

Gaudeul, Alexia; Giannetti, Caterina

Working Paper

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

Jena Economic Research Papers, No. 2012,031

Provided in Cooperation with:

Max Planck Institute of Economics

Suggested Citation: Gaudeul, Alexia; Giannetti, Caterina (2012) : The role of reciprocation in social network formation, with an application to blogging, Jena Economic Research Papers, No. 2012,031, Friedrich Schiller University Jena and Max Planck Institute of Economics, Jena

This Version is available at:

<https://hdl.handle.net/10419/70157>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



JENA ECONOMIC RESEARCH PAPERS



2012 – 031

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

Revised version, June 2012

by

**Alexia Gaudeul
Caterina Giannetti**

www.jenecon.de

ISSN 1864-7057

The JENA ECONOMIC RESEARCH PAPERS is a joint publication of the Friedrich Schiller University and the Max Planck Institute of Economics, Jena, Germany. For editorial correspondence please contact markus.pasche@uni-jena.de.

Impressum:

Friedrich Schiller University Jena
Carl-Zeiss-Str. 3
D-07743 Jena
www.uni-jena.de

Max Planck Institute of Economics
Kahlaische Str. 10
D-07745 Jena
www.econ.mpg.de

© by the author.

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

Revised version, June 2012

Alexia Gaudeul^a, Caterina Giannetti^b

^aMax Planck Institute of Economics, Jena

^bRoyal Holloway, University of London

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

[☆]Research was funded by the Federal Program "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 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, at the 38th Annual Conference of the European Association for Research in Industrial Economics (EARIE) in Stockholm in September 2011, at the Info-Metrics Institute workshop on the Information and Econometrics of Networks in Washington in March 2012, and at the 13th annual conference of the Association for Public Economic Theory (APET) in Taipei in June 2012. We benefited from discussions with Yann Bramoullé, Eric Darmon, François Gardes, Avi Goldfarb, Adelina Gschwandtner, Nicolas Jullien, Arianna Martinelli, Kevin Mellet, Raffaele Miniaci, 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.

Email addresses: alexia.gaudeul@econ.mpg.de (Alexia Gaudeul), caterina.giannetti@rhul.ac.uk (Caterina Giannetti)

June 22, 2012

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

Revised version

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. We therefore investigate how far one's efforts and willingness to initiate relations with others can help grow one's social network along with how attractive one is as a partner.

For that purpose, we chose to study bloggers' choices of which other blogs to subscribe to. Bloggers are emerging as significant players in the media market. They disseminate content from mainstream media and also serve as references for many newspapers and television stations. Blogging also has a social networking aspect, similar to that of Facebook and Google+. Blogging is therefore best seen as a special form of social networking where actors share knowledge along with forming and sustaining social links with others (Lussier et al., 2010).

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.* how many are gained through other activities, such as posting blog entries, making comments, joining communities, *etc.* We estimate an original economic model adapted from Glaeser et al. (2002) to include the mechanisms of reciprocation into the dynamics of the building of social capital over time. In this context, a blogger's social capital is measured by how many people subscribe his feed.¹ We take account of how bloggers build their social networks not only by providing attractive content which other bloggers will want to keep on reading over time, but also by interacting with other bloggers. Bloggers seek out other bloggers by adding them to their friends list, or "blogroll"², and they also reciprocate the attention of those who added them to their own blogroll. Sought out bloggers reciprocate reading, and those whose readership is reciprocated keep on reading. Reciprocation thus contributes to the accumulation of an audience over time.

¹"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, accessed 20th of January 2011.

²"A list of other blogs that a blogger might recommend by providing links to them (usually in a sidebar list)" (see Wikipedia, 2011).

Our paper is one of the first to follow the activity of bloggers over time along with their audience and the first to take account of the activity consisting in seeking out relations with other bloggers. Our research contributes to a better understanding of social network formation by exploiting fine-grained data collected online. Our work provides and tests an original conceptual and analytical tool that encompasses the variety of uses of social media. Indeed, our model allows us to locate the different manifestations of social media along the range between social networking, which is affinity-based and where reciprocation is important, and news reporting, which is focused on collecting, producing and disseminating information and where reciprocation plays less of a role. If bloggers' networks are primarily based on affinity – people read those they feel close to – then reciprocation will be particularly important in maintaining one's network while effort exerted in blogging will have little influence on one's number of readers. Alternatively, if bloggers' networks are primarily based on less personal factors such as how interesting, informative or influential a blogger is, then readers will not require reciprocation while better and more frequent posting will translate into higher audiences. Whether reciprocation plays a role in the process of bloggers' network formation therefore depends on the type of activity that is being led there.

Our work also refines the 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 model for the analysis of individual social capital. Those dynamic multipliers allow us to measure the long term impact of changes in a blogger's activity and in a blog characteristics (our exogenous variables) on that blogger's social capital (*i.e.* "readership", our endogenous variables).

Outline:

A first section explains why we chose blogging networks as our source of data. A second section presents and motivates a model of investment in social capital that takes account of the rules of reciprocation within social networks. A third section tests how well that model describes the relation between activity and network size for a panel of bloggers on LiveJournal whom we followed week by week over more than a year.

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 (ekiwi LLC, 2011) to extract information from websites.

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 (by definition) so that (almost) all relations between blogs, whether through their blogroll, comments by the bloggers or links to entries on other blogs, will leave a trace online.³ 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, and to the small subset of their interactions that emerge online. Second, blogs and their interconnections form a relatively self-contained world with a fairly clear unity of purpose. Generalist social networking tools such as Facebook are more difficult to make sense of because they 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 those networks mainly serve to formalize relations that were established outside their settings so that activity within those networks usually plays little role in establishing relations between their users. 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, thus making them more difficult to use when analyzing reciprocation processes.⁴

Given our choice to focus on blogging networks, we explain briefly what are blogs, who blogs and why they do so before going into the rules that most bloggers observe when blogging. This will be of use when motivating our modeling of their activity.

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 or music, and frequently includes links to and commentary 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.

³One still misses the interactions between bloggers that occur by mail or during occasional meetups though.

⁴The researcher would need access to non-public data on who made the first move to add someone else to their list, including cases where the move was not reciprocated and the relationship thus not established.

Of the variety of surveys that have been conducted to find out who the bloggers are, a few more reliable and recent surveys rely on random selection and direct contact with the bloggers (Technorati, 2009; Lenhart and Fox, 2006) and 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 host 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.

