Blogs and the Economics of Reciprocal Attention*

Alexia Gaudeul,† Laurence Mathieu‡ and Chiara Peroni§

October 28, 2008

Abstract

Blogs differ from other media in that authors are usually not remunerated and inscribe themselves in communities of similarly minded individuals. Bloggers value reciprocal attention, interaction with other bloggers and information from reading other blogs; they value being read but also writing itself, irrespective of an audience. A novel dataset from a major blogging community, LiveJournal, is used to verify predictions from a model of social networking. Content production and blogging activity are found to be related to the size and degree of asymmetry of the relational networks in which bloggers are inscribed.

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

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

This paper models the behavior of bloggers, a specific type of producers of media content on the Internet. Such content consists mainly of entries, or short articles, published in blogs. Blogs are websites which are either individually or collectively maintained and deal in a variety of topics, ranging from the most mundane to the political, scientific or cultural. Entries are usually rather short, frequent and arranged chronologically. Blogs feature many links with a number of other websites, and they are widely read by and engaged with by the Internet community. Bloggers spend considerable time interacting with their readers, mainly other bloggers. They also read the entries of other bloggers, with whom they establish relations maintained over long periods of time.

Blogs differ from other media as their authors are usually not remunerated and inscribe themselves in communities of similarly minded bloggers. There is little hierarchy and no clear distinction between writers and readers, *i.e.* between producers and consumers of information. Blogging

exhibits the same type of peer group relationships as those found in non virtual worlds, through links between individual weblogs which result in the creation of lively networks.

Bloggers respond to a mix of different motivations; they value interaction with other bloggers and information from reading other blogs. They value being read but they also value writing itself, irrespective of an audience.

The aim of this paper is to determine how a mix of agents with different motivations exchange reciprocal attention, and to measure the relative importance of different motivations in taking part in the newly emergent networks of 'blogs'. In doing so, the paper exploits a unique and novel dataset from the LiveJournal blogging community.

This paper thus devotes its attention to the moral codes, motivations and interactions of online participants into blogging networks. For example, if bloggers care primarily about expressing themselves, then their activity will not depend on the number in their audience. If instead bloggers care primarily about building relations with their readers and those they read, then there will be a fine balance between how much attention a blogger gets and the attention that she devotes to other bloggers.

This paper produces important insights into the motivational dynamics of this new public sphere (the 'blogosphere') and furthers the understanding of the motivations for participation in Internet-mediated non-monetary transactions. This is important because blogs and other community-led endeavors (Wikis, open source software) challenge the usual top-down approach of old industries, and make it easier for dispersed contributions to have a global impact without central coordination. The conclusions from the paper are relevant to those companies that manage blogging and social networking communities, such as Google's Blogger, Six Apart's Typepad, SUP's LiveJournal, Wordpress, Facebook, or News Corp's MySpace). This work is of interest to the media industry in general as it contributes to the study of a new source of expression and news. This work is also relevant to those companies that are involved in the management of communities of self-motivated individuals, such as eBay or Amazon with their networks of independent sellers and buyers. Finally, this study matters to those companies that want to involve their clients and employees in a more open way though the medium of blogs.

1 Background

Nowadays, individuals make extensive use of the Internet and release large amounts of information on the Web. They often do so through the medium of weblogs. A weblog, often referred to as a *blog*, is a personal journal arranged in reverse chronological order (from the most recent post to the oldest post) and made publicly accessible on the web, and characteristically updated on a daily basis. Blogs are now accessible to a large population, thanks to the availability of user-friendly tools for their updating. Statistics on blogs, however, are notoriously fickle (Bialik, 2005). Henning (2005) estimated the number of blogs at 53 million by the end of 2005 (see figure 4 in appendix B), and Technorati, an Internet search engine for blogs, claimed to track about 113

million blogs in May 2008. Perseus Development Corporation (2004) estimates that about one third of them are active. Age and gender characteristics of bloggers vary according to reports (Rainie, 2005), but most agree that weblog authors are dominated by youth while all genders are represented (Huffaker, 2004; Huffaker and Calvert, 2005).

Weblogs have many possible formats and can be classified according to their content and purpose. In general, we can divide weblogs into two distinct groups: personal and 'topic-oriented' weblogs (Herring et al., 2004). The former type relates to people's expression of opinion, ideas and emotions whereas the purpose of the later is to convey expert information on a specific topic, generally in the area of professional expertise of the blogger herself. Blogging is thus a form of leisure but also of economic, political and social action.

A number of views have been expressed about the role, value and future of blogs and bloggers. Some consider them as a newly emergent media form (Ribstein, 2005, 2006), leading Lemann (2006) to question its value to journalism. Others (Drezner and Farrell, 2008) consider blogs as a tool of political influence, leading Sunstein (2008) to ask what kind of public space is being created within this new media and to worry about how blogs may polarize opinions. Quiggin (2006) considers blogs as part of the 'creative commons' with Wikis and open source software, while Huck et al. (2008) are interested in how blogs help consumer choice and affect firms' reputations. Finally, others consider blogs simply as a tool for social networking, the online equivalent of networks of friends, relatives and acquaintances (Paolillo et al., 2005).

2   Related Literature

The focus of this paper is to analyse the occurrence, strength and significance of reciprocal (non-directed) or non-reciprocal (directed) links between agents in a network. In particular, this study aims to investigate the trade-off between attention given and content produced by bloggers. As noted by Drezner and Farrell (2008), blogs are 'a major topic for research' and offer 'extraordinarily fertile terrain for the social sciences'. Here, we adopt the view expressed in Bruni and Sugden (2008) that economics is not only concerned with market interactions, but also with the wider class of reciprocal relationships in civil society, which justifies applying an economic analysis to the study of blogging.

This paper is one of comparatively few that use statistical analysis on weblogs. Ribstein (2005) and Lassica (2001) have good introductions to the use, function and role of blogs and can serve as starting points to this literature. Furukawa et al. (2006) focuses on the analysis of relationships between weblogs and the effects of those relationships on the reading behavior of weblog authors in a weblog network. Bachnik et al. (2005) looks at weblog networks and the existence of social structures in those networks. Bar-Ilan (2005) monitored a set of weblogs and analyzed their content in order to determine the reason why people wrote and read weblogs.

Paolillo et al. (2005) used data from LiveJournal, the blog host, to correlate the change in Live-Journal users' interests and the change in their friends' networks. Kumar et al. (2004) is another researcher trying to relate cross-linking patterns by exploring the list of interests of bloggers on LiveJournal.

While there is some amount of research into the structural patterns of blogging relationships online (Backstrom et al., 2006; Bachnik et al., 2005; Finin et al., 2005), there is little quantitative or theoretical research linking social structure, activity and motivation of bloggers (Kumar et al., 2004; Furukawa et al., 2006; Lento et al., 2006; Mishne and Glance, 2006).

Qualitative studies of bloggers' motivations (Raynes-Goldie, 2004; Fono and Raynes-Goldie, 2006) reveal 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. Other papers distinguish between different types of bloggers by their position in a network and determine how this influences their activity. They differentiate between agents that are central or peripheral (Caffarelli, 2004), readers or writers, community or self-oriented, magnetic or unattractive (Brueckner, 2006).

This paper complements the existing literature on blogging with a different — network structural — economic perspective. The analysis of social networks draws from a long tradition in sociology and has recently also become a subject of study for economists (Watts, 2001; Jackson, 2003; Newman, 2003; Gui and Sugden, 2005; Brueckner, 2006). This paper uses existing models of network structure and formation of links among individuals (see Jackson (2003) and Newman (2003) for Internet applications). We are interested in the activity – content production – of each node – agent – and explore the relation between network structure and production in small networks. As seen in Jackson (2003), the literature on directed networks is still rather undeveloped. 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 is little explored (on weak vs. strong relations, see (Granovetter, 1973)). Most models assume relations are reciprocal and do not set out to determine the intensity of the relation between agents. This paper differs from those by measuring not only the number of links that each agent maintains, but also the intensity of the activity of each agent. This allows us to use devised proxies for the intensity of relations among bloggers.

The paper is thus unique in exploiting measures of a blogger's network size and composition with measures of the type and extent of his or her activities. The structure of the links connecting bloggers and the activity of each bloggers is drawn from a major blog hosting site, LiveJournal. Blogging activity is measured in various ways, for example: number of entries, number of links back to other weblogs, number of references to other blogger entries, number of comments made on other weblogs, *etc.* We show that bloggers attempt to balance attention received and attention given within their network. However, they are also ready to sacrifice attention received from a blogger if that blogger provided sufficient content. This study shows that one can obtain good results by distinguishing the act of reciprocation of friendship, and acts that are designed to attract

readership.

