NET Institute*

www.NETinst.org

Working Paper #10-13

September 2010

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

Chrysanthos Dellarocas
Boston University

Zsolt Katona
University of California at Berkeley

William Rand
University of Maryland

* The Networks, Electronic Commerce, and Telecommunications ("NET") Institute, http://www.NETinst.org, is a non-profit institution devoted to research on network industries, electronic commerce, telecommunications, the Internet, "virtual networks" comprised of computers that share the same technical standard or operating system, and on network issues in general.

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

Chrysanthos Dellarocas, Boston University (dell@bu.edu)
Zsolt Katona, University of California at Berkeley(zskatona@haas.berkeley.edu)
William Rand, University of Maryland (wrand@umd.edu)

30th September 2010

Abstract

A key property of the World Wide Web is the possibility for firms to place virtually costless links to third-party content as a substitute or complement to their own content. This ability to hyperlink has enabled new types of players, such as search engines and content aggregators, to successfully enter content ecosystems, attracting traffic and revenues by hosting links to the content of others. This, in turn, has sparked a heated controversy between content producers and aggregators regarding the legitimacy and social costs/benefits of uninhibited free linking. This work is the first to model the implications of *interrelated and strategic* hyper-linking and content investments. Our results provide a nuanced view of the much-touted "link economy", highlighting both the beneficial consequences and the drawbacks of free hyperlinks for content producers and consumers. We show that content sites can reduce competition and improve profits by forming links to each other; in such networks one site makes high investments in content and other sites link to it. Interestingly, competitive dynamics often preclude the formation of link networks, even in settings where they would improve everyone's profits. Furthermore, such networks improve economic efficiency only when all members have similar abilities to produce content; otherwise the less capable nodes can free-ride on the content of the more capable nodes, reducing profits for the capable nodes as well as the average content quality available to consumers. Within these networks, aggregators have both positive and negative effects. By making it easier for consumers to access good quality content they increase the appeal of the entire content ecosystem relative to the alternatives. To the extent that this increases the total traffic flowing into the content ecosystem, aggregators can help increase the profits of the highest quality content sites. At the same time, however, the market entry of aggregators takes away some of the revenue that would otherwise go to content sites. Finally, by placing links to only a subset of available content, aggregators further increase competitive pressure on content sites. Interestingly, this can increase the likelihood that such sites will then attempt to alleviate the competitive pressure by forming link networks.

Acknowledgment. The authors gratefully acknowledge financial support for this research by Google/WPP and the NET-Institute (www.NETinst.org).

1 Introduction

The advent of the Internet has thrown content industries into turmoil. Traditional content creators, such as newspapers, are witnessing their geographical monopolies erode and some of their most important sources of revenue, such as classified ads, migrate to specialized online marketplaces like eBay and Craigslist. To make matters worse, content creators are facing stiff competition from user-generated substitutes, such as blogs, as well as from new types of players, such as search engines and other kinds of content aggregators, that attract traffic and advertising revenues by hosting collections of links to others' original content.

As traditional content organizations are seeing their business models collapse, some have vocally accused aggregators of "stealing" their revenues, lobbying for copyright reform that will make it more difficult to place links to third-party content without payment to the producers of that content.[1] Others point out that in today's "link economy" links bring valuable traffic to target nodes so content creators should be happy that aggregators exist, because without aggregators no one would find the content in the first place (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, without good content to point to, links (and aggregators) are worthless.[2] The debate is complex, lively and often emotional. It has already attracted the attention of governments and regulatory bodies. For instance, during 2009-2010 the FTC hosted three workshops on the Future of Journalism and has published a controversial "discussion draft" that hints of copyright reform and protection for newspapers from aggregators (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 producers.[3]

A central aspect of the debate focuses on the complex implications of the process of placing (for the most part) free hyperlinks across content nodes. Hyperlinks between websites have been a defining feature of the World Wide Web right from its inception. Nevertheless, their full economic implications on content ecosystems are not yet fully understood. Linking has fundamentally transformed the ways in which content sites must compete for attention and revenue. In networked settings the traffic a site attracts not only depends on the quality of its own content but also on the pattern of links that connect it to other sites. The traditional wisdom has been that incoming links bring traffic and higher search engine ranking to a site (given most modern search algorithms) and are, thus, desirable. Increasingly, however, content producers are realizing that outgoing links are a costless way of providing one's readers with easy access to third party content. A site can thus use links to maintain or increase its appeal to visitors by what some have labeled as "free riding" on others' efforts. But links also allow sites to coordinate and divide labor, thus avoiding an unnecessary du-

[1] The recent spat 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

[2] "CEO Eric Schmidt wishes he could rescue newspapers", *Fortune* January 7, 2009.

