Economics Bulletin

Volume 30, Issue 3

Reciprocal attention and norm of reciprocity in blogging networks

Alexia Gaudeul
GSBC, Friedrich Schiller University, Jena

Chiara Peroni
Statec & Centre de Recherche Publique Henri Tudor

Abstract

Bloggers devote significant time not only producing content for others to read, watch or listen to, but also paying attention to and engaging in interactions with other bloggers. We hope to throw light not only on the factors that gain bloggers significant readership and lively interactions with their audience, but also on the rules that govern their relations with others. We relate bloggers' activity with the size and structure of their network of fellow bloggers. A blogger's readership increases with his activity, while bloggers who read back proportionally fewer of their readers tend also to be more active. We find evidence that those bloggers who read back proportionally fewer of their readers have less readers than bloggers who reciprocate more, but tend to receive more comments per posts.

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, at the Annual Conference of the Royal Economic Society in Guilford in April 2009, at the AEA workshop "Usage et Impact d'Internet" in Marseilles in November 2009 and at the CREM-EconomiX workshop "Digital Business Models: Understanding Strategies" in Paris in June 2010. We are grateful to Caterina Giannetti, Adelina Gschwandtner, Peter Moffatt and Paul Seabright for useful discussion. Views and opinion expressed in this article are of the authors solely and are not those of the Statec and CRP.

Citation: Alexia Gaudeul and Chiara Peroni, (2010) "Reciprocal attention and norm of reciprocity in blogging networks", *Economics Bulletin*, Vol. 30 no.3 pp. 2230-2248.

Submitted: Mar 22 2010. **Published:** August 26, 2010.

Bloggers devote significant time not only producing content for others to read, watch or listen to, but also paying attention to and engaging in interactions with other bloggers. We hope to throw light not only on the factors that gain bloggers significant readership and lively interactions with their audience, but also on the rules that govern their relations with others. We relate bloggers' activity with the size and structure of their network of fellow bloggers. A blogger's readership increases with his activity, while bloggers who read back proportionally fewer of their readers tend also to be more active. We find evidence that those bloggers who read back proportionally fewer of their readers have less readers than bloggers who reciprocate more, but tend to receive more comments *per* posts.

1. LiveJournal

Our object of study is LiveJournal.[1] LiveJournal provides a platform for bloggers to publish their blogs and form relations with other bloggers. Bloggers establish relations by reading each other, commenting on each other's entries and participating in communities of shared interests.

LiveJournalers ("LJers") tend to maintain personal journals, or "diaries" rather than thematic, or "filter" blogs (Shirky, 2003; Wei, 2009). LJ's social networking features encourage long term engagement not only with the practice of blogging but also with other LJers (Marwick, 2009). A LJer may add any blogger on LJ to the list of those he reads ("friends" in LJ's terminology). A friend's entries automatically appear on his reading page along with other friends' entries, in reverse chronological order. A friend may reciprocate readership by listing our blogger as a friend as well, in which case the new friend automatically joins our blogger's list of "readers" (LJers who read him). A blogger may read and comment freely on any blogs except if that blog's entries can be viewed only by friends or where commenting is restricted to friends.

Not all LJers understand the act of friending and of reciprocating friendship in the same way (Raynes-Goldie and Fono, 2006). Some will reciprocate readership from all and sundry. Others will reciprocate only after lengthy interactions. Friending may be conditional on reciprocation so those who do not reciprocate after a while may be dropped. Once a mutual relation has been established though, breaking it ("unfriending") will not go without explanation...

A large part of the motivation for writing is to gain readers or to maintain a circle of friends. The ecology of LJ is such that one's reading encourages others' blogging, and others' blogging encourages one's writing. For example, Danah Boyd, researcher and participant in LJ, writes that "most of my friends (ed: on LJ) who read what I write (...) produce very little but I would produce absolutely nothing if they weren't reading what I wrote".[2]

2. Literature

Our research deals with social networking, and more particularly blogging. We are interested in the relation between strength of interpersonal ties, individual network properties and individual activity within social networks (Granovetter, 1973). We explore the differences between bloggers who entertain mainly reciprocal (strong) relations and those who have many asymmetric (weaker) relations (for

[1] http://www.livejournal.com

[2] "grrr re: basic account decision", http://danahboyd.livejournal.com/1396.html, March 13th, 2008, accessed December 18, 2009.

example, many unreciprocating friends and unreciprocated readers). Studies of reciprocation patterns show that intimate, repeated relations are more likely to be reciprocal than occasional ones, as are relationships with close kins vs. friends (Marsden, 1990). Reciprocal friendships are more stable over time than those friendships that are not reciprocated (Runger and Wasserman (1979), commenting on Hallinan (1978)). Intransitive friendships appear to cause psychological distress (Vaquera and Kao, 2008).