Beyond works on blogging and network theory, this article is also of interest to behavioral economics in that sustained participation in blog networks requires bloggers to adhere to norms of reciprocity. It is also of interest to innovation economics as Weblogs change the dynamics of knowledge production and information exchange in a way that can be likened to how open-source methods change the way software and information infrastructures are developed and maintained (von Hippel, 1998). Users become participants in the development and management of the products they use (Quiggin, 2006). The micro-economic analysis of activity in blog network presented in this paper throws light on the details of the motivational structure for such activities. Finally, this paper is also of direct interest to the blogging and media world, and especially to those bloggers who are theorising on their own activities (Blood, 2000; Fono and Raynes-Goldie, 2006; Herring et al., 2004; Huffaker and Calvert, 2005; Raynes-Goldie, 2004; Ribstein, 2005, 2006). Public interest has been focused on the few atypical blogs that have achieved an audience that goes far beyond the world of blogging, such as for example the politically influential Huffington Post or Belle de Jour's blog, which was made into books and a TV series.

We argue that the model set out in this paper is better adapted to describe the majority of blogs, as these are part of a self-sustaining economy which does not require any exchange of money or other goods. Bloggers are integrated into tightly knit communities of like-minded individuals, who devote continued attention to each other. This imposes a limit to the size of anyone's audience: a blogger with many friends will not be able to devote much attention to each of them and will at some point have to limit the size of his/her network if he/she is to sustain relationship with its readers. In particular, this view is suited to the study of the LiveJournal community of bloggers, with its unique and sophisticated federating system. Regarding competition with other media, it would be difficult in our view for any one blog to emerge as a significant competitor to any one journal or other media. On one hand, blogging that is organised in communities contributes to audience fragmentation, as each group of bloggers concentrate on interests specific to that group. On the other hand, this type of blogging encourages the production of a diversity of views, as any type of views is able to federate an audience in a community of bloggers.

3 The Model

We ground the analysis on formal models such as Domon (2001) or Brueckner (2006). In those models, a value function for each agent belonging to the network is determined as a function of the number of their links with other bloggers, and whether those are reciprocated or not. The utility of one agent depends not only on whom they are linked to, but also on the links of those other agents with other agents.

Consider a simple model where N agents produce their own content and read content generated by others in the network. $e = (e_1, e_2, ..., e_N)$ denotes the vector of effort into producing

5

content by agents 1, 2, ..., N and $n = (n_{ij})_{i \neq j}$ denotes the vector of attentions, that is, for example, agent i devotes attention n_{ij} to the content produced by $j \neq i$.

A simple additive form for the total utility of a representative agent, i in this instance, is

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

with $\lambda_i \geq 0$. λ_i measures the propensity to enjoy being read compared to the propensity to enjoy reading others (normalized to 1 here). We do not consider the cost to reading others or producing content (see appendix E for a full model where such costs are considered).

In the following, we will establish the existence of a 'norm of reciprocity', whereby an agent that devotes less attention to others than others devote to her will produce more content than others, while an agent that devotes more attention to others than others devote to her will produce less content than others.

Consider a competitive market for attention, where an agent i has the choice between establishing a link with an agent j or an agent k. Agent i will prefer establishing the link with j if the net gain in utility from doing so, $\lambda n_{ji} e_i + n_{ij} e_j$ (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 net gain in utility from establishing a link with k, $\lambda n_{ki} e_i + n_{ik} e_k$.[1] In such a competitive market, the surplus from creating links should be the same across all agents — indeed, if that was not the case, then some agents who offer lower surplus would lose friends and ultimately be left out of the market, while those that offer more would accumulate friends at the expense of others. Therefore, it must be that both surplus are equal, so that

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

which can be rewritten as

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

In general, the level of attention displayed individually to each agent in a network is unobservable. It is also generally the case that, when an agent friends another, she does not know how much attention she will receive, and if this is more or less than what other friends receive. Therefore, one must assume in most cases that, a priori, agent i will devote the same attention to both j and k if both are listed as friend by i, so $n_{ij} = n_{ik}$. This is true if, for example, i cannot apportion attention individually and must thus devotes equal attention to all his friends.[2]

The interpretation of the formula of equation (3) can then be divided in two parts, one considering the situation where a friendship is reciprocated, the other the situation when the friendship

is not reciprocated.

1. Suppose i reciprocates friendship from both j and k so that a priori, $n_{ij} = n_{ik}$. Suppose that $e_j > e_k$ (agent j offers more or better content or interactions), then, from formula (3), I must have $n_{ki} > n_{ji}$: agent k, who produces less, must devote more attention to i than agent j, who produces more, does. In other terms, assuming again that agents cannot apportion attention individually and thus devote equal attention to all their friends, then agent k must have a lower number of friends than j.

   This allows us to define a **norm of reciprocity** as follows: *an agent that offers little content compared to others in a network will have to compensate this by devoting more attention to those others if she is to maintain her place in the network. Conversely, an agent that offers a lot of content compared to others in a network will be able to devote less attention to those others and still maintain her place in the network.*

   Such a norm of reciprocity will occur naturally as a result of the competitive equilibrium in a market where reciprocal attention is being exchanged. Agents will not build or sustain links that are not reciprocal. The 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 behavior that is to be encouraged in the setting in which the agents operate.

2. Note that reciprocity in this instance is defined not as reciprocity in attention or in content, but as reciprocity in net surplus. Under that norm, an agent may accept getting less attention from an agent with more content. Agents trade attention for content and vice versa. One could observe other forms of reciprocity, whereby agents would for example only link with agents that have the same number of friends as they have so as to ensure a balance of attention, or will only link with agents that display the same level of overall blogging activity, so as to ensure reciprocal exchange of content. Those other forms of reciprocity are potentially as valid as the one we consider here. However, those alternative norms do not have such a power of explanation as the norm of reciprocity we consider here; indeed, a typical blogger' reading list will include a variety of more and of less popular blogs, and among this reading list, some blogs will be more active than others.

3. Whether an agent reciprocate friendships individually is not observable. We only observe at the aggregate level whether an agent has many friends or not, and whether the agent has a balance between friends and friend of. However, we can hypothesise that an agent with many friends compared to the average will have more of his friends having less friends than himself or herself than the opposite (again, this is not necessarily true but rather likely). This agent with many friends should therefore according to the norm of reciprocity produce more than agents with less friends. This leads us to exposing hypothesis 1:

Hypothesis 1 *Bloggers who display higher levels of content production and general blogging activity have a higher number of readers (**H1**).*

H1 would seem quite intuitive, although we show in the full model in appendix E that if all agents are linked with each other and are identical, then content should be invariant with the number of friends (an hypothesis we denote H1', see appendix E). H1 therefore implicitly assumes asymmetries in the network structure, that is, agents differ in their type of motivation, do not link exclusively with agents that have the same number of friends as they have, and do not link exclusively with agents with the same quality or quantity of content as they have. Appendix E shows indeed that if agents differ in their motivation, then results similar to those posited in H1 will appear.

4. Consider now the case when at least some friendships are not reciprocated. For example, suppose one observes an agent who reciprocates only a few of its friendships (high ratio of friend of to friends). This is a clear case where at least some readers do not see their friendship reciprocated. Suppose i has a choice between friending j who reciprocates friendships and friending k, who does not, so $n_{ki} < n_{ji}$. Suppose as before that i cannot segregate between friends so we must have $n_{ij} = n_{ik}$. Then, from formula (3), this means that we must have $e_k > e_j$: an agent that does not reciprocate friendship must be producing more content than another agent that reciprocates. This leads us to spelling hypothesis 2 as follows:

Hypothesis 2 *Bloggers with less friends than readers (who are read by more bloggers than they read blogs) produce more content than others. Bloggers with more friends than readers (who read more blogs than other bloggers read them) have less content than others (**H2**).*

5. Note that as in H1, what we observe (aggregate level of reciprocity) is only an imperfect signal of individual levels of reciprocity. For example, one might very well have a blogger with a number of friends equal to his number of 'friend of' who sees none of his friends reciprocate while she reciprocates none of her friendships. This blogger will have an equal amount of friends than of friends of and will thus appear to reciprocate on aggregate. However, statistically, those with imbalances between friends and 'friend of' are more likely than others to have imbalances at the individual level, which justifies our spelling out hypothesis 2.

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.

4 Empirical analysis

The data analysed in this article are collected from LiveJournal ("LJ"), a web-based community where Internet users can keep a weblog (blog), journal or diary.

LiveJournal, at www.livejournal.com, is a service offering Internet users a fast and easy way to create blogs. Among blog hosts, LJ distinctive feature is the emphasis on the *community* aspect of blogging, in that it is based on an open-source code and is to a great extent maintained and kept alive by a community of volunteers.[3]

Created in 1999 by Brad Fitzpatrick, in January 2005 LJ was purchased by SixApart, the owners of Typepad – another popular blog host. Since the take-over, the profit making aspect of LJ has become prevalent: "sponsored" (advertising-bearing) accounts have been introduced and the discrepancy between services offered to free versus paying users has been widening ever since. In December 2007, Six Apart sold LiveJournal to SUP, a Russian company that was already managing LJ in Russia. This raised issues regarding the management of LJ's user base and communities, notably for those Russian bloggers who feared a crackdown by Russian authorities. The "privatisation" of LJ was furthered in March 2008, when the option for creating free accounts was removed. Widespread protests by users led to the option being re-instated in August 2008.[4]

In December 2005, the number of accounts (or blogs) recorded by LJ on its publishing system totaled 9,000,634, of which 1,333,839 (15%) had been updated in the previous 30 days. LJ was most popular in the United States, where 46% of the total number of users were located, followed by Canada, Russia and the UK, each with 3% of users.[5] As LJ is based on an open-source code, several LJ "clones" have been established, such as DeadJournal, a blog host aimed at the goth community.[6] These offshoots are not surveyed in our study, as they are not under the same umbrella as LJ.