[3] "France plans 'Google tax' on internet searches". http://www.telegraph.co.uk/technology/google/6947706/France-plans-Google-tax-on-internet-searches.html

plication of effort, e.g. situations where multiple sites end up producing very similar content on the same stories. If the links are chosen well, then they point to good quality content; as a result, they reduce the search costs of the entire consumer population, which might lead to an aggregate increase in content consumption and to more traffic for higher quality sites. As of today, a full understanding of the complex implications of hyperlinks for a site's traffic, revenue and content strategy does not exist.

This work offers a step in the direction of understanding the *complex interplay between content and links* in settings where a set of nodes compete for traffic and make strategic investments in both content and links to maximize their revenues. We are interested in answering both micro level questions of interest to firms (e.g., what is the optimal way in which content organizations can combine content creation and hyperlinking to optimize revenue? how can content organizations best leverage and respond to the presence of aggregators?) as well as macro level questions of interest to regulators (e.g., what are the implications of strategic hyperlinking for quality of content that is available to consumers? are aggregators beneficial or harmful to society?).

Our models focus on a set of content sites (e.g. news sites) that generate revenue from user visits (e.g. through advertising) and who compete for traffic among themselves as well as with an outside alternative that represents alternative media (e.g. TV, twitter feeds, traditional print newspapers). 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 who (if anyone) to link to. Users want to maximize the utility they obtain per unit of attention and are, therefore, more likely to patronize sites that provide them with access to better total 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 have access to and, thus, the number of readers they attract. On the other hand, original content increases the 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.

To fully appreciate the economic implications of hyperlinking, it is necessary to first understand how the Internet changed the content creation industry. Traditionally print newspapers and other forms of content were distributed on a geographic basis: if you were in a certain area then you read the content produced by your local newspaper for that area. In most cases, there was one or few firms that controlled all of the content distributed to a particular area. The Internet removed this geographic monopoly of content sites. As a result, any consumer could choose to read any content creator's website (i.e., consumers were no longer limited to their local geographic monopoly), because of this, the individual firms went from being essentially monopolies to a world in which they had to directly compete with every other content creation firm. As a result, competition for audience became fierce, with two important consequences. First, all sites are forced to over invest in essentially substitute content since no site wants to be left behind; this duplication of effort hurts everyone's profits. Second, sites that are substantially less efficient (skilled) in producing content relative to the top competitors are forced out of the market as it is not economically viable for them to produce content at any quality.

Against this fiercely competitive backdrop we find that the ability to place hyperlinks has qualitatively different implications when all sites are roughly equal in their ability to produce content than when sites have substantial disparities. When all content sites are roughly equally efficient, the ability to form links across sites plays an economically and, in many cases, socially beneficial role. By reducing competitive pressures, since there is no incentive to out-compete a site's content investment when one can simply link to it, linking helps reduce inefficiencies that arise from overinvestment in content. In some cases, it allows one site to produce better content than in the no link case, content that everyone else links to and which, thus, becomes available to all consumers. This makes the entire content ecosystem more attractive to consumers. In the presence of an outside alternative this increases the total traffic flowing into the content ecosystem, compounding the benefit. In contrast, when some sites are substantially different in their ability to produce content, linking allows the less efficient competitors to free ride on the more efficient site's content and capture substantial market share. In a world without linking these less efficient competitors would be forced out of the market, which means that hyperlinks increase the number of sites competing for user attention; this reduces the efficient competitor's profits, meaning that at equilibrium it produces less quality content. In turn, lower content quality reduces the attractiveness of the entire content ecosystem relative to the outside alternative, exacerbating the negative effect.

