

MPRA

Munich Personal RePEc Archive

Blogs and the Economics of Reciprocal Attention

Gaudeul, Alexia and Mathieu, Laurence and Peroni, Chiara

30 March 2008

Online at <https://mpra.ub.uni-muenchen.de/14309/>
MPRA Paper No. 14309, posted 06 Apr 2009 01:18 UTC

Blogs and the Economics of Reciprocal Attention^{*}

Alexia Gaudeul[†], Laurence Mathieu[‡] and Chiara Peroni[§]

March 30, 2009

Abstract

We argue in this paper that attention to one's blog is won by paying attention to other bloggers. We derive properties of blogging networks from a model where bloggers trade attention and content. The predictions from the model are then checked against a novel dataset from LiveJournal, a major blogging community. As predicted, the activity of bloggers is found to be related to the size and level of reciprocity within a blogger's relational network. We also find that bloggers who do not adhere to reciprocity norms are sanctioned with a lower number of readers.

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

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

This paper offers a model of blogging activity in which members of blogging communities derive utility from other people reading their blogs as well as from reading the blogs of other people. We argue that in this context, a *norm of reciprocity* occurs naturally as a result of a competitive equilibrium in an economy where the currency is *mutual attention*.

^{*}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 2009 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.

[†]Unaffiliated, email: a.gaudeul@gmail.com, website: <http://agaudeul.free.fr>.

[‡]Unaffiliated, email: laurencefmathieu@yahoo.co.uk.

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

Such a norm of reciprocity is defined as follows: “in a network, an agent that offers little content compared to others must 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 can devote less attention to others and still maintain her place in the network”. Attention is a cognitive process that is difficult to measure. It is linked to time spent selectively concentrating on one aspect of the environment while ignoring other aspects; it is also linked to the cognitive involvement in that activity. However, we are able to use a number of measures of attention from data gathered on the activity of 2767 bloggers drawn randomly from LiveJournal to check the model’s predictions. 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), the following figures 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), with a previous report (Perseus Development Corporation, 2003) estimating 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. 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.³ For example, the BBC controversially relied on (micro)bloggers in its coverage of the Mumbai terrorist attacks.⁴ As a sign of growing credibility, Austria Presse Agentur (APA) recently included blogs into its Medienbeobachtung 3.0 facility to monitor news in real time.⁵ The influence of blogs extends beyond news. For example, travel blogs are an increasingly important mechanism for exchanging information among tourists and are thus important for countries with a large tourism sector (Wenger, 2008).

1.1 LiveJournal

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. LiveJournal, at <http://www.livejournal.com>, offers its users a fast and easy way to create and maintain their own blog and to interact with and follow updates from other bloggers. LiveJournal is essentially an aggregation tool with lock-in effect, where bloggers write and read public and private entries, participate in communities and exchange comments, and thus develop relations with other LJ users that are not replicable on another blogging tool.

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 2008.⁶

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.⁷ Of the 72% of users who chose to reveal their gender to LJ on registration⁸, two-thirds were female. The average age of bloggers on LJ was 25, the median was 22 and the mode was 20.⁹

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 the “user information”

page, 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.¹⁰

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 and are formed and maintained by reading and posting comments on each other's blogs. Friending is a meaningful and potentially costly act, as friends are able to read "*friends-only*" entries, *i.e.* those entries that are not accessible unless one is logged in LJ and one is listed as a friend.¹¹ "Friendship" therefore acquires a meaning that is not present to the same extent in other blog networks or with RSS readers; it means confidence and readiness for closer intimacy. It is also a public act since other users can observe who is friend with whom *via* the *friend list* on the blogger's *user info* (public profile). This is an act that commits the blogger to at least browse through their friends' entries every time they read their own friends' page. Finally, friendships must be maintained over time since reciprocation of friendship is always at risk of being withdrawn (*unfriending*). This is why the term "friend" on LJ does carry its usual meaning of liking and being involved with someone.

All the above explains why many LJ users 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 Underlining again the importance of the act of friending, many users do not welcome unsolicited friendships, that is, users who list another as friend when that user does not wish to reciprocate.¹³

In the same vein, 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 friend-

ship, 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.¹⁴ 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 Lascica (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) shows 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 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. 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 networks are more likely to be cliques (*i.e.* with 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. The issue of whether links are reciprocated or not, and the related issue of the strength of relationships in a network, 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 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, *i.e.* an agent cannot 'pay' attention she received with attention she devoted to another agent. Our study allows us to develop insights into the relation between motivations and interactions of bloggers: we show that bloggers do not care primarily about expressing themselves, as then their posting activity would not depend on their audience. Instead, we show that they care about interactions, as the size of their network depends on measures of how many interactions they have (comments received and made). We also develop insights into blogging norms: a widespread expectation of reciprocity in individual relations between bloggers is reflected at the individual blogger's level through a relation between the attention she devotes to other bloggers and the attention she receives.

