Link to Success: How Blogs Build an Audience by Promoting Rivals

July 2008

Dina Mayzlin
dina.mayzlin@yale.edu
203-436-4262
Fax: 203-432-3003

Hema Yoganarasimhan
hema.yoganarasimhan@yale.edu

Yale School of Management[1]
135 Prospect Street
P.O. Box 208200
New Haven, CT 06520-8200

[1] Dina Mayzlin is an Associate Professor of Marketing at the Yale School of Management. Hema Yoganarasimhan is a Ph.D. student in the marketing department of the Yale School of Management. We thank David Godes, Birger Wernerfelt, Jose Silva, Jiwoong Shin, K. Sudhir, and Sachin Sancheti for comments that have improved the paper. We also thank participants at the Yale SOM Ph.D. seminar, Marketing Science Conference 2006, SICS 2006 conference, the Yale SOM Wednesday lunch seminar, the CMU's Heinz School seminar, the University of Minnesota marketing seminar, and the University of Maryland marketing camp, and the Fourth Symposium on Statistical Challenged in Electronic Commerce Research for helpful comments. Both authors contributed equally, and their names are listed in alphabetical order.

Link to Success: How Blogs Build an Audience by Promoting Rivals

Abstract

According to a survey by the Pew Internet Project, 12 million U.S. adults keep Web logs (or blogs), websites that provide commentary on various topics. In addition, 57 million adults read blogs. However, it is not obvious how readers find blogs that provide useful information given the multitude of blogs, the heterogeneity in their quality, and the lack of brand equity in this context. This article attempts to explain the popularity of blogs, given these obstacles. We suggest that linking may be the responsible for the success of blogs. We explain why a blogger may choose to link to another blog, though doing so may result in losing a reader to that linked blog. The proposed model distinguishes blogs along two dimensions: 1) the ability to break the news and, 2) the ability to link to other blogs with news-breaking stories. By linking, a blog demonstrates its ability to link to news-breaking blogs in the future, though it also indicates that a competing blog is likely to feature a news-breaking story in the future. In equilibrium, blogs that are more likely to break the news are more likely to have incoming links and hence receive more visitors. The empirical evidence is consistent with the theoretical predictions.

Keywords: networks and marketing, competitive strategy, game theory, media, e-commerce

1. Introduction

In 1994, a Swarthmore College student, Justin Hall, created an online personal journal called Links.net, now recognized as the first Web log or "blog" (Rosen 2004). Since then, creating (or "blogging") and reading blogs have become mainstream online activities. According to the Pew Internet Project, 12 million U.S. adults maintain blogs, and 57 million U.S. adults read them (Lenhart et al. 2007). Firms and politicians also engage with bloggers. For example, in 2003, Dr Pepper/7Up recruited six teenage bloggers to promote its new flavored milk product, Raging Cow, on their sites (Walker 2003). During the 2006 Fashion Week in New York, 40 fashion bloggers received coveted access to designer runway shows (Dodes 2006), and seven of the eight Democratic Party Presidential candidates gathered at the 2007 "Netroots" Forum in front of an audience of liberal bloggers (Zeleny 2007). Further, in 2007, U.S. advertising spending on blogs reached $94.1 million (eMarketer.com 2008).

Blogs represent part of the larger set of online social media, which include online forums and bulletin boards, social networking sites, and video sharing sites. Although both blogs and other social media involve user-generated content, blogs also share some characteristics of newspapers. For example, blogs provide information to readers, and the mode of transmission is often one-to-many. David Winer, a blogging pioneer, gives the following definition: "A blog is like a personal newspaper.… It's sort of publishing on a small scale." (Potier 2003). Blogs generally are chronological, updated regularly with the most recent posting at the top, and relative, that is, "You're often writing about something other people have written." (Potier 2003).

For example, consider the blog A VC, at www.avc.blogs.com ("Musings of a VC [venture capitalist] in NYC"), by Fred Wilson, a partner in Union Square Ventures. The blog's posts range

from the personal—"I've been in a funk for the past three days and I don't know why" (April 22, 2008)—to the general—"So why is Facebook worth $15bn and Wordpress is worth $200mm?" (April 18, 2008). Some posts break news, such as, "Disqus [which Union Square Ventures financed] announced a new feature release and an investment today" (March 18, 2008). Other posts contain information originally reported on another blog, for example – "Microsoft has apparently agreed to acquire Xobni [included a link to a post on TechCrunch]" (April 20, 2008). TechCrunch broke the news in the following post, "Two independent sources tell us that the Microsoft/Xobni deal is moving along and that Microsoft signed an acquisition LOI in the week" (April 20, 2008).

Although the nature of news-breaking events differ across domains, links to other blogs are common. For example, on May 11, 2006, www.daddytypes.com ("The Weblog for New Dads") posted an announcement about a two-day sale starting at Netto Collection, an upscale children's furniture store in Manhattan: "Looks like Netto Collection's having a sample sale. I have no idea what is there, but I do know that it's already been going for four hours…." The post then provided a link to www.daddydrama.com, which originally had posted the information on May 9.

These examples illustrate several interesting aspects of blogs. First, a blog can provide useful information that may not be available in the mainstream media. However, because most blogs lack the brand equity of a mainstream publication and because there are so many blogs, it is not obvious how readers find blogs that provide useful information. Second, from a consumer's perspective, distinguishing the useful from the useless is critical, because blogs often just provide outlets for the author rather than provide useful information to readers. Half of the bloggers

state that "express yourself creatively" represents their major reason for blogging (Lenhart and Fox 2006).[2]

We propose that linking between blogs may help solve consumers' search problem. In particular, we concentrate on dynamic links (or "permalinks") between blogs, which are links to specific posts in other blogs, as opposed to static links (or the "blog roll") that often appear on the right-hand side of a site[3] (see Figure 1). From a blogger's perspective, a one-way link to a potential rival seems to present a risky strategy, because a reader may defect to the linked blog. Moreover, recommending a rival is not confined to this space. For example, Amazon features links to other book sellers, which often have the same product at a discounted price. Although Amazon earns a commission from sales generated through such links, it nonetheless chooses to facilitate a relationship between its customer and another bookstore. Another example involves the decision by WashingtonPost.com to feature links to related articles and blog posts in other publications, next to the *Post*'s articles (Tedeschi 2006).

The common thread among blogs, Amazon, and WashingtonPost.com is that they all offer information to customers. We focus on a particular aspect of information, namely, the ability to deliver timely news.[4] To capture the heterogeneity between blogs, we allow bloggers to differ along two dimensions: (1) the ability to post news-breaking content, and (2) the ability to find news in other blogs. For example, TechCrunch's post about the Xobni deal reflects its ability to

[2] While readers may derive utility from reading about the blogger's opinions and interests, or other types of creative output, we abstract away from these considerations and focus on the information-production aspect of blogs.
[3] The "static" and "dynamic" terminology is due to Marlow (2005), who refers to links between blogs as "social" links to differentiate them from links to news sites, etc.
[4] A recent survey of journalists and editors by Brodeur, a unit of Omnicom, confirms that blogs are an important source of news, even to the professionals in the media industry: 46% of respondents indicate that they find blogs helpful in getting information about breaking news, and 57% read blogs at least two or three times a week. (Brodeur 2008).

break news, whereas A VC's link to TechCrunch reflects its ability to find news in other blogs. The downside of A VC's link is that it demonstrates that TechCrunch is likely to generate useful information; however, the link also allows Fred Wilson, (the author of 'A VC') to signal that he can direct readers to interesting information posted on other blogs. Similarly, Amazon's decision to link to other booksellers may enhance Amazon's image as a destination site, from which customers can either buy directly or find links to other booksellers.

As a byproduct of such incentives, consumers can learn more efficiently as to which blogs deliver news-breaking content. This may be especially important in light of the incentives that companies offer to influence bloggers, as exemplified by Wal-Mart's public relations campaign to get bloggers to post positive comments about the company (Barbaro 2006). Thus, through linking, blogs that are better at breaking news grow their readership more quickly than they would in the absence of linking. This effect gets further accentuated by search engines that commonly offer higher placement to sites with more incoming links.[5]

We organize the remainder of the paper as follows: In Section 2, we discuss previous literature. We present the main model in Section 3, followed by extensions in Section 4. In Section 5, we empirically test two of the model's predictions. Finally, we conclude in Section 6 and discuss some limitations and future work in Section 7.

2. Previous Work

Several literature streams relate to the problem studied herein. First, we turn to literature that deals with the role of hyperlinks on the Internet. Kleinberg 1999 proposes that hyperlinks offer

[5] An ongoing debate in computer science literature considers whether Google accentuates a bias toward popular sites or allows small sites to grow; that is, does Google create a Googlearchy or Googleocracy? See Hindman et al. 2003 and Fortunato et al. 2006.

valuable information because they reflect the subjective judgments of the author who created them. He further offers an algorithm, based on incoming links, to uncover the most authoritative Web pages for a given query. Brin and Page 1998 expanded this idea to develop PageRank, a more flexible algorithm that calculates the authority rank of sites as a function of their incoming links, the basic framework behind Google's search engine. The assumption about the informativeness of the link structure is also analogous to Surowiecki's 2004 "wisdom of the crowd" hypothesis.

While the computer science literature suggests that there is valuable information in links, it takes the structure of the network as a given. In contrast, Bala and Goyal 2000 and subsequent papers (see Demange and Wooders 2005) study network formation as an equilibrium in a non-cooperative game. These papers study issues such as network stability and efficiency, learning and coalitions. Despite the extensive literature in this area, to our knowledge, no research studies the issue of a third party (i.e., the reader) who makes inferences on the basis of the observed pattern of links.

In marketing literature, Katona and Sarvary 2008 investigate strategic linking online and propose a market for advertising, such that a website may sell advertising space or buy an incoming link from another site. Some similarities mark their article and the current work, in that they also find that a site with better content enjoys more incoming links. However, we differ with regard to the proposed mechanism driving this result; Katona and Sarvary 2008 focus on an explicit pricing scheme, whereas we address the role of inferences made by readers when they observe a link.

Finally, Garicano and Santos 2004 and Chen et al. 2002 study the referral process. The former authors show that when an expert diagnoses a problem and decides to address it or refer it to another expert, different revenue-sharing schemes have unique implications for efficiency. In contrast, in our setting, an explicit payment structure between sites is not possible. Chen et al. 2002 consider infomediaries, or Internet services that direct visitors to retailers that are members of their network. Their work therefore deals with an explicit contractual arrangement between the infomediary and its clients, without any inferences by clients about the infomediary's ability to refer to others.

3. Model

3.1 Set Up

We employ a finite-period game with an infinite number of risk-neutral consumers (we refer to a consumer as R, denoting reader) and an infinite number of blogs.[6] Bloggers obtain utility from the size of the readership, due to either increased ad revenue or social utility.[7] Furthermore, we model bloggers as producers, and readers as consumers of information. We use two dimensions to distinguish bloggers' abilities: (1) the ability to break news on their own site and (2) the ability to find news in other blogs. Although we initially assume these abilities are independent, we relax this assumption subsequently in an extension. A blog can be either a high (H) type (probability v of acquiring information) or a low (L) type (probability $w < v$) with regard to breaking news.[8] The high type's superior ability to break news derives either from its insider sources, such as appears to be the case for TechCrunch, or simply from residing at the right location, such as is

[6] This technical assumption simplifies the model. Qualitatively, the results do not change as long as we assume a finite but very large number of blogs. The model does not depend on the assumption that the number of consumers is infinite.

