Blogs and the Economics of Reciprocal Attention

Gaudeul, Alexia and Peroni, Chiara and Mathieu, Laurence

GSBC, Friedrich Schiller University, Jena, School of Economics, University of East Anglia

21 October 2009

Blogs and the Economics of Reciprocal Attention[1]

Alexia Gaudeul[2], Chiara Peroni[3] and Laurence Mathieu[4]

October 21, 2009

[1]This paper was presented at the Second FLOSS Workshop in Rennes in June 2008, at the Fifth bi-annual Conference on The Economics of the Software and Internet Industries in Toulouse in January 2009 and at the Annual Conference of the Royal Economic Society in Guilford in April 2009. We are grateful to Adelina Gschwandtner, Peter Moffatt and Paul Seabright for useful discussion. The support of the ESRC is gratefully acknowledged.

[2]Graduate School "Human Behaviour in Social and Economic Change" (GSBC), Friedrich Schiller University, Jena, Germany, email: a.gaudeul@gmail.com, website: http://agaudeul.free.fr.

[3]School of Economics, University of East Anglia, Norwich, UK and Rimini Center for Economic Analysis (RCEA), Rimini, Italy, email: c.peroni@uea.ac.uk.

[4]email: laurencefmathieu@yahoo.co.uk.

Abstract

Properties of blogging networks are derived from a model where bloggers devote attention to others, produce content for others, and exchange attention with content within their network of relations. The predictions from the model are tested with a novel dataset from LiveJournal, a major blogging community. The activity of a blogger is found to be related to the size of, and level of aggregate reciprocation within that blogger's relational network. Bloggers who do not adhere to reciprocity norms are found to have less readers than their activity might otherwise have predicted.

JEL Classifications: D63, D85, H41, L17, L82, L86, Z13.

Keywords: Blog; Community; Interaction; Internet; LiveJournal; Media; Network; Reciprocity; Social Network; Web 2.0.

This paper offers a model of blogging activity in which members of blogging communities derive utility from their blog being read as well as from reading others' blogs. We argue that in this context, an *inverse relation* between content produced and attention devoted to others occurs naturally as a result of a competitive equilibrium in an economy where the currency is *mutual attention*. Such an inverse relation is expressed as follows: "in a network, an agent that offers little content compared to others will need to compensate this by devoting more attention to others in order to maintain her place in the network. Conversely, an agent that offers a lot of content compared to others will devote less attention to others". We are able to to check the model's predictions by using a number of measures of activity and involvement in social relations from data gathered on the activity of 2767 bloggers drawn randomly from LiveJournal. We argue that the empirical patterns of mutual attention in that sample are broadly consistent with our model.

1 Context

In recent years, blogs established themselves as an important way to produce, promote and read content on the Internet, and also as a tool for social networking. Although statistics on blogs and bloggers are notoriously fickle (Bialik, 2005), a number of statistics suggest the importance of the blogging phenomenon. Henning (2005) estimated the number of blogs at 53 million by the end of 2005 (see figure 2 in appendix B). A previous report (Perseus Development Corporation, 2003) estimated that about one third of them were active. Technorati, which ranks blogs by popularity, claimed to track about 113 million blogs in May 2008. A seller of search targeted advertising, Chitika, estimated from its own data that the top 50 thousand blogs in terms of Technorati ranking generated a total of $500 million in ad revenues in 2006, with the top 5000 getting 80% of those revenues.[1] An emarketer survey estimated advertising on social networking sites at $1.2 billion in 2008.[2] A number of companies are involved in the development of blogging software and the management of blogging platforms. Among those are Google's Blogger, Six Apart's Typepad, SUP's LiveJournal, Wordpress, Facebook and News Corp's MySpace. Beyond those companies directly involved in blogging, the influence of blogs is wide ranging, for example in news reporting,[3] but also in a range of other economic activities, such as tourism (Wenger, 2008).

The empirical part of this paper relies on a novel dataset from LiveJournal ("LJ"), a web-based community where Internet users can maintain their blog. This study focuses on LiveJournal because it provides more detailed and easily accessible information on users' activity than other blog hosts. Information on LJ users is accessible on their *public profile*, the content of which is described in appendix A. Some data is provided by default and cannot be hidden by the user: user name, account number, date of creation, status of the account (*i.e.* early adopter, permanent, paid, free, sponsored), name and number of friends and readers, number of posts made, number of comments made and received, *etc*... Other data, such as the blogger's date of birth, location, list of interests, and any additional information, is provided on a voluntary basis.[4]

LiveJournal is essentially an aggregation tool with lock-in effect: LJ provides an environment in which bloggers can read each other's public and private entries, exchange comments, participate in communities, and thus develop relations with other LJ users that are not replicable on another blogging tool. Blogs on LiveJournal tend to belong to a specific genre in blogging, "online diaries" (Krishnamurty, 2002). Online diaries are those blogs that are individual and personal. Those account for the vast majorities of existing blogs (Herring et al., 2004). Our data and our analysis does not therefore necessarily apply to other types of blogs, such as topic-oriented or group blogs, as those tend to be independently hosted or hosted on other blogging platforms. Our analysis is meant to model the networking, community, sharing aspect of blogging. It is not meant as a "general theory" of blogging in all its diverse aspects.

Created in 1999 by Brad Fitzpatrick, LJ is based on open-source code and was initially maintained by a community of volunteers. LJ was purchased in January 2005 by Six Apart, the owners of Typepad – another popular blog host. The profit making aspect of LJ then became more important: "sponsored" (advertising-bearing) accounts were introduced and the discrepancy between services offered to free versus paying users widened. In December 2007, Six Apart sold LiveJournal to SUP, a Russian company that was already managing LJ in Russia, and which removed in March 2008 the option to create free accounts. Widespread protests by users led to the option being re-instated in August 2008.[5]

In February 2009, the number of blogs on LJ totalled more than 18

millions, of which 1.2 million (7%) had been updated in the previous 30 days. Of the top 15 countries, 63% of users were located in the United States, 13% in the Russian Federation, 6% in Canada and 5% in the UK.[6] The average age of bloggers on LJ was 25, the median was 22 and the mode was 20.[7] Of the 72% of users who chose to reveal their gender to LJ on registration[8], two-thirds were female. This is more than on many other blogging platforms and may be a reflection of the popularity of LJ among online diarists and among teens (teen bloggers and online diarists tend to be female (Herring et al., 2004)).

Much of the analysis in this paper focuses on the lists of *friends* and on the act of *friending*. Those words have a range of different meanings on LJ (Fono and Raynes-Goldie, 2006). At a technical level, a *friend* is a blog the user subscribes to, that is, whose updates appear on the user's "*friends' page*", a page where the entries made by the blogger's friends appear in reverse chronological order. Listing someone as a friend is what is referred to as *friending*. While some LJ friendships reflect "real world" friendships, many are exclusive to LJ. Those are formed and maintained by reading and posting comments on each other's blogs. Friending is a meaningful act. Friending someone means they are able to read one's "*friends-only*" entries, *i.e.* those entries that are not accessible unless one is logged in LJ and one is listed as a friend.[9] Friending is also a public act since other users can observe who is friend with whom *via* the *friend list* on the blogger's public profile. Friending is a costly act, as it commits the blogger to at least browse through and comment on their friends' entries – otherwise, friendships may be withdrawn ('*unfriending*').[10] "Friendship" on LJ therefore acquires a meaning that is not present to the same extent in other blog networks or with RSS readers; it is a public show of confidence, a commitment and a sign of readiness for closer intimacy. Friendships are generally established with the expectation of *reciprocity*. This means that a user usually expects a friend to read her back in return; it also means that an user may be reluctant to friend someone who is unlikely to reciprocate the friendship, and may drop from her friend list those who do not reciprocate her friendship after a while. The extent to which those norms and type of behaviour hold varies however from user to user and from situation to situation.[11] However, LJ users generally attach great significance to the act of 'friending' and of dropping other users from one's friend list. Status within LJ is often linked to the number of users who list you as friend, which leads some users to engage in 'popularity contests'.[12] Many users do not welcome unsolicited friendships, that is, users who list another as friend when that user does not wish

to reciprocate.[13] All this explains why the list of *'friends'* and *'friend of'* (that is, the list of bloggers who list one as *'friend'*, from now on, *'readers'*) is a variable of great interest in our study.

