Media, Aggregators and the Link Economy:
Strategic Hyperlink Formation in Content Networks

Chris Dellarocas - Boston University

Zsolt Katona - University of California, Berkeley

William Rand - University of Maryland

June 25, 2012

Abstract

A defining property of the World Wide Web is a content site's ability to place virtually costless hyperlinks to third-party content as a substitute or complement to its own content. Costless hyper-linking has enabled new types of players, usually referred to as content aggregators, to successfully enter content ecosystems, attracting traffic and revenue by hosting links to the content of others. This, in turn, has sparked a heated controversy between content creators and aggregators regarding the legitimacy and costs/benefits of uninhibited free linking. This work is the first to model the complex interplay between content and links in settings where a set of sites compete for traffic. We develop a series of analytical models that distill how hyperlinking affects (a) the incentives of content nodes to produce quality content vs. link to third-party content, (b) the profits of the various stakeholders, (c) the average quality of content that becomes available to consumers, and (d) the impact of content aggregators. Our results provide a nuanced view of the, so called, "link economy", highlighting both the beneficial consequences and the drawbacks of free hyperlinks for content creators and consumers.

1 Introduction

The ability to place hyperlinks across content is a defining feature of the World Wide Web (WWW). Hyperlinks have transformed the notion of content from a *collection* of loosely related items (e.g. books in a library) to a tightly woven *network* of web sites, blogs, etc. whose value comes not only from information stored in its nodes but also from the connections among those nodes.

Hyperlinks enable content creators to substitute or complement their own content with links to third-party content. Links are usually accompanied by a summary or commentary related to that content. Using links in this manner is common among bloggers, who use them as a mechanism for building a community and engaging with each other's ideas (Blood 2002). With few exceptions, such behavior is protected under the First Amendment.[1] Among other things this means that the

---

[1] See http://www.ala.org/ala/issuesadvocacy/copyright/copyrightarticle/hypertextlinking.cfm for a good discussion of U.S. law in this area. The law may be different in different countries.

Figure 1: Example of a news aggregator article entry

URL	Description
news.google.com	Algorithmically aggregated headlines and a search engine of many of the world's news sources. Search results group similar articles together.
huffingtonpost.com	Hybrid of news aggregator and original content creator.
drudgereport.com	Aggregator containing selected hyperlinks to news websites all over the world, each link carrying a headline written by the site's editors.
digg.com	User-generated news aggregator. Users post links to stories and votes promote stories to the front page.
techmeme.com	Blog aggregator tracking other technology blogs.
popurls.com	Meta-aggregator; aggregates links to the top stories chosen by other aggregators.
reddit.com	User-generated news aggregator. Users post links to stories and votes promote stories to the front page.
newsvine.com	Community news aggregator; member voting determines the position of news stories.
newsnow.co.uk	U.K.-based news portal.

Table 1: Examples of news aggregators.

link source does not need to seek permission, or pay royalties to the link target. On the commercial front, hyperlinking has enabled new types of players, commonly referred to *content aggregators* or *web portals*, to successfully enter professional content ecosystems, attracting traffic and revenue by hosting collections of links to the content of others (Dewan et al. 2004). Aggregators produce little or no original content; they usually provide titles and short summaries of articles they link to (Figure 1). Well known aggregators include Google News, the Drudge Report and the Huffington Post. Table 1 provides a more extensive list of examples.

The advent of the Internet has been disruptive to traditional content industries, such as newspapers and broadcast media, that have historically focused almost exclusively on content creation. Seeing their revenues collapse, some of them have turned against content aggregators, accusing them of "stealing" their revenues by free-riding on their content.[2] Some are even questioning the legitimacy of the unilateral free linking culture of the WWW, arguing that it might lead to a "tragedy of the commons" situation where content organizations are tempted to minimize the effort they spend on

---

[2] The recent dispute between the Associated Press and News Corporation with Google is perhaps the best example. See http://www.forbes.com/2009/04/06/google-ap-newspapers-business-media-copyright.html

original content production and rely on linking to third-party content. Other market actors point out that, in today's "link economy," links bring valuable traffic to target nodes, so content creators should be happy that aggregators exist and direct consumers to quality content sites (Karp 2007; Jarvis 2008). Key aggregator executives, such as Google's Eric Schmidt, assert that it is to their interest to see content creators thrive, since the value of links (and aggregators) is directly related to the quality of content that these point to.[3] The debate has already attracted the attention of governments and regulatory bodies. For instance, in 2009-2010 the Federal Trade Commission (FTC) hosted three workshops on the Future of Journalism and published a discussion draft that hints of copyright reform that will make it more difficult to place links to third-party content without payment to the creators of that content (FTC 2010). Moreover, the government of France has been discussing the idea of a "Google tax" to be imposed on search engines and aggregators and distributed to content creators.[4]

A central aspect of the debate focuses on the complex economic implications of the process of placing (for the most part) free hyperlinks across content nodes. As previously mentioned, links are a costless way of providing one's readers with access to third party content without incurring the cost of producing it. But links also allow sites to coordinate and divide labor, thus avoiding duplication of effort, e.g. situations where multiple sites end up producing similar content on the same stories. If links are chosen well, then they point to quality content; as a result, they reduce the search costs of the consumer population, which may lead to an aggregate increase in content consumption and to more traffic for higher quality sites. As these simple examples show, the aggregate economic impact of free hyperlinks on content ecosystems is not obvious, and a full understanding of the implications of hyperlinks for a site's traffic, revenue and content strategy is overdue.

This work aims to fill this gap by approaching the problem using a game-theoretic strategic network formation perspective. We develop a series of analytical models that distill how the ability to place free hyperlinks affects (a) the incentives of content nodes to produce quality content vs. link to third-party content, (b) the profits of the various stakeholders, (c) the average quality of content that becomes available to consumers, and (d) the impact of content aggregators. Our aim is to provide insights to both content creators as well as policy makers about how to behave and regulate respectively the new networked content ecosystem that has emerged.

Our models highlight the *complex interplay between content and links* in settings where a set of sites compete for traffic and make strategic investments in *both* content and links to maximize their revenues. Specifically, we model a set of content sites (e.g. news sites) that generate revenue from user visits (e.g. through advertising) and who compete for user attention among themselves as well as with alternative media (e.g. TV, blogs, Twitter feeds). Each site tries to maximize the number of visitors it receives and the amount of time they spend on the site. They do this by making decisions about what content (if any) to produce and what other sites (if any) to link to. Users want to maximize the utility they obtain per unit of attention and are, therefore, more likely to patronize

---

[3]"CEO Eric Schmidt wishes he could rescue newspapers", *Fortune* January 7, 2009.
[4]"France plans 'Google tax' on internet searches". http://www.telegraph.co.uk/technology/google/6947706/France-plans-Google-tax-on-internet-searches.html

sites that provide them with access to better content; this content could be directly produced by a site or simply linked to from a site. Through the use of links, sites can increase the total quality of content that their readers can access and, thus, the number of readers they attract, usually at lower cost than producing original content. On the other hand, original content increases the expected amount of time that a reader spends on a site, and thus, the likelihood that s/he will click on an ad and generate revenue for the site displaying the ad. We derive equilibrium content and linking strategies in such settings and compare the resulting site profits and content qualities to settings where hyperlinks are not possible. We also analyze how the entry of content aggregators affects site profits and content qualities.

2   Related Work

Although a substantial body of literature has studied the properties of Web-based content networks using empirical and simulation methods (for example: Huberman and Adamic 1999; Huberman et al. 1998; Pennock et al. 2002; Wu and Huberman 2008) most of this literature has either made no assumptions about individual agent behavior or has relied on ad-hoc and usually static agent-level specifications. Our work, in contrast, is approaching the formation of Web content networks from a strategic actor perspective, explicitly modeling both content organizations and users as strategic utility-maximizing agents. We are aware of few other papers that utilize a strategic content network formation approach, such as the one we are adopting in this paper. Katona and Sarvary (2008) investigate strategic linking between Web sites in a market for advertising links. Utility maximizing sites decide which other sites to buy (advertising) links from and how to price their own links. Their paper is among the first to model the evolution of the World Wide Web as a strategic network formation game. Building on the Katona-Sarvary model, Kominers (2009) examines the strategic production of sticky content in commercial sites that generate revenues from both selling services and selling links. A crucial question not fully addressed by the aforementioned papers is how links that are freely established contribute to the ecosystem and what motivates their formation. Mayzlin and Yoganarasimhan (2010) study a blogger's strategic decision to link or not link to a competitor's blog. Their work is focused on capturing a blogger's local link formation decision and does not attempt to analyze the system-level consequences of such decisions on network structure, content quality and social welfare. Ma (2010) conducts an empirical study of the implications of strategic linking within a website between individuals who post reviews. He finds that reciprocal linking is a natural outcome of the dynamic decisions since nodes have an incentive to increase viewership of others that link to them. This result is an empirical complement to some of our results, but does not examine the question of aggregators or different costs of production which are central to the examination of the more general web-based content creation industry. Chiou and Tucker (2011) conduct an empirical study of how the presence of news aggregators affects traffic to news websites. Their study recognizes some of the same trade-offs that motivate our own study.

Our results also contribute to the literature of strategic network formation (Jackson 2008). The majority of models that have been studied in that literature fall under one of two categories: first,

models where a set of fixed-attribute nodes make strategic link formation decisions (for example: Bala and Goyal 2000; Bloch and Dutta 2009; Kleinberg et al. 2008), and second, models where nodes that are connected together in a fixed network play a strategic game that requires them to make decisions about effort or some other node-specific strategic variable (for example: Bramoulle and Kranton 2007; Galeotti et al. 2006). In contrast, our models involve simultaneous and interdependent node-level strategic decisions about both node properties (e.g. effort to invest in original content) and link decisions (e.g. how many links to form and to whom to link to).

We are aware of only one paper that looks at interdependent content and link formation decisions: Galeotti and Goyal (2010) study a setting where identical nodes looking for information can either directly invest in acquiring it or form connections with others who already have it. They show that every robust equilibrium of this model exhibits the "law of the few": a small number of nodes invest in acquiring information and everyone else links to them. Our work has similarities but also important differences to theirs. In terms of similarities, we also find that, under certain conditions, competing nodes form link equilibria where one site makes high investments in content and other sites link to it. In our setting, however, nodes are in competition: node revenue has a positive relationship with the quality of information a node offers to consumers *relative to every other node*. This provides an incentive for nodes to abandon link equilibria and try to outdo one another by investing in original content. For that reason, we find that equilibria that exhibit "law of the few" properties often do not form. We also allow for nodes of heterogeneous ability as well as assume the presence of aggregators and alternative media; this enriches our results in important ways.

In a different context, Steiner (1952) examined quality competition among radio broadcasts in content creation and showed that this competition sometimes results in overproduction of content beyond what is socially optimal. Though our work is different in that it allows for linking and, as a result, dramatically alters the situation, some of the same results still hold.

Several of our results have analogies with results obtained in the literature on compatibility (see Katz and Shapiro 1994 and references contained therein). However, the mechanisms that drive those results do not hold in exactly the same way for linking as opposed to compatibility. Compatibility increases every consumer's willingness to pay due to network effects (consumers experience higher utility as part of a single large network than of two smaller networks). Linking increases consumers' willingness to pay due to the higher quality content that it now becomes rational for the link target to produce (this content becomes available to all via linking). Linking, further, increases joint profits because it also reduces the wasted effort of both sites investing in high quality content.

An important impact of aggregators on content ecosystems is the reduction of search costs, which points to the consumer search literature (e.g. Wolinsky 1984; Bakos 1997). Alba et al. (1997) showed that consumers seek out platforms which will decrease search costs. Though they were discussing physical products, our work supports this finding when it comes to content aggregators. Lal and Sarvary (1999) and Lynch and Ariely (2000) show that, in some contexts, the Internet increases competition due to the ability to quickly compare between different sites. We find that the market entry of aggregators increases competition among content sites, for similar reasons. Other work

(Ratchford et al. 2001; Johnson et al. 2004) has found that consumer search costs are affected by familiarity with the network and with a particular site. In this paper we do not specifically address consumer expertise, but it is one potential consumer motivation behind the notion of "anchor" sites that we will discuss in the next Section.

