

# MPRA

Munich Personal RePEc Archive

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

Gaudeul, Alexia and Giannetti, Caterina  
GSBC, Friedrich Schiller University, Jena, Royal Holloway,  
University of London

13. October 2011

Online at <http://mpra.ub.uni-muenchen.de/34094/>  
MPRA Paper No. 34094, posted 13. October 2011 / 20:41

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

ALEXIA GAUDEUL<sup>†</sup> AND CATERINA GIANNETTI<sup>‡</sup>

*ABSTRACT.* This paper deals with the role of reciprocation in the formation of individuals' social networks. We follow the activity of a panel of bloggers over more than a year and investigate the extent to which initiating a relation brings about its reciprocation. We adapt a standard capital investment model to study how reciprocation affects the build-up of the individual social capital of bloggers, as measured by their links and interactions with others. This allows us to measure the role of content production and relationship building in the dynamics of online social networks and to distinguish between the social networking and media aspects of blogging.

*Keywords:* Blogs, Friendship, LiveJournal, Reciprocation, Social Capital, Social Networks

*JEL classifications:* C33, D85, L82

---

*Date:* October 13, 2011.

<sup>†</sup> Graduate School "Human Behavior in Social and Economic Change" (GSBC), Friedrich Schiller University, Jena. Email: a.gaudeul@uni-jena.de.

<sup>‡</sup> Royal Holloway, University of London. Email: caterina.giannetti@rhul.ac.uk.

Research was funded by the Federal Programme "ProExzellenz" of the Free State of Thuringia. This paper was presented at the CREM-EconomiX workshop "Digital Business Models: Understanding Strategies" in Paris in June 2010, at a GSBC Brown Bag Seminar in Jena in October 2010, at the North American Summer Meeting of the Econometric Society in St Louis in June 2011, at the 17th International Panel Data Conference in Montreal in July 2011 and at the 38th Annual Conference of the EARIE in Stockholm in September 2011. We benefited from discussions with Yann Bramoullé, Eric Darmon, Avi Goldfarb, Adelina Gschwandtner, Nicolas Jullien, Arianna Martinelli, Kevin Mellet, Alessio Moneta, Chiara Peroni, Raphaël Suire and Xiaoquan (Michael) Zhang. Data collection was performed with screen-scrapers (ekiwi, LLC, 2011). Stata (StataCorp LP, 2011) was used for the statistical analysis.

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.

For that purpose, we study bloggers' choices of which other blogs to subscribe to. Bloggers are emerging as a significant player in the media market, as they disseminate content from mainstream media and also serve as references for many newspapers and television stations. However, blogging also has a social networking aspect, similar to that of Facebook and Google+. For that reason, 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 the first case, reciprocation would be particularly important in maintaining one's network and effort exerted in blogging would have little influence on one's number of readers, while in the second case, readers would not require reciprocation and better and more frequent posting would translate into higher audiences.

We use data collected from LiveJournal, a site where Internet users can keep a blog, and follow the activity of a panel of bloggers over more than a year. We measure how many of a blogger's new subscribers are gained through reciprocation (they read me because I read them), *vs.* what is gained through other activities, such as posting blog entries, making comments, joining communities, *etc.* . . . 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 people subscribe to a

blogger's feed.<sup>1</sup> Bloggers are shown 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 so they keep on reading.

This paper is one of the first to follow the activity of bloggers over time along with their audience. We contribute with this research to a better understanding of social network formation by exploiting fine-grained data collected online. 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, which is affinity-based and where reciprocity is thus important, and media activities, which are oriented towards collecting, producing and disseminating information and where reciprocity plays less of a role. Our work refines our understanding of how reciprocation contributes to the building of human relations by taking account not only of an individual's tendency to reciprocate readership but also of the willingness of others to do the same. Within this context, we compute dynamic multipliers in a multivariate regression models for the analysis of individual social capital. Those dynamic multipliers indicate the impact of changes in the blogger's activity and in blog characteristics (our exogenous variables) on a blogger's social capital (*i.e.* "readership", our endogenous variables).

**Outline:** The paper starts by setting out the context of the present study: why did we choose to study blogging, what is a blog, who are the bloggers, why do they blog, what does the activity of blogging involve, how are blogging networks structured? We pay special attention to the workings of the process of reciprocation in the context of blogging. We then consider various ways to model bloggers' activity, either as motivated by the need to gather information, by the desire to gain attention from others or as a way to access social support. We then present the model that we designed to exploit our data, a panel of bloggers on LiveJournal whose activity was followed week by week over more than a year.

---

<sup>1</sup>"A web feed (or news feed) is a data format used for providing users with frequently updated content", cf. [http://en.wikipedia.org/wiki/Web\\_feed](http://en.wikipedia.org/wiki/Web_feed), accessed 20th of January 2011.

## 2. WHAT IS BLOGGING, AND WHY STUDY IT?

The study of social networks has long been impeded by the difficulty of recording the interactions of individuals over time along with their activity. The emergence of tools for social networking and collaboration *via* the Internet such as Facebook, Google+, LinkedIn, Twitter, Wikipedia or Reddit, has made it possible to collect such data unobtrusively – that is, without the individual knowing their activity is being recorded – and cheaply – by using web-scraping software such as screen-scrapers to extract information from websites. It is now possible to monitor the activity of the nodes in a network along with their formation of new links, which makes it both necessary and possible to elaborate research tools and theories in order to deal with and make sense of such data.