2 Related Literature

As noted by Drezner and Farrell (2008), blogs are 'a major topic for research' and offer 'extraordinarily fertile terrain for the social sciences'. A number of views have been expressed about the role, value and future of blogs and bloggers, in the media, in politics, or as a tool for collaboration and information sharing. Ribstein (2005, 2006) and Lassica (2001) consider blogs as a newly emergent media form, while Lemann (2006) questions their value to journalism. Drezner and Farrell (2008) evaluate blogs as a tool of political influence, while Sunstein (2008) worries that blogs may contribute to a fracture in the political discourse. Schmidt (2007) considers blogging networks as communities of shared practices, with their own rules in selecting blogs to read, interacting with other bloggers and choosing what to publish. Huck et al. (2008) are interested in how blogs help consumer choice and affect firms' reputations. Quiggin (2006) argue that blogs are part of the 'creative commons', along with Wikis and open source software.

More closely related to this paper are qualitative studies of bloggers' motivations and quantitative studies of the relation between their activity and the structure of their network of relations. Raynes-Goldie (2004) and Fono and Raynes-Goldie (2006) find that bloggers are interested in producing their own content and opinions on current events, interacting with other bloggers and generating debate on their own opinions, as well as in joining communities of shared interests. Nardi et al. (2004) underline the social aspects and dynamics of online diaries, and speculate, as we do, that "blogging is as much about reading as writing, as much about listening as talking". Bar-Ilan (2005) shows that bloggers act as information hubs with links to a number of topical web sources. Furukawa et al. (2006) reveal that blog entries are primarily read through links from other blogs. Backstrom et al. (2006) observe that links between bloggers can be partly explained through common membership in communities on LiveJournal. Lento et al. (2006) explain that continued activity within blogging networks is positively related to the number of relations established with other bloggers. Mishne and Glance (2006) evidence a relationship between the popularity of a weblog and the number of comments it attracts.

Bachnik et al. (2005) establish that blog networks are only weakly connected, that they have small worlds properties and that large blogging networks have clique properties (*i.e.* have few relations with other networks). Paolillo et al. (2005) determine that LiveJournal users' interests and their network of friends are largely uncorrelated. On the other hand, Kumar et al. (2004) note that a combination of age, location and interests explains a large part of cross-linking patterns between users of LJ.

This paper contributes to the above literature with a network structural perspective inspired by insights from sociology (Granovetter, 1973) and motivated by the growing importance of this field to economics (Gui and Sugden, 2005). We present a model of network structure and formation of links among individuals along the lines of Watts (2001), Jackson (2003) and Newman (2003). We differ, however, from most those papers in that we are particularly interested in developing insights on the structure of *directed* networks, as in Caffarelli (2004) for example. We are interested in what drives reciprocation of links, and in the more general, related issue of the strength of relationships in a network. This is an area of study that has only recently been explored (Brueckner, 2006). The main contribution of this paper is to exploit measures of the *structural characteristics* of bloggers' network **along with** *measures of the type and extent of their activities*. We consider not only the structure of links that an agent maintains, but also their direction, their intensity and the intensity of the activity of the agent. We study the activity – content production and attention devoted to others – of each node – agent – in a context where money plays no role and there is no exchange currency – an agent cannot 'pay' attention she receives from someone with the attention she devotes to another agent. Our study allows us to develop insights into the motivations of bloggers and how those determine their interactions and relations with others: we show that bloggers' posting activity depends on their audience – which would not happen if they only cared about expressing themselves. We also develop insights into blogging norms: a widespread expectation of reciprocity in individual relations between bloggers means that the attention a blogger receives will be proportional to how much attention she devotes to other bloggers. However, we also show that an agent may exchange *attention* received with *content* produced: bloggers are ready to get less attention from a blogger if that blogger provides sufficient content, *up to some limit*. Indeed, we show that large deviations from a pattern of reciprocal *friendship* are sanctioned. This paper thus

contributes along the lines of Dohmen et al. (2009) or Gu et al. (2009) to the literature on how reciprocal behaviour influences the structure of human activity by showing how reciprocity matters empirically in an online setting.

Section 3 presents the model on which we ground our working hypotheses. Those are then tested empirically in section 5 using data described in section 4.

3 A model of reciprocal (in)attention

In the following, we consider a model in which agents derive utility from being paid attention to, and from reading the content of others. In a competitive equilibrium, each individual relations that an agent maintains must give her the same utility. This means that an agent that provides more content than others has to be 'paid' more attention. More content is thus reciprocated with more attention, and vice-versa. The model thus introduces a more general form of reciprocity than if agents were to link only with agents that have the same number of friends as they have, or only link with agents that display their same level of overall blogging activity, or exchange one set of interactions (discussion through comments) with another. The model is well designed for the case of blogging, where indeed a typical blogger's reading list will include a variety of more and of less popular blogs, and of blogs that vary in their level of activity,[14] and with which the terms of trade will thus vary.

We define a value function for each agent belonging to a network as a function of the number of other bloggers she is linked to, of the strength of those links, and of the bloggers' activity. Consider thus representative agent i who is part of a network of N agents who produce their own content and read content generated by others. $e = (e_1, e_2, ..., e_N)$ denotes the vector of content produced by agents in the set $\mathcal{N} = \{1, 2, ..., N\}$ consisting of all agents in the network. $n = (n_{ij})_{i,j\in\mathcal{N},\, j\neq i}$ denotes the vector of attentions – for example, agent i devotes attention n_{ij} to the content produced by $j \neq i$. Denote $n_{.i} = (n_{ji})_{j\in\mathcal{N},\, j\neq i}$ and define $N_{.i} = card(j \in \mathcal{N} \mid n_{ji} > 0)$ the cardinality of the subset of $\mathcal{N}$ consisting of agents in $\mathcal{N}$ who read content produced by i. We assume free entry and perfect information in the network, which implies that new agents may enter at no cost and all agents know $\mathcal{N}$, e and n.

A simple additive form[15] for the total utility of a representative agent i is

$$U_i(n,e) = \underbrace{\lambda_i \sum_{j \neq i} n_{ji} e_i}_{\text{Utility from being read}} + \underbrace{\sum_{j \neq i} n_{ij} e_j}_{\text{Utility from reading others}} - \underbrace{C(e_i)}_{\text{Cost of production}} \quad (1)$$

subject to $\sum_{j \neq i} n_{ij} \leq T_i$.

$\lambda_i \geq 0$ measures the propensity to enjoy being read compared to the propensity to enjoy reading others (normalised to 1 here). $C(.)$ is the cost of content production, increasing and convex in the content produced . T_i is the attention budget of agent i, *i.e.* the total attention that she can devote to her friends.[16]

3.1 Choice of effort

To begin with, consider the choice of effort by representative agent i, taking the vector of attention n as given. Maximising $U_i(n,e)$ with respect to e_i, we find that at the optimum, given the concavity of the maximisation problem, one obtains $C'(e_i) = \lambda_i \sum_{j \neq i} n_{ji}$. As the cost function is convex, this means that the higher the sum of attention devoted to agent i, the more the agent will devote effort to his blog. In order to draw further implications, from this, we will make the following assumption:

Assumption 1. *The sum of attention devoted to* i, $\sum_{j \neq i} n_{ji}$, increases in the number of her readers, $N_{.i}$.