3   Model

We study a setting where N content sites (e.g. news websites) are competing with each other to attract and monetize user traffic. To maintain tractability, we assume that in each period there is only one topic of interest (e.g. one newsworthy story, chosen by nature) and that site $i = 1, .., N$ produces content of quality $c_i \geq 0$ on this topic. In addition to content creation, sites can place links to selected content of other sites. In the setting of this paper sites have no incentive to link to more than one other site (see discussion at the beginning of Section 6). Sites will, therefore, contain at most one link. Links are usually accompanied by a *link description*, a snippet of text that describes something about the link target's content or an excerpt of the link target's content. In the context of a blogger, this link description could in fact be an entire paragraph or blog post describing and commenting on the link, or in the case of an aggregator, such as Google News, it could simply be the title of the article and 2-3 lines taken from either the text or the meta-data of the link target (Figure 1).[5] Within the model, site i's cost of producing content of quality c_i is $\frac{k_i}{2}c_i^2$. Linking, on the other hand, is free and links cannot be refused by link targets in our basic model.

Users and Traffic.   A population of users/consumers visits content sites and the consumers derive utility from reading the content on these sites. Each period, every consumer begins her session from a site that we will refer to as the consumer's *anchor site*. The notion of an anchor site is based on the fact that recent research has shown that consumers tend to use a small number of sites when they start consuming news (Purcell, 2010). At the anchor site, she spends some time on the site reading its content and potentially clicks on one or more links and reads the content at the link target. To keep the setup parsimonious we assume that consumers who visit link target sites do not click on any other links while there. Since in the model we assume that all content is on the same topic, we treat different pieces of content as substitutes. Therefore, if site i has own content c_i and links to site j, consumers can expect to gain utility $z_i = \max(c_i, \delta c_j)$ by choosing site i as their anchor site. Factor $\delta \in [0, 1]$ captures the disutility of accessing content via a link, as opposed to directly. For example, this can be related to the cognitive cost of clicking on a link and reorienting oneself to a different context, i.e., a new web page layout. In most cases this disutility is small. To reduce notational clutter in the rest of the paper we will, therefore, assume $\delta = 1$.[6]

---

[5]There exists controversy on whether the unauthorized reproduction of excerpts (or even the title) of content to which one links constitutes a violation of copyright or whether it is covered by the "fair use" provisions of copyright legislation. Our model captures the implications of this practice and allows us to make theory-driven arguments about its social benefits and costs.

[6]We cover the case of $\delta < 1$ in a separate Technical Appendix.

New consumers are unaware of the content quality offered by each site and choose a random anchor site. All consumers aim to maximize the utility they receive from content so they periodically switch anchor sites using an exploration-exploitation process akin to a multi-armed bandit problem (Dubins and Savage 1965). In reality users often combine random browsing, the use of search engines, recommendations from their social networks, etc. when deciding what sites to use for news consumption. Regardless of the specific mechanisms employed, as consumers become more experienced it is expected that they will spend more time anchored at high utility sites.

Our objective is to derive a static model that captures the steady state properties of the dynamic game between content producers and consumers so that we can focus our attention on the competition among content creators. If we make the assumption that every period some consumers leave the ecosystem, i.e., they switch to alternative forms of content consumption, whereas an equal number of new (uninformed) consumers enter, then in the steady state, the population will include consumers at different stages of their exploration of the content ecosystem. Under this assumption, at steady state every node will have some traffic and the number t_i^A of consumers who anchor themselves at node i (the *anchor traffic* of site i) will be an increasing function of node i's user utility z_i and a decreasing function of every other node's utility. One function that satisfies these properties and lends itself to analytical tractability is a Tullock contest success function (Tullock 1980), commonly used to relate the probability of winning a contest (in this case, attracting traffic to one's site) to the resources each contestant devotes to it:

$$t_i^A(z_i) = \frac{z_i}{\sum_{j=1}^{N} z_j + \mu} \tag{1}$$

In the rest of the paper we will use the above function as our specification of anchor traffic. Factor $\mu \geq 0$ represents the utility that consumers expect to get *outside* the content ecosystem (e.g. by watching TV, exchanging Twitter messages, or simply having a live discussion on current events with friends). The presence of this factor underlies the fact that, however we delimit them, media ecosystems are almost invariably in competition with something else (an *outside alternative*) for consumer attention. The outside alternative can be another medium or simply another activity that consumers have the option of engaging in instead of consuming content. Therefore, if a site increases the utility z_i it offers to users, not only will it attract users away from other sites within the same ecosystem but also from the outside alternative. For example, if an online news site offers a revolutionary new way of accessing interesting content, it will attract visitors not only from competing online news sites but also from traditional TV. In fact, the above specification implies that, the higher the μ, the higher the percentage of additional traffic for site i that will come from the outside alternative versus from other sites of the content ecosystem. Restating this from the perspective of every other site, the higher the μ, the lower the relative impact of the change in any site's strategy on everybody else's traffic. The presence of an outside alternative, thus, tends to soften the competition among sites of the same ecosystem. As we will see in the following sections, this softening of competition plays an important role in enabling the formation of linking equilibria

that improve the profits of some or all content sites.

Site Revenues. We assume that site revenue (e.g. from advertising) is proportional to the total time visitors spend there. Once a consumer arrives at her anchor site, if there are no links she stays there for time proportional to the quality of available content. Let m_i denote the marginal revenue that site i earns. As this is a function of time spent per user and time spent is a function of content, it is $m_i = m(c_i)$. To keep the presentation simple we assume that $m(c) = c$. With this, if there are no links in the system, the total revenue of a site becomes:

$$R_i = t_i^A(c_i)c_i \tag{2}$$

The situation changes if we allow sites to place links to each other. Suppose that there is a single link from site i to site j. As we will show, in our setting it is only rational to place links to content of better quality. Therefore, the presence of a link implies that $c_j > c_i$. We assume that consumers behave as follows: Upon visiting site i, with probability ρ a consumer stays on the site and consumes its content without clicking the link, whereas with probability $1 - \rho$ she clicks a link and consumes site $j's$ content without consuming site i's content.

We can justify such consumer behavior as follows: Assume that there are two different types of readers: Readers who only care about (i.e. receive maximum utility from reading) a summary of a topic and readers who are interested in every single detail. Consider, for example, an article on a football game. Some readers would be fully satisfied from the article by learning the game's final score and would get no extra utility from reading more details. Others might want a detailed description of how well the two teams played. Let us call the former type *shallow* readers and the latter type *deep* readers. Assume that, whenever sites link to other sites, they include a summary of the target site's additional content (i.e. a summary of the content that is present at the target site and not at the source site). This assumption is consistent with the way that links to third-party content are used in news articles and news aggregators (see, for example, Figure 1). Shallow readers who have read the source content and link content summary would get no additional utility from visiting the link target.[7] Deep readers, on the other hand, will be better off if they click and move to the link target immediately.

We assume that a reader's type (shallow/deep) during a particular reading session depends on a variety of factors, such as how interested the reader is on the topic, how busy the reader is at the moment, how extensive and descriptive is the content summary that accompanies the link, etc. The same person can, thus, sometimes behave as a shallow reader and, at other times, as a deep reader. Let ρ denote the expected value of a reader behaving as a shallow reader when we take into consideration the entire population of readers, as well as the probability distribution of topics, reading times, etc. We, then, obtain the hypothesized behavior. Such behavior is consistent with

---

[7]In fact, if we assume that the cognitive cost of reading an article is proportional to its content quality, shallow readers would be strictly better off reading the (lower quality) anchor node content plus the (cognitively inexpensive) link summary than the link target content.

rational consumers who recognize that the link target always offers better content than the link source.

For the sake of parsimony, we assume that the factor ρ is an exogenous constant.[8] For a given distribution of readers and topics, ρ also depends on the amount of information contained in the excerpt/snippet that accompanies the link (Figure 1). Palme et al. (2012) have experimentally shown that, the longer the snippet length, the higher the ρ. Our comparative statics with respect to ρ can, thus, be interpreted as showing the implications of allowing aggregators to place longer snippets of articles they link to. The limiting case $\rho = 1$ models settings where site i reproduces all salient aspects of the content of site j.[9]

The preceding discussion shows that linking has both advantages and disadvantages. By placing a link, the source site can become a more attractive anchor node to consumers since it can now offer them access to better content, even if that content is a click away. Specifically, whereas without links the expected consumer utility from visiting site i would be $z_i = c_i$, placement of a link to site j whose content satisfies $c_j > c_i$ allows site i to increase its expected utility to $z_i' = \max(c_i, c_j) = c_j > z_i$. By (1), higher utility implies higher anchor traffic, i.e. $t_i^A(c_j) > t_i^A(c_i)$. (This argument shows that it is only rational to place links to sites that offer better content.) The main disadvantage of placing a link to a site of better content is that a fraction $1 - \rho$ of visitors will now click through directly to the better content, and these visitors will leave no revenue to the source site.

The trade-offs for the link target are exactly the opposite. The advantage of being a link target is that additional visitors arrive through that link. The disadvantage is that, as discussed in the previous paragraph, the link source can free ride on the link target's superior content, decreasing the target's *relative* attractiveness as an anchor node and, by (1), its anchor traffic. The number of visitors that reach j through a link from i to j (where $c_i < c_j$) is equal to the traffic $(1 - \rho)t_i^A(c_j)$ leaving site i. Thus, the total incoming *link traffic* of site j is:

$$t_j^L(c_j) = (1 - \rho) \sum_{k=1}^{l} t_{i_k}^A(c_j)$$

where $i_1, ..., i_l$ are the sites linking to j. We assume that visitors that arrive to site j through links consume its content and do not click on any further links present on j. The total revenue of a site in the possible presence of links then becomes:

$$R_i = \begin{cases} t_i^A(c_i)c_i & \text{if site } i \text{ is neither a link source nor a target} \\ \rho t_i^A(c_j)c_i & \text{if site } i \text{ links to another site } j \text{ but is not a link target} \\ \left(t_i^A(c_i) + t_i^L(c_i)\right)c_i & \text{if site } i \text{ is a link target but not a source} \\ \left(\rho t_i^A(c_j) + t_i^L(c_i)\right)c_i & \text{if site } i \text{ links to another site } j \text{ and is also a link target} \end{cases} \tag{3}$$

The following sections analyze the competitive interactions among content sites in the above setting. In all cases we study a simultaneous move game where sites simultaneously decide how much to invest in content, as well as if and which other sites to link to.

4   Two sites

We derive our first set of insights by studying a setting with just two content sites. Our objective is to examine how the option of placing free links to third-party content affects competition, content quality and site profits in such a simple setting.[10]

4.1   Payoff functions

Depending on the context we will refer to the two sites either using subscripts 1, 2 or S, T, the latter indicating the source and target of a link respectively. When there are no links between the two sites the expected utility for site i is $z_i = c_i$, the link traffic for site i is $t_i^A = c_i/(c_i + c_j + \mu)$, and, given that site i's cost of creating content is increasing in the square of the content created by the site then its payoff function is given by:

$$\pi_i = \frac{c_i}{c_i + c_j + \mu} c_i - \frac{k_i}{2} c_i^2. \tag{4}$$

Once we introduce the possibility of placing free links to other sites' content is it easy to see that it is never individually rational for any site S to place a link to a site T of equal or lower content. Specifically, placing a link to content $c_T \leq c_S$ does not change the utility $z_S = \max(c_S, c_T)$ consumers get from making site S their anchor, and therefore does not increase site S's anchor traffic. At the same time, per (3), the presence of the link decreases the source site's revenue per visitor by a factor ρ. Therefore, either no site will link to the other or the site with (strictly) lower content will link to the peer with better content. Under these assumptions, when site S links to site T the relationship between content for the two sites must be governed by $c_S < c_T$, and the utility expected by consumers for the two sites is: $z_S = c_T$ and $z_T = c_T$. Together with (1), these relationships imply that the traffic for the two sites will be: $t_S^A = t_T^A = c_T/(2c_T + \mu)$. From (3) the corresponding payoff functions then take the form:

$$\pi_S = \frac{c_T}{2c_T + \mu} \rho c_S - \frac{k_S}{2} c_S^2 \qquad \pi_T = \frac{c_T + (1 - \rho)c_T}{2c_T + \mu} c_T - \frac{k_T}{2} c_T^2 \tag{5}$$

For expositional clarity we first look at the case where the two sites have identical cost parameters $k_i = 1$. In Section 4.3 we study the more general case where one site is more efficient than the other.

