," *Information Systems Research*, 30 (1), 117–132.

Boguñá, Marian, Romualdo Pastor-Satorras, Albert Díaz-Guilera, and Alex Arenas (2004), "Models of Social Networks Based on Social Distance Attachment," *Physical Review E*, 70 (5), 56–122.

Bollinger, Bryan, Jesse Burkhardt, and Kenneth Gillingham (2020), "Peer effects in residential water conservation:Evidence from migration," *American Economic Journal: Economic Policy*, 12 (3), 107–133.

Boucher, Vincent (2016), "Conformism and self-selection in social networks," *Journal of Public Economics*, 136 (C), 30–44.

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

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

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

Claussen, Jorg, Benjamin Engelstätter, and Michael Ward (2014), "Susceptibility and Influence in Social Media Word-of-Mouth," *Working Paper*.

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

Dubé, Jean-Pierre, Günter Hitsch, and Peter Rossi (2010), "State Dependence and Alternative Explanations for Consumer Inertia," *The RAND Journal of Economics*, 41 (3), 417–445.

Gargiulo, Martin and Mario Benassi (2000), "Trapped in Your Own Net? Network Cohesion, Structural Holes, and the Adaptation of Social Capital," *Organization Science*, 11 (2), 183–196.

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

Greenan, Charlotte (2015), "Diffusion of Innovations in Dynamic Networks," *Journal of the Royal Statistical Society: Series A*, 178 (1), 147–166.

Hanaki, Nobuyuki, Alexander Peterhansl, Peter Dodds, and Duncan 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 Ccatherine Tucker (2008), "Modeling Social Interactions: Identification, Empirical Methods and Policy Implications," *Marketing Letters*, 19 (3-4), 287–304.

Heckman, James (1981), "Statistical models for discrete panel data," *Structural analysis of discrete data with econometric applications*, 114, 178.

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

Hoff, Peter, Adrian Raftery, and Mark Handcock (2002), "Latent Space Approaches to Social Network Analysis," *Journal of the American Statistical Association*, 97 (460), 1090–1098.

Hsieh, Chih-Sheng and Lung Lee (2016), "A Social Interactions Model with Endogenous Friendship Formation and Selectivity," *Journal of Applied Econometrics*, 31 (2), 301–319.

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

Jackson, Matthew (2008), *Social and Economic Networks*, Princeton University Press.

Jackson, Matthew, Tomas Rodriguez-Barraquer, and Xu Tan (2012), "Social Capital and Social Quilts: Network Patterns of Favor Exchange," *American Economic Review*, 102 (5), 1857–97.

Katona, Zsolt (2013), "Competing for Influencers in a Social Network," *Working Paper*.

Katona, Zsolt and Tamás Móri (2006), "A New Class of Scale Free Random Graphs," *Statistics & Probability Letters*, 76 (15), 1587–1593.

Keane, Michael (1997), "Modeling Heterogeneity and State Dependence in Consumer Choice Behavior," *Journal of Business and Economic Statistics*, 15 (3), 310–327.

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

Lanz, Andreas, Jacob Goldenberg, Daniel Shapira, and Florian Stahl (2019), "Climb or Jump: Status-Based Seeding in User-Generated Content Networks," *Journal of Marketing Research*, 56 (3), 361–378.

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.

Lewis, Kevin, Marco Gonzalez, and Jason Kaufman (2012), "Social Selection and Peer Influence in an Online Social Network," *Proceedings of the National Academy of Sciences of the United States of America*, 109 (1), 68–72.

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

Lovett, Mitch and Richard Staelin (2016), "The Role of Paid and Earned Media in Building Entertainment Brands: Reminding, Informing, and Enhancing Enjoyment," *Marketing Science*, 35 (1), 142–157.

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.

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

Mele, Angelo (2017), "A Structural Model of Dense Network Formation," *Econometrica*, 85 (3), 825–850.

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

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

Perry, Brea, Bernice Pescosolido, and Stephen Borgatti (2018), *Egocentric Network Analysis: Foundations, Methods, and Models*, Cambridge University Press.

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

Seetharaman, P.B., Andrew Ainslie, and Pradeep Chintagunta (1999), "Investigating Household State Dependence Effects Across Categories," *Journal of Marketing Research*, 36 (4), 488–500.

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

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

Shum, Matthew (2004), "Does Advertising Overcome Brand Loyalty? Evidence from the Breakfast Cereal Market," *Journal of Economics and Management Strategy*, 13 (2), 241–272.

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

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

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

Steglich, Christian, Tom Snijders, and Michael Pearson (2010), "Dynamic Networks and Behavior: Separating Selection from Influence," *Sociological Methodology*, 40 (1), 329–393.

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

Sun, Tianshu and Sean Taylor (2020), "Displaying Things in Common to Encourage Friendship Formation: A Large Randomized Field Experiment," *Quantitative Marketing and Economics*, 18 (3), 237–271.

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.

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

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

Trusov, Michael, Anand Bodapati, and Randy 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 and David Johnson (1972), *Social Psychology: Issues and Insights*, Lippincott Philadelphia.

Web Appendix A: Country Distribution

Table A-1: Country Distribution

Country	Frequency	Percentage	Country	Frequency	Percentage
US	380	27.4%	UAE	3	0.2%
Great Britain	64	4.6%	Ukraine	3	0.2%
Poland	60	4.3%	Argentina	2	0.1%
Canada	50	3.6%	Bangladesh	2	0.1%
Netherlands	40	2.9%	Dominican Republic	2	0.1%
Germany	34	2.5%	Hong Kong	2	0.1%
Israel	34	2.5%	New Zealand	2	0.1%
Brasil	29	2.1%	Peru	2	0.1%
Italy	24	1.7%	Taiwan	2	0.1%
France	22	1.6%	Tunisia	2	0.1%
Sweden	22	1.6%	Bosnia and Herzegovina	1	0.1%
Australia	18	1.3%	Bulgaria	1	0.1%
Norway	18	1.3%	Chile	1	0.1%
Finland	16	1.2%	Colombia	1	0.1%
Lithuania	16	1.2%	Cyprus	1	0.1%
Philippines	16	1.2%	Czech Republic	1	0.1%
Hungary	15	1.1%	Estonia	1	0.1%
Mexico	14	1.0%	India	1	0.1%
Russia	13	0.9%	Indonesia	1	0.1%
Singapore	12	0.9%	Kuwait	1	0.1%
Malaysia	10	0.7%	Maldives	1	0.1%
Belgium	8	0.6%	Malta	1	0.1%
Latvia	8	0.6%	Panama	1	0.1%
Romania	8	0.6%	Saudi Arabia	1	0.1%
Denmark	7	0.5%	Slovenia	1	0.1%
Japan	6	0.4%	South Korea	1	0.1%
Portugal	5	0.4%	Spain	1	0.1%
Austria	4	0.3%	Switzerland	1	0.1%
Slovakia	4	0.3%	Missing	386	27.8%
Greece	3	0.2%			

