Individual (ir)rationality? An empirical analysis of behavior in an emerging social online-network*

Johannes Koenen† Steffen Reik

University of Mannheim University of Mannheim

May 26, 2009

Very preliminary draft, please do not circulate without permission of the authors.

Abstract

The last decade has seen important advances in the theoretical literature on the economics of social networks. More recently, experimental economists have started to examine the quality of predictions derived from the theory in highly controlled surroundings with mixed results. In this article, we look at individual behavior within an actually existing social online-network with additional utility generating functions. We find that theory regarding network formation fails to explain the behavior of users in a significant number of interactions, when they decline to enter into or sever purely beneficial links. Further, we show that in the face of public-good like provision and organization of information users do undersupply these services for higher *levels* of other users' efforts on the one hand, as standard theory would predict. On the other hand, users immediately react to other users' *additional provision* of public goods in a reciprocal manner.

*We are very grateful for helpful comments from the participants of the ZEW Web 2.0 Workshop and the Spring 2009 Conference of the SFB TR 15 in Caputh. The study enjoyed financial support from the DFG.

†corresponding author: Johannes Koenen, Fakultät für VWL und Rechtswissenschaften, Abteilung VWL, Universität Mannheim, 68131 Mannheim; johannes(dot)koenen(at)gmail(dot)com.

1 Introduction

1.1 Motivation

Actions and interactions in social networks involve highly complex decisions. The resulting payoffs can be immense, such as acquiring a job offer that one would otherwise not have received or founding long-lasting mutually beneficial partnerships. Regarding the timing, these benefits or detriments can be immediate or gradual, lasting or short-lived. Their source may be linked to an individual directly or through any number of intermediaries. The last decade has seen major breakthroughs in economic thinking on these subjects. Works such as Jackson and Wolinsky (1996), Bala and Goyal (2000) and Jackson and Watts (2002) have made many of these concepts tractable in a graph-theoretic setup. In these articles, rational agents maximize the payoffs generated from links to others, which in turn allows predictions concerning the structure of the underlying networks. Yet in many social networks one encounters potentially emotionally charged terms, such as "friendships" in social online-networks or "partnerships" in business relations. Consequentially, a more recent strand of the literature therefore has realized, that concepts such as inequality aversion or trust may play an important role in interactions between agents in social networks. Experimental economists have therefore introduced these concepts into simple network models and tested their behavioral predictions under highly controlled settings, providing evidence on why some of the "rational" predictions for equilibrium network structure are not observed in reality as often as expected.[1]

The contrast to these controlled settings highlights some of the challenges that non-experimental empirical work on individual behavior in social networks faces: e.g., it is impossible to disentangle the motives of individual persons or entities entering into relationships, costs and utilities of decisions are mostly unobservable and may differ substantially between individuals, etc. Most importantly, though: There is next to no data that combines micro- (individual decisions and characteristics) and macro-level (structure of the network overall, locations of links inside a network) observations.

Our study aims exactly at this gap, using a completely unique data-set. We were able to obtain data from an emerging online-social network: on the one hand on user behavior of more than 30,000 registered users, for example with regard to provision of a public good type effort, number of private messages sent, number of public comments

[1] See Kosfeld (2003) for a survey of the early literature in this vein.

posted, entry into new and severance of existing friendships, and on the other hand on the structure of the underlying network itself, i.e. number and location of "friendships" between users and overall network size.[2] Using this information, we try to address two issues that could barely be studied empirically so far:

The first is the provision of local public goods by users. Users in the network can provide and organize information as described below in detail, from which both they themselves and their direct friends derive utility. Standard theory, such as Bramoullé and Kranton (2007) in the specific context of a network with information provision, would lead us to expect to find evidence of free riding in the sense that users should provide less effort themselves if they have friends who provide more effort, all else given. Interestingly, we find only weak evidence for this type of free riding. More pronouncedly, users seem to react in a positively reciprocal way to their friends' effort provision - that is they react to their friends' efforts by providing more of their own, a phenomenon that has been noted before in the context of peer-to-peer networks among others by Gu, Huang, Duan, and Whinston (2009).

This leads us to the second issue of our study: Direct negatively reciprocal behavior as an enforcement device. Studying the formation and stability of individual links, we find strong evidence that users are willing to incur costs to actively punish free riding by their friends by severing (otherwise purely beneficial) relationships.

Clearly, the provision of public goods in Web 2.0 services is a highly relevant matter. Microsoft's partial acquisition of Facebook values the latter company at multiple billion dollars[3] - which is somewhat surprising, because apart from the infrastructure, the entire content of the social-networking site as well as most applets are created and provided by users. The same holds true for 3 other purely social networks that together with Facebook form 4 of the 20 most frequented sites on the internet.[4] If one further considers massively popular sites with a strong social-network character such as Youtube.com or flickr, it becomes obvious that social online-networks are encroaching on many people's

[2] As part of the agreement concerning the data, the name of the company and its service is not used in this article. All user-related information was provided to us in a way to completely assure users' confidentiality and protection of their personal data. We were only provided with randomized user-IDs that cannot be linked to individuals in any way and we have no access whatsoever to individuals' personal characteristics.

[3] See, for example http://www.time.com/time/business/article/0,8599,1675658,00.html

[4] The sites are MySpace, FaceBook, Orkut and Hi5.com, according to alexa.com as of June 2008.

daily lives and that a major building block of this development is free user-provided content. In the course of this article, we seek to provide new economic insights into the behavior - and potential motivation and payoffs - of individuals in these kinds of networks.

We proceed as follows: After we introduce the most closely related literature, in **Section 2** we introduce and describe the online-network we wish to study as well as the features that make it especially interesting from an economic standpoint. **Section 3** relates these features to some simple theoretical concepts that will form the basis for our further inquiry. **Section 4** contains the data description and our empirical analysis. **Section 5** concludes and proposes resulting new venues for future research.

1.2 Related Literature

The literature relevant to our questions can be separated into theoretical, experimental and empirical studies.

From the **theoretical** perspective, a set of by now canonical articles has enabled economists to come to term with the complexity of the subject, allowing them to make predictions about the structure of networks and to judge their efficiency. Perhaps the initiators of this development are Jackson and Wolinsky (1996). Their graph theoretic framework, depicting individuals in a network as the nodes in a graph and the connections or links between these individuals as the arcs, has become the de-facto standard approach. Individuals derive utility from the links they are involved in (and indirectly also from the links of those they are connected to), while it is costly to maintain direct links. A network is "stable" if the utility of agents cannot be increased by creating a new or severing an existing link. The predictions concerning stable networks from this model range from complete interconnection (high benefits to costs ratio) over a "star network", in which one node is connected to everyone else, while the peripheral nodes maintain only a single connection each, to a completely disconnected network (lowest benefits to costs ratio). Bala and Goyal (2000) enrich the setup by explicitly modeling the linking-strategies of players in a non-cooperative game and distinguishing between cases in which information (and thereby benefits) flow in only one or both directions. In addition to the stable network structures mentioned above, they find that a symmetric circle or wheel network in which each agent is linked to exactly two other players. Of further importance in their setup is the concept of decay of benefits - the more intermediate links separate two indirectly connected players, the lower the benefits they derive from each others' information.[5]

There are a number of important extensions to these basic models. Goyal (2005) notes that depicting the existence of a link between research affiliates, for example, as a purely binary variable may often be too much of a simplification. Instead, partners can invest specifically into the stability of individual links, so that heterogenous link strengths may arise in equilibrium. Bloch and Dutta (2009) adopt this concept and model it explicitly. They find that under relatively broad conditions, the stable and efficient network again will have a star structure.