If the above assumption is verified, then the effort by i into blogging will be increasing in the number of his readers, which may be tested rather easily using the data available. Is that assumption reasonable however? That is, will an agent with more readers necessarily be receiving more attention in the aggregate? This depends on how elements in the vector of attention $n_{.i}$ vary with $N_{.i}$. Suppose for example that the distribution of the elements of $n_{.i}$ is *independent* of i's number of readers. Then, the higher one's number of readers, the higher the attention one receives, and the higher one's effort level. However, this independence assumption will not be verified if for example agents link exclusively with agents that have the same number of readers, or the

same quality or quantity of content as they have. Suppose for example blog networks were perfect cliques, *i.e.* all agents within the network were linked with each other, and none have links outside the clique. Assume also all agents in the network devote the same time budget T to blogging and devote equal attention to each of their friends. Then $n_{ji} = T/N$ for any j in the clique. We then have $C'(e_i) = \lambda_i T$, and therefore, effort in a clique is unrelated to the number of members of that clique. Note however that in reality, blogging networks have 'small world' properties that combine heavily interlinked individuals and links with a number of individuals in other networks (Bachnik et al., 2005). There may be other reasons why the distribution of elements in $n_{.i}$ would depend on i's number of readers. Some readers may devote less attention to more popular blogs and more attention to obscure blogs – out of a form of snobism, due to the attraction of exclusivity, or because those offer better chances of reciprocated attention. However, one would expect that such individual effect would be more than compensated by the additional attention devoted by new readers as the number of readers increases. Subject to the above caveats, we will therefore express hypothesis *H1,* according to which one should observe a positive relation between content produced and number of readers:

Hypothesis H1 (Network size): *Bloggers with more readers display higher levels of content production and general blogging activity*.

Note that *H1* could also be derived from a model of audience whereby readers are attracted to bloggers who produce more and better content. However, such a model would not in our opinion adequately reflect the dynamics of blogging. For example, such a model would neglect relevant issues that make the expression of *H1* non-trivial, such as the issue of how elements in $n_{.i}$ vary with $N_{.i}$. Furthermore, such a model would not allow us to derive further implications as below.

3.2 Mutual attention

Consider now the intensity of relations between agents. Consider entrant i who has the choice between establishing a link with agent j or an agent k. Suppose i takes the vector of effort e as given, *i.e.* i neglects the effect of his choice on the effort exerted by each agent, as well as how their choice to reciprocate or not will affect his own effort.[17] Agent i prefers establishing the link with j if the gain in utility from doing

so, $\lambda_i n_{ji} e_i + n_{ij} e_j$ (the first part is what is gained from being read by j, the second is what is gained by reading j), is more than the gain in utility from establishing a link with k, $\lambda_i n_{ki} e_i + n_{ik} e_k$. With free entry and perfect information about attention and effort exerted by all agents in the network, the surplus gained from creating a link should be the same across all agents. If that was not the case, then any agent who offers better surplus would keep on gaining friends at the expense of others until a new equilibrium was reached where equality was restored. Therefore, it must be that the surplus obtained from j and from k is equal, so that $\lambda_i n_{ji} e_i + n_{ij} e_j = \lambda_i n_{ki} e_i + n_{ik} e_k$ which can be rewritten as

$$\lambda_i (n_{ki} - n_{ji}) e_i = n_{ij} e_j - n_{ik} e_k \tag{2}$$

Note that since this relation holds at the margin ('marginal' friend), it also determines the relation between number of friends, readers and activity in the aggregate. In order to draw some implications from this this expression, we will make the following assumption:

Assumption 2. *For any i and j, $n_{ij} e_j$ is non-decreasing in e_j.*

This assumption simply states that the utility derived by i in reading the content produced by j does not decrease as j puts more effort into his blog. Note that this does not necessarily mean that n_{ij}, the attention devoted by i to j, will increase in e_j, only that their product will increase in e_j. Note also that an increase in e_j does not necessarily mean more posts or more comments – in which case, beyond a certain point, saturation sets in and i may stop actively following j altogether –, but rather more effort into producing better, more interesting (not necessarily longer) posts, or more perceptive and helpful comments. From this point of view, it is hard to argue against this assumption.

Suppose this assumption is verified, and suppose that $e_j > e_k$ (agent j offers more or better content or interactions). Then, from formula (2), I must have $n_{ki} > n_{ji}$. This means that agent k, who produces less content than agent j, must devote more attention to i than agent j does in order to be kept in the network of i. Conversely, agent k, who devotes more attention to i than agent j does, need not produce as much content than agent j in order to be kept in the network of i. Note that this relation also holds in the extreme case in which $n_{ji} = 0$ (j does not reciprocate i's friendship). In that case, the non-reciprocating j should be observed to produce more or better content than an otherwise similar reciprocating agent.

One should thus observe the following inverse relation between content produced and attention devoted to others: *an agent that offers comparatively little content will devote more attention to others than an agent that offers a lot of content compared to others.* Such an inverse relation occurs naturally as a result of the competitive equilibrium in a market where mutual attention is being exchanged. Agents do not build or sustain links that are not balanced in the way expressed above. Note however that this inverse relation may also emerge not through a competitive process, but out of a sense of fairness, whereby each blogger is required to contribute equally, though in different ways, to the maintenance of the network of relations they belong to.

What we observe is how many bloggers follow a blogger *vs.* how many that blogger follows. The number of blogs one follows is only an imperfect signal of how much attention is devoted to each blogger on one's reading list, but it is reasonable to expect, everything else being equal, that bloggers with more friends will have less time to devote to each than bloggers with less friends. A bloggers who maintains a balance between readers and friends (readees) is also more likely to be reciprocating at the individual level than an agent with more readers than friends. Consider indeed two agents, i and j, similar in all respects except that while both have N friends, i has M readers while j has $M+1$. This means that j reciprocates the readership of at least one less reader than i does. This means that everything else being equal, failing to reciprocate at the individual level is reflected in an increase in the number of 'readers' *vs.* the number of 'friends' at the aggregate level. We can thus spell out hypothesis *H2* as follows:

Hypothesis H2: (Aggregate reciprocity): *Bloggers with a high ratio of readers to friends will be observed to produce more content than otherwise identical bloggers with a lower ratio.*

Note that we do not attempt to explain why some agents would produce more content than others. It might be that some agents are more proficient at or have more of an inclination in doing so. As hypothesized in *H1*, this gains them a bigger audience. For various reasons, a bigger audience may tend to be less reciprocated than a smaller audience – because of limits on the attention a blogger can devote to other bloggers, for example. What *H2* states however is quite different from this: we say that bloggers who have a positive imbalance between readers and friends will produce more than those with a balanced friend list, *irrespective of how many readers or friends they have*. That is, a blogger

with 10 friends and 15 readers will produce more than a blogger with 10 friends and 10 readers, in exactly the same way as a blogger with 100 friends and 150 readers will produce more than a blogger with 100 friends and 100 readers.[18] This means that *H2* is not merely a consequence of *H1*, but is rather a separate consequence of our model.

In what follows, we identify relations between network size, structure and content production in a sample of users of LiveJournal, and check whether hypotheses *H1* and *H2* are verified. We consider the relation between how many readers a blogger has and how much content she produces, and between the level of aggregate reciprocity within an agent's network and how much content is produced by that agent.

4 The data

The data used in this study are observations on the list of friends, readers and posting activity of 2767 bloggers on LiveJournal. This sample was selected using a script that chooses bloggers at random.[19] The data was then collected using `Screen-Scraper`, a software that extracts content from websites and adds it to a database.[20]

This article uses regression techniques to study the relationship between network size, bloggers' activity, and reciprocity. Network size is measured by one bloggers' friends and subscribers (readers). Measures of bloggers' activity are as follows:

1. Commitment, as measured by the length of time the blog has been active (*duration*).
2. Content production, as measured by the number of posts per day (*entries per day*).
3. Intensity of interactions, as measured by
   (a) How many comments are received (per post).
   (b) How many comments are posted (per friend).
   (c) How many communities the blogger belongs to.