We model aggregators as sites for whom content production is extremely costly and, as a result, produce virtually no original content and rely exclusively on links for attracting traffic. The rational behavior of an aggregator is, thus, to place links to the best available content. Our analysis shows that aggregators have three distinct effects on content ecosystems: First, they make it easier for consumers to access high quality content (since they link to it). As a result, they increase the average attractiveness of the entire content ecosystem relative to the outside alternative, which in turn attracts visitors away from the outside alternative. Such additional visitors are eventually directed (through links) to the best content sites. This increases the profits of the best content sites. Second, aggregators are nodes that would not be viable in a setting without links. Since they do not have any of their own content, in order to exist they must appropriate some advertising revenue that would otherwise go to the content sites. This effect reduces the profits of content creating sites and, in turn, can reduce the incentives of these sites to produce good content. Third, aggregators typically form links to only a subset of available content (the "best" content). Their presence then intensifies the competition among content sites to produce the "best" content and receive aggregator links. This increases the quality of available content, which is good for consumers, but in most cases further decreases the profits of content sites because it forces them to overproduce. Interestingly, however, the additional competitive pressure that is brought forth by the presence of aggregators makes it more likely that sites will form link equilibria to alleviate the pressure.

Our results provide a nuanced view of the much-touted "link economy", highlighting both the "good" and the "bad" implications of free hyperlinks. Our analysis shows that content sites can alleviate competition and improve profits by forming links if one site specializes on producing very high quality content and other sites agree to link to it. However, such networks are efficiency-improving only when all their members are of roughly equal efficiency (ability, skill). Links formed from less efficient to more efficient sites are a form of free-riding that harms both high quality content producers as well as consumers. If it was possible, high quality content sites should, therefore, consider selectively forming links with other high quality sites and refusing links that come from low quality sources.

Aggregators have both positive and negative consequences. By making it easier for consumers to find good quality content they increase the appeal of the entire content ecosystem, relative to alternative media. To the extent that this increases the total traffic flowing into the content ecosystem, aggregators can help increase the traffic and profits of the highest quality content sites (and induce them to produce even better content). At the same time, it is important to observe that aggregators are a new type of node that would not be viable in a setting without links. Their market entry inevitably takes away some of the advertising revenue that would otherwise go to content sites, and this is a negative impact that proponents of the "link economy" tend to downplay. Aggregators are beneficial to content sites only to the extent that their presence attracts sufficient additional visitors to the content ecosystem to offset the advertising revenue they appropriate from content creators. Finally, by selectively forming links to the best content, aggregators further increase the competitive pressure on content nodes. This is good for consumers but bad for content sites.

The rest of this paper is structured as follows. Section 2 discusses related work. Section 3 introduces our model. Section 4 derives our first set of results by analyzing a stylized setting with two content nodes. Section 5 distills the impact of aggregators in a setting where there are two content nodes and one aggregator. Section 6 discusses the implications of allowing link targets to refuse incoming links. Finally, section 7 concludes. All proofs are delegated to the Appendix.

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 adaptive 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 (2008) study a blogger's strategic decision to link or not link to a competitor's blog. Their models are focused on capturing a blogger's local link formation decision and do not attempt to analyze the system-level consequences of such decisions on network structure, content quality and social welfare.

Our results also contribute to the literature of strategic network formation. In the last ten years the network economics literature has emerged as an important and very vibrant subfield of economics (Jackson 2008). This literature has made important advances in our understanding of the processes of strategic network formation. However, 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 2008; Kleinberg et al. 2008). 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). With the exception of Cabrales et al. (2007) and Galeotti and Goyal (2007), who study very special cases of simultaneous node-level effort and link formation decisions, such models have received no attention in the network economics literature.

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 there is only one topic of interest and that, each period, site $i = 1, .., N$ produces content of quality $c_i \geq 0$ on this topic. In addition, sites can place links to the content of other sites. 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.[4] 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.

[4] Currently 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.

Users and Traffic. A population of users/consumers visits content sites and derives utility from reading their content. 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 have a small number of sites from which they start their news consumption (Purcell, 2010). Once a consumer arrives at her anchor site, she spends some time on the site reading its content or clicks on one or more links and reads the content at the link target sites. 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 basic 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 sites $j_1, ..., j_l$, consumers can expect to gain utility $z_i = \max(c_i, \delta c_{j_1}, ..., \delta c_{j_l})$ 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.