In this context, we decided to focus on bloggers' networks because they have properties that make them particularly well suited for empirical analysis: First, 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,<sup>2</sup> 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. Second, 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.* . . . Third, most of the activity that plays a role in establishing relations between bloggers, such as posting entries, making comments or joining communities, can be tracked. This is not the case for activity within Facebook or LinkedIn as activity there usually plays little role in establishing relations – as opposed to maintaining them – since those networks (for the most part) only formalize relations that were established outside their settings. Fourth, there is the option within blogging *not* to reciprocate readership by others, that is, there is a distinction between outlinks and inlinks. In comparison, social networking sites usually require reciprocation for a relation between users to be established, meaning that their network graph is undirected.

---

<sup>2</sup>“A list of other blogs that a blogger might recommend by providing links to them (usually in a sidebar list)” (see Wikipedia, 2011).

Given our choice, we now need to explain what is blogging, who blogs and why they do so. This leads us into an exposition of the of the rules that most bloggers observe when blogging and of the characteristics of blogger networks. This will be of use when discussing how to model their activity.

**What are blogs?** 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 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.

**Who are the bloggers?** A variety of surveys have been conducted to find out who the bloggers are. While many 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, a few recent ones rely on random selection and direct contact with the bloggers (Technorati, 2009; Lenhart and Fox, 2006). 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.

**What is in a blog?** 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.<sup>3</sup> A lot of attention has focused on thematic blogs, how

---

<sup>3</sup>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.

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 on-line diary genre however (Herring et al., 2005b; Technorati, 2009).

**Why do some people blog?** 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.<sup>4</sup> 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. Strategies that help one gain attention include posting interesting content, being the first to cover a topic, differentiating from others' coverage of a topic, or covering original topics (Shen, 2009). 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. Whether higher audience leads to more effort or *vice-versa* is not clear however, though, by using instrumental variables, 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. In this, blogging exhibits the same, more general mechanics that are at play in other settings where users contribute freely to media content on the Internet. Huberman, Romero, and Wu (2009), for example, identify the same effects as Hofstetter et al. (2009) among YouTube contributors. Zhang and Zhu (2011) show that Wikipedia contributors appear to respond to lower audiences by reducing their output. Contributors to user-generated content thus generally respond positively to attention, while those who make more effort are rewarded with more attention.

---

<sup>4</sup>A blogroll is a list of links to other blogs, usually shown in a sidebar but also on the blogger's profile page along with other information.

**What are the rules of blogging?** In the same way as any social behavior, 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.

**How do bloggers relate to each other? The role of reciprocation.** In terms of link formation, maintenance and dissolution, blogging rules have many similarities with the more general “rules of friendships” (Argyle and Henderson, 1984). Of particular interest to us are rules that relate to linking with (“making friend with”, “friending”) other bloggers. Publicly adding a blogger to one’s blog roll was at the root of the emergence of the blogging community (Ammann, 2010). Marlow (2006) reports that bloggers read more than 80% of the blogs in their blog roll in any given month, and over 60% in any given week . On LiveJournal, adding a blogger to one’s friend list not only means it will appear on one’s reading list but also gives that blogger access to “friends-only” entries (see LiveJournal, 2011a). The depth of feeling attached to the term “friend” on LiveJournal is reported in Marwick (2008), which also mentions that the vast majority of users of LiveJournal make at least occasional use of the “friends-only” privacy setting for their posts. This means that while signaling a relation between one’s blog and another is costless technically, it is by no mean an insignificant or neutral act.

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,<sup>5</sup> or of searching for bloggers with similar interests to one's own. Reactions to being referred to, commented upon or added vary, but generally involve 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.

Reciprocation processes may be strategically exploited by bloggers so as to maximize their audience. Some bloggers may be paying attention to others only 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. This is because reciprocation may indeed occur even if the blogger who initiated the reading relation is of little interest because of a *norm of reciprocity* which makes one feel obligated to reciprocate attention when paid attention to. This norm of reciprocity is "a universal structure of human morality" (Gintis et al., 2008; Henrich

---

<sup>5</sup>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).

et al., 2001), and was of particular relevance to the emergence of bloggers' social networks (Ammann, 2010). Indeed, 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 (Gouldner, 1960). 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.

**What is the structure of blogging networks?** Specific types of network structures come to be established as a result of the above mix of blogging motivations and norms. 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 that 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.<sup>6</sup> This can be seen as the result of a rich-get-richer dynamic as in Barabási and Albert (1999), or of a fitter-get-richer process as in Bianconi and Barabási (2001). However, this is only part of the story, and 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.<sup>7</sup> 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. Those are what we would consider as full participants in blogging networks, and those are the bloggers whose activity we are most interested in analyzing.

---

<sup>6</sup>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.

<sup>7</sup>See also Herring et al. (2005a) for similar results.

### 3. ONLINE SOCIAL NETWORKS AND SOCIAL CAPITAL

From the above, blogging is best seen 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). Bloggers are both producers and recipients of information, which means that blogging cannot be represented as a sender-receiver games where senders compete for the attention of a passive audience (Falkinger, 2007). 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, for example information (Galeotti and Goyal, 2010), and those that consider them as a tool for combining individual contributions to a productive effort (Jackson and Wolinsky, 1996). Galeotti and Goyal 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 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. Results thus 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. The network structure thus depends on the function the network plays, *i.e.* in which way it is used, and thus, on the objectives of the participants in the network. It is thus interesting to see what happens when the network can be used in different ways by different participants. For example, Harmsen-van Hout, Herings, and Dellaert (2010) combines social and informational value from link formation and shows this allows for a wider variety of equilibrium outcomes.