This study focuses on LiveJournal because it provides detailed and easily accessible information on users' activity. Other blog hosts do not adopt the same policy of wide transparency, and the same type of information would be more difficult and costly to access. 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 creation number, date of creation, status of the account (*i.e.* early adopter, free, paid, sponsored), name and number of friends, number of posts, number of comments made and received belong to this type of information. Other data, such as date of birth, location, list of interests, and any additional information, is provided on a voluntary basis.[7] Other information, such as gender, is not shown but it is collected by LJ for internal statistical purposes.[8]

LJ offers users several ways of interacting with other bloggers, namely *friend* lists and communities. Information on the relational network of a blogger, such as name and number of bloggers who list the user as friend and the communities the user is member of, can be hidden. Alternatively, the blogger can choose to list as friends only those users who list her as friend, and

hide those for whom she does not return, or reciprocate, the friendship.

4.1 Friends and communities

Much of the analysis in this paper focuses on the lists of *friends*. This word has different meanings on LJ (Fono and Raynes-Goldie, 2006). At a technical level, a *friend* is another blogger the user subscribes to. Listing someone as a friend is referred to as *friending*. LJ users list their friends on *friends' pages*; such lists record all entries made by the friends, in reverse chronological order. Listing another blogger as 'friend' implies the user will be following the blogger's entries over some period of time. In turn, friends are able to read those entries the user chooses not to make public (*i.e.* those entries that are not accessible unless one is logged in LJ and one is listed as a friend). Not all users choose to make such 'filtered' entries, but a large portion restrict access to at least some of their posts. "Friendship" therefore acquires a meaning that is not present in other blog networks; in this case means confidence and readiness for closer intimacy. Finally, the term "friend" carries its usual meaning, *i.e.* one of liking and positive interest in other users. In this sense, many LJ friendships reflect "real world" friendships. Others are "LJ-only" friendships, formed and maintained by reading and posting comments.

Many LJ users attach great significance to the act of 'friending', or of dropping other users from the friend list. Some engage in 'popularity contests', whereby status is attached to the number of users who list you as friend. Some do not welcome unsolicited friendships. The latter occurs when a user is listed as friend by users she does not wish to be-friending. Options were made available on LJ to prevent unwanted interaction.

Friendships are generally established with the expectation of *reciprocity*. This means that a user expects a friend will read her back in return; it also means that the same user will not friend someone who is not expected to reciprocate the friendship, and will drop from the list those who do not reciprocate. The extent to which those norms hold, however, varies from user to user and from situation to situation, and they are therefore not universally followed. More on LJ and its social dynamics can be found in Raynes-Goldie (2004).

4.2 The Data

The data used in this study are observations on list of friends, readers, and posting activity of 250 bloggers, which were selected from the LiveJournal blog host using a script that chooses bloggers at random.[9]

Table 1 compares features of the randomly selected bloggers to those of LJ, for an informal check of the representativeness of the sample:[10]

	Random sample	LiveJournal
Weblogs updated in the last 30 days	69%	15%
Weblogs updated in the last 7 days	51%	9%
Weblogs updated in the last 24 hours	28%	3%
Countries	US: 48% UK: 12% Russia: 6% Canada: 5%	US: 46% UK: 3% Russia: 3% Canada: 3%
Age (mode)	20-22 years	18 years

Table 1: Comparison table

One can see that bloggers are young, between 20 and 22 years of age, and predominantly located is the US. The random sample appears to reflect age and location of the wider community, although several locations appears to be over-represented. The level of activity is higher in the sample than in LJ. There are several possible explanations for this. The random script could be in fact designed to select active blogs, in order to spare the user having to sift through inactive blogs. Perhaps as a result of this *'bias'* towards active journals, UK, Russian and Canadian blogs appear to be over-represented. Noticeably, LJ was adopted only recently in those countries, which makes it more likely that bloggers will be more active due to a "novelty" effect. Alternatively, the random script might be programmed to output bloggers closer to the user's location — the UK, in our case.

The descriptive statistics of the sample are given below:

Variables	Mean	Median	St. dev	Min	Max	obs
Friends	64.00	38	91.32	1	717	250
Friend of	70.06	34	120.23	1	746	203
Number of entries	391.82	199	526.81	0	4271	249
Comments received	1705.80	661	2923.73	0	19580	249
Comments posted	1973.37	768	2926.18	0	14447	249
Member	9.63	3	14.99	0	94	203
Community	9.06	3	12.92	0	79	251
Entries per day	0.45	0.335	0.40	0	2.51	246
Comments per post	4.31	2.85	5.17	0	44.06	246
Comments per friend	41.19	19.54	110.86	0	1615	248
Duration	767.57	734.50	474.58	1	2143	246
Reciprocity	0.029	-0.048	0.77	-2.15	6.53	202

Table 2: Summary Statistics.

There is clear evidence that data are skewed: variables' median value are always lower than the mean. So, in what follows, the "typical" user is described by median values. (It is often pointed out that 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.) Our blogger lists 38 *friends* (*i.e.* she reads the blogs of 38 LJ users) and, in turn, she is read by 34 bloggers (*friendof*), which highlights a considerable level of reciprocity. The number of *comments made/received* is also remarkably balanced.[11] The blogger follows, and is *member* of, 3 *communities*.[12] She created 199 *entries* (posts) since the blog's inception. The blog's lifetime, which is measured by *duration*, the length of time between the creation of the blog and its last update, is 734 days (*i.e.* more than 2 years). A new entry is typically added every three days (*entries per day*). Each post receives approximately 3 comments (*comments per post*), whereas the blogger makes about 20 comments on the journals of each of her friends (*comments per friend*).[13] So, these data evidence bloggers' considerable commitment, although the frequency of updating and the posting activity varies greatly among users.

The variable named *reciprocity* measures the degree of asymmetry in the blogger's network, and is computed by taking the logarithm of the ratio of readers to friends, as follows:

(4) $\text{reciprocity} = \ln(\text{friendsof}/\text{friends})$

Such measure implies that, when reciprocity is zero, the number of friends equals the number of readers (in other words, the blogger's network is symmetric). Any departure from zero indicates some asymmetry in the network of the blogger. Furthermore, movements from negative towards zero or positive values indicates that the number of readers increases relatively to the number of friends.

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

Appendix A describes the variables in the dataset in greater detail.

4.3   Results

This section studies the relationship among bloggers' activity, number of readers and friends, and reciprocity, 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 a higher number of 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 various measures of bloggers' effort (activity). The measures, or proxies, of bloggers' activity are:

1. The length of time the blog has been active.

2. The number of posts the blogger has been making per day.

3. How long ago the blog was last updated.

4. The level of *interactivity* in the blog, as measured by how many comments are received and posted in other bloggers' journals.

5. The number of communities the blogger belongs to.

We consider that a blog that has been active for longer does evidence a higher level of commitment. This first proxy, however, does not fully capture bloggers' effort. Indeed, a blog that is active for a long time *and* presents a low level of posting activity may be considered less active than a blog that has been updated for a shorter amount of time but has been posting more actively. Furthermore, inactivity may lead other bloggers to drop the blog from their list of friends. So, the second and third proxy aims to capture these latter effects.

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 because there is a high level of inertia in the friending process on LiveJournal: LJers tend to keep a blogger on their list even long after that blogger has stopped updating, as long as that blogger does not drop them. Indeed, most bloggers like to inflate their list of friends and friend of and will maintain relations long after they have ceased being active.