In the analysis, posting activity is normalized by length of activity (entries per day). This allows us to distinguish the genuine "intensity" of the posting activity from the mere effect of blogs' duration. (For example, a blog that is active for a long time may accumulate a high number of posts, even though the posting activity is low, whereas a blog that

has been updated for a shorter amount of time may post more actively but total a much lower number of entries.) [21] Among measures of interaction, one should note that the number of communities one belongs to may have an ambiguous effect. This is because a blogger involved in many communities may have less or more readers: communities may draw attention away from personal blogs, but may also represent a way to set up relations based on common interests.

The descriptive statistics of the variables used in the analysis are given below:

Variables	Mean	Median	St. dev	Min	Max	obs
Friends	95	41	175	0	1944	2340
Readers	124	39	350	1	7855	2092
Number of entries	871	390	1596	0	24094	2340
Comments received	4041	705	10910	1	239564	2340
Comments posted	3855	1022	7624	0	106505	2337
Member	27	14	39	1	506	1463
Entries per day	1.42	0.45	2.74	0	19.90	2340
Comments received per post	3.65	2	5.57	0	75	2336
Comments made per friend	44.21	20.70	98.03	0	3109	2296
Duration	1198	1165	818	10	3185	2336

Table 1: Summary Statistics.

Two noticeable features of the data are skewness and large standard deviations. So, in what follows, the "typical" user is described by median values.[22] Our blogger lists 41 *friends* (*i.e.* she reads the blogs of 41 LJ users) and, in turn, she is subscribed by 39 bloggers (*readers*), which highlights a considerable level of aggregate reciprocity. The number of *comments posted / received* is also remarkably balanced. The blogger follows, and is *member* of, 14 *communities*. She created 390 *entries* (posts) since the blog's inception. The *duration*, a measure of the blogs' lifetime, is the length of time between the creation of the blog and its last update. Duration is about 1165 days (*i.e.* more than 3 years). A

new entry is typically added every 2 days (*entries per day*). Individual posts receives 2 comments (*comments received per post*), whereas the blogger makes about 20 comments on the journals of each of her friends (*comments made per friend*). This data evidences bloggers' considerable commitment, although the frequency of updating and the posting activity varies greatly among users.

To filter out mechanical agents, we have restricted attention to those blogs with a duration of at least 10 days, which post less than 20 journal entries per day. (Appendix A describes the variables in the dataset in greater detail and compares the sample's characteristics with those of all accounts created on LJ since its beginning.)

In view of their skewness, data were transformed on the logarithm scale for performing regression analysis. Computations were carried out using the econometrics and statistical softwares `Stata` and `S-Plus`.

5 Empirical analysis

This section analyses bloggers' network properties, and verifies implications of the theoretical model presented in section 3, namely hypothesis H1 and H2, which we recall below:

H1 Bloggers who display higher levels of content production and general blogging activity have more readers.

H2 Bloggers with relatively more friends than readers produce less content than other bloggers (viceversa, bloggers with relatively less friends than readers produce more content than others).

To verify these hypotheses, we estimate a set of *activity equations* in which the number of readers and the *ratio* of readers to friends are regressed on measures of bloggers' effort, such as commitments to one's blog, content production, and intensity of interaction with other bloggers. We also consider the role of measures of reciprocal attention and network symmetry.

Figure 1 presents a scatterplot of the number of friends against the number of readers. This graph offers a preliminary idea of the structure of the network, and helps to motivate the analysis. The relation between the two variables is very strong, with the observation points concentrated at a — nearly — 45 degrees line.[23]Most bloggers thus have similar number of friends and readers, which we can interpret as evidence of reciprocal attention in the aggregate. We also observe that

close to zero, observations are more dispersed and that there are more values of *readers* associated to a specific value of *friend* than conversely.

The graph highlights several data points (denoted by letters A-B-C-D) which corresponds to agents whose relative location motivates the following analysis. From hypothesis 1, agent B, with more readers than agent A, should be more active than agent A. From hypothesis 2, agent C, with the same number of readers than agent D but less aggregate reciprocation (less friends), should be more active than agent D. The following investigates determinants of a bloggers' network size, *i.e.* their position on the regression line; it also investigates exceptions from the reciprocity of attention, and checks whether such departures can be accounted for by content production.

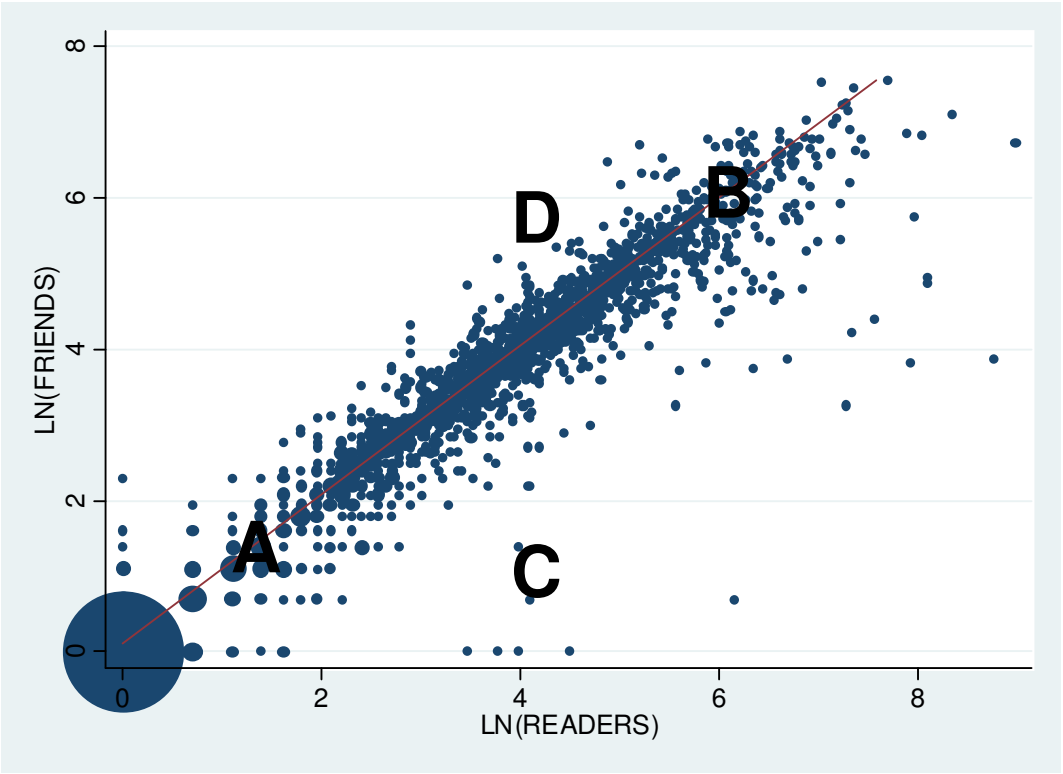

Figure 1: Scatterplot of friends vs. readers in a sample of users of LiveJournal

Note: Expressed in natural logarithm. Circle size indicates number of observations at that point. For reference, there are 335 observations at data point (0,0).

5.1 Testing the first hypothesis

In what follows, we regress the numbers of *readers* on several indicators of bloggers' activity, to check whether higher levels of content production and activity increase the network size (H1).The model is as follows:

$$ln(readers) = \alpha + \beta_1\ ln(member) + \beta_2\ ln(entries) + \\ +\beta_3\ ln(comments\ received) + + \beta_4\ ln(comments\ posted) + \beta_5\ ln(duration) + \epsilon; \quad (3)$$

Here, *member* denotes the number of communities one is a member of, *entries* the number of entries per day, *comments received* the number of received comments (per post), *comments posted* the number of comments posted (per friend), and *duration* the duration in days; ϵ is an *iid* error term. A preliminary analysis of residuals and leverage revealed several outlier observations, which we removed.[24]

The first column of table 2 presents results for the regression of equation 3.