This paper focuses on 1) why bloggers read each other and according to which criteria, 2) why some bloggers have more readers than others and finally, 3) how bloggers who entertain only reciprocal relations differ from those who attach little importance to reciprocity. Regarding point 1, Nardi et al. (2004) underline the social aspects and dynamics of online diaries and speculate that "blogging is as much about reading as writing, as much about listening as talking". In this, blogging differs from pure social networking, where the appeal of a potential friend appears to increase with its number of friends (Tong et al., 2008), or from other media where the distinction producer/consumer is clearly delineated. Kumar et al. (2004) find that bloggers who read each other tend to be of similar age, live in the same city or region and have similar interests, though Paolillo et al. (2005) find that LiveJournal users' interests and their network of friends are largely uncorrelated. Backstrom et al. (2006) show that bloggers who belong to the same blogging communities tend also to read each other. In relation to point 2, Hofstetter et al. (2009) show that bloggers with more readers produce more content and *vice-versa*. Lento et al. (2006) explain that bloggers with many friends and readers are more likely to maintain their activity within blogging networks. Mishne and Glance (2006) show that popular blogs attract more comments than others. Finally, on the topic of reciprocity, Herring et al. (2005) show that A-list bloggers (those that have many readers) tend to link to each other though they experience low rates of reciprocation. Some recent studies also consider how reciprocity norms play out in online activities. Gu et al. (2009) demonstrate that a norm of (indirect) reciprocity helps maintain file sharing networks, as participants are ready to punish or reward others based on their past contributions. Sadlon et al. (2008) show how submitters on Digg, a social news website, are able to attain prominence by voting up each others' link submissions.

To the best of our knowledge, little research is concerned with reciprocity in blogging and with investigating the relation between an agent's number of links and the likelihood those are reciprocated. We thus explore in this project, among other questions, whether being perceived as of a 'not reciprocating' or of a 'not reciprocated' type influences one's likelihood to attract readers.

3. MODEL

Consider two unrelated bloggers, i and j, exerting effort e_i and e_j respectively and with n_i and n_j friends respectively. Define u_{ij}, the utility derived by i from reading j, as $f(e_j, n_j) - \theta_{ij}\eta + \epsilon_{ij}$. $f(.)$ is a continuous function taking its value over $\mathbb{R}$, increasing in its first argument and decreasing in its second. θ_{ij} is zero if the friendship is reciprocated, one otherwise. η is a positive number. ϵ_{ij} is the realization of a random variable ϵ continuously distributed over $\mathbb{R}$, drawn independently for each pair i and j. The first part of the utility function means that a blogger who exerts high levels of effort e_j and has only few friends n_j will be attractive because he offers plentiful content as well as a high level of individuated attention

to his friends.[3] The second part of the equation expresses the effect of not adhering to the reciprocity norm: not only is i not read back in return, but he also suffers the indignity of not being acknowledged as a friend. The third part expresses how inherently attractive j is to i.

i will read j only subject to $u_{ij} > A_i$ with A_i the best available alternative for i (another blogger, for example). Suppose the blog market is such that bloggers are free to friend or unfriend whom they like and have decent search tools to look for the best available alternatives. Blogger j with more friends than alternative friend k may still be preferred to k as long as he exerts sufficiently higher effort than k. For e_j sufficiently high, he may even be able not to reciprocate i's reading and still not lose her to a reciprocating k. Bloggers who exert higher effort are thus able to attract more readers and maintain more friendships than bloggers who are less active. However bloggers who attract more readers must either reciprocate, which dilutes their attention to their friends, or not reciprocate. Both options reduce their attractiveness. The process of friendship formation thus reaches an equilibrium.

This model is a variation on Jackson and Wolinsky (1996)'s co-authorship model. In that model, an author wishes his co-author to limit the number of his co-authorships so he can devote more time to their common project. However, to the difference of this co-authorship model, higher content production can at least partially compensate for the low intensity of the relation. Bloggers who gain more friends have the choice between maintain the intensity in their posting of comments on their friends' blogs, or provide more content by writing more entries. A blogger might indeed be happy reading a very productive other blogger without receiving much attention in exchange, in the same way as for example, an academic may be happy collaborating with a very productive other, even if that collaboration is sporadic and unpredictable.

We also consider in our model the impact of not reciprocating a friending act. LJers are uncomfortable not reciprocating readership and are reluctant to establish a link unless it is very likely to be reciprocated. The norm of reciprocity is thus a social norm as *per* Elster (1989).[4] Among other consequences, this means it is important for a blogger to project an image as a "reciprocator" in order to attract readers. The ratio of readers to friends is information that is readily available on LJ. Our empirical analysis will consider a high ratio as a proxy for non-adherence to the reciprocity norm.

4. DATA DESCRIPTION

The data used in this paper is a sample of 2445 Anglo-Saxon blogs[5] hosted on LiveJournal and selected at random on the 10th of February 2010. Data was collected using `Screenscraper`, which is software for website data extraction. The list of friends and readers were processed with `R` to assign them into mutual friends, unreciprocated readers and unreciprocating friends. Data analysis and regressions were performed with `Stata`.

[3] In a fuller model, we specify i's total utility along with his cost function $C(e_i + n_i a_i)$, with a_i the level of individuated attention devoted to each friend, and determine under what condition a_i is indeed decreasing in n_i.

[4] pp. 99-100: "For norms to be social, they must be shared by other people and partly sustained by their approval and disapproval. They are also sustained by the feelings of embarrassment, anxiety, guilt and shame that a person suffers at the prospect of violating them."

[5] Blogs giving as their location either the US (1926), Canada (217), the UK (211), Australia (80) or Ireland (11).

Our model is an equilibrium model, which means it applies only to bloggers who have had the time to find friends to read, attract readers, join communities, learn about the options available on LJ, and more generally learn about the norms of behavior in this setting. In order to evaluate how long it takes for bloggers to become established on LJ, we looked at when bloggers start to change their default options, such as distinguishing mutual friends from other readers on their profile (the default is to list all readers together), or getting a paid account (most LJers start with a free account). The trend for "mutual shown" is increasing after about three months before stabilizing after about one year (figure D.1). We thus chose to consider only the 2019 bloggers that have been active for more than 100 days.