We show that, to some extent, an agent can exchange attention received with content she produces, and there is thus a trade-off between attention and content production; bloggers are ready to sacrifice attention received 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 shows that reciprocity matters in an empirical online setting, and 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.

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 – in both a figurative and quite literal sense, attention being the exchange currency among bloggers. More content is thus reciprocated with more attention, and vice-versa. The model thus exposes 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 comment for a comment. We show the model is well designed for the case of blogging. Indeed, a typical blogger’s reading list includes a variety of more and of less popular blogs,¹⁵ and of blogs that vary in their level of activity. This variety would not occur under less general forms of reciprocity than the one we expose here.

We define a value function for each agent belonging to the network as a function of the number of other bloggers she is linked to, of whether those links are reciprocated, 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. $e = (e_1, e_2, \dots, e_N)$ denotes the vector of content produced by agents in the set $\mathcal{N} = \{1, 2, \dots, N\}$ and $n = (n_{ij})_{i \neq j}$ denotes the vector of attentions (For example, agent i devotes attention n_{ij} to the content produced by $j \neq i$). I (respectively J, K) denotes the number of friends of representative agent i (respectively j, k). 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 form¹⁶ for the total utility of a representative agent i is

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

with $\lambda_i \geq 0$, $C(\cdot)$ increasing, convex, and subject to $\sum_{j \neq i} n_{ij} \leq T_i$. λ_i measures the propensity to enjoy being read compared to the propensity to enjoy reading others (normalised to 1 here). T_i is the attention budget of agent i , *i.e.* the total attention that she can devote to her friends. $C(\cdot)$ is the cost of content production.

Consider entrant i who has the choice between establishing a link with agent j or an agent k . 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 , $\lambda_i n_{ki} e_i + n_{ik} e_k$.¹⁷ 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

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

which can be rewritten as

$$\lambda_i (n_{ki} - n_{ji}) e_i = n_{ij} e_j - n_{ik} e_k \quad (3)$$

Note that while 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 simplify this expression, we make two assumptions:

Assumption 1. *If agent i lists both j and k as friend, then $n_{ij} = n_{ik}$.*

This equality holds if for example i cannot vary attention individually and must thus devote equal attention to all his friends. On Live-Journal, agents can put some of their friends on special filters so as

to read those friends' entries separately from others, and also to post entries available only to those friends. However, the use of such tools requires a certain degree of sophistication and is usually not publicised by the blogger for fear of alienating others. It is therefore reasonable to assume that attention is shared equally among friends on LiveJournal.

Assumption 2. *All agents devote the same amount of attention to others, that is, $T_i = T$ for any i in the set \mathcal{N} .*

Combined with the above assumption, this means that since agent i has I friends, then she devotes attention T/I to each of them. While in practise, time available for blogging may vary among bloggers, this does not necessarily mean a blogger with less time devotes noticeably less attention to others. We tried empirically to proxy time available for blogging with age or location, but those variable turned out not to be significant.

The interpretation of the formula of equation (3) can then be divided in two sections, (A) and (B), one considering the situation where both j and k reciprocate i 's friendship, the other the situation where one of the friendships is not reciprocated.

(A) Suppose both j and k reciprocate i 's friendship and suppose that $e_j > e_k$ (agent j offers more or better content or interactions). Then, from formula (3), and assumption 1, 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 needs to 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 .

In other terms, using assumption 2 as well as 1, agent k who produces less content than agent j has a lower number of friends than i (i.e. she has $K = T/n_{ki}$ friends, while j has $J = T/n_{ji}$ friends).

(B) Suppose now j reciprocates the friendship but k does not, so $n_{ki} < n_{ji}$. Under assumption 1, and from formula (3), this means that we must have $e_k > e_j$: an agent that does not reciprocate a friendship must be producing more content than another agent that reciprocates.