Hypothesis H1 suggests that the measures listed above, as proxies for activity, should be positively related to the number of readers. In this respect, however, one should note that the effect of the number of community one belongs to looks ambiguous: 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 bloggers which a user lists as friends (*friends*) and number of bloggers that list the same user as a friend (*friendsof*). This should give a preliminary idea of the structure — and degree of aggregate reciprocity — in the network. To do so, we fit a simple regression model to blogs data, as follows:

(5) $ln(friendsof) = \alpha + \beta\ ln(friends) + u;$

Here, the response variable is the logarithm of *friendsof*, or readers, and the covariate is the logarithm of the number of *friends*. Residuals and sensitivity analysis lead us to remove two outlier observations.[14] Table 3 presents regression estimates from the model:

	ln(friendsof)	se	$t-$value	$P > |t|$
ln(friends)	0.932***	0.034	27.14	0.000
constant	0.214	0.135	1.59	0.115
obs	200			
RSS (df)	62.750 (198)			
R^2	0.789			
F (1, 198)	736.41***			
BP	6.73	(0.0095)		
RESET	0.87	(0.4569)		

Table 3: Simple Regression

Legend: *obs* is number of observations; *RSS* is the residuals' sum of squares; *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 rejects its null of constant variance in the model, evidencing the presence of heteroscedasticity in the data, as it is often the case with individual observations; we thus report robust standard errors.[15]

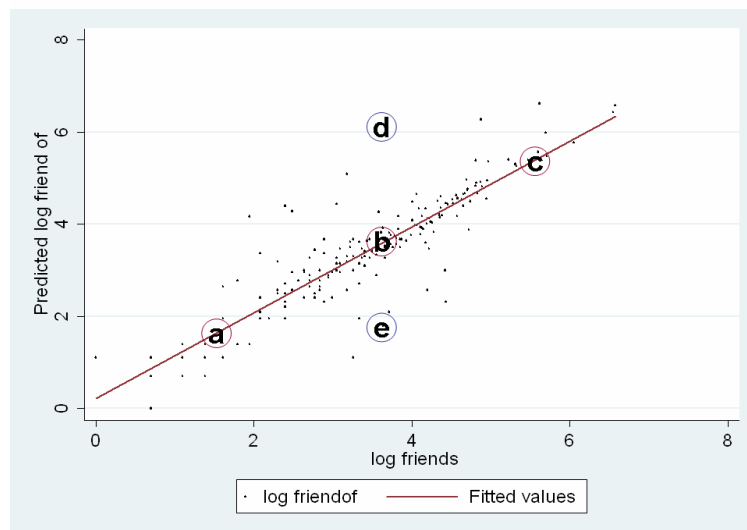Figure 1 presents the scatterplot of the data and the fitted regression line:

Figure 1: Scatterplot of friendof versus friends and regression line.

From hypothesis 1, agent (c), who has more friends than agent (a), should also be more active, while from hypothesis 2, agent (d), who has the same number of friends but is read by more people than agent (e), should also be more active. We will see in the following whether network sizes and imbalances are indeed related in the way we hypothesize.

4.3.1 Testing the first hypothesis

In what follows, we regress the numbers of readers (*friendof*) on several indicators of bloggers' activity. The model is as follows:

$$ln(friendsof) = \alpha + \beta_1 \, ln(member) + \beta_2 \, ln(entries) + \beta_3 \, ln(comments) + \beta_4 \, ln(commade) +$$

(6) $$ + \beta_5 \, ln(duration) + \beta_7 \, active + \varepsilon;$$

here, *member* denotes the number of communities one is member of, *entries* the number of entries per day, *comments* the number of (received) comments per post, *commade* the number of comments posted per friend, and *duration* the duration (in days); finally, *active* is a dummy that takes value one when the blog has been updated in the last 8 weeks (prior to the data collection date) and zero otherwise. All variables are in natural logs.

Table 4 presents results for the regression of equation 6.

Dependent variable:	ln(friendsof)			
ln(member)	0.140***	0.145***	0.186***	0.188***
	(0.048)	(0.045)	(0.036)	(0.035)
ln(entries)	0.390***	0.395***	0.414***	0.417***
	(0.072)	(0.074)	(0.068)	(0.071)
ln(comments per post)	0.766***	0.767***	0.758***	0.758***
	(0.060)	(0.060)	(0.048)	(0.048)
ln(comments per friend)	-0.271***	-0.272***	-0.440***	-0.440***
	(0.052)	(0.060)	(0.055)	(0.056)
active	0.061		0.028	
	(0.152)		(0.013)	
ln(duration)	0.537***	0.547***	0.613***	0.618***
	(0.050)	(0.046)	(0.048)	(0.048)
reciprocity			0.656***	0.657***
			(0.094)	(0.096)
constant	0.351	0.340	0.304	0.298
	(0.303)	(0.301)	(0.248)	(0.273)
obs	141	141	141	141
RSS (df)	36.04 (134)	36.10 (135)	25.87 (133)	25.88 (134)
R^2_{adj}	0.755	0.756	0.832	0.824
F stat	73.00***	88.06***	101.90***	102.45***
BP	22.41 (0.001)	16.40 (0.006)	47.95 (0.000)	39.32 (0.000)
RESET	2.76 (0.001)	2.70 (0.001)	2.05 (0.012)	2.06 (0.012)

Table 4: Multiple regression.

Legend: robust standard errors are in parentheses; p-values for BP, RESET in parentheses.

One can see that all coefficients are significant, with the exception of the dummy variable for activity. A possible explanation for this, is that a blogger with no entries in the last 8 weeks may be very active in communities or in commenting on other people's entries. As noted at the beginning of this section, the LJ inertia makes us reluctant to label any account as 'inactive'. So, we re-estimate the model without the dummy. Results for this regression are reported in the second column of table 4. The exclusion of the dummy does not worsen the overall fit of the model, and changes only marginally the value of the coefficients, without altering signs.

The chosen measures of activity have a positive and statistically significant effect on the number of readers. In particular, the largest coefficients are associated with the number of comments per post and duration: a 1% increase in these variables lead, respectively, to about a 0.8% and a 0.5% increase in the number of readers, *cæteris paribus*. Smaller, but positive and significant effects are those of the number of entries and the number of communities the blogger belongs to. Otherwise, an increase in the number of comments posted induces a reduction in the number of readers.

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 *friendsof*, 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. There is, however, some evidence of heteroscedasticity in residuals, as shown by the Breusch-Pagan (BP) test statistic, and the value of the RESET test suggests misspecification in the model.

Figure 2 reports partial effects of the model's covariates on the dependent variable (or added-variable plots). These plots link the response variable to each regressor, emphasising the contribution of each covariate to the model, in isolation from the others; so, they are useful in evidencing departures from linearity/normality and outlying values.[16] By looking at the graphs, one can notice the good fit produced by the measures related to posting activities. Other covariates have more dispersed observations, highlighting the heterogeneity in the data.
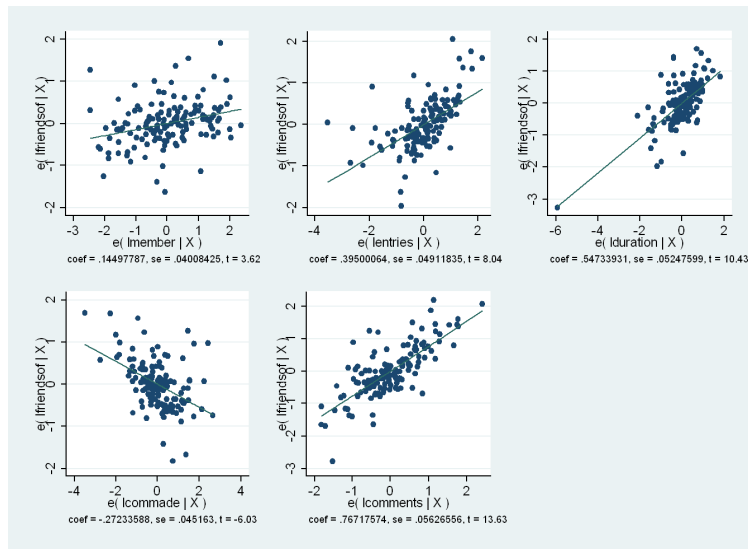
Figure 2: Multiple regression model: partial effects without reciprocity as an independent variable.

Before estimating the second activity equation, there is an additional effect that should be taken into account. Bloggers who maintain a balance between *friends* and *friends of* are not regarded by other bloggers in the same way as bloggers who do not maintain that balance. Notably, bloggers are deterred from friending those bloggers who have more readers than friends, because the chances of reciprocation are low. Similarly, bloggers with more friends than readers may give the impression they do not require reciprocation and may thus end up reinforcing the imbalance. This is why we expect most blogs to have an equal number of friends than friends of in the aggregate, since there is likely to be a stigma to not doing so. This motivates extending the model of equation (6) to include *reciprocity*, which captures network imbalances, in the vector of explanatory variable.

The third and fourth column of table 4 reports estimates of the activity equation with reciprocity. Results shows that reciprocity has a positive and statistically significant effect on the dependent variable. Controlling for the effect of reciprocity improves the overall goodness-of-fit of the model. The value of the Breusch-Pagan test statistic, however, increases considerably, and the partial effects plots in figure 3 casts several doubts on the explanatory power of the added variable:
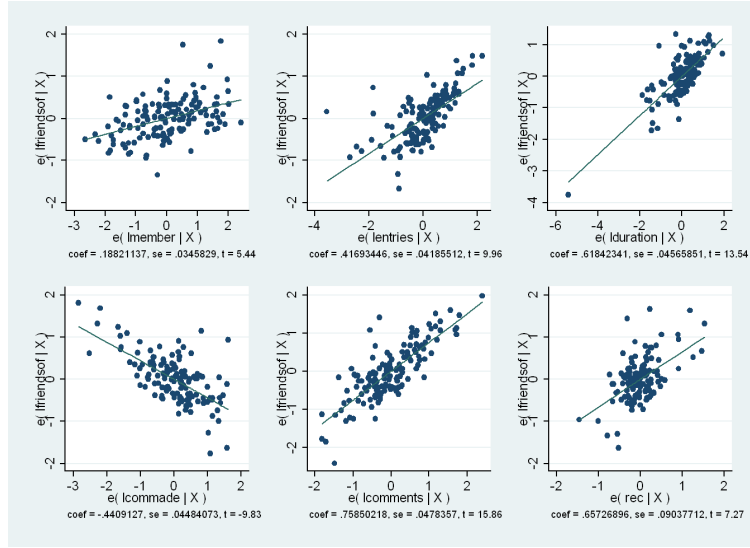
Figure 3: Multiple regression model: partial effects with reciprocity as an independent variable.