Three papers of which we are aware of try to model the provision of public goods in social networks. In Cho (2006) ex-ante symmetric agents agree to a binding contract that covers both the amount of a local public good each agent provides as well as the form of the network through which the benefits of the public good are transmitted. In this setup, equilibria exist, in which agents almost surely provide efficient levels of effort if they are patient enough. In Bramoullé and Kranton (2007) agents only specify their effort levels, taking the structure of the network as given. Effort provided by an agent's neighbors is a substitute for her own efforts. They show that under these assumptions free riding exists in every social network in equilibrium and adding links between players reduces individual incentives to contribute. In the model of O'Dea (2008), depending on the outcome, one can find both equilibria in which agents voluntarily provide public goods and equilibria which resemble cost sharing for an excludable club good.

These clear and distinct prediction concerning network formation and structure have piqued the interest of a number of **experimental** economists, who find various deviations from the rational depictions. Falk and Kosfeld (2003) directly test the hypotheses developed in the Bala and Goyal (2000) model and show that in settings in which the Nash-equilibrium provides for symmetric payoffs and little coordination efforts, the predictions are quite accurate, while in settings in which the predictions would be an asymmetric payoff generating star network, rational equilibrium behavior is rather unlikely. They show that this is only to a small extent explained by the higher coordination requirements and to a larger extent explained by individuals' aversion to unequal payoffs. Cagno and Sciubba (2007) attempt to dissect the various drivers of individual decision

[5]For an excellent survey of the theoretical literature on network development and stability, see Jackson (2004).

in network formation. They identify two conflicting driving forces: While optimizing best-response behavior explains much of the behavior, they find that reciprocity and inertia play an important explanatory role as well. Corbae and Duffy (2008) also test the rational behavior hypothesis in an experimental setting, in which individuals can form risk-sharing partnerships through a non-cooperative proposal game with idiosyncratic risk. They find that individual behavior is generally well predicted by rational theory, which agent achieving near-efficient outcomes. Goeree, Riedl, and Ule (2008) analyze a setting in which agents with heterogeneous linking costs can form links under different informational settings. While for homogenous agents equilibrium predictions cannot be validated, the heterogeneous agent predictions are well suited to explain individual linking behavior. Interestingly, star networks do develop in this experimental setting as frequently as expected. Finally, in a highly surprising study, Cagno and Sciubba (2008) combine a network formation with a trust experiment. They find that the levels of trust in a trust game played after an endogenous network formation game are significantly lower than in the simple trust game. Subjects hereby tend to trust players less on average with whom they have interacted previously than they trust complete strangers. The authors deduce from this that potentially existing reciprocal behavior in the network formation game does not foster trust.

There is a large number of **empirical** studies regarding networks in general that mainly focus on locating and quantifying network effects, such as prominently Saloner and Shepard (1995) or Gowrisankaran and Stavins (2004). For a survey of this strain of the literature, we refer the reader to Birke (2008). There is a limited amount of empirical economic research on social networks, though. Prompted by Newman (2001) who looks at the properties of research collaborations from a physicist's perspective, Goyal, Leij, and Moraga-Gonzlez (2006) and Rosenblat and Mobius (2004) both analyze the structure of research networks in economics. Goyal, Leij, and Moraga-Gonzlez (2006) find a collaboration structure composed of a number of stars (authors that write a relatively large number of papers together with coauthors) and a large periphery (composed of authors who only published papers together with one of the star-authors), which suits the structural predictions of theoretical models. Based on a model that they develop, Rosenblat and Mobius (2004) look at the rise of the internet and electronic communications as a kind of natural experiment which lowered the costs of communication between potential collaborators. They found that the degrees of separation between individuals on average decrease and each author is more likely to coauthor a paper with a distant author from a similar field. On the other hand, the likelihood of coauthoring a paper with a coauthor from a dissimilar field significantly decreases. Further, Kossinets and Watts (2006) use the email-communications inside a large US college to draw up a social network structure and discover various characteristics that determine the likelihood that two individuals within the college will be linked directly.

A wave of more recent yet to be published papers proceeds in a direction that is almost a hybrid between economic and information systems research - in their subject, these are most closely related to our paper, even though they strongly differ in methodology and focus.

Kumar, Novak, and Tomkins (2006) try to characterize user types in two huge social online-networks, Flickr and Yahoo!360. They characterize 3 kinds of users: passive members, inviters who entice others to join, and "linkers" who make full use of the social networking capabilities of the services. Gu, Huang, Duan, and Whinston (2007) look at at peer-to-peer IRC music sharing service and show that it resembles a two-sided market in many ways, in which contribution and consumption are highly complementary.

Three studies specifically scrutinize public good problems and free riding: Asvanund, Clay, Krishnan, and Smith (2004) try to quantify the extent of free riding in a peer-to-peer file sharing network (Gnutella) on a macroscopic level. They find that a substantial share of users (42%) behave as free riders, i.e. access others' files without providing any of their own. Xia, Huang, Duan, and Whinston (2007) discover that users are more likely to share files in online sharing communities if he has himself benefited from the community or has a recognized social status. Finally Gu, Huang, Duan, and Whinston (2009) consider the effects of indirect reciprocity on the public good provision behavior of members of a peer-to-peer file sharing network. They find that the propensity of an individual user to free-ride depends on the prevalent behavior of other users in the community, which they construe as evidence for an indirect reciprocal behavior. Further they find evidence for reciprocity as a social norm, as free riders are discriminated against through voluntarily enforced settings on communal servers.

Against this extensive background, our contribution is the following: We make use of a unique data-set in which we observe both individual characteristics and behavior as well as the link structure among all individuals within an actually existing social online-network. We provide evidence in this context on whether network formation and linking decisions are motivated by strategic behavior which complements the existing experimental literature. Then we focus on public (or club-) good provision within the network. Looking at the entire linking and continuation behavior over approximately four months, we show that negative reciprocity, i.e. unilaterally severing existing friendships, even though they are purely beneficial, is used to punish free-riding behavior. Finally, we look at the overall effects of this. Using a user-fixed effects panel estimation procedure, we show that there is only very weak evidence for the existence of free riding behavior, despite an almost textbook public good problem. On the contrary, we are able to identify signs of positively reciprocal behavior.

2 The Social Online-Network under Scrutiny

2.1 Function and Features of the Network

In this article, we study an emerging European social online-network, which we will refer to from now on as the Network. Various features make the Network especially interesting for economic analysis. The basic infrastructure is very similar to well-known sites. Each user has an individual profile page which is accessible to others and provides information about himself. As a further feature familiar from other networks, individuals can become "friends" with other users - for this to occur, one of the users has to propose a friendship to another and the second user has to accept this proposal. Note that for legal reasons, the number of friendships a user is allowed to enter into is capped at 150.

The main differentiating feature of the Network is that users can upload digital copies of music that they legally own to an online music library. We call this activity *provision of information*. Via a player embedded in their web-browser, users can listen to (but not download) music that they either own themselves or that was uploaded by their direct, i.e. first-degree, friends. A user therefore has access to all the music that either he or one of his direct friends has uploaded.