This leads us to defining the **norm of reciprocity** as follows: *an agent that offers little content compared to others in her network ($e_k < e_j$) must compensate this by devoting more attention to those others ($n_{ki} > n_{ji}$) in order to maintain her place in her network. Conversely, an agent that offers a lot of content compared to others in a network is*

able to devote less attention to those others and still maintain her place in the network.

Such a norm of reciprocity occurs naturally as a result of the competitive equilibrium in a market where reciprocal attention is being exchanged. Agents do not build or sustain links that are not reciprocal in the sense expressed above. Note however that this norm of reciprocity may also emerge not through a competitive process, but from an innate sense of ‘justice’, or because agents consider it is a ‘desirable’ norm of behaviour that is to be encouraged in the setting in which the agents operate.

Individual vs. aggregate data: Our data does not allow us to determine whether an agent reciprocate a friendship at the individual level, as we only observe at the aggregate level how many friends an agent has, and whether there is a balance between friends and readers.

We show that under general conditions, we can hypothesise from (A) that an agent with many friends compared to the average produces more than agents with less friends. This insight relies on the assumption that the friends of a blogger who has more friends than the average have less friends on average than that blogger. To prove this, let us find out the expression relating effort with the total level of attention received from others in one’s network. Maximising $U_i(n, e)$ with respect to e_i , we find that at the optimum, assuming concavity of the maximisation problem, one obtains $C'(e_i) = \lambda_i \sum_{j \neq i} n_{ji}$. Suppose n_{ji} is drawn at random, with the distribution of n_{ji} ’s *independent* of i ’s number of friends. Then the higher one’s number of friends, the higher the attention one receives, and the higher one’s effort level. This independence assumption would not be verified if agents linked exclusively with agents that have the same number of friends as they have, or with agents with the same quality or quantity of content as they have. Consider for example the case where all blog networks are perfect cliques, *i.e.* all agents within the network are linked with each other, and none have links outside the clique. Then content produced would be invariant with the number of friends an agent has. Indeed, in a perfect clique with N agents, $n_{ji} = T/N$ for any j in the clique, since all agents in the clique have N friends. We would then have $C'(e_i) = \lambda_i T$, and therefore effort in a clique is unrelated to the number of members of that clique. A relation between content produced and number of friends therefore occurs only in networks that are not perfectly connected. This is the case in blogging networks with ‘small world’ properties that combine

heavily interlinked individuals and links with individuals in other networks (Bachnik et al., 2005). We thus contrast two hypotheses, *H1* and *H1'* as follows:

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

Hypothesis H1': *There is no relation between a blogger's number of friends and her content production.*

Support for *H1* indicates that bloggers do not form perfect cliques and/or do not adhere to narrow forms of reciprocity based on one factor alone (number of friends or comments, content production) and/or differ in their motivations. It rather indicates that bloggers form links based on a wider view of reciprocity where content can be exchanged for attention.

In the same manner as above, we can hypothesise from (*B*) that an agent who is observed not to reciprocate friendships at the aggregate level (high ratio of readers to friends) produces more content than another agent that reciprocates. As before, what we observe (aggregate level of reciprocity) is only an imperfect signal of individual levels of reciprocity. One might very well indeed observe a blogger who appears to reciprocate on aggregate, even though none of her friendships are reciprocated while she reciprocates none of her friendships. However, failing to reciprocate at the individual level is reflected in an increase in 'readers' vs. 'friends' at the aggregate level. Consider indeed agent *i* with *N* friends and *M* readers, and suppose this agent is 'friended' by one additional agent (reader) whose friendship he does not reciprocate. Then, everything else being equal, agent *i*'s network aggregate level of reciprocity went down. This justifies our spelling out hypothesis *H2* below:

Hypothesis H2: (Aggregate reciprocity): *Bloggers with more readers than friends produce more content than others.*

In what follows, we analyse patterns of relationship and content production, exemplified by the theoretical framework detailed in this section, using real data. Our objective is to identify relations between network size, structure and content production, and check whether hypothesis *H1* and *H2* are verified. We determine if there is a positive correlation between how many readers one has and how much content one produces, and between the level of reciprocity within an agent's network¹⁸ 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, which were selected from the LiveJournal blog host using a script that chooses bloggers at random.¹⁹ The data was collected using `Screen-Scraper`, which is software that extract content from websites and add it to a database.²⁰

The descriptive statistics of the sample are given below:

Variables	Mean	Median	St. dev	Min	Max	obs
Friends	81	26	165	1	1944	2767
Readers	104	25	322	1	7855	2515
Number of entries	1049	361	3134	0	62369	2767
Comments received	3439	361	10193	1	239564	2764
Comments posted	3261	517	7148	0	106505	2767
Member	27	14	39	1	506	1481
Entries per day	33.50	0.56	211.74	0	3123	2767
Comments rec. per post	3.09	1.45	5.29	0	76	2763
Comments made per friend	37.38	13.66	91.50	0	3109	2718
Duration	1019	1003	865	1	3185	2767

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.²¹ Our blogger lists 26 *friends* (i.e. she reads the blogs of 26 LJ users) and, in turn, she is read by 25 bloggers (*readers*), which highlights a considerable level of aggregate reciprocity. The number of *comments made/received* is also remarkably balanced. The blogger follows, and is *member* of, 14 *communities*. She created 361 *entries* (posts) since the blog’s inception. The blog’s lifetime, which is measured by its *duration*, the length of time between the creation of the blog and its last update, is about 1000 days (i.e. approximately 3 years). A new entry is typically added every 2 days (*entries per day*). Individual posts receives 1.45 comments on average (*comments received per post*), whereas the blogger makes 13.66 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.

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*.

Appendix A describes the variables in the dataset in greater detail and compares its characteristics with those of all accounts created on LJ since its beginning.

5 Empirical analysis

This section studies the relationship among bloggers’ activity, number of readers and friends, and a reciprocity measure using regression techniques. This is done to verify 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 less/more friends than readers produce more/less 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 the following measures of bloggers’ effort, divided in three parts: commitment to one’s blog, posting activity and intensity of interactions with others:

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

We consider that a blog that has been active for longer does evidence a higher level of commitment. Posting activity is normalized by length of activity (posts per day). This allows us to make the difference between a blog that is active for a long time *and* presents a low level of posting activity, and a blog that has been updated for a shorter amount of time but has been posting more actively.²²

Comments posted and received are normalized per friend and per post. The number of communities one belongs to has an ambiguous effect, as a blogger involved in many communities may have less (communities draw attention away from personal blogs) or more readers (communities are a way to set up relations based on common interests).

Before estimating the activity equations, we analyse the relationship between number of friends and readers by fitting a simple regression model as follows:

$$\ln(\text{readers}) = \alpha + \beta \ln(\text{friends}) + u; \quad (4)$$

Here, the response variable is the logarithm of the number of *readers*, and the covariate is the logarithm of the number of *friends*. Table 2 presents regression estimates from the model:

	<i>ln(readers)</i>	se	<i>t</i> (p-value)
<i>ln(friends)</i>	0.985***	0.005	208.98 (0.000)
<i>constant</i>	0.103	0.017	6.07 (0.000)
obs	2496		
R^2	0.942		
F (1, 2494)	43673	(0.000)	
BP	9.39	(0.002)	
RESET	3.41	(0.017)	

Table 2: Simple Regression

Legend: obs is number of observations; *t* is the t-ratio of the coefficients; F is the F statistics for the significance of the regression; BP is the Breusch-Pagan test for heteroscedasticity; RESET is Ramsey's test for the omission of relevant variables. Test p-values are in parentheses.

*** : < 1%; ** : < 5%; * : < 10%.

Those results suggest a strong relation between the two variables. The value of the coefficient on *friends* is significant and close to 1, indicating that a (percentage) unit increase in the number of friends induce a nearly 1% increase in the number of readers. The Breusch-Pagan (BP) test evidence a certain degree of heteroskedasticity in the error term (This is hardly surprising, as models with individual data often encounter errors that have heteroskedasticity of unknown form). We thus report robust standard errors.²³

Figure 1 presents the scatterplot of the data and the fitted regression line:

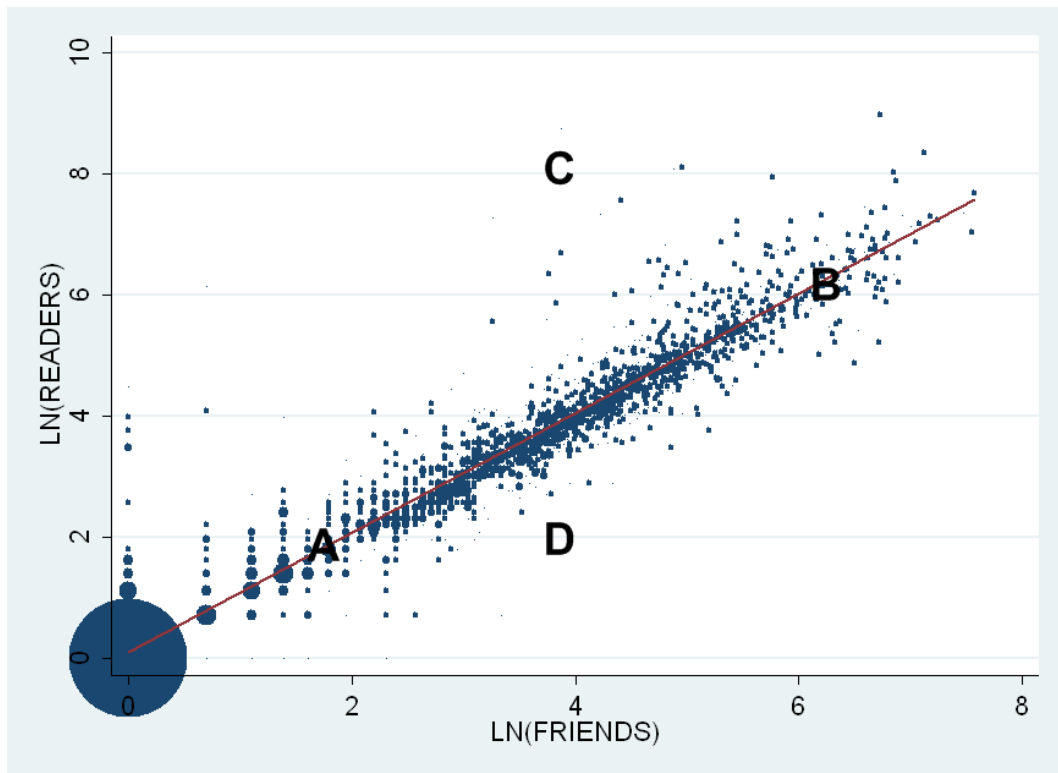


Figure 1: Scatterplot of friends vs. readers in a sample of 2767 users of LiveJournal

Note: Expressed in natural logarithm. Circle size indicates number of observations at that point. For reference, there are 672 observations at data point (0,0).

We observe that close to zero, observations are more dispersed and there are more values of *readers* associated to a specific value of *friend* than conversely. Results from the simple regression above also offer a preliminary idea of the structure – and degree of reciprocity – in the network. From hypothesis 1, agent (b), who has more friends than agent (a), should also be more active, while from hypothesis 2, agent (c), who has the same number of friends but is read by more people than agent (d), should also be more active. We see in the following whether network size and imbalances are indeed related in the way we hypothesize.

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 number of readers (H1).

The model is as follows:

$$\begin{aligned} \ln(\textit{readers}) = & \alpha + \beta_1 \ln(\textit{member}) + \beta_2 \ln(\textit{entries}) + \beta_3 \ln(\textit{comments received}) + \\ & + \beta_4 \ln(\textit{comments posted}) + \beta_5 \ln(\textit{duration}) + \epsilon; \end{aligned} \quad (5)$$

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 3 presents results for the regression of equation 5.

<i>Dependent variable:</i>	<i>ln(readers)</i>	<i>ln(readers)</i>	<i>reciprocity</i>
<i>ln(member)</i>	0.152*** (15.66)	0.170*** (18.17)	-0.030*** (-4.29)
<i>ln(entries)</i>	0.579*** (41.88)	0.576*** (44.39)	-0.007 (0.78)
<i>ln(comments received)</i>	0.776*** (55.49)	0.742*** (57.46)	0.070*** (7.06)
<i>ln(comments posted)</i>	-0.463*** (-34.07)	-0.525*** (-37.04)	0.138*** (13.82)
<i>ln(duration)</i>	0.601*** (36.53)	0.603*** (36.11)	-0.025** (-2.12)
<i>reciprocity</i>		0.474*** (15.26)	
<i>constant</i>	0.781*** (8.03)	0.950*** (9.34)	-0.240*** (-3.36)
obs	1334	1334	1357
R^2_{adj}	0.880	0.902	0.275
F stat	1550 (0.000)	1646 (0.000)	83 (0.000)
BP	29.27 (0.000)	65.68 (0.000)	6.28 (0.280)
RESET	4.84 (0.000)	5.40 (0.001)	0.92 (0.431)

Table 3: Multiple regressions with measure of reciprocity.

Legend: robust t-ratios are in parentheses; p-values for F, BP and RESET statistics in parentheses.