We consider two main measures of a blogger's activity: number of entries made *per* day, which is a measure of content produced and comments posted *per* friends *per* year, which is a measure of attention given to friends. We also consider the number of communities joined as those facilitate finding new friends, and the longevity of the blog ("duration"), as blogs maintained for longer have more time to establish new relations. We relate those with measures of the blogger's network: the number of his readers ("readers") and the number of blogs he reads ("friends"). Those can be divided in "mutual friends" (friend and reader), "unreciprocated readers" (not friend but reader) and "unreciprocating friends" (friend but not reader) (table 1 in appendix A). We also consider an alternative measure of readership: the number of comments received *per* reader *per* year. This measures the active involvement of readers. Finally, we consider the effect of having a paid account as this gives access to many features on LJ.[6] More details on variables are given out in appendixA.

Table 2 in appendix B reports summary statistics. The median blogger maintained her blog for about four years and a half. Over that period, she made one entry every three days, received about 5 comments from each of her readers each year, and posted about 8 comments on the blogs of each of her friends each year. She had 33 friends (blogs she reads) and 35 readers (bloggers that read her). She followed and thus probably participated in about 6 communities. 27 of her friendships were mutual (both read and read by). The median ratio of friends to readers ("balance") was 1, but 13% of her readers were not read back in return (% unreciprocated) and 13% of her friends did not read her back (% unreciprocating). On average, 43% of the bloggers in our sample chose to distinguish mutual friends from other friends on their profile, while 27% had a paid account.

In view of their skewness, variables were transformed on the logarithm scale for performing regression analysis (appendix A.4). Statistics on the regressions variables, as well as their correlations, are provided in tables (3) and (4) in appendix B.

5. HYPOTHESES

The perspectives introduced in the model section allow us to express the following set of hypotheses:

The first hypothesis ("H1") deals with the relationship between activity and active/passive readership (as measured by comments received and number of readers respectively):

[6] "What are the Paid Account benefits?", http://www.livejournal.com/support/faqbrowse.bml?faqid=131, accessed March 16, 2010.

H1: Bloggers that have many readers and/or receive many comments are more active, in terms of comments posted, entries written, communities participated in and length of activity.

H1 is derived directly from our model: Bloggers who have many readers and/or receive lots of comments must either devote more effort in reading back and replying to each of them or compensate lower attention with higher activity, either in their own blogs or in communities. Against H1, bloggers with more readers may be inherently more attractive because they give access to a wider network of relations than others, irrespective of their activity.

The second hypothesis ("H2") deals with the relationship between network composition and activity:

H2: Bloggers with high balances of readers to friends and/or with many unreciprocated readers, are more active. Bloggers with many unreciprocating friends are less active.

H2 is derived from our model: A blogger with a high balance or many unreciprocated readers does not read back a portion of his readership. This portion then maintains its reading only if the blogger is producing more content than another blogger who would reciprocate would. Similarly, a blogger with many unreciprocating friends is not read by some of his friends, which presumably means he is unattractive (low level of activity). Against H2, bloggers might not care about being read back, or might apply a different logic to very popular blogger whom they understand are not able to reciprocate their readership.

The third hypothesis ("H3") deals with the how not reciprocating readership (interpreted as a failing to adhere to the reciprocity norm) affects one's ability to gain readers or elicit comments:

H3: Bloggers with high balances of readers to friends and/or with many unreciprocated readers, have less readers and receive less comment than bloggers with lower balances and/or less unreciprocated readers. Bloggers with many unreciprocating friends have more readers and receive more comment than bloggers with less unreciprocating friends.

H3 is also derived from the model: Everything else being equal, bloggers who value attention get more value from reading a blogger who reads them back than from a blogger who does not. Therefore, a blogger who fails to reciprocate will be less attractive and thus have less readers and/or receive less comments than one who does. Reciprocation is a function of activity according to H2, and we therefore will have to instrument measures of reciprocation in order to determine their true effect on one's number of readers and comments received. We will thus use two-stage least-square regressions to test H3.

H3, if verified, means that bloggers who initiate friendships are likely to trigger automatic reciprocation. Bloggers that are derisively called "adders" on LJ seem to exploit such a norm by randomly adding many friends and seeing many of those friendships automatically reciprocated. However, and this leads us to doubt the applicability of H3 to the number of comments received, it is unlikely that the quality of the relations thus established is high. We expect that adding friends only after a number of successful interactions and a sustained period of mutual observation, while generating lower number of readers and translating into an unwillingness to reciprocate unbidden friendships, will also induce closer relations (more comments exchanged) within one's network.

6. RESULTS

We first test hypotheses 1 and 2 by performing linear regressions of the measures of readership structure and readers' involvement on measures of activity. As in all further regressions, outlier observations were removed along the procedure proposed by Belsley (1980), which identifies highly influential observations as those characterized by either a high leverage or a high residual.[7] The results are presented in table 5 in appendix C. The Breusch-Pagan test statistics reveals heteroskedasticity in all the regressions, so we report t-statistics using White's robust covariance matrix estimator.