---

[10]Section 6 shows that the results obtained in this Section remain qualitatively robust in settings with multiple content sites.

4.2 Homogeneous sites

When sites are homogeneous and if there are no links, then sites simply maximize the profit function described in (4), yielding $t_1^A = t_2^A = \frac{c_{NL}}{2c_{NL}+\mu}$ and

$$c_1^* = c_2^* = c_{NL} = \frac{3 - 4\mu + \sqrt{9 + 8\mu}}{8} \tag{6}$$

When $\mu > 2$, this expression becomes negative. We will thus assume $\mu < 2$ throughout the analysis to avoid this situation. Further examining the equilibrium described by (6), one can derive that it is Pareto dominated by the symmetric outcome that maximizes sites' profits:

$$c_1 = c_2 = c_P = \frac{2 - 4\mu + \sqrt{4 + 16\mu}}{8} < c_{NL}$$

The above outcome arises e.g. in settings with geographical segregation, where each site has exclusive access to one half of the audience and only competes with the outside option.[11]

In common with many similar settings, it follows that, when

two homogeneous content sites compete for the same audience, they produce better content, but end up with lower profits, relative to a setting where each site has exclusive access to one half of the audience.

This result captures an important consequence for the news industry (and other, previously geographically segregated, content industries) of technologies, such as the Web, that allow individuals to gather information for free (or cheaply) from any content-producing site around the world. Sites that previously had monopoly power over their respective audience segments now have to compete with each other for the entire audience. Competition induces higher investment in, what is essentially duplicate content. Because increased content investments are symmetric, they do not change relative market shares and result in reduced profits for all sites.

We will show that under certain circumstances sites can use the option of linking to alleviate counter-productive investments in duplicate content. The following proposition characterizes the form of the resulting equilibria.

Proposition 1. *There exist thresholds $\underline{L}(\rho), \overline{NL}(\rho) \in [0, 2]$ such that:*

1. *If $\mu \leq \overline{NL}(\rho)$ then sites do not establish links in equilibrium and $c_i^* = c_j^* = c_{NL}$.*

2. *If $\mu \geq \underline{L}(\rho)$ then there are two asymmetric equilibria where one site links to the other and*

$$c_T = \frac{1 - \mu}{2} - \frac{\rho}{4} + \frac{\sqrt{(2 - \rho)(4\mu + 2 - \rho)}}{4} >$$

$$c_S = \frac{\rho c_T}{2c_T + \mu}$$

---

[11]The payoff functions of this alternative setting are given by: $\pi_i = \frac{c_i}{2c_i+\mu}c_i - \frac{k_i}{2}c_i^2$.
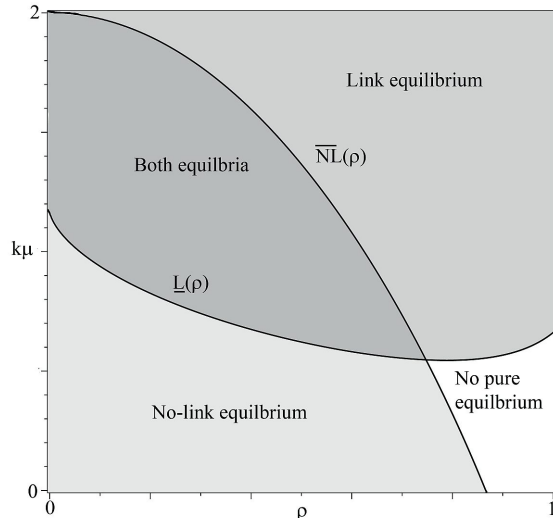
Figure 2: Equilibrium regions when the two content sites have identical cost parameters and $\delta = 1$.

3. *There is no equilibrium in pure strategies otherwise.*

The most striking part of the above result is the possibility of an asymmetric equilibrium, in which one site produces high quality content, whereas the other site gives up content competition and links to the high content site. Section 6 shows that this result generalizes in settings with multiple nodes. In real world terms, this result corresponds to situations where one target newspaper invests heavily in quality content on a topic, and all other newspapers link to it, without trying to compete by producing what would, essentially, be duplicate content on the same topic. Although the concept of linking to competitors has, traditionally, been an anathema to most newspapers, in recent years we are witnessing new players, such as the Huffington Post, who have built efficient and successful news organizations by covering few stories themselves and linking to high quality (competing) news outlets for the rest.

Figure 2 depicts the equilibrium regions. Symmetric no-link equilibria exist in the region *below* curve $\overline{NL}(\rho)$; asymmetric link equilibria exist in the region *above* curve $\underline{L}(\rho)$. As ρ grows, linking becomes an attractive option for the link source because it retains a higher percentage of anchor traffic. Accordingly, as ρ grows, the region where no-link equilibria are sustainable shrinks and the region where link equilibria are sustainable grows (except for large ρ, see below). Observe, also, that asymmetric link equilibria are sustainable only for sufficiently large values of μ. The intuition for this, as well as the details of the linking equilibria are different when ρ is small and when ρ is large.

Small ρ. For $\delta = 1$, when one site links to the other, then both sites offer the same expected utility to consumers, and end up sharing the total traffic that comes to the content ecosystem. This reduces the competition between them for market share. Furthermore, in the presence of an outside alternative ($\mu > 0$), higher content investments by the link target benefit both the target and the source, because the new customers who will be attracted away from the outside alternative will be split between the link source and target. When ρ is small, many of the visitors of the link source
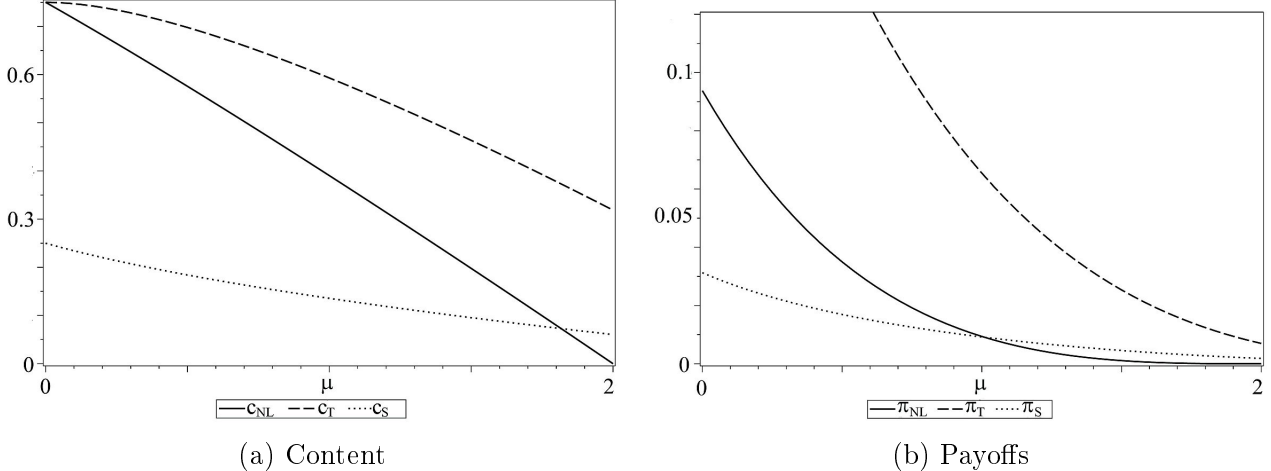
(a) Content   (b) Payoffs

Figure 3: Equilibrium content and payoffs when $\rho = 0.5$ ($k = \delta = 1$)

end up clicking through and generating revenue for the link target. For every unit of additional investment in content, the link target thus (a) increases the number of visitors that come from the outside alternative both to itself and to the link source, and (b) is able to capture additional revenue both from the new visitors that come to itself as well as from a fraction $1 - \rho$ of the new visitors that come to the link source. Therefore, the presence of an outside alternative allows the link target to invest in substantially better content than in a no-link equilibrium. This, in turn, makes it very difficult for the link source to compete with the link target in a content (no-link) equilibrium. If μ is sufficiently large, the additional traffic that the link source receives (thanks to the increase in the link target's content) compensates it for the fact that it only retains a fraction ρ of the corresponding revenue. Both these forces make it more profitable for the link source to form a link, albeit with lower content and lower profits than the link target, than to attempt to compete directly with the link target on content.

Figure 3 depicts the equilibrium content and payoffs for $\rho = 0.5$. For these parameters the asymmetric equilibria become sustainable for $\mu > 0.62$. Observe that $c_S < c_{NL} < c_T$ and $\pi_{NL} < \pi_T$, and therefore consumers and the link target are better off in the presence of links. In fact, the link target is even better off than in the Pareto optimal setting where each site has a monopoly over half the consumer population. The link source is always worse off relative to the link target. Interestingly, however, if μ is sufficiently large ($\mu > 1$ in this example), the link source also ends up better off relative to the no link case because it receives revenue from a lot of additional visitors (attracted away from the outside alternative) while spending less on original content costs.

Large ρ. When ρ is large (implying that the link source captures a large portion of the advertising revenue) then the potential link target does not benefit much from link traffic. However, the reduced competition between itself and the other site allows the link target to reduce its content investment and bring its content levels down to the Pareto optimal c_P, a move that in the end results in increased profits for the link target. If there is no outside alternative ($\mu = 0$) such a move would not be sustainable: Lower content investment by the link target would give an incentive to the link source

13

to abandon the link and invest heavily in content itself. However, if μ exceeds a certain level (i.e., the outside alternative is significantly high), then the potential link target has an incentive to maintain a higher level of investment in order to not lose traffic to the outside alternative. This makes it more difficult for the potential link source to compete without linking to its competitor, making it more attractive for the link source to free-ride on the target's content as opposed to deviating.[12]

The following corollary states the general form of these relationships:

Corollary 1. *The equilibrium content levels and profits satisfy:*

$$c_S \leq c_P \leq c_T \text{ and } \pi_{NL} \leq \pi_P \leq \pi_T$$

For low values of ρ it is also the case that:

$$c_{NL} \leq c_T, \quad \pi_S \leq \pi_T \text{ and } \pi_S \leq \pi_{NL}$$

In summary, in settings where there are two evenly matched competitors, the option of placing links across sites may lead to equilibria where one or both sites are better off relative to a no-link setting. Even though some details depend on the magnitude of ρ, in all cases linking helps reduce the inefficiency present in the no linking case that is due to, from the perspective of content sites, excessive investment on substitute content, driven by competition.

4.3 Heterogeneous sites

In the more general case the two competing sites have different abilities to produce content. We capture this by assuming different cost parameters $k_1 = 1$ and $k_2 = k > 0$. Therefore, if $k_2 < 1$, then site 2 is more efficient than the site 1, and if $k_2 > 1$ then site 1 is more efficient.

If there are no links, then sites maximize profit as described in (4), yielding the following result:

Proposition 2. *When two heterogeneous sites compete for the same audience and cannot form links then:*

1. *If the difference in their cost parameters is not too large then both competitors produce content and capture a positive market share; the more efficient (lower cost) competitor produces better content and captures a higher market share.*

2. *If the difference in their cost parameters is large, market entry of the less efficient content producer is not viable; the more efficient competitor then becomes a monopolist.*

The most striking property of the above result is that there is no room for a second competitor whose content production cost is substantially higher than that of the most efficient content producer.

---

[12]The link target's reduced content investment and the source's correspondingly increased temptation to abandon the link and compete on content explains the non-monotonicity of $\underline{L}(\rho)$ for large ρ.