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. Bloggers with more readers write more entries per day (more active), write less comments per friends (reduction in attention given) but receive more comments per post (increase in attention received). The signs of the coefficients are thus consistent with intuition and the pre-

diction of $H1$.

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-ratio using White's robust covariance matrix estimator.

The reciprocity norm:

Before estimating the second activity equation, we consider whether adherence to a reciprocity norm affects a blogger's number of friends. Bloggers may be deterred from friending those bloggers who have more readers than friends, because the chances of reciprocation are low. Reciprocity is measured by taking the logarithm of the ratio of readers to friends, as follows:

$$\textit{reciprocity} = \ln(\textit{readers}/\textit{friends}) \quad (6)$$

Reciprocity is zero when the number of friends equals the number of readers. Increases in *reciprocity* indicate that the number of readers increases relatively to the number of friends. Estimates of the activity equation (5) with this added variable are reported in the second column of table 3: The size and significance of other variables' effects is not substantially modified by the inclusion of the new variable, which suggests that the information contained in *reciprocity* is incremental to that provided by other regressors. Controlling for the effect of reciprocity 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.

Interesting to note is that $\ln(\textit{readers}) = \hat{\alpha} + \hat{\beta}X + \hat{\gamma}(\textit{reciprocity})$ implies

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

This means while reciprocity has a positive and statistically significant effect on the number of readers, the effect of reciprocity on the number

of friends is negative ($0.474 - 1 = -0.526$). We do not extend the analysis of those effects in this part, since we show in the next section that *reciprocity* is a function of activity as well, and thus needs to be instrumented.

5.2 Testing the second hypothesis

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

$$\begin{aligned} reciprocity = \alpha + \beta_1 \ln(member) + \beta_2 \ln(entries) + \beta_3 \ln(comments\ received) + \\ + \beta_4 \ln(comments\ posted) + \beta_5 \ln(duration) + \epsilon; \end{aligned} \quad (8)$$

Explanatory variables are as in the model of equation 5. Estimation results are given in the last column of table 3. Once again, results highlight the explanatory power of the activity measures for the network's properties. All coefficients are significant, with the exception of the number of entries. Compared to the regression of *readers* over activity, it is notable that comments posted per friends and reciprocity are positively related. It may be that bloggers who entertain closer relations with their friends (more comments posted) are more attractive as potential friends, thus attracting more readers, but are also less likely to reciprocate readership as this would involve making a considerably higher number of comments. The blog's lifetime and community membership have negative coefficients, but those coefficients are much smaller in size than those reported for the first activity equation.

The regression explains about 27% of the variation in the dependent variable, which is not at all 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. The variable *comments posted* displays the most significant and largest effect on reciprocity. Noticeably, 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.

In this analysis, reciprocity entered the activity equations both as an explanatory and a dependent variable. Indeed, we argued that adherence to the reciprocity norm may affect 'friending' patterns. This requires the investigation of the possibly endogenous effect of reciprocity in the equation for the number of readers:

5.3 Instrumental Variable Estimation

This section applies a version of the instrumental variable technique to the estimation of the equation for the number of readers. 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 5, in appendix C, along with OLS estimates for comparison.

IV coefficients are consistently higher than those produced by the corresponding OLS regression, with the exception of the variable *member*, which is no longer significant.²⁵ The most striking result refers to the effect of reciprocity: this variable enters the IV activity equation with a large elasticity of about 3.5, and its effect is negative in sign. *Hausman's* test statistics rejects its null of exogeneity of reciprocity at every significance level, which supports the adequacy of the IV procedure.

Interestingly, these 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 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 examine whether some of the effect of reciprocity is rather due to imbalances in the bloggers' networks – net-

works that are balanced being more (or less) attractive, while imbalances in any direction are stigmatized or valued.

5.4 The effect of imbalances

We explore in this part whether there might be an effect from having *unbalanced* friendships, i.e. not only from not reciprocating friendships, but also from maintaining unreciprocated friendships. Indeed, the analysis so far has not explicitly considered how imbalances relates to blogging activity's measures. This is because *reciprocity* measures the number of readers *relatively* to the number of friends. It does *not* tell us, however, whether the network becomes more or less asymmetric. As a result, the interpretation of the effect of reciprocity in the activity equations is 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:

$$asymmetry = \ln(1 + |readers - friends|) \quad (9)$$

Here, one can see that any departure from zero signals an increase in the asymmetry of the network. Note that this measure differs from taking, for example, the absolute value of *reciprocity*, as it depends not on the *proportion* of readers *vs.* friends, but on the *number* of friends who don't read you or the *number* of readers whom you don't read in return.

