JENA ECONOMIC RESEARCH PAPERS

2011 – 015

The role of reciprocation in social network formation, with an application to blogging

by

Alexia Gaudeul
Caterina Giannetti

THE ROLE OF RECIPROCATION IN SOCIAL NETWORK FORMATION, WITH AN APPLICATION TO BLOGGING

ALEXIA GAUDEUL[†] AND CATERINA GIANNETTI[‡]

ABSTRACT. *This paper deals with the role of reciprocation in the formation of individuals' social networks, that is to what extent initiating a relation brings about its reciprocation. Following the activity of a panel of bloggers over more than a year, we seek to establish whether bloggers are mainly involved in social networking or are part of the media industry. We adapt a standard capital investment model to study the effect of reciprocation on the building of social capital. Results of our analysis confirm that activity and reciprocation both play a role in the dynamics of social media.*

Keywords: Bloggers, Friendship, LiveJournal, Media, Panel Data, Reciprocation, Reciprocity, Social Capital, Social Media, Social Networks

JEL classifications: C33, D85, L82

In all friendships implying inequality, the love also should be proportional,

i.e. the better should be more loved than he loves (...)

Aristotle, Nicomachean Ethics, Book VIII, Chapter 7

1. INTRODUCTION

This paper deals with the role of *reciprocation* in the formation of individuals' social networks. Our study focuses on the extent to which initiating a relation brings about its reciprocation, that is whether, for example, me saying "I like you" induces you to like me. To the extent this is the case, we investigate how far one's willingness and efforts in initiating relations with others can help grow one's social network. We concentrate on such patterns of reciprocation because, while reciprocity has been shown to be "a universal structure of human morality" (Gintis et al., 2008; Henrich et al., 2001), and while there is abundant evidence on the prevalence of reciprocal behavior among humans (mostly experimental, *e.g.* Fehr and Gächter, 2000), there is less empirical evidence on the role of reciprocation in the formation of human relations (see section 3).

We study bloggers' choices of which other blogs to subscribe to. We seek to determine to what extent bloggers' networks are primarily based on affinity – people read those they feel close to –, or whether other, less personal factors come into play, such as how interesting, informative, influential, *etc...* a blogger is. In that sense, we want to establish whether bloggers are mainly involved in social networking, whereby their readers are also friends, or whether, as in the traditional media industry, bloggers, like most journalists, have little emotional attachment to their readers. In the first case, reciprocation would be particularly important in maintaining one's network while effort exerted in blogging would have little influence on one's number of readers, while in the second case, readers would not require reciprocation but better and more frequent posting would translate into higher audience.

The economics of friendship. We apply economic reasoning to one's choice of whom to entertain relations with and offer a model whereby the utility derived from a new relation, as well as the perceived obligation to reciprocate the interest shown by others in oneself, both play a role. While it may appear unusual to apply grim economic reasoning

to one's choice of friends, there is a long tradition thinking in those terms: Aristotle distinguished three types of relations between individuals: those that arise because of how good, virtuous or perfect an individual is (first type), those that are based on how *useful* the relation is (second type), and those that survive only as long as they are *pleasant* (third type) (Nicomachean Ethics, VIII, 3). Only in the first type of relationships is there no necessary balance between how much is contributed by each party in the relationship. Indeed, "genuine" friends *"loathe and banish from their conversation these words of division and distinction, benefits, obligation, acknowledgment, entreaty, thanks, and the like"* (de Montaigne, 1588). However, as noted by Aristotle, most relationships are of the second or third type. Examples of such relationships are those that allow individuals to gain social support, information and opportunities for a number of activities such as going out, doing sport, getting a job or making business (Lin, 2001). For this type of relations, individuals with qualities such as beauty, intelligence, wealth, status, power, breadth of experience or wit are more attractive. In this context, individuals that are pleasant or useful face high demand so that individuals that are less attractive may have to incur large investments – in courtship for example – to establish a mutual relation with them, or may have to content themselves with not seeing their interest reciprocated. Relations that occur between unequal parties are thus driven by their relative status and contributions, with the more attractive, productive, prestigious partners not necessarily reciprocating fully the attention of those with lower status: *"In all friendships implying inequality, the love also should be proportional, i.e. the better should be more loved than he loves (...)"* (Nicomachean Ethics, VIII, 7). However, we recognize that one's decision to entertain a relation with someone else cannot be reduced to a calculus in terms of costs and benefits: social skills and attitudes also matter. Not all similarly endowed individuals are able to or want to maintain many relations. Individuals that are open to making new acquaintances and willing to devote time to others ought to have wider networks, irrespective of their inherent qualities. Some dimensions of personality are important as well: Ozer and Benet-Martínez (2006) cite evidence that extroversion is "the most important [personality] predictor of popularity and status among adults".

Why blogging. The study of social networks has been impeded by the difficulty of recording every interactions of an individual over time along with his activity. The emergence of tools for social networking and collaboration *via* the Internet such as Facebook, LinkedIn, Twitter, Wikipedia or Reddit, makes it possible to collect such data unobtrusively – that is, without the individual knowing their activity is being recorded – and cheaply – by extracting information from blogging sites using Web-scraping software.

Blogger networks have properties that make them particularly well suited for empirical analysis. All blogs are online so it is possible to have a complete picture of all blogs an individual blog is linked to, whether through its blog roll,[1] comments by the blogger or links to entries on other blogs. In contrast, networking tools for professionals such as LinkedIn only reflect a part of those professionals' networks – limited to those individuals that also use the same tool. Another advantage is that blogs and their interconnections form a relatively self-contained world with a fairly clear unity of purpose, while generalist social networking tools such as Facebook mix many different types of relations – friends, acquaintances, classmates, colleagues, family, celebrities, romantic interests, *etc*... Finally, activity that plays a role in establishing relations between bloggers, such as posting entries, making comments or joining communities, can be tracked. On the other hand, activity within Facebook or LinkedIn usually plays little role in establishing relations – as opposed to maintaining them –, as those networks (for the most part) only formalize relations that were established outside their settings.

The role of reciprocation. Posting interesting content, being the first to cover a topic, differentiating from others' coverage of a topic, or covering original topics, are all strategies that help one gain attention (Shen, 2009). However, we also think that reciprocation processes are strategically exploited by bloggers so as to maximize their audience. We are interested in strategies that rely on paying attention to others so as to be paid attention to. A blogger may thus seek bloggers and read their blogs in order to get attention reciprocated rather than merely to keep updated on topics of interest. Reciprocation may occur not only out of interest for the blogger who initiated the reading relation, or in order to encourage the initiator to keep on reading one's blog, but also because a *norm of reciprocity* may

---

[1]"A list of other blogs that a blogger might recommend by providing links to them (usually in a sidebar list)" (see Wikipedia, 2011).

make one feel obliged to reciprocate attention when paid attention to. We think that the working of this norm of reciprocity is of particular relevance to the emergence of bloggers' social networks because, as Gouldner (1960) points out, the norm of reciprocity helps in the creation of stable social systems by providing a starting mechanism for relations in situations where there are no established rules for social interactions. In the presence of such a norm, the one who starts giving attention is confident his gift will be reciprocated, thus resolving an impasse over who should start giving attention first.

We adapt an economic model proposed by Glaeser, Laibson, and Sacerdote (2002) to include the mechanisms of reciprocation into the dynamics of the building of social capital over time. In this context, social capital is measured by how many readers a bloggers has and/or how many communities he belongs to.[2] Bloggers are assumed to build their social networks not only by providing content and interacting with other bloggers, but also by seeking out other bloggers (reading them) so they then reciprocate reading, and by reciprocating the attention (reading) of others.

Outline. The paper starts by setting out the context of the present study: what is a blog, who are the bloggers, why do they blog, how are blogging networks structured, how do they evolve? We then elaborate on the workings of the process of reciprocation in the context of blogging. Different ways to model bloggers' activity are subsequently introduced, either as motivated by the need to gather information, by the desire to gain attention or as a way to access social support. Of those models, we choose our preferred one and test it by analyzing our data, a panel of bloggers on LiveJournal whose activity was followed week by week over more than a year.

2. BLOGS: WHAT, WHO, WHY AND HOW

Blogs are websites that are updated regularly with content posted in units, called "posts" or "entries". The more recent entries appear at the top of the web page. Content posted usually consists of text, but can also include pictures, videos, speech and music, and frequently includes links to and commentaries on other content on the web. Each blog post

---

[2]The concept of "community" is specific to LiveJournal, the blog hoster we are extracting data from. "A LiveJournal community is a journal where many users post entries about a similar topic." (see LiveJournal, 2011b). This corresponds more or less to the concept of a collaborative blog (see Wikipedia, 2011).

can typically be commented upon by readers in a space beneath the entry itself. Those comments may themselves be commented upon, leading to threads of discussion among the readers or with the author. Other interactions may occur if a blogger's post is mentioned in another blog.