Web Appendix B: Log-Likelihood Derivation and Estimation

In this section, we show the steps in deriving the model log-likelihood and explain the techniques used for estimation. Given the conditional independence assumption of user i's decision to become friends with each user j (as discussed in the Model section) and given the need for mutual agreement to become friends, the probability of a tie forming between individual i and individual j is given by

$$\Pr\left(m_{ijt}=1\right)=\Pr\left(U_{ijt}^{m}>0\right)\cdot\Pr\left(U_{jit}^{m}>0\right). \tag{B1}$$

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

$$L_{ijt\,|\,\alpha_i,\alpha_j,\zeta_{ij},\epsilon_{it},\epsilon_{jt}}^{m}=\left[\left[\Pr\left(m_{ijt}=1\right)\right]^{m_{ijt}}\left[1-\Pr\left(m_{ijt}=1\right)\right]^{1-m_{ijt}}\right]^{1-m_{ij,t-1}} \tag{B2}$$

where $\zeta_{ij}=\zeta_{ij}^{m}$, $\alpha_i=\{\alpha_i^{m},\alpha_i^{pa},\alpha_i^{cg}\}$ and α_j is defined similarly. Note that $L_{ijt\,|\,\alpha_i,\alpha_j,\zeta_{ij},\epsilon_{it},\epsilon_{jt}}^{m}$ conditions on the two users not being friends before time t through the exponent $1-m_{ij,t-1}$. The likelihoods for the other two types of activities, i.e., product adoption and content generation, at time t are given by

$$\begin{aligned}
L_{it\,|\,\alpha_i,\epsilon_{it}}^{pa}&=\left[\Pr\left(A_{it}^{pa}=1\right)\right]^{A_{it}^{pa}}\left[1-\Pr\left(A_{it}^{pa}=1\right)\right]^{1-A_{it}^{pa}}\\
L_{it\,|\,\alpha_i,\epsilon_{it}}^{cg}&=\left[\Pr\left(A_{it}^{cg}=1\right)\right]^{A_{it}^{cg}}\left[1-\Pr\left(A_{it}^{cg}=1\right)\right]^{1-A_{it}^{cg}}
\end{aligned} \tag{B3}$$

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

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

$$\begin{aligned}
L_{it\,|\,\alpha_i,\alpha_j,\zeta_{ij},\epsilon_{it},\epsilon_{jt}}=&\left[\Pr\left(A_{it}^{pa}=1\right)\right]^{A_{it}^{pa}}\left[1-\Pr\left(A_{it}^{pa}=1\right)\right]^{1-A_{it}^{pa}}\\
&\cdot\left[\Pr\left(A_{it}^{cg}=1\right)\right]^{A_{it}^{cg}}\left[1-\Pr\left(A_{it}^{cg}=1\right)\right]^{1-A_{it}^{cg}}\\
&\cdot\prod_{j=i+1}^{N}\left[\left(\Pr\left(m_{ijt}=1\right)\right]^{m_{ijt}}\left[1-\Pr\left(m_{ijt}=1\right)\right]^{1-m_{ijt}}\right]^{1-m_{ij,t-1}}\qquad i\neq j.
\end{aligned} \tag{B4}$$

Note that α_j and ϵ_{jt} also enter the above equation for user i because they are part of the probability of user i and user j becoming friends at time t (see Equation (B1)).

The full likelihood can be written as

$$L = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \prod_{t=1}^{T} \prod_{i=1}^{N} \left[\Pr\left(A_{it}^{pa}=1\right)\right]^{A_{it}^{pa}} \left[1-\Pr\left(A_{it}^{pa}=1\right)\right]^{1-A_{it}^{pa}} \cdot \left[\Pr\left(A_{it}^{cg}=1\right)\right]^{A_{it}^{cg}} \left[1-\Pr\left(A_{it}^{cg}=1\right)\right]^{1-A_{it}^{cg}} \cdot \prod_{j=i+1}^{N} \left[\left[\Pr\left(m_{ijt}=1\right)\right]^{m_{ijt}} \left[1-\Pr\left(m_{ijt}=1\right)\right]^{1-m_{ijt}}\right]^{1-m_{ij,t-1}} d\epsilon d\alpha d\zeta \tag{B5}$$

and the log likelihood of the model is given by

$$LL = \log \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \prod_{t=1}^{T} \prod_{i=1}^{N} \left[\Pr\left(A_{it}^{pa}=1\right)\right]^{A_{it}^{pa}} \left[1-\Pr\left(A_{it}^{pa}=1\right)\right]^{1-A_{it}^{pa}} \cdot \left[\Pr\left(A_{it}^{cg}=1\right)\right]^{A_{it}^{cg}} \left[1-\Pr\left(A_{it}^{cg}=1\right)\right]^{1-A_{it}^{cg}} \cdot \prod_{j=i+1}^{N} \left[\left[\Pr\left(m_{ijt}=1\right)\right]^{m_{ijt}} \left[1-\Pr\left(m_{ijt}=1\right)\right]^{1-m_{ijt}}\right]^{1-m_{ij,t-1}} d\epsilon d\alpha d\zeta. \tag{B6}$$

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

For computational reasons, the conventional approach of estimating a model via MLE and

SMLE involves taking the logarithm of the model likelihood to convert an extremely-small-in-value product of probabilities to a sum of the logarithms of these probabilities. This approach cannot be applied to the likelihood of our model for three reasons. First, recall that, at any time t, the error terms in the three utility functions are correlated (see Equation (5)). Therefore the integral taken over $f(\epsilon)$ has to include user i's likelihood of all three activities at time t. Second, recall that the probability of a friendship formation depends on both user i's and user j's utilities for the tie formation, i.e. a friendship is only formed if both users derive positive utilities from doing so (see Equation (B1)). Since at each time t, all users can become friends with any other user with whom they are not friends yet, all friendship formation decisions of all users at time t are connected through the pair-specific random effect ζ_{ij} and the error terms in users' friendship formation utilities. In other words, due to the second reason, the integrals over $f(\epsilon)$ and $f(\zeta)$ have to to include all friendship formation probabilities of all users at time t. Combining the first and second reason, it is evident that the integrals over $f(\epsilon)$ and $f(\zeta)$ have to include the probabilities of all actions of all users at time t.

Third, recall that our model includes time-invariant individual-specific intrinsic propensities for each type of activity, i.e., α_i^m, α_i^{pa}, and α_i^{cg}, and a time-invariant pair-specific random effect ζ_{ij}. Therefore, for each user and each type of activity, the integrals over $f(\alpha)$ and $f(\zeta)$ have to include all activities of that type over all time periods. Given that the first two reasons necessitate that the integrals over $f(\epsilon)$ and $f(\zeta)$ contain the probabilities of all actions of all users at each time t and given that the third reason necessitates that the integrals over $f(\alpha)$ and $f(\zeta)$ contain all probabilities over all time period for a specific type of activity and a specific user, the integrals over $f(\epsilon)$, $f(\alpha)$, and $f(\zeta)$ have to contain the probabilities of all actions of all users over all time periods (see Equation (B5)). As a result of these three issues, when we take the logarithm of the model likelihood, we *cannot* convert the product of the probabilities into a sum of the logarithms of these probabilities (see Equation (B6)). 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. To make the likelihood estimation computationally tractable, we use the following transformation of the logarithm of a sum of variables to a function of the logarithm of those variables.