**Reciprocation in social network formation.** While the above models do provide interesting insights into blogging and go a long way towards explaining the variety in the structural patterns of blogging networks, only few, more recent papers attempt to look into the role of reciprocity in maintaining links between agents (Rivas, 2009; Jackson, Barraquer, and Tan, 2011), the tendency of individuals to link with agents that are similar to them (homophily) (Bramoullé and Rogers, 2010; Currarini, Jackson, and Pin, 2009), or the tendency to free-ride on the effort of others in the network (Bramoullé and Kranton, 2007). We find this strand of literature particularly interesting because there is a wealth of empirical evidence on the importance of such factors in online social networks.<sup>8</sup> 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.<sup>9</sup> Gu et al. (2009) investigate a peer-to-peer music sharing network and show that a pattern of indirect reciprocity holds, whereby free-riders, who only download music but do not offer music to download, are sanctioned with lower download speeds. Koenen and Reik (2010), looking at a similar service, find that users punish free-riding and respond positively to the provision of (useful!) effort by others. From other contexts, we also know that reciprocation is one of the main drivers in network formation along with popularity (making friends with those who have many friends) and triadic closure (making friends with friends of friends) (Schaefer et al., 2010). Reciprocated links appear very early in the formation of networks, while triadic closure plays a role only later (Doreian et al., 1996). Reciprocation is not only important in the emergence of relations

---

<sup>8</sup>There is also a small but growing literature presenting experimental evidence that underline the role of reciprocation (Conte, Di Cagno, and Sciubba, 2009) and of inequity aversion (Falk and Kosfeld, 2003) in the dynamics of social networks.

<sup>9</sup>For another paper dealing with the dynamics of user submission and upvoting on Digg, see Lussier, Raeder, and Chawla (2010).

but also makes them more stable than unilateral ones (see Hallinan, 1978; Runger and Wasserman, 1979, for data on primary school children).

This leads us to believe that while there will be a relation between activity at the level of a node and its ability to attract links, as per the first part of our literature review, this will be moderated by the need to maintain a balance in bloggers' relations. As seen above, this may be out of concern for maintaining reciprocal relations with other bloggers, because of a dislike for entertaining links with bloggers that have higher status, because bloggers will link with those they feel close to irrespective of their activity, *etc.* . . . Specifically, we focus on how the reciprocation process impacts how bloggers build readership. Two factors come into play: When a blogger adds another to his blogroll, then that other is likely to reciprocate (add back), while when a blogger is added by another, he is also likely to reciprocate. Readers may thus be gained from two sources, those bloggers that add another on their own, and those bloggers that merely reciprocate when another adds them. Others' actions influence one's actions, and vice versa, so our estimates of the role of different facets of the activity of bloggers must take account of this. We will use a model of social capital formation (Glaeser, Laibson, and Sacerdote, 2002) to take account of the tendency to reciprocate offers of friendship (or link, or readership, as befits the context) *vs.* the willingness of others to reciprocate one's offer of friendship. This model will allow us to answer some questions we think are still pending: do bloggers gain audience mainly by seeking out other bloggers, by freely reciprocating offers of friendship, or by providing content for others to consume? Is there a correlation between activity and reciprocation, such that those individuals that reciprocate less compensate by being more active so as to keep their audience? Does the tendency to reciprocate evolve over time, so that individuals become less ready to reciprocate as their social network grows? Finally, do individuals who reciprocate easily also obtain easier reciprocation of their own offers?

The social capital perspective is mainly popular in sociology and political science but exhibits very neat analogies with standard economic models of investment in physical and human capital (Glaeser, Laibson, and Sacerdote, 2002). In that perspective, social networks are an embodiment, a representation of social capital. Individual social capital is what is gained from maintaining relationships that give individuals access to social support and

to opportunities for a number of activities such as going out, doing sport, getting a job or making business (Lin, 2001). Actors that differ in terms of social capital will entertain relations that reflect this level of inequality. Relations that occur between unequal parties will be 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.

Individual social capital, social capital at the level of a node, may be measured in network theoretic terms: the number of degrees of that node, the density of its network, or how critical it is to the network as a whole (betweenness) (Borgatti, Jones, and Everett, 1998). Under this perspective, 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. We are of course not alone in studying the role of online social networks in accumulating, using and retaining social capital. Ellison, Steinfield, and Lampe (2007) show that Facebook works well in maintaining “weak ties”, which are sources of information rather than of social support (Granovetter, 1973). 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. Ahn and Watson (2010) study how the use of an online social network interact with the level of social capital accumulated there. Finally, Lento et al. (2006) show that social capital gained online can motivate continued activity in blogging.

**A model of investment in social capital.** Let us now present the model that we will be evaluating empirically. 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 Readers”, 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:

$$(3.1) \quad R_t = f(I_t) + (1 - \delta)R_{t-1}$$

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

$$(3.2) \quad r_t = f(I_t) - \delta R_{t-1}$$

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 readers gained in period  $t$ ,  $r_t = R_t - R_{t-1}$ , with  $R_t$  the number of readers at time  $t$ . This can be subdivided into  $r_t^a$ , the number of readers that added 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 reciprocated my readership. Thus,  $r_t = r_t^r + r_t^a$ .