The results of the regression of the number of readers on activity support H1: duration, number of communities joined and the number of entries *per* day all contribute positively to the number of readers. The effect of comments posted *per* friend is negative, but small and less significant. This indicates that there is only a limited decrease in attention paid *per* friend even as readership increases. We will see that this is at the cost of reciprocating fewer readerships. Paid accounts accumulate more readers, either because they have access to more functionality that increase the attractiveness of their blog, or because paying indicates they are more involved in their blogging practice.

There is partial support for H1 in the regression of active readership (comments received *per* readers) on activity. The number of entries posted *per* day and of comments posted both have a large positive effect. The effect of the number of communities participated in, while smaller, is however negative. It may be that bloggers who participate in more communities are thus distracted from interacting with their friends. The effect of duration is also small and negative. It is possible that interactions decrease over time because their main role is to establish rather than maintain relation.

Columns 3, 4 and 5 give only partial support for H2. As should be expected, regressions for the ratios of unreciprocated readers to mutual friends ("unreciprocated", column 4) and unreciprocating friends to mutual friends ("unreciprocating", column 5) are almost perfect mirror images of each other, while the results of the regressions for the ratios of unreciprocated readers to mutual friends and of readers to friends ("balance") are very similar.[8] Bloggers that exert more effort in posting comments on their friends' blogs have higher balance and unreciprocated ratios, and lower unreciprocating ratios. This means that bloggers who pay more attention to their friends (more comments posted) are unwilling to add back new readers and/or add friends to their friend list as this would dilute their attention to existing friends and/or require extra effort. Bloggers with high balances tend to participate in fewer communities, maybe for the same reasons. The balance increases with duration (column 3). This effect seems mainly driven by lower ratios of unreciprocating friends over time (large negative coefficient in column 5) rather than by more unreciprocated readers (smaller positive coefficient in column 4). The balance thus increases over time mainly because unreciprocating friends are dropped or add the blogger back (see figure D.2). The negative coefficient on mutual shown indicates that bloggers who choose to distinguish between mutual friends and unreciprocated readers tend to have lower balance ratios. The possibility to do such a distinction was introduced by LJ after protest by those bloggers

[7] Median regressions, which are more robust to outliers than OLS, yield similar results.

[8] To understand this, note that the balance ratio can be expressed as the ratio of the unreciprocated ratio over the unreciprocating ratio (appendix A.3).

who did not appreciate being friended by unknown people, and wanted a way to single them out. Singling out unreciprocated readers appears to be effective deterrence.

We then test hypothesis 3, which has to do with the effect of the reciprocity norm (or the failure to abide by it) on popularity and levels of interactions (number of readers and/or comments received). The first set of regressions (Table 6 in appendix C) deals with the number of readers. The balance, unreciprocated and unreciprocating ratios were instrumented with the variables "mutual shown" and "comments posted *per* friends". This is because, as argued before, "mutual shown" seems to indicate that one attaches higher value to reciprocal relations, while exerting effort in commenting on friends' blogs appears to limit one's willingness to reciprocate readership. The results show that instrumented reciprocation ratios have a significant influence on one's number of readers. As hypothesized in H3, bloggers with lower balances of readers to friends have more readers. Tests of the strength of the instruments (Shea's partial R^2), of the endogeneity of the reciprocation ratios (Durbin-Wu-Hausman, robust to heteroskedasticity), of the orthogonality of the instruments (Sargan-Hansen) and of the relevance of both instruments (Lagrange Multiplier), all confirm the validity of our IV regressions.[9]

The second set of instrumental regressions deals with the number of comments received *per* reader *per* year (table 7 in appendix C). "Mutual shown" is our only instrument because "comments posted" is a strong influence on "comments received" so that its exclusion would lead to biased estimates. Tests confirm the endogeneity of the reciprocation ratios, but also show that "mutual shown" is a weak instrument. Contrary to our hypothesis H3, bloggers with higher ratios of readers to friends tend to receive more comments from their readers. This confirms the doubts we expressed regarding the applicability of H3 to the number of comments received (section 5).

7. Conclusion

Based on a range of evidence from studies of bloggers and of reciprocation patterns, we formulated a model whereby bloggers care not only about the content offered by their friends but also about getting attention from them. The model's predictions were tested against a novel dataset from a major blogging community, LiveJournal. We showed a positive relation between a blogger's activity, the number of his readers and the active involvement of those readers through comments. We also observed that those bloggers who failed to reciprocate readership were sanctioned with lower number of readers than their activity would otherwise warrant. However, we also found that bloggers who were loathe to reciprocate obtained more involvement from their readers through comments.

This study thus illustrates that bloggers who wish to gain an audience must not only exert persistent effort in producing content but also return the attention of their readers. Adherence to the reciprocity norm increases one's number of readers but does not foster more involvement from those readers. It thus appears that bloggers face a choice between reciprocating freely and accumulating many readers, or being more discriminating so as to cultivate existing mutual links. Note however that our sample was drawn from LiveJournal, where blogs have social networking features and often take the form of diaries. Reciprocity may

[9]Tests were performed using the `first, endog, orthog` and `redundant` options of `Stata`'s `ivreg2` package (Baum et al., 2003).

play less of a role on other blogging platforms and our results may not generalize to thematic, A-list blogs.

Further work will exploit weekly observations of a sample of 3000 bloggers, gathered over more than a year, in order to measure to what extent gaining an audience motivates effort, as opposed to the extent in which effort attracts an audience. Another sample comparing the characteristics of a sample of 100 bloggers and of the members of their network of friends will help us clarify what leads bloggers to reciprocate readership.