Let R denote the number of random draws. Given this estimation method and the law of large numbers, the log-likelihood can be written as follows:

$$\begin{aligned} LL = \; & \log \frac{1}{R} \sum_{r=1}^{R} \Big[\prod_{t=1}^{T} \prod_{i=1}^{N} (\Pr(A_{it}^{pa}=1))^{A_{it}^{pa}} (1-\Pr(A_{it}^{pa}=1))^{1-A_{it}^{pa}} \\ & \qquad \cdot (\Pr(A_{it}^{cg}=1))^{A_{it}^{cg}} (1-\Pr(A_{it}^{cg}=1))^{1-A_{it}^{cg}} \\ & \qquad \cdot \prod_{j=i+1}^{N} \left[(\Pr(m_{ijt}=1))^{m_{ijt}} (1-\Pr(m_{ijt}=1))^{1-m_{ijt}} \right]^{1-m_{ij,t-1}} \Big] \Big|_r \\ = & -\log R + \log \sum_{r=1}^{R} Q_r \end{aligned} \tag{B7}$$

with

$$\begin{aligned} Q_r = & \prod_{t=1}^{T} \prod_{i=1}^{N} (\Pr(A_{it}^{pa}=1))^{A_{it}^{pa}} (1-\Pr(A_{it}^{pa}=1))^{1-A_{it}^{pa}} \cdot (\Pr(A_{it}^{cg}=1))^{A_{it}^{cg}} (1-\Pr(A_{it}^{cg}=1))^{1-A_{it}^{cg}} \\ & \cdot \prod_{j=i+1}^{N} \left[(\Pr(m_{ijt}=1))^{m_{ijt}} (1-\Pr(m_{ijt}=1))^{1-m_{ijt}} \right]^{1-m_{ij,t-1}} \Big|_r \end{aligned} \tag{B8}$$

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

$$\log \sum_{i=0}^{N} a_i = \log a_0 + \log \left(1 + \sum_{i=1}^{N} e^{(\log a_i - \log a_0)} \right). \tag{B9}$$

Thus we can write the model log-likelihood as

$$LL = -\log R + \log Q_1 + \log \big(1 + \sum_{r=2}^{R} e^{(\log Q_r - \log Q_1)} \big), \tag{B10}$$

with

$$
\begin{aligned}
&\log Q_r = \\
&\sum_{t=1}^{T}\sum_{i=1}^{N}\Bigg[A_{it}^{pa}\cdot\log\left(\Pr\left(A_{it}^{pa}=1\right)\right)+\left(1-A_{it}^{pa}\right)\cdot\log\left(1-\Pr\left(A_{it}^{pa}=1\right)\right) \\
&\qquad +A_{it}^{cg}\cdot\log\left(\Pr\left(A_{it}^{cg}=1\right)\right)+\left(1-A_{it}^{cg}\right)\cdot\log\left(1-\Pr\left(A_{it}^{cg}=1\right)\right) \\
&\qquad +\sum_{j=i+1}^{N}\Big[\left(1-m_{ij,t-1}\right)\cdot\left[m_{ijt}\cdot\log\left(\Pr\left(m_{ijt}=1\right)\right)+\left(1-m_{ijt}\right)\cdot\log\left(1-\Pr\left(m_{ijt}=1\right)\right)\right]\Big]\Bigg].
\end{aligned}
\tag{B11}
$$

Note that theoretically, the choice of either of the draw sets as a_0 does not matter. However, in practice, one needs to be careful when choosing the base of a_0 so that taking the exponent does not produce numbers beyond computer limits. To achieve this goal, one method is choosing the maximum value for a_0 in such a way that it minimizes the chance of underflow. However, this method does not work for our estimation. This is the case because the loglikelihood value is relatively large and the values of $\log(a_i)$'s produced by the final estimates in our main model are typically within a range of about 2,000 units from each other. Since the exponent of a number larger than $\sim \pm 700$ is beyond the computer limit, the method of choosing the maximum a_0 does not prevent overflow. Instead, in each step of the calculation, we monitor the calculated values to ensure that they do not go beyond computing limits and set them to valid values if they they go beyond the limit. This adjustment is extremely minute compared to the loglikelihood of our model and does not materially affect the estimation process.

Web Appendix C: Super Computing Estimation

In our model, the probability of a tie formation depends on both individual i's and individual j's utilities for the tie formation. The errors in user i's and user j's tie formation utilities are simultaneously correlated with user i's and user j's errors in the utilities for the activities in the other two areas, i.e., product adoption and content generation. As a result, the likelihood is given by the product of over 136,000,000 probabilities calculated using about 68 million observations.

Recall that we estimate our model using Simulated Maximum Likelihood Esmiation (SMLE) taking 30 draws for each user from a standard normal distribution to estimate the standard deviations of user random effects and taking 30 draws for each user in each time period from a standard normal distribution to estimate the covariance matrix of the three error terms. Further, we use the Cholesky decomposition of the error covariance matrix in the estimation. The initial data set containing the independent variables and sets of 30 random draws for different REs and error matrices took up about 30GB of RAM memory. As detailed in the previous paragraphs, the calculations of friendship utility need to be done over 68 million observations and 30 times (once for each set of draws). Furthermore, since we use probit probabilities, for each friending decision, we need two draws from a normal distribution CDF based on utilities of users i and user j. As a result, for calculating the final likelihood, several calculations need to be done over 4 billion times. We use parallel computing to speed up the process.

There are two ways to approach the parallel computing of over 4 billion calculations. The first one is to do the calculations of user i's and user j's utilities using each of the 30 different sets of draws separately, using 30 (or its factor) number of CPU cores. The issue with this approach is that each CPU core still needs to handle 68 million lines of observations and reserve memory for them. Using each CPU core would take up an additional 5-8GB of memory. For example, using 10 CPU cores would require a system with more than 100GB of RAM. This high requirement restricts the computing options significantly.

The second approach to parallel computing is to divide the 68 million observations among different CPU cores. In this case, each CPU core would do the calculations over a set of observations, for all 30 sets of draws. This approach is possible in our case because the calculations of each user's utility and the draw from a normal distribution based on that utility are done independently from other observations. Using this method, each CPU core is handling a smaller matrix and, as a result, requires less memory.