Using the measure of imbalance given above, we re-estimated the two activity equations of sections 5.1 and 5.2. Table 6 in appendix D presents results from this estimation. Magnitude, sign and significance of individual coefficients in column 2 compare to those reported in column 2 of table 3, except that the coefficient on *asymmetry* is lower than the coefficient on *reciprocity*. Regressions using both the measure of reciprocity and of asymmetry as regressors show that both measures are significant, with a higher coefficient on *reciprocity* (0.368) than on *asymmetry* (0.148). Care should be exercised however when interpreting those results, due to collinearity between the two variables. Further research is needed to disentangle the effect of both measures. One

should note that the measure of asymmetry enters the readers' equation with a positive sign, even when estimating using the instrumental variable method.²⁶ The effect of imbalances is thus positive on the number of *readers*.²⁷ This is not what we expected and should be the subject of further research.

Column 3 shows that the degree of asymmetry of the network is positively related to activity. Results are considerably better when using the measure of *asymmetry* than when using the measure of *reciprocity* (column 3 of table 3), both in terms of magnitude and significance of individual effects, and of overall significance of the regression. We checked that coefficients were similar whether considering positive or negative values of (*readers* – *friends*). Why this relation between activity and asymmetry would hold is a matter for speculation. It may be for example 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.

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. We also found that failing to reciprocate was sanctioned with a lower popularity than other measures of activity might normally warrant. This can be interpreted in terms of bloggers sanctioning deviations from the norm of reciprocity, when a blogger does not return friendship as expected.

The ‘endogenous’ character of network relationships, however, makes their empirical analysis difficult. For example, it is hard to determine which comes first: having many readers or producing a lot of content. More research is needed on what determines the reciprocation of relationships in the network. Future work will rely on the collection of individual 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<http://www.brodeurmediasurvey.com>, accessed February 25, 2009.

⁴<http://tinyurl.com/62ja4e>, accessed February 25, 2009.

⁵<http://tinyurl.com/acbwcn>, accessed February 21, 2009.

⁶More information on LJ and its history can be found in its Wikipedia entry (<http://en.wikipedia.org/wiki/livejournal>, accessed February 21, 2009).

⁷Source: <http://www.livejournal.com/stats.bml>, accessed February 9, 2009.

⁸Data 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.

⁹Bloggers do not have to display their age publicly, but their birth date must be provided to LJ for legal reasons.

¹⁰Users can provide additional information in the "bio", a space where bloggers present themselves.

¹¹Not all users choose to make such 'filtered' entries, but a large portion restrict access to at least some of their posts.

¹²Other 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'.

¹³A 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...*

¹⁴For more on the social dynamics of LiveJournal, see Raynes-Goldie (2004) and Marwick (2009).

¹⁵We found that pattern to be generally observed on LiveJournal (not reported here for lack of space).

¹⁶More general utility representations could be adopted and would generate the same set of insights.

¹⁷We assume 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 he or she establishes the link. This expression of net surplus thus takes 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.

¹⁸proxied by $\ln \frac{\text{Friend of}}{\text{Friends}}$ or by $\text{sgn}(\text{Friend of -Friends}) \ln(1 + |\text{Friend of -Friends}|)$.

¹⁹<http://www.livejournal.com/random.bml>

²⁰<http://www.screen-scraper.com>

²¹As it is often pointed out, 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.

²²When 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.

²³Robust standard errors are computed using the option `robust` in Stata, which implements the estimator proposed by White (1980).

²⁴The 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.

²⁵The 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.

²⁶Results for IV regressions are not reported for reasons of space. They are available from the authors on request.

²⁷Deducing the effect of imbalances on the number of *friends* cannot be computed in the same way as in equation (7). Results from the regression replacing *readers* with *friends* show the same positive effect, whether using the instrumental variable method or not.

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 U. ACM:New York, NY (Ed.), *Proceedings of the 12th ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 44–54.
- 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>.
- Brueckner, J. (2006). Friendship 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.
- 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.
- 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.
- Mishne, G. and N. Glance (2006). Leave a reply: An analysis of weblog comments. In *3rd Annual Workshop on the Weblogging Ecosystem*.
- 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.
- 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.
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48, 817–38.