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 a more general form such users might combine random browsing, the use of search engines, recommendations from one's social network, etc. Regardless, as a result of this process, 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 producers. If we make the assumption that every period some consumers leave the ecosystem, i.e., switching to alternative forms of content consumption, whereas an equal number of new (uninformed) consumers enter, 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:

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

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 or by exchanging Twitter messages instead of reading the news). The presence of an *outside alternative* plays an important, yet subtle, role in our models. The most obvious impact of the outside alternative is that it reduces the total traffic $t^A = \sum_{i=1}^{N} t_i^A$ that flows into our content ecosystem; the higher the μ the lower the traffic. However, the presence of the outside alternative also implies that sites within the content ecosystem are not only in competition with one another but also with the outside alternative. Therefore, if a site increases the utility z_i it offers to users, not only will it attract users away from other sites but also from the outside alternative. In fact, if a site does increase its utility, z_i, the above specification implies that, the higher the μ, the higher the percentage of additional traffic 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 a site's change in 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 enables them to form linking equilibria that improve the profits of some or all content sites.

Site Revenues. Sites make money from advertising; their revenue 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:

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

The situation changes if we allow sites to place links to each other. Suppose that there is a 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 the link and consumes site j's content without consuming site i's content. We can justify such consumer behavior as follows: Consumers want to maximize the utility they obtain from content while minimizing the cognitive cost they must expend. We assume that consumers exhibit some degree of heterogeneity with respect to what aspects of an article fulfill their needs. For example, when it comes to a article about a soccer game, some consumers only care about the final score, and so they do not click the link, whereas other consumers want a detailed description of how well their team played, and therefore do click the link. As a result, a fraction ρ of consumers assesses that the combination of site i's own content plus the link description can satisfy their needs just as well as if they clicked the link; these consumers stay at the anchor site and do not click the link. The remaining consumers make the opposite assessment and click the link immediately. Recalling that $c_j > c_i$, another way of stating the argument is that, for a fraction ρ of consumers, the link description captures the crucial information that is contained in the link target's content (relative to the content of the link source). To keep things as simple as possible we assume that factor ρ is a constant.[5] The magnitude of ρ depends on

[5] Factor ρ can arguably also depend on the difference between the content quality offered by sites i and j. We

the content topic, as well as on the amount of information contained in the link description. The more information a site provides about the content of sites it links to, the higher the ρ. The limiting case $\rho = 1$ models settings where site i reproduces all salient aspects of the content of site j.

The preceding discussion shows that linking has both advantages and disadvantages. From the perspective of the link source, 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, leaving no revenue to the source site. On the other hand, 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 $\delta c_j > c_i$ allows site i to increase its expected utility to $z_i' = \max(c_i, \delta c_j) = \delta c_j > z_i$. By (1), higher utility implies higher anchor traffic.

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, thus, its anchor traffic. The number of visitors that reach j through a link from i to j is equal to the traffic $(1 - \rho)t_i^A$ leaving site i. Thus, the total incoming *link traffic* of site j is:

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

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 presence of links then becomes:

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

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.

experimented with such modeling specifications and concluded that they introduced substantial additional complexity without offering fundamental new insights. We chose to stay with the simplest possible model.

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.

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 it is $z_i = c_i$, $t_i^A = c_i/(c_i + c_j + \mu)$, $i = 1, 2$. Site i's 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 $c_T \leq c_S/\delta$ does not change the utility $z_S = \max(c_S, \delta c_T)$ consumers get from making site S their anchor, therefore does not help 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 one with (strictly) lower content will link to its peer. Under these assumptions, when site S links to site T it will be $c_S < \delta c_T$, $z_S = \delta c_T$, $z_T = c_T$, which, by (1), implies $t_S^A = \delta c_T/(\delta c_T + c_T + \mu)$ and $t_T^A = c_T/(\delta c_T + c_T + \mu)$. From (3) the corresponding payoff functions then take the form:

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

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

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 $t_1^A = t_2^A = \frac{1}{2}$ and

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

When $k > 2/\mu$, or, equivalently, $k\mu > 2$, this expression becomes negative. Given the competition they are facing from the outside option, for large k it is too expensive for sites to invest in even a little bit of content. We will thus assume $k\mu < 2$ 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:

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

The above equilibrium corresponds to a setting where a single monopolist owns both sites. The following benchmark result ensues:

Proposition 1. *When two homogeneous content sites compete for the same audience they overproduce content and end up with lower profits than in a setting where a single monopolist owns both sites.*

The above result captures an important consequence of Internet technologies for the press and for other, previously geographically segregated, content-based businesses. Sites that previously had monopoly power over their respective audience segments now have to compete with each other for the entire audience. Competition induces overinvestment in content and, even if it does not change relative market shares, results in reduced profits for all sites.