Blogging is very much of a rule driven activity. Bloggers follow a range of informal rules for making the selection of their own content – "Can I speak about other topics if the advertised theme of my blog is religion?", or "Can I speak about the personal affairs of close relatives?"–, providing references to their sources –"Should I provide links to all content I refer to?", or "Does an article in the Daily Mail count as a valid reference?"–, relating to others –"Can I delete comments made by others on my blog?", or "Do I have to explain when I stop reading another blog?"–, and so on (Schmidt, 2007; Marwick, 2008). While there is no consensus on what specific rules are valid, and while bloggers may change the rules that hold on their blogs over time, most have some idea of what is acceptable behavior and what is not, on their blog or in the wider community of bloggers.

Typology of blogging. Blogs may be classified along many dimensions – topic, popularity, type of content, language, *etc...* –, but a particularly prevalent distinction is made between filter blogs (also: thematic blogs) and personal journals (also: diaries) (Wei, 2009). Filter blogs focus on a specific topic, often within the professional expertise of the blogger, while personal journals mainly deal with events in the blogger's life and are used as a tool for self-expression. The distinction is not always clear cut however, as authors of thematic blogs often mention events in their own life while diarists often share expertise on their own job, regularly speak about their hobbies or express their political views and their positions on contemporary social debates. A lot of attention has focused on thematic blogs, how they change the way information is being spread (Bar-Ilan, 2005), how such bloggers compare with journalists (Lemann, 2006) and how they influence politics (McKenna and Pole, 2008). The majority of blogs belong to the online diary genre however (Herring et al., 2005b; Technorati, 2009). In our work, we differentiate between those bloggers with less than *150* "friends" (people they read), who then might conceivably be composed of those people with whom the blogger maintains stable social relationships (Dunbar, 1992), and those with more than *150* "friends", with whom relations are then less likely to be driven

by personal factors but rather by the content produced and the information exchanged within the blog.

Sociology of blogging. A variety of surveys have been conducted to find out who the bloggers are. Many of those surveys rely on bloggers' self-selection into filling online questionnaires, cover a limited range of blog hosters, or rely on the blogger being listed at sites such as Technorati or using other specific tracking tools. Among more recent surveys however, a few rely on random selection and direct contact with the bloggers (Technorati, 2009; Lenhart and Fox, 2006). For all their limitations, surveys tend to agree that bloggers are better educated and more affluent than the average, and are majoritarily males – though gender composition can vary greatly depending on the blog hoster or the type of blogs surveyed. Bloggers are still mainly concentrated in the US, though the Russian (Gorny, 2006) and Chinese (Yu, 2007) blogosphere develop fast and with little links to the English language blogging community.

Motivations for blogging. Motivations for blogging are varied: expressing one's self, documenting one's life, commenting on current events, participating in community forums, and searching for information (Huang et al., 2007). According to a survey by Technorati (2009), bloggers say they blog to "speak their mind" and "share their expertise and experience with other people", but also to "meet and connect with like minded people". They measure their blog's success first by the personal satisfaction they derive from it, followed by how many people read it, how many comment on it, link to it, or add it to their blog roll. Getting attention for their opinions and expertise, and building relations with others, notably to share experiences and obtain social support, are therefore important to bloggers. Empirical work confirms the link between getting attention and producing content. Marlow (2006) finds that time spent maintaining a blog pays off in terms of audience size and feedback. It is not clear from their work whether higher audience leads to more effort or *vice-versa*. Using instrumental variables however, Hofstetter et al. (2009) show that bloggers who gain readers increase their content production as a result, but also that more content production gains more readers. Beyond the world of blogging, the mechanics of free, user-generated contributions to media content on the Internet have been investigated in Huberman, Romero, and Wu (2009), who identify the same effects as in Hofstetter et al.

(2009) among YouTube contributors, and Zhang and Zhu (2010), who show that Wiki contributors appear to respond to lower audiences by reducing their output. Contributors to user-generated content thus respond positively to attention, while those who make more effort are rewarded with more attention.

Structure of blogging networks. Insights into the social structure of blogging networks were first expressed by Shirky (2003), who predicted that blogs would soon come to diverge into two types, A-list blogs who attract so much attention that they cannot possibly reciprocate all of it, thus ending up as part of the mainstream media, and conversational blogs, part of the "long tail" (Anderson, 2004), who spend time cultivating their links with a few others.[3] Kumar, Novak, and Tomkins (2010) provide further insights into the structure of blogging networks. They distinguish three types of blogger networks: singletons, which are isolated bloggers with no links to others, isolated communities with a star like structure centered on a single blogger, and giant components, centered on a core of well connected bloggers.[4] This type of structure is shown to be consistent with the existence of three types of bloggers: "passive" ones who only read others, "inviters" who are the core of star shaped networks made of those they invited and "linkers", who both link to and read other blogs.

In the next part, we provide some insights into the mechanisms that govern the establishment of relations between bloggers.

3. THE RULES OF RECIPROCATION

Reciprocation is an important mechanism in the formation of social networks. Schaefer et al. (2010) show that reciprocation is one of the main drivers in network formation among preschool children, along with popularity (making friends with those who have many friends) and triadic closure (making friends with friends of friends). Doreian et al. (1996) find that reciprocated links appear very early in the formation of networks, while triadic closure plays a role only later. Reciprocation is not only important in the emergence of relations but also makes them more stable and profitable. For example, reciprocated

---

[3]Rui and Whinston (2010) provide some conditions under which an even more extreme form of differentiation occurs, whereby some bloggers only post content and do not read others', while others only read blogs and do not produce any content.

[4]See also Herring et al. (2005a) for similar results.

friendships among children are more stable than unilateral ones (Hallinan, 1978). Perhaps contributing to the stability of reciprocated relations, agents tend to maintain reciprocated links even when this is no longer optimal (see experiment in Conte, Di Cagno, and Sciubba, 2009). Finally, adolescents with many mutually named best friends obtain better educational outcomes than others (Vaquera and Kao, 2008). Our innovation with respect to the above literature consists in being able to monitor the activity of the nodes along with their formation of new links, something that would be exceedingly difficult in the contexts evoked above.

In this part, we want to focus on describing the precise working of the rules for reciprocation in blogging networks and how those rules combine with other rules for selecting people to read. Of particular interest to us are rules that relate to linking with ("making friend with", "friending") other bloggers, that is, publicly adding a blogger to one's blog roll. This is not an insignificant or neutral act. (Marlow, 2006) reports that bloggers read more than *80%* of the blogs in their blog roll in the last month, and over *60%* in the last week . Adding a blogger to one's friend list on LiveJournal – the blog hoster we are extracting data from – not only means it will appear on one's reading list but also gives that blogger access to one's "friends-only" entries (see LiveJournal, 2011a). Marwick (2008) reports on the depth of feeling attached to the term "friend" on LiveJournal, and mentions that the vast majority of users of LiveJournal make at least occasional use of the "friends-only" privacy setting for their posts.

Raynes-Goldie and Fono (2006) set out the different meanings that are attached to adding a blogger to one's "friend" list – the list of journals read – on LiveJournal. Among different reasons guiding the choice of whom to read, a blogger might read people he likes or feels a connection to, people he enjoys reading, or, through the norm of reciprocity, people who chose to read him. The act of "friending" may be initiated through a comment on a post by another blogger, or a reference to a blogger in one's post. It may also occur without need for any prior contact, as a result of having seen the blogger referred to on another blog, seeing him post in a community one belongs to as well, or of searching for bloggers with similar interests to one's own. Reactions to being referred to, commented upon or added vary, but there is at least some level of reciprocation: If a post of mine is

being linked to and reviewed favorably, I may acknowledge this in some fashion, by for example referring to the linker's blog in a later post. If I receive a comment on my blog, I may reply to it if appropriate, and could also leave a comment on the commenter's blog. Such interactions, repeated over time, may result in establishing a stable reading relation with the linker. Conversely, many bloggers do not add back a blogger who "friended" them without prior interactions.

The rules for "adding" and "adding back" a blogger to one's blogroll are dependent on the context. A prestigious blogger (many readers, often linked to, well written) may "friend back" less easily. A recently established blog may have to "face the test of time" before being added back. A blogger may add back another only after a length of time, to make sure the adding was not a random fluke. Some bloggers may add back only people they know in real life, others may exclude just such people to preserve their own anonymity. Conversely, there are also rules for "dropping" other bloggers from one's blog roll: many bloggers systematically "drop" another if that other drops them first. It is often considered good manner to give some explanation for not reading another blogger anymore.

As seen above, the norm of reciprocity is particularly prevalent in the initiation and continuation of a relationship between bloggers. While rules of friendship reciprocation apply to blogging, the norm of reciprocity has also been shown to be important in many types of social networks on the Internet. Chun et al. (2008) observe comments left on the "guestbooks" of users of Cyworld, "the largest social networking site in (South) Korea", and show that those are highly reciprocated. Chan and Li (2009) consider a Chinese co-shopping site where users share shopping tips and product information and plan bulk purchases, and show that reciprocity, or the expectation of reciprocity, has a "critical effect on social system maintenance by enhancing commitment to the community and intention to co-shop". Sadlon et al. (2008) study Digg, a social news website where users vote on submitted links, and show that top submitters (those with popular submissions) tend to form an exclusive group that "upvote" each others' stories in a reciprocal fashion.[5] Gu et al. (2009) investigate a peer-to-peer music sharing network and show that a pattern of

---

[5]For another paper dealing with the dynamics of user submission and upvoting on Digg, see Lussier, Raeder, and Chawla (2010).

indirect reciprocity holds, whereby free-riders, who only download music but do not offer music to download, are sanctioned with lower download speeds.