Dependent variable:	*ln(readers)*	*ln(readers)*	*reciprocity*
ln(member)	0.153^{***}	0.170^{***}	-0.035^{***}
	(15.80)	(18.16)	(-4.20)
ln(entries)	0.585^{***}	0.580^{***}	0.012
	(42.21)	(44.45)	(1.05)
ln(comments received)	0.776^{***}	0.742^{***}	0.072^{***}
	(56.42)	(57.73)	(4.98)
ln(comments posted)	-0.466^{***}	-0.526^{***}	0.130^{***}
	(-30.18)	(-36.91)	(9.56)
ln(duration)	0.597^{***}	0.600^{***}	-0.007
	(30.20)	(31.25)	(-0.40)
reciprocity		0.474^{***}	
		(14.64)	
constant	0.830^{***}	0.975^{***}	-0.305^{***}
	(6.72)	(8.07)	(-2.57)
obs	1321	1321	1321
R^2_{adj}	0.880	0.900	0.230
F stat	1493 (0.000)	1621 (0.000)	46.70 (0.000)
BP	29.56 (0.000)	70.46 (0.000)	29.70 (0.280)
RESET	4.81 (0.000)	5.40 (0.001)	6.25 (0.431)

Table 2: Multiple regressions with measure of reciprocity.
Legend: robust t-statistics are in parentheses; p-values for F, BP and RESET statistics in parentheses.

Measures of goodness-of-fit suggest the model is quite successful at describing the data: the R^2 shows that the equation explains a great proportion of the variation in *readers*, and the F statistic (F stat) for the overall significance of the regression rejects the null that the slope coefficients are jointly zero at any conventional level of significance. The analysis of regression errors reveals certain degrees of non-normality and heteroskedasticity in the data, which is confirmed by the Breusch-Pagan (BP) test statistic. To correct for the potential loss of efficiency, we compute coefficients' t-statistics using White's robust covariance

matrix estimator.

One can see that all coefficients are significant. The largest coefficients are associated with number of comments per post and duration: a 1% increase in these variables lead, respectively, to about a 0.8% and a 0.6% increase in the number of readers, *cæteris paribus*. The smallest effect is that of the number of communities the blogger belongs to. This seems to confirm the ambiguous effect of community membership on the intensity of interaction, which was noted in the previous section 4. Bloggers with more readers write more entries per day (more active), write less comments per friends (reduction in attention given, showing that indeed, as posited, bloggers who are more popular can devote less attention on average to each of their friends than other bloggers can[25]) but receive more comments per post (increase in attention received). The signs of the coefficients are thus consistent with intuition and the prediction of *H1*: an important characteristic of one blogger's network size — ie the number of subscribers, or readers — is positively related to content production and activity.

The norm of reciprocity

Before estimating the second activity equation, we consider whether adherence to a possible *norm of reciprocity* may affect a blogger's number of readers. Bloggers who do not reciprocate readerships may have less readers because bloggers expect readership to be reciprocated and do not therefore maintain unreciprocated relations, or because bloggers are deterred from subscribing to those blogs which have more readers than readees because this means that chances of reciprocation are low. To do so, we estimate a version of the activity equation (3) which includes a measure of reciprocal attention, *reciprocity*. This is computed by taking the logarithm of the ratio of readers to friends, as follows:

$$\textit{reciprocity} = \ln(\text{readers}/\text{friends}) \tag{4}$$

One can see that reciprocity is zero when the number of friends equals the number of readers. Furthermore, increases in *reciprocity* indicate that the number of readers increases relatively to the number of friends.

The second column of table 2 reports estimates of the activity equation (3) with the added variable. Reciprocity has a positive coefficient, suggesting that lower aggregate reciprocation does not negatively affect network sizes. Sign, size and significance of other variables' coef-

ficients are not substantially altered by the inclusion of the new variable. This suggests that the information contained in *reciprocity* is incremental to that provided by other regressors. Controlling for its effect slightly improves the goodness-of-fit of the model. The value of the Breusch-Pagan test statistic, however, increases considerably, and the analysis of partial residuals casts doubts on the explanatory power of the added variable.[26]

5.2 Testing the second hypothesis

To examine the second hypothesis (*H2*), which relates content production and ratio of readers to friends, we estimate a second activity equation, in which *reciprocity* is regressed on the various activity measures listed in section 4. The model is as follows:

$$reciprocity = \alpha + \beta_1\, ln(member) + \beta_2\, ln(entries) + \beta_3\, ln(comments\ received) + \\ + \beta_4\, ln(comments\ posted) + \beta_5\, ln(duration) + \epsilon; \qquad (5)$$

Explanatory variable are as in the model of equation 3. Estimation results are given in the last column of table 2. Compared to the regression of *readers* over activity, coefficients are smaller in size, and number of posts and duration are not significant. Interestingly, however, measures of posting activity and number of entries (albeit the latter not significant) are positively related with reciprocity, as predicted by H2. In particular, the coefficients of comments posted, which displays the most significant and largest effect on reciprocity, enters the estimated equation with a positive sign. We note that this variable is mostly related to the level of blogs' *interactivity*, in that it measures the activity of the blogger in *other users*' blogs, as opposed to indicators of activity in her own blog (such as, for example, entries posted). This can be interpreted as evidence that willingness to interact does affect network patterns, and increases the number of readers relatively to number of friends. Duration and community membership have negative coefficients, but those are small and, for duration, insignificant.

The regression explains about 27% of the variation in the dependent variable, which is perhaps disappointing. The F test statistic for the regression decisively rejects the null of joint lack of significance of the regressors. Notably, the Breusch-Pagan test does not reject its null of constant variance.

In summary, results from this analysis show that there exist a positive and statistically significant relation between level of activity and

number of readers, confirming hypothesis (*H1*). The evidence in favor of hypothesis (*H2*) is less favorable, but offers some interesting insights. Noticeably, measures of posting activity and number of entries (albeit the latter not significant) are positively related with reciprocity, as predicted by H2.

One problem with the analysis of this section is that network structures are characterised by a large degree of endogeneity (Bramoullé and Fortin, 2009; Mihaly, 2007; Weinberg, 2007). For example, the estimation of the first activity equation showed that a *norm of reciprocity* may affect 'friending' patterns. In turn, this variable, by definition, depends on readership and friendship. This circularity may lead to biased estimates. To investigate the possibly endogenous effect of reciprocity, and handle the endogeneity bias, the following uses instrumental variables' techniques.

5.3 Instrumental Variable Estimation

This section applies a version of the instrumental variable technique to the estimation of the first activity equation (eq. 3). This allows us to test for the endogeneity of the reciprocity measure, and to treat it as an exogenous variable in the model of the number of readers. Results are shown in table 4, in appendix C, along with OLS estimates for comparison.[27]

Results evidence that endogeneity is a concern in this dataset.[28] IV coefficients are consistently higher than those produced by OLS estimation, although the sign of the activity measures' coefficients does not change. The most striking result refers to reciprocity: not only does its effect turn positive, but it also enters the activity equation with a large elasticity of about 3.5. *Hausman*'s test statistics rejects its null of exogeneity of reciprocity at every significance level, which supports the adequacy of the IV procedure.

Interestingly, the above results are consistent with bloggers attaching a 'stigma' to failing to reciprocate. Indeed, the negative sign of reciprocity implies that, when the number of readers increases relatively to the number of bloggers who are befriended, then the number of readers (and of friends) is lower than what measures of activity would predict. It may be, as conjectured previously, that bloggers do not want to friend bloggers who do not adhere to the norm of reciprocity and thus appear unlikely to reciprocate. However, it may also be that a blogger with many readers may reach a limit on how many readers she can add

back as friends and reasonably follow, and thus be less likely to reciprocate beyond that limit. It could also be that those bloggers who do not adhere to the norm do not care about how many friends they get, which is why they get less of them.

In the following, we consider a different indicator of network structure to check the robustness of our results: We examine whether imbalances in the bloggers' networks may explain more than aggregate reciprocity. It may indeed be that networks that are balanced are more attractive, while imbalances in any direction are stigmatized.