4.3.2 Testing the second hypothesis

To examine the second hypothesis (H2), which relates content production and network imbalances, we estimate a second equation, in which *reciprocity* is regressed on the various proxies of activity listed in the beginning of this section. As already noted, reciprocity captures the imbalance in the network, but also measures the number of readers relatively to the the number of friends. This is relevant to the verification of hypothesis H2.[17] The model is as follows:

$$reciprocity = \alpha + \beta_1 \; ln(member) + \beta_2 \; ln(entries) + \beta_3 \; ln(comments) + \beta_4 \; ln(commade) +$$

(7) $$+ \beta_5 \; ln(duration) + \beta_7 \; active + \varepsilon;$$

(Here, explanatory variable are as in the model of equation 6.)

The estimated equations, for the models with an without the dummy variable, are given in table 5 below:

Dependent variable:	reciprocity	
ln(member)	−0.069*	−0.065*
	(0.084)	(0.040)
ln(entries)	−0.037	−0.033
	(0.092)	(0.042)
ln(comments per post)	0.012	0.013
	(0.127)	(0.052)
ln(comments per friend)	0.258***	0.256***
	(0.099)	(0.050)
active	0.050	
	(0.252)	
ln(duration)	−0.116**	−0.108**
	(0.090)	(0.051)
constant	0.073	0.062
	(0.299)	(0.272)
obs	141	141
RSS (df)	23.612 (134)	62.84 (135)
R^2_{adj}	0.267	0.271
F stat	9.52***	11.56***
BP	43.94 (0.000)	37.20 (0.000)
RESET	2.15 (0.012)	2.10 (0.014)

Table 5: Multiple regression: reciprocity equation.
Legend: robust standard errors are in parentheses; p-values for BP, RESET in parentheses.

The number of comments posted has the largest coefficient (in absolute value), followed by duration. These coefficients, however, have opposite signs. An increase in the number of comments posted leads to an increase in the number of readers/friends ratio. Conversely, an increase in the blog's lifetime is associated to a reduction in the same ratio. Community membership, number of entries, and comments received, have small and scarcely significant coefficients.

The regression explains about 30% of the variation in the dependent variable, which is not at all disappointing; it is indeed well known that R^2s tend to be low when using individual data. Once again, the Breusch-Pagan test rejects its null of constant variance at any significance level. The F test statistic for the regression, however, decisively rejects the null of joint lack of significance of the regressors.

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 per friends* 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 posting activity of the blogger in *other users*' blogs, as opposed to indicators of mere activity (such as, for example, entries posted). This can be interpreted as evidence that willingness to interact does affect network patterns, and increases number of readers relatively to number of friends. Other variables do not have sizeable effects, whereas the life span of blogs, albeit significant, has a negative effect.

In what follows, we attempt at combining information from the two activity equations by specifying and estimating systems of regression equations. Furthermore, because reciprocity enters the activity equations both as an explanatory and a dependent variable, we apply an Instrumental Variable (IV) technique to the equation for the number of readers. Indeed, we have argued that network imbalances may affect 'friending' patterns, and certainly, such variables are both determined within the same network of relationships.

Before investigating the possibly endogenous effect of reciprocity, we estimate a SURE model, in which reciprocity and number of readers are regressed on the same activity indicators.

4.3.3 The SURE model

Here, we present results from the estimation of a seemingly unrelated regression (SURE) model. SURE (Zellner, 1962) is an attractive estimation method, in that it allows us to exploit information in the DGP of the error terms to improve the efficiency of the OLS estimates. This is done by allowing the error terms in the equations to be correlated. Results from the estimate of the system of equations is shown in table 6 below:

Dependent variable:	ln(friendsof)	reciprocity
ln(member)	0.144***	−0.065**
	(0.039)	(0.031)
ln(entries)	0.395***	−0.033
	(0.048)	(0.038)
ln(comments per post)	0.767***	0.013
	(0.055)	(0.044)
ln(comments per friend)	−0.272***	0.256***
	(0.035)	(0.044)
ln(duration)	0.547***	−0.108***
	(0.051)	(0.041)
constant	0.339	0.062
	(0.286)	(0.231)
obs	141	141
R^2_{adj}	0.765	0.291
F stat (χ^2)	459.89***	59.78***
BP ($\chi^2_{(1)}$)	39.903 (0.000)	
corr	0.532	

Table 6: SURE.

Legend: Standard errors are in parentheses; BP is the Breusch-Pagan test for independence (p-values in parentheses); corr is the off-diagonal term in the correlation matrix of residuals.

Here, as we expect, the estimated coefficients are the same as in the separate OLS regressions presented in the previous section. Interestingly, however, a sizeable correlation appears in the variance-covariance matrix of the residuals – The only covariance term is displayed in the bottom line of table 6. Indeed, some efficiency gain appears to be confirmed by lower estimated standard errors than those reported for the OLS equations. Parameters of community membership and duration are now more significant than in the single equation model for reciprocity. The Breusch-Pagan test for the diagonality of the variance matrix decisively rejects its null of independence of these residual series. In view of the fact that both equations have the same regressors, this seems to indicate that there is indeed some "common factor", perhaps unobservable, which affects both reciprocity and the number of readers and is omitted by the model.

The validity of the SURE method is based on the condition that no dependent variable should enter one of the other equations in the system as an explanatory variable. When estimating the first activity equation, however, we found that reciprocity has a significant effect on the number of 'friends of'. So, in what follows, we attempt at combining information from the two activity equations in a manner that allow reciprocity and number of friends to be simultaneously determined.

4.3.4   Instrumental Variable Estimation

This section looks at the problem of the suspected endogeneity of the reciprocity measure by estimating an equation for the number of readers where reciprocity is "instrumented" by the measures of bloggers' activity. This allow us to test for the endogeneity of reciprocity, and to treat is as an exogenous variable in the model of the number of readers.

Table 7 below presents results from the estimation of such model, fitted by 2 Stage Least Squares (2SLS), where the explanatory variable *reciprocity* is "instrumented" by measures of activity:

ln(friendsof)	IV	OLS
ln(member)	0.075	0.123***
	(0.072)	(0.046)
ln(entries)	0.359***	0.268***
	(0.094)	(0 .057)
ln(comments per post)	0.781***	0.682***
	(0.104)	(0.060)
ln(comments per friends)	(instrument)	
ln(duration)	0.432***	0.373***
	(0.073)	(0 .094)
reciprocity	-1.061***	0.198***
	(0.391)	(0.117)
constant	0.405	0.637
	(0.482)	(0.634)
obs	141	141
R^2	0.375	0.700
Wald χ^2	154.73***	49.96***
Hausman test	15.05** (0.010)	

Table 7: IV regression: model estimates and comparison with OLS regression.
Legend: Robust standard errors in parentheses.

Identification is achieved by excluding *comments per friend* from the regression.[18] Hausman's test statistics rejects its null of exogeneity of reciprocity at the 1% significance level. So, there is some indication that OLS estimation yields inconsistent results, supporting the adequacy of the IV procedure. IV coefficients are slightly higher than those produced by the corresponding OLS regression (these are reported in the second column of table 7, for comparison). The variable member, however, is no longer significant. Noticeably, reciprocity enters the activity equation with an elasticity greater than 1, and its effect is now negative in sign. Results from the IV estimation using Generalised Method of Moments (GMM), which produces consistent *and* efficient estimates when errors are non-*iid*, yields identical results.

Interestingly, these results are consistent with bloggers attaching a 'stigma' to network imbalances. Indeed, the negative sign of reciprocity implies that when the number of readers increases, relatively to the number of bloggers who are befriended, the number of readers decreases. In future work, we will examine whether it is the degree of symmetry in the bloggers' networks that influences the number of readers (networks that are more symmetric are more attractive, while asymmetries in any direction tend to be stigmatized), or whether bloggers simply prefer befriending a blogger who has less friend of compared to friends because those bloggers then can devote more attention to themselves.