4. ONLINE SOCIAL NETWORKS AND SOCIAL CAPITAL

We justify in this part our choices in the modeling of the activity of bloggers. We believe that blogging systems can only be partially represented through standard sender-receiver games where senders compete for the attention of a passive audience, whereby "he who shouts loudest has the floor" (Falkinger, 2007). Indeed, bloggers are both producers and receivers of information. It is therefore more appropriate to see blogging as a special form of social networking, where in addition to knowledge sharing, actors are also involved in the formation and sustenance of social links (Lussier, Raeder, and Chawla, 2010). Formal models in the economic literature on social networks – networks where nodes choose with whom to form and maintain links – can be distinguished between those that consider social networks as a tool for gathering and transmitting value (*e.g.* information), and those that consider them as a tool for combining individual contributions to a productive effort. Galeotti and Goyal (2010) consider a situation where information can be acquired at some cost by any agent and links between agents are hard to maintain, but those links can be used to transmit information freely once it has been acquired by any agent in a network. Core-periphery network architectures – such as star-shaped networks – then emerge as equilibrium outcomes. Jackson and Wolinsky (1996) focus rather on the combination of individuals' activity within a network. Their co-authorship model is such that individuals divide their time across members of their network. Combinations of individual efforts generate added value, that is, it is more efficient to work with co-authors than by oneself. This results in a number of strongly connected network components with no links to each other. In this literature, the network structure and activity of participants depends on what function a network plays, *i.e.* in which way it is used, and by extension, on the objectives of the participants in the network. Results differ starkly: In Galeotti and Goyal (2010), only a few participants are active in gathering information while others form links with information gatherers and passively receive information from them, while in Jackson and Wolinsky (1996) individuals connect with all those that have the same number of co-authors as they themselves have and devote to each of their relations the same amount of attention as

those devote to them.[6] We believe that blogging networks fall in-between those two types of predictions. There will be a relation between activity at the level of a node and its ability to attract links, as in Galeotti and Goyal (2010), but this will be moderated by the need to maintain a balance in bloggers' relations as in Jackson and Wolinsky (1996), which means bloggers will devote less time and attention to bloggers that do not reciprocate by reading them back. The literature on social capital formation provides more insights on the value derived from entertaining mutual reading relations among bloggers. In that view, blogging networks are the *result* of social capital accumulation, in the sense that they are the result of past favors that have led to "friending" – for example making a positive reference to one's blog, giving advice in a comment, sharing relevant life experiences –, and a *source* of social capital, in the sense that mutually beneficial exchange of social support and information is expected to take place within one's network in the future. The role of online social networks as a way to accumulate, use and retain social capital has been investigated in a number of recent papers. Ellison, Steinfield, and Lampe (2007) show that Facebook works well in maintaining "weak ties", which are sources of information rather than of social support. Marlow (2006) find that "social" bloggers (diarists) maintain personal contacts with their readers and thus gain social capital through their online activity, while "professional" (thematic) bloggers, who entertain larger audiences and invest more time in their blogs, do not gain much in terms of social capital. As for the use of social capital online, Antin and Earp (2010) suggest that MySpace is used by lesser known musicians to express affiliation with better known ones – and thus attract traffic –, and by better known musicians to express support for lesser known ones, thus making use of their power of patronage. Finally, social capital gained online can motivate continued activity in blogging (Lento et al., 2006).

Inspired by the above, we modify Glaeser, Laibson, and Sacerdote's model of social capital formation [2002] to investigate the role that reciprocation plays in the building of social networks. In doing so, a distinction is made between one's tendency to reciprocate offers of friendship *vs.* the willingness of others to reciprocate one's offer of friendship.

---

[6]A recent paper combines social and informational value from link formation, which allows for a wider variety of equilibrium outcomes (Harmsen-van Hout, Herings, and Dellaert, 2010).

A model of investment in social capital. Glaeser, Laibson, and Sacerdote (2002) present "a simple model of investment in social capital" that is "almost identical to the standard models of investment in physical and human capital". In this, it departs from the "bulk of the modern literature on social capital, which treats social capital as the characteristic of a community". An individual's social capital at time t, denoted R_t, for "number of **R**eaders", as the number of readers is our proxy for social capital, evolves as a function of investment in social capital formation, denoted I_t, and of last period's social capital R_{t-1}, which depreciates at a rate $\delta > 0$ each period, so the following equation obtains:

$$R_t = f(I_t) + (1 - \delta)R_{t-1} \tag{1}$$

This equation is rewritten to represent the increase in social capital in period t, $r_t = R_t - R_{t-1}$ as:

$$r_t = f(I_t) - \delta R_{t-1} \tag{2}$$

This expresses how two factors are at work in the evolution of social capital: investments in social capital formation and depreciation of existing social capital. Depreciation reflects the tendency of existing readers to drop the blogger out of their reading list, due to several factors, such as boredom, lack of attention received, conflict and disagreements with the blogger, and so on. Investment I_t combines several aspects of one's activity: seeking out and adding "friends", publishing content in one's blog, making comments on others' blogs, joining communities and so on.

Integrating reciprocity into the model. Denote the total number of "friends" gained in period t, $f_t = F_t - F_{t-1}$, with F_t the number of "friends" at time t. This can be divided into those "friends" I sought out on my own and **a**dded, f_t^a, and those "friends" who first sought me out (added me to their reading list) and whose readership I **r**eciprocated, f_t^r. Thus, $f_t = f_t^a + f_t^r$.

In the same manner, denote the total number of readers gained in period t, $r_t = R_t - R_{t-1}$. This can be subdivided into r_t^a, the number of readers that **a**dded me in period t without

me first adding them to my reading list, and r_t^r, those bloggers whom I first added to my reading list and who then **r**eciprocated my readership. Thus, $r_t = r_t^r + r_t^a$.

Denote θ my likelihood to reciprocate the readership of those bloggers who add me as "friends". Then $f_t^r = \theta r_t^a$. Similarly, denote ρ the likelihood with which bloggers I add to my reading list reciprocate and add me back. Then $r_t^r = \rho f_t^a$.

The following system of two equations results from the above:

$$(3) \qquad\qquad f_t = f_t^a + \theta r_t^a$$

$$(4) \qquad\qquad r_t = \rho f_t^a + r_t^a$$

Solving to express r_t as a function of f_t and vice versa, one obtains:

$$(5) \qquad\qquad r_t = \rho f_t + (1 - \rho\theta) r_t^a$$

$$(6) \qquad\qquad f_t = \theta r_t + (1 - \rho\theta) f_t^a$$

Denote A_t all aspects of investment I_t other than seeking out and adding "friends", and let us rewrite $f(I_t)$ as $\beta A_t + \rho f_t^a$. Replacing $f(I_t)$ by this expression in (2) and comparing with (4), one obtains that

$$(7) \qquad\qquad r_t^a = \beta A_t - \delta R_{t-1}$$

so (5) can be rewritten as follows:

$$(8) \qquad\qquad r_t = \rho f_t + (1 - \rho\theta)\beta A_t - (1 - \rho\theta)\delta R_{t-1}$$

This equation explicitly takes into account my investment in seeking "friends", which through reciprocation increases my number of readers. A naive regression of r_t on the elements in A_t and on R_{t-1}, on the other hand, would lead to incorrect estimates of the influence of activity and depreciation on the evolution of one's readership.

In the same way as estimation of (8) requires indicators for A_t, which were spelled out previously on the preceding page, estimation of (6) requires indicators for f_t^a, that is, how active the blogger is in his search for new "friends". We consider investment variables B_t

such as the number of communities joined per sampling period or the number of communities one participates in, as those put one in contact with more people. Our stock variable will be the number of existing "friends", which will depreciate by factor λ each period as the blogger drops less interesting friends. We thus rewrite f_t^a as $\gamma B_t - \lambda F_{t-1}$, so (6) can be expressed as follows:

$$(9) \qquad\qquad f_t = \theta r_t + (1 - \rho\theta)\gamma B_t - (1 - \rho\theta)\lambda F_{t-1}$$

The number of friends F_{t-1} may have an ambiguous effect in the friends equation because while friendships "depreciate" over time, they also put one in contact with the "friends" of their "friends", thus potentially contributing to one's "friending" activity. One may thus find λ to actually be negative, i.e. more friends beget more friends. In the same way, having more readers may make it easier to obtain further readers, either because existing readers refer to one's blog in their own posts, or because having many readers is seen as a signal of quality and thus increase one's attractiveness. One may thus find δ to actually be negative, i.e. more readers beget more readers.