A feature of the Network that is something of a nuisance to users is a blessing to our study: The access to individual music files is somewhat complicated. For one, the music available to a given user is distributed into different music libraries (e.g. music uploaded by himself is one library, but music uploaded by each friend is stored in an individual separate library). The music files in each of these libraries are sorted alphabetically, which makes it potentially time-consuming to find specific songs in large libraries or to find something interesting while browsing through them. While the service provides users with a search function, searches always cover the library of the entire Network, i.e. both files that a user can access and files that she cannot. From the search results, a user can only establish whether she has access to a given file by adding it to her player and waiting for an error message to appear.

This is where the second kind of effort provision comes into play. This inconvenience is eased by a further option for users. They can create play-lists by adding multiple songs to the embedded music player. Any number of songs can be added from the libraries that a user has full access to. These play-lists are temporary in principle, i.e. they are deleted whenever the cache is cleared. But a user can also save and give a name to a play-list that she has compiled, which stores the list with the Network and makes it available to other users. Clearly, this will be most useful to the user's friends, who can access many of the same songs as her. For these reasons, we call the creation and storing of play-lists *organization of information*.

A user can prevent one of his friends from accessing music-files only by ending their friendship. Therefore provision and (indirectly) organization of information have the characteristics of club goods.

2.2 Network Size and Development

The web-site came online in December 2007 with 2,000 registered users. This number grew to 9,000 in March 2008, more than 20,000 in September 2008 and exceeded 30,000 in November 2008, when the data for this study were mainly collected. For many - maybe even most - questions of interest concerning online networks, the number of registered users is not a very good indicator, as it may include individuals who only registered at some point of time but never actually used the service or have ceased to do so at some point in the past; therefore we will briefly discuss the most widely used alternative measure.

According to the common practice of sites such as facebook, myspace or similar sites such as the German studi-vz, a user is considered active if she has logged into her account in the course of the past 30 days. At first glance, it is not obvious whether and how this measure is meaningful. **Table 1** sheds some light on this question. In it, we have gathered descriptive statistics for various such cutoff levels of user activity as of October

15th 2008.[6]

	All Users	Act30	Act20	Act10
Friends	0.92	2.16	2.50	3.91
Songs Listened	35.57	100.82	121.50	208.63
Songs uploaded	28.62	50.23	56.93	96.88
Play-lists created	1.63	2.55	2.80	3.80
Messages sent	0.38	1.30	1.67	3.02
Comments left	0.03	0.12	0.16	0.28
Days logged in	2.92	6.03	7.01	10.57
#Observations:	31,455	6,013	4,466	2,291

Table 1: Descriptive statistics for various activity measures

Some important observations can be gleaned from these numbers, both about the activity measures in general and about the Network in particular. First of all, each more restrictive sub-sample includes users that are on average highly significantly more active than the users in the more inclusive sample. For example, users that have been active in the last 30 days (Act30) upload almost twice as many songs on average as the general population, i.e. those users that are left out of the sample are significantly less active. Note that one loses a large number of observations in the process, from all registered users to Act30, the drop is precipitous, from 31,455 to 6,013. These findings indicate that the frequently employed measure does have some merit. **Figure 1** depicts the development of registered vs. active users in the period from December 2007 to November 2008:

Two findings from Table 1 concerning the Network are particularly important for our study. On the one hand, note that "social" activities, such as posting comments or sending messages to other users appear to play a very minor role in the network, with even the most active users having sent no more than a total of 3.02 messages and

[6]Act30 designates user who have logged into their accounts in the last 30 days, the other columns have analogous interpretations. By line, the sample means of the following are reported: number of friends per user, number of songs listened to per user, number of songs uploaded per user, number of play-lists created per user, number of messages sent per user, number of public comments left per user, number of days on which the user has logged into her account at least once per user. All respective sample means differ at 1% significance level (t-tests).

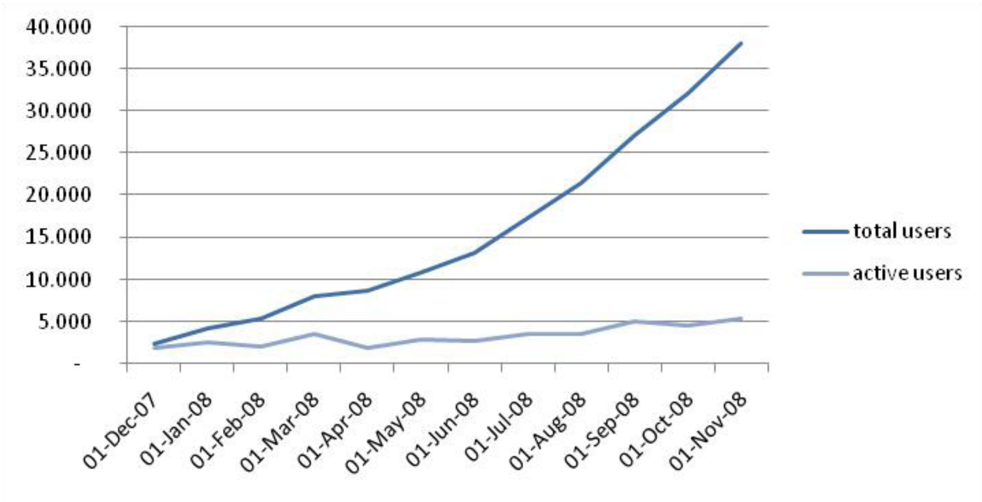

Figure 1: Development of Act30 Users over time.

posting no more than 0.28 comments on average. On the other hand, listening to and uploading songs take up a much larger share of users' activities, which points in the direction that these music-consumption related activities are the main source of utility for those involved in the Network.

3 Analytical Framework

The aim of this article is to provide some evidence for strategic and utility maximizing behavior among users in the Network. In order to facilitate the understanding of readers and help us formulate our hypotheses more precisely, we will sketch a simple formal model of link formation and individual behavior within the Network. This will form the basic framework for our empirical analysis. We hereby draw on the notation and concepts introduced by Bala and Goyal (2000), Jackson and Wolinsky (1996) and Bramoullé and Kranton (2007). As we do not intend to explain or analyze the equilibrium network structure, but individuals' decision behavior instead, we will keep the complexity of the exposition to the absolute minimum.

Following the by now established graph-theoretic approach, we think of the Network as a set of nodes $N = \{1, ..., n\}$. In our context, each individual agent (user) is represented by one node. Further, for each agent i, there is a vector $g_i = \{g_{i,1}, ..., g_{i,i-1}, g_{i,i+1}, ..., g_{i,n}\}$, such that each $g_{i,j} \in \{0, 1\}$ indicates whether or not a direct link exists between nodes i and j. These links can be interpreted as friendships between users in our context. A network g therefore is defined by the set of vectors g_i, $i \in N$. Note at this point, that all friendships have the same strength in this framework, as they are depicted as purely binary relationships.

3.1 Creation and Severance of Individual Links

Assume that each link g_i, j between two users or players i and j is undirected, i.e. both can benefit. To form a new link, both players have to consent to the formation. One player (the "sender") initiates the link formation process through an invitation while the other (the "receiver") decides whether to accept, ignore or decline the invitation. If the receiver accepts the invitation, a new link is formed and $g_{i,j}$ becomes 1. We denote this change in the network structure by $g' = g + g_{i,j}$. If the receiver declines or ignores the invitation, the structure of the network does not change.

Conversely, each user inside a friendship can sever an existing link unilaterally. If user i or user j cancels their friendship, then $g_{i,j}$ becomes zero. Analogously to above, we denote this change in the network structure as $g' = g - g_{i,j}$.