In the same manner, denote the total number of "friends" (bloggers who are on a blogger's blog roll (reading list)) 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 added,  $f_t^a$ , and those "friends" who first sought me out (added me to their reading list) and whose readership I reciprocated,  $f_t^r$ . Thus,  $f_t = f_t^a + f_t^r$ .

Denote  $\theta$  my likelihood to reciprocate the readership of those bloggers who add me as "friends". Then  $f_t^r = \theta r_t^a$ . Similarly, denote  $\rho$  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:

$$(3.3) \quad f_t = f_t^a + \theta r_t^a$$

$$(3.4) \quad r_t = \rho f_t^a + r_t^a$$

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

$$(3.5) \quad r_t = \rho f_t + (1 - \rho\theta)r_t^a$$

$$(3.6) \quad f_t = \theta r_t + (1 - \rho\theta)f_t^a$$

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 (3.2) and comparing with (3.4), one obtains that

$$(3.7) \quad r_t^a = \beta A_t - \delta R_{t-1}$$

so (3.5) can be rewritten as follows:

$$(3.8) \quad r_t = \rho f_t + (1 - \rho\theta)\beta A_t - (1 - \rho\theta)\delta R_{t-1}$$

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 (3.8) requires indicators for  $A_t$ , which were spelled out on the facing page, estimation of (3.6) 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  $\lambda$  each period as the blogger drops less interesting friends. We thus rewrite  $f_t^a$  as  $\gamma B_t - \lambda F_{t-1}$ , so (3.6) can be expressed as follows:

$$(3.9) \quad f_t = \theta r_t + (1 - \rho\theta)\gamma B_t - (1 - \rho\theta)\lambda F_{t-1}$$

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  $\lambda$  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  $\delta$  to actually be negative, i.e. more readers beget more readers.

Estimation of both equations obtains estimates of  $\theta$  and  $\rho$ , 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.

#### 4. 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.<sup>10</sup> 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.<sup>11</sup>

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)

<sup>10</sup>Yandex Press release, September 26, 2006 ([http://company.yandex.com/press\\_center/press\\_releases/2006/2006-09-26.xml](http://company.yandex.com/press_center/press_releases/2006/2006-09-26.xml), accessed October 6, 2010).

<sup>11</sup>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).

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 Screenshoter (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 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:

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>i.e.</i> how many comments each entry receives on average.
Inactive	The number of weeks since the blog's last update, <i>i.e.</i> 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) reports descriptive statistics (average during the collection period):

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
$\Delta\log\_Readers$	.004	0.041	0	-1.43	1.783
$\Delta\log\_Friends$	.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.

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):

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
$\Delta\log\_Readers$	0.006	0.049	0	0.002	0.025	0
$\Delta\log\_Friends$	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

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 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:

TABLE 4. Blog age, size and nationality

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

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.

## 5. THE ECONOMETRIC MODEL

Estimating equation (3.8) and (3.9) 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. Since we cannot rely on any external instruments (e.g. there is no sources of exogenous variation in our sample of bloggers), we need to rely on “internal instruments” and blog-characteristics by applying system generalized method of moments (GMM) (Arellano and Bover, 1995; Blundell and Bond, 1998; Roodman, 2009a)). In addition, since for several bloggers, readers and friends present high persistence,<sup>12</sup> thus causing an weak-instruments concern, we rely on various specification to check the validity to our set of instruments (Bobba and Coviello, 2007) and to avoid instrument proliferation (Roodman, 2009b).

We consider the following dynamic specification to identify the reciprocity parameter  $\rho$  for readership, that is, the proportion of one’s new “friends” who reciprocate readership:

(5.1)

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

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 (3.8) (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),

<sup>12</sup>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).

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,  $\mu_t$  is a period-specific intercept,  $a_i$  is an unobserved time-invariant blog-specific effect, and  $\epsilon_{it}$  reflects serially uncorrelated errors.

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

(5.2)

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

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).

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.



TABLE 5. Friendship reciprocation. Dependent variable: Readers

VARIABLES	(1) Pooled OLS	(2) FE	(3) SYS GMM
L.Readers	1.004*** (0.000)	0.986*** (0.004)	0.999*** (0.001)
$\Delta \log\_Friends$	17.780** (7.720)	16.000** (7.067)	10.400*** (3.060)
Functionality	0.026 (0.028)	0.469*** (0.177)	9.784*** (1.104)
Age_blog	-0.003*** (0.000)	0.020*** (0.004)	0.010* (0.006)
Community_joined	0.238** (0.119)	0.207* (0.107)	0.680* (0.409)
Community_left	0.033 (0.039)	0.039 (0.038)	0.414*** (0.082)
Entries	0.009** (0.005)	0.010* (0.005)	-0.015 (0.014)
Posted	0.011*** (0.002)	0.022*** (0.003)	0.064*** (0.003)
Extroversion	-0.014*** (0.003)	-0.015*** (0.003)	-0.018 (0.018)
Engagement	0.011 (0.012)	0.012 (0.011)	0.030* (0.016)
Inactive	0.004 (0.004)	-0.018*** (0.004)	-0.388 (0.282)
Other	0.164 (0.176)		0.110 (1.866)
Russian	0.049 (0.134)		0.112 (1.362)
English	0.120 (0.091)		-3.511** (1.606)
Constant	-0.046 (0.186)	-3.038** (1.356)	-26.570*** (3.919)
Observations	75432	75432	75432
Number of user	1347	1347	1347
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*** ( 0.001)	0.932*** ( 0.015)	0.952*** (0.012)
$\Delta\log\_Readers$	54.740*** (12.800)	48.870*** (11.910)	15.520*** (4.975)
Functionality	0.063** ( 0.030)	0.756*** ( 0.260)	0.770** (0.386)
Age_blog	-0.000 ( 0.001)	0.028** ( 0.014)	0.011** (0.004)
Community_joined	0.595* ( 0.314)	0.469* ( 0.284)	0.352** (0.159)
Community_left	0.058 ( 0.283)	0.067 ( 0.282)	-0.074 (0.120)
Entries	0.005 ( 0.004)	0.005 ( 0.005)	-0.000 (0.006)
Posted	0.003 ( 0.002)	0.008** ( 0.004)	0.014*** (0.002)
Extroversion	-0.012* ( 0.006)	-0.013* ( 0.007)	-0.014*** (0.003)
Engagement	0.005 ( 0.006)	0.006 ( 0.006)	-0.003 (0.004)
Inactive	-0.005 ( 0.005)	-0.028*** ( 0.007)	-0.107*** (0.036)
Other	0.237 ( 0.253)		0.880 (1.160)
Russian	0.362 ( 0.225)		5.668*** (1.857)
English	-0.025 ( 0.105)		-1.293 (0.868)
Constant	-0.100 ( 0.159)	1.307 ( 3.527)	0.545 (1.111)
Observations	75432	75432	75432
Number of user	1347	1347	1347
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