Estimation of both equations obtains estimates of θ and ρ, from which the influence of activity on r_t^a, i.e. the number of readers gained through activity other than simply adding "friends", is obtained. Indeed, consider a regression of r_t as a linear function of f_t, the elements in A_t and R_{t-1}. Comparing (8) and (7), one sees that estimates of the coefficients on the different aspects composing A_t must be divided by $(1 - \rho\theta)$ to obtain a true estimate of parameters in (7). Similarly, the rate of depreciation in the number of readers, δ, is obtained by dividing the estimate of the coefficient on R_{t-1} by $(1 - \rho\theta)$. We are particularly interested in how those coefficients vary depending on the stage in the life in the blog, its popularity and the language community it belongs to.

5. Data collection and dataset description

We followed the activity of a sample of bloggers on LiveJournal ("LJ"). LJ was created by Brad Fitzpatrick in 1999. The first users of LJ were US high school and college students. LJ's growth in the US slowed down in the second half of the noughties as its original users

either left for Facebook for pure social networking, or for other blog hosts and tools that were better integrated into the more general blogosphere (Wordpress, Blogger, Movable Type, ...). The site found a second breath in Russia, where LJ is the most popular blogging site and social media platform.[7] Reflecting this change, the company is now owned by SUP, an online media company based in Moscow.

A blog on LJ ("a LJ") can be used in many ways, as a private journal, a blog, a discussion forum or a social network. This illustrates the flexibility and breadth of potential uses of a LJ: some users have accounts to post their diaries and choose to keep their content private or limit its access to a close circle of "friends", others make the whole of their journal public, link to content on others' blogs and comment on others' entries. Finally, some have an account on LJ mainly to be able to join and contribute to communities and discuss the content posted there.[8]

Our sample consists of a cross-section of more than two thousands of LiveJournal users. The bloggers were originally selected randomly among those bloggers that had displayed some recent activity (within the last three days) on January 30, 2009. Their activity and audience was measured every six days (+ or - one day, and with some gaps, henceforth "period") from January 30, 2009 to March 30, 2010. Data collection was performed using Screenscraper (ekiwi, LLC, 2011) under an academic license. In this paper, we limit ourselves to the analysis of the *1,347* bloggers with complete data on their number of readers over the *59* weeks of data collection and who showed some activity (either adding "friends", making comments or posting entries) in at least *90%* of the periods. This therefore excludes *463* bloggers that showed activity in less than *10%* of the sample periods. This also excludes *487* blogs that did not show their number of readers, either because they elected to show only readers whom they also read back (*258*) or because they chose to hide this statistic (*229*).

Over the collection periods, we gathered data on the number of "friends" (*Friends*), that is blogs read by a user on LJ, and number of "friend of" (*Readers*), that is blogs reading

---

[7]Yandex Press release, September 26, 2006 (http://company.yandex.com/press_center/press_releases/2006/2006-09-26.xml, accessed October 6, 2010).

[8]Communities that are particularly popular include ohnotheydidnt (celebrity gossip), customers_suck (rant community), adayinmylife (picture diaries, most posts visible to members only), saucydwellings or abandonedplaces (pictures) or bakebakebake (cooking).

the user on LJ. We were not able to gather information about individual characteristics of the bloggers. However, we were able to collect information about the activity of the bloggers as well as on some of the characteristics of the blog (*e.g.* if it is a paid account). More precisely, in relation to a blogger's activity, we collected every period the number of communities joined (*Communities_joined*) or left (*Communities_left*) by the user, the number of entries written by the user (*Entries*), the number of comments made by the user either in communities or on entries in other blogs (*Posted*), the number of comments made by the blog's readers on its entries (*Received*) and the number of weeks since the blog's last update, *i.e.* since the last entry was made (*Inactive*). To control for different blog characteristics, we rely on: the date on which the account was set up to compute the age of the blog, in weeks, (*Age_blog*), the range of an account's functionalities (*Functionality*), which depends on whether the account is Basic, Early, Sponsored, Plus, Paid, or Permanent, and the country where the blogger is located.

Table (1) contains our description of the variables used in the analysis while Table (2) reports descriptive statistics (average during the collection period):

[INSERT TABLE (1) HERE]

[INSERT TABLE (2) HERE]

From these descriptive statistics, it is already interesting to note that while the median number of friends and readers are equal, there is a greater variation in the numbers of one's readers.

The two main communities on LJ are Russian and English speaking blogs, which we define according to their location. Blogs from Australia, Canada, the UK and the US are classified as English-speaking (*"English"*), while blogs from Ukraine, Russia and Belarus are classified as Russian-speaking (*"Russian"*). Other countries of origin constitute the residual category *"Other"*, while those bloggers that did not reveal their location are categorized as *"Unknown"*. The averages in the descriptive statistics of the two main communities differ substantially (see Table 3): tests for the mean (not reported) suggest that Russian blogs are larger (both in term of readers and friends), younger, and have less (and cheaper) functionalities. In terms of activity, Russian blogs are on average more active in terms of posting comments and entries, as well as in in joining communities and in eliciting

engagement (*i.e.* comments per entries) from other bloggers, and they tend to stay inactive for shorter periods of time.

These differences can also be seen by further dividing the sample according to the age of the blog: *young, mature* and *old*. Table (4) shows the number of blogs in each language category and age-group, as well as their average number of readers and friends during the sample period:

[INSERT TABLE (4) HERE]

As can be seen above, the sample is about equally divided overall between Russian (*434*) and English (*473*) language bloggers, but Russian blogs tend to have been more recently created, reflecting the later emergence of blogging in Russia, and also tend to have more readers than English blogs. Overall, older blogs tend to have more readers, a pattern which is repeated across both language communities.

6. THE ECONOMETRIC MODEL

Estimating equation (8) and (9) in order to get the reciprocity parameters is quite challenging. In fact, the main regressors of interest (*i.e.* the number of friends and readers) cannot be considered as being exogenous (or weakly exogenous), as they are jointly determined by the activity of the blogger. In addition, for several bloggers, these variables have high persistence, thus causing an additional weak-instruments concern.[9]

Nevertheless, by relying on various specification of system generalized method of moments (GMM) (Arellano and Bover, 1995; Blundell and Bond, 1998; Roodman, 2009a), and by paying attention to weak identification (Bobba and Coviello, 2007) and instruments proliferation (Roodman, 2009b), we obtain results in line with theoretical predictions.

We consider the following dynamic specification to identify the reciprocity parameter ρ for readership, that is, the proportion of one's new "friends" who reciprocate readership:

(10)

$$Readers_{it} = \alpha_r Readers_{i(t-1)} + \rho f_{it} + \beta_1 Activity_{it} + \beta_2 Blog\_Characteristics_{it} + a_i + \mu_t + \epsilon_{it}$$

---

[9]We perform various tests in order to check the presence of unit roots in the data: Im, Pesaran, and Shin (2003)'s, Choi (2001)'s and Harris and Tzavalis (1999)'s statistics. Although we can reject the hypothesis of unit-root for the Friends and Readers series in various specification, the series are highly-persistent (with the autocorrelation parameter being above 0.8 and 0.9 respectively).

where $Readers_{it}$ is the number of readers at time t and $f_{it} = log(Friends_{it}) - log(Friends_{i(t-1)})$ is the variation in the total number of friends between time $t-1$ and t as defined in equation (8) (in logs). The variables in $Activity_{it}$ aim to capture the investment activity (A_{it}) of the blogger other than seeking out and adding "friends", and consist of the number of community joined ($Community\_joined_{it}$) or left ($Community\_left_{it}$) by the user, and the number of comments posted ($Posted_{it}$) and entries made by the user ($Entries_{it}$). In this group of variables, two indicators are also included in order to measure the extent of a blogger's interactions with his/her readers, namely $Engagement_{it} = Received_{it}/Entries_{it}$ and $Extroversion_{it} = Posted_{it}/Received_{it}$. The former measures the number of comments the blogger received from his/her readers on each of his/her entries, whereas the latter compares the number of comments posted by the user to the number of comments received from the blogger's readers. The aim of the first indicator is to capture how interesting the blogger's entries are and/or how engaged the readers are (in so far as interesting entries attract more comments and engaged readers make more comments), while the second indicator signals how extroverted the blogger is (in so far as extroverted bloggers will post many comments on other bloggers entries or in communities rather than merely replying to comments received on their own entries). Finally, we also include in the regression a variable ($Inactive_{it}$) which captures how many weeks went by since a blogger's last post. The group of variables $Blog\_Characteristics_{it}$ include characteristics of the blog that may affect its activity, and are either fixed or slow-changing variables. Specifically this group comprises of a categorical variable related to the type of account (i.e $Functionality_{it}$), the language of the blog (*Russian, English, Other* or *Unknown*) and the length of time since its creation (Age_blog_{it}). Of the error components, μ_t is a period-specific intercept, a_i is an unobserved time-invariant blog-specific effect, and ϵ_{it} reflects serially uncorrelated errors.