With the help of XSEDE Extended Collaborative Support Service (ECSS), we implemented the above algorithm by splitting the process into two programs: One program prepares the data, and the second program performs optimization and gathers the results. The first program contains a data preparation step that only needs to be run once for each hybrid multinode configuration of nodes and CPUs used. It first creates the large matrix of network related variables (68 million rows), then breaks it into the desired number of parts (along the rows) and stores them on the hard drive. By testing different number of nodes, we found that the use of 8 to 12 computing nodes and all CPU cores on each node provides the best tradeoff in parallelization and memory usage in terms of shortest computation time and minimal error in file reading from and saving on the hard drive. The following code is specifically doing the splitting step:

```
nparts=12   # Number of parts along the row the data is broken into
  N =nrow(X_f_i)   # Get number of observatins
  cby = ceiling(N/nparts)   # Determine how many observations go into each part
  cparts=seq(0,N,cby)   # A vector containing the boundary row number of each part
  if (cparts[length(cparts)]<N) {  # Add the last row as the finishing boundary
    cparts[length(cparts)+1]=N
    }
  for (cpi in seq(1:(length(cparts)-1))) {   # Loop for separating and saving each data part
    bei=cparts[cpi]+1;eni=cparts[cpi+1];   # Beginning and end row numbers of each data part
    X_f_i_part=X_f_i[bei:eni,];    # Selecting the appropriate section of data
    X_f_j_part=X_f_j[bei:eni,];
    save(X_f_i_part,X_f_j_part,Distributer_i,Distributer_j,Distributer_a,bei,eni,cpi,X_a,X_c,
          REs, REDistributer_i,REDistributer_j,REDistributer_a,
          Obs_f, Obs_a,
          N,T,K,error,TieDummy,ActionDummy,ContentDummy,Adoption,Content,
          pairRE_ij,DistPair_ij,
          Tdistributer_f, Tdistributer_a,
          file=paste('81M_Data_Px_',nparts,'_',cpi,'.Rdata',sep=''));   #Saving the necessary
variables for calculating friending utilities.
    }
```

The discussed step needs to be done only once, because the created data parts are saved on the hard drive.

The second part of the process is the execution of the optimization. For this step, a number of computing nodes corresponding to each data part are used. One of the nodes is assigned as main node, and the others are marked as secondary nodes. When initializing the nodes using `mpirun`, a Perl code is ran on each node. This code takes the number of nodes and cores to be used on each node as input, assigns a number to each node and then runs the R script for optimization with the assigned number(`task_index`) as one of its inputs.

```
#!/usr/bin/perl
  use strict;
  use warnings;
  my $ntasks = $ARGV[0]; #input
  my $numparts = $ARGV[1]; #input
  my ($myid, $numprocs) = split(/\$s+/, `./getid`); #getid is a user-written function that
will return cpu id and total processor in MPIjob
  #for (my $i=0; $i<$ntasks; $i++) {
    # if($myid == $i % $numprocs) {
      # if($numprocs % $numtasks == 0) { #make sure appropriate number of processors are in
each node
        my $task_index = $myid+1;
        `module load R;Rscript Optimization.R $task_index $numparts $ntasks >
                XP_hyb_out.$task_index.txt`; # run the R script with assigned node number,
number of data parts, and number of tasks to run as input
      # }
    #}
  #}
```

Each node runs the R optimization code separately. However, within the code, there are sections that will run only if the task index of the node is the index of the main node. The main steps in the optimization function are as follows: First, all nodes read the data part that was created previously and that corresponds to their task index, i.e., each node has access to and loads only one part of the main data. Then all nodes run the utility calculations that require parallel computing on their own data part. Note that the computation of utilities of each pair can be done independently from other pairs. The random draws can be matched with each data part using distributer variables already created and included with all data parts (`REDistributer_i`, `REDistributer_j`, `Distributer_i`, `Distributer_j`, `DistPair_ij`). The appropriate section of the distributor variables are selected using the beginning row and end row indices of the part. Then each node calculates the friending loglikelihood and saves the numbers in a text file on the hard drive.

```
optimizationfunction = function(theta) {
#### General calculations until it reaches the friending utility calculation that requires parallel computing
  ...
#### Friending utility calculations for the 30 draws.  Each node only does the calculation on their corresponding data part
  res_xfor = foreach(k = 1:K) %dopar% { #each node does parallel computing within its cores for calculations on its data part
    fcovmat = as.vector(cov_mat[,k,1]) #k^th draw for error matrix
    fRE = as.vector(re_mat[,k,1]) #k^th draw for individual RE
    pair_ij=as.vector(re_ij[,k]) #k^th draw for pair RE
#note that the xt_i is already only selected with indices bei to eni
    fP_Ui = pnorm(fRE[REDistributer_i[bei:eni]]+fcovmat[Distributer_i[bei:eni]]+xt_i
                  + pair_ij[DistPair_ij[bei:eni]])
    fP_Uj = pnorm(fRE[REDistributer_j[bei:eni]]+fcovmat[Distributer_j[bei:eni]]+xt_j
                  + pair_ij[DistPair_ij[bei:eni]])
    fP_U = fP_Ui*fP_Uj
    fP_U_sum <- sum(log(fP_U*TieDummy[bei:eni]+1-fP_U*1-TieDummy[bei:eni]),na.rm=TRUE)
  }
#Every node write out value for their part of matrix calculation
  res_part=matrix(unlist(res_xfor),nrow=K,ncol=1,byrow=TRUE) # likelihood of friending probabilities related to corresponding data part
  str(res_part)
  save(res_part,file=fn2use)
```

The rest of the work is done by only the main node. The main node stays on standby, constantly checking whether all the nodes have created the file with their calculated portion of the friending likelihood. Once all the files have been found, the main node exits the standby state, reads the loglikelihood matrix written in each file, combines the matrices, and then deletes the files. Then the main node proceeds with the rest of the model likelihood calculations. While the main file does the remaining calculations, secondary nodes remain idle and wait for the main node to finish its job. Once the main node has finished the calculations, it creates a file for each of the secondary nodes, signaling the end of the calculations and entering the next step of the optimization process.

```
if (myrank==1){ #rank 1 is the main node and this part is only done by main node
  allfnd =0 # initializing existing secondary node created file counter
  fn2rec =‘‘Res_xp_’’ # format of the name for the files with calculated portion of the friending likelihood
  fndres=wait_for_workerfiles(); # user-written wait function that counts the number of files on hard disk with the specified name format
  allfnd=fndres[1]
  if (allfnd==1){
    res_xp_xn=matrix(0,K,1)  #set up space to hold results across workers
    for (cpi in seq(1:nparts)){  #first get sum across workers
      fn2get =paste(‘‘Res_xp_’’,cpi,‘‘.Rdata’’,sep=‘‘’’)
      load(fn2get)
      res_xp_xn =res_xp_xn+res_part
```

```
      file.remove(fn2get)
      }
# rest of the likelihood calculation for Adoption and Content Generation part for get final
prob of the 30 draws only run by rank 1
    ...
    print(loglikelihood)
# now save a copy of final total sum for each secondary node node to signal end of calculations
for one step of optimization
    for (cpi in seq(1:nparts)){
      fn2save =paste("Res_xp_sum_",cpi,".Rdata",sep="")
      save(res_xp_xn,loglikelihood,file=fn2save)
    }
  }else{ #if not all secondary node files found, then quit
    ...# printing error messages
    quit()
  }
# end of rank 1 (main node) gathering of final sums from secondary nodes
} else{ # nodes other than main node remain idle until the main node is done with the calculations
    sumfnd =0 # initializing existing main node created file counter
    master_results = wait_for_masterfile() # user-written wait function that counts the number
of files on hard disk with the specified name format
    sumfnd = master_results[1]
    if (sumfnd==0) { #if not all files found, then quit
      ...  # printing error messages
      quit()
    }
  }
}
```

Using this hybrid programming algorithm, each of the models used in this paper requires about 170 to 200 hours for the optimization process to converge.