We are interested in consistent estimation of the parameters  $(\theta, \rho)$ . 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  $\rho$  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  $\theta$  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  $\rho$  and  $\theta$  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  $\alpha_f$  and  $\alpha_r$ . The estimate for  $\rho$  (10.40) is closer to that of  $\theta$  (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 variables 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.<sup>13</sup>

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 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.

---

<sup>13</sup>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).

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).

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*** (0.000)	0.995*** (0.001)	0.990*** (0.013)	0.971*** ( 0.006)
L.Readers_square				4.310 · 10 <sup>-06</sup> *** ( 5.350 · 10 <sup>-07</sup> )
Δlog_Friends	9.792*** (3.150)	21.480*** (3.238)	42.040*** (7.085)	20.170*** ( 4.440)
Functionality	10.410*** (1.049)	17.810*** (1.802)	0.035 (0.037)	-4.724 ( 9.745)
Age_blog	0.035*** (0.008)	0.022 (0.015)	0.002 (0.002)	0.036 ( 0.069)
Community_joined	0.054 (0.033)	-0.342 (0.293)	-0.011 (0.026)	0.123* ( 0.065)
Community_left	0.003 (0.021)	0.752*** (0.089)	-0.001 (0.036)	-0.007 ( 0.018)
Entries	-0.005 (0.004)	0.479*** (0.055)	-0.003** (0.001)	0.017* ( 0.009)
Posted	0.008*** (0.002)	0.024*** (0.005)	0.004*** (0.001)	0.020*** ( 0.003)
Extroversion	0.018 (0.011)	0.078*** (0.017)	-0.001 (0.002)	-0.005 ( 0.006)
Engagement	-0.002 (0.007)	0.137*** (0.019)	0.001 (0.007)	0.008*** ( 0.003)
Inactive	-2.095*** (0.315)	-0.673 (0.453)	-0.013 (0.013)	-0.262 ( 0.369)
Other		15.460*** (4.986)	0.167 (0.123)	41.020 (136.200)
Russian		8.596** (3.490)	0.277 (0.326)	65.920 ( 49.710)
English		-15.660*** (5.516)	-0.018 (0.032)	17.960 ( 61.830)
Constant	-25.470*** (3.930)	-53.270*** (7.913)	0.096 (0.284)	-22.740 ( 33.390)
Observations	24304	20384	55048	75432
Number of user	434	364	983	1347
Hansen test of overid.	0.392	0.135	0.317	0.446
AR(1) in first differences	0.019	0.002	2.020 · 10 <sup>-05</sup>	0.006
AR(2) in first differences	1.660 · 10 <sup>-05</sup>	0.323	0.682	0.248
diff Hansen level	0.633	0.294	0.558	0.971
diff Hansen 2	0.570	0.001	0.014	0.110
diff Hansen 3	0.210	0.094	0.565	0.074

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

VARIABLES	(1)	(2)	(3)	(4)
	SYS GMM Russian	SYS GMM Big Blog	SYS GMM Small-Med Blog	SYS GMM Network Effect
L.Friends	1.000*** ( 0.005)	0.997*** ( 0.023)	0.993*** (0.005)	1.022*** ( 0.046)
L.Friends_square				$-1.170 \cdot 10^{-05}$ ( $2.690 \cdot 10^{-05}$ )
$\Delta \log\_Readers$	28.450 (22.280)	394.300*** (78.640)	6.627*** (2.166)	12.160* ( 7.188)
Functionality	-0.102 ( 0.264)	1.127 ( 4.347)	0.019 (0.030)	-0.338 ( 1.666)
Age_blog	-0.002 ( 0.004)	0.077 ( 0.144)	0.000 (0.001)	-0.058 ( 0.132)
Community_joined	0.097 ( 0.299)	0.119 ( 0.300)	0.136** (0.055)	0.254 ( 0.216)
Community_left	-0.117 ( 0.285)	-0.166 ( 0.239)	0.161 (0.125)	0.343 ( 0.245)
Entries	-0.002 ( 0.015)	-0.082*** ( 0.030)	0.002 (0.003)	0.002 ( 0.005)
Posted	0.013*** ( 0.004)	-0.001 ( 0.005)	0.005*** (0.001)	0.010*** ( 0.003)
Extroversion	-0.026*** ( 0.007)	-0.006 ( 0.008)	-0.002 (0.002)	-0.012* ( 0.007)
Engagement	-0.009 ( 0.011)	-0.004** ( 0.001)	0.011 (0.009)	0.001 ( 0.005)
Inactive	0.001 ( 0.024)	-0.257 ( 0.454)	-0.008* (0.004)	-0.069 ( 0.198)
Other		-42.490 (60.720)	0.062 (0.103)	-0.098 ( 7.919)
Russian		-43.570 (40.180)	0.281 (0.181)	-1.343 ( 6.325)
English		-48.660 (38.310)	-0.018 (0.069)	3.749 (11.220)
Constant	0.534 ( 0.562)	16.030 (53.200)	0.321* (0.190)	13.230 (36.610)
Observations	24304	20384	55048	75432
Number of user	434	364	983	1347
Hansen test of overid.	0.377	0.782	0.304	0.612
AR(1) in first differences	0.063	0.013	0.001	0.011
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