Similarly, we consider the following dynamic specification to identify the reciprocity parameter θ for friendship, that is the proportion of one's new readers whose readership one reciprocates:

(11)

$$Friends_{it} = \alpha_f Friends_{i(t-1)} + \theta r_{it} + \beta_3 Activity_{it} + \beta_4 Blog\_Characteristics_{it} + a_i + \mu_t + \epsilon_{it}$$

where $r_{it} = log(Readers_{it}) - log(Readers_{i(t-1)})$ is the variation in the total number of readers between time $t - 1$ and t (in logs). We are interested in consistent estimation of the parameters (θ, ρ). Table (5) and (6) report results for the estimation of the reciprocity coefficients for a range of estimators with known properties in dynamic panel data. In particular, in the presence of individual-specific effects, OLS levels is expected to give an upwards-biased estimate of the coefficient on the lagged dependent variable, whereas the fixed-effects estimator is expected to give a downwards-biased estimate of this coefficient. These estimates in columns (1) and (2) provide a value for the reciprocity parameter ρ that is between *16.00* and *17.78*, (that is, of *100* additional added friends, at least *16* will reciprocate the user's friendship), whereas the value for the reciprocity parameter θ is between *48.87* and *54.74* (that is, of *100* additional readers of one's blog, at least *49* will see their readership reciprocated by being added to the user's list of friends). Those values also suggest that bloggers in our sample are more likely to reciprocate readership than they are to see their readership reciprocated. The reciprocity parameters ρ and θ ought to be the same in a closed network where all members are identical, but our bloggers differ in type (some may be passive, others may be centers of a network of friends), and our system is not closed, that is, users keep on joining or leaving LJ over our collection period. This might explain the discrepancy.

Blundell and Bond (1998) showed that weak instruments could cause large finite-sample biases when using the first-differenced GMM procedure to estimate autoregressive models for moderately persistent series. Therefore, in columns (3) we focus on "System-GMM" estimation, which uses lagged first-differences as instruments for equations in levels in addition to the usual lagged levels as instruments for equations in first-differences. Relying on these estimators we find reasonable parameter estimates. The estimated coefficient on the lagged dependent variable is higher than the fixed-effect estimator, and below the OLS levels estimate for both α_f and α_r . The estimate for ρ (*10.40*) is closer to that of θ (*15.52*) when estimating with System-GMM.

To check the validity of our instruments, we rely on the Hansen *J*-test statistics, which is not only a test of instrument validity but can also be viewed as a test of structural specification. Whenever important explanatory variable are left out, important components

of variation are moved into the error terms making them correlated with the instruments. According to this statistics, instruments (lagged level and differences) dated *t-3* up to *t-5* are accepted for *Readers* and *Friends*, and dated *t-1* up to *t-3* for *Activity*. However, by being numerous, instruments can overfit instrumented variables, thus failing to wipe out the endogenous components and biasing coefficient estimates (Roodman, 2009b). That is, the Hansen test may be vitiated by instrument proliferation, which is signaled by too high *p*-value of the statistics (sometime as high as implausible *p*-values of *1.000*). We therefore carefully check the value of the statistics across different specification of the model. In our preferred specification, the test of common restrictions (see Hans test of overid.) is passed in System-GMM results at the *10%* level.

Closely related to the Hansen *J*-test for validity of the full instrument set is the *difference-*in-Hansen test, which allows to test the validity of a *subset* of instruments by computing the increase in *J* when the given subset of instruments is added to the estimation set-up. This difference test can also be weakened by a high instrument count. We check the robustness of our specification by testing different subset of instruments (diff level, diff Hansen 2, diff Hansen 3 and diff Hansen 4). These tests suggest the validity of our specification at the *10%* level.[10]

The parameters relating to the activity of the bloggers are also in line with expectations from the model, that is, activity contribute to increases in one's number of readers and friends. In particular, in the readers equation, the coefficient for the variable accounting for the number of comments posted (*Posted*) is positive and significant. Relatedly, the coefficient on extroversion is negative (although not significant in the GMM specification), suggesting that those bloggers who are too active compared with their readers in posting comments will then end up having fewer readers. Similarly, the number of comments made by the readers *per* entries made by the user (*Engagement*) turn out to be an important variable: the more interesting or provocative the blogger's entries, the higher the number of readers. The number of comments posted and the degree of a blogger's extroversion have a similar impact in the *"Friends"* equation, although in this case they completely offset each

---

[10]The Sargan and difference-in-Sargan tests are not so vulnerable to instrument proliferation as they do not depend on an estimate of the optimal weighting matrix. However, they require homoskedastic errors for consistency, which can hardly be assumed in this context (Roodman, 2009b).

other: the higher the number of comments posted, the higher the number of friends, but making too many comments compared to those received has an equivalent negative effect. Making more comments is thus more effective in gaining readers when it is compensated by receiving more comments as well.

Also in line with the expectations, we consistently observe across specifications a negative effect of the variable measuring how long the blogger has been inactive (*Inactive*): The longer a blogger has been inactive, the lower his number of friends and readers. This effect is significant in the *"Friends"* equation, probably because inactivity has a more direct effect on one's adding of new friends: Those who do not post entries are also likely not to be using their account at all, and thus not to add friends. The effect on one's number of readers is less direct, as readers will "drop" one only after a long period of inactivity. There is robust evidence a blog's level of functionalities having an effect: the higher the number of functionality in the blog (which also means the account becomes more costly), the higher the numbers of readers and friends. This latter result may also mean that having a costly account captures a blogger's commitment to the activity of blogging, that is, bloggers that are more strongly invested in the activity of blogging will be readier to pay for their account, and will also have more readers and friends.

A possible concern with the estimated results for the reciprocity parameters is that they may be driven by the presence in our sample of country or blog-size effects. We investigate this issues by splitting the sample in different sub-groups. We report estimations for the subgroup of Russian blogs in column (*1*) of Tables (7) and (8), in columns (*2*) for big blogs (*i.e.* blogs with an average number of Readers above *150*), in columns (*3*) for small and medium blogs (*i.e.* blogs with an average number of Readers equal or below *150*), and we control for network effects in columns (*4*) by introducing the (lagged) square level of readers (or friends). Overall, these results are consistent with the previous ones. A Wald test confirms that there is no significant difference in the reciprocity parameters between Russian and English blogs. This means that despite differences between Russian, more media-oriented blogs, *vs.* English, more social-networking-oriented blogs, those differences do not translate in higher or lower willingness to reciprocate readership. In terms of differences between bigger and smaller blogs (columns (*2*) and (*3*)), bigger blogs seem

to be less successful in getting new friends to reciprocate readership (Table 7), but are more likely to reciprocate the friendship of new readers (Table 8). This would seem to indicate that bigger blogs thrive not so much by adding friends (those new friends tend not to reciprocate), but by being readier to reciprocate the readership of others. With reference to our introductory quote, we could tentatively say, therefore, that more popular blogs are not so much "more loved" as "more loving". The discrepancy in the reciprocation ratios may also be interpreted in view of the greater variability in readership than in friendship among bigger blogs. Big blogs would not actively search for new friends to add, resulting in low success rates in terms of obtaining reciprocation when they do add friends, but instead would rely on new readers finding them, whose readership they would then automatically reciprocate. Finally, looking at column (*4*) in both tables, blog readership seems to benefit from network effects (more readers begets more readers), while the number of friends does not seem to increase at a higher rate as the number of friends increases.

7. CONCLUSION

This paper combines an original data set – the first panel following the activity of bloggers over time, to our knowledge – with an original model – an adaptation of a standard capital investment model to study the effect of reciprocation on social capital building. Results of our analysis confirm the intuition behind this paper. Both activity and reciprocity play a role in the formation of social capital so that studying one without data on the other leads to incorrect evaluation of their importance. This research is therefore an important contribution to a better understanding of social network formation, which would not have been possible without the fine-grained data we collected online. We expect our approach to be of use when comparing different kind of social networks; those that are primarily affinity-based, where reciprocity is important, *vs.* those that are primarily activity-based, where reciprocity would play less of a role. Already, we were able to evidence differences between popular and less popular blogs in that respect, though not when comparing Russian and English-speaking blogs. Being the first to develop an analysis of reciprocation as a factor in the growth of social network, we do not have references points for our reciprocation parameters ρ and θ, so that we do not know if they are high or low compared to other social media. However, we think that LiveJournal is merely a point on a continuum

that spans from Twitter to Facebook in terms of how important activity (media aspect) is compared to reciprocation (social network aspect). We would expect Twitter account to exhibit low levels of reciprocation parameters while Facebook users would presumably display high levels of reciprocation – though Facebook only reveals reciprocated links so that one would need to have access to company data to verify the above. In that sense, our work provides a conceptual and analytical tool to better understand variety in social media and locate its different manifestations along the range between social networking and media activities.

REFERENCES

Anderson, Chris. 2004. "The long tail." *Wired* 12 (10). (ref. p. 8).

Antin, Judd and Matthew Earp. 2010. "With a little help from my friends: Self-interested and prosocial behavior on MySpace Music." *Journal of the American Society for Information Science and Technology* 61 (5):952–963. (ref. p. 12).