The utility that a user obtains from the network depends on its structure, therefore $u_i = u_i(g)$. It appears sensible to assume that the formation of a new friendship is associated with costs, which may vary with the characteristics of the individuals involved: The sender must first locate the receiver, then click on a button and compose a message to send out a friendship invitation, and finally, he looses one free slot among his friendships if the receiver accepts the invitation. Note the difference: Once a friendship offer has been extended, the search costs are sunk and only the costs of the free slot remain. The receiver has to click on one button, whether he accepts or rejects a friendship (she has to take no action to ignore the invitation) and she looses one free friendship slot if she accepts. As the formation of new friendships requires mutual consent, the following two conditions have to hold at the same time for a new friendship to be created, assuming that i is the sender and j is the receiver, and denoting their respective costs of inviting/accepting as $\kappa_{i/j}$:

$$u_i(g + g_{i,j}) - \kappa_i > u_i(g) \Leftrightarrow \Delta u_i > \kappa_i \tag{1}$$

$$u_j(g + g_{i,j}) - \kappa_j > u_j(g) \Leftrightarrow \Delta u_j > \kappa_j \tag{2}$$

The way the Network is set up, additional links are close to purely beneficial. They allow users to access more music and to make better use of additional play-lists. As indicated above, users in the Network are mainly interested in music consumption, therefore these considerations should be their main priority. On the other hand, each additional friendship costs next to nothing as long as users are not close to the upper limit of friends.

Analogously, for user i to be willing to sever an existing link, denoting the costs of severance as λ_i the following condition must hold:

$$u_i(g - g_{i,j}) - \lambda_i > u_i(g) \Leftrightarrow \Delta u_i > \lambda_i \tag{3}$$

For the same reasons as stated above, we would expect Δu_i to be negative in the case of link separation, unless the number of friends is 150 (or close to the upper limit and users anticipate reaching it with certainty). As the costs of link separation λ_i are non-negative, in this simple setting users should be expected not to separate links unless they are close to the upper bound of friends allowed.

3.2 Provision of Club Goods in Network

One of the main features of the Network is that users provide different kinds of effort, uploading music or compiling play-lists, that are potentially useful both to themselves and to their friends. We treat this in the easiest imaginable form: Consider that user i can exert effort e_i at cost $c(e_i)$. We denote the set of user i's first degree friends as J_i, each of whom also exerts effort. Expanding the utility function of user i above to incorporate these considerations, it now becomes $u_i(g, e_i) = u_i(e_i, \{e_j\}_{j \in J_i}) - c(e_i)$. Clearly, the optimal level of effort of i is determined by

$$\frac{\partial u_i(e_i, \{e_j\}_{j \in J_i})}{\partial e_i} = c'(e_i) \tag{4}$$

Therefore, if the effort of user i and his friend user j are substitutes (i.e. the cross-partials are negative) and the costs of effort provision are non-negative, the optimal effort of i should decrease with the effort that his first degree friends provide. If in addition the utility of user i increases in the effort of his friend user j, i.e. $\frac{\partial u_i(e_i, \{e_j\}_{j \in J_i})}{\partial e_j} > 0$, we can talk of a local public good. [7] It would be straight-forward to derive the expectation of

[7] Note that in the Network, we can distinguish two dimensions of effort, provision and organization of information, which are complements.

free-riding by users from this. In a similar setup, Bramoullé and Kranton (2007) find a stable equilibrium with complete specialization, i.e. the agents can be cleanly separated into two groups, one of which provides effort, the other of which free rides completely.

There are two straightforward reasons why this should not be so clear cut in the Network we are considering. For one, users do not take the link-structure as given, but can instead strategically form friendships and sever existing links. Therefore free-riding can be punished by severing links with free riders (negative reciprocity) on the one hand and users may only enter into new friendships with others who have already demonstrated that they are good citizens (strategic link formation). The second reason not to expect free riding is the fact that many of the peer-groups within the network are relatively small, with even the most active group of users only having somewhat less than 4 friends on average. In such a close (if still potentially anonymous) setting, one may expect that social forces such as shame or guilt, combined with the relative ease of monitoring a small number of connections, may prevent users from engaging in free-riding behavior. Experimental evidence shows, though, that even in small groups and repeated settings, the play in repeated public-good games converges neither to a pure free-riding equilibrium (the rational equilibrium), nor to purely cooperative behavior (the efficient outcome), but instead behavior is mixed, with play converging towards a situation in between. See for example the seminal contribution of Andreoni (1988).

From these simple assumptions and considerations, we derive the following hypotheses:

Hypothesis 1a - Strategic Linking Behavior regarding Invitations: Users that are invited to join a friendship have provided higher efforts on average than the peer-group of active users.

Hypothesis 1b - Strategic Linking Behavior regarding Acceptance: For friendship invitations that are accepted, the inviter has provided higher efforts on average than her peer-group.

Note that for the formation of new friendships, both users have to agree, therefore the behavior and characteristics of both parties are relevant for the formation. As opposed to this, for the severance of friendships it is sufficient for one of the users to make the decision. This is reflected in the following hypothesis, which will be more closely specified further below:

Hypothesis 2 - Negative Reciprocity: Friendships from which one of the users derives little utility are more likely to be severed.

The hypotheses concerning free riding behavior are very straightforward:

Hypothesis 3a - Free Riding for Information Provision: Users upload fewer songs themselves, the more songs are available to them through their friends.

Hypothesis 3b - Free Riding for Information Organization: Users generate fewer play-lists themselves, the more play-lists are available to them through their friends.

Finally, we account for the mirror-behavior to negative reciprocity: Users are informed by the system on their welcoming-screen whenever one of their friends have provided new music or play-lists. Seeing this provision of effort, they may feel induced to reply in kind, i.e. provide additional effort of their own.

Hypothesis 4a - Positive Reciprocity for Information Provision: Users respond in kind whenever their friends provide additional information/upload new songs, i.e. users are more likely to upload new songs if their friends have recently uploaded new songs.

Hypothesis 4b - Positive Reciprocity for Information Organization: Users respond in kind whenever their friends provide additional organization of information information/create new play-lists, i.e. users are more likely to compile new play-lists if their friends have recently generated new play-lists.

4 Empirical Analysis

4.1 Data Description

We use two different data sources to approach the questions that we wish to study. For the questions related to link formation and severance, we mainly use data covering every change of user friendship status between July 10th and November 11th 2008. From this we create two different data-sets. To test Hypotheses 1a and 1b, we look at all friendship invitations that were issued during this time, as well as what happened to them (acceptance, declination, revocation, nothing). We observe 3,657 friendship invitations in this time span. Importantly for our topic of interest, we also observe certain user characteristics for those involved in these exchanges. In order to observe the probability for a friendship being severed for Hypothesis 2, we limit our analysis to those invitations that were accepted in this time span and which we observe for at least 50 days. This limits our number of observations to 1,521 observations. We call these data sets A.

For the remaining hypotheses on free riding and positive reciprocity, we make use of a panel data set. We currently have 10 weekly cross-sections of user data in which we observe various characteristics and behavior variables as described below. The data cover the period from September 14th until November 18th 2008. Due to attrition and new arrivals, this results in an unbalanced panel with ca. 25,000 users and an average of 8.1 observations per user. We will refer to these data as data set B.

In the following, we briefly describe the variables used in our study and indicate in which of the data sets they are available to us.