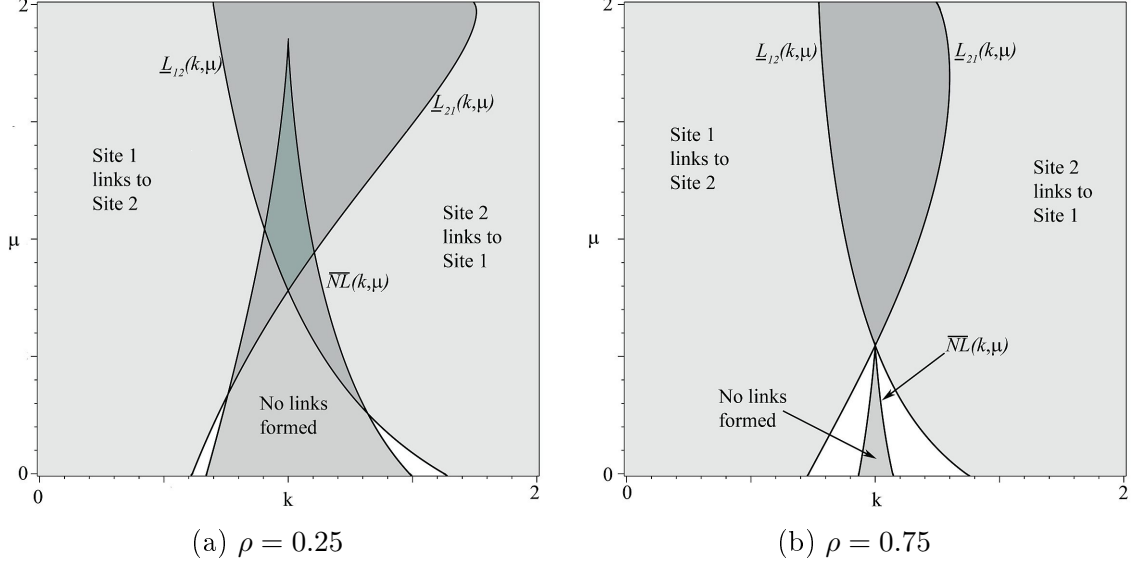|	
(a) $\rho = 0.25$	(b) $\rho = 0.75$

Figure 4: Equilibrium regions when sites have heterogeneous costs and links are possible. The white regions are regions where no pure equilibria exist.

The ability to place links to the other site's content becomes even more important in this setting as it now allows inefficient sites, that would otherwise not be able to viably enter the market, to generate positive profits. The following proposition provides the details.

Proposition 3. *There exist thresholds $\underline{L}_{12}(k,\rho), \underline{L}_{21}(k,\rho), \overline{NL}(k,\rho)$ such that:*

1. *If $\mu \leq \overline{NL}(k,\rho)$, an equilibrium exists where sites do not establish links in equilibrium. Their content levels are then given by (10).*

2. *If $\mu \geq \underline{L}_{21}(k,\rho)$, an equilibrium exists where site 2 links to site 1 and*

$$c_1 = c_T = \frac{1-\mu}{2} - \frac{\rho}{4} + \frac{\sqrt{(2-\rho)(4\mu + 2 - \rho)}}{4} >$$

$$c_2 = c_S = \frac{\rho c_T}{k\,(2c_T + \mu)}$$

3. *If $\mu \geq \underline{L}_{12}(k,\rho)$, an equilibrium exists where site 1 links to site 2 and*

$$c_2 = c_T = \frac{1-k\mu}{2k} - \frac{\rho}{4k} + \frac{\sqrt{(2-\rho)(4k\mu + 2 - \rho)}}{4k} >$$

$$c_1 = c_S = \frac{\rho c_T}{2c_T + \mu}$$

4. *There is no equilibrium in pure strategies otherwise.*

15

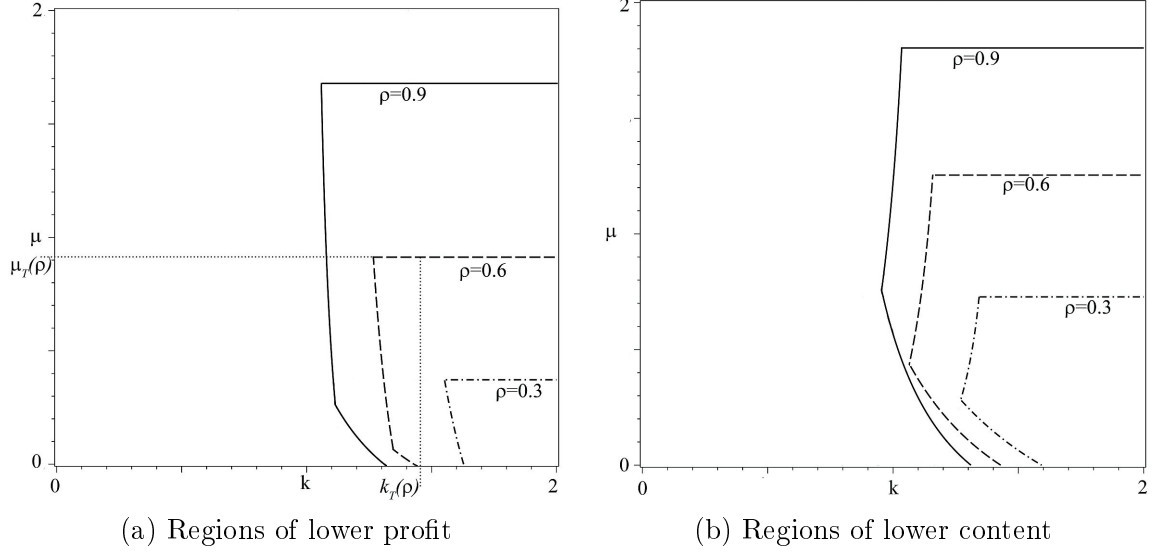(a) Regions of lower profit   (b) Regions of lower content

Figure 5: The areas that are *below* and *to the right* of each of the above curves represent parameter regions when site 1 ends up worse off and produces lower content when links are allowed relative to a setting where no links are possible.

Figure 4 depicts the parameter regions where each of the above equilibria become possible. As expected, high cost (inefficient) sites will link to low cost (efficient) sites, especially when the cost differential is high. For example, we see that site 2 links to site 1 when $k > 1$ and that site 1 links to site 2 when $k < 1$. An important observation is that the ability to link to a more efficient site *always* makes it individually rational for a site, no matter how inefficient, to enter the market and capture positive market share. This is in stark contrast to the situation without links, where market entry is not viable for very inefficient sites.

Figure 5 depicts the parameter regions where link equilibria result in higher or lower profits and content for the link target (relative to equilibria where no links are possible).[13] It is interesting to contrast these to the corresponding results of the previous section. In settings with sites of equal capability to produce content, when linking is sustainable, it *always* results in higher profits for the link target, because it reduces the inefficiency of duplicate content investment when the two (evenly matched) sites compete head-on. When sites have different capabilities this result only holds (a) when the two sites do not have large differences in their cost parameters, or (b) when μ is sufficiently large compared to ρ. Specifically, observe that in Figure 5 for each ρ there exist thresholds $k_T(\rho), \mu_T(\rho)$ such that, when $k > k_T(\rho)$ (i.e. when site 2 is sufficiently less efficient than site 1) the link target realizes higher profits only when $\mu > \mu_T(\rho)$. Furthermore, it appears that $\mu_T(\rho)$ is an increasing function of ρ. Similar patterns govern the production of content.

The intuition behind the result is that, without linking, if a site is substantially more efficient than its competitor it will capture the entire market and will produce the monopoly content levels. Linking allows sites that would otherwise not be viable competitors to stay on the market, free-

---

[13]We only depict results for equilibria where site 2 links to site 1. Equilibria where site 1 links to site 2 have symmetric properties.

riding on the efficient site's content and (for $\rho > 0$) capturing some revenue that would otherwise go to the link target. This decreases the efficient site's profits, as well as its incentives to produce content. At the same time, however, the market entry of the new site attracts some traffic away from the outside alternative and thus expands the total traffic flowing into the content ecosystem. To see why this happens observe that the new site links to content of quality c_T. When site T is the only content site, the total traffic that flows into the content ecosystem is equal to $\frac{c_T}{c_T+\mu}$. After the entry of the less efficient site the total traffic that flows into the new ecosystem (i.e. to sites S and T) is $\frac{2c_T}{2c_T+\mu} \geq \frac{c_T}{c_T+\mu}$, with the inequality strict for $\mu > 0$. This specification models the fact that every additional node entering the ecosystem represents an additional path to good content. The existence of more paths makes it more likely that consumers will "stumble upon" such content, effectively reducing their search costs. This, in turn, increases the attractiveness of the entire content ecosystem relative to the outside alternative, attracting some traffic (e.g. audience and/or time spent) away from it . If ρ is small compared to μ, most of the additional traffic that flows to the link source clicks on its link and lands on the link target, compensating the target for the loss of market share and revenue incurred by the entry of the link source site. If ρ is large, however, the net effect of the new node's entry for the link target's profits and content is negative.

The trade-off discussed in the previous paragraph is at the core of the controversy surrounding aggregators. We explore it more formally in the next section.

5 Aggregators

In this section, we consider the entry of sites that have a very high cost of content production and, thus, can only attract visitors by linking to content created by others. These sites are usually called aggregators. To better highlight the complex impact of aggregators on content networks we first examine the direct effect of an aggregator in a setting where the incumbent content sites do not change their content and linking behavior as a response to the aggregator's entry. Then, we study how incumbents react to aggregator entry by changing their content level and how this affects the competition between content creators.

5.1 The main effect of aggregators on traffic and revenue

We assume three sites and fix their content decisions. Site 3 is a new entrant, an aggregator with a very high content creation cost. Such a site will produce almost no content and will place a link to the site that produces the highest content since that will attract the largest audience to the aggregator. We study how the presence of this player changes the payoffs of the other two sites. To do this we look at two scenarios:

1. Sites 1 and 2 produce the same amount of content $c_1 = c_2$ and there are no links between them

2. One site (e.g. site 2) produces lower quality content and links to the other site (site 1).

In both scenarios we assume that the aggregator links to content of quality c_1 and, thus, offers utility c_1 to consumers. Without the aggregator, the total traffic that flows into the content ecosystem (sites 1 and 2) is equal to $\frac{c_1+c_2}{c_1+c_2+\mu}$. With the aggregator the total traffic that flows into the *new* ecosystem (sites 1, 2 and 3) is $\frac{2c_1+c_2}{2c_1+c_2+\mu} > \frac{c_1+c_2}{c_1+c_2+\mu}$. This specification models the fact that, by providing additional entry points for accessing good content, the presence of aggregators increases the attractiveness of the content ecosystem relative to outside alternatives (e.g. it reduces consumer search costs). This, in turn, increases the total traffic that flows into the entire content ecosystem.

The impact of aggregators on individual content site traffic is two-fold. On the one hand $\frac{c_i}{2c_1+c_2+\mu} < \frac{c_i}{c_1+c_2+\mu}$, which means that aggregator entry reduces the anchor traffic of *all* incumbent content sites. Aggregators, therefore, attract anchor traffic away both from the outside alternative as well as from every other content site. On the other hand, a fraction $1-\rho$ of the aggregator's anchor traffic clicks on its link and eventually lands at the high content site. If ρ is not very high, the link from the aggregator thus allows the high content site to recover most of the original traffic that it lost to the aggregator and also part of the additional traffic that flows to the aggregator from the outside alternative. The high content site then ends up better off in the presence of the aggregator. In contrast, as ρ approaches one, the aggregator retains almost all its anchor traffic, and this results to a net loss for all incumbent content sites. Of course, the low content site is always worse off since it loses audience share to the aggregator and gains none of it back. The following proposition formalizes the above discussion:

Proposition 4. *(1) If sites 1 and 2 produce the same content and do not link to each other, then they are better off in the presence of an aggregator iff*

$$\rho < \hat{\rho}_{NL} = \frac{\mu}{2c_1 + \mu}$$

(2) If site 1 produces higher content and site 2 links to it then site 1 is better off in the presence on an aggregator iff

$$\rho < \hat{\rho}_T = \frac{\mu}{c_1 + \mu}$$

whereas site 2 is always worse off. The sum of the profits that sites 1 and 2 make is higher in the presence of an aggregator iff

$$\rho < \hat{\rho} = \frac{\mu}{c_2 + c_1 + \mu} < \hat{\rho}_T$$

Proposition 4 can help explain the current tension between content producers and news aggregators, e.g. the controversy between the Associated Press and Google News, discussed in the paper's introduction. Aggregators claim that their presence ought to be beneficial to high quality content producers because they link and, thus, direct traffic to them. Nevertheless, all aggregator visitors were originally (before aggregators appeared) visitors of one of the content producers, so aggregators indeed "stole" some anchor traffic away from content producers. Furthermore, a fraction ϱ of aggregator anchor traffic is satisfied reading article titles and snippets posted by the aggregator (Figure 1), and does not click through to visit any content sites. This lost traffic represents a net loss for content producers, and the main source of their complaints.