Arellano, Manuel and Olympia Bover. 1995. "Another look at the instrumental variable estimation of error-components models." *Journal of Econometrics* 68 (1):29–51. (ref. p. 18).

Aristotle. ca. 350 B.C.E. *The Nicomachean Ethics*. Oxford University Press. (1908 translation by W. D. Ross). (ref. p. 3).

Bar-Ilan, Judit. 2005. "Information hub blogs." *Journal of Information Science* 31 (4):297–307. (ref. p. 6).

Blundell, Richard and Stephen Bond. 1998. "Initial conditions and moment restrictions in dynamic panel data models." *Journal of Econometrics* 87 (1):115–143. (ref. p. 18 and 20).

Bobba, Matteo and Decio Coviello. 2007. "Weak instruments and weak identification, in estimating the effects of education, on democracy." *Economics Letters* 96 (3):301–306. (ref. p. 18).

Chan, Kimmy Wa and Stella Yiyan Li. 2009. "Understanding consumer-to-consumer interactions in virtual communities: The salience of reciprocity." *Journal of Business Research* 63 (9-10):1033–1040. (ref. p. 10).

Choi, In. 2001. "Unit root tests for panel data." *Journal of International Money and Finance* 20 (2):249–272. (ref. p. 18).

Chun, Hyunwoo, Haewoon Kwak, Youg-Ho Eom, Yong-Yeol Ahn, Sue Moon, and Hayoong Jeong. 2008. "Comparison of online social relations in terms of volume vs. interaction: A case study of Cyworld." In *Proceedings of the 8th ACM SIGCOMM conference on Internet measurement*. 57–70. (ref. p. 10).

Conte, Anna, Daniela Di Cagno, and Emanuela Sciubba. 2009. "Strategies in social network formation." *Jena Economic Research Papers, 2009-095* . (ref. p. 9).

de Montaigne, Michel E. 1588. *Essays*, chap. XXVII - Of Friendship. Project Gutenberg. URL `http://www.gutenberg.org/ebooks/3600`. (1877 translation by C. Cotton). (ref. p. 3).

Doreian, Patrick, Roman Kapuscinski, David Krackhardt, and Janusz Szczypula. 1996. "A brief history of balance through time." *The Journal of Mathematical Sociology* 21 (1-2):113–131. (ref. p. 8).

Dunbar, Robin I.M. 1992. "Neocortex size as a constraint on group size in primates." *Journal of Human Evolution* 22 (6):469 – 493. (ref. p. 6).

ekiwi, LLC. 2011. *Screen-scraper, version 5.0, enterprise edition*. Provo, UT, USA. URL `http://www.screen-scraper.com`. (ref. p. 1 and 16).

Ellison, Nicole B., Charles Steinfield, and Cliff Lampe. 2007. "The benefits of Facebook "Friends": Social capital and college students' use of online social network sites." *Journal of Computer-Mediated Communication* 12 (4). (ref. p. 12).

Falkinger, Josef. 2007. "Attention economies." *Journal of Economic Theory* 133 (1):266–294. (ref. p. 11).

Fehr, Ernst and Simon Gächter. 2000. "Fairness and retaliation: The economics of reciprocity." *Journal of Economic Perspective* 14 (3):159–181. (ref. p. 2).

Galeotti, Andrea and Sanjeev Goyal. 2010. "The law of the few." *American Economic Review* 100 (4):1468–1492. (ref. p. 11 and 12).

Gintis, Herbert, Joseph Henrich, Samuel Bowles, Robert Boyd, and Ernst Fehr. 2008. "Strong reciprocity and the roots of human morality." *Social Justice Research* 21 (2):241–253. (ref. p. 2).

Glaeser, Edward L., David Laibson, and Bruce Sacerdote. 2002. "An economic approach to social capital." *The Economic Journal* 112 (483):437–458. (ref. p. 5 and 12).

Gorny, Eugene. 2006. "Russian LiveJournal: The impact of identity on the development of a virtual community." In *Control + Shift. Public and Private Usages of the Russian Internet*, edited by Henrike Schmidt, Katy Teubener, and Natalja Konradova. Norderstedt: Books on Demand. (ref. p. 7).

Gouldner, Alvin W. 1960. "The norm of reciprocity: A preliminary statement." *American Sociological Review* 25 (2):161–178. (ref. p. 5).

Gu, Bin, Yun Huang, Wenjing Duan, and Andrew B. Whinston. 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* . (ref. p. 10).

Hallinan, Maureen T. 1978. "The process of friendship formation." *Social Networks* 1 (2):193–210. (ref. p. 9).

Harmsen-van Hout, Marjolein J.W., P. Jean-Jacques Herings, and Benedict G.C. Dellaert. 2010. "The structure of online consumer communication networks." *FCN Working Paper No. 3/2010* . (ref. p. 12).

Harris, Richard D. F. and Elias Tzavalis. 1999. "Inference for unit roots in dynamic panels where the time dimension is fixed." *Journal of Econometrics* 91 (2):201–226. (ref. p. 18).

Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, and Richard McElreath. 2001. "In search of Homo Economicus: Behavioral experiments in 15 small-scale societies." *The American Economic Review* 91 (2):73–78. (ref. p. 2).

Herring, Susan C., Inna Kouper, John C. Paolillo, Lois A. Scheidt, Michael Tyworth, Peter Welsch, Elijah Wright, and Ning Yu. 2005a. "Conversations in the blogosphere: An analysis "From the bottom up"." In *Proceedings of the 38th Hawaii International Conference on System Science*. (ref. p. 8).

Herring, Susan C., Lois Ann Scheidt, Elijah Wright, and Sabrina Bonus. 2005b. "Weblogs as a bridging genre." *Information Technology & People* 18 (2):142–171. (ref. p. 6).

Hofstetter, Reto, Scott K. Shriver, Harikesh S. Nair, and Klaus Miller. 2009. "Social ties and user generated content: Evidence from an online social network." *NET Institute, Working Paper 09-28* . (ref. p. 7).

Huang, Chun-Yao, Yong-Zheng Shen, Hong-Xiang Lin, and Shin-Shin Chang. 2007. "Bloggers' motivations and behaviors: A model." *Journal Of Advertising Research* 47 (4):472–484. (ref. p. 7).

Huberman, Bernardo A., Daniel M. Romero, and Fang Wu. 2009. "Crowdsourcing, attention and productivity." *Journal of Information Science* 35 (6):758–765. (ref. p. 7).

Im, Kyung So, M. Hashem Pesaran, and Yongcheol Shin. 2003. "Testing for unit roots in heterogeneous panels." *Journal of Econometrics* 115 (1):53–74. (ref. p. 18).

Jackson, Matthew O. and Asher Wolinsky. 1996. "A strategic model of social and economic networks." *Journal of Economic Theory* 71 (1):44–74. (ref. p. 11 and 12).

Kumar, Ravi, Jasmine Novak, and Andrew Tomkins. 2010. "Structure and evolution of online social networks." In *Link Mining: Models, Algorithms, and Applications*, edited by Philip S. S. Yu, Jiawei Han, and Christos Faloutsos. Springer: New York, 337–357. (ref. p. 8).

Lemann, Nicholas. 2006. "Journalism without journalists." *The New Yorker* URL `http://www.newyorker.com/archive/2006/08/07/060807fa_fact1`. Accessed January 18, 2010. (ref. p. 6).

Lenhart, Amanda and Susannah Fox. 2006. "Bloggers: A portrait of the Internet's new storytellers." URL `http://www.pewinternet.org/Reports/2006/Bloggers.aspx`. Accessed October 18, 2010. (ref. p. 7).

Lento, Thomas, Howard T. Welser, Lei Gu, and Marc Smith. 2006. "The ties that blog: Examining the relationship between social ties and continued participation in the Wallop weblogging system." In *3rd Annual Workshop on the Weblogging Ecosystem*. (ref. p. 12).

Lin, Nan. 2001. "Building a network theory of social capital." In *Social Capital: Theory and Research*, edited by Nan Lin, Karen S. Cook, and Ronald S. Burt, chap. 1. Aldine de Gruyter, 6–7. (ref. p. 3).

LiveJournal. 2011a. "What are Friends? How do I manage my Friends list? Frequently Asked Question #61." URL `http://www.livejournal.com/support/faqbrowse`.

`bml?faqid=61`. Accessed January 17, 2011. (ref. p. 9).

———. 2011b. "What is a community? Frequently Asked Question #77." URL `http://www.livejournal.com/support/faqbrowse.bml?faqid=77&view=full`. Accessed January 17, 2011. (ref. p. 5).

Lussier, Jake, Troy Raeder, and Nitesh Chawla. 2010. "User generated content consumption and social networking in knowledge-sharing OSNs." In *Advances in Social Computing, Lecture Notes in Computer Science*, vol. 6007, edited by Sun-Ki Chai, John Salerno, and Patricia Mabry. Springer Berlin / Heidelberg, 228–237. (ref. p. 10 and 11).