Days Online (A,B) - This count variable logs each day that a user logged into her account or used one of the functions on the Network. The value of this variable is the aggregated number of days that a user has visited the site. Note that this variable does not measure how many actions a user has performed on a given day or how much time he has spent using the service.

Days since Registration (A,B) - This variable describes how many days ago a user registered with the service.

Last Login (A,B) - Measured in days, this variable denotes the time since the last activity of the user. An activity is a login or any other action within the user interface.

Friends (A,B) - The number of currently active friendships a user is involved in. There is a limit of 150 friendships that a user can enter into, but this limit was binding for only one user at the time of our analysis.[8]

Music Uploaded (A,B) - This variable states the number of uploaded songs by a specific user. These songs are collected in the users own music library, which is accessible for the user's friends.

[8]Due to a technical glitch in the system, this user was actually able to enter into more than 150 friendships.

Songs Listened (A,B) - This is the total number of songs the user has listened to on the Network. A limitation to this variable is that it includes both songs that were skipped as well as songs that were listened to in full length. Nevertheless, assuming that users derive utility from listening to music through the server, this is one of the better proxies for user utility available to us.

Own Play-lists (A,B) - The number of play-lists that a user has created and that exist at the current point of time, i.e. if a user deletes a play-list, this variable declines by one.

Play-lists Friends (B) - The total number of current play-lists of all of a user's friends.

Common Friends (A) - All users that are friends of both the sender and the receiver of a friendship invitation at the given point of time are counted in this variable, illustrating the overlap of senders' and receivers' friends.

Common Music (A) - This variable counts songs that are owned by both sender and receiver. Important to note is that this variable is not able to completely capture the overlap of the both music libraries, since some songs may vary in quality or length and are thus tagged with different internal IDs by the system.

Songs from Friend's library (A) - This counts the number of songs from the other users library that a user has actually listened to.

4.2 Strategic Linking Behavior and Negative Reciprocity

One criticism that one could easily point at our project is that users do not act out of rational objectives, i.e. that they interact with people they consider friends totally irrespective of their effort provision. The aim of the first hypothesis is to if not repudiate this argument then at least to put it into perspective. **Table 2** shows the average characteristics of four different groups: First, those who received friendship invitations throughout the period we observe. We partition the senders of these invitations into two subgroups, those whose friendship was accepted and those whose friendship was declined or left pending for more than 30 days. [9]. We compare the characteristics of these groups to the group of those that have logged in within the past 10 days, the most restrictive "conventional" activity measure.

The results are quite striking. People who have received an invitation are on average by far the most active group. They have uploaded far more music (3 times the average of the second most active group), provided more play-lists and are also more active concerning the social network functions. Both mean- and median-equality tests are rejected at the 1% significance level for each of these variables. Therefore we cannot reject **Hypothesis 1a** and consider the evidence in support of the hypothesis to be strong.

	Invitees	Inviters Accepted	Inviters Not Accepted[a]	Act10 users[b]
Own music	1936.02	667.30	316.28	368.71
	(5416.59)	(1879.08)	(1326.79)	(1299.06)
Own play-lists	40.65	20.95	7.79	11.12
	(126.97)	(90.41)	(37.37)	(47.00)
Friends	38.07	21.5	20.61	14.48
	(39.76)	(27.50)	(26.69)	(27.93)
Messages sent	36.41	10.40	5.61	11.47
	(150.83)	(62.61)	(33.44)	(88.03)
Comments left	2.82	1.17	.59	1.06
	(7.82)	(4.16)	(2.88)	(4.78)
obs.	3657	2,076	930	602

[a] An invitation was considered non-accepted if it was outright declined or if it was left pending for more than 30 days.

[b] Here, we exclude those users that have registered less than 10 days ago, which explains the differences to table 1.

Table 2: The average characteristics of invitees, accepted and non-accepted inviters and the most active group of the Network's users according to the most widely used definition, (means reported, std. deviations in parentheses)

[9] The date was chosen as more than 95% of acceptances were issued within 28 days, but the results are highly robust to moving the cutoff date.

Similar arguments hold with respect to **Hypothesis 1b**. Those users whose invitation was accepted, have provided twice as much music as their rejected (or ignored) counterparts, almost three times as many play-lists and have sent twice as many messages or comments on average. Again, both mean- and median-tests show that the differences are significant at the 1% level. So the evidence is strongly in favor of the hypothesis. For the more general point, we also conclude that users do behave in a way that appears to be driven by rational (utility maximizing) motives.

Next, we focus on the unilateral decision to sever existing friendships. We distinguish between two parties, the person who severed a friendship ("Severer"), and the person who passively had to endure the severance ("Severed"). Even a cursory glance at the descriptives displays a very clear pattern. In **Table 3**, we present the average characteristics of a Severer, a Severed and an "average" user in the sub-sample. Comparing the latter to the other measures of user characteristics presented above shows us, that the new friendships we observe were formed between relatively active and involved users.

The comparison between Severer and Severed is very stark: From a static viewpoint, the Severer had provided 22-times the music (4472.15 vs. 206.14) and 48-times the number of play-lists (358.62 vs. 8.47) at the time that the friendship was formed. Looking at their individual behavior during the time that the friendship lasted, the Severer uploaded 110-times as many songs per day (33.71 vs. 0.28) and 12-times as many play-lists (0.65 vs. 0.051) as the Severed. Finally, the Severer was far less likely to listen to a song of the Severed than vice-versa (0.02 vs. 1.29 songs of the other listened to per day). The small, yet positive, number of songs that the Severer listens to from the Severed's music library is a reminder that the decision to end the friendship is costly: You lose access to those songs that the other person has provided. In a static world, which takes the utility derived from a given friendship as given, this appears not to be rational, unless a user derives an immediate satisfaction from link severance ("revenge" for example). On the other hand, it is relatively easy to imagine a dynamic game, in which an equilibrium exists, in which punishment prevents free riding. Both concepts have in common that free riding is the cause of dis-utility to the other party, therefore we modify Hypothesis 2 to account for this observation:

Hypothesis 2' - Negative Reciprocity: Friendships in which one of the users displays behavior resembling free-riding are more likely to be severed.

Importantly for our analysis, no user reached the maximum number of friends (150) during the time that we observe, the maximum number of friends that was reached in the severed friendships was 125.

	Severer	Severed	Average User [a]
Own music beginning	4472.15	206.14	885.84
	(3111.36)	(987.39)	(1335.54)
Average songs added per day	33.71	0.28	4.08
	(34.44)	(1.03)	(11.05)
Own play-lists beginning	358.62	8.47	31.52
	(317.54)	(36.25)	(71.19)
Average play-lists added per day	0.65	0.051	0.29
	(1.80)	(0.11)	(0.84)
Friends beginning	12.82	15.61	28.96
	(29.15)	(22.90)	(22.68)
Friends end[b]	22.27	17.63	34.48
	(35.36)	(23.24)	(24.34)
Number of friend's songs listened per day	.020536	1.29	0.15
	(0.13)	(3.65)	(0.74)
obs.	114	114	3,042

[a] Average of users in the entire sample of new friendships being formed.

[b] The maximum number of friends observed among severed friendships was 125, well below the limit of 150. For non-severed friendships, we report the number after 50 days.

Table 3: Mean characteristics of the severer's of friendships compared to those who were severed and the average of all users involved in "new" friendships.)