At the same time, by making it easier for consumers to access good content, aggregators increase the attractiveness of the entire content ecosystem and, thus, also attract traffic away from alternative media. A fraction $1 - \varrho$ of this additional traffic is directed, via the aggregator, to the high quality content producer's site. If this additional traffic is substantial enough, it can compensate high quality content producers for the anchor traffic lost to the aggregator. This can happen when μ is large and ϱ is small. In such cases the net impact of the aggregator to high quality content producers can be positive.

It is interesting to observe that the presence of an outside alternative, and, thus, the opportunity to expand the total traffic that flows into the content ecosystem, is essential for aggregators to be beneficial. Aggregators can never be beneficial to incumbent sites in "closed universe" settings where there is no outside audience to be attracted. This is easy to see in our model: when $\mu = 0$, Proposition 4 predicts that incumbent sites are better off in the presence of aggregators if and only if $\rho < 0$, which can never occur.

5.2   The competitive effect of aggregators

In this section, we allow sites 1 and 2 to endogenously adjust their content and link formation decisions when an aggregator enters the market. As before, our key assumption is that aggregators create very little content of their own but contain a *single* link that points to the best available content. This assumption captures the content filtering role of aggregators: in real-life aggregators do not link to all available content on a topic; instead they make selections for a number of reasons. For example, consumers might have limited cognitive budgets and thus are willing to follow one link at most. Alternatively, there may be screen size constraints that limit the number of possible links, especially in mobile browsing environments. In summary, the main added value of aggregators is as filters.

We assume that the aggregator cannot perfectly determine the content levels of the incumbents, but links to the high content site with high probability. When there is no link between the two incumbents, the aggregator will link to site 1 with probability $\frac{c_1^s}{c_1^s + c_2^s}$, where $s \geq 0$ is the amount of search that the aggregator does. When $s = 0$ the aggregator is unable to determine quality and randomly chooses between the two sites. When $s = 1$, the aggregator is only as good as consumers in finding the best sites. If $s \to \infty$ the aggregator can find the top site with perfect precision. Since the two incumbents know each other's content well, we further assume that a link between them indicates that the link target has a higher content. Thus, if links exist, the aggregator will *always* link to the link target.[14]

Under the above assumptions, aggregator entry induces competition among the two incumbent sites for the aggregator's link and its associated traffic. We will now perform an analysis similar to that of Section 4.2 in this new setting and investigate how the competition induced by the aggregator affects the results. Our main objective is to explore how the aggregator's search parameter s affects

---

[14]The assumption that aggregators use the link structure as a cue to content quality is consistent with actual practice. For example, the PageRank algorithm that forms the backbone of the Google search engine ranks sites on the basis of how many incoming links they receive from other, similarly highly ranked, sites.

content investments, profits and the propensity of content sites to form links. To keep things simple, we only study the case where the two incumbent sites have identical cost parameters $k_i = 1$.

5.2.1 Incumbents are pure content creators

First, we look at the case when the two incumbents are pure content creators who (for strategic or policy reasons) do not consider placing links to each other. Player i expects that it will be linked to from the aggregator with probability $\frac{c_i^s}{c_i^s + c_j^s}$. If so, the aggregator's utility to consumers will be c_i and thus the content ecosystem will effectively consist of three nodes with total content $c_i + c_i + c_j$ competing against the outside alternative μ. Node i's anchor traffic will then be equal to $\frac{c_i}{2c_i + c_j + \mu}$. The aggregator's anchor traffic will also be $\frac{c_i}{2c_i + c_j + \mu}$. A fraction $1 - \rho$ of that traffic will click the aggregator's link and will visit node i. The total traffic of node i will thus be equal to $\frac{c_i + (1-\rho)c_i}{2c_i + c_j + \mu}$. With probability $\frac{c_j^s}{c_i^s + c_j^s}$ the aggregator will choose to link to node j. In that case (a) the total content offered by the content ecosystem will become $c_i + c_j + c_j$, (b) node i's anchor traffic will be $\frac{c_i}{c_i + 2c_j + \mu}$, and (c) there will be no link traffic flowing to node i. Putting everything together the expected payoff of player i is given by:

$$\pi_i = \frac{c_i^s}{c_i^s + c_j^s} \cdot \frac{c_i + (1-\rho)c_i}{2c_i + c_j + \mu} c_i + \frac{c_j^s}{c_i^s + c_j^s} \cdot \frac{c_i}{c_i + 2c_j + \mu} c_i - \frac{1}{2}c_i^2. \tag{7}$$

We determine the symmetric equilibrium.

Lemma 1. *When $\mu < (3-\rho)\left(1 - s\frac{3(1-\rho)}{2(5-2\rho)}\right)$ or, equivalently, $s < \frac{2(5-2\rho)(3-\rho-\mu)}{3(1-\rho)(3-\rho)}$ and content creators do not have the option of linking to each other, equilibrium content levels are*

$$
\begin{aligned}
c_1^* = c_2^* &= c_{NL} \\
&= \frac{13}{36} + s\frac{1-\rho}{24} - \frac{3\mu+\rho}{9} + \frac{\sqrt{9(1-\rho)^2 s^2 + 12(1-\rho)(13-4\rho)s + 64\rho^2 - (192\mu+416)\rho + 480\mu + 676}}{72}
\end{aligned}
\tag{8}
$$

and the profits are

$$\pi_1^* = \pi_2^* = \pi_{NL} = \frac{c_{NL}^2(3 - \mu - \rho - 3c_{NL})}{2(3c_{NL} + \mu)}.$$

Otherwise, there is no equilibrium in pure strategies and $\pi_{NL} = 0$.

The lemma reveals the disruptive effect of an aggregator on the competition between content creators. It is easy to see that, as long as $\rho < 1$, equilibrium content levels are increasing and profit levels are decreasing in s. The intuition is that the more effective the aggregator is at finding the better site, the more the incumbent sites compete for the incoming link by investing in content above the level that is optimal for them. In fact, if $s > \frac{2(5-2\rho)(3-\rho-\mu)}{3(1-\rho)(3-\rho)}$, the only equilibria are mixed equilibria that leave both content sites with zero profits.

The following proposition generalizes these intuitions and also examines the rather surprising impact of ρ on content and profits.

Proposition 5. *The symmetric equilibrium of Lemma 1 exhibits the following properties:*

1. *Equilibrium content is monotonically increasing in s and decreasing in ρ.*

2. *Equilibrium profits are monotonically decreasing in s.*

3. *For low (high) values of s profits are monotonically decreasing (increasing) in ρ. For inter-mediate values of s profits exhibit an inverse U-shaped relationship, first increasing and then decreasing with ρ.*

We already discussed how s affects content levels: Aggregators that can more effectively pick the highest content site increase competitiveness between the content sites, leading to higher content and lower profits. The rather complex effect of the aggregator's click-through rate $1 - \rho$ on profits is a consequence of ρ's dual impact on traffic and competitiveness: The more visitors an aggregator sends to its link target (the lower the ρ) the higher the revenue of the link target. This is the *traffic effect* of ρ which increases content levels and increases profits. At the same time, the higher the link traffic, the higher the competition for this link between the two incumbent sites. This, secondary, *competition effect* of ρ further increases content but *decreases* profits. When s is low, a site's content does not significantly affect its probability of being linked to from the aggregator. The competition effect is then weak and the traffic effect dominates, resulting in a reduction of profits as ρ grows. In contrast, when s is high the competition effect dominates and leads to the surprising result that an increase in ρ may lead to higher profits: As ρ increases the aggregator sends fewer visitors to content sites through its link, but this in turn can decrease competitiveness, leading to lower content and higher profit overall.

5.2.2 Incumbents can place links

Assume that site S produces own content c_S and links to site T who produces content $c_T > c_S$. Per our assumption, the aggregator will then also link to site c_T. In such a setting the two incumbent sites and the aggregator will each offer utility c_T to their consumers. Site T will receive anchor traffic $\frac{c_T}{3c_T + \mu}$ plus link traffic $\frac{(1-\rho)c_T}{3c_T + \mu}$ from site S and the same link traffic from the aggregator. Site S will receive anchor traffic $\frac{c_T}{3c_T + \mu}$ and no link traffic. The corresponding payoff functions take the form:[15]

$$\pi_S = \frac{c_T}{3c_T + \mu}\rho c_S - \frac{1}{2}c_S^2 \qquad \pi_T = \frac{c_T + 2(1 - \rho)c_T}{3c_T + \mu}c_T - \frac{1}{2}c_T^2 \tag{9}$$

The following proposition characterizes the form of the resulting equilibria when free linking is allowed in a setting with two content sites and an aggregator.

Proposition 6. *There exist thresholds $\underline{L}(\rho, s), \overline{NL}(\rho, s)$ such that:*

1. *If $\mu \leq \overline{NL}(\rho, s)$, an equilibrium exists where sites do not establish links in equilibrium. Their content levels are then given by (8).*

---

[15]Contrast these functions to equation (5), which gives the payoff functions in a setting with two sites and no aggregator.

(a) $\overline{NL}(\rho)$ for different s (b) $\underline{L}(\rho)$ for different s
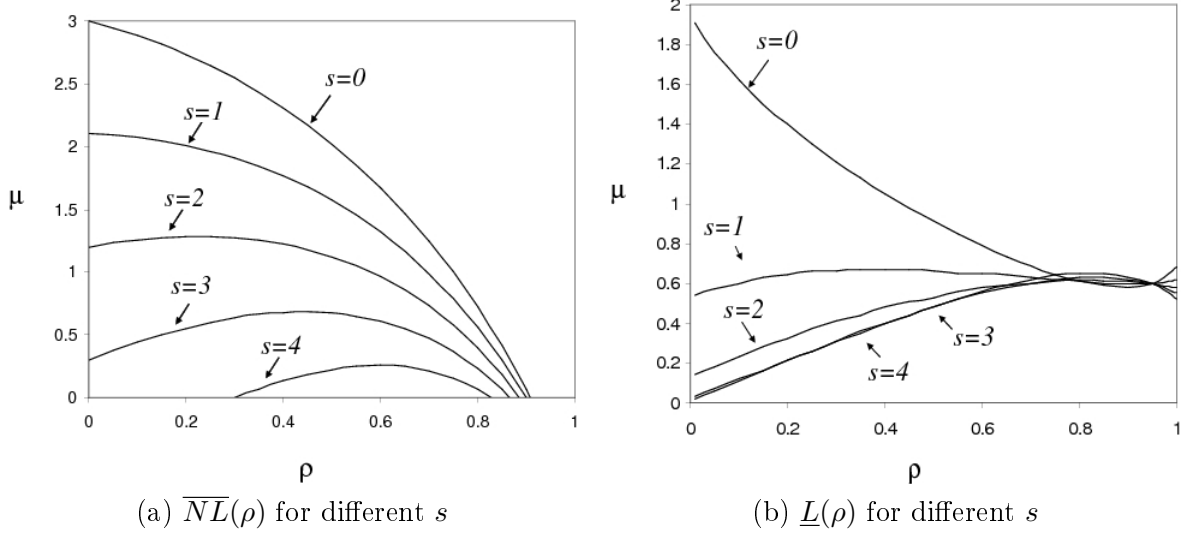
Figure 6: As the aggregator's search parameter s increases, the parameter region where no-link equilibria are sustainable (the area *below* curves $\overline{NL}(\rho)$) shrinks, whereas (for low ρ) the parameter region where link equilibria are sustainable (the area *above* curves $\underline{L}(\rho)$) expands.