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.

**The social multipliers and the reflection effects.** The presence of positive spillovers or strategic complementarities between individuals and their peers creates a “social multiplier”, that is the individual’s results will be affected by social interactions with individuals in the “same-group”. Econometrically, this implies a simultaneity problem (Manski, 1993). It is possible to deal with this problem and to identify exogenous effects whenever the social network has a rich (non-linear) structure or there was a random assignment to peers’s group (Bramoullé and Fortin, 2009).

In our setting, we do not have information about “groups”. Since there is no limits in the number of friends that could be added to one’s list, our main assumption is that every blogger in LJ is a potential *peer* for each other blogger. That is, we consider each blogger’s peers’ group endogenous. However, to check the robustness of our results, we split our sample



according to the language of the blog and obtained similar results (see again Tables (7) and (8)). Moreover, in contrast with many other studies, we have access to a set of variables that can proxy a blogger's tendency to join different groups (*i.e.* *Community*) and a blogger's behaviour with respect to his peers (*i.e.* *Engagement* and *Extroversion*). We can thus capture the different ways in which bloggers interact with each other without having to introduce arbitrary constraints on the size of the group.

As in Glaeser, Laibson, and Sacerdote (2002), another advantage of our analysis is that, by focusing on individuals, we can study how they make endogenous decisions about social capital accumulation instead of having to rely on aggregate group-outcomes. Moreover, we also account for group interactions by considering the role of reciprocation and thus taking into account how others' actions (in this case, reading one's blog) influence one's actions.

Therefore, in order to fully assess the impacts of a blogger's activity on their readership while taking into account the effects of friends' reciprocation, we compute in this section the dynamic multipliers based on the reduced-form parameters derived from the specification in Column (3) of Table (5) and (6).

In particular, we compute two types of dynamic multipliers: *interim multipliers* and *long-run multipliers*. The interim multiplier gives the effects of a unit increase in an exogenous variable on an endogenous variable when this effect is sustained for a specific amount of time. A long-run multiplier gives the effect of a unit increase in an exogenous variable on an endogenous variable when sustained in the indefinite future.<sup>14</sup>

The results are reported in Table (9) for a selection of variables. Specifically, the table reports the effects of: a unit increase of the blog functionality (e.g. from an Early account to a Sponsored account), an increase of ten in

<sup>14</sup>Starting from the representation of the multivariate dynamic regression model,  $Y_t = BY_{t-1} + \Gamma X_t + V_t$ , and solving by iteration one obtains  $Y_t = B^t Y_0 + A_t \Gamma X_t + A_t V_t$ , where  $A_t = \sum_{k=0}^{t-1} B^k = (I - B^t)(I - B)^{-1}$ . If  $B^t \rightarrow 0$  as  $t \rightarrow \infty$ , we obtain the *long-run reduced form*  $Y_t = \Pi X_t + \Delta_t$ , where  $\Pi = (I - B)^{-1} \Gamma$  and  $\Delta_t = (I - B)^{-1} V_t$  are the reduced form coefficients and disturbance terms respectively. The matrix of s-period (*interim*) multipliers can be expressed as  $\frac{\partial Y_t}{\partial X_{t-s}} = A_t \Gamma$ , whereas the matrix of *long-run* multipliers (with  $t \rightarrow \infty$ ) can be expressed as  $\frac{\partial Y}{\partial X} = \Pi$ . See Helmut (2005) for more details. In our case, after algebraic manipulation,  $B = (1 - \theta\rho)^{-1} \begin{bmatrix} \alpha_r - \theta\rho & \rho(\alpha_f - 1) \\ \theta(\alpha_r - 1) & \alpha_f - \theta\rho \end{bmatrix}$ .

the number of posts entered by the blogger per week, and a unit (in our case, a week) increase in the length of inactivity of the blogger.