4.4 Robustness analysis

To examine the robustness of our conclusions, we have considered different versions of the activity equations estimated in sections 4.3.1 and 4.3.2. In particular, we have replaced the variable *member* with two different measures of bloggers' involvement in communities: the variable *community*, which measures the number of communities read by a blogger, and a dummy which takes value zero when the subject does not read or belong to any community. Tables 8 and 9, in Appendix C, present estimates from the fitting of the activity equations with, respectively, the variable *community* and the dummy in place of *member*. Results provide clear evidence that significant activity effects come out as robust results. Magnitude, sign and significance of individual coefficients do not change substantially in those models, compared to the original ones. The same conclusions are achieved when re-estimating the first activity equations using IV regression. In particular, reciprocity maintain its significant and negative effect. Furthermore, results do not change when inactive blogs are excluded from the analysis.[19]

To check for the correct specification of the functional form, and examine possible 'kinks' in the response to changes in activity, the original models have also been fitted using a non-linear version of the multiple regression equations 6 and 7, namely a generalised additive framework (GAM). This works by fitting non-linear, non-parametric, components in addition to linear terms. Non-linear components, however, are not significant, supporting the linear specification of the models.[20]

5 Conclusion

This paper analysed patterns of relationship and content production among bloggers from a theoretical and empirical perspective. The analysis has identified statistically significant relationships between network size, structure of the network, and content production using real data. In particular, levels of content production and blogging activity have been found to be related to the size of and level of imbalance in the network of relationships in which a blogger is inscribed.

Future work will be based on the collection of data over several periods in order for example to determine which comes first: having many readers or producing a lot of content. We will also be exploring more general functional forms for the utility an agent derives from belonging to a network of bloggers. We hope this will allow us to define some flexible formula relating attention given and received at the aggregate level for each agent. We will also be collecting additional data on the type of relationship entertained between bloggers on LiveJournal, notably on reciprocation of friendship at the individual level. We will also exploit data that identifies the most popular LJ users based on their Google rank, a score Google attributes to websites depending on the number of websites they are linked from.

Another interesting extension to this study is to measure not only quantity but quality of content. This would enable us to distinguish between a vast majority of blogs that are only of

'local' interest for their friends and a minority of blogs whose content (and audience?) evidences a level of achievement of the blogger in his/her chosen topic. We would also like to consider whether bloggers might not have different 'ethos'; for example, some bloggers will be very careful with and will attach a lot of meaning to friending when this means letting access to screened, more personal entries. Some others will consider their list of friend as a reading list, devoid of any further meaning. Obviously, those two types of bloggers may display the same level of activity and yet have widely different number of friends. This type of difference in 'ethos' is not something that we can capture with our measures, but it could be proxied by the level of imbalances in bloggers' networks. Finally, there is valuable work to do exploiting data on communities on LJ, and the way they federate agents around themes without their members necessarily establishing direct links.

Notes

[1] This expression of net surplus does not take account of the fact that additional attention by a new friend will lead the blogger to increase his own activity, thus also leading in an increase in his cost. Similarly, the new friend will herself increase her own activity. This expression therefore considers only the short term benefit of establishing a new link. The long term benefit depend on the result of a whole chain of reaction and counter-reaction to the establishment of this new friendship. For a full model, see appendix E.

[2] On LiveJournal, agents can put some of their friends on special filters, allowing them to read those friends' entries separate 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 their use is usually not publicized by the blogger for fear of alienating others.

[3] Other blog hosts include Google's Blogger and WordPress. Blogging features are also offered in social networking sites such as MySpace and Facebook, but these are not widely used.

[4] More information on LJ and its history can be found on its Wikipedia entry, at http://en.wikipedia.org/wiki/livejournal.

[5] Source: www.livejournal.com/stats.bml (accessed on 13 December 2005).

[6] www.deadjournal.com

[7] Users can provide additional information in the "bio", a space where bloggers present their profiles.

[8] Users have the option not to reveal their gender, an option that is taken by about a third of users.

[9] http://www.livejournal.com/random.bml

[10] Source: http://www.livejournal.com/stats.bml.

[11] Note, however, that "comments posted" includes comments posted in communities, while "comments received" includes only comments made on the user's journal. This implies that, on average, a blogger will have received more comments than she makes on another's blog.

[12] *Member* denotes the number of communities one is member of, whereas *Community* is the number of communities read by the blogger; this distinction is relevant here because a blogger cannot post in a community unless she is a member of that community, but she can follow a community's debate without being a member.

[13] Comments per post are those comments *received* by the blogger, averaged by the number of posts, whereas comments per friend are comments *made* by the blogger in other blogs, and averaged by the number of friends.

[14] This is performed by computing squared-residuals and a measure of leverage associated to the data-points. The observations removed are the 109th and 223nd. These bloggers presents the highest difference between the number of *friends* and *friendsof* in the sample, with just one blogger listed as a friend.

26

[15]Robust standard errors are computed using the option `robust` in `Stata`, which implements the sandwich estimator proposed by Huber and White.

[16]In practice, added variable plots present residuals from a regression of the covariate x_j on the set of remaining covariates in the model (the vector X_{-j}) versus residuals of the regression of X_{-j} on y, the response; in doing so, they extract the genuine contribution of each explanatory variable to the model.

[17]It should be noted that the effects of the activity variables on the degree of asymmetry of the network, in the context of this linear regression model, is ambiguous. For example, a positive sign in an estimated coefficients indicates that the ratio friendsof/friends increases when the value of the variable associated to that coefficients increases. It does *not* say, however, whether the network becomes more or less asymmetric.

[18]Results of the Sargan's over-identifying restriction test does not support the inclusion of further instruments.

[19]Results for IV regressions and active-only blogs are not reported for reasons of space. They are available form the author on request.

[20]Additive models (Hastie and Tibshirani, 1990) are flexible regression frameworks, which allow departures from linearity while preserving the additive structure of a standard linear regression model. Non-linear components are non-parametric in the sense that functional forms are not specified and the components are fitted by non-parametric algorithms, such as kernels or splines. As a result, they provide an ideal setting to explore functional forms of relationship among variables of interests, and can be used for checking regression models. Results are available from the authors upon 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 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.

Bialik, C. (2005). Measuring the impact of blogs requires more than counting. *The Wall Street Journal*. May 26.

Blood, R. (2000). Weblogs: A history and perspective. *Rebecca's Pocket*. September 7, http://www.rebeccablood.net/essays/weblog_history.html.

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

Bruni, L. and R. Sugden (2008). Fraternity: Why the market need not be a morally free zone. *Economics and Philosophy 24*, 35–64.

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

Domon, K. (2001). Content competition vis-à-vis viewer externalities. Working Paper, School of Social Sciences, Waseda University.

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

Finin, T., L. Ding, L. Zhou, and A. Joshi (2005). Social networking on the semantic web. *The Learning Organization 12*(5), 418–435.

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.

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.

Hastie, T. J. and R. J. Tibshirani (1990). *Generalized Additive Models*. Chapman & Hall / CRC.

Henning, J. (2005). The blogging geyser. *Newsletter of the Web Marketing Association*. April 8.

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*.

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

Huffaker, D. (2004). *Gender Similarities and Differences in Online Identity and Language Use among Teenage Bloggers*. Ph. D. thesis, Georgetown University.

Huffaker, D. A. and S. L. Calvert (2005). Gender, identity, and language use in teenage blogs. *Journal of Computer-Mediated Communication 10*(2).

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.

Lemann, N. (2006). Journalism without journalists. *The New Yorker*. August 7.

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*.

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 (2004). The blogging iceberg. http://perseus.com/survey/resources/perseus_blogging_iceberg.pdf.

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

Rainie, L. (2005, January). The state of blogging. Pew Internet & American Life Project.

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).

Ribstein, L. E. (2005). Initial reflections on the law and economics of blogging. Working Paper, University of Illinois College of Law.

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

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

von Hippel, E. (1998). Economics of product development by users: The impact of 'sticky' local information. *Management Science 44*(5), 629–644.

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

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests of aggregation bias. *Journal of the American Statistical Association 57*, 500–509.

A Data Description

A.1 Original data

- User: User name (pseudonym)

- Location: Region and /or country where the blogger is based. 1 if Australia; 2 if Canada; 3 if Czech Republic; 4 if India; 5 if Philippines; 6 if Russian Republic; 7 if Scotland; 8 if Singapore; 9 if Sweden; 10 if UK; 11 if US.

- Birth date: Blogger's date of birth

- Friends: Number and list of weblogs read by the blogger. Limited to other blogs on LJ.

- Friend of: List of those bloggers with an account on LJ who read one's weblog

- Mutual friends: A subgroup of 'friend of' : 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 'friend of' : number and list of those bloggers whose friendship is not reciprocated. Note that Friend of = Mutual friends + Also friend of

- 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.

- Syndicated accounts: Number and list of those weblogs not on LiveJournal that are read by the blogger via LJ

- 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.

- Account type: 1 if 'free'; 2 if 'paid'; 3 if 'early adopter'. 'Early adopter' are the first few members of LJ. There are also some 'permanent accounts', paid for life, grouped with 'paid'. The introduction of 'sponsored accounts' with advertising was made subsequent to our data collection.