References

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

Baum, C. F., M. E. Schaffer, and S. Stillman (2003). Instrumental variables and GMM: Estimation and testing. *Stata Journal 3*(1), 1–31.

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

Elster, J. (1989). Social norms and economic theory. *The Journal of Economic Perspectives 3*(4), 99–117.

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

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*.

Hallinan, M. T. (1978). The process of friendship formation. *Social Networks 1*(2), 193–210.

Herring, S. C., I. Kouper, J. C. Paolillo, and L. A. Scheidt (2005). Conversations in the blogosphere: An analysis "from the bottom up". In *Proceedings of the 38th Hawaii International Conference on System Science*.

Hofstetter, R., S. K. Shriver, H. S. Nair, and K. Miller (2009). Social ties and user generated content: Evidence from an online social network. *NET Institute, Working Paper 09-28*.

Jackson, M. O. and A. Wolinsky (1996). A strategic model of social and economic networks. *Journal of Economic Theory 71*, 44–74.

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

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*.

Marsden, P. V. (1990). Network data and measurement. *Annual Review of Sociology 16*, 435–463.

Marwick, A. (2009). LiveJournal users: Passionate, prolific and private. 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*.

Nardi, B. A., D. J. Schiano, and M. Gumbrecht (2004). Blogging as social activity, or, Would you let 900 million people read your diary? In *Proceedings of the 2004*

ACM conference on Computer Supported Cooperative Work, pp. 222–231. ACM Press.

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*.

Raynes-Goldie, K. and D. Fono (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.

Runger, G. and S. Wasserman (1979). Longitudinal analysis of friendship networks. *Social Networks 2*(2), 143–154.

Sadlon, E., Y. Sakamoto, H. Dever, and J. Nickerson (2008). The karma of Digg: Reciprocity in online social networks. In *Proceedings of the 18th Annual Workshop on Information Technologies and Systems*.

Shirky, C. (2003, February 8). Power laws, weblogs, and inequality. `http://shirky.com/writings/powerlaw_weblog.html`.

Tong, S. T., B. V. D. Heide, L. Langwell, and J. B. Walther (2008). Too much of a good thing? The relationship between number of friends and interpersonal impressions on Facebook. *Journal of Computer-Mediated Communication 13*(3), 531–549.

Vaquera, E. and G. Kao (2008). Do you like me as much as I like you? Friendship reciprocity and its effects on school outcomes among adolescents. *Social Science Research 37*(1), 55–72.

Wei, L. (2009). Filter blogs vs. personal journals: Understanding the knowledge production gap on the Internet. *Journal of Computer-Mediated Communication 14*(3), 532–558.

APPENDIX A. DESCRIPTION OF VARIABLES

A.1. **Data in bloggers' profile.**

- User: User name (pseudonym)
- Location: Country where the blogger is based.
- Friends: Number of weblogs read by the blogger. Limited to other blogs on LJ.
- Readers (or "friend of" in LJ's terminology): List of those bloggers with an account on LJ who read one's weblog.
- Mutual shown: Dummy variable, equal to one if the blogger chose the option to lists mutual friends and unreciprocated readers separately on his profile, rather than list all his readers together as is the default.
- 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.
- Communities: Number of communities the blogger reads or is member of.
- 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).

- Date collected: Date on which data on the weblog was collected.
- 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.

A.2. **Mutual friends, unreciprocated readers and unreciprocating friends.** Friends and readers were divided in:

- Mutual friends: A subgroup of 'readers'; those bloggers whose reading is reciprocated.
- Unreciprocated readers ("also friend of" in LJ lingo): A subgroup of 'readers'; those bloggers whose reading is not reciprocated.
- Unreciprocating friends: A subgroup of 'friends'; those bloggers who do not reciprocate reading.

In the following table, showing the profile of an hypothetical blogger called "Alexia", with her list of friends and readers side by side, Chiara is a mutual friend, Gabby is an unreciprocated reader and Wade an unreciprocating friend. Alexia therefore has one mutual friend, one unreciprocated reader and one unreciprocating friend. Also, mutual shown is equal to zero since Alexia does not distinguish in her profile between mutual friends and unreciprocated readers.

Alexia's profile	
Friends	**Readers**
Chiara	Chiara
Wade	Gabby

TABLE 1. Alexia's list of friends and readers, as shown on her profile.

A.3. **Descriptive statistics:**

- Duration: Difference between the date of creation and the date of last update (in days).
- Inactive: Difference between the date of collection and the date of last update (in days).
- Entries *per* day: Number of journal entries divided by duration.
- Comments received *per* readers *per* year: Comments received divided by number of readers, divided by duration, multiplied by 365.
- Comments posted *per* friends *per* year: Comments posted divided by number of friends, divided by duration, multiplied by 365.
- Reciprocation ratios:
  - Balance: Readers divided by friends. This is easily available to a blogger who considers whether to friend a bloggers, since the blogger's public profile includes his number of friends and friend of.
  - Unreciprocated: unreciprocated readers divided by mutual friends
  - Unreciprocating: unreciprocating friends divided by mutual friend.