TABLE 9. Dynamic Multiplier: Results

TIME	Functionality		Posted		Inactive	
	Readers	Friends	Readers	Friends	Readers	Friends
0	0.0019	0.0027	0.0001	0.0004	-0.0001	-0.0003
1	0.0026	0.0054	0.0002	0.0008	-0.0001	-0.0006
2	0.0033	0.0078	0.0003	0.0011	-0.0002	-0.0008
3	0.0051	0.0102	0.0004	0.0015	-0.0003	-0.0011
4	0.0063	0.0124	0.0005	0.0018	-0.0004	-0.0013
5	0.0075	0.0146	0.0006	0.0021	-0.0004	-0.0015
⋮						
10	0.0132	0.0237	0.0011	0.0034	-0.0007	-0.0025
⋮						
Long-run	0.1034	0.0530	0.0072	0.0078	-0.0045	-0.0058

We observe that the long-run effect of functionality on the number of readers, holding constant prior number of friends, is about 75 times larger than the multiplier at time 0. Similarly, the long-run run effect on readers of increasing by ten the number of posts is about 72 times larger than the multiplier at time 0, whereas this values is about 45 times larger for an increase of a week of inactivity of the blogger. The long-run effects on friends are much smaller, and about 19 times larger then the multiplier at time 0, for all these three variables. Overall, the effects of social multipliers are slow in coming and appear only if sustained over long periods of time.

More precisely, the effects of varying the blog characteristics and the blogger's activity appear quite early when it comes to accumulating friends (blogs one reads), whereas the effect on one's number of readers appear only later on. In particular, increasing the functionalities of one's blog, for example by paying for it, will initially have a stronger effect on the number of blogs one reads ("friends"), but will result in a stronger effect on readership in the long term. Similarly, making more comments will initially be associated

with reading more blogs, but the effect on one's audience will equalize in the long-term. Finally, being inactive for longer periods will initially mean one will add fewer people to one's reading list, but will in the end also result in a decrease in one's readership.

Our analysis therefore confirms that blogging is best thought of as a long term investment requiring sustained effort over long periods of time, whereby the flow of investment and benefits will not be evenly matched in time. As in many other human endeavours, patience and persistence are of the essence.

## 6. 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. Our model implies that 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.

We applied that insight to our data and found that indeed, adding bloggers to one's reading list translated in a significant increase in one's audience, and conversely that being added by others was associated with an increase in the number of bloggers on one's reading list. We also found that a blogger's social capital (here, number of readers) was affected by how much attention they devoted to other bloggers through comments posted on their blogs, a finding that is all the more striking as writing entries, which is often seen as the main activity of bloggers, did not seem to impact readership to the same extent. Finally, we showed that while increases in blogging activity had little impact on readership in the short-term, and were in fact initially associated mainly with adding new blogs to one's reading list, they translated in at least commensurate increases in terms of readership over the long-term thanks to the combined effect of increased content production and reciprocation by others when paid attention to.

Being the first to develop an analysis of reciprocation as a factor in the growth of social network, we do not yet have reference points for our reciprocation parameters  $\rho$  and  $\theta$ . 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 in the reciprocation parameters while Facebook users would presumably display high levels of reciprocation. Further research might require access to company data however since LiveJournal is quite unique among social networks in the amount of public information it provides about the activity of its users.

## REFERENCES

- Ahn, Dae-Yong and Randal Watson. 2010. "A dynamic model of user behavior in a social network site." University of Technology, Sydney, and University of Texas at Austin. (ref. p. 13).
- Ammann, Rudolf. 2010. "Reciprocity, social curation and the emergence of blogging: A study in community formation." In *7th Conference on Applications of Social Network Analysis*. (ref. p. 7 and 9).
- Anderson, Chris. 2004. "The long tail." *Wired* 12 (10). (ref. p. 9).
- 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. 22).
- Argyle, Michael and Monika Henderson. 1984. "The Rules of Friendship." *Journal of Social and Personal Relationships* 1 (2):211–237. (ref. p. 7).
- Aristotle. ca. 350 B.C.E. *The Nicomachean Ethics*. Oxford University Press. (1908 translation by W. D. Ross). (ref. p. 2).
- Bar-Ilan, Judit. 2005. "Information hub blogs." *Journal of Information Science* 31 (4):297–307. (ref. p. 6).
- Barabási, Albert-László and Réka Albert. 1999. "Emergence of scaling in random networks." *Science* 286:509–512. (ref. p. 9).
- Bianconi, Ginestra and Albert-László Barabási. 2001. "Competition and multiscaling in evolving networks." *Europhysics Letters* 54 (4):436–442. (ref. p. 9).
- 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. 22 and 26).
- 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. 22).

- Borgatti, Stephen P, Candace Jones, and Martin G. Everett. 1998. "Network measures of social capital." *Connections* 21 (2):online. (ref. p. 13).
- Bramoullé, Yann and Bernard Fortin. 2009. "The Econometrics of Social Networks." *Cahiers de recherche 0913, CIRPEE* . (ref. p. 31).
- Bramoullé, Yann and Rachel Kranton. 2007. "Public goods in networks." *Journal of Economic Theory* 135 (1):478–494. (ref. p. 11).
- Bramoullé, Yann and Brian W. Rogers. 2010. "Diversity and popularity in social networks." Department of Economics, CIRPÉE and GREEN, Université Laval and MEDS, Kellogg School of Management, Northwestern University. (ref. p. 11).
- Chan, Kimmy Wa and Stella Yiyang 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. 11).
- Choi, In. 2001. "Unit root tests for panel data." *Journal of International Money and Finance* 20 (2):249–272. (ref. p. 22).
- 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. 11).
- Conte, Anna, Daniela Di Cagno, and Emanuela Sciubba. 2009. "Strategies in social network formation." *Jena Economic Research Papers, 2009-095* . (ref. p. 11).
- Currarini, Sergio, Matthew O. Jackson, and Paolo Pin. 2009. "An Economic Model of Friendship: Homophily, Minorities and Segregation." *Econometrica* 77 (4):1003–1045. (ref. p. 11).
- 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. 11).
- ekiwi, LLC. 2011. *Screen-scraper, version 5.0, enterprise edition*. Provo, UT, USA. URL <http://www.screen-scraper.com>. (ref. p. 1 and 17).
- 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. 13).