Marlow, Cameron A. 2006. "Linking without thinking: Weblogs, readership and online social capital formation." In *Proceedings of the 56th Annual Conference of the International Communication Association*. (ref. p. 7, 9, and 12).

Marwick, Alice. 2008. "LiveJournal users: Passionate, prolific and private." URL `http://www.livejournalinc.com/LJ_Research_Report.pdf`. Accessed January 18, 2010. (ref. p. 6 and 9).

McKenna, Laura and Antoinette Pole. 2008. "What do bloggers do: an average day on an average political blog." *Public Choice* 134 (1):97–108. (ref. p. 6).

Ozer, Daniel J. and Verónica Benet-Martínez. 2006. "Personality and the prediction of consequential outcomes." *Annual Review of Psychology* 57 (1):401–421. (ref. p. 3).

Raynes-Goldie, Kate and David Fono. 2006. "Hyperfriends and beyond: Friendship and social norms on LiveJournal." In *Internet Research Annual Volume 4: Selected Papers from the Association of Internet Researchers Conference*, edited by M. Consalvo and C. Haythornthwaite. Peter Lang: New York, USA. (ref. p. 9).

Roodman, David. 2009a. "How to do xtabond2: An introduction to difference and system GMM in Stata." *Stata Journal* 9 (1):86–136. (ref. p. 18).

———. 2009b. "A note on the theme of too many instruments." *Oxford Bulletin of Economics and Statistics* 71 (1):135–158. (ref. p. 18 and 21).

Rui, Huaxia and Andrew Whinston. 2010. "Social media as an innovation - the case of Twitter." University of Texas at Austin. (ref. p. 8).

Sadlon, Elliot, Yasuaki Sakamoto, Hannah J. Dever, and Jeffrey V. Nickerson. 2008. "The karma of Digg: Reciprocity in online social networks." In *Proceedings of the 18th Annual*

Workshop on Information Technologies and Systems. (ref. p. 10).

Schaefer, David R., John M. Light, Richard A. Fabes, Laura D. Hanish, and Carol L. Martin. 2010. "Fundamental principles of network formation among preschool children." *Social Networks* 32 (1):61 – 71. (ref. p. 8).

Schmidt, Jan. 2007. "Blogging practices: An analytical framework." *Journal of Computer-Mediated Communication* 12 (4):1409–1427. (ref. p. 6).

Shen, Wenqi. 2009. "Competing for attention in online reviews." *AMCIS 2009 Doctoral Consortium. Paper 2* URL `http://aisel.aisnet.org/amcis2009_dc/2`. Accessed January 18, 2010. (ref. p. 4).

Shirky, Clay. 2003. "Power laws, weblogs, and inequality." URL `http://shirky.com/writings/powerlaw_weblog.html`. Accessed January 18, 2010. (ref. p. 8).

StataCorp LP. 2011. *Stata Statistical Software: Release 11*. College Station, TX, USA. URL `http://www.stata.com`. (ref. p. 1).

Technorati. 2009. "State of the Blogosphere 2009." URL `http://technorati.com/blogging/feature/state-of-the-blogosphere-2009`. Accessed October 18, 2010. (ref. p. 6 and 7).

Vaquera, Elizabeth and Grace Kao. 2008. "Do you like me as much as I like you? Friendship reciprocity and its effects on school outcomes among adolescents." *Social Science Research* 37 (1):55–72. (ref. p. 9).

Wei, Lu. 2009. "Filter blogs *vs.* personal journals: Understanding the knowledge production gap on the Internet." *Journal of Computer-Mediated Communication* 14 (3):532–558. (ref. p. 6).

Wikipedia. 2011. "Glossary of blogging." URL `http://en.wikipedia.org/w/index.php?title=Glossary_of_blogging&oldid=407505604`. Accessed January 17th, 2011. (ref. p. 4 and 5).

Yu, He. 2007. *Exploring the Chinese blogosphere: The motivations of blog authors and readers*. Master's thesis, National University of Singapore. URL `https://scholarbank.nus.edu.sg/handle/10635/13159`. Accessed January 27, 2011. (ref. p. 7).

Zhang, Xiaoquan (Michael) and Feng Zhu. 2010. "Group size and incentives to contribute: A natural experiment at Chinese Wikipedia." *NET Institute Working Paper No. 07-22,*

forthcoming AER . (ref. p. 8).

TABLE 1. Variable Description

Variable	Description
Readers	The number of blogs reading the user at time t.
Friends	The number of blogs read at time t by the user.
Community joined	The number of communities joined by the user between time $t-1$ and t.
Community left	The number of communities left by the user between time $t-1$ and t.
Entries	The number of entries made by the user between between time $t-1$ and t.
Posted	The number of comments posted by the user between between time $t-1$ and t.
Received	The number of comments received by the user between between time $t-1$ and t.
Functionality	A categorical variable equal to 1 if the blog is a Basic account (free, limited advertising), to 2 if the blog is an Early account (created before mid-September 2000), to 3 if the blog is Sponsored by a company that is in partnership with LiveJournal or if the blog is a Plus account (free but with advertising, more features than Basic but less than Paid), to 4 if the blog is a Paid account (no advertising, access to all features of LJ), and to 5 if the blog is Permanent (either paid forever in a lump sum, or given for services to the LiveJournal project)
Age of the blog	Weeks since the date of creation of the blog
Extroversion	The number of comments posted by the user relative to the number of comments received from the blogger's readers.
Engagement	The number of comments received from the blogger's readers relative to the number of entries made by the user, *i.e.* how many comments each entry receives on average.
Inactive	The number of weeks since the blog's last update, i.e. since the last entry was made.
English	A dummy variable equal to 1 if the language of the blog is English, 0 otherwise.
Russian	A dummy variable equal to 1 if the language of the blog is Russian, 0 otherwise.

TABLE 2. Summary statistics

Variable	Mean	Std Dev.	Median	Min	Max
Readers	189.683	477.492	68	1	10921
Friends	140.174	217.885	69	0	1958
Δlog_Readers	.004	0.041	0	-1.43	1.783
Δlog_Friends	.004	.058	0	-6.957	6.970
Functionality	2.579	1.230	3	1	5
Age_blog	230.452	102.491	220	1	514
Community_joined	.151	.999	0	0	161
Community_left	.099	2.087	0	0	261
Entries	5.540	18.786	2	0	1139
Posted	25.40	64.018	5	0	2982
Extroversion	2.547	10.827	1	0	967
Engagement	3.765	14.256	1	0	3362
Inactive	1.928	5.763	0	0	48
Russian	.421	.494	0	0	1
English	.458	.498	0	0	1
Other	.121	.326	0	0	1

Numbers of bloggers: *1347*.

Number of weeks: *59*.

TABLE 3. Summary statistics by nationality

Variable	Russian			English		
	Mean	Std Dev.	Median	Mean	Std Dev.	Median
Readers	330.781	661.201	132	100.335	203.583	55
Friends	232.892	294.346	127	85.292	97.285	56
Δlog_Readers	0.006	0.049	0	0.002	0.025	0
Δlog_Friends	0.005	0.090	0	0.002	0.023	0
Functionality	2.407	1.093	3	2.769	1.341	3
Age_blog	205.483	85.154	200	276.827	101.077	284
Community_joined	0.179	1.304	0	0.116	0.677	0
Community_left	0.128	2.422	0	0.083	2.226	0
Entries	7.186	22.74	2	4.356	10.117	2
Posted	35.979	70.398	10	17.545	42.001	4
Extroversion	2.330	8.492	1	2.824	14.166	1
Engagement	4.953	8.439	2	2.808	6.129	1
Inactive	1.550	5.012	0	1.947	5.772	0
Russian	1	0	1	0	0	0
English	0	0	0	1	0	1
Other	0	0	0	0	0	0

Number of bloggers: 434 Russian, 473 English

Number of weeks: 59

TABLE 4. Blog age, size and nationality

BLOG AGE		Unknown	Other	Russian	English	*Total*
			NATIONALITY			
Young	*Readers**	68.22	87.43	152.09	46.38	
	Friends*	63.24	80.67	153.65	58.75	
		(49)	(21)	(43)	(24)	(137)
Mature	*Readers**	189.33	167.73	372.86	91.78	
	Friends*	120.29	153.54	264.99	84.30	
		(238)	(95)	(382)	(352)	(1067)
Old	*Readers**	136.00	409.22	1741.00	174.97	
	Friends*	112.18	223.22	382.78	111.33	
		(28)	(9)	(9)	(97)	(143)
Total		(315)	(125)	(434)	(473)	(1347)

Young is a blog created no more than *125* weeks ago by the end of the sample period.

Mature is a blog created between *125* weeks and *400* weeks ago by the end of the sample period.

Old is a blog created more than *400* weeks ago by the end of the sample period.

* Average size over the sample period.

( ) Number of bloggers in the category.