[7] According to Lenhart and Fox (2006), 61% of bloggers listed the desire to "motivate others to action" as a reason for blogging, and 51% listed the desire to "influence the way others think" as a reason

[8] Alternatively, we could differentiate bloggers according to the costs of cultivating insider sources or searching. For example, a blogger with lower costs would be able to break the news with a higher probability. Assuming an exogenous differentiation of costs as opposed to probabilities does not change the results qualitatively but does complicate the model.

the case for Daddydrama, a resident of Manhattan who lives near upscale baby stores. Similarly, a blog can be either H type (probability p) or L type (probability q < p) with regard to finding news in other blogs.[9] Fred Wilson of the A VC blog, appears to be an avid reader of other venture capital and high-tech blogs, which enhances his ability to find news and enables him to link to interesting posts elsewhere. At the beginning of the game, the blogger's type is not observable to the game's players (including the blogger). The assumption of symmetric uncertainty at the start of the game is not critical to our results, as we show in Lemma 1. The prior probability that blogger is H type on ability to break news is γ, and the prior that the blogger is H type on ability to find news in other blogs is δ. Thus, the prior probability that a blog breaks news is $\alpha_0 = \gamma v + (1-\gamma)w$, and the prior probability that it finds news in other blogs is $\beta_0 = \delta p + (1-\delta)q$. In order to clarify the exposition, we henceforth refer to the blogger as "he," and to the reader as "she."

We also assume that a unique piece of information is released every period.[10] Although information contained in two blogs (during the same period) is substitutable, a blog that has broken the news may still link to signal its ability to find news in other blogs. We could alternatively model "information" as any content potentially useful to readers, which would decrease the substitutability between blog content. While there may be complementarities between blogs' content, constraints on a reader's time creates competition for readers. In addition, because introducing complementarities would increase the incentive to link, the stark assumption about substitutability biases the results against finding links in equilibrium.

[9] The probability that a blog can find news on another blog is conditional on the event that at least one other blog breaks the news. However, with the assumption that the number of blogs is infinite, the probability of such an event is 1, and we can treat p and q as independent of v and w, which greatly simplifies the calculations.

[10] If we assumed that information were released probabilistically, a blogger who was not able to break news might not want to link to conceal that information had been released. By assuming information is released with certainty, we rule out this possibility, which simplifies our analysis. We thank the Associate Editor for noting this point.

The game consists of two periods of three stages each (see Figure 2). In the first period, blogs break the news and/or post links to other blogs that broke the news, which influences the visitor's beliefs about the focal and the linked blogs' abilities. This in turn affects his or her choice of blogs in the following period. In the second period, the game repeats.

In stage 1 of period 1, a reader (R) visits a random blog (A).[11] At stage 2, bloggers may gain access to valuable information, which we assume to be verifiable. That is, a blogger must tell the truth in posting a news story since a blogger who lies would be distrusted by other bloggers and readers in the future. Of course, this assumption cannot extend to information that is "soft" or subjective. We assume that all bloggers who obtain news post that information, because there is no strategic reason for doing otherwise. Then, if R visits blog A, which breaks news during stage 2, she derives utility u from the post and updates her prior about A's ability to break news upward. Otherwise, R derives no utility and updates the prior on A's ability to break the news downward.

During stage 3 of period 1, bloggers search other blogs for information, and all those who find a blog with news may link simultaneously. We require that credit be given to the news-breaking blog, which represents a reasonably realistic assumption for two reasons. First, we observe attribution; for example, the Smoking Gun Web site received almost universal credit in the blogosphere for exposing James Frey's memoir *A Million Little Pieces* as largely fictional.[12] Second, a blog that consistently plagiarizes may be exposed by other bloggers. Further, news-breakers are highly motivated to prevent others from copying their content without attribution. In the model,

[11] To simplify the exposition, we focus on a particular reader (R), and we assign names to blogs (A, B, and C). However, all readers and blogs face the same game.
[12] http://www.thesmokinggun.com/jamesfrey/0104061jamesfrey1.html.

plagiarism is not feasible, because a blogger who falsely claims credit for breaking the news would post it later than expected (i.e., at stage 3 instead of stage 2). Finally, we assume that a blogger can not fabricate a link, because readers can easily verify the link's authenticity by clicking on it.

A reader who has not yet seen the news derives utility $u - c$ if she sees a link to a news-breaking blog (B). Note that the value of the information should decline over time. For example, if the information is the name of a store that is having a sale, then readers experience urgency in discovering this information, because the store may sell out of the consumer's size. We can alternatively interpret c as the travel (click) cost readers must pay to travel to blog B. Also, a reader who has already seen the news derives no extra direct utility from seeing a link to blog B. After seeing a link to blog B, the reader updates her belief about B's ability to break the news upward but also updates upward her belief about A's ability to find the news in other blogs. This is the key tension in the model: By linking, blog A looks better but also makes blog B, a potential competitor in the next period, look good.[13]

To simplify the analysis, we assume that the reader does not update her priors about B's ability to find news in other blogs, either because she does not observe B's links (i.e., information from B may be consumed from A's post) or because she visits blog B and observes its links only after the news has become stale and the links have no signaling value. The results of our analysis remain qualitatively the same if we assume that the reader can resolve uncertainty about B's ability to find the news, but the analysis becomes much more cumbersome. We further assume

[13] Here, we do not consider links to blogs with no news since the reader derives no utility from these links. These types of non-strategic links could of course occur in reality, but they are outside the scope of our model. We thank an anonymous reviewer for pointing this out to us.

that the reader does not learn about the abilities of any other blog during this time period, due to time constraints or because information quickly becomes stale in this environment.

At the beginning of the second period (stage 4), R decides which blog to visit next. If A had linked to B at stage 3, then R chooses between A and B. If A hadn't linked to any blog at stage 3, then R chooses between A and a random blog (C). When making this choice, the reader also experiences a reader-blog–specific random shock to her utility. At stage 5, a piece of information that offers the reader utility u gets released to some blogs, which post it. At stage 6, blogs link to news-breaking blogs if they can find them. That is, in the absence of strategic considerations (i.e., the last round of the game), all blogs link if they find news. If R had not seen the news in stage 5, she obtains utility $u - c$ from the news, and the game ends after stage 6.

The assumption that linking takes place in the last round of the game is important, in that signaling in the first round is motivated by reader's desire to learn about the blogger's ability to link in the future. Instead, we could specify an overlapping generations model, in which bloggers have a constant incentive to signal their abilities to new readers. Thus, the assumption about terminal period play represents a simplification that enables us to solve a less complicated model.

3.2 Perfect Bayesian Nash Equilibrium

Because of the symmetry in the model, all blogs have identical incentives, and all readers face the same decision at the beginning of the game. Thus, we can focus on the decision faced by a random reader R and a random blog A, and then generalize the findings to all blogs and all readers. We look for a perfect Bayesian Nash equilibrium with respect to the linking behavior of a blogger during stage 3, at which point the blogger has either broken or failed to break news and has either found a news-breaking blog or failed to do so. By assumption, if a blogger cannot find

the news in other blogs, he is not able to link. So we concentrate on the scenario in which the blogger finds the news in another blog (B), which creates four possible equilibria: (L, L), (L, DL), (DL, L), and (DL, DL), where L stands for **l**ink and DL indicates **d**o not **l**ink, and the cells refer to the actions when the blog has or has not broken news respectively. For example, in (DL, L), the blogger does not link if he has broken the news but does link if he has not broken the news.

Each equilibrium generates a different set of posterior beliefs by R. For example, the absence of a link in (L, L) is a negative signal about the blog's ability to find news, whereas in (DL, DL) it has no negative effect, because linking is not expected. In addition, as is the case for most signaling models, some actions are not in equilibrium, so we must specify off-equilibrium beliefs. Specifically, if linking is observed but is not played in equilibrium, then Bayes's rule does not apply. In that case, we assume that R *always* updates beliefs about the blogger's ability to find information upward when observing a link, because a link occurs only if the blogger finds information in other blogs, which is a positive signal about that ability. Whether the blogger plays an out-of-equilibrium strategy or not should not interfere with that inference. We summarize these posterior beliefs in Table 1.

We start to solve for the equilibrium by analyzing R's problem when she chooses a blog to visit during stage 4. The utility from blog A is the sum of the expected utility based on R's updated beliefs about the blog's abilities and a random i.i.d. shock $\varepsilon_{A,R}$, distributed on the real line with the cumulative distribution function (C.D.F.) F (where density is non-zero everywhere):[14]

[14] There are two reasons to introduce the error term in the model. First, it explains why a blog with a negative outcome for either breaking the news or linking still may attract readers in the next period. Second, the error term allows us to consider how linking affects the difference in the expected utilities between blog A and its primary rival, $EU_A - EU_j$, a continuous incentive, rather than a discrete incentive, as would be the case in a model without noise. Further, the results (unless otherwise noted) are independent of the exact distribution of the error term.

$$U_{A,R} = \alpha_A(a_1)u + [1-\alpha_A(a_1)]\beta_A(a_2)(u-c) + \varepsilon_{A,R} \equiv EU_A(a_1,a_2) + \varepsilon_{A,R}\,, \tag{1}$$

$$\text{where } a_1 = \begin{cases} 1 \text{ if A broke the news} \\ 0 \text{ otherwise} \end{cases},\ a_2 = \begin{cases} 1 \text{ if A linked to B} \\ 0 \text{ otherwise} \end{cases}, \tag{2}$$

α and β are the updated probabilities that the blog will break and find the news respectively, and EU is the expected utility based on the firm's actions (which influence R's beliefs). Then,

$$\begin{aligned} &P(\text{R will choose A over j}) = \\ &P\left[U_{j,R} < U_{A,R}\right] = P\left[e_{j,R} - e_{A,R} < EU_A(a_1,a_2) - EU_j(L^{A\to j})\right] \equiv G\left[EU_A(a_1,a_2) - EU_j(L^{A\to j})\right] \end{aligned} \tag{3}$$

where $L^{A\to j} = \begin{cases} 1 \text{ if A linked to j} \\ 0 \text{ otherwise} \end{cases}$ and G is the C.D.F. of the random variable $\varepsilon_{j,R} - \varepsilon_{A,R}$.

The primary rival is endogenous to the linking decision. If A links to another blog ($j = B$), the news-breaking blog B becomes its primary rival. If A does not link to another blog, its primary rival is a random blog ($j = C$), perceived by the reader as average along both dimensions.