We will show that under certain circumstances sites can use the option of linking to alleviate counterproductive overinvestment in content. The following proposition characterizes the form of the resulting equilibria.

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

1. *If* $k\mu \leq \overline{NL}(\delta, \rho)$ *then sites do not establish links in equilibrium and* $c_i^* = c_j^* = c_{NL}$.
2. *If* $k\mu \geq \underline{L}(\delta, \rho)$ *then there are two asymmetric equilibria where one site links to the other and*

$$c_S = \frac{\delta\rho}{k(1+\delta)} \cdot \frac{1 + \delta(1-\rho) - 2k\mu + \sqrt{(1+\delta(1-\rho))^2 + 4(1+\delta(1-\rho))k\mu}}{1 + \delta(1-\rho) + \sqrt{(1+\delta(1-\rho))^2 + 4(1+\delta(1-\rho))k\mu}} \leq$$

$$\leq c_T = \frac{1}{k(1+\delta)} \cdot \frac{1 + \delta(1-\rho) - 2k\mu + \sqrt{(1+\delta(1-\rho))^2 + 4(1+\delta(1-\rho))k\mu}}{2}.$$

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

Figure 1 depicts the equilibrium regions when $\delta = 1$. We observe that when ρ and $k\mu$ are small then the only equilibrium is the symmetric no-link equilibrium. When ρ and $k\mu$ are large the only equilibria are the asymmetric linking equilibria. There is a parameter region where both equilibria co-exist as well as a small region where no equilibrium in pure strategies exists.

Fixing k, the proposition tells us that if the two sites are competing against a sufficiently strong outside option (μ) then there exist equilibria where one site will link to the other and will produce less content than its competitor. Recall that the presence of an outside option ($\mu > 0$) implies that

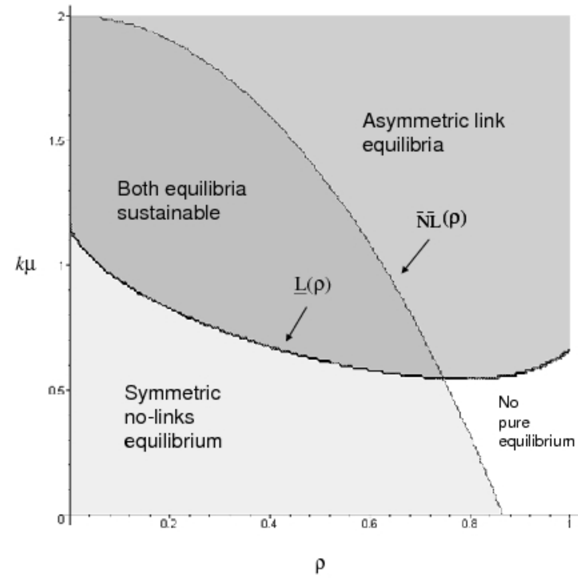


Figure 1: Equilibrium regions when the two content sites have identical cost parameters and $\delta = 1$.

an increase in a site's content investment not only attracts visitors from the other site but also from the outside option. In fact, the stronger the outside option the larger the fraction of new visitors that comes from the outside option (versus from the other site).

The intuition and details of the linking equilibria are different when ρ is small and when ρ is large.

Small ρ. Recall that, for δ close to one, when one site links to the other then both sites offer roughly 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. This also implies that, in the presence of an outside option, 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 option will be split between the link source and target. When ρ is small, many of the visitors of the link source 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 option 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 option 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 link target's better 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 remain a link source, albeit with lower content and lower profits than the link target, than to attempt to compete head-on with the link target on content.