  Note that since

  $$friends=(mutual\ friends+unreciprocating\ friends)$$

  while

  $$readers=(mutual\ friends+unreciprocated\ readers)$$

then

$$balance = \frac{1+unreciprocated}{1+unreciprocating}$$

so that in effect, “unreciprocated” and “unreciprocating” are a way to divide up the “balance” ratio into two parts.

However, the decomposition of this statistic into the two other statistics requires painstakingly comparing the list of names of friends and of readers so as to find who is in both and who is in only one. A blogger does not generally have access to this statistic, except when a blogger chooses to display his mutual friends separately from his unreciprocated readers (mutual shown=1), in which case “unreciprocated” is known.

For some reason (!), bloggers on LJ never choose to display how many bloggers do not reciprocate their readership. As a matter of fact, this option is not even provided by LJ...

A.4. **Regression variables.**

- lreaders=ln(readers)
- lreceived=ln(comments received *per* reader *per* year)
- llbalance=ln(ln(1+balance))
- llunreciprocated=ln(ln(1+unreciprocated))
- llunreciprocating=ln(ln(1+unreciprocating))
- lentries=ln(entries *per* day)
- lposted=ln(comments posted *per* friends *per* year)
- lcommunities=ln(communities)
- lduration=ln(duration)

Appendix B. Summary statistics

variable	mean	median	sd	skewness	min	max	N
friends	62.61	33.00	92.54	3.99	0.00	1159.00	2019
readers	76.94	35.00	173.28	9.99	1.00	3151.00	2000
mutual friends	51.41	27.00	78.76	4.73	0.00	1158.00	1998
unreciprocated readers	25.60	4.00	141.20	15.37	0.00	3091.00	1998
unreciprocating friends	11.84	4.00	25.61	6.84	0.00	384.00	1998
balance	1.79	1.00	17.26	35.42	0.06	690.00	1987
% unreciprocated	0.20	0.13	0.22	1.42	0.00	1.00	1998
% unreciprocating	0.18	0.13	0.18	1.35	0.00	1.00	1985
paid account	0.27	0.00	0.45	1.01	0.00	1.00	2019
mutual shown	0.43	0.00	0.50	0.28	0.00	1.00	2019
communities	18.37	6.00	56.78	26.25	0.00	2121.00	2019
duration	1615.36	1774.00	935.53	-0.16	100.00	3575.00	2019
inactive	2.10	1.00	2.05	1.14	0.00	9.00	2019
entries/day	0.70	0.33	4.74	39.92	0.00	204.81	2019
received/readers/year	9.11	5.39	18.07	13.29	0.00	443.70	2000
posted/friends/year	15.10	7.66	35.72	10.77	0.00	776.77	1988

Table 2. Descriptive statistics

variable	mean	median	sd	skewness	min	max	N
lreaders	3.32	3.56	1.57	-0.40	0.00	8.06	2000
lreceived	1.68	1.76	1.16	-0.54	-4.29	6.10	1848
llbalance	-0.35	-0.37	0.33	0.28	-2.92	1.88	1987
llunreciprocated	-1.62	-1.58	1.14	-0.23	-5.88	1.83	1532
llunreciprocating	-1.73	-1.64	1.01	-0.51	-7.05	1.28	1596
lentries	-1.22	-1.10	1.29	-0.55	-7.24	5.32	2019
lposted	2.01	2.13	1.25	-0.33	-3.09	6.66	1865
lcommunities	2.05	1.95	1.35	0.25	0.00	7.66	1847
lduration	7.06	7.48	0.98	-1.17	4.61	8.18	2019

Table 3. Regressions variables

TABLE 4. Correlations between regressions variables

Variables	lreaders	lreceived	llbalance	llunreciprocated	llunreciprocating	lentries	lposted	lcommunities	lduration
lreaders	1.0000								
lreceived	0.1566	1.0000							
	(0.0000)								
llbalance	0.3523	0.1523	1.0000						
	(0.0000)	(0.0000)							
llunreciprocated	0.0363	0.0171	0.6251	1.0000					
	(0.1956)	(0.5414)	(0.0000)						
llunreciprocating	-0.3355	-0.2518	-0.4020	0.1312	1.0000				
	(0.0000)	(0.0000)	(0.0000)	(0.0000)					
lentries	0.4375	0.6541	0.1388	0.0296	-0.1766	1.0000			
	(0.0000)	(0.0000)	(0.0000)	(0.2923)	(0.0000)				
lposted	0.1125	0.7206	0.3963	0.2416	-0.2612	0.3981	1.0000		
	(0.0001)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)			
lcommunities	0.4085	0.0480	-0.0720	-0.0652	0.0230	0.2066	0.1374	1.0000	
	(0.0000)	(0.0868)	(0.0102)	(0.0201)	(0.4133)	(0.0000)	(0.0000)		
lduration	0.3260	-0.0682	0.1425	0.0128	-0.1544	0.2094	-0.1161	0.1282	1.0000
	(0.0000)	(0.0150)	(0.0000)	(0.6483)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	
mutual shown	0.0267	-0.0043	-0.1200	-0.0820	0.0775	0.0928	-0.0021	0.0749	0.0235
	(0.3416)	(0.8771)	(0.0000)	(0.0034)	(0.0057)	(0.0009)	(0.9392)	(0.0075)	(0.4015)