To find a pure strategy perfect Bayesian Nash equilibrium, we derive the blogger's optimal strategy at stage 3, assuming that R's beliefs are consistent with the given equilibrium. Therefore, A links, if doing so increases the probability that it will be chosen over the primary rival[15] or if:

$$G\left[EU_A(a_1, a_2 = 1) - EU_B(L^{A\to B} = 1)\right] > G\left[EU_A(a_1, a_2 = 0) - EU_C(L^{A\to C} = 0)\right] \tag{4}$$

Because $\varepsilon_{A,R}$'s density function is assumed to be non-zero, $(\varepsilon_{j,R} - \varepsilon_{A,R})$'s density is also non-zero on the real line. This, along with the fact that G is a C.D.F., implies that G is a strictly increasing function. Hence, A links if:

$$EU_A(a_1, a_2 = 1) - EU_B(L^{A\to B} = 1) > EU_A(a_1, a_2 = 0) - EU_C(L^{A\to C} = 0)\,. \tag{5}$$

[15] At stage 4, A has three potential sources of readers. First, some readers may choose it randomly. Second, if A had broken news at stage 2 and other blogs had linked to it at stage 3, A may gain these readers. Third, each one of A's stage 3 readers returns to A with probability $G(EU_A - EU_j)$. Note that only the latter probability is affected by A's outgoing links, which is why we focus on this expression.

To simplify this notation, we redefine EU as a function of updated beliefs, such that $EU_A(u,d)$ is the expected utility of blog A when the prior belief about the ability to break news gets updated **u**pward while prior about the ability to find news is updated **d**ownward. We denote the prior belief by **n**eutral. For example, $EU_A(u,n) \equiv EU_A(\alpha_U, \beta_0) = \alpha_U u + (1-\alpha_U)\beta_0(u-c)$.

In (L, L), A chooses to link if it breaks news (see (6) below) and if it did not break news (see (7)):

$$EU_A(u,u) - EU_B(u,n) > EU_A(u,d) - EU_C(n,n) \quad (6)$$

$$EU_A(d,u) - EU_B(u,n) > EU_A(d,d) - EU_C(n,n) \quad (7)$$

In (L, DL), A links only if it is able to break news:

$$EU_A(u,u) - EU_B(u,n) > EU_A(u,d) - EU_C(n,n) \quad (6)$$

$$EU_A(d,u) - EU_B(u,n) \leq EU_A(d,n) - EU_C(n,n) \quad (8)$$

In (DL, L), A links only if it is not able to break news:

$$EU_A(u,u) - EU_B(u,n) \leq EU_A(u,n) - EU_C(n,n) \quad (9)$$

$$EU_A(d,u) - EU_B(u,n) > EU_A(d,d) - EU_C(n,n) \quad (7)$$

In (DL, DL), A never chooses to link:

$$EU_A(u,u) - EU_B(u,n) \leq EU_A(u,n) - EU_C(n,n) \quad (9)$$

$$EU_A(d,u) - EU_B(u,n) \leq EU_A(d,n) - EU_C(n,n) \quad (8)$$

As we discussed previously, the absence of a link prompts a negative perception if linking is expected in equilibrium (see $EU_A(u,d)$ in (6) and $EU_A(d,d)$ in (7)) but a neutral response if linking is not expected (see $EU_A(d,n)$ in (8) and $EU_A(u,n)$ in (9)). Equations (6) – (9) can also be rewritten as the difference between the marginal benefit (increase in own utility) and marginal cost (increase in rival's utility) from linking; for example (6) becomes $[EU_A(u,u) - EU_A(u,d)] - [EU_B(u,n) - EU_C(n,n)] > 0$, where the left hand side can be interpreted as the incentive to link.

Before we characterize the equilibria regions, we show that the symmetric uncertainty assumption is not critical. That is, even if a blogger were to initially observe his own type, he cannot credibly signal it by posting (or withholding) links. A blogger's utility depends on the size of the readership, which is determined by the readers's posterior beliefs about the blogger's ability. Hence, conditional on finding the news in another blog, posting a link is costless, and all types find information in other blogs with non-zero probability. Therefore, if one type benefits from linking, all types equally benefit from linking (and vice versa), which rules out a separating equilibrium based on type.

Lemma 1

No separating equilibrium based on privately observed type is possible; if a blogger of type i, where $i \in \{HH, HL, LH, LL\}$, prefers to link/not link, then all types prefer to link/not link.

Proof: See Appendix

Next, we describe the equilibria in the different regions of the parameter space. In the Appendix, we demonstrate the existence of all the sub-regions under certain parameters.

Proposition 1

The following regions (each with a different set of equilibria) partition the space:

Region I: Equation 6 holds, 8 does not hold, and 9 holds: (DL, L) and (L, L) exist.

Region II: Equation 9 does not hold: (L, L) is unique.

Region III: Equation 6 does not hold, 8 holds, and 7 holds: (DL, DL) and (DL, L) exist.

Region IV: Equation 7 does not hold: (DL, DL) is unique.

Region V: Equation 6 does not hold, and 8 does not hold: (DL, L) is unique.

Region VI: Equation 6 holds, and 8 holds: all four equilibria exist.

Proof: See Appendix.

A blogger who is more likely to break news earns a smaller marginal benefit from being able to direct readers to news in other blogs, because the content on the blogs is substitutable. Therefore, a blogger should have less incentive to link when he has broken news, yet an added difficulty results from the change in beliefs across information states. For example, in (L, DL), the reader's inference following 'no link' is more punishing if the blog *had* broken the news. As Proposition 1 implies, (L, DL), the equilibrium in which the blogger only chooses to link if he has broken the news, only occurs in Region VI. Since this is the only equilibrium that defies the intuition that the incentive to link is greater if the blog had *not* broken the news, it is comforting to see that it is relatively rare and never unique. However, all other equilibria may be unique.

Next, we illustrate the relative placement of the six regions in the (v – w) – (p – q) space (see Figure 3). That is, we fix w and q and vary v and p. By varying these two parameters, we alter the informativeness of the signal associated with either breaking the news or linking. For example, if v and w are very close, there is little difference between the types in terms of the ability to break news, so a link to another blog would not significantly change the prior about the other blog's ability to break news (marginal cost of linking is minimal). However, if $p >> q$, linking greatly improves the prior for own ability to find news in other blogs (marginal benefit of linking is high). The boundaries of the regions are defined by the iso-curves derived from Equations 6–9. For example, we define $6^{=}$ as $EU_A(u,u) - EU_A(u,n) = EU_B(u,d) - EU_C(n,n)$. Proposition 2 summarizes the relative placement of regions.

Proposition 2

$6^{=}$ and $9^{=}$ are increasing in v, $8^{=}$ and $7^{=}$ are either increasing in v or increasing and then decreasing in v. For all $0 \le q < p \le 1, 0 \le w < v \le 1$, $9^{=}$ lies above $6^{=}$, $8^{=}$, and $7^{=}$; $7^{=}$ lies below $6^{=}$, $8^{=}$ and $9^{=}$.

1. (L, L) is unique in the *non-empty* northwest region described by

$$(1-\alpha_U)(\beta_U-\beta_0)(u-c)-(\alpha_U-\alpha_0)[(1-\beta_0)u+\beta_0 c]>0\,.$$

2. (DL, DL) is unique in the *non-empty* southeast region described by

$$(1-\alpha_D)(\beta_U-\beta_D)(u-c)-(\alpha_U-\alpha_0)[(1-\beta_0)u+\beta_0 c]\le 0\,.$$

3. (DL, L) and (L, DL) exist between these two regions.

Proof: See Appendix.

There always exists a unique (L, L) region in which blogs prefer to link, irrespective of whether they break news. As Proposition 2 and Figure 3 show, this occurs when p is high relative to v. In this region, the signal about own ability to find news is relatively more informative than the signal about the rival's news-breaking ability; therefore, the marginal benefit of linking is higher than the marginal cost. In contrast, when v is high relative to p, there is a unique (DL, DL) region. The other two equilibria, (DL, L) and (L, DL), exist at the margin, when v and p are relatively similar. In that case, the outcome depends on the equilibrium beliefs and whether the blog breaks news.

If linking occurs in equilibrium, the blogs which are High type on the ability to break news, are more likely to attract incoming links. Thus, our model provides a micro-foundation for why incoming links may serve as a measure of quality. From a consumer's perspective, linking increases the attractiveness of the blogosphere, because links enable readers to locate information more efficiently. From bloggers' perspective, outgoing links enable them to signal their ability to locate information. The desire to signal generates an incentive for blogs to promote their high-quality rivals, which is a byproduct of selfish behavior and not the result of altruism.

3.3 Comparative Statics

We now examine how the incentive to link in equilibrium changes with the level of heterogeneity across types and with changes in the delay cost c. The differences between this analysis and Proposition 2 are twofold. First, we consider the changes within (as opposed to across) equilibria. Second, we perform a slightly different exercise – we spread the posterior probabilities but keep the prior probabilities (α_0, β_0) fixed (previously, we varied v or p, which changes both the prior and posterior probabilities). This approach enables us to concentrate purely on the change in learning associated with the signal. That is, we compare the incentive to link at (p, q) and $(p' = p + \varpi, q' = q - \frac{\gamma\varpi}{1-\gamma})$,[16] where ϖ is a small change, as well as at (v, w) and $(v' = v + \varpi, w' = w - \frac{\delta\varpi}{1-\delta})$.

Proposition 3

1. The incentive to link is greater at (p', q') than at (p, q).
2. If the blog breaks the news, the incentive to link is lower at (v', w') than at (v, w).
3. If the blog does not break the news,
    - In (L, L) and (DL, L), the incentive to link at (v', w') is lower iff

      $\frac{\alpha_0}{1-\alpha_0} < \frac{(1-\beta_0)u + \beta_0 c}{(\beta_U - \beta_D)(u-c)}$ and higher otherwise.

    - In (DL, DL) and (L, DL), the incentive to link at (v', w') is lower iff

      $\frac{\alpha_0}{1-\alpha_0} < \frac{(1-\beta_0)u + \beta_0 c}{(\beta_U - \beta_0)(u-c)}$ and higher otherwise.
4. The incentive to link is decreasing in c.

Proof: See Appendix

[16] Note that this approach constitutes a mean-preserving spread, because the prior for breaking the news is the same at (v, w) and (v', w'): $\alpha_0' = \gamma p + (1-\gamma)q + \gamma\varpi - (1-\gamma)\frac{\gamma\varpi}{1-\gamma} = \gamma p + (1-\gamma)q = \alpha_0$.

Spreading the posteriors on the ability to find news in other blogs makes both the positive signal (a link) and negative signal (no link) more informative. Because this increases the marginal benefit of linking and has no effect on marginal cost, the overall incentive to link increases. In contrast, the effect of spreading the posteriors on news-breaking ability depends on the model parameters. A more informative signal improves the reader's assessment of the rival's ability and hence increases the marginal cost of linking. The effect on the marginal benefit depends on whether the blog breaks news; if it does, the marginal benefit decreases, because the reader is less likely to rely on a link to obtain news in the next period. However, if the blog has not broken news, the marginal benefit increases, because the reader attaches greater value to the ability to link. Thus, the overall effect depends on the relative importance of these two factors.

Finally, we consider the change in the incentive to link as c (delay cost) increases. Increasing c decreases the utility that the reader expects to receive from information obtained through a link, which decreases the marginal benefit of linking. In addition, increasing c increases the benefit of receiving information directly versus through a link, which makes the linked blog look particularly attractive because of its news-breaking ability, which increases the marginal cost of linking. Hence, an increase in c decreases the incentive to link. An interesting implication of this result, which we test empirically in Section 5, is that linking is *more* likely to occur in categories in which information is *less* urgent.