Functions used in the discussed procedure are as follows:

getid.c

```
#include <stdio.h>
#include <mpi.h>
#include <stdlib.h>
int main(int argc, char **argv)
{
int numprocs, my_id;
MPI_Status status;
MPI_Init(&argc, &argv);
MPI_Comm_size(MPI_COMM_WORLD, &numprocs);
MPI_Comm_rank(MPI_COMM_WORLD, &my_id);
MPI_Finalize();
printf("%d %d\n",my_id,numprocs);
} /* end main() */
```

Wait functions

```
args = commandArgs(trailingOnly=TRUE) # Getting the following parameters from Slurm mpirun
command:  n-th-core-task num-fileparts total-core-tasks
dompi =1
whdel = .10  #delay each while loop iteration to wait this num-secs for files
whmax = 1500 #total delay
```

```
 filedel = 1.5  #delay a moment after file-exist=true search, to let file fill up
nc2do =10;  #num of cores per node to use
#------------------------------------------
# function to check for matrix mult.  total sum
# called by workers to wait for sum file from main node
#------------------------------------------
wait_for_masterfile=function() {
  sumfnd=0
  whcnt =0
  while (whcnt<whmax && sumfnd==0) {
    if (file.exists(fnsum)) {
      Sys.sleep(filedel)
      load(fnsum)
      sumfnd=1
      print(paste(Sys.time(),'r:',myrank,' 2.1r:',myrank,'del fnsum',fnsum))
      file.remove(fnsum)
      break
    } else {whcnt=whcnt+1
      Sys.sleep(whdel)
  }# -------- end while
  return(list(sumfnd,loglikelihood))
} #res_xp_xn,loglikelihood
#-------------------------------------------------------
# function to check for file-triggers from all workers
# each file-trigger is the matrix sum for the file part from the worker
#------------------------------------------
whcnt =0 #set this up for wide scope
allfnd =0
sumfnd =0
wait_for_workerfiles = function() {
  whcnt =0
  allfnd=0
  while (whcnt<whmax && allfnd==0) {  #loop until all results xparts found
    ffnd =0
    ffndlist={}
    for (cpi in seq(1:nparts)){  #for each part (node) get their result sum
      fn2get=paste("Res_xp_",cpi,".Rdata",sep="")
      if (file.exists(fn2get)) {
        ffnd =ffnd+1
        ffndlist[cpi]=1
      } else {
        ffndlist[cpi]=0
      }
    } #end for
    if (ffnd==nparts) {allfnd=1
    } else {whcnt=whcnt+1
      Sys.sleep(whdel)}
  } #end while
return(list(allfnd,ffnd))
}
```

Web Appendix D: Active and Inactive Friends

Table D-1: Results Allowing for Different Effects of Active and Inactive Friends

	Egos	Alters
Friendship Formation		
Network Properties		
	(0.0067)	(0.0096)
j's Number of Friends by $t-1$[a]	0.5331***	0.6449***
	(0.0067)	(0.0096)
Number of Friends in Common with j by $t-1$[a]	0.0643***	0.188***
	(0.0114)	(0.0163)
Similarity		
Number of Animes in Common with j by $t-1$[a]	0.1893***	0.3079***
	(0.0275)	(0.0388)
Dummy for Whether i and j Are Within 5 years of Age	0.241***	0.1849***
	(0.0185)	(0.0266)
Dummy for Whether i and j Have the Same Gender	0.1455***	0.0551
	(0.0292)	(0.0426)
Dummy for Whether i and j Are from Same Country	0.2174***	0.1016*
	(0.0296)	(0.0427)
Standard Deviation of Pair-Specific Random Effect	0.0044	
	(0.0038)	
Expertise		
j's Number of Written Posts by $t-1$[a]	0.0348***	0.0031
	(0.0078)	(0.0113)
Control Variables		
Number of Membership Days by t[a]	−0.5974***	−0.5958***
	(0.0022)	(0.0032)
j's Number of Watched Animes by $t-1$[a]	−0.0275***	−0.0524***
	(0.0038)	(0.0055)
Dummy for Whether t Is A Weekend	−0.5321	0.0095
	(0.0047)	(0.0068)
Dummy for Whether j was Active from $t-7$ to $t-1$	0.1686***	0.3256***
	(0.0127)	(0.0185)
Dummy for Whether Both i and j Indicate Their Country	−0.7123***	−0.06
	(0.0291)	(0.0409)
Dummy for Whether Both i and j Indicate Their Age	−0.396***	−0.1256**
	(0.0321)	(0.0457)
Dummy for Whether Both i and j Indicate Their Gender	−0.2587***	−0.1204***
	(0.0185)	(0.0262)
Dummy for i Having joined before July 2007	0.1152***	
	(0.0012)	
Standard Deviation of Individual-Specific Random Effect	0.0328***	
	(0.0000)	
Week Dummies	yes	

	Egos	Alters
Anime Watching		
Number of POPULAR Friends by $t-1$[a]	−0.0291	−0.0423
	(0.0266)	(0.0328)
Number of UNPOPULAR Friends by $t-1$[a]	−0.0127	0.0021
	(0.0266)	(0.0328)
Number of Animes Watched by ACTIVE Friends in $t-1$[a]	0.1369***	0.0727
	(0.0309)	(0.0431)
Number of Animes Watched by INACTIVE Friends in $t-1$[a]	0.1637***	0.0223
	(0.0099)	(0.0135)
Number of Posts Published in ACTIVE Friends in $t-1$[a]	−0.023	0.058
	(0.0313)	(0.0442)
Number of Posts Published in INACTIVE Friends in $t-1$[a]	−0.0572***	−0.017***
	(0.0004)	(0.0006)
Dummy for Whether i Watched An Anime in $t-1$[a]	0.9622***	0.913***
	(0.0067)	(0.0085)
Dummy for Whether t Is A Weekend	0.1118***	0.0358***
	(0.0121)	(0.0158)
Constant	−1.5959***	−1.4067***
	(0.0101)	(0.0121)
Standard Deviation of Individual-Specific Random Effect	0.0134	
	(0.009)	
Week Dummies	yes	
Content Generation		
Number of POPULAR Friends by $t-1$[a]	0.0753***	−0.0134
	(0.0291)	(0.0426)
Number of UNPOPULAR Friends by $t-1$[a]	−0.0323	0.0185
	(0.0239)	(0.0334)
Number of Animes Watched by ACTIVE Friends in $t-1$[a]	0.2011	−0.0101
	(0.0295)	(0.0406)
Number of Animes Watched by INACTIVE Friends in $t-1$[a]	0.006	0.006
	(0.0076)	(0.0105)
Number of Posts Published in ACTIVE Friends in $t-1$[a]	0.2442***	0.2741***
	(0.0205)	(0.0282)
Number of Posts Published in INACTIVE Friends in $t-1$[a]	0.2261***	0.3601***
	(0.0003)	(0.0005)
Dummy for Whether i Published A Post in $t-1$[a]	2.2397***	2.1143***
	(0.0035)	(0.004)
Number of Animes Watched by $t-1$[a]	−0.0269	0.0622
	(0.0305)	(0.0432)
Dummy for Whether t Is A Weekend	0.0974***	0.0092
	(0.0115)	(0.016)
Constant	−3.1291***	−3.0318***
	(0.0076)	(0.0138)
Standard Deviation of Individual-Specific Random Effect	0.0117*	
	(0.0058)	
Week Dummies	yes	

Standard errors in parentheses.