Appendix C. Regressions tables

	(1) lreaders	(2) lreceived	(3) llbalance	(4) llunreciprocated	(5) llunreciprocating
lentries	0.3711***	0.4294***	-0.0106*	-0.0667*	0.0153
	(18.68)	(29.52)	(-2.04)	(-2.39)	(0.74)
lposted	-0.0403*	0.5255***	0.0967***	0.3334***	-0.2755***
	(-2.13)	(39.70)	(18.70)	(12.86)	(-14.59)
lcommunities	0.2842***	-0.0976***	-0.0295***	-0.0500*	0.1113***
	(19.39)	(-10.26)	(-7.52)	(-2.49)	(6.97)
lduration	0.5700***	-0.1778***	0.0801***	0.0899**	-0.3181***
	(23.65)	(-10.27)	(12.51)	(2.59)	(-12.16)
paid account	0.2371***	0.0272	0.0215+	-0.2022***	-0.2826***
	(5.29)	(1.00)	(1.93)	(-3.41)	(-5.89)
mutual shown			-0.0579***	-0.1582**	0.1179**
			(-5.83)	(-3.04)	(2.83)
_cons	-0.6488**	2.6432***	-1.0642***	-2.8371***	1.0198***
	(-3.25)	(18.66)	(-20.32)	(-10.00)	(4.92)
N	1599	1545	1597	1272	1402
R^2	0.6026	0.7792	0.2734	0.1164	0.2312
F	522.1970	1065.5642	83.2251	33.4054	84.7339
RESET	1.9713	0.3965	0.9266	3.3991	1.6274
p-value	0.1163	0.7556	0.4271	0.0173	0.1812

t statistics in parentheses

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5. H1 and H2: Activity, readership and network structure, robust without outliers

	(1) lreaders	(2) lreaders	(3) lreaders	(4) lreaders
lentries	0.3545***	0.3614***	0.3413***	0.3577***
	(19.68)	(18.64)	(16.67)	(13.29)
lcommunities	0.2824***	0.2628***	0.2443***	0.2226***
	(19.24)	(16.80)	(14.96)	(11.22)
lduration	0.5824***	0.6030***	0.4930***	0.5236***
	(25.12)	(24.97)	(17.92)	(13.18)
paid account	0.2264***	0.2572***	0.1799***	0.2687***
	(5.06)	(5.47)	(3.63)	(4.36)
llbalance		-0.4388**		
		(-2.73)		
llunreciprocated			-0.2397***	
			(-3.89)	
llunreciprocating				0.2551**
				(2.70)
_cons	-0.8333***	-1.0885***	-0.4242+	0.2881
	(-4.75)	(-5.47)	(-1.91)	(1.09)
N	1599	1580	1332	1211
R^2	0.6015	0.5782	0.4680	0.3324
F	651.1122	501.1557	319.3329	194.9772
RESET	2.2532	0.3077	1.8104	0.2142
p-value	0.1333	0.5791	0.1785	0.6435

t statistics in parentheses

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

TABLE 6. H3: Readers, with mutual shown and lposted as instrument, robust without outliers

	(1) lreceived	(2) lreceived	(3) lreceived	(4) lreceived
lentries	0.4294***	0.4530***	0.4742***	0.4328***
	(29.52)	(19.50)	(12.28)	(19.07)
lposted	0.5255***	0.3949***	0.3669***	0.3910***
	(39.70)	(5.43)	(4.09)	(5.54)
lcommunities	-0.0976***	-0.0670***	-0.0494+	-0.0404
	(-10.26)	(-3.30)	(-1.93)	(-1.49)
lduration	-0.1778***	-0.2738***	-0.2385***	-0.3331***
	(-10.27)	(-4.80)	(-5.25)	(-4.08)
paid account	0.0272	-0.0233	0.0668	-0.1140
	(1.00)	(-0.53)	(1.24)	(-1.35)
llbalance		1.0964+		
		(1.86)		
llunreciprocated			0.4833+	
			(1.79)	
llunreciprocating				-0.5112*
				(-1.97)
_cons	2.6432***	3.9671***	4.1424***	3.0596***
	(18.66)	(5.35)	(4.68)	(9.25)
N	1545	1545	1327	1217
R^2	0.7792	0.6323	0.3544	0.5432
F	1065.5642	498.7282	265.0197	374.3156
RESET	1.2075	1.3353	2.6219	0.0362
p-value	0.2718	0.2479	0.1054	0.8490

t statistics in parentheses

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

TABLE 7. H3: Comments received *per* reader *per* year, with mutual shown as instrument, robust without outliers