4. Extensions

Here we extend the main model to allow for correlation in abilities and heterogeneity on the prior beliefs across blogs. The former is necessary because empirically there may be positive correlation in abilities if they are manifestations of the same construct (e.g., expertise), and this

correlation may affect the basic trade-off in the model. The latter extension allows us to explore the extent to which bloggers' incentive to link and readers' benefit from linking are aligned.

4.1 Correlated Abilities

Assume a positive correlation between the two abilities. Specifically, $P(\text{H}-\text{type find news} \mid \text{H}-\text{type break news}) \equiv \rho$,

$P(\text{H}-\text{type find news} \mid \text{L}-\text{type break news}) \equiv 1-\rho$, where $0.5 \le \rho \le 1$ (also see Figure 4).

That is, since $\rho \ge 0.5$, we assume that a blog that is high type on the ability to break news is more likely to be high type on the ability to find news in other blogs. Note that at $\rho = 0.5$, the two abilities are independent; an increase in ρ implies an increase in positive correlation; and at $\rho = 1$, the two abilities are perfectly correlated. We can show that the reverse inference holds as well: a blog that is high type on the ability to find news in other blogs is more likely to be high type on the ability to break news.

We solve this model in Technical Appendix A but discuss the main results here. First, allowing correlation between abilities increases the cost of linking, because readers perceive the rival as not only more likely to break the news, but also more likely to find news in other blogs. The benefit of linking, however, may increase or decrease. On the one hand, a link increases the likelihood that the blogger is high type on news-breaking ability, which increases the incentive to link. On the other hand, a blog that has broken news has a higher posterior on both abilities, which implies that a link may have a lower marginal impact.

At an arbitrary level of positive correlation, the space of equilibria becomes more complicated; instead of six sub-regions, nine become possible, including a region in which (L, DL) is unique. See Proposition 1A in the Technical Appendix A for the complete characterization of the

equilibrium regions. Generally, if the two abilities are significantly correlated, the incentive to link might no longer be lower following a news-breaking story. To illustrate the intuition behind this result, suppose that blogs have a very low prior on the ability to find news. Following a news-breaking event, a blog is judged to be more likely to break news in the future, which implies a lower marginal benefit from a link as before. However, if the two abilities are highly correlated, following a news-breaking story, a blog may enter the intermediate region on the posterior on the ability to find news in other blogs. This in turn implies that another positive signal (a link) would result in a significant jump in this posterior: a higher marginal benefit from a link. Hence, the incentive to link may be higher or lower, depending on the relative strength of these forces.

Consistent with our intuition, if the positive correlation is small, the main results are not qualitatively affected. While the exact equations for the iso-curves differ (see Technical Appendix A), the number of regions that partition the space and the set of equilibria in each sub-region remain the same as in Proposition 1.

Proposition 4

There exists a $\bar{\rho} > 0.5$, such that for all $0.5 < \rho \leq \bar{\rho}$, Regions I–VI partition the space.

Proof: See Technical Appendix A.

4.2 Consumer Learning and Bloggers' Incentives

In the main model, we show that linking speeds up consumer learning compared to a scenario where linking is not in equilibrium. However, there are some limits to this finding. In the model, linking benefits the reader in two ways. First, there is a direct benefit due to information obtained through a link. Second, there is an indirect benefit due to the learning on the linked site's type which hurts the linking blog. Hence, a blogger may link in a way that is not welfare-maximizing for the reader. In particular, we allow for heterogeneity in the prior on the news-

breaking ability (perhaps due to differences in past actions). Specifically, we assume that A can link to (1) a type D blog that has a prior of γ_D on being high type on news-breaking ability or (2) a type E blog that has a prior γ_E. Assuming that A wants to link, which blog would it choose?

To simplify this analysis, we assume that the shock to the deterministic part of the utility ($\varepsilon_{j,R}$) is i.i.d., according to the double exponential distribution, with $\mu = 1$. At t = 4, R chooses among three blogs: blog A (focal blog), a type D blog, and a type E blog. Given the assumption about the distribution of the error term, the resulting choice probabilities are $P(i) = \frac{\exp(EU_i)}{\sum_j \exp(EU_j)}$. The expected surplus for R from a choice set that contains these three alternatives is $S = \ln(\exp(EU_A) + \exp(EU_D) + \exp(EU_E))$ (Anderson et al. 1992).

Proposition 5

Let $F(\gamma_j) = \exp[EU_j(u,n) \mid \gamma_J] - \exp[EU_j(n,n) \mid \gamma_J]$, where j is a blog with prior γ_j.

1. F is initially increasing and then decreasing in γ_j.

2. A blogger prefers to link to a rival k s.t. $F(\gamma_k)$ is minimized, whereas a reader benefits most from a link to a blog m s.t. $F(\gamma_m)$ is maximized: the bloggers' and the readers' incentives are perfectly misaligned.

Proof: See Technical Appendix B.

A link is a signal on the ability of the rival blog, and the signal is most informative when it points to a blog in the medium range of the prior distribution, where the readers' uncertainty is maximized. However, though resolving this uncertainty in favor of the rival blog is costly, linking to a blog at the ends of the distribution does not greatly change R's prior and is thus less costly. Hence, bloggers prefer to link to blogs that are obviously high or low on news-breaking ability.

5. Empirical Evidence

We test two predictions related to strategic linking. First, in Proposition 3, we find that an increase in c decreases the incentive to link. Because we interpret c as the decay rate of information, we predict a lower rate of linking in a category in which information is more urgent. Second, Proposition 5 implies that bloggers tend to link to other blogs with the least uncertainty pertaining to the ability to break news.

These predictions are not consistent with several other salient alternative explanations for the existence of links in the blogosphere. For example, if bloggers linked because of altruism toward readers, they would link to blogs with maximum uncertainty to maximize consumer learning. Alternatively, links could be part of reciprocal giving among blogs (Resnick and Zeckhauser 2002; Cabral and Hortascu 2004; Narayan and Yang 2006), in which case blogs with the greatest incentive to reciprocate are those whose quality is uncertain, which contradicts our prediction. Finally, an alternative strategic reason suggests that bloggers may link to demonstrate their interests, which results in better targeting, or to grow the market for blogs, which would ultimately benefit the blogger.[17] However, none of these alternative explanations focuses on the competitive downside of linking and therefore do not make the same empirical predictions as our model.

To test these predictions, we constructed a data set of blogs in eight categories: politics, health, fashion, food, business, music, sports and movies. We used the popular blog search engine Technorati to collect a list of blogs for sampling. In particular, we used Technorati's blog

[17] We thank an anonymous referee and the Associate Editor for suggesting these alternative explanations.

directory feature, where about 1.5 million blogs are classified based on self-reported tags.[18] For each category, we sorted all blogs according to the recency of their updates and selected the most recently updated 150 blogs at the time of the data collection (April 19, 2007)[19] (see Table D1 in the Technical Appendix for examples of blogs in our sample). We then collected the last 10 posts (as of June 25, 2007) for each selected blog using blog feeds available through the Bloglines aggregator.[20] Because not all feeds provide the full content of the posts and due to technical issues, we obtained complete post data for 602 of the 1,200 blogs, including the date and time of each post and the outgoing links (permalinks) embedded in the text. We refer to these blogs as "parent" blogs, whereas the 2,173 linked blogs[21] are the "children" blogs. Using the same algorithm, we also collected data about the date and time of each post on children blogs during July 11–12, 2007.

We present the summary statistics in Table 2. The variables calculated on the collected sample of posts are inter-post time (INTERPOST), or the average time (in days) between posts; total number of outgoing links (NO_LINKS) to other sites; and average post length (POST_LENGTH), excluding link-related characters (or tags). We argue below that INTERPOST relates to the blogger's ability to break news. Length may be associated with nuance (Chevalier and Mayzlin 2006). Finally, in Table D3 and Figure D1 of the Technical

[18] See http://support.technorati.com/support/siteguide/directory for a more detailed description.
[19] This algorithm skews the sample toward more active blogs. Because this possible bias is constant for all blogs, it can not explain the variance in the propensity to link across categories. Moreover, we expect that blogs that break news more often (which we later argue correlates with activity level) have a lesser incentive to link, which should bias against our hypothesized results.
[20] A blog's content can be syndicated automatically through a feed, which then can be read through a blog aggregator. For a subset of blogs, the feed presents all post content (including outgoing links). A feed is much easier to process technically, because the data are more structured that original blog content.
[21] The link data obtained from the posts had to be cleaned because they contained links to tags, navigation links, links to self, repeated observations, and so forth. A linked site in our sample also must be available for subscription on Bloglines, which excludes non-blog links such as tags but may not exclude a link to a Web site. We cleaned the data further before comparing the parent and child blogs.

Appendix, we present the distribution of NO_LINKS, which follows a power law (Herring et al. 2005; Marlow 2005): A large percentage of blogs (38%) do not link at all, and a few blogs link frequently. We thus find heterogeneity in linking across blogs.

We further cleaned the parent sample to address concerns about Technorati's classification.[22] Specifically, we recruited three independent raters to visit each site and decide whether it was a blog and whether it belonged in the category to which it had been tagged by Technorati. In case of conflicts, we chose the majority rating. Of the original list of 1,200 parent blogs, the raters agreed with the Technorati categorization of 442 blogs, which leaves a set of 262 blogs for which we have reliable category information and post information. We also excluded 4 sites rated as 'not blogs'.

Finally, we conducted an online survey with 421 participants to obtain a demand-based measure of readers' perceptions of the urgency of the information in each category (see Technical Appendix E). In the survey, subjects were asked if they were interested in receiving a free online newsletter in each category. If they indicated interest, they chose between an aggregated weekly version and a daily version. We calculate the urgency score as the fraction of the participants who prefer the daily version, such that Urgency Index = Number Daily/(Number Daily + Number Weekly) (see Table 3). A higher score implies a more urgent category. We explicitly calibrate this measure for a subsample of respondents who express an interest in the category to replicate the self-selection that takes place online. For example, readers who visit political blogs probably perceive political information as very urgent.

[22] Technorati derives its classification from blogger's self-reports and others' reports (social tagging). Unfortunately, some bloggers blog on a variety of topics, which makes classification into a single category problematic.

The summary data in Table 3 give some indication of a relationship between urgency and linking in the hypothesized direction. The most urgent category, sports, contains the lowest fraction of blogs with links and the lowest average number of links. Movies, the least urgent category, reveals the highest average number of links, though politics provides an outlier—despite its urgency, it has a high proportion of blogs that link and a high average number of links, as well as the lowest inter-post time and the second highest length average. This category thus appears very active. Table 4 presents the results of an estimation of the relationship between the propensity to link and category urgency on the blog level. In Model 1, the dependent variable equals 1 if we observe a link in our sample and is 0 otherwise. Thus, we estimate a probit model. The main coefficient of interest, category urgency, is significant and negative, as theorized. We also find a negative relationship between INTERPOST and linking; blogs that post often are more likely to link, which implies a positive correlation between the ability to post information and the ability to link to other sites. Finally, we find a positive relationship between the length of the post and linking.