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.⁵ The majority of blogs belong to the online diary genre (Herring et al., 2005b; Technorati, 2009) but the public focus has been mainly on thematic blogs and their influence on journalism and politics (Bar-Ilan, 2005; Lemann, 2006; McKenna and Pole, 2008).

Another way to categorizes blogs, along Shirky (2003), is between *A-list blogs* whose authors attract so much attention that they cannot possibly reciprocate all of it, thus ending up as part of the mainstream media, and *conversational blogs* whose authors spend time cultivating their links with a few others.⁶⁷ Kumar et al. (2010) provide more detailed insights into the variety of blogs by looking at their network structure. 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.⁸ 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. The later are those that we would consider as full participants in blogging networks, and they are

⁵The 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.

⁶Rui 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.

⁷The emergence of A-list blogs 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). In that view, conversational blogs would then be said to be part of a “long tail” of less successful blogs (Anderson, 2004). We reject that view however since the goals of those who maintain conversational blogs may not be the same as of those who maintain A-list blogs.

⁸See also Herring et al. (2005a) for similar results.

the bloggers whose activity we mean to model.

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 mean to "speak their mind" and "share their expertise and experience with other people", but also mean 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.⁹ Getting attention for their opinions and expertise, along with building relations with others, notably to share experiences and obtain social support, is therefore an important motivation for bloggers. They will therefore strive to post interesting content, be the first to cover a topic, differentiate from others' coverage of a topic, or cover original topics (Shen, 2009). Empirical work confirms that getting attention motivates content production and that time spent maintaining a blog pays off in terms of audience size and feedback (Marlow, 2006). 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 both directions of causality hold. Beyond blogging, the same general mechanics are at play whenever one can contribute freely to media content on the Internet. Huberman et al. (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 and those who make more effort are rewarded with more attention.

3. Integrating reciprocity into a model of social capital formation

The next two sub-sections introduce two building blocks for the model that will be exposed in section 3.3. The first section underlines the strength of the rule of reciprocity within blogging networks. The second section argues that blogging networks can be considered at least in part as a repository of social capital. This motivates the third section where we integrate reciprocity processes into Glaeser et al. (2002)'s model of investment in social capital.

3.1. *The rules of reciprocity within blogging networks*

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,

⁹A 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.

providing references to their sources, and relating to others (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.

We focus in this paper on the specific rules that guide *reciprocation* within blogging networks. 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 other drivers such as triadic closure play a role only later (Doreian et al., 1996). Not only is reciprocation very important in the emergence of relations but it also makes them more stable than unilateral ones (see Hallinan, 1978; Runger and Wasserman, 1979, for data on primary school children). Experimental evidence also underline the role of reciprocation (Conte et al., 2009), and relatedly of inequity aversion (Falk and Kosfeld, 2003), in the dynamics of social network formation. Reciprocation processes draw their strength from the *norm of reciprocity*, which is “a universal structure of human morality” (Gintis et al., 2008; Henrich et al., 2001) and 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). This norm of reciprocity translates very well online and has indeed been shown to be a crucial component in the maintenance of online social systems. (Chan and Li, 2009; Chun et al., 2008; Gu et al., 2009; Koenen and Reik, 2010; Lussier et al., 2010; Sadlon et al., 2008).

Linking with other bloggers (alternatively, “blogrolling” or “friending”) is a decision that bloggers take very seriously. One should not be fooled into thinking that signaling a relation between one’s blog and another is insignificant because it is technically costless. Indeed, the blogging community can be rightly said to have emerged only once bloggers publicly added other bloggers to their blog roll (Ammann, 2011). Linking is not only a matter of signaling a relation between two blogs however. 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. Furthermore, there is a depth of feeling attached to “friending” on LiveJournal (Marwick, 2008), where 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). “Friending” rules on LiveJournal have therefore many similarities with the more general “rules of friendships” (Argyle and Henderson, 1984).

Among different reasons guiding the choice of whom to “friend”, 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 (Raynes-Goldie and Fono, 2006). 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,¹⁰ 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 also depend 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 privacy. 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 and it is often considered good manner to give some explanation for ceasing to follow another blogger.

3.2. *Blogging networks as social capital*

Models in the economic literature on *social* networks – networks where nodes choose with whom to form and maintain links – focus on discerning the equilibrium properties of networks under a variety of assumptions regarding the use that is made of them or the tendencies that are at play in the behavior of their members. For example, Galeotti and Goyal (2010) focuses on social networks as a tool for gathering and transmitting value, e.g information, among its members, while Jackson and Wolinsky (1996) consider the role of social networks as a tool for combining individual contributions to a productive effort. Other papers look into variations of those models where reciprocity plays a role in maintaining links between agents (Rivas, 2009; Jackson et al., 2011), where individuals

¹⁰The concept of “community” is specific to LiveJournal, the blog host 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).

tend to link with agents that are similar to them (homophily) (Bramoullé and Rogers, 2010; Currarini et al., 2009), or where agents tend to free-ride on the effort of others in the network (Bramoullé and Kranton, 2007). Those models would be well-adapted for explaining the static properties of bloggers' social graphs, but are not fit for our purpose as our main interest is in identifying properties of the dynamics of social network formation. To understand the difference between a static and a dynamic analysis in our setting, consider a blogger who would add 100 bloggers to his reading list every period, see 10% reciprocate, and drop all those who did not reciprocate. This blogger would accumulate a balanced reading list over time and reach an equilibrium whereby all those he reads also read him. A static analysis would thus give very different results than a dynamic analysis like ours. To give an example, *static* analysis of social graphs on Twitter shows that relations are often not reciprocated there (Kwak et al., 2010, section 3.3). Facebook relations on the other hand can appear only if they are reciprocal. Whether this translates when doing a dynamic analysis is not guaranteed since there might be many friending requests on Facebook that are rejected but are not captured when doing a static analysis – so the reciprocation ratio might actually be low there – and conversely, the low number of reciprocated relations on Twitter might be due to a high rate of depreciation in relationships – that is, it may be that Twitterers reciprocate all incoming requests but quickly drop those that provide lower quality content.