2. *If* $\mu \geq \underline{L}(\rho, s)$, *an equilibrium exists where site S links to site T and*

$$c_T = \frac{1}{2} - \frac{\mu + \rho}{3} + \frac{\sqrt{(3 - 2\rho)(4\mu + 3 - 2\rho)}}{6} >$$

$$c_S = \frac{\rho c_T}{3 c_T + \mu}$$

3. *There is no equilibrium in pure strategies otherwise.*

The above result is analogous to Proposition 1, which describes a similar situation in a setting without aggregators. Our main interest in this section is to explore how the effectiveness of the aggregator in discovering quality content (s) affects the incentives of the two sites to form links vs. to compete head-on on content. This is best accomplished by examining how the parameter regions where no-link/link equilibria are sustainable shift as the aggregator's search parameter s increases.

Figure 6(a) plots the curve $\overline{NL}(\rho)$ for several values of s. The area *below* each curve corresponds to the parameter region where no-link equilibria are sustainable for the corresponding value of s. Observe that, for $s = 0$, the shape of the curve is similar to that of the corresponding curve in settings without the aggregator (see Figure 2). As s grows, the region where it is an equilibrium for sites to compete head-on on the basis of content *shrinks*. The explanation is straightforward in light of the results of the previous section: As s increases, so does competition among content sites. This increases content levels but reduces profits. As profits get squeezed, each content site finds it increasingly attractive to deviate from the equilibrium, reduce its own content production and simply place a link to the other site.

Figure 6(b) similarly plots the curve $\underline{L}(\rho)$ for several values of s. The area *above* each curve

corresponds to the parameter region where link equilibria are sustainable. Observe that, for $s = 0$, the shape of the curve is similar to that of the corresponding curve in settings without the aggregator (see Figure 2). When ρ is small, as s grows the curve moves down. This means that the area where link equilibria are sustainable *expands*, or, equivalently, that the area where it is profitable for the link source to deviate from the link equilibrium shrinks. To see why this happens, let us enumerate the reasons why the link source (say, site 2) might want to deviate from a link equilibrium. The first reason is independent of the presence of the aggregator: when ρ is small, the link source retains little revenue from its anchor traffic and is tempted to drop the link and compete head-on on content. The presence of the aggregator adds an additional motivation to deviate from a link equilibrium: By not placing a link, site 2 withholds information about site 1's superior quality from the aggregator. If the aggregator cannot figure out quality on its own (i.e. when s is low), in the absence of this link it would link to site 2 with higher probability. This would increase the expected traffic and revenue flowing into site 2 and might make deviation from the link equilibrium attractive. As s increases, the aggregator becomes more and more capable of identifying the site with the best content on its own. Deviation from the link equilibrium will then not substantially change the expected traffic flowing into site 2 from the aggregator, reducing the attractiveness of such deviation.

When ρ is close to 1, changes in the aggregator's search parameter have a minimal effect on the parameter region where link equilibria are sustainable. This can be explained by observing that high ρ means a low fraction of traffic is flowing from the link source to the link target. The presence of the aggregator then has a small effect on both the equilibrium content investment of the link target, as well as on the signaling implications of site 2's decision to link or not link to site 1. For that reason, changes in the aggregator's ability to discern content quality have similarly small effects on site 2's strategic behavior.

6 Multiple sites

We have demonstrated the most important forces governing incentives to link using models with two or three sites. Here, we consider the case of $N > 2$ homogeneous sites which can simultaneously invest in content and link to each other. The preceding sections have examined the role of the outside option extensively, thus, to simplify the analysis, we assume $\mu = 0$ here. At the same time, we consider linking costs and assume that establishing a link costs $K_L > 0$. Our first observation regarding the link structure is that a site has no incentive to link to multiple other sites, since only the highest quality link target affects the source's attractiveness.[16] Therefore, each site has at most one outgoing link and only the sites with the highest content quality can have incoming links. This leads to a network structure where a number of sites have one outgoing link each (N_S link sources) to a set of N_T link targets. Note that all the link targets have to have the same level of content since

---

[16]If link creation is costless then sites are indifferent between placing one or multiple links. If we assume an arbitrarily small cost of adding a new link (e.g. the cost of writing the link summary) then sites in our setting strictly prefer placing a single link to the best available content. In this Section we focus on small, but positive costs for linking. We examine the case of higher linking costs in the Technical Appendix.

links only point to the highest content site(s). Furthermore, we show that, in equilibrium, all link targets have the same number of incoming links and that, as long as there is one link in the network, all sites have to have either an incoming or outgoing link.

Lemma 2. *In equilibrium, either there are no links or sites can be divided into N_S link sources and N_T link targets, where $N_S + N_T = N$, each link source has one outgoing link, and each link target has the same number (N_S/N_T) of incoming links.*

The preceding lemma narrows down the network structure but we still need to determine when such an equilibrium network is feasible. We are primarily interested in determining whether an equilibrium with links is a possible outcome or no links are formed. Recall that, when $\mu = 0$, a linking equilibrium was not possible in 2-node settings (Section 4.2). We show that when $N > 3$ linking equilibria are possible even when $\mu = 0$. The following proposition provides details.

Proposition 7. *There exist functions $\overline{NL}(\rho)$, $\underline{L}_1(\rho)$, and $\overline{L}_i(\rho)$ (for $i \geq 2$) such that the following equilibria are the only ones possible:*

1. *No linking is an equilibrium for any $K_L > 0$ iff $N \leq \overline{NL}(\rho)$, where $NL(\rho)$ is decreasing in N.*

2. *A linking equilibrium with $N_T = 1$ exists for some $K_L > 0$ iff $N \geq \underline{L}_1(\rho)$,*

3. *As $K_L \to 0$, a linking equilibrium with $N_T = i \geq 2$ exists if $2i \leq N \leq \overline{L}_i(\rho)$, where $\overline{L}_i(\rho)$ is decreasing in ρ and the only feasible networks are the following: $(N_T = 2; \ 4 \leq N \leq 14)$, $(N_T = 3; \ N = 6)$, $(N_T = 3; \ N = 9)$, $(N_T = 4; \ N = 8)$, $(N_T = 5; \ N = 10)$, $(N_T = 6; \ N = 12)$, $(N_T = 7; \ N = 14)$.*

4. *Content levels of sites that do not link, that are sources, and that are targets are respectively:*

$$c_{NL} = \frac{2N-1}{N^2}, \quad c_S = \frac{\rho}{N}, \quad c_T = \frac{(2N_T - 1)(\rho N_T + (1 - \rho)N)}{N \cdot N_T^2}$$

The results show that when N is sufficiently high, an equilibrium with a single target is possible, whereas an equilibrium without links is less likely to occur. Figure 7 depicts the regions where the different equilibria are possible as $K_L \to 0$. We focus on linking equilibria with a single target since the results indicate that equilibria with more than one target are limited to a few cases. Indeed when linking is cheap and $N > 7$, the only equilibrium with links has a single target. The intuition behind this result is that when there is more than one link target, they have to compete for anchor traffic and it is often appealing to give up being a link target by simply linking to one of the other targets.

Considering the outcome with no links and the outcome with one link target, we see that, as N increases, the different equilibria constitute a straightforward generalization of the basic case with two sites (Section 4.2): Linking generally occurs if ρ is high enough so that link sources can capture some of the traffic that they attract using the link target's content. However, if ρ is close to 1, link targets are discouraged by the lack of traffic through links leading to a lower content investment
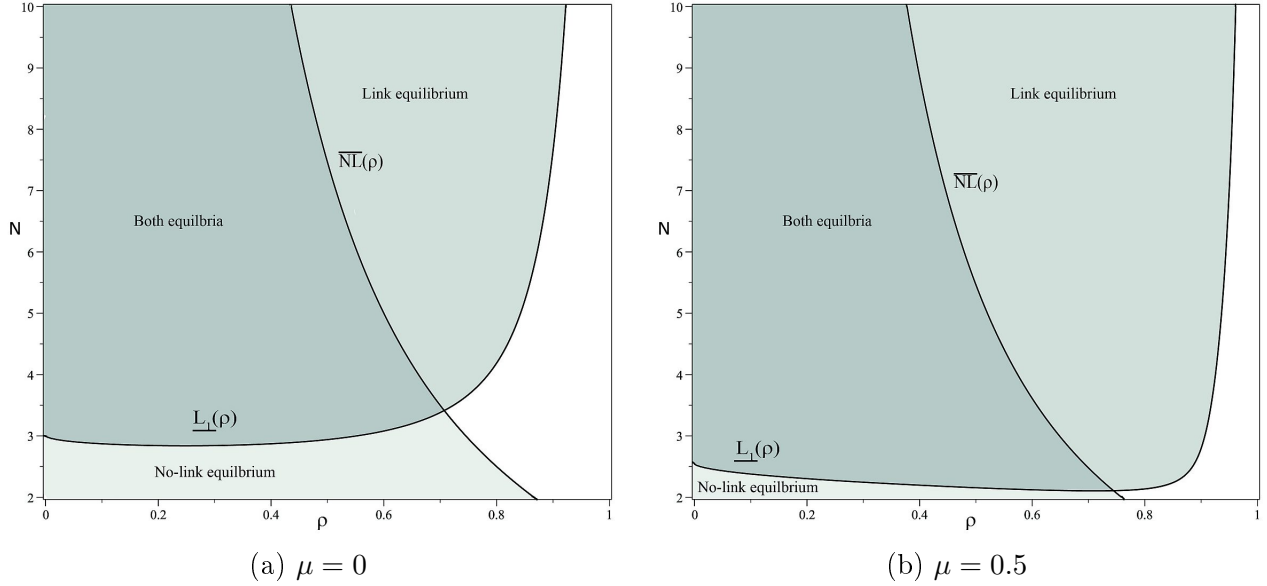
(a) $\mu = 0$ (b) $\mu = 0.5$

Figure 7: Feasible no-linking and single-target linking equilibria

and no links. The results also show that, as N increases, linking is more likely. This is due to weaker competition between sites, which has a similar effect to a stronger outside option. Although we assume $\mu = 0$ to derive the proposition, numerical analysis indicates that linking becomes more likely as μ increases just as in the basic model with two sites. [17]

7 Allowing targets to veto links

The dominant practice in today's Internet is that links are formed unilaterally and that link targets have no say about another site creating a link that points to them. This is the mechanism that we have assumed so far when analyzing the effects of links on content investments and profits. However, there are technical and legal means by which a target can refuse a link or indicate that it does not desire to be a link target. Examples include news sites (such as the Wall Street Journal) that do not display some articles to visitors who do not reach such articles via the site's front page or a paywall. Other sites do not allow search engines to crawl their content, effectively refusing to receive links from a large aggregator. It is not surprising that such practices are becoming prevalent in light of complaints of content creators about unfair linking practices.[18] It is, thus, worthwhile to assess the economic implications of this link refusal policy.

In this section we explore how our results change in settings where link formation requires the

---

[17]Increasing the cost of linking changes the equilibrium patterns somewhat. As K_L increases, the region where no linking is an equilibrium expands and the region with a single link target shrinks. In fact, when $K_L > 1/32$, no linking is always an equilibrium. The regions where linking with multiple targets is possible changes in a more complex manner. If linking is costly enough $\left(K_L > \frac{1}{2}\left(\frac{N_T - 1}{N_T^2}\right)^2 \right)$, then an equilibrium with N_T link targets always exists when N is high. We provide more details in the Technical Appendix.