These basic results are robust against different specifications. In Model 2 in Table 4, the dependent variable is the total number of links, whereas in Model 3, it is the natural log of the number of links, to address the potential scale effects due to differences in blog size. In Model 4, we estimate a negative binomial model, because our dependent variable consists of integers and a count model may be more appropriate. In Model 5, we estimate a model with fewer functional form restrictions on urgency. That is, since our survey reveals "Sports" to be by far the most urgent category, we test whether we observe different linking behavior for sports blogs. We find that the "Sports" dummy is negative and significant. Finally, in Model 6, we control for the average INTERPOST for that category. Categories with higher average INTERPOST have a

lower supply of posts to which a blogger can link, and hence we expect to see fewer links in these categories. We do indeed find that the sign on Category INTERPOST is negative. Despite this added control, the effect of urgency remains negative and significant. In summary and consistent with our theory, *less* linking occurs in more urgent categories.

Next, we test whether bloggers tend to link to other blogs with the least uncertainty on the news-breaking ability. Theoretically, "uncertainty" is expressed as the prior belief and is minimized at the extremes. The operationalization of this construct requires certain assumptions. We assume that bloggers who are high type on ability to break news are more likely to have access to news and to post. In Technical Appendix C we present a simple model that shows that this assumption implies that bloggers who post very often (INTERPOST is low) are almost certainly high type on the ability to break news, whereas those who post very infrequently (INTERPOST is high) are almost certainly low type on this ability. Hence, the uncertainty is maximized for bloggers in the middle range and, according to our hypothesis, bloggers should be less likely to link to them.[23]

As with the parent sample, we performed some further data-cleaning on the child sample. In particular, we excluded the sites of professional news organizations. Two independent raters classified a random subsample (N = 438) of child sites into either blogs or news organization sites (the other two categories were "not sure" and "other"). The raters achieved a reasonably high agreement of 78%, and classified 72% of the 438 sites as blogs (see Table D2 in the Technical Appendix).

[23] Agarwal et al. 2008 find that active blogs are more likely to produce influential posts (or posts that generate discussion and a large number of incoming links) than inactive blogs. Hence, to the extent that influential posts are more likely to be news-breaking, there is some empirical evidence for the correlation between blog activity and ability to break news.

To demonstrate the tendency to link to blogs with either very low or very high INTERPOST, we compare the INTERPOST distribution of the 338 child sites with that of the 566 parent sites, such that both samples include only those sites classified as blogs by the raters. In Table 5, we place all blogs into three categories (low, middle, and high) on the basis of their INTERPOST. The value cutoffs are derived from the parent distribution, such that the first set of values separates the parent blogs into the top 5%, bottom 5%, and middle 90%. As Table 5 demonstrates, for the cut-offs ranging from 5% to 20%, there is a relatively higher percentage of child blogs in the top and the bottom bins. For example, 13% of child blogs have INTERPOST $\geq$ 6.53 compared with 10% of parent blogs, and 18% of child blogs have INTERPOST $\leq$ .227 compared with 10% of parent blogs. The results are not simply because the child blogs have shorter INTERPOSTs than parent blogs; we find evidence of linking to other blogs with very long INTERPOSTs, which suggests an effect beyond selection of blogs that post often. For the different cutoffs, the chi-squared of equivalence of the distributions gets rejected at $p < .05$. In the Technical Appendix (Figure D2), we illustrate the same point by comparing the cumulative density functions of the two samples, which does not require the selection of cutoffs. Note that there are a few caveats to this analysis. First, INTERPOST is not a perfect measure of news-breaking ability. Second, some blogs may have a higher awareness level than others, which is not currently modeled. However, we consider this analysis a necessary first step and leave the exploration of better metrics to future research.

In summary, less linking exists in more urgent categories, and linked blogs are more likely to exhibit very low or very high INTERPOST. Both of these findings indicate that bloggers' behavior is driven by strategic considerations.

6. Conclusion

In the blogosphere, bloggers are providers and readers are consumers of information. In such an expansive space, without quality controls, it remains unclear how readers find blogs that provide high-quality information. We suggest that the selfish desire of bloggers to link plays a key role in consumer learning about blog quality. While links are referrals to competing sites, recommending rivals may actually be profitable in certain circumstances. These individual incentives result in a system in which high-quality blogs gain prominence, which enhances consumer learning.

Several interesting marketing implications result from our findings. First, by revealing that links enable consumer learning, we hope to dispel some criticism surrounding blogs as a potential advertising medium. Second, we find that both the number of incoming and outgoing links may serve as a metric of blog quality. Third, we point out the trade-offs inherent to recommending a rival. Moreover, we believe that our analysis can apply to other instances in which a firm might provide referrals to potential competitors in exchange for gaining a reputation as a destination site. For example, Tweeter offered an automatic price protection that sent checks to customers if the retailer found a better deal elsewhere (see Gourville and Wu 1996). In this scenario, referral ability serves as a marketing tactic for the firm.

7. Limitations and Future Research

We note several limitations to this study that offer opportunities for further research. First, we assume that the abilities of bloggers are exogenously determined. Instead we could allow bloggers to invest in both abilities and thus determine whether bloggers prefer to invest in both or specialize. Moreover, we do not model other moderators of quality, such as choice of topic,

personality and style of blogging, or offline brand equity.[24] Another potentially interesting area pertains to the dynamics of information diffusion in blogs.[25] In our model, we largely abstract away from these issues by assuming a simple information dissemination structure. A richer model that allows for news to be released stochastically or multiple pieces of information to be released simultaneously could include a wider range of phenomena, such as herding or word of mouth. Blogs may be an especially interesting context for studying these effects, because they share characteristics with both newspapers and social media. Finally, our model assumes that bloggers truthfully credit their sources, and some mutual monitoring may enforce this norm (Varian 1990; Knez and Simester 2001). In particular, we believe that a blogger who creates original content has an incentive to monitor credit.[26] Formally modeling such incentives might explain why users invest in creating content, even in the absence of legal copyright protection.

Various opportunities for future empirical work also exist in this area. An analysis of the determinants of linking behavior in the blogosphere might provide general insights into network formation, such as the dynamics of linking. While our current analysis may be limited in its goals and methods, we hope that it provides a first step toward a more sophisticated empirical model of linking.

[24] We thank an anonymous reviewer for this suggestion.
[25] We thank an anonymous reviewer for this suggestion.
[26] Bloggers often track news they have broken, which sets a higher bar for plagiarism, because the news-breaking blogger can take steps to discredit the plagiarizer. For example, consider the post "McCain 'Family Recipes' Lifted from the Food Network" (April 14, 2008) on the Huffington Post, a political blog. An addendum to story read, "UPDATE: The Cindy McCain 'Family Recipe' plagiarism story, broken here, has been picked up by Politico. Wonkette and Gawker also covered the Cindy McCain recipe story. Even the L.A. Times mentioned the McCain family recipes."