5.4 The effect of imbalances

This section explores how imbalances in friendships relates to blogging activity's measures and network size. *Reciprocity* measured the number of readers *relatively* to the number of friends. It did *not* tell us, however, whether the network became more or less asymmetric. As a result, the interpretation of the effect of reciprocity in the activity equations was ambiguous. Quite apart of the relative number of friends vs readers, bloggers may react to whether a fellow blogger maintains a balance between *friends* and *readers* or not. For example, a blogger who friends too many bloggers relative to how many read her back in return could be seen as too eager, or indifferent to the act of reciprocation, and thus get less readers than her activity would suggest.

We therefore consider the effect of a measure of network imbalances, called *asymmetry*, and defined as follows:

$$\textit{asymmetry} = \ln(1 + |readers - friends|) \tag{6}$$

Here, one can see that any departure from zero signals an increase in the asymmetry of the network.

Using the measure of imbalance given above, we re-estimated the activity equations of previous sections. Table 5 in appendix D presents results from this estimation. One can see that results for the first activity equation are comparable with those reported in table 2, except that the coefficient on *asymmetry* is lower than the coefficient on *reciprocity*.[29] On the contrary, results for the second activity equations differ substantially when the different measure is considered. When considering asymmetries, effects of activity measures are sizeable and all significant. One can see that the coefficient on comments posted is now negative, indicating that its positive effect on reciprocity was not a robust result.

In summary, the above results shows that the degree of asymmetry of the network is positively related to activity, and negatively related to the number of comments posted on other bloggers' article. So, the more a blogger posts comments on other bloggers entries, the least the network is asymmetric, which seems reasonable. It is not clear, however, why other activity and interaction variables enter the equation with a positive sign. A possible explanation is that some bloggers "friend" many bloggers with the hope of attracting readership through a high level of activity, while reciprocation of friendship comes only with a lag. This could be checked with a panel data set, and is left for future research.

6 Conclusion

This paper analysed patterns of relationship and content production among bloggers from a theoretical and empirical perspective. The analysis has identified statistically significant positive relationships between the size of and degree of reciprocity within a blogger's network of relations, and her blog's durability, intensity of activity and degree of interactivity. Main results are as follows:

- Posting activity and intensity of interaction are positive determinants of network size;
- Departures from aggregate reciprocity can be accounted for by content production;
- Failure to reciprocate attention is sanctioned with a lower popularity than other measures of activity might normally warrant.

These results suggest that bloggers who produce more content devote less attention to others. Furthermore, bloggers sanction deviations from the norm of reciprocity, which occur when a blogger does not return friendship as expected.

This analysis has several limitations. Firstly, because it is not possible to observe the entire network, the empirical analysis relies on a random sample, albeit representative of the wider community. Second, stylized facts summarised above concern aggregate behaviour. In addition to aggregate data, the availability of individual data would help assessing the predictions of the theoretical model concerning reciprocal attention. For example, one could verify directly whether blogs'

subscriptions are indeed reciprocal. Another problem is to determine the direction of causality between number of readers, or network size, and content production. This can be addressed by the estimation of panel regressions, and is left for future research. Furthermore, more research is needed on what determines the reciprocation of relationships in the network. Future work will rely on the collection of data over several periods, and will also rely on the gathering of further quantitative and qualitative information, such as blogs' rankings on search engines and differences in bloggers' attitudes and objectives. This will hopefully enable us to address those difficulties.

Notes

[1] `http://www.scribd.com/doc/219285/Blogging-Revenue-Study`, accessed February 21, 2009.

[2] `http://www.emarketer.com/Article.aspx?id=1006799`, accessed February 21, 2009.

[3] Mainstream news coverage has controversially relied on (micro)bloggers in its coverage of the Mumbai terrorist attacks or of the Green Revolution in Iran. A May 2008 survey by Brodeur, a unit of Omnicom Group, found that journalists made use of blogs for their news report, felt that blogs influenced the focus and brought diversity to news, but also felt that they lowered the quality and accuracy of news reports, as well as the tone of the coverage (`http://www.brodeurmediasurvey.com`, accessed February 25, 2009).

[4] Users can provide additional information in the "bio", a space where bloggers present themselves.

[5] More information on LJ and its history can be found in its Wikipedia entry (`http://en.wikipedia.org/wiki/livejournal`, accessed February 21, 2009).

[6] Source: `http://www.livejournal.com/stats.bml`, accessed February 9, 2009.

[7] Bloggers do not have to display their age publicly, but their birth date must be provided to LJ for legal reasons.

[8] Data on individual bloggers' gender is not available publicly but is collected by LJ for internal statistical purposes, with an option for the user not to disclose gender on registration.

[9] Not all users choose to make such 'filtered' entries, but a large portion restrict access to at least some of their posts.

[10] "Unfriending" without explanation and due care commonly leads to "drama" on LJ!

[11] For more on the social dynamics of LiveJournal, see Raynes-Goldie (2004) and Marwick (2009).

[12] Other forms of status are associated to the length of time one has been on LJ, to the design of one's LJ, to the identity of one's friends, to the popularity of the communities one maintains, and occasionally, to the quality of one's entries! Status may also be imported from the 'real world'.

[13] A number of tools are available on LJ to prevent unwanted interaction, for example making one's entries friends only, preventing or screening comments by people

other than friends, listing unwanted (unreciprocated) friends in a separate list, banning unwanted friends from commenting in one's journal, *etc*...

[14]This at least is the pattern we observed on LiveJournal (data not reported here for lack of space).

[15]More general utility representations could be adopted and would generate the same set of insights.

[16]We could also define T_i as the total time budget available for blogging, including both content production and attention. We would then express the cost of production in terms of time spent producing and write $\sum_{j \neq i} n_{ij} + C(e_i) \leq T_i$. This is of no consequence in the subsequent analysis however.

[17]Alternatively, one may assume less realistically that agent i is able to predict the result of a whole chain of reaction and counter-reaction to the establishment of this new friendship, and thus knows how e and n come out after she establishes the link. The later expressions of net surplus from a new relation would then thus take account of the fact that additional attention by a new friend i may lead a blogger to increase his or her own activity and modify the attention she gives to other agents in the network.

[18]We will see that both the ratio of readers over friends (1.5 in both cases) and the difference between readers and friends (5 in the first case, 50 in the second) are related to content production.

[19]`http://www.livejournal.com/random.bml`

[20]`http://www.screen-scraper.com`

[21]When considering the effect of blogs' lifetime, one should also note that a blogger who has been updating for a long time is likely to accumulate many friends, irrespective of his or her level of activity. This is because there is some *inertia* in the friending process on LiveJournal: LJers tend to keep a blogger on their list even after that blogger has stopped updating and as long as that blogger does not drop them. Indeed, some bloggers like to inflate their list of friends and readers and thus may maintain reciprocal links long after they ceased being active.

[22]The median offers a better description of the center of a distribution than the mean when data are skewed, because it is robust to extreme values.

[23]A fitted simple regression line is also reported. The simple regression coefficient on *ln(readers)* equals 0.94, which is highly significant with a t-statistic of 161.63.

[24]The removal of outlier observations follows the procedure proposed by Belsley (1980), which identifies highly influential observations as those characterised by either a high leverage or a high residual.

[25]This partially settles a common proposition according to which some bloggers are more popular and active than others merely because they devote more time to blogging than others do, or because they are better able to maintain relations with many people at the same time, through faster typing for example!

[26]It is possible, from this estimates, to deduce the effect of activity measures and reciprocity on the number of friends. This is done as follows. Consider again the estimated activity equation (3) with reciprocity:

$$ln(readers) = \hat{\alpha} + \hat{\beta}X + \hat{\gamma}(reciprocity),$$

Where X is the set of regressors other than reciprocity. Interestingly, the equation above implies

$$ln(friends) = \hat{\alpha} + \hat{\beta}X + (\hat{\gamma} - 1)(reciprocity) \quad (7)$$

This means that, while reciprocity has a positive and statistically significant effect on the number of readers, its effect on the number of friends is negative ($0.474 - 1 = -0.526$).