TABLE 5. Friendship reciprocation. Dependent variable: Readers

VARIABLES	(1) Pooled OLS	(2) FE	(3) SYS GMM
L.Readers	1.004***	0.986***	0.999***
	(0.000437)	(0.00449)	(0.000556)
Δlog_Friends	17.78**	16.00**	10.40***
	(7.720)	(7.067)	(3.060)
Functionality	0.0256	0.469***	9.784***
	(0.0281)	(0.177)	(1.104)
Age_blog	-0.00312***	0.0195***	0.0101*
	(0.000450)	(0.00440)	(0.00609)
Community_joined	0.238**	0.207*	0.680*
	(0.119)	(0.107)	(0.409)
Community_left	0.0325	0.0392	0.414***
	(0.0392)	(0.0378)	(0.0821)
Entries	0.00948**	0.0101*	-0.0147
	(0.00477)	(0.00527)	(0.0139)
Posted	0.0109***	0.0218***	0.0640***
	(0.00247)	(0.00334)	(0.00332)
Extroversion	-0.0136***	-0.0152***	-0.0175
	(0.00280)	(0.00307)	(0.0178)
Engagement	0.0111	0.0119	0.0302*
	(0.0123)	(0.0106)	(0.0163)
Inactive	0.00446	-0.0182***	-0.388
	(0.00446)	(0.00399)	(0.282)
Other	0.164		0.110
	(0.176)		(1.866)
Russian	0.0493		0.112
	(0.134)		(1.362)
English	0.120		-3.511**
	(0.0912)		(1.606)
Constant	-0.0461	-3.038**	-26.57***
	(0.186)	(1.356)	(3.919)
Observations	75,432	75,432	75,432
Number of user	1,347	1,347	1,347
Hansen test of overid.			0.135
AR(1) in first differences			0.003
AR(2) in first differences			0.162
diff Hansen level			0.144
diff Hansen 2			0.082
diff Hansen 3			0.035
diff Hansen 4			0.710

TABLE 6. Readership reciprocation. Dependent variable: Friends

VARIABLES	(1) Pooled OLS	(2) FE	(3) SYS GMM
L.Friends	1.000***	0.932***	0.952***
	(0.00114)	(0.0152)	(0.0122)
Δlog_Readers	54.74***	48.87***	15.52***
	(12.80)	(11.91)	(4.975)
Functionality	0.0632**	0.756***	0.770**
	(0.0296)	(0.260)	(0.386)
Age_blog	-0.000460	0.0284**	0.0105**
	(0.000547)	(0.0138)	(0.00442)
Community_joined	0.595*	0.469*	0.352**
	(0.314)	(0.284)	(0.159)
Community_left	0.0576	0.0669	-0.0742
	(0.283)	(0.282)	(0.120)
Entries	0.00450	0.00485	-0.000393
	(0.00414)	(0.00480)	(0.00617)
Posted	0.00271	0.00791**	0.0139***
	(0.00175)	(0.00366)	(0.00219)
Extroversion	-0.0115*	-0.0126*	-0.0139***
	(0.00586)	(0.00670)	(0.00338)
Engagement	0.00499	0.00576	-0.00250
	(0.00581)	(0.00586)	(0.00364)
Inactive	-0.00469	-0.0280***	-0.107***
	(0.00530)	(0.00737)	(0.0361)
Other	0.237		0.880
	(0.253)		(1.160)
Russian	0.362		5.668***
	(0.225)		(1.857)
English	-0.0253		-1.293
	(0.105)		(0.868)
Constant	-0.100	1.307	0.545
	(0.159)	(3.527)	(1.111)
Observations	75,432	75,432	75,432
Number of user	1,347	1,347	1,347
Hansen test of overid.			0.334
AR(1) in first differences			0.010
AR(2) in first differences			0.191
diff Hansen level			0.201
diff Hansen 2			0.084
diff Hansen 3			0.141
diff Hansen 4			0.398
diff Hansen 5			0.270

TABLE 7. Friendship reciprocation: robustness checks. Dependent variable: Readers

VARIABLES	(1) SYS GMM Russian	(2) SYS GMM Big Blog	(3) SYS GMM Small-Med Blog	(4) SYS GMM Network Effect
L.Readers	1.001***	0.995***	0.990***	0.971***
	(0.000398)	(0.000793)	(0.0128)	(0.00613)
L.Readers_square				4.31e-06***
				(5.35e-07)
Δlog_Friends	9.792***	21.48***	42.04***	20.17***
	(3.150)	(3.238)	(7.085)	(4.440)
Functionality	10.41***	17.81***	0.0345	-4.724
	(1.049)	(1.802)	(0.0374)	(9.745)
Age_blog	0.0352***	0.0216	0.00155	0.0357
	(0.00837)	(0.0153)	(0.00155)	(0.0692)
Community_joined	0.0538	-0.342	-0.0107	0.123*
	(0.0332)	(0.293)	(0.0264)	(0.0645)
Community_left	0.00335	0.752***	-0.000705	-0.00698
	(0.0212)	(0.0886)	(0.0357)	(0.0182)
Entries	-0.00495	0.479***	-0.00277**	0.0166*
	(0.00368)	(0.0552)	(0.00137)	(0.00913)
Posted	0.00807***	0.0236***	0.00441***	0.0197***
	(0.00238)	(0.00452)	(0.00118)	(0.00255)
Extroversion	0.0177	0.0784***	-0.00125	-0.00510
	(0.0112)	(0.0173)	(0.00233)	(0.00557)
Engagement	-0.00187	0.137***	0.00149	0.00758***
	(0.00725)	(0.0186)	(0.00682)	(0.00262)
Inactive	-2.095***	-0.673	-0.0130	-0.262
	(0.315)	(0.453)	(0.0134)	(0.369)
Other		15.46***	0.167	41.02
		(4.986)	(0.123)	(136.2)
Russian		8.596**	0.277	65.92
		(3.490)	(0.326)	(49.71)
English		-15.66***	-0.0179	17.96
		(5.516)	(0.0317)	(61.83)
Constant	-25.47***	-53.27***	0.0959	-22.74
	(3.930)	(7.913)	(0.284)	(33.39)
Observations	24,304	20,384	55,048	75,432
Number of user	434	364	983	1,347
Hansen test of overid.	0.392	0.135	0.317	0.446
AR(1) in first differences	0.0191	0.00217	2.02e-05	0.00640
AR(2) in first differences	1.66e-05	0.323	0.682	0.248
diff Hansen level	0.633	0.294	0.558	0.971
diff Hansen 2	0.570	0.00109	0.0142	0.110
diff Hansen 3	0.210	0.0943	0.565	0.0744

TABLE 8. Readership reciprocation: robustness check. Dependent variable: Friends

VARIABLES	(1) SYS GMM Russian	(2) SYS GMM Big Blog	(3) SYS GMM Small-Med Blog	(4) SYS GMM Network Effect
L.Friends	1.000***	0.997***	0.993***	1.022***
	(0.00483)	(0.0225)	(0.00513)	(0.0457)
L.Friends_square				-1.17e-05
				(2.69e-05)
Δlog_Readers	28.45	394.3***	6.627***	12.16*
	(22.28)	(78.64)	(2.166)	(7.188)
Functionality	-0.102	1.127	0.0190	-0.338
	(0.264)	(4.347)	(0.0297)	(1.666)
Age_blog	-0.00166	0.0770	0.000218	-0.0576
	(0.00360)	(0.144)	(0.000537)	(0.132)
Community_joined	0.0965	0.119	0.136**	0.254
	(0.299)	(0.300)	(0.0547)	(0.216)
Community_left	-0.117	-0.166	0.161	0.343
	(0.285)	(0.239)	(0.125)	(0.245)
Entries	-0.00217	-0.0819***	0.00158	0.00220
	(0.0147)	(0.0297)	(0.00335)	(0.00460)
Posted	0.0133***	-0.00135	0.00493***	0.0100***
	(0.00419)	(0.00499)	(0.00127)	(0.00322)
Extroversion	-0.0261***	-0.00643	-0.00212	-0.0121*
	(0.00733)	(0.00822)	(0.00167)	(0.00679)
Engagement	-0.00940	-0.00351**	0.0106	0.000883
	(0.0113)	(0.00146)	(0.00932)	(0.00466)
Inactive	0.000871	-0.257	-0.00779*	-0.0692
	(0.0244)	(0.454)	(0.00408)	(0.198)
Other		-42.49	0.0620	-0.0976
		(60.72)	(0.103)	(7.919)
Russian		-43.57	0.281	-1.343
		(40.18)	(0.181)	(6.325)
English		-48.66	-0.0177	3.749
		(38.31)	(0.0690)	(11.22)
Constant	0.534	16.03	0.321*	13.23
	(0.562)	(53.20)	(0.190)	(36.61)
Observations	24,304	20,384	55,048	75,432
Number of user	434	364	983	1,347
Hansen test of overid.	0.377	0.782	0.304	0.612
AR(1) in first differences	0.0625	0.0131	0.000879	0.0110
AR(2) in first differences	0.218	0.201	0.296	0.191
diff Hansen level	0.156	0.974	0.779	0.472
diff Hansen 2	0.151	0.220	0.492	0.861
diff Hansen 3	0.045	0.984	0.222	0.064