Figure 1: An example of a blog

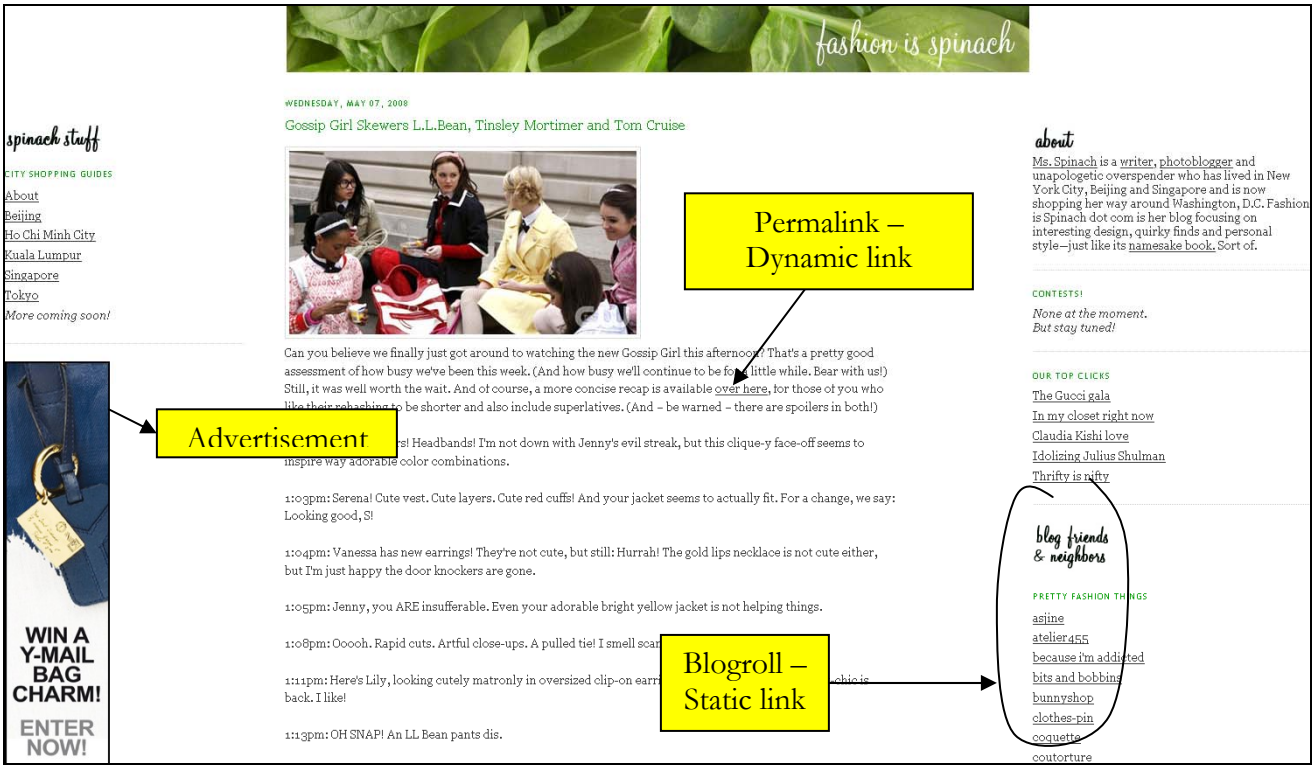

Figure 2: Time Line

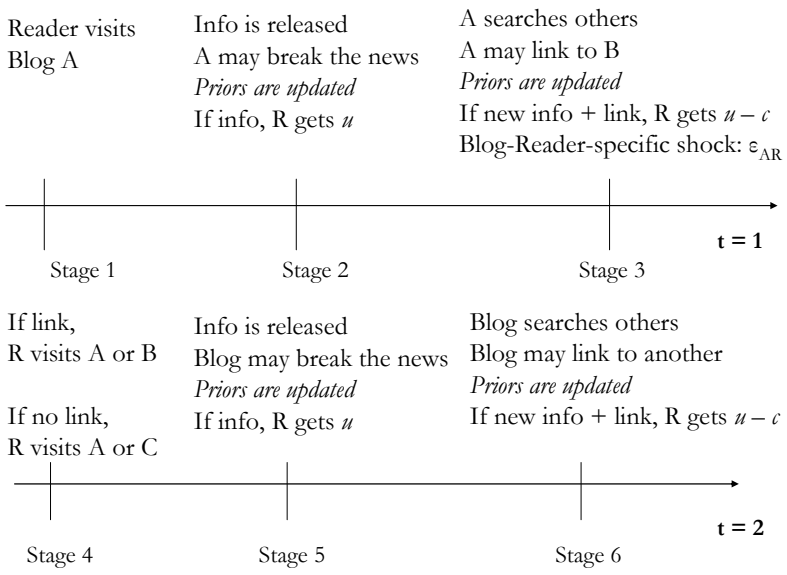

Figure 3: Equilibria

$(\gamma = \delta = 0.5,\ q = 0.2,\ w = 0.1,\ u = 10\ , c = 1)$

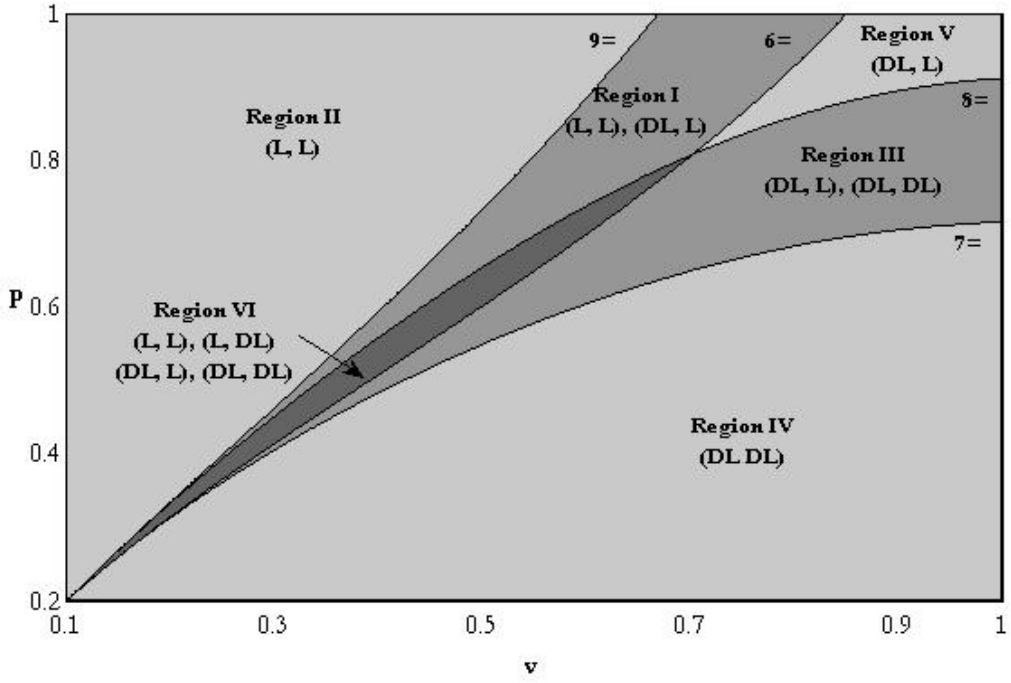

Figure 4: Positive Correlation between Abilities

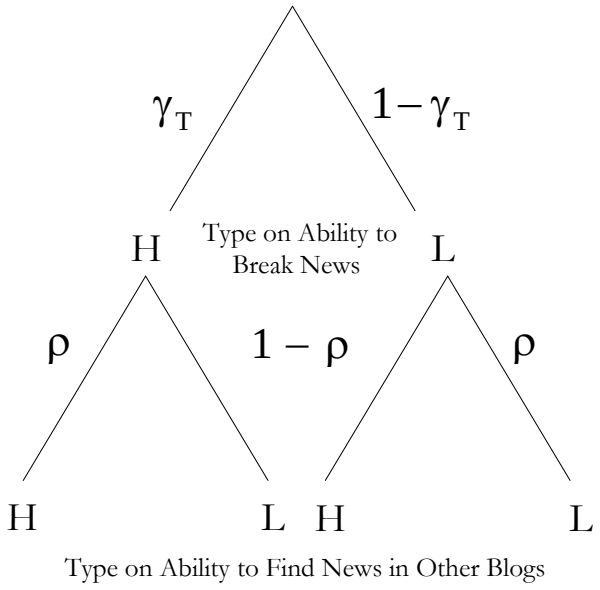

Table 1. Posteriors

Cases	α_A	β_A	α_{OTHER}	β_{OTHER}
A breaks the news A links to B	↑	↑	↑	Prior
A breaks the news A doesn't link	↑	↓ if (L,L) or (L,DL) Prior otherwise	Prior	Prior
A doesn't break the news A links to B	↓	↑	↑	Prior
A doesn't break the news A doesn't link	↓	↓ if (L,L) or (DL,L) Prior otherwise	Prior	Prior

Notes: α_A refers to the probability that A will break the news during the next period, and β_A refers to the probability that A will link to another blog during the next period.

Table 2. Summary Statistics: Parent and Child Blogs

Parent Blogs

Variable Name	Description	No. obs	Min	Mean	Median	Max	Std Dev
INTERPOST	Avg interpost time	602	0	2.86	1.60	29.58	3.81
NO_LINKS	No. outgoing links	602	0	3.61	1	221.00	10.65
POST_LENGTH	Avg post length (chars)	602	0	1299.28	1041.40	10876.80	1127.17

Child Blogs

Variable Name	Description	No. obs	Min	Mean	Median	Max	Std Dev
INTERPOST	Avg interpost time	1848	0	4.37	1.10	290.09	38.70
POST_LENGTH	Avg post length (chars)	1134	0	1596.21	1180.50	13784.00	1504.68

Table 3. Means by Category

Category	N	URGENCY	FRACTION LINK	NO_LINKS	INTERPOST	POST_LENGTH
Sports	38	0.343	0.38	1.18	3.02	1688
Politics	46	0.224	0.87	5.57	1.98	1717
Food	27	0.223	0.57	1.79	3.72	1315
Health	28	0.219	0.57	3.27	3.43	1452
Business	40	0.178	0.68	3.98	3.05	1808
Music	18	0.155	0.67	3.67	2.21	890
Fashion	38	0.138	0.50	1.34	3.01	702
Movies	19	0.118	0.58	5.74	2.11	938

Table 4. Relationship between Linking and Category Urgency

	(1)	(2)	(3)	(4)	(5)	(6)
Model	Probit	OLS	OLS	Neg Binomial	OLS	OLS
Dependent Variable	Link? (0/1)	No. links	Ln(No. links)	No. links	No. links	No. links
Independent Variables						
URGENCY	-2.86 **	-13.67 **	-2.11 ***	-5.97 ***		-10.69 **
	-2.22	-2.40	-2.80	-2.95		-2.08
INTERPOST	-0.06 **	-0.24 ***	-0.05 ***	-0.11 ***	-0.24 ***	-0.20 ***
	-2.02	-3.48	-3.78	-3.23	-3.44	-2.93
POST_LENGTH	0.0003 ***	0.0014 ***	0.0003 ***	0.0004 ***	0.0013 ***	0.0014 ***
	3.27	3.38	4.97	4.35	3.18	3.35
Sports dummy					-2.87 ***	
					-4.58	
Category INTERPOST						-1.69 **
						-2.36
N	254	254	254	254	254	254
R-squared	0.06	0.08	0.14		0.08	0.10
Log-likelihood	-159.01			-543.82		

Notes: (1) t and z-statistics are presented below the coefficient estimates
(2) all models are estimated with robust standard errors
(3) ** : $p < 0.05$
(4) *** : $p < 0.01$

Table 5. Parent and Child INTERPOST Distributions

	Parent Distribution		Child Distribution	
	No. Blogs	% Blogs	No. Blogs	% Blogs
INTERPOST <= .087	29	5%	29	9%
.087< INTERPOST < 9.69	509	90%	284	84%
INTERPOST >= 9.69	28	5%	25	7%
Pearson chi2 (k = 2)	6.95 (p = 0.031)			
INTERPOST <= .227	57	10%	60	18%
.227< INTERPOST < 6.53	453	80%	233	69%
INTERPOST >= 6.53	56	10%	45	13%
Pearson chi2 (k = 2)	15.30 (p = 0.000)			
INTERPOST <= .52	114	20%	102	30%
.52 < INTERPOST < 4.67	339	60%	172	51%
INTERPOST >= 4.67	113	20%	64	19%
Pearson chi2 (k = 2)	12.07 (p = 0.002)			

References

Agarwal, N., Liu, H., Tang, L., and P. Yu (2008), "Identifying the Influential Bloggers in a Community," *Proceedings of the First ACM International Conference on Web Search and Data Mining.*

Anderson, S., de Palma, A., and J-F Thisse (1992) *Discrete Choice Theory of Product Differentiation*. Cambridge, MA: MIT Press.

Bala, V. and S. Goyal (2000) "A Noncooperative Model of Network Formation," *Econometrica,* Vol. 68, No. 5, 1181-1229.

Barbaro (March 7, 2006), "Wal-Mart Enlists Bloggers in P.R. Campaign," *New York Times*.

Brodeur (2008) "Journalism, Social Media and the Blogosphere" http://www.brodeur.com)

Brin, S. and L. Page (1998) "The Anatomy of a Large-Scale Hypertextual Web Search Engine," *Computer Networks and ISDN Systems*, Vol. 30, Issues 1-7, 107-117.

Cabral, L. and A. Hotracsu (2004) "Dynamics of Seller Reputation: Theory and Evidence from eBay," Working paper, Stern Business School, New York University.

Chen, Y., Iyer, G., and V. Padmanabhan (2002) "Referral Infomediaries," *Marketing Science*, Vol. 21, No. 4, 412-434.

Chevalier, J., and D. Mayzlin (2006) "The Effect of Word of Mouth on Sales: Online Book Reviews," Journal of Marketing Research, Vol. 43, 345 – 54.

Demange, G. and M. Wooders (2005) *Group Formation in Economics*. Cambridge: Cambridge University Press.

Dodes, (Sept 12, 2006), "Bloggers Get Under the Tent," *Wall Street Journal.*

Fortunato, S., Flammini, A., Menczer, F., and A. Vespignani (2006) "Topical Interests and the Mitigation of Search Engine Bias," *Proceedings of the National Academy of Sciences of the United States of America*,Vol. 103, No. 34, 12684-12689.

Garicano, L. and T. Santos (2004) "Referrals," *American Economic Review*, Vol. 94, No. 3, 499-525.

Herring, S., Kouper, I., Paolillo, J., Scheidt, L., Tyworth, M., Welsch, P., Wright, E. and N. Yu (2005), "Conversations in the Blogosphere: An Analysis 'From the Bottom Up," *Proceedings of the 38th Hawaii International Conference on System Sciences*.

Hindman, M, Tsioutsiouliklis, K. and J. Johnson (2003) "'Googlearchy': How a Few Heavily-Linked Sites Dominate Politics on the Web," Working paper, Princeton Computer Science.

Katona, Z. and Sarvary, M. (2008) "Network Formation and the Structure of the Commercial World Wide Web," *Marketing Science*, forthcoming.