From the descriptive statistics alone, it is hard to disentangle whether it is the sheer amount of effort provided by a user or the additional provision of effort over time that the other party values. In order to shed more light on this question, we consider the following simple discrete decision model. Let us call the probability that a given friendship is severed $p(x) \equiv P(y = 1|x)$, where x is a vector that captures both parties' characteristics. Using the latent variable approach, we estimate simple logit-models, see **Tables 4 and 5** in the Appendix for the exact results.

In model 1a, we only include the "static" characteristics, i.e. the effort provided by both users at the beginning of their friendship (here again, we talk of a severer and a severed user - if the friendship was not severed, then these two roles were assigned to the users in a friendship randomly). We would expect that the effort provided by the Severed should lower the probability that the friendship will be ended, as it raises the costs of the Severer (more information that he can no longer access), and the effect does have the expected sign and is (though weakly) significant for the provision of music, while the provision of play-lists is not significant. Interestingly, the effort provision of the Severer is significant and increases the probability of severance along both effort dimensions - this means that users that have contributed a lot of effort themselves are more likely to end a friendship (while they are also more likely to encounter free riders, according to the analytical framework presented above).

In model 2a, we only include the "dynamic" characteristics, i.e. the number of songs and play-lists that the individual users have added over time on average. We encounter the same pattern: The more music the Severer adds, the more likely he is to end a friendship, while adding more music makes it less likely that a user will have his friendship ended by the other party. The effect of play-lists added by the Severer is more interesting. Adding more play-lists makes it much *less* likely that a user will end a friendship. On second glance, this is rather intuitive: When adding a play-list a user organizes the information available to herself for easier access. This may (or may not) include information provided by the other user. In any case, this makes the information more valuable (as it is more accessible) and therefore raises the costs of ending a friendship.

In model 3a, we include both the static and dynamic characteristics. Encouragingly, the effects retain their signs and significance levels.

In models 1b and 2b, we add those variables that capture free riding in the potentially most intuitively appealing way, by including the number of songs of the other person that the respective users have listened to per day. Model 1b only includes these two variables and, as one would expect, songs listened by the severer decrease the probability that a friendship is severed. Interestingly, though, the number of songs listened per day by the severed highly significantly increase the probability that a friendship is severed. In mode 2b we add all the explanatory variables used in the previous models and the effect that this intuitive form of free-riding strongly increases the probability of a friendship being severed persists.

To conclude the section, let us briefly summarize our findings. There are strong indications that users make their link-formation decision contingent on the utility they expect to derive from a given friendship. Invitations are sent out to very active individuals, and invitations by users who have provided little effort themselves are far more likely to be rejected. Regarding the severance of links, we find that users are willing to incur severance costs (not being able to access the other's information) in ca. 7.5% of friendships in the interval of time that we observe. Patterns and regularities, as well as our logit regression results, suggest that this behavior is caused by the desire to punish the other users' free riding. This corresponds with experimental findings especially in the literature on the ultimatum game. Interestingly, this punishment behavior can itself be interpreted as a form of public-good provision, e.g. enforcement of social norms.

What remains now is to show how prevalent free-riding is in this community, given the existence of control-mechanisms on an individual friendship level.

4.3 Free Riding and Positive Reciprocity

As discussed above, we consider a user to be free-riding if she provides less effort ceteris paribus given that her friends have exerted themselves more. A naive approach to this problem therefore would be to simply regress all users levels of uploaded songs and generated play-lists onto the levels of their friends in a cross-sectional approach. At second glance, though, this approach is clearly not conductive to answering the question due to a reverse causality issue. Assume that one finds that people whose friends provide a lot of effort provide a lot of effort themselves. One can clearly imagine a story in which people who have provided a lot of effort in the past attract more friends, especially considering our findings on strategic linking behavior.

To circumvent this issue, we therefore don't consider the total levels of effort that users provide as a dependent variable. Instead, we look at their weekly efforts, i.e. how many songs/play-lists they add in a given week given the explanatory variables and different sets of controls, which gives us the following regression equations:

$$e_{it} = \beta_1 X_{it} + \beta_2 Z_{it} + u_{it} \tag{5}$$

where e_{it} is the effort an individual exerts, X_{it} describes the explanatory variables and Z_{it} designates the controls. Clearly, there may be unobserved individual effects that are correlated with the individual error terms, e.g. some individuals may be compulsive collectors of music, while others are casual listeners only. In order to capture these differences, we estimate a user fixed-effects model, and adjust the standard errors for potential user-cluster effects.

We estimate three different models (with two different samples) for information provision and organization respectively, in order to shed some light on hypotheses 3 and 4. In each specification, X is composed of three variables:

The determinants for the effort that a user provides for **information provision** in a given week should be determined by a) the total amount of music that his friends have made available to him, b) the total amount of music that he himself has uploaded in the past and c) the amount of music that his friends have added in the recent past. For c), we use the additional music friends have uploaded in the course of the past week - if anything, this should under-estimate the effect we are trying to find. Relating these variables to our hypotheses we would expect the following: a) should have a negative influence on effort provision (free riding, hypothesis 3a), b) should have a negative effect on effort provision due to decreasing marginal utility from effort and c should have a positive influence on effort provision (positive reciprocity, hypothesis 4a).

The three models differ by the controls that we add. Model 1 is bare, i.e. without additional controls. In model 2 we add a control for the user's time since registration by adding a dummy that takes the value 1 if the user has been registered for longer than 60 days (experienced dummy). Further we add a control for time. In order to control for users whose main motivation to join the network is mainly social, we add controls for the amount of public comments that users have left and private messages they have written. In model 3 we give credit to the consideration that the two kinds of efforts are complementary. One would expect users to add more music if their existing music is well organized and vice versa. Therefore in model 3 we add controls for the number of own play-lists and number of friends' play-lists available to a user. We estimate each of the models once for the entire population of registered users and once for the sub-sample of active users only.

Analogously, for **information organization** efforts in a given week, the explanatory variables in X are a) the total number of play-lists that a user's friends have provided, b) the total number of play-lists that he himself has compiled and c) the number of play-lists that his friends have added in the past week. From our hypotheses, we would expect the same signs as above. Again, model 1 is bare, model 2 adds controls for time and social interactions and model 3 takes into account the amount of music that the user and her friends have respectively uploaded.

For the detailed regression results, please refer to tables 2 through 5 in the Appendix. We find very weak evidence for hypothesis 3a, free riding with regard to information provision. Users upload weakly significantly less music only for the sub-sample of active users in model 3, i.e. taking the effects of the complementary effort provision into account. On the other hand, the evidence for positive reciprocity with regard to information provision (Hypothesis 4a) is comparatively robust. In each specification of the model there is a small but significant and positive reaction to other users' providing more music. Some other observations for the regressions regarding music files added are highly interesting. For one, the level effect of the own music uploaded is never significant - this could either be due to the fact that the marginal utility from additional music is non-decreasing, i.e. now matter how much music is already available, more music is yet better. Or there is so much music available freely in any case, that additional own music files from the beginning on have little value to users.

The regressions for the additional provision of play-lists (tables 4 and 5) on the other hand resemble our assumptions more closely. There is a significant negative level effect for the amount of own play-lists already compiled, which resembles initially positive and decreasing marginal utility of organization of information. Model 3 shows that the more music a user himself provides, the more likely he or she is to exert more effort in organization. The effect of friends' music on the other hand is surprisingly significant and negative (if economically not very meaningful). Again, this might resemble the fact that more music in questionably accessible libraries may actually reduce the utility a user derives from the service because it makes things harder to find. Regarding our central hypotheses, we find robust evidence for hypothesis 4b, users react to their friends' organization of information in a positively reciprocal manner. Again, the evidence for free riding is limited to one of the sub-samples and only visible in model 3.