APPENDIX D. EFFECTS OF DURATION

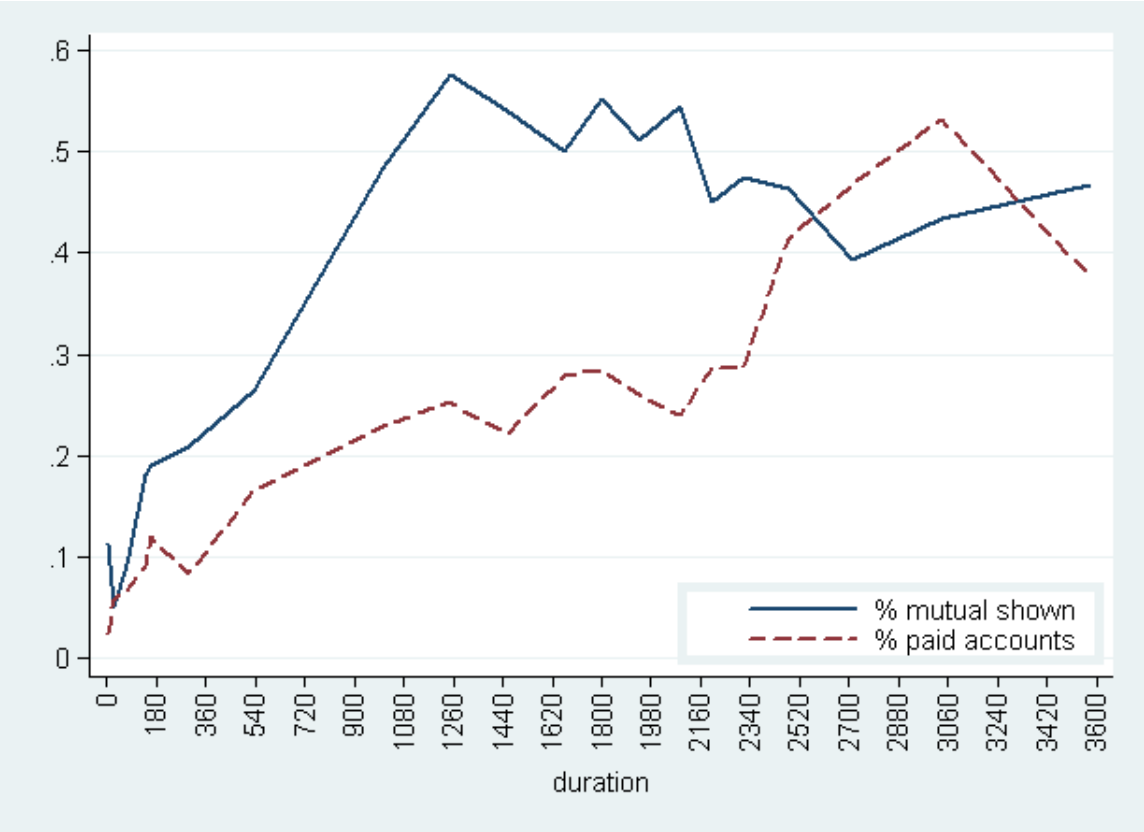

FIGURE D.1. % of bloggers showing mutual friends and % with a paid account, by duration vigintiles.

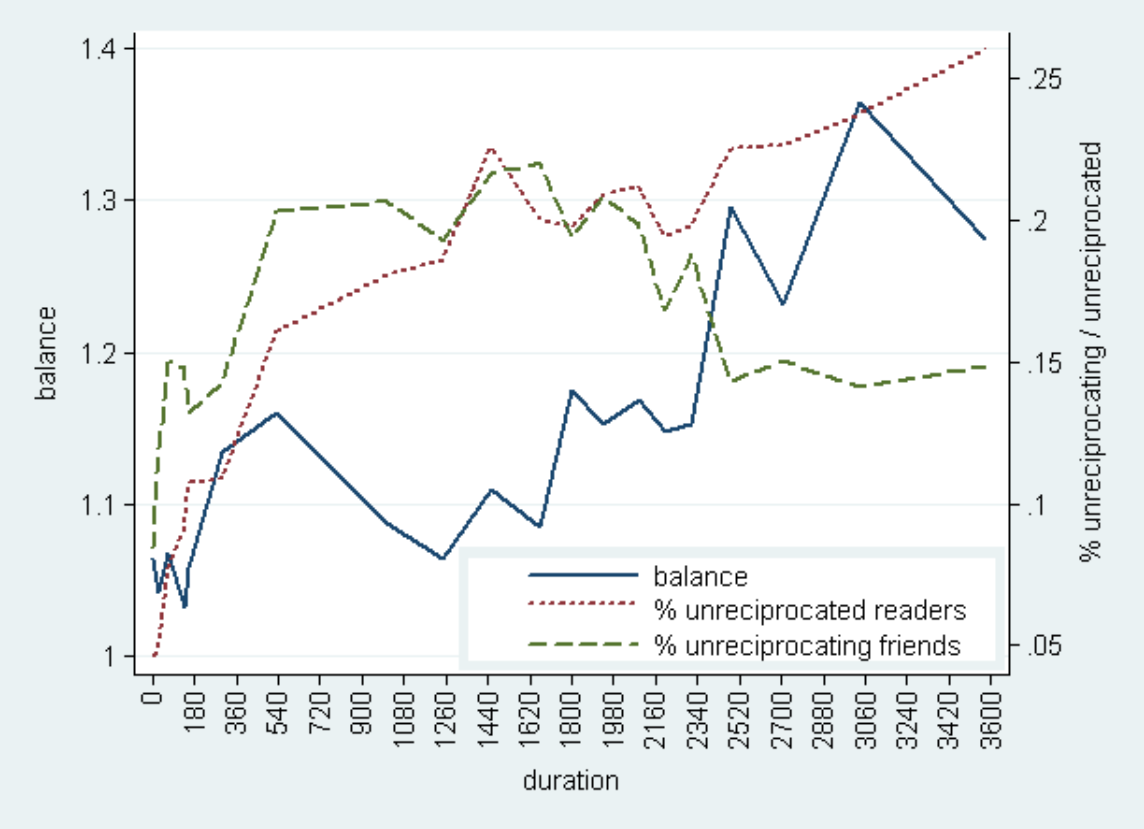

FIGURE D.2. Balance and % unreciprocated readers and unreciprocating friends, by duration vigintiles.