The social capital perspective, which is mainly popular in sociology and political science, proves to be more amenable to our purpose. That perspective exhibits very neat analogies with standard economic models of investment in physical and human capital (Glaeser et al., 2002). In that perspective, social networks are an embodiment, a representation of social capital. Individual social capital is the product of maintaining relationships with others, which gives 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). Individual social capital, social capital at the level of a node, can 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 et al., 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. A number of papers, foremost among them Ellison et al. (2007), establish the role of online social networks as tools to accumulate, make use of and retain social capital. Marlow (2006) is one of the few papers in that literature which focuses

specifically on blogging network. It finds that it is mainly “social” bloggers (diarists) that gain social capital through their online activity by maintaining personal contacts with their readers, while “professional” (thematic) bloggers, who entertain larger audiences, do not gain much in terms of social capital even though they invest more time in their blogs. The interaction between online social capital and activity is explored in Lento et al. (2006) and Ahn (2009), which both take account of the feedback effect whereby activity in social networks builds up social capital online which itself in turn motivates further activity. However, those papers do not take account as we do of those activities that consist in seeking blogs to read and reciprocating the readership of others. As we will argue, taking account of those activities is important if one is to obtain correct estimates of the effect of different types of activities on the rate of social capital accumulation in blogging networks.

3.3. Integrating reciprocation into a model of investment in social capital

Having established the strength of the rule of reciprocation within blogging networks, as well as the role of blogging networks as a repository of social capital, we now adapt Glaeser et al. (2002)’s model of social capital formation to take account of the tendency to reciprocate offers of friendship (or link, or readership, as befits the context) and of the willingness of others to reciprocate one’s offer of friendship. The model will allow us to take account of 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 will thus be gained from two sources, those bloggers that add me on their own and those bloggers who reciprocate because I added them. Estimates of the role of different facets of the activity of bloggers must take account of this later source of readership.

A model of investment in social capital

Glaeser et al. (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 δ each period so the following equation obtains:

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

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

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

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. 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. 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. Note however that 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. We will attempt to correct for such potential network effects in some specifications of our econometric model.

Integrating reciprocation 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 θ my likelihood to reciprocate the readership of those bloggers who add me as "friends". Then $f_t^r = \theta r_t^a$. The parameter θ is a function not only of the quality of the content provided by those bloggers, but also of my policy with respect to reciprocation. For example, I may add back anyone who adds me (in which case $\theta = 1$), or I may be much more careful (see section 3.1). We will investigate whether this propensity to reciprocate depends on the blogging culture in which I evolve (Russian vs. English-speaking blogs) or on the stage in the blog's life.

Similarly, denote ρ the likelihood with which bloggers I add to my reading list reciprocate and add me back. Then $r_t^r = \rho f_t^a$. This propensity may again be a function of the

blogging culture in which I evolve, or of the stage in the development and/or popularity of my blog.

The following system of two equations results from the above:

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

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

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

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

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

As evoked above, investment I_t can be divided into A_t (publishing content in one's blog, making comments on others' blogs, joining communities and so on) and f_t^a (the activity of seeking out and adding "friends"). Let us assume that $f(\cdot)$ is a linear function of A_t and f_t^a such that $f(I_t) = \beta A_t + \rho f_t^a$.¹¹ Replacing $f(I_t)$ with this expression in (3.2) and comparing with (3.4), one obtains that

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

so (3.5) can be rewritten as follows:

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

This can be further rewritten as

$$R_t = \rho f_t + (1 - \rho\theta)\beta A_t + (1 - (1 - \rho\theta)\delta)R_{t-1} \quad (3.9)$$

which is the specification that we will estimate econometrically. 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 we wrote that $r_t^a = \beta A_t - \delta R_{t-1}$, one can write $f_t^a = \gamma B_t - \lambda F_{t-1}$. B_t includes indicators for how active the blogger is in his search for new "friends", such as the number of communities joined per sampling period or the number of communities

¹¹More general specifications allowing for interactions between various aspects of activity were also tested for.

one participates in, as those put one in contact with more people. λ is the factor by which our stock variable, the number of existing “friends” F_{t-1} , depreciates each period as the blogger drops less interesting friends (for example). Note however that while friendships “depreciate” over time, friends put one in contact with their friends, thus potentially contributing to the increase in one’s number of friends. One may thus find λ to actually be negative, i.e. more friends beget more friends. This being said, we can now rewrite (3.6) as follows:

$$f_t = \theta r_t + (1 - \rho\theta)\gamma B_t - (1 - \rho\theta)\lambda F_{t-1} \quad (3.10)$$

This can be further rewritten as

$$F_t = \theta r_t + (1 - \rho\theta)\gamma B_t + (1 - (1 - \rho\theta)\lambda)F_{t-1} \quad (3.11)$$

which is the specification that we will estimate econometrically.

Estimation of equations 3.9 and of 3.11 obtains estimates of θ and ρ , the reciprocation ratios. Knowing θ and ρ allows one to have correct estimates of β (resp. γ), which measure the number of readers (resp. friends) that are gained through activity other than simply adding “friends” (resp. readers). Similarly, one also obtains correct estimates of the depreciation rates in the number of friends and readers over time.

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.¹² 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,

¹²Yandex Press release, September 26, 2006 (http://company.yandex.com/press_center/press_releases/2006/2006-09-26.xml, accessed October 6, 2010).

some have an account on LJ mainly to be able to join and contribute to communities and discuss the content posted there.¹³

Our sample consists of a cross-section of more than two thousand 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 who 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 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 blogger’s readers on his entries (*Received*), the net number of communities joined or left by the user (if positive, this is our variable *Communities_joined*, if negative, this is our variable *Communities_left*), 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:

¹³Communities that are particularly popular include ohnotheydidnt (celebrity gossip), customers.suck (rant community), adayinmylife (picture diaries, most posts visible to members only), saucydwelling or abandoned-places (pictures) or bakebakebake (cooking).

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 who 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 equations (3.9) and (3.11) 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 the system generalized method of moments (GMM) procedure (Arellano and Bover, 1995; Blundell and Bond, 1998; Roodman, 2009a). We acknowledge that there could be several issues

when using this technique, some of which have yet to be determined when the length of the panel (T) is large (Han and Phillips, 2010, p. 120). However, in the following analysis, we try to take care of its main weaknesses as identified in the literature for the type of data at hand. First, we perform various tests in order to check the presence of unit roots in the data. These indeed highlight that for several bloggers, readers and friends present high persistence.¹⁴ Second, since high-persistence causes a weak-instruments concern, we rely on various specifications of system-GMM (e.g variables in logs) and check the validity (with the Hansen tests) of our set of instruments (Bobba and Coviello, 2007). Since some of these checks (i.e. Hansen test) may be vitiated by instrument proliferation and be inconsistent against some local alternatives (Newey, 1985), we checked the validity of different subset of instruments (e.g. differences in the Hansen test). We also made sure to avoid instrument proliferation by limiting (see further below) the lag length of our instruments (Roodman, 2009b). Finally, we check the robustness of our results across different subsamples.