[27]The OLS model given in table 4 does not include the variable "comments posted", so that OLS results are comparable to IV results. However, this leads to the exclusion of a relevant variable, therefore to biased results. In interpreting IV results, one should consider that the bias in OLS estimates is induced not only by the endogeneity, but also buy the exclusion of comments posted. This is due to the limited choice of instruments available in the dataset.

[28]The model is estimated using 2 Stage Least Squares procedure. Identification is achieved by excluding *comments posted* from the IV regression, as this variable is highly correlated with the reciprocity measure. More details on this estimation method can be found in Greene (1980), chapter 5.

[29]One should note that the measure of asymmetry enters the readers' equation with a positive sign, even when estimation uses the instrumental variable method. Results for IV regressions are not reported for reasons of space. They are available from the authors on request.

References

Bachnik, W., S. Szymczyk, P. Leszczynski, R. Podsiadlo, E. Rymszewicz, L. Kurylo, D. Makowiec, and B. Bykowska (2005). Quantitative and sociological analysis of blog networks. *Acta Physica Polonica B 36*(10), 3179–3191.

Backstrom, L., D. Huttenlocher, J. Kleinberg, and X. Lan (2006). Group formation in large social networks: membership, growth and evolution. In *Proceedings of the 12th ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 44–54. ACM:New York, NY, USA.

Bar-Ilan, J. (2005). Information hub blogs. *Journal of Information Science 31*(4), 297–307.

Belsley, D. (1980). *Conditioning diagnostics: collinearity and weak data in regression*. Wiley.

Bialik, C. (2005). Measuring the impact of blogs requires more than counting. *The Wall Street Journal*. May 26, `http://tinyurl.com/7luge`.

Bramoullé, Y. and B. Fortin (2009). The econometrics of social networks. CIRPEE Working Paper.

Brueckner, J. (2006). Frienship networks. *Journal of Regional Science 46*(5), 847–865.

Caffarelli, F. (2004). Non-cooperative network formation with network maintenance costs. Working Paper ECO 2004/18, European University Institute.

Dohmen, T., A. Falk, D. Huffman, and U. Sunde (2009). Homo Reciprocans: Survey evidence of behavioural outcomes. *The Economic Journal 119*, 592–612.

Drezner, D. and H. Farrell (2008). Introduction: Blogs, politics and power: a special issue of Public Choice. *Public Choice 134*, 1–13.

Fono, D. and K. Raynes-Goldie (2006). Hyperfriends and beyond: Friendship and social norms on LiveJournal. In M. Consalvo and C. Haythornthwaite (Eds.), *Internet Research Annual Volume 4: Selected Papers from the Association of Internet Researchers Conference*. Peter Lang: New York, USA.

Furukawa, T., T. Matsuzawa, Y. Matsuo, K. Uchiyama, and M. Takeda (2006). Analysis of user relations and reading activity in weblogs. *Electronics and Communications in Japan (Part 1: Communications) 89*(12), 88–96.

Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology 78*(6), 1360–1380.

Greene, W. (1980). *Econometric Analysis*. Prentice Hall.

Gu, B., Y. Huang, W. Duan, and A. B. Whinston (2009). Indirect reciprocity in online social networks - a longitudinal analysis of individual contributions and peer enforcement in a peer-to-peer music sharing network. *McCombs Research Paper Series No. IROM-06-09*. http://ssrn.com/paper=1327759.

Gui, B. and R. Sugden (2005). Why interpersonal relations matter for economics. In B. Gui and R. Sugden (Eds.), *Economics and Social Interactions*, pp. 1–22. Cambridge University Press: Cambridge, UK.

Henning, J. (2005). The blogging geyser. *Newsletter of the Web Marketing Association*. April 8, `http://www.webmarketingassociation.org/wma_newsletter05_05_iceberg.htm`.

Herring, S., L. Scheidt, S. Bonus, and E. Wright (2004). Bridging the gap: A genre analysis of weblogs. In *Proceedings of the 37th Annual Hawaii International Conference on System Sciences*, pp. 101–111.

Huck, S., G. Lünser, and J.-R. Tyran (2008). Consumer networks and firm reputation: A first experimental investigation. Technical Report 6624, CEPR.

Jackson, M. O. (2003). A survey of models of network formation: Stability and efficiency. In G. Demange and M. Wooders (Eds.), *Group Formation in Economics: Networks, Clubs, and Coalitions*. Cambridge University Press: Cambridge.

Krishnamurty, S. (2002). The multidimensionality of blog conversations: The virtual enactment of September 11. In *AOIR Internet Research 3.0*.

Kumar, R., J. Novak, P. Raghavan, and A. Tomkins (2004). Structure and evolution of blogspace. *Communications of the ACM 47*(12), 35–39.

Lassica, J. (2001). Blogging as a form of journalism. *Online Journalism Review*. May 24, `http://www.ojr.org/ojr/workplace/1017958873.php`.

Lemann, N. (2006). Journalism without journalists. *The New Yorker*. August 7, `http://www.newyorker.com/archive/2006/08/07/060807fa_fact1`.

Lento, T., H. Welser, L. Gu, and M. Smith (2006). The ties that blog: Examining the relationship between social ties and continued participation in the Wallop weblogging system. In *3rd Annual Workshop on the Weblogging Ecosystem*.

Marwick, A. (2009). LiveJournal users: Passionate, prolific and private. Research Report. December 19, `http://www.livejournalinc.com/press_releases/20081219.php`.

Mihaly, K. (2007). Too popular for school? Friendship formation and academic achievement. Department of Economics, Duke University Working Paper NC27708.

Mishne, G. and N. Glance (2006). Leave a reply: An analysis of weblog comments. In *3rd Annual Workshop on the Weblogging Ecosystem*.

Nardi, B. A., D. J. Schiano, and M. Gumbrecht (2004). Blogging as social activity, or, Would you let 900 million people read your diary? In *Proceedings of the 2004 ACM conference on Computer Supported Cooperative Work*, pp. 222–231. ACM Press.

Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review 45*, 167.

Paolillo, J., S. Mercure, and E. Wright (2005). The social semantics of LiveJournal FOAF: Structure and change from 2004 to 2005. In G. Stumme, B. Hoser, C. Schmitz, and H. Alani (Eds.), *Proceedings of the ISWC 2005 Workshop on Semantic Network Analysis*.

Perseus Development Corporation (2003). The blogging iceberg. October 6, `http://tinyurl.com/ceepn4`.

Quiggin, J. (2006). Blogs, wikis and creative innovation. *International Journal of Cultural Studies 9*(4), 481–496.

Raynes-Goldie, K. (2004). Pulling sense out of today's informational chaos: LiveJournal as a site of knowledge creation and sharing. *First Monday 9*(12). `http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1194/1114`.

Ribstein, L. E. (2005). Initial reflections on the law and economics of blogging. University of Illinois, `http://law.bepress.com/uiuclwps/papers/art25/`.

Ribstein, L. E. (2006). From bricks to pajamas: The law and economics of amateur journalism. *William & Mary Law Review 48*, 185–249.

Schmidt, J. (2007). Blogging practices: An analytical framework. *Journal of Computer-Mediated Communication 12*, 1409–1427.

Sunstein, C. (2008). Neither Hayek nor Habermas. *Public Choice 134*(1-2), 87–95.

Watts, A. (2001). A dynamic model of network formation. *Games and Economic Behavior 34*, 331–334.

Weinberg, B. (2007). Social interactions with endogenous association. NBER Working Paper 13038.

Wenger, A. (2008). Analysis of travel bloggers' characteristics and their communication about Austria as a tourism destination. *Journal of Vacation Marketing 14*(2), 169–176.