Figure 2 depicts the equilibrium content and payoffs for $k = 1$, $\delta = 1$ and $\rho = 0.5$. For these

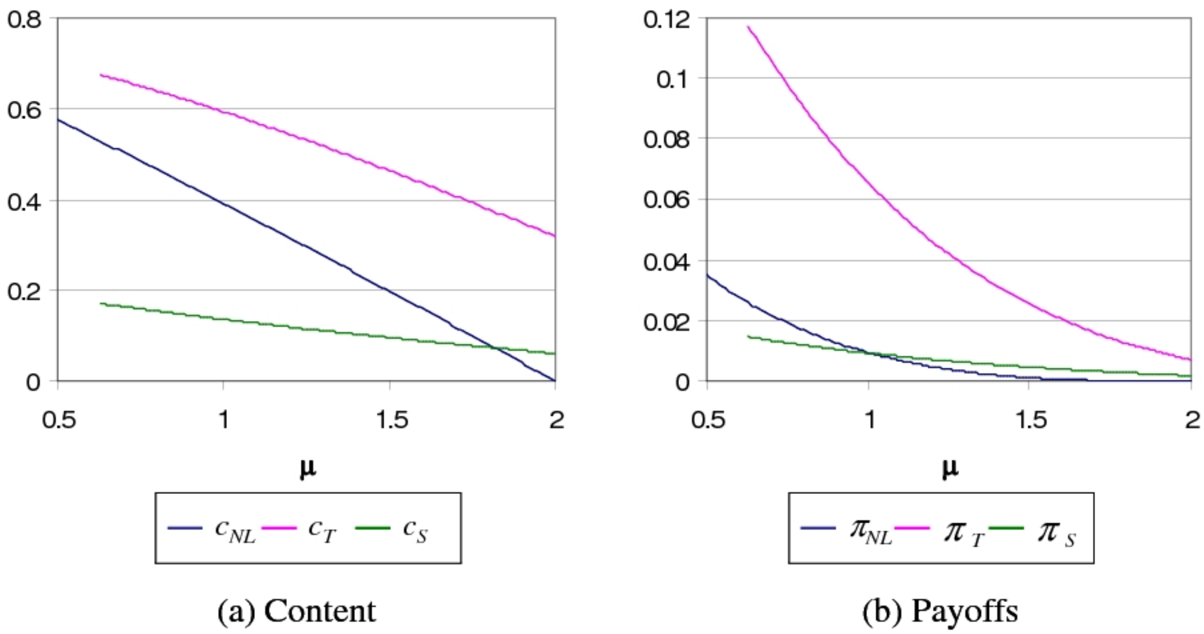
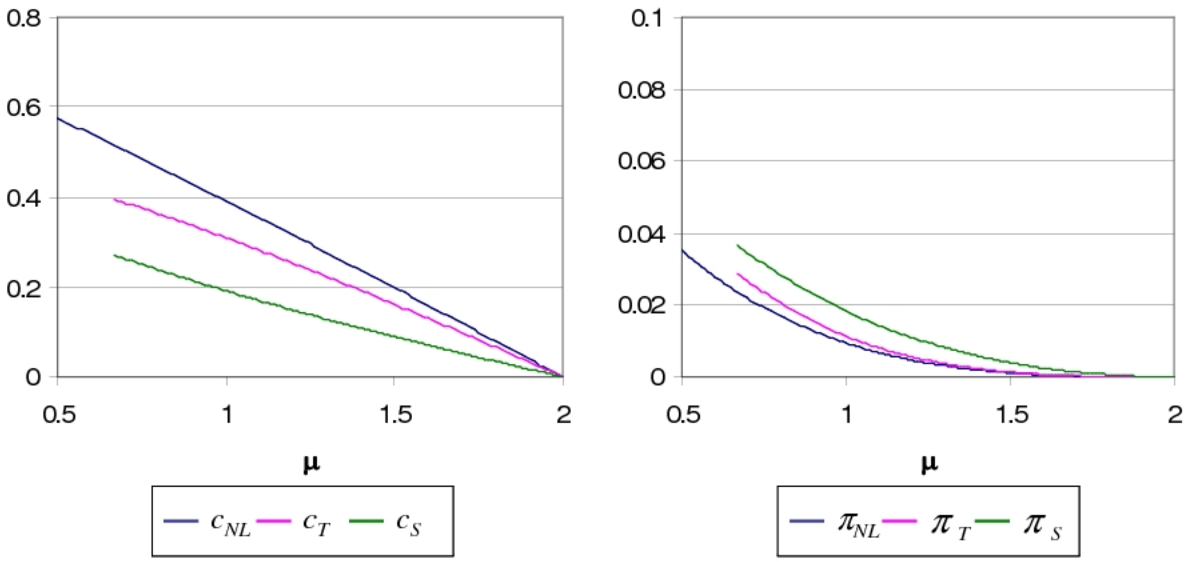


Figure 2: Equilibrium content and payoffs when $\rho = 0.5$ ($k = \delta = 1$)

parameters the asymmetric equilibria become sustainable for $\mu > 0.62$. Observe that it is $c_S < c_{NL} < c_T$ and $\pi_{NL} < \pi_T$, 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 ("pre-Internet") setting where each site has a monopoly over half the consumer population. The link source, in contrast might end up worse off than in the no-link equilibrium (in this particular setting it is $\pi_S < \pi_{NL}$ when $\mu < 1$).