5 Conclusion and Outlook

In this article, we attempted to empirically assess the rationality of user behavior, analyzing unique data from a real social online-network. The major advantage of this data is that we observe both individual user characteristics as well as the structure of the entire network, which allows us to study issues that were not easily approachable up until now empirically.

We find that despite mixed experimental evidence, the existing theory on link-formation allows straightforward and useful predictions: Users systematically pursue more valuable "friends" in the process of link formation. On the other hand, individuals are willing to sever at first glance purely beneficial links in 7.5% of the friendships that we observe. There is substantial evidence that this is linked to punishment-behavior when users encounter free riding, i.e. a direct form of negative reciprocity that in the past has been most prominently encountered in ultimatum games. Viewed statically, this kind of behavior would not be predicted by network theory (or economic theory in general). Interestingly, this can in itself be construed as a form of public good provision, as it may help to enforce social norms, from which users in general benefit. Finally, we focused on whether free-riding is detectable, or whether on the contrary we observe a form of positively reciprocal behavior, in which individual users respond to their friend's additional effort provision in kind. There is clear evidence for the latter, while free-riding behavior appears to be close to non-existent.

We believe that our research raises some interesting theoretical questions. Evidentially, there are interactive effects between the formation of links, the provision of club goods within a network and unilateral link-severance, which may influence the prevalence of free-riding within a social network. These effects have, to our knowledge, not been captured in theoretical network models up until now, even though they may yield predictions that are far closer to the empirically observed facts. In addition to the individual behavior, it would be interesting to generate predictions concerning network formation and equilibrium network structure in an in this sense richer context.

For the specific network, there are a number of issues that we want to address in future work with additional data. The questions include, but are not limited to: Is the punishment mechanism that we observe effective, i.e. do individual users, who have been "kicked out" of friendships change their behavior? What are the effects of the size of neighborhoods (or the number of friends of an individual user) on his free-riding behavior - e.g. are members of larger neighborhoods more likely to exhibit this kind of behavior, as potentially monitoring costs increase and punishment is less likely? Finally, it would be wonderful to see whether our findings hold true for comparable social networks, if similar data becomes available in the future.

In closing, we would like to restate that we are well aware of the drawbacks that empirical work in this field suffers from in general. We do not observe and cannot perfectly control for individual users' actual motivation in their actions. There may be substantial and unpredictable differences in their utility functions, "network savvy" and computer literacy, which may systematically bias our results. While experimental are not nearly as prone to these problems, they are subject to certain shortcomings of their own. We hope that our study is viewed as a strongly complementary building block that enriches our understanding of the behavior of individuals in complex social network settings.

References

Andreoni, J. (1988): "Why Free Ride?," *Journal of Public Economics*, 37, 291–304.

Asvanund, A., K. Clay, R. Krishnan, and M. D. Smith (2004): "An Empirical Analysis of Network Externalities in Peer-to-Peer Music-Sharing Networks," *Information Systems Research*, 15(2), 155–174.

Bala, V., and S. Goyal (2000): "A Noncooperative Model of Network Formation," *Econometrica*, 68(5), 1181–1229.

Birke, D. (2008): "The economics of networks - A survey of the empirical literature," *Nottingham University Occasional Paper Series*, pp. 1–33.

Bloch, F., and B. Dutta (2009): "Communication networks with endogenous link strength," *Games and Economic Behavior*, 66, 39–56.

Bramoullé, Y., and R. Kranton (2007): "Public Goods in Networks," *Journal of Economic Theory*, 135, 478–494.

Cagno, D. D., and E. Sciubba (2007): *Advances in Decision Making Under Risk and Uncertainty*chap. The Determinants of Individual Behaviour in Network Formation: Some Experimental Evidence, pp. 219–241. Springer-Verlag Berlin Heidelberg.

——— (2008): "Social Networks and Trust: not the Experimental Evidence you may Expect," working paper.

Cho, M. (2006): "Endogenous Formation of Networks for Local Public Goods," working paper.

Corbae, D., and J. Duffy (2008): "Experiments with Network Formation," *Games and Economic Behavior*, 64, 81–120.

Falk, A., and M. Kosfeld (2003): "It's all about Connections: Evidence on Network Formation," *IZA Discussion Paper*, (777).

Goeree, J. K., A. Riedl, and A. Ule (2008): "In search of stars: Network formation among heterogeneous agents," *Games and Economic Behavior*, article in press.

Gowrisankaran, G., and J. Stavins (2004): "Network Externalities and Technology Adopotion: Lessons from Electronic Payments," *The RAND Journal of Economics*, 35, 260–276.

Goyal, S. (2005): "Strong and Weak Links," *Journal of the European Economic Association*, 3, 608–606.

Goyal, S., M. J. v. d. Leij, and J. L. Moraga-Gonzlez (2006): "Economics: An Emerging Small World," *Journal of Political Economy*, 114, 403–412.

Gu, B., Y. Huang, W. Duan, and A. B. Whinston (2007): "Online Peer-to-peer Communities: An Empirical Investigation of a Music Sharing Network as a Dynamic Two-sided Network," working paper.

——— (2009): "Indirect Reciprocity in Online Social Networks - A Longitudinal Analysis of Individual Contributions and Peer Enforcement in a Peer-to-Peer Music Sharing Network," *McCombs Research Paper Series No. IROM-06-09*, pp. 1–40.

Jackson, M. O. (2004): *Group Formation in Economics; Networks, Clubs and Coalitions*chap. A Survey of Models of Network Formation: Stability and Efficiency. Cambridge University Press.

Jackson, M. O., and A. Watts (2002): "The Evolution of Social and Economic Networks," *Journal of Economic Theory*, 106, 265–295.

Jackson, M. O., and A. Wolinsky (1996): "A Strategic Model of Social and Economic Networks," *Journal of Economic Theory*, 71, 44–74.

Kosfeld, M. (2003): "Network Experiments," *IERE Working Paper*, (152).

Kossinets, G., and D. J. Watts (2006): "Empirical Analysis of an Evolving Social Network," *Science*, 311, 88–90.

Kumar, R., J. Novak, and A. Tomkins (2006): "Structure and Evolution of Online Social Networks," in *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, Philadelphia 2006*, pp. 611–617.

Newman, M. (2001): "The structure of scientific collaboration networks," *Proceedings of the national Academy of Sciences of the United States of America*, 98, 404–409.

O'Dea, D. (2008): "Heterogeneity in Social Networks and the Classical Theory of Public Goods," working paper.

Rosenblat, T. S., and M. M. Mobius (2004): "Getting Closer or Drifting Apart?," *Quarterly Journal of Economics*, 119, 971–1009.

Saloner, G., and A. Shepard (1995): "Adoption of Technologies with Network Effects: An Empirical Examination of the Adoption of Automated Teller Machines," *The RAND Journal of Economics*, 26, 479–501.

Xia, M., Y. Huang, W. Duan, and A. B. Whinston (2007): "To Share or Not To Share? - An Empirical Analysis on User Decisions in Online Sharing Communities," working paper.