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:

$$Readers_{it} = \alpha_r Readers_{i(t-1)} + \rho f_{it} + \beta_{r1} Activity_{it} + \beta_{r2} Blog_Characteristics_{it} + a_i + \mu_t + \epsilon_{it} \quad (5.1)$$

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 (3.9) (in logs). We checked alternative specifications (available upon request) in which we consider Readers and Friends either in logs or in levels and results are largely consistent. We decided to present the results of this specification as it allows us to smooth (through log) the "Friends" data series and to include (ratio) indicators of activities. 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

¹⁴Specifically, we perform the following tests in order to check the presence of unit roots in the data: Im et al. (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).

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 that are either fixed or slow-changing. Specifically this group includes a categorical variable related to the type of account (i.e. $Functionality_{it}$), dummies for the language of the blog ($Russian$, $English$, $Other$ or $Unknown$) and a variable measuring 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:

$$Friends_{it} = \alpha_f Friends_{i(t-1)} + \theta r_{it} + \beta_{f1} Activity_{it} + \beta_{f2} Blog_Characteristics_{it} + a_i + \mu_t + \eta_{it} \quad (5.2)$$

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 as defined in equation (3.11) (in logs, under the same justification as above).

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)	0.985*** (0.007)
Δlog(Friends)	17.783** (7.720)	16.001** (7.067)	25.786*** (9.585)
Functionality	0.026 (0.028)	0.469*** (0.177)	0.776 (1.463)
Age blog	-0.003*** (0.000)	0.020*** (0.004)	0.237* (0.129)
Community joined	0.238** (0.119)	0.207* (0.107)	0.567 (0.928)
Community left	0.033 (0.039)	0.039 (0.038)	0.497** (0.226)
Entries	0.009** (0.005)	0.010* (0.005)	-0.080 (0.092)
Posted	0.011*** (0.002)	0.022*** (0.003)	0.148*** (0.008)
Extroversion	-0.014*** (0.003)	-0.015*** (0.003)	-0.136*** (0.044)
Engagement	0.011 (0.012)	0.012 (0.011)	0.071** (0.031)
Inactive	0.004 (0.004)	-0.018*** (0.004)	-0.016 (0.597)
Other	0.164 (0.176)		-123.660* (67.280)
Russian	0.049 (0.134)		-33.676* (19.120)
English	0.120 (0.091)		-51.766** (22.318)
Constant	-0.046 (0.186)	-3.038** (1.356)	-22.337 (39.733)
Observations	75 432.000	75 432.000	75 432.000
Number of user	1 347.000	1 347.000	1 347.000
Hansen test of overid.			0.575
AR(1) in first differences			0.762
AR(2) in first differences			0.222
diff Hansen level			0.696
diff Hansen 2			0.297
diff Hansen 3			0.248
diff Hansen 4			0.561
diff Hansen 5			0.892

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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)	0.959*** (0.009)
Δlog(Readers)	54.736*** (12.801)	48.869*** (11.907)	43.031* (22.805)
Functionality	0.063** (0.030)	0.756*** (0.260)	0.285 (1.146)
Age blog	-0.000 (0.001)	0.028** (0.014)	0.106 (0.108)
Community joined	0.595* (0.314)	0.469* (0.284)	4.261*** (0.197)
Community left	0.058 (0.283)	0.067 (0.282)	1.097*** (0.365)
Entries	0.005 (0.004)	0.005 (0.005)	0.035 (0.047)
Posted	0.003 (0.002)	0.008** (0.004)	0.050*** (0.007)
Extroversion	-0.011* (0.006)	-0.013* (0.007)	0.003 (0.072)
Engagement	0.005 (0.006)	0.006 (0.006)	-0.035 (0.061)
Inactive	-0.005 (0.005)	-0.028*** (0.007)	-0.963** (0.464)
Other	0.237 (0.253)		-0.022 (4.349)
Russian	0.362 (0.225)		3.273 (3.040)
English	-0.025 (0.105)		-7.911 (8.676)
Constant	-0.100 (0.159)	1.307 (3.527)	1.104 (27.133)
Observations	75 432.000	75 432.000	75 432.000
Number of user	1 347.000	1 347.000	1 347.000
Hansen of overid.			0.139
AR(1) in first differences			0.006
AR(2) in first differences			0.300
diff level			0.399
diff Hansen 2			0.519
diff Hansen 3			0.003
diff Hansen 4			0.555
diff Hansen 5			0.974

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

We are interested in consistent estimation of the parameters (θ, ρ) . In particular, in the presence of individual-specific effects, OLS in 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 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. Interesting to note is how low our reciprocation parameter appears to be compared to static analysis which indicated that 87% of relations were reciprocated on LiveJournal (Gaudeul and Peroni, 2010) (see page 9 for why static and dynamic analysis might obtain different results).

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. Their approach, which combines moment conditions based on the level equations together with the usual Arellano and Bover (1995) type conditions for the difference equations, yields consistent estimates even with persistent series. However, there are still open issues in regard to the limit distribution when T is large and the series contain unit roots (Han and Phillips, 2010). 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. In particular, we use as set of instruments the lagged levels and differences dated $t-2$ up to $t-4$ of *Readers* and *Friends*, and dated $t-1$ up to $t-3$ of *Activity*. We additionally include in our set of instruments the characteristics of the blog such as the language and the type of functionality, and we further distinguish the instruments between the equations expressed in differences and in levels.¹⁵

Relying on these estimators we find reasonable parameter estimates. The estimated coefficient on the lagged dependent variable is not below the fixed-effect estimator, and

¹⁵More specifically, we exclude the second lag for the equations in differences. In a previous version of the paper, we did not distinguish the instruments for the equation in levels and differences. Although results are qualitatively consistent, estimations improve considerably in comparison with the OLS and fixed-effects estimations when we distinguish between the two types of instruments.

above the OLS levels estimate for both α_f and α_r . The estimate for ρ (25.79) is closer to that of θ (43.03) 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 $t-2$ up to $t-4$ 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 (sometimes as high as implausible p -values of 1.000). We therefore carefully check the value of the statistics across different specifications of the model. In our preferred specification, the test of common restrictions (see “Hans test of overid.” line in tables) 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 and to deal with the loss of power of the Hansen’s statistics against local alternatives (Newey, 1985), 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 subsets of instruments (diff level, diff Hansen 2, diff Hansen 3, diff Hansen 4 and diff Hansen 5). These tests suggest the validity of our specification at the 10% level.¹⁶