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

[a] Measured on logarithmic scale.

Web Appendix E: Results with Smaller Samples

Table E-1: Model Estimation Results for Smaller Samples

	(i) Random Sample of 100		(ii) Random Sample of 200	
	Egos	Alters	Egos	Alters
Friendship Formation				
Network Properties				
j's Number of Friends by $t-1$[a]	0.265***	0.402***	0.401***	0.641***
	(0.053)	(0.075)	(0.014)	(0.020)
Ratio of Number of Friends in Common with j	0.009***	0.019***	0.012	0.006
to $i's$ Number of Friends by $t-1$[a]	(0.009)	(0.013)	(0.001)	(0.002)
Similarity				
Ratio of Number of Animes in Common with j	0.030***	−0.013***	0.022	−0.006
to $i's$ Number of Animes by $t-1$[a]	(0.007)	(0.010)	(0.002)	(0.003)
Dummy for Whether i and j Are	0.115***	0.594***	0.032	0.364
Within 5 Years of Age	(0.220)	(0.321)	(0.008)	(0.011)
Dummy for Whether i and j	0.291***	−0.031	0.298	0.095
Have the Same Gender	(0.182)	(0.271)	(0.014)	(0.020)
Dummy for Whether i and j Are	0.018***	0.130	0.046	0.070
from Same Country	(0.147)	(0.204)	(0.011)	(0.017)
Standard Deviation of Pair-Specific	0.001		0.006***	
Random Effect	(0.003)		(0.001)	
Expertise				
j's Number of Written Posts by $t-1$[a]	−0.020***	−0.148	−0.044	−0.102
	(0.070)	(0.097)	(0.013)	(0.019)
Control Variables				
Number of Membership Days by t[a]	−0.568***	−0.380***	−0.634***	−0.489***
	(0.034)	(0.047)	(0.009)	(0.012)
j's Number of Watched Animes by $t-1$[a]	−0.161***	−0.013***	−0.155***	−0.110
	(0.023)	(0.032)	(0.005)	(0.007)
Dummy for Whether t Is a Weekend	−8.632***	0.070	−8.607***	0.105***
	(0.001)	(0.001)	(0.000)	(0.000)
Dummy for Whether j was Active from	2.901***	2.730***	2.762***	2.493***
$t-7$ to $t-1$	(0.123)	(0.175)	(0.022)	(0.032)
Dummy for Whether Both i and j	0.108***	−0.137	0.068	−0.182
Indicate Their Gender	(0.145)	(0.217)	(0.018)	(0.026)
Dummy for Whether Both i and j	−0.044***	−0.397*	0.079	−0.235
Indicate Their Age	(0.177)	(0.257)	(0.016)	(0.023)
Dummy for Whether Both i and j	−0.355***	−0.017	−0.323*	0.005
Indicate Their Country	(0.185)	(0.258)	(0.014)	(0.019)
Constant	−3.125***	−4.505***	−3.126***	−4.418***
	(0.122)	(0.169)	(0.015)	(0.021)
Dummy for i Having joined before	0.001***		−0.000***	
July 2007	(0.001)		(0.000)	
Standard Deviation of Individual-Specific	0.025***		0.024***	
Random Effect	(0.002)		(0.001)	
Week Dummies	yes		yes	

Standard errors in parentheses.

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

[a] Measured on logarithmic scale.

Table E-1: Model Estimation Results for Smaller Samples (Continued.)

	(i) Random Sample of 100		(ii) Random Sample of 200	
	Egos	Alters	Egos	Alters
Anime Watching				
Number of Friends by $t-1$[a]	-0.035	-0.009	0.028	-0.053
	(0.233)	(0.297)	(0.012)	(0.016)
Number of Animes Watched by Friends in $t-1$[a]	0.135***	0.179	0.105	0.186
	(0.285)	(0.330)	(0.002)	(0.003)
Number of Posts Published by Friends in $t-1$[a]	0.003	0.069	0.021	0.071
	(0.141)	(0.194)	(0.005)	(0.007)
Dummy for Whether i Watched an Anime in $t-1$	0.883***	0.698***	1.020**	0.659
	(0.316)	(0.457)	(0.002)	(0.003)
Dummy for Whether t Is a Weekend	-0.005***	0.063	0.034	0.048
	(0.219)	(0.296)	(0.003)	(0.004)
Constant	-1.490***	-1.359***	-1.602***	-1.192***
	(0.202)	(0.327)	(0.005)	(0.007)
Standard Deviation of Individual-Specific Random Effect	0.000		0.003**	
	(0.274)		(0.001)	
Week Dummies	yes		yes	
Content Generation				
Number of Friends by $t-1$[a]	0.133***	-0.001	0.098	0.002
	(0.149)	(0.214)	(0.006)	(0.008)
Number of Animes Watched by Friends in $t-1$[a]	0.249***	0.070	0.123	0.117
	(0.229)	(0.305)	(0.001)	(0.002)
Number of Posts Published by Friends in $t-1$[a]	0.057***	0.244***	0.024	0.220
	(0.173)	(0.233)	(0.003)	(0.005)
Dummy for Whether i Published a Post in $t-1$	1.314***	1.800***	1.755***	1.633***
	(0.314)	(0.444)	(0.001)	(0.001)
Number of Animes Watched by $t-1$[a]	-0.047	0.023	0.118	0.002
	(0.172)	(0.254)	(0.011)	(0.015)
Dummy for Whether t Is a Weekend	0.016***	-0.021	0.027	0.025
	(0.246)	(0.339)	(0.001)	(0.002)
Constant	-2.415***	-2.274***	-2.768***	-2.071***
	(0.140)	(0.208)	(0.003)	(0.006)
Standard Deviation of Individual-Specific Random Effect	0.018		0.002**	
	(0.463)		(0.001)	
Week Dummies	yes		yes	
Error Correlation Matrix				
Correlation between Friendship and Adoption	0.310***		0.621***	
	(0.003)		(0.001)	
Correlation between Friendship and UGC	-0.025		-0.779***	
	(0.021)		(0.000)	
Correlation between Adoption and UGC	0.001		-0.975	
	(0.047)		(0.000	

Standard errors in parentheses.