A Data Description

A.1 Original data

- User: User name (pseudonym)
- Location: Region and /or country where the blogger is based.
- Friends: Number and list of weblogs read by the blogger. Limited to other blogs on LJ.
- Readers (or 'friend of' in LJ terminology): List of those bloggers with an account on LJ who read one's weblog. This can be divided between:
  - Mutual friends: A subgroup of 'readers': Number and list of those bloggers whose friendship is reciprocated. This statistic is not provided as a default and must be activated by the user.
  - Also friend of: A subgroup of 'readers': number and list of those bloggers whose friendship is not reciprocated. Again, this statistic is not provided as a default and must be activated by the user.
- Communities: Number and list of communities the blogger reads. Communities are blogs with a specific theme to which all members can contribute posts and comments.
- Member of: Number and list of the communities one is member of. Differs from 'communities' in that one can read a community without being a member of it (but one generally cannot contribute if one is not a member).
- Posting access: Differs from 'member of' in that one can be a member of a community but not have access to posting there.

- Feeds: Number and list of those weblogs not on LiveJournal that are read by the blogger via LJ. Those can be read via their RSS feed and appear on the blogger's 'friends' page' (list of entries by friends).
- Account type: Accounts, can be 'free', 'sponsored', 'paid'; 'permanent' or belong to 'early adopters'. 'Early adopters' are the first few members of LJ. 'Paid' accounts give access to the full range of LJ's services and do not display any advertising. 'Permanent accounts' are accounts that are paid for life. 'Sponsored' accounts' display advertising. 'Free' accounts displays less advertising than sponsored accounts but have reduced functionality.
- Date created: Date on which the weblog was created.
- Date updated: Last date on which the weblog was updated (*i.e.* when an entry was last posted).
- Journal entries: Number of posts written since the weblog was created
- Comments posted: Number of comments made on entries in other weblogs or communities.
- Comments received: Number of comments made by other bloggers on one's own entries, and own comments in reply to those.

A.2 Processed data:

- Days since creation: Difference between date of data collection and date of creation of the blog (in days).
- Days since update: Difference between date of data collection and date of the last update (days).
- Duration: Difference between date of creation and date of last update (days).
- Active: 1 if weblog was updated less than 8 weeks ago, 0 otherwise.
- Entries per day: Number of journal entries divided by duration

- Comments per post: Comments *received* divided by number of posts
- Comments per friends: Comments *made* divided by number of friends.
- Reciprocity: Readers divided by friends, expressed in logarithm.

A.3 Representativity of the sample

Table 3 compares features of the randomly selected bloggers to those of LJ, for an informal check of the representativeness of the sample:[30]

	Random sample	LiveJournal
Updated last month	100%	7%
Updated last week	100%	3%
Updated on the day	61%	1%
Countries	US: 38%, Russia: 31%, Ukraine: 8%, Canada: 3%, UK: 3%.	US: 63%, Russia: 13%, Canada: 6%, UK: 5%.
Age (in years)	Average: 30, Median: 27, Mode: 24.	Average: 25, Median: 22, Mode: 20.

Table 3: Comparison table

One can see that the stock of bloggers on LJ are young and predominantly located in the US. The random sample is essentially a representation of active bloggers on LJ. This is because the random script provided by LJ is designed to select active blogs in order to spare the user having to sift through inactive blogs. The distribution of nationality thus reflects countries in which LJ is presently popular (Russia and Ukraine), rather than LJ's country of origin (the US), where competition from Bebo and Facebook dented LJ's popularity among high school and college students respectively. This is also why the average age of bloggers in our sample is higher than in LJ's stock.

B Figures

Figure 2 represents the evolution of the number of blogs from 2000 to 2005 (in logarithmic scale).

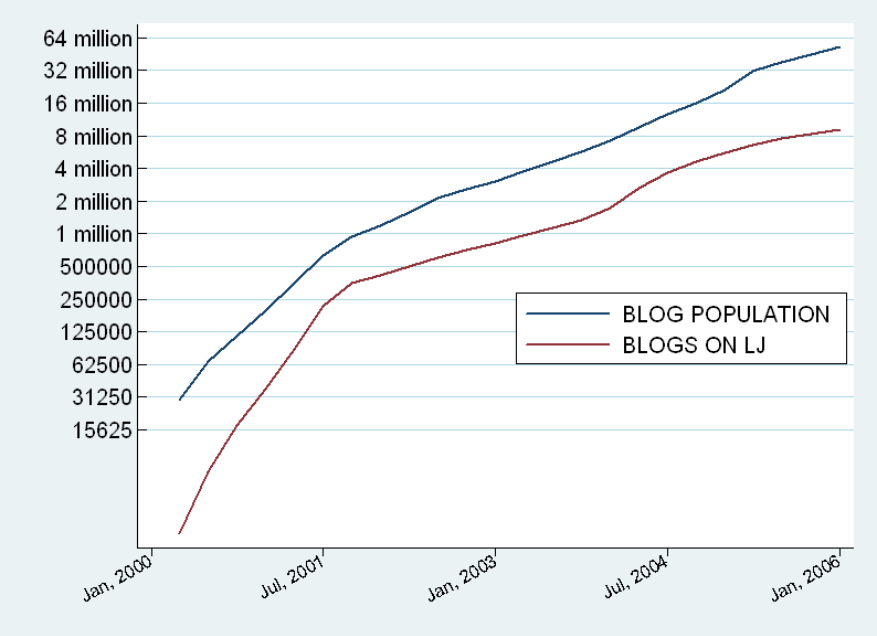

Figure 2: Number of hosted weblogs created between 2000 and 2005.
Source: Henning (2005) and LJ statistics (http://www.livejournal.com/stats/stats.txt), both accessed February 20, 2009.

C Instrumental variable estimation of number of readers

Table 4 shows the result of the instrumental variable estimation of the number of readers.

ln(readers)	IV	OLS
ln(member)	0.024	0.190***
	(0.62)	(12.99)
ln(entries)	0.630***	0.390***
	(13.20)	(21.41)
ln(comments received)	1.040***	0.607***
	(17.53)	(27.89)
ln(comments posted)	(instrument)	
ln(duration)	0.570***	0.236***
	(7.71)	(8.97)
reciprocity	−3.680***	0.060
	(-10.40)	(1.12)
constant	−0.293	1.80***
	(-0.56)	(9.38)
obs	1321	1321
Hausman test	113.80 (0.000)	

Table 4: IV regression of first activity equation: model estimates and comparison with OLS regression.
Legend: Robust $z-$ and $t-$ratios in parentheses.

D The effect of network imbalances

Table 5 shows the results of the estimation of the activity equations using the asymmetry measure.

Dependent variable:	ln(readers)	ln(readers)	asymmetry
ln(member)	0.153^{***}	0.120^{***}	0.193^{***}
	(15.80)	(13.64)	(7.53)
ln(entries)	0.585^{***}	0.490^{***}	0.499^{***}
	(42.20)	(36.32)	(14.90)
ln(comments received)	0.776^{***}	0.672^{***}	0.600^{***}
	(56.40)	(49.60)	(16.75)
ln(comments posted)	-0.466^{***}	-0.375^{***}	-0.526^{***}
	(-34.50)	(-27.80)	(-14.70)
ln(duration)	0.600^{***}	0.515^{***}	0.472^{***}
	(30.20)	(28.00)	(8.99)
asymmetry		0.174^{***}	
		(18.56)	
constant	0.830^{***}	0.780^{***}	0.291
	(6.72)	(6.90)	(0.88)
obs	1321	1321	1321
R^2_{adj}	0.880	0.908	0.403
F stat	1493 (0.000)	2055 (0.000)	156 (0.000)
BP	29.50 (0.000)	23.00 (0.001)	58.83 (0.000)
RESET	4.80 (0.000)	4.70 (0.001)	5.75 (0.000)

Table 5: Multiple regressions with measure of asymmetry.
Legend: robust t-ratios are in parentheses; p-values for F, BP and RESET statistics in parentheses.