Kleinberg, J. (1999) "Authoritative Sources in a Hyperlinked Environment," *Journal of the ACM*, Vol. 46, No. 5, 604-632.

Knez, M. and D. Simester (2001), "Firm-Wide Incentives and Mutual Monitoring at Continental Airlines," *Journal of Labor Economics*, Vol. 19, No. 4, 743–772.

Lenhart, A. and S. Fox (2006) "Bloggers: A Portrait of the Internet's New Storytellers," *Pew Internet and American Life Project*, http://www.pewinternet.org.

Lenhart, A., Madden, M., Macgill, A. and A. Smith (2007) "Teens and Social Media," *Pew Internet and American Life Project*, http://www.pewinternet.org.

Marlow, C (2005). "The Structural Determinants of Media Contagion," M.S Dissertation, MIT.

Narayan, V. and S. Yang (2006) "Trust Between Consumers in Online Communities: Modeling the Formation of Dyadic Relationships," Working paper, Stern School of Business, New York University.

Potier, (April 17, 2003), "Harvard University fellow Dave Winer wants to get Harvard blogging" *Harvard University Gazette*.

Resnick, P. and R. Zeckhauser (2002). "Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System," in *The Economics of the Internet and E-Commerce,* Michael R. Baye, editor. Volume 11. Amsterdam, Elsevier Science.

Rosen, (Dec 19, 2004), "Your Blog or Mine?" *New York Times Magazine*

Surowiecki, J. (2004) *The Wisdom of Crowds: Why the Many Are Smarter Than The Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations.* New York, Doubleday.

Tedeschi (July 31, 2006), "Newspapers to Use Links to Rivals on Web Sites,", *NYTimes E-Commerce Report*.

Varian, H. R., (1990) "Monitoring Agents with Other Agents," *Journal of Institutional and Theoretical Economics*, 146, 53-174.

Walker, (April 14, 2003), "Blogging for Milk" *Slate.com*, http://www.slate.com/id/2081419.

Zeleny, (August 5, 2007), "Democratic Candidates Spar at 'Netroots' Forum," *New York Times*.

Appendix

Proof of Lemma 1

Suppose that there exists a (semi) separating equilibrium where types {i … k} link if they find the news and types {k+1 … m} do not link if they find the news. In such an equilibrium, let $\{\alpha_{Link}, \beta_{Link}\}$ be the posteriors following a link and $\{\alpha_{NoLink}, \beta_{NoLink}\}$ be the posteriors following no link. Since types {i … k} link, $V_i = EU_i(\alpha_{Link}, \beta_{Link}) - EU_B(.) > EU_i(\alpha_{NoLink}, \beta_{NoLink}) - EU_C(.)$, where B is a linked blog (that was able to break the news), and C is a random blog. Since V_i is independent of i's type and is a function of posteriors only, the trade-offs remain the same for all types. This implies that types {k+1 ... m} would prefer to link if they find the news in other blogs, which happens with non-zero probability. This contradicts the assumption that {k+1 ... m} do not link in equilibrium. Note that allowing the blogger to strategically post or withhold breaking news also would not enable a separating equilibrium since the same logic would hold, albeit with more complicated posteriors, such as $\{\alpha_{News,Link}, \beta_{News,Link}\}$, etc.

Proof of Proposition 1

We first re-write (6) – (9) by substituting the expressions for expected utilities and simplifying:

1) (L, L): $(1-\alpha_U)(\beta_U - \beta_D)(u-c) > (\alpha_U - \alpha_0)[(1-\beta_0)u + \beta_0 c]$ (6)

$(1-\alpha_D)(\beta_U - \beta_D)(u-c) > (\alpha_U - \alpha_0)[(1-\beta_0)u + \beta_0 c]$ (7)

2) (L, DL): $(1-\alpha_U)(\beta_U - \beta_D)(u-c) > (\alpha_U - \alpha_0)[(1-\beta_0)u + \beta_0 c]$ (6)

$(1-\alpha_D)(\beta_U - \beta_0)(u-c) \le (\alpha_U - \alpha_0)[(1-\beta_0)u + \beta_0 c]$ (8)

3) (DL, L): $(1-\alpha_U)(\beta_U - \beta_0)(u-c) \le (\alpha_U - \alpha_0)[(1-\beta_0)u + \beta_0 c]$ (9)

$(1-\alpha_D)(\beta_U - \beta_D)(u-c) > (\alpha_U - \alpha_0)[(1-\beta_0)u + \beta_0 c]$ (7)

4) (DL, DL): $(1-\alpha_U)(\beta_U - \beta_0)(u-c) \le (\alpha_U - \alpha_0)[(1-\beta_0)u + \beta_0 c]$ (9)

$(1-\alpha_D)(\beta_U - \beta_0)(u-c) \le (\alpha_U - \alpha_0)[(1-\beta_0)u + \beta_0 c]$ (8)

$$\alpha_0 = \gamma v + (1-\gamma)w,\ \alpha_U = \frac{\gamma v^2 + (1-\gamma)w^2}{\gamma v + (1-\gamma)w},\ \alpha_D = \frac{\gamma v(1-v) + (1-\gamma)w(1-w)}{\gamma(1-v) + (1-\gamma)(1-w)},$$

$$\beta_0 = \delta p + (1-\delta)q,\ \beta_U = \frac{\delta p^2 + (1-\delta)q^2}{\delta p + (1-\delta)q},\ \beta_D = \frac{\delta p(1-p) + (1-\delta)q(1-q)}{\delta(1-p) + (1-\delta)(1-q)}.$$

We can show that $\alpha_U > \alpha_0 > \alpha_D$ and $\beta_U > \beta_0 > \beta_D$.

Since there are 4 different constraints, theoretically there could be up to 16 different regions. However, these constraints are not independent. We can show that (6) → (7), (6) does not hold → (9), (8) → (9), (8) does not hold → (7). Below we provide a partition of the space:

a) Assume *(6) holds and (8) doesn't hold.* If (8) doesn't hold → (DL, DL) and (L, DL) don't exist. (6) → (7). If also (9) → (L,L) and (DL,L) exist. If (9) doesn't hold → only (L,L) exist.

b) *(6) doesn't and (8) holds.* (6) doesn't hold → (L,L) and (L,DL) don't exist. (6) doesn't hold → (9). If (7) → (DL, DL) and (DL,L) exist. If (7) doesn't hold → only (DL,DL) exists.

c) *(6) holds and (8) holds*. (6) → (7). (8) → (9). Here all four equilibria exist.

d) *(6) doesn't hold and (8) doesn't hold.* (6) doesn't hold rules out (L,L) and (L, DL). (8) doesn't hold rules out (DL,DL). (6) doesn't hold → (9). (8) doesn't hold → (7). Only (DL,L) exists.

Below we provide a numerical example (let $\gamma = \delta = 0.5,\ q = 0.2,\ w = 0.1,\ u = 10, c = 1$)

Region I: (6) holds, (8) doesn't hold, (9) holds: **(DL, L)** & **(L,L)**. Example: p=0.9, v=0.7.

Region II: (6) holds, (8) doesn't hold, (9) doesn't hold: **(L,L).** Also note that (9) doesn't hold → (6) and (8) doesn't hold: it is sufficient to describe the region by stating that (9) doesn't hold. E.g.: c p=0.8, v=0.2

Region III: (6) doesn't hold, (8) holds, (7) holds: **(DL,DL)** & **(DL, L)**. E.g.: p=0.8, v=0.9.

Region IV: (6) doesn't hold, (8) holds, (7) doesn't: **(DL,DL).** Note that (7) doesn't hold → (6) doesn't hold and (8) holds, so it is sufficient to describe the region by stating that (7) doesn't hold. E.g.: p=0.3, v=0.4.

Region V: (6) doesn't hold, (8) doesn't hold: **(DL,L)**. E.g.: p=0.95, v=0.9.

Region VI: (6) holds, (8) holds: all four equilibria exist. E.g.: p=0.64, v=0.5.

Proof of Proposition 2

$6^{=}$, $9^{=}$, $8^{=}$ and $7^{=}$ are the iso-curves of the inequalities (6) – (9) in the $v-p$ plane. That is,

$$f_6 = (1-\alpha_U)(\beta_U-\beta_D)(u-c)-(\alpha_U-\alpha_0)[(1-\beta_0)u+\beta_0 c]$$

$$f_7 = (1-\alpha_D)(\beta_U-\beta_D)(u-c)-(\alpha_U-\alpha_0)[(1-\beta_0)u+\beta_0 c]$$

$$f_8 = (1-\alpha_D)(\beta_U-\beta_0)(u-c)-(\alpha_U-\alpha_0)[(1-\beta_0)u+\beta_0 c]$$

$$f_9 = (1-\alpha_U)(\beta_U-\beta_0)(u-c)-(\alpha_U-\alpha_0)[(1-\beta_0)u+\beta_0 c]$$

$6^{=}$ is defined as $f_6 = 0$; $7^{=}$ is defined as $f_7 = 0$, and so on.

The derivatives of the iso-curve are obtained using the implicit function theorem and are:

$$\left.\frac{dp}{dv}\right|_6 = -\frac{\frac{df_6}{dv}}{\frac{df_6}{dp}} = \frac{(\beta_U-\beta_D)(u-c)\frac{d\alpha_U}{dv}+[u-\beta_0(u-c)]\frac{d(\alpha_U-\alpha_0)}{dv}}{(1-\alpha_U)(u-c)\frac{d(\beta_U-\beta_D)}{dp}+(\alpha_U-\alpha_0)(u-c)\frac{d\beta_0}{dp}}$$

$$\left.\frac{dp}{dv}\right|_7 = -\frac{\frac{df_7}{dv}}{\frac{df_7}{dp}} = \frac{(\beta_U-\beta_D)(u-c)\frac{d\alpha_D}{dv}+[u-\beta_0(u-c)]\frac{d(\alpha_U-\alpha_0)}{dv}}{(1-\alpha_D)(u-c)\frac{d(\beta_U-\beta_D)}{dp}+(\alpha_U-\alpha_0)(u-c)\frac{d\beta_0}{dp}}$$

$$\left.\frac{dp}{dv}\right|_8 = -\frac{\frac{df_8}{dv}}{\frac{df_8}{dp}} = \frac{(\beta_U-\beta_0)(u-c)\frac{d\alpha_D}{dv}+[u-\beta_0(u-c)]\frac{d(\alpha_U-\alpha_0)}{dv}}{(1-\alpha_D)(u-c)\frac{d(\beta_U-\beta_0)}{dp}+(\alpha_U-\alpha_0)(u-c)\frac{d\beta_0}{dp}}$$

$$\left.\frac{dp}{dv}\right|_9 = -\frac{\frac{df_9}{dv}}{\frac{df_9}{dp}} = \frac{(\beta_U-\beta_0)(u-c)\frac{d\alpha_U}{dv}+[u-\beta_0(u-c)]\frac{d(\alpha_U-\alpha_0)}{dv}}{(1-\alpha_U)(u-c)\frac{d(\beta_U-\beta_0)}{dp}+(\alpha_U-\alpha_0)(u-c)\frac{d\beta_0}{dp}}$$

where all the following terms are positive:

$$\frac{d\alpha_U}{dv} = \frac{\gamma[\gamma(v-w)^2+(2v-w)w]}{[\gamma v+(1-\gamma)w]^2} > 0,\ \frac{d(\alpha_U-\alpha_0)}{dv} = \frac{\gamma(1-\gamma)(v-w)[\gamma(v-w)+2w]}{[\gamma v+(1-\gamma)w]^2} > 0$$

$$\frac{d(\beta_U - \beta_0)}{dp} = \frac{\delta(1-\delta)(p-q)[\delta(p-q)+2q]}{[\delta p+(1-\delta)q]^2} > 0$$

$$\frac{d\beta_0}{dp} = \delta > 0, \quad \frac{d(\beta_U - \beta_D)}{dp} = \frac{\delta(1-\delta)(p-q)[2(1-q)q+\delta(p-q)(1-2q)]}{[\delta p+(1-\delta)q]^2[\delta(1-p)+(1-\delta)(1-q)]^2}.$$

Note that $\frac{d(\beta_U - \beta_D)}{dp}$ is positive if $q \le 0.5$. If we assume that $q > 0.5$, we can show that

$2(1-q)q+\delta(p-q)(1-2q) > 2(1-q)q+(1-q)(1-2q) = 1-q > 0$. Hence, $\frac{d(\beta_U - \beta_D)}{dp} > 0$.