[18]See, for example "Rupert Murdoch Begins Blocking News Aggregators, Search Engines", January, 9, 2010, http://www.mediaite.com/online/rupert-murdoch-begins-blocking-new-aggregators-search-engines/

agreement of both the source and the target. Our main result shows that the increased competition among content sites that is induced by the presence of aggregators makes it less likely that content nodes will unilaterally refuse links coming from aggregators, suggesting that, in most practical settings, the technical ability to veto links is somewhat of a moot point.

We consider a setup with an aggregator and two incumbents. The setup is equivalent to that of Section 5.2 except that the two content creators have the ability to refuse links. We assume that after the content decisions have been made, the two sites simultaneously decide whether to allow or refuse a potential link from the aggregator. According to our model, the aggregator will create at most one link. Thus, if both sites allow links the aggregator will choose one of them with probabilities that depend on the sites' content and the aggregator's search parameter s (see Section 5.2.1). If only one site allows links the aggregator will link to it. If neither site accepts links, the aggregator will not be able to link and will attract zero traffic. We assume that $\mu = 0$ in order to examine the system in settings where there are high incentives to refuse links.[19] We determine the subgame-perfect Nash-equilibria that are symmetric in content choice.

Proposition 8. *There always exists an equilibrium in which neither site refuses links and the content levels are*

$$c_1^* = c_2^* = c_{NL} = \frac{(3s+8)(1-\rho)+18}{36}.$$

This equilibrium is unique iff $\rho < 1/2$. Otherwise, there exists a second equilibrium in which both sites refuse the link from the aggregator and build a content level of $3/8$. Profits in the latter equilibrium are always higher.

If ρ is not very high, then, at equilibrium, neither site refuses aggregator links. In contrast, when ρ is high, sites have an incentive to refuse the link from the aggregator as it attracts traffic away from content creators and sends little traffic back to them. However, sites have to coordinate so that they both refuse the links. If one site does allow the aggregator to link to it, the other site is under pressure to allow the link as well. Note that when refusal is a possible outcome, sites make higher profits when they can coordinate to refuse links since, for $\mu = 0$, the presence of the aggregator is a net negative on the content sites (see Section 5.1).

8 Managerial implications and research opportunities

This paper is the first to take a comprehensive look at the economic implications of free hyperlinks in settings where content sites compete for traffic and revenue and are, thus, inclined to make *interrelated strategic investments in both content and links*. Our models have produced a number of insights of relevance to industry practitioners, which, in turn suggest several opportunities for future research:

1. Links among peer content creators can increase joint profits and content quality. One of the most disruptive effects of the Internet to content industries has been the elimination of geographical

---

[19]When $\mu = 0$ aggregators always generate revenue at the content creator's expense (See Section 5.1). When $\mu > 0$, the results are similar, but link refusal is even less likely.

monopolies; all content sites now directly compete for every consumer. Direct competition induces what, from the perspective of sites, are excessive investments in duplicate content. This hurts everyone's profits without substantially benefiting the quality of content that becomes available to consumers. We show (Sections 4.2 and 6) that linking allows sets of roughly equally capable sites to coordinate content production in ways that increase their joint profits as well as consumer utility. In such an equilibrium, one site invests in high quality content and all other sites link to this site. Expecting to receive both anchor and link traffic, the site that invests in content is usually able to produce much higher quality content than in a no-link equilibrium. Through links, this content becomes available to all consumers, increasing the attractiveness of the entire content ecosystem and drawing additional visitors away from outside alternatives. Although we derived this result in a simplified setting where we assumed a single topic, one can envision an extension of this result in a more realistic, multiple-topic setting where each site specializes in producing content on some topics and sites link to each other to provide their readers with coverage of the remaining topics. Our results suggest that such an industry structure, built on complementary content production, would be better, for firms as well as for consumers, than the current situation where a number of large media organizations are trying to be all things to all people.

2. *Linking can sustain market entry of inefficient players.* We find (Section 4.3) that the ability to place free hyperlinks allows inefficient players, who would otherwise not be viable, to remain in the market by free-riding on the content of efficient sites. If the amount of revenue that is retained by link source nodes is substantial, this free-riding lowers incentives for quality content production and represents an argument against the culture of barrier-free unilateral linking decisions that currently pervades the web (see, for example http://www.right2link.org/). It has been argued that content sites have the technical means to avoid this by selectively refusing incoming links from substantially less efficient competitors. However, the analysis of Section 7 suggests that, unless all competing content sites commit to such selective refusal policies, competitive pressures often force them to accept all incoming links. This underlies the need for research on better mechanisms and policies around the use of hyperlinks that allow content sites to harness the positive consequences of hyperlinks (e.g. coordination among peers, avoidance of unnecessary duplicate effort) while minimizing their negative consequences (free riding by less capable competitors).

3. *The main benefit of aggregators to content creators comes from traffic expansion.* Under the natural assumption that content aggregators form links to the best available content, their presence makes it easier for consumers to access good content, and increases the attractiveness of the entire content ecosystem. To the extent that there exists an outside alternative that the focal content ecosystem competes with, aggregators increase the total traffic flowing into the content ecosystem. Most of that new traffic is directed to the highest quality content sites, increasing their profits (Section 5.1).

4. *The presence of aggregators incurs industry costs that must not be overlooked.* Aggregators represent an additional type of node that did not exist in traditional content ecosystems. Their market entry inevitably appropriates some of the attention and revenue that would otherwise be

shared among content creators. Their net effect is positive for content creators only if the traffic expansion they induce is sufficient to offset the loss of attention and advertising revenue (Section 5.1). An interesting and timely avenue for future research would be to provide an empirical assessment of the relative magnitudes of the two effects (traffic expansion, retention of ad revenue) of well-known aggregators. Such an assessment will be a particularly important input in the current debate between aggregators and content creators. Our work has laid the theoretical foundations along which such an assessment can be made.

5. Aggregators increase competition among content sites. In most cases aggregators place links to a subset of the available content (the "best" content). Since links drive traffic to their target nodes, this creates competition among content nodes. Such competition induces them to produce better content but the impact on profits is negative (Section 5.2). This interesting second order effect of aggregators that our study has uncovered presents an opportunity for a closer empirical investigation.

We close by noting that this work is a first step toward understanding an arguably under-researched area. Despite the richness of its insights, our analysis has only scratched the surface of the full complexity of strategically formed content networks. In order to capture the fundamental strategic processes at play we focused on relatively simple settings with a single content topic and homogeneous consumers. We also abstracted away the search costs borne by content nodes and aggregators in order to discover content to which they might want to form links. Last, but not least, we restricted our attention to links that are not accompanied by side payments between the link source and target. Although the current legal regime does not require the link source to pay the link target (or vice versa), sponsored or paid links are commonplace in many settings and an important source of revenue for companies like Google and Yahoo (Katona and Sarvary 2008). Understanding how the option of side payments affects the incentives to produce content and place links in content industries is an interesting question for future research. There are several opportunities for our results to be extended to larger networks with more realistic features (e.g. multiple topics, consumers with heterogeneous tastes, costly search for content by aggregators, etc.) and to study the implications of the identified strategic interactions on the structure of the emergent content networks where the creation and deletion of nodes themselves is endogenous on their ability to compete.

References

Alba, J., Lynch, J. Weitz, B., Janiszewski, C., Lutz, R. Sawyer, A., and Wood, S. (1997) Interactive Home Shopping: Consumer, Retailer and Manufacturer Incentives to Participate in Electronic Marketplaces. Journal of Marketing 61(3):38-53

Bakos, Y. (1997) Reducing Buyer Search Costs: Implications for Electronic Marketplaces. Management Science 43 (12), 1676-1692.

Bala V. and Goyal. S. (2000) A non-cooperative model of network formation. Econometrica 68, 1181-1229.

Bloch, F. and Dutta, B. (2009) Communication networks with endogenous link strength. Games and Economic Behavior 66 (1), 39-56.

Blood, R. (2002) The Weblog Handbook: Practical Advice on Creating and Maintaining Your Blog. Cambridge MA: Perseus Publishing.

Bramoulle, Y. and Kranton, R. (2007) Public goods in networks. Journal of Economic Theory, 135 (1) 478-494.

Chiou, L. and Tucker, C. (2011) News and Online Aggregators. Working Paper.

Dewan, R. M., Freimer, M. L., Seidmann, A. and Zhang, J. (2004) Web portals: Evidence and analysis of media concentration. J. Mgmt Info. Sys., 21, 2, 2004, 181-199.

Dubins, L. E., and Savage, L. J. (1965) How to gamble if you must: Inequalities for stochastic processes. McGraw-Hill.

FTC (2010). Potential Policy Recommendations to Support the Reinvention of Journalism. Federal Trade Commission Staff Discussion Draft. (http://www.ftc.gov/opp/workshops/news/jun15/docs/new-staff-discussion.pdf.)

Galeotti, A. and Goyal, S. (2010) The Law of the Few. American Economic Review, 100 (4), 1468–92

Galeotti, A., Goyal, S., Jackson, M.O., Vega-Redondo, F. and Yariv, L. (2006) Network Games, European University Institute Working Paper ECO 2008/07. Huberman B.A. and Adamic, L.A. (1999) Growth dynamics of the World Wide Web. Nature 401:131.

Huberman, B.A., Pirolli, P., Pitkow, J. and Lukose, R. (1998) Strong regularities in world wide web surfing. Science 280:94-97.

Jackson, M.O. (2008) Social and Economic Networks. Princeton University Press.

Jarvis, J. (2008) The link economy vs. the content economy. Buzzmachine blog, June 18, 2008. (http://www.buzzmachine.com)

Johnson, E. J., Moe, W. M., Fader, P. S., Bellman, S. and Lohse, G. L. (2004) On the Depth and Dynamics of Online Search Behavior. Management Science. 50(3): 299-308.

Karp, S. (2007) The Web's Link-Driven Attention Economy. Publishing 2.0, December 15, 2007. (http://publishing2.com/)

Katona, Z. and Sarvary, M. (2008) Network Formation and the Structure of the Commercial World Wide Web. Marketing Science 27(5) 764-778

Katz, M. L. and Shapiro, C. (1994) Systems Competition and Network Effects. The Journal of Economic Perspectives 8 (2), 93-115

Kleinberg, J., Suri, S., Tardos, E. and Wexler, T. (2008) Strategic Network Formation with Structural Holes. ACM Conference in Electronic Commerce (ACM EC).

Kominers, S.D. (2009) Sticky content and the structure of the commercial web. 2009 Workshop on the Economics of Networks, Systems and Computation (NetEcon'09).

Lal, R. and Sarvary, M. (1999) When and How is the Internet Likely to Decrease Price Competition. Marketing Science. 18(4):485-503.

Lynch, J. G. and Ariely, D. (2000) Wine Online: Search Costs Effect Competition on Price, Quality and Distribution. Marketing Sicence. 19(1): 83-103.

Ma, Liye (2010) A Dynamic Competitive Analysis of Content Production and Link Formation of Internet Content Developers, Working Paper, Carnegie Mellon University

Mayzlin, D. and Yoganarasimhan, H. (2010) Link to Success: How Blogs Build an Audience by Promoting Rivals. Working Paper, Yale School of Management.

Palme, E., Dellarocas, C., Calin, M, and Sutanto, J. (2012) Attention Allocation in Information-Rich Environments: The Case of News Aggregators. Working Paper.

Pennock, D. M., Flake G.W., Lawrence, S. Glover, E. J., Giles C. L. (2002) Winners don't take all: Characterizing the competition for links on the Web. Proc Nat Acad Sci (PNAS) USA, 103:12684-12689.

Purcell, K., Rainie, L., Mitchell, A., Rosenstiel, T., and Olmstead, K. (2010) Understanding the Participatory News Consumer, Pew Internet & American Life Project, Research Report.

Ratchford, B. T., Talukdar, D., and Lee, M.-S. (2001) A Model of Consumer Choice of the Internet as an Information Source. International Journal of Electronic Commerce. 5(3):7-21.

Steiner, P. O. (1952) Program Patterns and Preferences, and the Workability of Competition in Radio Broadcasting. The Quarterly Journal of Economics. 66(2): 194-223.

Tullock, G. (1980) Efficient rent seeking, in: J. Buchanan, R. Tollison and G. Tullock, (eds.), Towards a Theory of the Rent-Seeking Society, College Station, Texas A&M University Press, 97-112.