The parameters relating to the activity of the bloggers are also in line with expectations from the model, that is, activity contributes 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, suggesting that those bloggers who are too active compared with their readers in posting comments will not gain readers. Bloggers who simply make comments and do not receive many will thus not gain as many readers as those who both make and receive comments. The coefficient on the number of comments made by the readers *per* entries made by the user (*Engagement*) is positive, and can be interpreted as meaning that the more interesting or provocative the blogger’s entries, the higher will be his number of readers. On the other hand, simply writing more entries will not gain readers. The number of comments posted has a similar impact in the “*Friends*”

¹⁶The 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).

equation: The higher the number of comments posted, the higher the number of friends, possibly because bloggers who pay more attention to their friends also add more bloggers to their reading list. The degree of a blogger's extroversion turned out not be single significant in this equation.

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 also some evidence that a blog's level of functionalities has an effect (although not significant in this specific GMM specification): the higher the level of functionality in the blog (which may also mean 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 who 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)	(2)	(3)	(4)
	SYS GMM Russian	SYS GMM Big Blog	SYS GMM Small-Med Blog	SYS GMM Network Effect
L.Readers	0.988*** (0.003)	0.995*** (0.001)	0.984*** (0.014)	0.979*** (0.006)
L.Readers_square				0.008*** (0.000)
Δlog_friends	12.543** (4.897)	21.069*** (3.034)	16.952*** (3.624)	23.609*** (4.658)
Functionality	0.739 (0.549)	17.727*** (1.737)	0.045 (0.295)	-4.705 (8.342)
Age blog	0.015* (0.009)	0.018 (0.014)	0.028 (0.023)	-0.129 (0.164)
Community joined	6.089*** (0.734)	-0.302 (0.292)	-0.010 (0.420)	0.131** (0.060)
Community left	0.986*** (0.115)	0.723*** (0.084)	0.118 (0.138)	-0.000 (0.019)
Entries	-0.080* (0.044)	0.480*** (0.055)	0.005 (0.014)	0.015* (0.008)
Posted	0.068*** (0.011)	0.024*** (0.004)	0.027*** (0.010)	0.019*** (0.003)
Extroversion	0.044 (0.033)	0.076*** (0.017)	-0.048 (0.042)	-0.005 (0.005)
Engagement	0.353*** (0.094)	0.131*** (0.018)	-0.055 (0.048)	0.008*** (0.003)
Inactive	-0.864*** (0.286)	-0.615 (0.437)	-0.060 (0.063)	-0.058 (0.175)
Other		16.175*** (4.937)	-22.558** (10.635)	-4.158 (99.063)
Russian		9.414*** (3.423)	-4.878* (2.849)	52.121 (42.977)
English		-13.274** (5.230)	-7.575** (3.517)	27.668 (54.907)
Constant	15.416*** (4.913)	-53.122*** (7.476)	0.990 (5.172)	22.915 (47.832)
Observations	24 304.000	20 384.000	55 048.000	75 432.000
Number of user	434.000	364.000	983.000	1 347.000
AR(1) in first differences	0.025	0.002	0.001	0.006
AR(2) in first differences	0.334	0.349	0.115	0.343
Hansen test of overid.	0.761	0.142	0.706	0.531
diff Hansen lev	0.546	0.311	0.989	0.992
diff Hansen 2	0.289	0.390	0.631	0.189
diff Hansen 3	0.877	0.935	0.736	0.706
diff Hansen 4	0.318	0.076	0.505	0.320
diff Hansen 5	0.535	0.457	0.944	0.320

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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	0.964*** (0.016)	0.970*** (0.023)	0.989*** (0.003)	0.997*** (0.018)
L.Friends square				0.000 (0.000)
$\Delta \log$ readers	37.913*** (10.476)	192.211*** (22.112)	6.386*** (2.076)	9.312 (5.666)
Functionality	14.565** (7.415)	3.448 (4.135)	0.037 (0.024)	-0.528 (0.484)
Age blog	-0.062 (0.166)	-0.092 (0.160)	0.001 (0.000)	-0.068 (0.055)
Community joined	6.317*** (1.599)	0.128 (0.273)	0.141** (0.055)	0.205 (0.181)
Community left	10.203*** (0.458)	-0.081 (0.176)	0.157 (0.124)	0.296 (0.224)
Entries	0.434*** (0.104)	-0.017 (0.024)	0.003 (0.003)	0.002 (0.005)
Posted	0.025 (0.025)	0.006** (0.003)	0.005*** (0.001)	0.010*** (0.002)
Extroversion	0.067 (0.183)	-0.006 (0.006)	-0.002 (0.002)	-0.011** (0.005)
Engagement	0.659*** (0.219)	-0.001 (0.001)	0.016** (0.008)	0.001 (0.004)
Inactive	0.530 (0.381)	-1.195 (0.796)	-0.011*** (0.003)	-0.065 (0.082)
Other		-122.902 (90.922)	0.069 (0.100)	0.272 (1.914)
Russian		-136.138** (61.110)	0.426*** (0.126)	-0.288 (1.795)
English		-153.054** (75.744)	-0.012 (0.067)	4.892 (4.342)
Constant	-38.993 (57.686)	152.306* (88.176)	0.448*** (0.158)	17.550 (15.293)
Observations	24 304.000	20 384.000	55 048.000	75 432.000
Number of user	434.000	364.000	983.000	1 347.000
Hansen of overid.	0.761	0.693	0.293	0.716
AR(1) in first differences	0.002	0.015	0.001	0.011
AR(2) in first differences	0.920	0.175	0.298	0.197
diff Hansen lev	0.551	0.880	0.744	0.744
diff Hansen 2	0.494	0.094	0.814	0.814
diff Hansen 3	0.328	0.218	0.203	0.203
diff Hansen 4	0.639	0.830	0.460	0.460
diff Hansen 5	0.833	0.796	0.475	0.475

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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 (at least on LiveJournal), 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 as 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 derive some of their readership not so much by being more successful when adding friends 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 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 amplified through social interactions with individuals in the “same-group” (Glaeser et al., 2003). As in Glaeser et al. (2002), an 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. However, one still has to deal econometrically with the simultaneity problem, whereby an individual’s actions are affected by behavior in his reference group (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’ groups (Bramoullé and Fortin, 2010).