- Falk, Armin and Michael Kosfeld. 2003. "It's All About Connections: Evidence on Network Formation." *IZA Discussion Paper No. 777* . (ref. p. 11).
- Falkinger, Josef. 2007. "Attention economies." *Journal of Economic Theory* 133 (1):266–294. (ref. p. 10).
- Galeotti, Andrea and Sanjeev Goyal. 2010. "The law of the few." *American Economic Review* 100 (4):1468–1492. (ref. p. 10).
- 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. 8).
- Glaeser, Edward L., David Laibson, and Bruce Sacerdote. 2002. "An economic approach to social capital." *The Economic Journal* 112 (483):437–458. (ref. p. 2, 12, 13, and 32).
- 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. 5).
- Gouldner, Alvin W. 1960. "The norm of reciprocity: A preliminary statement." *American Sociological Review* 25 (2):161–178. (ref. p. 9).
- Granovetter, M. 1973. "The Strength of Weak Ties." *American Journal of Sociology* 78 (6):1360–1380. (ref. p. 13).
- 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. 11).
- Hallinan, Maureen T. 1978. "The process of friendship formation." *Social Networks* 1 (2):193–210. (ref. p. 12).
- 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. 10).
- 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. 22).
- Helmut, Lütkepohl. 2005. *New Introduction to Multiple Time Series Analysis*. Springer-Verlag. (ref. p. 32).

- 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. 8).
- 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. 9).
- 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. 6).
- 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. 6).
- Huberman, Bernardo A., Daniel M. Romero, and Fang Wu. 2009. "Crowdsourcing, attention and productivity." *Journal of Information Science* 35 (6):758–765. (ref. p. 6).
- 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. 22).
- Jackson, Matthew O., Tomás Rodríguez Barraquer, and Xu Tan. 2011. "Social capital and social quilts: Network patterns of favor exchange." *American Economic Review, forthcoming*. (ref. p. 11).
- 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. 10).
- Koenen, Johannes and Steffen Reik. 2010. "Individual (ir)rationality? An empirical analysis of behavior in an emerging social online-network." CDSE Discussion Paper No. 97, University of Mannheim. (ref. p. 11).
- 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. 9).

- Lemann, Nicholas. 2006. "Journalism without journalists." *The New Yorker* URL [http://www.newyorker.com/archive/2006/08/07/060807fa\\_fact1](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. 5).
- 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. 13).
- 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. 13).
- 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. 7).
- . 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. 8).
- 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).
- Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies* 60 (3):531–42. (ref. p. 31).
- 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. 6, 7, and 13).
- Marwick, Alice. 2008. "LiveJournal users: Passionate, prolific and private." URL [http://www.livejournalinc.com/LJ\\_Research\\_Report.pdf](http://www.livejournalinc.com/LJ_Research_Report.pdf). Accessed January 18, 2010. (ref. p. 7).



- 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).
- 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. 7).
- Rivas, Javier. 2009. "Friendship selection." *International Journal of Game Theory* 38 (4):521–538. (ref. p. 11).
- Roodman, David. 2009a. "How to do xtabond2: An introduction to difference and system GMM in Stata." *Stata Journal* 9 (1):86–136. (ref. p. 22).
- . 2009b. "A note on the theme of too many instruments." *Oxford Bulletin of Economics and Statistics* 71 (1):135–158. (ref. p. 22 and 27).
- Rui, Huaxia and Andrew Whinston. 2010. "Social media as an innovation - the case of Twitter." University of Texas at Austin. (ref. p. 9).
- Runger, George and Stanley Wasserman. 1979. "Longitudinal analysis of friendship networks." *Social Networks* 2 (2):143–154. (ref. p. 12).
- 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. 11).
- 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. 11).
- Schmidt, Jan. 2007. "Blogging practices: An analytical framework." *Journal of Computer-Mediated Communication* 12 (4):1409–1427. (ref. p. 7).
- Shen, Wenqi. 2009. "Competing for attention in online reviews." *AMCIS 2009 Doctoral Consortium. Paper 2* URL [http://aisel.aisnet.org/amcis2009\\_dc/2](http://aisel.aisnet.org/amcis2009_dc/2). Accessed January 18, 2010. (ref. p. 6).
- Shirky, Clay. 2003. "Power laws, weblogs, and inequality." URL [http://shirky.com/writings/powerlaw\\_weblog.html](http://shirky.com/writings/powerlaw_weblog.html). Accessed January 18, 2010. (ref. p. 9).
- 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. 5 and 6).
- 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. 5).
- Wikipedia. 2011. "Glossary of blogging." URL [http://en.wikipedia.org/w/index.php?title=Glossary\\_of\\_blogging&oldid=407505604](http://en.wikipedia.org/w/index.php?title=Glossary_of_blogging&oldid=407505604). Accessed January 17th, 2011. (ref. p. 4 and 8).
- 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. 5).
- Zhang, Xiaoquan (Michael) and Feng Zhu. 2011. "Group size and incentives to contribute: A natural experiment at Chinese Wikipedia." *The American Economic Review* 101 (4):1601–1615. (ref. p. 6).