- Date created: Date on which the weblog was created.

- Date updated: Last date on which the weblog was updated (*i.e.* an entry was posted).

- Journal entries: Number of posts written since the weblog was created

- Comments posted: Number of comments made on entries made in other weblogs or communities.

- Comments received: Number of comments received from other bloggers or in response to them on one's own entries.

A.2 Processed data:

- Age: Age of the blogger

- 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

- Reciprocity: Friend of divided by friends, expressed in logarithm.

B Figures

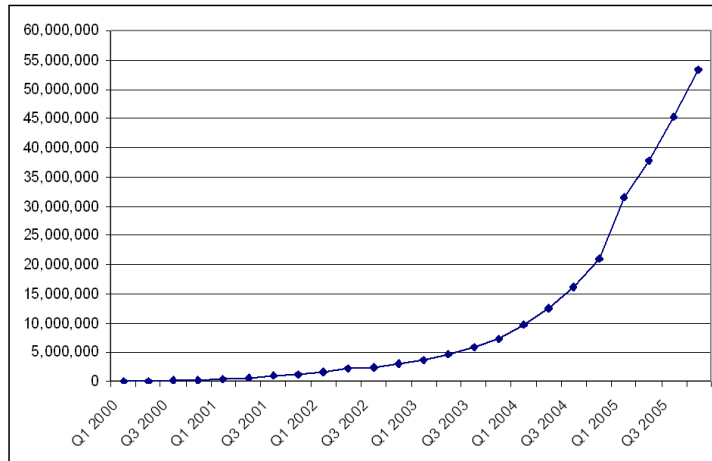Figure 4: Number of hosted weblogs created between 2000 and 2005.
Source: Perseus, The Blogging Geyser, available at: http://www.perseus.com/blogsurvey/geyser.html (accessed 04 January 2006).

C   Robustness analysis

Dependent variable:	ln(friendsof)	ln(friendsof)	reciprocity
ln(community)	0.082*	0.112***	−0.045
	(0.048)	(0.042)	(0.033)
ln(entries)	0.414***	0.440***	−0.039
	(0.085)	(0.085)	(0.035)
ln(comments per post)	0.781***	0.757***	0.036
	(0.065)	(0.060)	(0.055)
ln(comments per friend)	−0.247***	−0.402***	0.235***
	(0.068)	(0.057)	(0.047)
active	0.194	0.095	0.151
	(0.155)	(0.142)	(0.104)
ln(duration)	0.546***	0.626***	−0.122**
	(0.057)	(0.053)	(0.055)
reciprocity		0.658***	
		(0.097)	
constant	0.191	0.221	-0.046
	(0.326)	(0.300)	(0.280)
obs	153	153	153
RSS (df)	47.484 (146)	35.640 (146)	27.325 (146)
R^2_{adj}	0.701	0.774	0.303
F stat	58.66***	70.27***	6.75***
BP	33.81 (0.000)	86.25 (0.000)	3.86 (0.695)
RESET	2.00 (0.019)	1.73 (0.042)	1.63 (0.074)

Table 8: Multiple regression: robustness analysis (I). Model with logarithm of *community*. Legend: robust standard errors are in parentheses; p-values for BP, RESET in parentheses.

Dependent variable:	ln(friendsof)	ln(friendsof)	reciprocity
community	0.277***	0.211**	0.127
	(0.104)	(0.095)	(0.089)
ln(entries)	0.441***	0.488***	-0.088
	(0.066)	(0.068)	(0.042)
ln(comments per post)	0.819***	0.820***	-0.001
	(0.054)	(0.044)	(0.047)
ln(comments per friend)	-0.233***	-0.382***	0.284***
	(0.053)	(0.054)	(0.045)
active	0.105	0.080	0.049
	(0.128)	(0.118)	(0.114)
ln(duration)	0.560***	0.640***	-0.153***
	(0.052)	(0.049)	(0.042)
reciprocity		0.523***	
		(0.086)	
constant	0.053	0.089	-0.069
	(0.300)	(0.285)	(0.217)
obs	191	191	191
RSS (df)	58.139 (184)	46.862 (183)	41.129 (184)
R^2_{adj}	0.755	0.801	0.318
F stat	105.39***	112.57***	9.06***
BP	33.52 (0.000)	62.53 (0.000)	17.47 (0.007)
RESET	2.01 (0.026)	1.25 (0.243)	2.30 (0.009)

Table 9: Multiple regression: robustness analysis (II). Model with dummy *community*.
Legend: robust standard errors are in parentheses; p-values for BP, RESET in parentheses.

D SURE

Table 10 presents results on the estimation of the two equation for *friends* and *friends of* in the same system, to gather information on the data generating processes of number of readers and number of friends. Notably, the correlation between the error terms of the two equations is high and significant.

Dependent variable:	ln(friendsof)	ln(friends)
ln(member)	0.144***	−0.210***
	(0.039)	(0.034)
ln(entries)	0.395***	−0.428***
	(0.048)	(0.042)
ln(comments per post)	0.767***	0.753***
	(0.055)	(0.049)
ln(comments per friend)	−0.272***	−0.528***
	(0.035)	(0.039)
ln(duration)	0.547***	0.655***
	(0.051)	(0.045)
constant	0.339	0.276
	(0.286)	(0.254)
obs	141	141
R^2_{adj}	0.765	0.815
F stat (χ^2)	459.89***	623.17***
BP	$\chi^2(1) = 57.585 \ (0.000)$	
corr	0.640	

Table 10: SURE.
Legend: standard errors are in parentheses; BP is the Breusch-Pagan test for independence (p-values in parentheses).

E   Solving a full model of reciprocal attention

In this appendix, we consider a full model of reciprocal attention. The model we present below, while accounting for a diversity of motivations for agents in the network, allow us to examine how these different motivations translate into different outcomes.

Consider a simple model with three agents who produce their own content and read content generated by others in the network. We call 'Friends of i' the agents that are read by i. $e = (e_i, e_j, e_k)$ denotes the vector of effort into producing content by agents i, j and k respectively. $n = (n_{ij})_{i \neq j}$ denotes the vector of attentions. (Agent i devotes attention n_{ij} to the content produced by $j \neq i$.)

A simple additive form for the total utility of an agent, i in this instance, is

$$(8) \quad U_i(n,e) = \underbrace{\lambda_i \sum_{j \neq i} n_{ji} e_i}_{\text{Utility from being read}} - \underbrace{\frac{1}{2} e_i^2}_{\text{Cost of producing content}} + \underbrace{\sum_{j \neq i} n_{ij} e_j}_{\text{Utility from reading others}}$$

with $\lambda_i \geq 0$. λ_i measures the propensity to enjoy being read compared to the propensity to enjoy reading others (normalized to 1 here). There is no cost to reading others (the cost of reading content produced by others is likely to be negligible compared to the cost of producing such content).

Each agent faces a time budget K_i in the attention they can devote to others, so that for example i faces constraint $\sum_{j \neq i} n_{ij} \leq K_i$.

We assume an agent decides first on how much effort e she will put into writing. This determines her schedule of effort as a function of the attention she expects to be given to her. In a second stage, the agent decides on how much attention n she will allocate to others. She knows by backward induction how this will affect their effort in the first stage[21]. The Nash equilibrium of this two stage game is such that the vector of attention in the second stage corresponds to what was anticipated in the first stage when effort was made, and no agent wishes to deviate from the equilibrium schedule (e,n) of effort and attention.

In what follows we consider two polar cases:

A   In the first case, $\lambda_i = \lambda_j = \lambda_k \equiv \lambda$ and $K_i = K_j = K_k \equiv K$: agents have the same level of motivation between reading and writing and they all have the same time budget.

B   In the second case, $\lambda_i = \lambda_j \equiv \lambda$ and $\lambda_k = 0$ while $K_i = 0$ and $K_j = K_k \equiv K$: agent i is a 'writer', that is, she has no interest and thus no time for reading others. Agent k is a 'reader', that is, she derives no utility from being read but has got time to read others. Agent j is a reader/writer who derives utility from writing but also has time for others.

We show that symmetric agents devote equal attention to each members of their network and produce the same amount of content (case A). The impact of asymmetries in agents' motivations

is also analysed (case B), which leads us to postulate the existence of a norm of reciprocity. We show that, if the norm of reciprocity is such that a link between an agent i and an agent j is established only if it results in i gaining as much as j from the establishment of the relationship, then an agent i that has few friends will link with an agent j who has more friends only if that agent j offers more content than what agent i offers.

A) Let us consider a network where all agents are motivated the same way, that is, $\lambda_i = \lambda_j = \lambda_k \equiv \lambda$ and $K_i = K_j = K_k \equiv K$.

Proposition 1 *The unique equilibrium of the game of reciprocal attention with three symmetric agents is such that each agent produces effort $e = \lambda K$ and devotes attention $n = \frac{K}{2}$ to each of the other agents.*

Proof. In appendix F ∎

Each agent thus allocates half of its time budget to each of the other agents, and produces the same effort; she also devotes the same level of attention to each other agent as those other agents devote to itself: $n_{ij} = n_{ji}$.