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 peer group as being endogenous. However, to check the robustness of our results, we split our sample according to the language of the blog and obtained results that were consistent with aggregated 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 mode of and extent of interactions with his peers, namely his joining of different groups (*i.e. Community*) and the level of interaction with 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 his peer group. For example, we control how far a blogger's comments are determined by how many comments he receives, to what extent readers respond to his content, and how many new relations are determined by joining new communities of joint interest. Finally, our model is designed to take account of a major way in which we think others' actions affect mine, namely how far others adding me to their blog roll determines my adding them to mine.

Analysis of the long-term impact of changes in activity

Our approach allows us to fully assess the impacts of changes in a blogger's activity on their readership by taking into account the effect of friends' reciprocation. We compute in this section the dynamic multipliers based on the reduced-form parameters derived from the system-GMM 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.¹⁷

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 a default, "Plus" account to a Paid account), an increase of one in the number of comments posted by the blogger per week, and a unit (in our case, a week) increase in how long the blogger last wrote an entry on his blog.

¹⁷Starting 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 $\frac{\partial Y_t}{\partial X_{t-s}} = A_s \Gamma$, whereas the matrix of *long-run* multipliers (with $t \rightarrow \infty$) can be expressed as $\frac{\partial Y}{\partial X} = \Pi$. See Lütkepohl (2005) for more details. In our case, after algebraic manipulation, $B = (1 - \theta\rho)^{-1} \begin{bmatrix} \alpha_r - \theta\rho & \rho(\alpha_f - 1) \\ \theta(\alpha_r - 1) & \alpha_f - \theta\rho \end{bmatrix}$ while $\Gamma = (1 - \theta\rho)^{-1} \begin{bmatrix} \beta_r + \rho\beta_f \\ \beta_f + \theta\beta_r \end{bmatrix}$. One then obtains long run effect $\Pi = \begin{bmatrix} \beta_r(1 - \alpha_r)^{-1} \\ \beta_f(1 - \alpha_f)^{-1} \end{bmatrix}$.

Table 9: Dynamic Multiplier: Results

TIME	Functionality		Posted		Inactive	
	Readers	Friends	Readers	Friends	Readers	Friends
0	0.956	0.696	0.181	0.128	-0.297	-1.091
1	1.887	1.353	0.357	0.248	-0.577	-2.129
2	2.794	1.972	0.529	0.362	-0.839	-3.118
3	3.679	2.557	0.697	0.469	-1.086	-4.059
4	4.542	3.109	0.861	0.571	-1.316	-4.955
5	5.384	3.628	1.020	0.666	-1.533	-5.808
⋮						
10	9.302	5.805	1.764	1.064	-2.421	-9.497
⋮						
Long-run	51.733	6.951	9.867	1.220	-1.067	-23.488

The short-term effect (time 0) of increasing one's level of activity or functionality on one's number of readers is a composite of a direct effect, how many readers one attracts with this change, and of an indirect effect, which is due to reciprocation by the additional friends resulting from this change in activity. The long-run effect, on the other hand, does not depend on the level of reciprocation, that is, the indirect effect becomes less and less important over time compared to the main effect (see note 17).

The long run effect of functionality on the number of readers, holding constant the prior number of friends, is about 54 times larger than the multiplier at time 0. Going from the default type of account (Plus) to a Paid account, for example, increases one's number of reader by 52 in the long-term. Similarly, the long-run effect of increasing by one the number of comments on others' blogs on one's number of readers is about 54 times larger than the multiplier at time 0, whereas this values is only about 3 times larger if a bloggers increases by a week the interval between blog posts. The long-run effects on friends are much smaller, for the blog functionality and the number of comments, whereas they are much higher for the inactivity of the blogger. In this latter case, the long run effect of an additional week of inactivity is about 21 times larger than the multiplier at time zero.

Overall, the effects of social multipliers are slow in coming, that is, effects appear only if changes are sustained over long periods of time. The effects of varying the blog characteristics and the blogger's activity appear quite early when it comes to accumu-

lating friends (blogs one reads), whereas the disproportionate effect on one's number of readers is apparent only later on. In particular, increasing the functionalities of one's blog, for example by paying for it, is initially associated with parallel increases in the number of readers and friends, but results in a stronger effect on readership in the long term. Similarly, making more comments will initially be associated with reading more blogs, but the main effect will be on one's audience in the long-term. Overall, those findings fit well with the observation that older blogs tend to have blog rolls that are much more unbalanced than younger ones, that is, they typically have many more readers than friends as can be seen in table 3 and elaborated on in Gaudeul and Peroni (2010). Interesting is the effect of inactivity, which results mainly in not accumulating friends at the same rate as before while the effect on one's readership will be very limited in the long term. This underlines the high degree of inertia among readers, that is, one is not likely to lose subscribers merely because one does not update one's blog – indeed, there is no cost for readers in keeping a blogger on their reading list.

Our analysis 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 endeavors, 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 the process of social network formation. Our model implies that both activity and reciprocation 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 mainly affected by how much attention they devoted to other bloggers through comments posted on their blogs. Writing entries, which is often seen as the main activity of bloggers, impacted readership only to the extent that those entries attracted comments, while merely increasing the rate at which a blog was updated did not result in higher audiences. Finally, we showed that while increases in blogging activity were initially associated with commensurate increases in one's number of readers and the number of

blogs on one's reading list, they translated in large and disproportionate increases in terms of readership over the long-term.

Our analysis of reciprocation as a factor in the growth of social network provides a framework for a general understanding of social media and online social networking and could thus be extended from our analysis of blogging on LiveJournal to other social networks. We think that LiveJournal is merely a point on a continuum that spans from Twitter to Facebook in terms of how important content production (the "media" aspect) is compared to reciprocation (the "social networking" aspect). Only further empirical research using data from those networks could however confirm the value of our approach in determining how much is "social" and how much is "media" in social media.