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

[a] Measured on logarithmic scale.

Web Appendix F: Complete Estimation Results

Table F-1: Complete Estimation Results for Egos and Alters

	(i) Independent		(ii) Homogenous		(iii) Main	
	Egos	Alters	Egos	Alters	Egos	Alters
Friendship Formation						
Network Properties						
j's Number of Friends by $t-1^{\mathrm{a}}$	0.4434***	0.5443***	0.4601***	0.5598***	0.4512***	0.5523***
	(0.0071)	(0.0103)	(0.0067)	(0.0097)	(0.0082)	(0.0119)
Ratio of Number of Friends in Common with j	0.0727***	0.1591***	0.0768***	0.1644***	0.1147***	0.1288***
to $i's$ Number of Friends by $t-1^{\mathrm{a}}$	(0.0100)	(0.0143)	(0.0106)	(0.0151)	(0.0106)	(0.0150)
Similarity						
Ratio of Number of Animes in Common with j	0.1842***	0.2757***	0.1828***	0.2770***	0.1652***	0.1288***
to $i's$ Number of Animes by $t-1^{\mathrm{a}}$	(0.0051)	(0.0074)	(0.0084)	(0.0121)	(0.0311)	(0.0150)
Dummy for Whether i and j Are	0.2030***	0.1717***	0.1904***	0.1574***	0.2526***	0.1628***
Within 5 Years of Age	(0.0191)	(0.0279)	(0.0185)	(0.0267)	(0.0216)	(0.0311)
Dummy for Whether i and j	0.1394***	0.0636***	0.1405***	0.0580	0.1263***	0.0479
Have the Same Gender	(0.0056)	(0.0085)	(0.0225)	(0.0324)	(0.0349)	(0.0506)
Dummy for Whether i and j Are	0.2240***	0.0922***	0.0034	0.0855***	0.2338***	0.0875
from Same Country	(0.0058)	(0.0084)	(0.0107)	(0.0155)	(0.0327)	(0.0473)
Standard Deviation of Pair-Specific	0.0001		0.0069		0.0004	
Random Effect	(0.0043)		(0.0040)		(0.0044)	
Expertise						
j's Number of Written Posts by $t-1^{\mathrm{a}}$	0.0413**	0.0074	0.0500***	0.0144	0.0540***	0.0142
	(0.0127)	(0.0183)	(0.0117)	(0.0169)	(0.0131)	(0.0190)
Control Variables						
Number of Membership Days by t^{a}	−0.5192***	−0.5287***	−0.5396***	−0.5535***	−0.5384***	−0.5591***
	(0.0027)	(0.0039)	(0.0024)	(0.0035)	(0.0028)	(0.0040)
j's Number of Watched Animes by $t-1^{\mathrm{a}}$	−0.0198***	−0.0470***	−0.0154***	−0.0435***	−0.0202***	−0.0504***
	(0.0051)	(0.0074)	(0.0043)	(0.0062)	(0.0049)	(0.0071)
Dummy for Whether t Is a Weekend	−0.3105***	0.0280**	−0.3038***	0.0169*	−0.5239***	0.0144
	(0.0072)	(0.0103)	(0.0055)	(0.0079)	(0.0076)	(0.0108)
Dummy for Whether j was Active from	0.0801***	0.2384***	0.0824***	0.2428***	0.1126***	0.2510***
$t-7$ to $t-1$	(0.0188)	(0.0275)	(0.0176)	(0.0257)	(0.0194)	(0.0283)
Dummy for Whether Both i and j	−0.2631***	−0.0970***	−0.2756***	−0.1217***	−0.2007***	−0.0918
Indicate Their Gender	(0.0047)	(0.0070)	(0.0101)	(0.0140)	(0.0341)	(0.0482)
Dummy for Whether Both i and j	−0.3187***	−0.0953***	−0.3199***	−0.1095***	−0.3907***	−0.1019*
Indicate Their Age	(0.0051)	(0.0072)	(0.0113)	(0.0160)	(0.0336)	(0.0475)
Dummy for Whether Both i and j	−0.3731***	−0.0647***	−0.3504***	−0.0375	−0.7357***	−0.0640
Indicate Their Country	(0.0069)	(0.0096)	(0.0154)	(0.0215)	(0.0260)	(0.0367)
Constant	−1.6586***	−1.0497***	−1.6650***	−1.0740***	−1.5335***	−1.1202***
	(0.0071)	(0.0071)	(0.0136)	(0.0140)	(0.0209)	(0.0215)
Dummy for i Having Joined Before	0.1364***		0.1404***		0.1538***	
July 2007	(0.0018)		(0.0017)		(0.0017)	
Standard Deviation of Individual-Specific	0.0303***				0.0290***	
Random Effect	(0.0004)				(0.0004)	
Week Dummies	yes		yes		yes	

Standard errors in parentheses.

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

[a] Measured on logarithmic scale.

Table F-1: Complete Estimation Results for Egos and Alters (Continued.)