Wolinsky, A. (1984) Product Differentiation with Imperfect Information. The Review of Economic Studies 51 (1), 53-61.

Wu, F. and Huberman, B.A. (2008) Popularity, Novelty and Attention. Working Paper.

Proofs

Proof of Proposition 1: There are two possible types of equilibrium with respect to linking: (i) one where there is no link between the two sites and they invest equally in content (c_{NL}), and (ii) one where one site invests less in content and links to the other site. As we have already determined the potential equilibria of the first type, we will now identify the candidates for linking equilibria, then check when neither site has an incentive to deviate from a potential equilibrium. When site i links to site j, then its profit becomes $\pi_{i \to j} = \frac{c_j \rho}{c_j + c_j + \mu} c_i - \frac{1}{2} c_i^2$. Comparing these two yields that site i will link to site j iff $c_i \leq \frac{\rho c_j (c_j + \mu)}{(1 + (1-\rho))c_j + \mu}$.

Note that the right hand side of the above equation is increasing in c_j and always less than or equal to c_j, yielding that *only the lower quality site will establish a link and only if its quality is sufficiently low relative to its competitor*. Given the above described linking behavior, sites will choose their content investments to maximize profits. Although the site that ends up with a higher content does not consider linking, its profit function changes if its low content competitor decides to link to it: $\pi_{j \leftarrow i} = \frac{c_j + (1-\rho)c_j}{c_j + c_j + \mu} c_j - \frac{1}{2} c_j^2$. Differentiating $\pi_{j \leftarrow i}$ with respect to c_j yields that site j will invest c_T in content if site i links to it (as given in the proposition). Then, differentiating $\pi_{i \to j}$ with respect to c_i yields that site i will invest $b_{i \to j}(c_j) = \frac{\rho c_j}{(c_j + c_j + \mu)k}$ in content if it links to j, yielding the stated $c_i = c_S$ if we plug $c_j = c_T$.

To check whether sites have no incentives to deviate from the potential equilibria, we examine whether the no linking best response would yield higher profits in the linking case and whether the linking best response would yield higher profits in the no-link case. In the first case, the linking equilibrium holds iff $\pi_i(b_i(c_T), c_T) \leq \pi_S := \pi_{i \to j}(c_S, c_T)$.

Let $\underline{L}(\rho)$ denote the value of μ where the above holds with equality when $k = 1$. The above inequality holds for high values of μ, yielding that the linking equilibria exists iff $\mu \geq \underline{L}(\delta, \rho)$. Similarly, let $\overline{NL}(\rho)$ denote the value of μ for which $\pi_i(c_{NL}, c_{NL}) = \pi_{i \to j}(b_{i \to j}(c_{NL})_S, c_{NL})$. Sites do not have an incentive to deviate from the no-link equilibrium iff $\mu \leq \overline{NL}(\delta, \rho)$.

Proof of Corollary 1: It is useful to start with examining the comparative statics with respect to ρ. One can check that $c_S(\rho)$ is increasing in ρ, yielding that $\pi_S(\rho)$ is also increasing. Then it is

enough to check that $c_S(1) \le c_P$, yielding $c_S(\rho) \le c_P$ for any ρ. Similarly, one can check that $c_T(\rho)$ is decreasing in ρ, and that $c_P \le c_S(1)$. For the results depending on whether ρ is low or high, we check the derivatives for $\rho = 0$ and $\rho = 1$ and get the stated results. Since all the functions are continuously differentiable, we get the same results for a region of small values of ρ as for $\rho = 0$ and the same results for a region of high values as for $\rho = 1$.

Proof of Proposition 2: If there are no links, then sites maximize the profit function described in (4), yielding:

$$
\begin{bmatrix} c_1^{NL}, \ c_2^{NL} \end{bmatrix} = \begin{cases} [0 & , \ \frac{1}{2k} - \mu + \frac{\sqrt{4k\mu+1}}{2k}] & \text{if } k < \frac{2+\mu}{4} \\ [\frac{(2k-1)(c_2^{NL}+\mu)-\mu}{2-k} & , \ \frac{6-k(2\mu(k+1)+3)+\sqrt{(2-k)^2(4\mu(k+1)+9)}}{2(k+1)^2}] & \text{if } \frac{2+\mu}{4} \le k \le \frac{-1+\sqrt{1+4\mu}}{\mu} \\ [\frac{1}{2} - \mu + \frac{\sqrt{4\mu+1}}{2} & , \ 0] & \text{if } k > \frac{-1+\sqrt{1+4\mu}}{\mu} \end{cases}
$$
(10)

Equation (10) shows that, for very small and very large k, that is, when the ratio of the high to low-cost producer rises above a threshold, only the low-cost producer can stay in the market.

Proof of Proposition 3: The proof follows the exact same steps as the proof of Proposition 1. However, due to the asymmetric cost, the profits and the best responses are different for the two sites, yielding the two different thresholds for the linking equilibrium.

Proof of Proposition 4: When an aggregator enters the ecosystem in which two sites produce the same amount of content and do not link to each other, the aggregator will link to one of the two with equal probability. The aggregator will therefore have $z_3 = c_1$. The expected anchor traffic of a content producer will decrease from $\frac{c_1}{2c_1+\mu}$ to $\frac{c_1}{3c_1+\mu}$. However, they will receive some traffic through the link from the aggregator in the amount of $\left(\frac{1-\rho}{2}\right)\frac{c_1}{3c_1+\mu}$, yielding a total traffic of $\left(1 + \frac{1-\rho}{2}\right)\frac{c_1}{3c_1+\mu}$ which is higher than $\frac{c_1}{2c_1+\mu}$ iff $\rho < \frac{\mu}{2c_1+\mu}$. Since content decisions are fixed in this setting a higher traffic is equivalent to higher profits, completing the proof of part 1. In case of an aggregator entering a market in which a low content site links to a high content site, the aggregator will link to the higher content site to maximize the utility consumers can expect. Similarly to the previous case, we can determine how the amount of traffic changes at the two sites. Before the aggregator enters, sites 1 and 2 receive traffic of $\frac{(2-\rho)c_1}{2c_1+\mu}$ and $\frac{\rho c_1}{2c_1+\mu}$, respectively. When the aggregator enters, these change to $\frac{(1+2(1-\rho))c_1}{3c_1+\mu}$ and $\frac{\rho c_1}{3c_1+\mu}$. Determining the sign of the change in traffic for the two sites and comparing the profits yields the thresholds for ρ.

Proof of Lemma 1: We differentiate site 1's profit function with respect to c_1. We note that the profit function is concave, thus the f.o.c. provides the maximum. Since, we are searching for a symmetric equilibrium, it is enough to solve $\frac{\partial \pi_1(x,x)}{\partial c_1} = 0$ and obtain

$$
c_1^* = c_2^* = \frac{13}{36} + s\frac{1-\rho}{24} - \frac{3\mu+\rho}{9} + \frac{\sqrt{9(1-\rho)^2s^2+12(1-\rho)(13-4\rho)s+64\rho^2-(192\mu+416)\rho+480\mu+676}}{72}.
$$

Plugging into the profit function yields the equilibrium profits. When profits would be negative (ρ is below the stated threshold), sites do not invest in content.

Proof of Proposition 5: Straightforward analysis of the expressions derived in Lemma 1 show

the relationship between the equilibrium content profit and the model parameters.

Proof of Proposition 6: We start by proving part 2. In the possible equilibrium where the lower quality content creator and the aggregator both link to the higher quality content site, (9) describes the payoff the source and the target. Differentiating π_T with respect to c_T yields that the target will invest c_T^* in content (as given in the proposition). Then, differentiating π_S with respect to c_S yields the expression for c_S^*. To prove the remaining, as in the proof of proposition 1, let $\underline{L}(\rho, s)$ denote the value of μ where $\pi_i(b_i(c_T^*), c_T^*) \leq \pi_S(c_S^*, c_T^*)$ holds and similarly, let $\overline{NL}(\rho, s)$ denote the value of μ for which $\pi_i(c_{NL}, c_{NL}) = \pi_S(b_S(c_{NL}), c_{NL})$, using the expression given in (7) for $\pi_i(c_i, c_j)$.

Proof of Lemma 2: As shown in the discussion preceding the Lemma, each site has at most one outgoing link. The N_S sites that have outgoing links do not have incoming links and the N_T sites that have incoming links all have the same level of content. First, we show that all the link sources have the same level of content. Since their attractiveness is determined by the quality of the site they link to, they all attract the same amount of anchor traffic. Furthermore, their profit function does not depend on the content level of other link sources, therefore they have the same best response (we will see that the functional forms yield a unique best response). Given that all link sources attract the same amount of traffic, we can show that each link target has the same number of incoming links. Assuming that one of the link targets (T_1) has more incoming links than another one (T_2) would yield a different best response for T_1 than for T_2 since T_1 would have more incoming link traffic., but a different content level would imply that T_1 and T_2 cannot be both link targets.

Proof of Proposition 7: We first determine the equilibrium content level when there are no links. Sites maximize $\pi_i = \frac{x^2}{x + (N-1)c_{NL}^*} - \frac{x^2}{2}$. Differentiating yields $\frac{2}{N} - \frac{1}{N^2} = c_{NL}^*$, leading to profits $\pi_{NL}^* = \frac{2N-1}{2N^4}$. For an equilibrium with no links to be feasible we have to check that no site has an incentive to deviate by linking to another site. The optimal profit to be made in this case would be $\pi' = \frac{\rho^2}{2N^2}$. Let $\overline{NL}(\rho) = \frac{1 + \sqrt{1 - \rho^2}}{\rho^2}$. It is exactly below this threshold that a deviation is not profitable as $K_L \to 0$. As K_L increases, this threshold increases as well. In order to determine the feasibility of equilibria with links, we examine the profit function that all the N_T link targets are maximizing:

$$\pi_T = \frac{x^2 \left(1 + \frac{(1-\rho)(N - N_T)}{N_T}\right)}{xN/N_T + c_T^*(N - N/N_T)} - \frac{x^2}{2},$$

where c_T^* is the equilibrium content level that all targets set. Differentiating and setting to zero yields $c_T^* = \frac{(N(1-\rho) + \rho N_T)(2N_T - 1)}{N \cdot N_T^2}$. Link sources simply maximize $\pi_S = \frac{\rho x}{N} - \frac{x^2}{2}$, yielding $c_S = \frac{\rho}{N}$. Possible deviations are different when $N_T = 1$ and when $N_T > 1$. When there is only one link target, it cannot profitably deviate by linking to one of the sources, since they have low very low content. We only have to consider whether sources want to give up their link and compete directly in content. The threshold below which this is not profitable as $K_L \to 0$ is denoted by $\underline{L_1}(\rho)$. Note that this deviation is always profitable for $N = 2$, thus $\underline{L_1}(\rho) > 2$, but for certain ρ values it is less than 3. When $N_T > 1$, we need to check an additional deviation, that is, when one link target decides to become a link source and link to another link target. This deviation is always more profitable than the one discussed before, restricting the regions where an equilibrium with $N_T > 1$ is possible. The

detailed analysis is covered in the Technical Appendix.

Proof of Proposition 8: We start by analyzing the last stage of the game in which sites decide whether or not to refuse a link from an aggregator. Since we are looking for equilibria that are symmetric in content, we can assume that the aggregator would link to both sites with equal probability if its link is not refused by either one. Therefore, if site j decides to allow a link, the aggregator will already have an accumulated content to attract traffic away and as long as $\rho < 1$, site i can only benefit from also allowing a link. That is, both sites allowing linking is always an equilibrium of the subgame. If site j decides to refuse the link, site i has two options. If it also refuses the link its revenue from traffic is $c_i/2$, whereas if it allows the link its revenue becomes $\frac{2-\rho}{3}c_i$. It is easy to check that the former is greater iff $\rho \geq 1/2$, making the refusal-refusal setting an equilibrium of the subgame. One can then determine the equilibrium content levels in the two cases using the results of proposition 1 and lemma 1.