References

- Ahn, D.-Y., 2009. A dynamic model of usage behavior and network effects in social network sites. Ph.D. thesis, The University of Texas at Austin. (ref. p. 10).
- Ammann, R., 2011. Reciprocity, social curation and the emergence of blogging: A study in community formation. *Procedia - Social and Behavioral Sciences* 22 (0), 26 – 36. (ref. p. 7).
- Anderson, C., 2004. The long tail. *Wired* 12 (10). (ref. p. 5).
- Arellano, M., Bover, O., 1995. Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics* 68 (1), 29–51. (ref. p. 18 and 23).
- Argyle, M., Henderson, M., 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, J., 2005. Information hub blogs. *Journal of Information Science* 31 (4), 297–307. (ref. p. 5).
- Barabási, A.-L., Albert, R., 1999. Emergence of scaling in random networks. *Science* 286, 509–512. (ref. p. 5).
- Bianconi, G., Barabási, A.-L., 2001. Competition and multiscaling in evolving networks. *Europhysics Letters* 54 (4), 436–442. (ref. p. 5).
- Blundell, R., Bond, S., 1998. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87 (1), 115–143. (ref. p. 18 and 23).
- Bobba, M., Coviello, D., 2007. Weak instruments and weak identification in estimating the effects of education on democracy. *Economics Letters* 96 (3), 301–306. (ref. p. 19).
- Borgatti, S. P., Jones, C., Everett, M. G., 1998. Network measures of social capital. *Connections* 21 (2), online. (ref. p. 9).
- Bramoullé, Y., Fortin, B., 2010. The econometrics of social networks. In: Durlauf, S. N., Blume, L. E. (Eds.), *The New Palgrave Dictionary of Economics*. Palgrave Macmillan, Basingstoke. (ref. p. 28).
- Bramoullé, Y., Kranton, R., 2007. Public goods in networks. *Journal of Economic Theory* 135 (1), 478–494. (ref. p. 9).
- Bramoullé, Y., Rogers, B. W., 2010. Diversity and popularity in social networks, Université Laval and Northwestern University. (ref. p. 9).
- Chan, K. W., Li, S. Y., 2009. Understanding consumer-to-consumer interactions in virtual communities: The salience of reciprocity. *Journal of Business Research* 63 (9-10), 1033–1040. (ref. p. 7).
- Choi, I., 2001. Unit root tests for panel data. *Journal of International Money and Finance* 20 (2), 249–272. (ref. p. 19).

- Chun, H., Kwak, H., Eom, Y.-H., Ahn, Y.-Y., Moon, S., Jeong, H., 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. pp. 57–70. (ref. p. 7).
- Conte, A., Di Cagno, D., Sciubba, E., 2009. Strategies in social network formation. Jena Economic Research Papers, 2009-095. (ref. p. 7).
- Currarini, S., Jackson, M. O., Pin, P., 2009. An economic model of friendship: Homophily, minorities and segregation. *Econometrica* 77 (4), 1003–1045. (ref. p. 9).
- Doreian, P., Kapuscinski, R., Krackhardt, D., Szczypula, J., 1996. A brief history of balance through time. *The Journal of Mathematical Sociology* 21 (1-2), 113–131. (ref. p. 7).
- ekiwi LLC, 2011. Screen-scraper, version 5.0, enterprise edition. Provo, UT, USA.
URL <http://www.screen-scraper.com> (ref. p. 1, 4, and 14).
- Ellison, N. B., Steinfield, C., Lampe, C., 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. 9).
- Falk, A., Kosfeld, M., 2003. It's all about connections: Evidence on network formation. IZA Discussion Paper No. 777. (ref. p. 7).
- Galeotti, A., Goyal, S., 2010. The law of the few. *American Economic Review* 100 (4), 1468–1492. (ref. p. 8).
- Gaudeul, A., Peroni, C., 2010. Reciprocal attention and norm of reciprocity in blogging networks. *Economics Bulletin* 30 (3), 2230–2248. (ref. p. 23 and 31).
- Gintis, H., Henrich, J., Bowles, S., Boyd, R., Fehr, E., 2008. Strong reciprocity and the roots of human morality. *Social Justice Research* 21 (2), 241–253. (ref. p. 7).
- Glaeser, E. L., Laibson, D., Sacerdote, B., 2002. An economic approach to social capital. *The Economic Journal* 112 (483), 437–458. (ref. p. 2, 6, 9, 10, and 28).
- Glaeser, E. L., Scheinkman, J. A., Sacerdote, B. I., 2003. The social multiplier. *Journal of the European Economic Association* 1 (2/3), 345–353. (ref. p. 28).
- Gorny, E., 2006. Russian LiveJournal: The impact of identity on the development of a virtual community. In: Schmidt, H., Teubener, K., Konradova, N. (Eds.), *Control + Shift. Public and Private Usages of the Russian Internet*. Norderstedt: Books on Demand. (ref. p. 5).
- Gouldner, A. W., 1960. The norm of reciprocity: A preliminary statement. *American Sociological Review* 25 (2), 161–178. (ref. p. 7).
- Gu, B., Huang, Y., Duan, W., Whinston, A. B., 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. 7).
- Hallinan, M. T., 1978. The process of friendship formation. *Social Networks* 1 (2), 193–210. (ref. p. 7).
- Han, C., Phillips, P. C. B., 2010. Gmm estimation for dynamic panels with fixed effects and strong instruments at unity. *Econometric Theory* 26 (01), 119–151. (ref. p. 19 and 23).
- Harris, R. D. F., Tzavalis, E., 1999. Inference for unit roots in dynamic panels where the time dimension is fixed. *Journal of Econometrics* 91 (2), 201–226. (ref. p. 19).
- Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., McElreath, R., 2001. In search of Homo Economicus: Behavioral experiments in 15 small-scale societies. *The American Economic Review* 91 (2), 73–78. (ref. p. 7).
- Herring, S. C., Kouper, I., Paolillo, J. C., Scheidt, L. A., Tyworth, M., Welsch, P., Wright, E., Yu, N., 2005a. Conversations in the blogosphere: An analysis "From the bottom up". In: Proceedings of the 38th Hawaii International Conference on System Science. (ref. p. 5).
- Herring, S. C., Scheidt, L. A., Wright, E., Bonus, S., 2005b. Weblogs as a bridging genre. *Information Technology & People* 18 (2), 142–171. (ref. p. 5).
- Hofstetter, R., Shriver, S. K., Nair, H. S., Miller, K., 2009. Social ties and user generated content: Evidence from