	(i) Independent		(ii) Homogenous		(iii) Main	
	Egos	Alters	Egos	Alters	Egos	Alters
Anime Watching						
Number of Friends by $t-1$[a]	-0.0193^{**}	-0.0567^{***}	-0.0170	-0.0514^{***}	-0.0088	-0.0458
	(0.0072)	(0.0099)	(0.0134)	(0.0155)	(0.0229)	(0.0326)
Number of Animes Watched by Friends in $t-1$[a]	0.1477^{***}	0.0857^{***}	0.1396^{***}	0.0860^{**}	0.1506^{***}	0.0775
	(0.0064)	(0.0089)	(0.0206)	(0.0305)	(0.0361)	(0.0491)
Number of Posts Published by Friends in $t-1$[a]	-0.0394^{***}	0.0541^{***}	-0.0398^{**}	0.0563^{**}	-0.0388	0.0518
	(0.0026)	(0.0039)	(0.0143)	(0.0211)	(0.0375)	(0.0529)
Dummy for Whether i Watched an Anime in $t-1$	1.1275^{***}	1.0620^{***}	1.1236^{***}	1.0523^{***}	1.1169^{***}	1.0559^{***}
	(0.0020)	(0.0024)	(0.0045)	(0.0054)	(0.0197)	(0.0264)
Dummy for Whether t Is a Weekend	0.1465^{***}	0.0402^{***}	0.1479^{***}	0.0399^{**}	0.1570^{***}	0.0412
	(0.0039)	(0.0020)	(0.0097)	(0.0122)	(0.0477)	(0.0623)
Constant	-1.8350^{***}	-1.6613^{***}	-1.8094^{***}	-1.6326^{***}	-1.9329^{***}	-1.7504^{***}
	(0.0046)	(0.0055)	(0.0178)	(0.0208)	(0.0219)	(0.0264)
Standard Deviation of Individual-Specific Random Effect	0.0099^{***}				0.0099	
	(0.0016)				(0.0336)	
Week Dummies	yes		yes		yes	
Content Generation						
Number of Friends by $t-1$[a]	0.1701^{***}	0.0076	0.1723^{***}	0.0163	0.2038^{***}	0.0072
	(0.0107)	(0.0148)	(0.0105)	(0.0150)	(0.0175)	(0.0268)
Number of Animes Watched by Friends in $t-1$[a]	0.1278^{***}	-0.0064	0.1329^{***}	-0.0037	0.1101^{***}	-0.0028
	(0.0134)	(0.0187)	(0.0154)	(0.0221)	(0.0248)	(0.0329)
Number of Posts Published by Friends in $t-1$[a]	0.1848^{***}	0.2721^{***}	0.1872^{***}	0.2697^{***}	0.1666^{***}	0.2753^{***}
	(0.0152)	(0.0212)	(0.0146)	(0.0206)	(0.0237)	(0.0337)
Dummy for Whether i Published a Post in $t-1$	1.9848^{***}	2.1531^{***}	1.9013^{***}	2.0702^{***}	2.2277^{***}	2.1161^{***}
	(0.0005)	(0.0009)	(0.0017)	(0.0022)	(0.0064)	(0.0080)
Number of Animes Watched by $t-1$[a]	-0.0526^{**}	0.0561^{*}	-0.0276	0.0657^{**}	-0.0169	0.0548
	(0.0175)	(0.0220)	(0.0185)	(0.0225)	(0.0379)	(0.0554)
Dummy for Whether t Is a Weekend	-0.0049^{***}	0.0228^{***}	-0.0008	0.0664^{***}	0.0765^{***}	0.0141
	(0.0011)	(0.0017)	(0.0040)	(0.0052)	(0.0094)	(0.0137)
Constant	-2.7046^{***}	-2.7995^{***}	-2.6870^{***}	2.8395^{***}	-3.0452^{***}	-2.9219^{***}
	(0.0024)	(0.0108)	(0.0027)	(0.0110)	(0.0110)	(0.0231)
Standard Deviation of Individual-Specific Random Effect	0.0142^{***}				0.0114	
	(0.0019)				(0.0125)	
Week Dummies	yes		yes		yes	
Error Correlation Matrix						
Correlation between Friendship and Adoption			0.3245^{***}		0.3095^{***}	
			(0.0028)		(0.0032)	
Correlation between Friendship and UGC			-0.0243^{**}		-0.0247	
			(0.0086)		(0.0210)	
Correlation between Adoption and UGC			-0.0036		0.0007	
			(0.0192)		(0.0472)	
Model Summary Statistics						
Number of Observations	69,020,774		69,020,774		69,020,774	
AIC	359,515.60		359,501.20		359,180.20	
BIC	361,698.39		361,683.99		361,411.14	
LogLikelihood	$-179,621.80$		$-179,614.60$		$-179,451.10$	

Standard errors in parentheses.

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

[a] Measured on logarithmic scale.

Web Appendix G: Institutional Details and Descriptive Evidence

G.1 Institutional Details

Nowadays, on many platforms such as YouTube, Instagram or Twitter, users benefit monetarily from having many friends/followers through advertising revenue sharing, sponsorships, product promotions, free merchandise, tips, etc. Thus, users have a strong incentive to add more and more friends/followers to earn more money and these users are commonly referred to as "influencers." These influences are also very active on platforms, creating and posting a lot of UGC to keep their friends/followers engaged.

However, there were no monetary incentives for users to add more friends on MyAnimeList.net during our study period.[50] Friendships were formed based on "traditional" friendship considerations. To put it differently, there were no influencers on MyAnimeList.net who would have a strong incentive to have many/add more friends. Related to this, while users post UGC in our data, they post a relatively modest amount: for example, among egos, the 3rd quartile only posts on 1 and the maximum posts on 95 days out of 184 days.

G.2 Friendship Formation

As shown in Figure G-1 for different user cohorts, we observe a decreasing trend in the formation of new friendships over the course of a user's membership. That is, users are most likely to add friends shortly after joining the website and they slow down in their friend making over time after they have already acquired some friends. After the first 12 membership months, a user's friendship network is essentially fixed and new friends are rarely added. Thus, it does not appear in our data that older users (who tend to have more friends since friendship formation takes time) have a high desire to expand their network.

[50] To the best of our knowledge, there still are no monetary incentives.

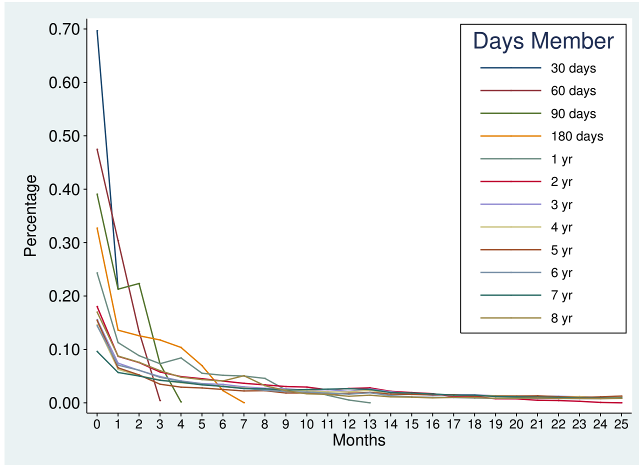


Figure G-1: Portion of the Final Friendship Network Added in Each Month

While Figure G-1 shows the averge *proportion* of newly formed friendships at different points in the platform membership for different cohorts, we also examined the *number* of friendship formations and the amount of heterogeneity in friendship formations across users. Overall, we find that – after the first year of platform membership – users only add very few new friends in absolute terms and that the amount of heterogeneity in friend additions is limited, i.e., there is not a small number of users who form a lot of new friendships. For example, users who joined the platform before 2007 added, on average, about 0.3 friends annually with a median of 0 and a maximum of 11 after 2008. Even if we only consider the 10% of users with the most friends among those who joined before 2007, these users, on average, only added 12 friends annually with a maximum of 23 after 2008.

To summarize, the descriptive data patterns on friendship formation do not suggest that users who already have many friends (are more likely to) form many new friendships.

G.3 UGC Posting

We examined whether users who post a lot of UGC form more friendships. To do so, we looked at three subsamples of egos: those who joined the platform during the first (i) 10 days, (i) 15 days, and (iii) 20 days of our study period. We then conducted a median split of users in each group based on the number of UGC posts during the study period and compared their number of friends at the end of the study period. We do not find a statistically significant difference (at $p < 0.05$) between users who post a lot and post a little UGC for any of the three subsamples of egos. Next, we re-did the same exercise for the three subsamples of ego, but split each group based on the 75th percentile in the number of UGC posts during the study period and compared their number of friends at the end of the study period. We do not find a statistically significant difference (at $p < 0.05$) between users who post a lot and post a little UGC for any of the three subsamples of egos.

To summarize, the descriptive data patterns on UGC posting do not suggest that users who already make a lot of UGC posts (are more likely to) form many new friendships.