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 4 compares features of the randomly selected bloggers to those of LJ, for an informal check of the representativeness of the sample:²⁸

	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 4: 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).

C INSTRUMENTAL VARIABLE ESTIMATION OF NUMBER OF READERS³²

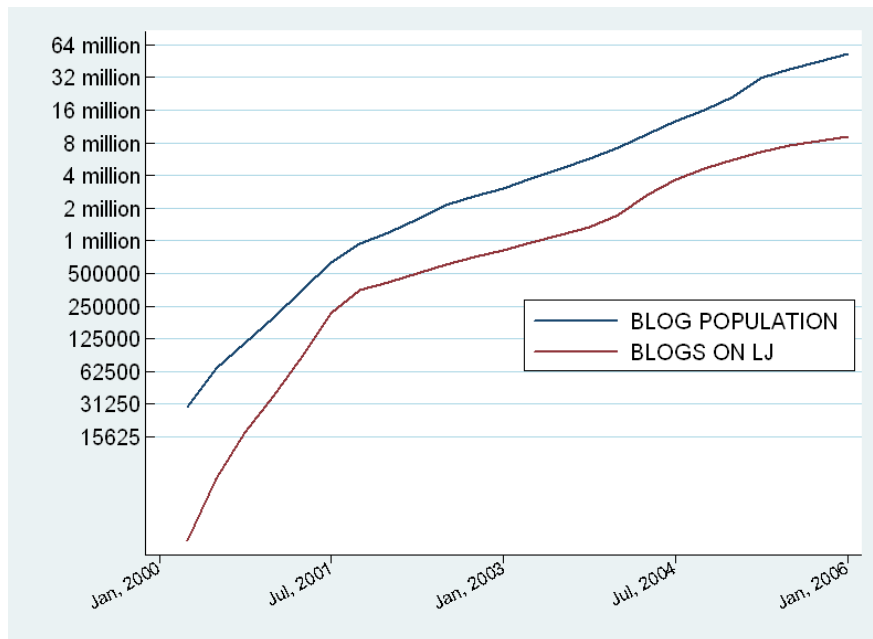


Figure 2: Number of hosted weblogs created between 2000 and 2005.
Source: Henning (2005) and LJ statistics (<http://www.livejournal.com/stats/stats.txt>), both accessed February 20, 2009.

C Instrumental variable estimation of number of readers

Table 5 shows the result of the instrumental variable estimation of the number of readers.

ln(readers)	IV	OLS
ln(member)	0.024 (-10.45)	0.200*** (13.50)
ln(entries)	0.610*** (13.10)	0.375*** (22.90)
ln(comments received)	1.034*** (17.69)	0.606*** (34.70)
ln(comments posted)	(instrument)	
ln(duration)	0.595*** (9.15)	0.272*** (11.87)
reciprocity	-3.545*** (-10.45)	0.070*** (1.64)
constant	-0.473*** (-1.05)	1.507*** (9.28)
obs	1334	1334
Hausman test	114.88 (0.000)	

Table 5: IV regression of first activity equation: model estimates and comparison with OLS regression.

Legend: Robust standard errors in parentheses.

D The effect of network imbalances

Table 6 shows the results of the estimation of the activity equations using the asymmetry measure.

Dependent variable:	ln(readers)	ln(readers)	asymmetry
ln(member)	0.152*** (15.66)	0.119*** (13.64)	0.238*** (9.59)
ln(entries)	0.579*** (41.88)	0.493*** (36.32)	0.450*** (13.82)
ln(comments received)	0.776*** (55.49)	0.671*** (49.06)	0.537*** (16.18)
ln(comments posted)	-0.463*** (-34.07)	-0.372*** (-27.54)	-0.507*** (-14.53)
ln(duration)	0.601*** (36.53)	0.518*** (33.39)	0.512*** (12.58)
asymmetry		0.176*** (18.81)	
constant	0.781*** (8.03)	0.734*** (8.23)	-0.161 (-0.67)
obs	1334	1334	1357
R_{adj}^2	0.880	0.908	0.411
F stat	1550 (0.000)	2127 (0.000)	177 (0.000)
BP	29.27 (0.000)	22.20 (0.001)	54.86 (0.000)
RESET	4.84 (0.000)	4.76 (0.001)	5.00 (0.000)

Table 6: Multiple regressions with measure of asymmetry.

Legend: robust t-ratios are in parentheses; p-values for F, BP and RESET statistics in parentheses.