- an online social network. NET Institute, Working Paper 09-28. (ref. p. 6).
- Huang, C.-Y., Shen, Y.-Z., Lin, H.-X., Chang, S.-S., 2007. Bloggers' motivations and behaviors: A model. *Journal of Advertising Research* 47 (4), 472–484. (ref. p. 6).
- Huberman, B. A., Romero, D. M., Wu, F., 2009. Crowdsourcing, attention and productivity. *Journal of Information Science* 35 (6), 758–765. (ref. p. 6).
- Im, K. S., Pesaran, M. H., Shin, Y., 2003. Testing for unit roots in heterogeneous panels. *Journal of Econometrics* 115 (1), 53–74. (ref. p. 19).
- Jackson, M. O., Barraquer, T. R., Tan, X., 2011. Social capital and social quilts: Network patterns of favor exchange. *American Economic Review*, forthcoming. (ref. p. 8).
- Jackson, M. O., Wolinsky, A., 1996. A strategic model of social and economic networks. *Journal of Economic Theory* 71 (1), 44–74. (ref. p. 8).
- Koenen, J., Reik, S., 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. 7).
- Kumar, R., Novak, J., Tomkins, A., 2010. Structure and evolution of online social networks. In: Yu, P. S. S., Han, J., Faloutsos, C. (Eds.), *Link Mining: Models, Algorithms, and Applications*. Springer: New York, pp. 337–357. (ref. p. 5).
- Kwak, H., Lee, C., Park, H., Moon, S., 2010. What is Twitter, a social network or a news media? In: *Proceedings of the 19th international conference on the World Wide Web*. ACM, pp. 591–600. (ref. p. 9).
- Lemann, N., 2006. Journalism without journalists. Accessed January 18, 2010.
URL http://www.newyorker.com/archive/2006/08/07/060807fa_fact1 (ref. p. 5).
- Lenhart, A., Fox, S., 2006. Bloggers: A portrait of the Internet's new storytellers. Accessed October 18, 2010.
URL <http://www.pewinternet.org/Reports/2006/Bloggers.aspx> (ref. p. 5).
- Lento, T., Welsch, H. T., Gu, L., Smith, M., 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. 10).
- Lin, N., 2001. Building a network theory of social capital. In: Lin, N., Cook, K. S., Burt, R. S. (Eds.), *Social Capital: Theory and Research*. Aldine de Gruyter, Ch. 1, pp. 6–7. (ref. p. 9).
- LiveJournal, 2011a. What are Friends? How do I manage my Friends list? Frequently Asked Question #61. Accessed January 17, 2011.
URL <http://www.livejournal.com/support/faqbrowse.bml?faqid=61> (ref. p. 7).
- LiveJournal, 2011b. What is a community? Frequently Asked Question #77. Accessed January 17, 2011.
URL <http://www.livejournal.com/support/faqbrowse.bml?faqid=77&view=full> (ref. p. 8).
- Lussier, J., Raeder, T., Chawla, N., 2010. User generated content consumption and social networking in knowledge-sharing OSNs. In: Chai, S.-K., Salerno, J., Mabry, P. (Eds.), *Advances in Social Computing*. Vol. 6007 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg, pp. 228–237. (ref. p. 2 and 7).
- Lütkepohl, H., 2005. *New Introduction to Multiple Time Series Analysis*. Springer-Verlag. (ref. p. 29).
- Manski, C. F., July 1993. Identification of endogenous social effects: The reflection problem. *Review of Economic Studies* 60 (3), 531–42. (ref. p. 28).
- Marlow, C. 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 9).
- Marwick, A., 2008. LiveJournal users: Passionate, prolific and private. Accessed January 18, 2010.
URL http://www.livejournalinc.com/LJ_Research_Report.pdf (ref. p. 7).
- McKenna, L., Pole, A., 2008. What do bloggers do: an average day on an average political blog. *Public Choice* 134 (1), 97–108. (ref. p. 5).
- Newey, W., September 1985. Generalized method of moments specification testing. *Journal of Econometrics* 29 (3), 229–256. (ref. p. 19 and 24).

- Raynes-Goldie, K., Fono, D., 2006. Hyperfriends and beyond: Friendship and social norms on LiveJournal. In: Consalvo, M., Haythornthwaite, C. (Eds.), *Internet Research Annual Volume 4: Selected Papers from the Association of Internet Researchers Conference*. Peter Lang: New York, USA. (ref. p. 8).
- Rivas, J., 2009. Friendship selection. *International Journal of Game Theory* 38 (4), 521–538. (ref. p. 8).
- Roodman, D., 2009a. How to do xtabond2: An introduction to difference and system GMM in Stata. *Stata Journal* 9 (1), 86–136. (ref. p. 18).
- Roodman, D., 2009b. A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics* 71 (1), 135–158. (ref. p. 19 and 24).
- Rui, H., Whinston, A., 2010. Social media as an innovation - the case of Twitter, University of Texas at Austin. (ref. p. 5).
- Runger, G., Wasserman, S., 1979. Longitudinal analysis of friendship networks. *Social Networks* 2 (2), 143–154. (ref. p. 7).
- Sadlon, E., Sakamoto, Y., Dever, H. J., Nickerson, J. V., 2008. The karma of Digg: Reciprocity in online social networks. In: *Proceedings of the 18th Annual Workshop on Information Technologies and Systems*. (ref. p. 7).
- Schaefer, D. R., Light, J. M., Fabes, R. A., Hanish, L. D., Martin, C. L., 2010. Fundamental principles of network formation among preschool children. *Social Networks* 32 (1), 61 – 71. (ref. p. 7).
- Schmidt, J., 2007. Blogging practices: An analytical framework. *Journal of Computer-Mediated Communication* 12 (4), 1409–1427. (ref. p. 7).
- Shen, W., 2009. Competing for attention in online reviews. AMCIS 2009 Doctoral Consortium, accessed January 18, 2010.
URL http://aisel.aisnet.org/amcis2009_dc/2 (ref. p. 6).
- Shirky, C., 2003. Power laws, weblogs, and inequality. Accessed January 18, 2010.
URL http://shirky.com/writings/powerlaw_weblog.html (ref. p. 5).
- 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. Accessed October 18, 2010.
URL <http://technorati.com/blogging/feature/state-of-the-blogosphere-2009> (ref. p. 5 and 6).
- Wei, L., 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. Accessed January 17th, 2011.
URL http://en.wikipedia.org/w/index.php?title=Glossary_of_blogging&oldid=407505604 (ref. p. 2 and 8).
- Yu, H., 2007. Exploring the Chinese blogosphere: The motivations of blog authors and readers. Master's thesis, National University of Singapore, accessed January 27, 2011.
URL <https://scholarbank.nus.edu.sg/handle/10635/13159> (ref. p. 5).
- Zhang, X. M., Zhu, F., 2011. Group size and incentives to contribute: A natural experiment at Chinese Wikipedia. *The American Economic Review* 101 (4), 1601–1615. (ref. p. 6).