6 Appendix - Tables and Regression Results

Dep. Variable: Prob(Severance)	Model 1a	Model 2a	Model 3a
Own music of severer at $t = 0$	**.0000139*****	-	**.0000169*****
	(2.72e-06)		(2.65e-06)
Own music of severed at $t = 0$	**-.0000182***	-	**-.0000161****
	(7.71e-06)		(7.10e-06)
Play-lists of severer at $t = 0$	**0.00017*****	-	.0000592
	(0.0000301)		(.000043)
Play-lists of severed at $t = 0$	.0000641	-	.000207
	(.000183)		(.0000197)
Music uploaded by severer per day	-	**.00483*****	**.00136*****
		(.00053)	(.000382)
Music uploaded by severed per day	-	**-.0120***	**-.00966***
		(.00634)	(.00555)
Play-lists created by severer per day	-	**-.0449*****	**-.0186*****
		(.00652)	(.00465)
Play-lists created by severed per day	-	.0775	.0754
		(.055)	(.0467)
observations	1,521	1,521	1,521
pseudo R^2	.429	.312	.469
Chi^2	347.12	252.54	379.64

Table 4: Free Riding and Negative Reciprocity a): Logistic regressions on the probability that a given friendship is severed. Marginal effects and standard errors at mean reported.

Dep. Variable: Prob(Severance)	Model 1b	Model 2b
Songs listened per day by severer	**-.130***	-.0450
	(.009)	(.0455)
Songs listened per day by severed	**.00952*****	**.00519*****
	(.00244)	(.00154)
Own music of severer at $t = 0$	-	**.0000145*****
		(2.38e-06)
Own music of severed at $t = 0$	-	**-.0000215***
		(.0000115)
Play-lists of severer at $t = 0$	-	.0000623
		(.0000451)
Play-lists of severed at $t = 0$	-	.0003717
		(.000234)
Music uploaded by severer per day	-	**.000746*****
		(.000286)
Music uploaded by severed per day	-	**-.00862****
		(.00366)
Play-lists created by severer per day	-	**-.0279*****
		(.00707)
Play-lists created by severed per day	-	**.0901****
		(.00707)
observations	1,477	1,477
pseudo R^2	.0473	.358
Chi^2	26.64	201.58

Table 5: Free Riding and Negative Reciprocity b): Logistic regressions on the probability that a given friendship is severed. Marginal effects and standard errors at mean reported.

Dep. Variable: music files added	Model 1	Model 2	Model 3
music friends	-5.44e-06	-.0000134	-.0000292
	(9.66e-06)	(.0000138)	(.0000199)
own music	-.0602665	-.0666011	-.0920428
	(.0585873)	(.0827902)	(.0928703)
Δmusic friends	**4.24e-06***	**5.56e-06****	**6.75e-06****
	(2.29e-06)	(2.49e-06)	(2.91e-06)
experienced dummy	-	-.0423019	.4954036
		(.0632985)	(.5668963)
time	-	.0234988	.0301578
		(.0155074)	(.016886)
comments	-	1.478684	1.352071
		(4.517221)	(4.346862)
messages	-	.0867128	-.1812929
		(1.362083)	(1.33175)
play-lists friends	-	-	.0073593
			(.0072321)
own playlists	-	-	.4954036
			(.5668963)
constant	3.364962	3.906991	5.863737
	(2.266137)	(2.804955)	(3.610217)
observations	177,828	177,828	177,828
groups	25,031	25,031	25,031
overall R^2	.0294	.0241	.0208
F	2.96	5.52	9.15
rho	.924	.933	.962

Table 6: Free Riding and Positive Reciprocity for Information Provision, FE-regressions controlling for potential user-cluster heteroskedasticity, entire sample

Dep. Variable: music files added	Model 1	Model 2	Model 3
music friends	-.0000124	-.0000293	**-.0000572***
	(.0000198)	(.0000197)	(.0000345)
own music	-.0593032	-.0661795	-.1171941
	(.0590668)	(.0836462)	(.1071663)
Δmusic friends	3.87e-06	**9.61e-06****	**.0000105****
	(2.81e-06)	(3.71e-06e-06)	(4.06e-06)
experienced dummy	-	-1.636376	-1.314448
		(2.713547)	(2.485865)
time	-	.3370503	.2793232
		(.17308)	(.1722218)
comments	-	1.48062	1.352071
		(4.511961)	(4.346862)
messages	-	.0533741	-.1812929
		(1.376997)	(1.33175)
play-lists friends	-	-	.0126578
			(.0123971)
own playlists	-	-	1.011472
			(1.176967)
constant	11.12651	5.51823	14.21924
	(7.765481)	(9.963463)	(13.35873)
observations	24,413	24,413	24,413
groups	9,428	9,428	9,428
overall R^2	.0472	.0409	.0317
F	2.27	5.93	6.67
rho	.725	.757	.896

Table 7: Free Riding and Positive Reciprocity for Information Provision, active user sub-sample

Dep. Variable: play-lists added	Model 1	Model 2	Model 3
play-lists friends	-.0002687	-.0007544	**-.0013166****
	(.0005835)	(.0007107)	(.0006179)
own play-lists	**-.1671713*****	**-.2086505*****	**-.2901718*****
	(.0393094)	(.0453584)	(.0489084)
Δplay-lists friends	**.0009297****	**.0010383****	**.0010542*****
	(.000391)	(.0004011)	(.0003998)
experienced dummy	-	.0046399	.0072075
		(.0084659)	(.0085653)
time	-	**-.0046172*****	.0019167
		.0010961	(.002046)
comments	-	-.0715414	-.0329556
		(.4122087)	(.3545851)
messages	-	.1556359	.0831986
		(.1184353)	(.1029976)
dummy active	-	-.0000785	.0040834
		(.0079719)	(.0073418)
music friends	-	-	**-5.21e-06*****
			(1.97e-06)
own music	-	-	**.0123697****
			(.0056172)
constant	.3138434	.4337222	.5619726
	(.0710501)	(.0549475)	(.2757704)
observations	177,828	177,828	177,828
groups	25,031	25,031	25,031
overall R^2	.0435	.0000	.0005
F	9.43	16.51	22.27
rho	.797	.923	.975

Table 8: Free Riding and Positive Reciprocity for Information Organization, entire sample

Dep. Variable: play-lists added	Model 1	Model 2	Model 3
play-lists friends	-.0003837	-.0006297	-.0012703
	(.0007793)	(.0009335)	(.0009274)
own play-lists	**-.123411*****	**-.1525017*****	**-.2550607*****
	(.029512)	(.0836462)	(.0808259)
Δplay-lists friends	**.0012131****	**.0012077****	**.0012261****
	(.0005418)	(.0005203)	(.0005189)
experienced dummy	-	.0961424	-.0053575
		(.1195154)	(.1435919)
time	-	-.0199061	.0036746
		(.0116495)	(.011876)
comments	-	-.0874014	-.0446645
		(.3545696)	(.3362418)
messages	-	.0924774	.0663375
		(.1224277)	(.1116464)
music friends	-	-	**-.0000168*****
			(3.25e-06)
own music	-	-	**.0108558***
			(.0064666)
constant	.5901143	.979095	1.46338
	(.1505223)	(.353222)	(.9223387)
observations	24,413	24,413	24,413
groups	9,428	9,428	9,428
overall R^2	.0892	.0026	.0008
F	12.18	11.53	24.35
rho	.524	.651	.914

Table 9: Free Riding and Positive Reciprocity for Information Organization, active user sub-sample