Generalising from the above, it is possible to show that in a network with n agents with time budget K then each agent will devote attention $\frac{K}{n-1}$ to every other agent and will produce effort λK. Each agent derives utility $\lambda(1 + \frac{1}{2}\lambda)K^2$ from participating in the network irrespective of its size. As the network size grows, each agent has less time to devote to other agents but their overall attention remains constant. Since the agent produces content in proportion of the sum of attention she gets, content is invariant with the number of friends. This leads us to spell out the following hypothesis:

Hypothesis 3 *If bloggers are all identical, then bloggers' total effort devoted to writing and interacting with other agents does not depend on the number of readers or friends they have (H1').*

Note however the limitations in the setting in which H1', which is the converse of H1, can be expected to hold:

1. All agents are supposed to be linked with each other. This will obviously not be the case in a network such as LiveJournal where several million bloggers are present. However, several studies (references) show a 'small world' effect in blogging networks, with small networks of heavily interlinked individuals (high in-degrees) with little links with other networks (small out-degrees). If all agents belong to such 'small worlds', then the model tells us that their effort should be invariant with the size of the network they belong to.

2. All agents devote the same time budget to blogging. There is little way for us to check this is the case, though it is possible to some extent to proxy time budget with age.

3. All agents are motivated the same way. We will see in the following what happens if agents respond to different motivations.

4. There are no network economies of scale, whereby for example agents would benefit from the diversity of content provided by their friends or would learn to better manage their relations or be more effective in producing content as their network size grows.

One should also note how counter-intuitive proposition 1 appears to be. Contrary to a competitive model where agents try to accumulate more friends (and influence and/or prestige) by producing more writing and interacting heavily with each other, this model assumes that agents know they are part of a system which subsistence depends on them devoting as much attention to others as is devoted to them. This contributes to limit any asymmetries between agents.

B) Consider now the asymmetric network whereby $\lambda_i = \lambda_j \equiv \lambda$ and $\lambda_k = 0$ while $K_i = 0$ and $K_j = K_k \equiv K$. The reason to posit that an agent that enjoys being read (high λ) will devote less attention to others (low K) is that this agent will devote a lot of effort to writing which will give her less time to devote attention to others. Obviously, this is not necessarily the case, and one could very well argue that λ and K would increase together, for example because a blogger would get inspiration from reading others. However, it is still interesting for our purpose, which is to illustrate the dynamics of networks of bloggers, to look at the above case.

Proposition 2 *The reader (agent k) produces no content and devotes most of her attention to the reader/writer (agent j). The writer (agent i) devotes no attention to others and produces more content than all others. The reader/writer (agent j) devotes all her attention to the writer (agent i) and produces a moderate amount of content.*

Proof. In appendix G ∎

The above findings are summarised in table 11.

	Attention given	Attention received	Effort	Friends	Friend of	$\ln \frac{\text{Friend of}}{\text{Friends}}$
Writer (i)	0 to j 0 to k	K by j $\frac{1}{4}K$ by k	$\frac{5}{4}K$	0	2	∞
Reader/Writer (j)	K to i 0 to k	0 by i $\frac{3}{4}K$ by k	$\frac{3}{4}K$	1	1	0
Reader (k)	$\frac{1}{4}K$ to i $\frac{3}{4}K$ to j	0 by i 0 by j	0	2	0	$-\infty$

Table 11: Attention and effort in an asymmetric network

From table11, one can see that, in an unbalanced network with varying representative motivations, the presence of a writer reduces the total amount of content produced in the network even

the writer himself produces more content than in a balanced network. The presence of a reader contributes to re-balancing somewhat the network, as she allocates more of her attention to the reader/writer than to the writer. This effect comes about not because the reader/writer pays any attention to the reader, but because the reader wants to encourage the reader/writer to produce content, knowing that the writer produces content anyway since she gets the whole attention of the reader/writer.

In the table, we assume that the number of friends of an agent is the number of agents she pays attention to, and the number of friend of is the number of agents that pay attention to her. From the table, there would be a positive correlation between how unbalanced an agent's network is (proxied by $\ln \frac{\text{Friend of}}{\text{Friends}}$) and how much content is produced by that agent, and there would also be a positive correlation between how many readers (friend of) one has and how much content one produces.

F   Proof of proposition 1

Consider agent i's decision in the first stage. She is maximizing her objective function $U(n,e)$ with respect to e_i. Denote e_i^* her optimal choice of effort. From the first order conditions of her maximization problem we obtain

(9) $$e_i^* = \lambda \sum_{j \neq i} n_{ji}$$

This means that the agent will produce content in proportion to the total attention she expects to receive. More content is produced as λ, which measures the level of utility from being read, increases.

Consider now agent i's decision in the second stage. She maximizes her objective function $U(n,e^*)$ with respect to $n_{i.}$, the attention she devotes to others, s.t. the constraint $\sum_{j \neq i} n_{ij} \leq K$. The Lagrangian of this maximization problem is

(10) $$\lambda \sum_{j \neq i} n_{ji} e_i^* + \sum_{j \neq i} n_{ij} e_j^* - \frac{1}{2} e_i^{*2} - \nu (\sum_{j \neq i} n_{ij} - K)$$

which agent i maximizes w.r.t. to both terms in $n_{i.}$: n_{ij} and n_{ik}. Replacing e_i^* with their expression in (9) and taking the first derivative of the above expression, we obtain the following first order conditions of the maximization problem:

(11) $\quad \lambda(2n_{ij} + n_{kj}) - v = 0$

(12) $\quad \lambda(2n_{ik} + n_{jk}) - v = 0$

(13) $\qquad\qquad n_{ij} + n_{ik} = K$

Replicating the same FOC for j and k, we obtain the following system of equation

(14) $\quad 2n_{ij} + n_{kj} = 2n_{ik} + n_{jk}$

(15) $\quad 2n_{ki} + n_{ji} = 2n_{kj} + n_{ij}$

(16) $\quad 2n_{jk} + n_{ik} = 2n_{ji} + n_{ki}$

(17) $\qquad n_{ij} + n_{ik} = K$

(18) $\qquad n_{ki} + n_{kj} = K$

(19) $\qquad n_{jk} + n_{ji} = K$

This is a system of six equations with six unknowns, which is solved for

(20) $\quad n_{ij}^* = n_{ik}^* = n_{ki}^* = n_{ji}^* = n_{kj}^* = n_{jk}^* = \dfrac{K}{2}$

Putting this back in (9) we conclude that

(21) $\quad e_i^* = e_j^* = e_k^* = \lambda K$

Each agent derives utility $\lambda(1 + \frac{1}{2}\lambda)K^2$ from their participation in the network.

G   Proof of proposition 2

In the first stage, both i and j will produce content in proportion to the attention they will receive from others, while agent k (the reader), will produce no content, so

(22) $\quad e_i^* = \lambda \sum_{j \neq i} n_{ji}$

(23) $\quad e_j^* = \lambda \sum_{i \neq j} n_{ij}$

(24) $\quad e_k^* = 0$

Consider now agent i 's decision in the second stage. Agent i will devote no attention to others since her time budget is 0, so $n_{ij} = n_{ik} = 0$.

Agent j will maximizes her objective function $U_j(n, e^*)$ with respect to $n_{j\cdot}$, the attention she

devotes to others, s.t. the constraint $\sum\limits_{i \neq j} n_{ji} \leq K$. The Lagrangian of this maximization problem is

(25) $\quad \frac{1}{2}\lambda^2 n_{kj}^2 + n_{ji}\lambda(n_{ji} + n_{ki}) - \nu(n_{ji} + n_{jk} - K)$

which agent j maximizes w.r.t. to n_{ji} and n_{jk}. This results in j optimally setting $n_{ji} = K$ and $n_{jk} = 0$; there is no point devoting attention to k since k produces no content.

Agent j will maximizes her objective function $U_k(n, e^*)$ with respect to $n_{k.}$, the attention she devotes to others, s.t. the constraint $\sum\limits_{i \neq k} n_{ki} \leq K$. The Lagrangian of this maximization problem is

(26) $\quad n_{ki}\lambda(n_{ji} + n_{ki}) + n_{kj}\lambda(n_{ij} + n_{kj}) - \nu(n_{ki} + n_{kj} - K)$

which agent k maximizes w.r.t. to n_{ki} and n_{kj}. This results in the following system of two equations:

(27) $\quad \lambda(n_{ji} + 2n_{ki}) = \lambda(n_{ij} + 2n_{kj})$

(28) $\quad\quad n_{ki} + n_{kj} = K$

Knowing from the above that $n_{ji} = K$ and $n_{ij} = 0$, one obtains that $n_{kj} = \frac{3}{4}K$ and $n_{ki} = \frac{1}{4}K$. k devotes more attention to j than to i in order to encourage production by j and thus better balance content produced across the agents she reads.

41