The only possible negative term is, $\frac{d\alpha_D}{dv} = \frac{\gamma\{1+\gamma v^2-(1-\gamma)w^2-2v[1-(1-\gamma)w]\}}{[\gamma(1-v)+(1-\gamma)(1-w)]^2}$.

$\frac{d\alpha_D}{dv}$ is negative when the numerator is negative. Solving for the roots of the numerator,

$$v_1 = \frac{1-w+\gamma w-\sqrt{1-\gamma-2w+2\gamma w+w^2-\gamma w^2}}{\gamma} \quad v_2 = \frac{1-w+\gamma w+\sqrt{1-\gamma-2w+2\gamma w+w^2-\gamma w^2}}{\gamma}$$

The first root lies between 0 and 1. The second root is also always positive but lies above 1, and hence is excluded from the analysis. Hence, $\frac{d\alpha_D}{dv}$ is positive for $v < v_1$ and is negative after that.

1. $\left.\frac{dp}{dv}\right|_6$ and $\left.\frac{dp}{dv}\right|_9$ are always positive. Hence $6^=$ and $9^=$ are increasing in v.

2. In case of $\left.\frac{dp}{dv}\right|_7$ and $\left.\frac{dp}{dv}\right|_8$, the denominators are always positive, but the numerator consists of a term with the multiplier $\frac{d\alpha_D}{dv}$, which could be negative for large values of v. The other term in the numerator is $[u-\beta_0(u-c)]\frac{d(\alpha_U - \alpha_0)}{dv}$, which is always positive. Neither of

these terms dominates the other. So, $\left.\frac{dp}{dv}\right|_7$ and $\left.\frac{dp}{dv}\right|_8$ are always increasing for small values of v (till v_1) and may decrease after that.

Relative Positions of the Iso-curves

We can show that the left-hand-sides of $6^=$, $7^=$, $8^=$, and $9^=$ are increasing in p. Thus, if $f(p,v) > g(p,v) \rightarrow g(p,v) = 0$ lies above $f(p,v) = 0$ in the $v - p$ plane ($f(p,v) = g(p',v) = 0$ where $p' > p$). We can also show that the LHS of $6^=$ is always greater than the LHS of $9^= \rightarrow 9^=$ lies above $6^=$. Similarly, LHS of $6^=$ < LHS of $7^=$, which implies that $6^=$ lies above $7^=$. Finally, LHS of $7^= > 8^= \rightarrow 8^=$ always lies above $7^=$. In summary,

1) (L, L) exists above $6^=$ and (DL, L) exists between $9^=$ and $7^=$. Since, $9^=$ lies above $6^=$, both (DL, L) and (L, L) exist between $9^=$ and $6^=$, whereas (L, L) uniquely exists above $9^=$. Since $9^=$ goes through the origin (v = w, p = q) and has a positive slope, the region where (L, L) is unique is non-empty.
2) (DL, DL) exists below $8^=$ and (DL, L) exists between $9^=$ and $7^=$. Since $7^=$ lies below $8^=$, both (DL, L) and (L, L) exist between $8^=$ and $7^=$, whereas (DL, DL) uniquely exists below $7^=$. Since $7^=$ goes through the origin (v = w, p = q) and has a positive slope for small v, the region where (DL,DL) is unique is non-empty.
3) (DL, L) exists between unique regions of (L, L) and (DL, DL), i.e. in between $9^=$ and $7^=$.

Proof of Proposition 3

i) First, we define all the priors, posteriors under both (v', w') and (v, w) as follows:

$$\alpha_0 = \alpha_0' = \gamma v + (1-\gamma)w,\ \alpha_U = \frac{\gamma v^2 + (1-\gamma)w^2}{\gamma v + (1-\gamma)w},\ \alpha_U' = \frac{\gamma(v+\varpi)^2 + \gamma(w - \frac{\gamma\varpi}{1-\gamma})^2}{\gamma v + (1-\gamma)w}$$

$$\alpha_D = \frac{\gamma v(1-v)+(1-\gamma)w(1-w)}{1-\gamma v-(1-\gamma)w},\ \alpha_D' = \frac{\gamma(v+\varpi)(1-v-\varpi)+(1-\gamma)(w-\frac{\gamma\varpi}{1-\gamma})(1-w+\frac{\gamma\varpi}{1-\gamma})}{1-\gamma v-(1-\gamma)w}$$

$$\beta_0 = \beta_0' = \delta p+(1-\delta)q,\ \beta_U = \frac{\delta p^2+(1-\delta)q^2}{\delta p+(1-\delta)q},\ \beta_U' = \frac{\delta(p+\varpi)^2+(1-\delta)(q-\frac{\delta\varpi}{1-\delta})^2}{\delta p+(1-\delta)q}$$

$$\beta_D = \frac{\delta p(1-p)+(1-\delta)q(1-q)}{1-\delta p-(1-\delta)q},\ \beta_D' = \frac{\delta(p+\varpi)(1-p-\frac{\delta\varpi}{1-\delta})+(1-\delta)(q-\varpi)(1-q+\frac{\delta\varpi}{1-\delta})}{1-\delta p-(1-\delta)q}$$

ii) Next, we derive the corresponding differences between the posteriors as follows:

v) $\alpha_D'-\alpha_D = \frac{-2\gamma\varpi(v-w)-\frac{\gamma\varpi^2}{1-\gamma}}{1-\gamma v-(1-\gamma)w} < 0,\ \alpha_U'-\alpha_U = \frac{2\gamma\varpi(v-w)+\frac{\gamma\varpi^2}{1-\gamma}}{\gamma v+(1-\gamma)w} > 0,$

$$(\beta_U'-\beta_D')-(\beta_U-\beta_D) = \frac{2\delta\varpi(p-q)+\frac{\delta\varpi^2}{1-\delta}}{[\delta p+(1-\delta)q][1-\delta p-(1-\delta)q]} \equiv K > 0$$

iii) The incentive to link is defined by a subset of f_6, f_7, f_8, and f_9, depending on the equilibrium under consideration.

iv) First, we can easily show that the incentive to link is always higher under (p',q') than under (p,q) by considering the differences in incentive to link defined below:

$$f_6(p',q')-f_6(p,q) = [(\beta_U'-\beta_D')-(\beta_U-\beta_D)](1-\alpha_U)(u-c) = K(1-\alpha_U)(u-c) > 0$$
$$f_7(p',q')-f_7(p,q) = [(\beta_U'-\beta_D')-(\beta_U-\beta_D)](1-\alpha_D)(u-c) = K(1-\alpha_D)(u-c) > 0$$
$$f_8(p',q')-f_8(p,q) = [(\beta_U'-\beta_0')-(\beta_U-\beta_0)](1-\alpha_U)(u-c) = K(1-\alpha_U)(u-c) > 0$$
$$f_9(p',q')-f_9(p,q) = [(\beta_U'-\beta_0')-(\beta_U-\beta_0)](1-\alpha_U)(u-c) = K(1-\alpha_U)(u-c) > 0$$

iv) Next, we compare the incentive to link under (v',w') and under (v,w)

a)
$$f_6(v',w')-f_6(v,w) = (\alpha_U-\alpha_U')[(\beta_U-\beta_D)(u-c)+(1-\beta_0)u+\beta_0 c]$$
$$\Rightarrow f_6(v',w')-f_6(v,w) = \frac{-2\gamma\varpi(v-w)-\frac{\gamma\varpi^2}{1-\gamma}}{\gamma v+(1-\gamma)w}[(\beta_U-\beta_D)(u-c)+(1-\beta_0)u+\beta_0 c] < 0$$

b)

$$f_7(v',w') - f_7(v,w) = (\alpha_D - \alpha_D')[(\beta_U - \beta_D)(u-c) + (\alpha_U - \alpha_U')[(1-\beta_0)u + \beta_0 c]$$

$$\Rightarrow f_7(v',w') - f_7(v,w) = \frac{2\gamma\varpi(v-w) + \frac{\gamma\varpi^2}{1-\gamma}}{1-\gamma v-(1-\gamma)w}(\beta_U - \beta_D)(u-c) - \frac{2\gamma\varpi(v-w) + \frac{\gamma\varpi^2}{1-\gamma}}{\gamma v+(1-\gamma)w}[(1-\beta_0)u + \beta_0 c]$$

$$\Rightarrow f_7(v',w') - f_7(v,w) = \frac{2\gamma\varpi(v-w) + \frac{\gamma\varpi^2}{1-\gamma}}{[1-\gamma v-(1-\gamma)w][\gamma v+(1-\gamma)w]}\{\alpha_0(\beta_U - \beta_D)(u-c) - [1-\alpha_0][(1-\beta_0)u + \beta_0 c]\}$$

c) $$f_8(v',w') - f_8(v,w) = \frac{2\gamma\varpi(v-w) + \frac{\gamma\varpi^2}{1-\gamma}}{[1-\gamma v-(1-\gamma)w][\gamma v+(1-\gamma)w]}\{\alpha_0(\beta_U - \beta_0)(u-c) - [1-\alpha_0][(1-\beta_0)u + \beta_0 c]\}$$

d) $$f_9(v',w') - f_9(v,w) = \frac{-2\gamma\varpi(v-w) - \frac{\gamma\varpi^2}{1-\gamma}}{v+w}[(\beta_U - \beta_0)(u-c) + (1-\beta_0)u + \beta_0 c] < 0$$

As can be seen from the equations above, while $f_6(v',w') - f_6(v,w) < 0$ and $f_9(v',w') - f_9(v,w) < 0$, $f_7(v',w') - f_7(v,w)$ and $f_8(v',w') - f_8(v,w)$ may be positive or negative depending on the parameters.

v) Finally, $\frac{df_6}{dc} = -(1-\alpha_U)(\beta_U - \beta_D) - (\alpha_U - \alpha_0)\beta_0 < 0$. The same holds for f_6, f_7, and f_8.