Large ρ. When ρ is large then the link target does not benefit much from link traffic. However, the reduced competition between itself and the other site allows it to reduce its content overinvestment and bring its content levels down to the Pareto optimal c_P, a move that results in increased profits. If there is no outside option ($\mu = 0$) such a move would not be sustainable: Lower content investment by the link target would give an incentive to the link source to abandon the link and invest heavily in content itself. However, the presence of a sufficiently strong outside option provides incentives to the link target to maintain a level of overinvestment so as not to lose a lot of traffic to the outside option. This, in turn, increases the traffic to the link source as well as makes it more difficult for the link source to compete with the link target in a no-link setting. Both forces make it more attractive for the link source to free-ride on the target's content as opposed to deviating.

Figure 3 depicts the equilibrium content and payoffs for $k = 1$, $\delta = 1$ and $\rho = 1$. This is the extreme case where link targets get no link revenue. For these parameters the asymmetric equilibria become sustainable for $\mu > 0.67$. Observe that it is now $c_S < c_T = c_P < c_{NL}$. Interestingly, it is also $\pi_S > \pi_T = \pi_P > \pi_{NL}$: the link target makes higher profits than in the no-link equilibrium by avoiding overinvestment in content. However, the link source now makes *even higher profits than the link target* by free riding on the link target's content.



(a) Content (b) Payoffs

Figure 3: Equilibrium content and payoffs when $\rho = 1.0$ $(k = \delta = 1)$

The following proposition states the general form of these results:

Corollary 3. *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:

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

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. The link target is always better off, although the reasons for this differ depending on the value of ρ.

- For low values of ρ the link target capitalizes on the additional revenue that is expected to come through link traffic and invests in higher quality content. This, in turn, results in an increase of the total audience that visits the content ecosystem (that further boosts the incentives to invest in content) and also acts as a deterrent that discourages the other site from attempting to compete on content. The link source then invests in lower content and makes lower profits than without links.

- For high values of ρ linking helps sites to avoid the profit-reducing overinvestment in content that is inherent in the no-link equilibrium. Even when the link target does not receive substantial additional revenue through the link, the fact that it is linked to from another site removes the competition for market share and allows it to choose a content level that is close to the Pareto-optimal without links, a content level that is lower and more profitable than the competitive level with no links. In the right range of parameters the link source will find that harnessing the free traffic that linking offers it with minimal content investments is preferable to trying to out-compete the link target on content. In fact, in the extreme case $\rho = 1$, by doing so the source makes higher profits than the link target. This unusual outcome is the result of the strategic force that makes the link target tougher when it comes to investing in content, since that decision influences both the number of visitors and the marginal revenues. The link source, on the other hand, takes the effective content levels that attract consumers as given and only invests to increase marginal revenues.

Even though the details differ, in both cases linking helps reduce the inefficiency that is due to the competitive overinvestment on substitute content. In the first case (low ρ) only one of the two sites makes substantial investments in content but, thanks to linking, these investments become available to all consumers visiting the content ecosystem. The system becomes more similar to a situation where there is a single monopolist site. In the second case (high ρ) both sites reduce their content investments close to the Pareto-optimal levels without links. The system becomes more similar to a situation where there are two sites each having a monopoly over its respective segment. Both cases are better for the link target, but only the low ρ case is better for consumers and only the high ρ case better for the link source.

The next result summarizes the relevant comparative statics:

Corollary 4. *The following statements are true:*

1. *The link source functions $c_S(\delta, \rho)$ and $\pi_S(\delta, \rho)$ are increasing in both δ and ρ.*
2. *The link target functions $c_T(\delta, \rho)$ and $\pi_T(\delta, \rho)$ are decreasing in ρ and are decreasing (increasing) in δ when ρ is high (low).*

These results tell us that, as expected, the link source is better off when it can attract more revenue using the other site's content (higher ρ) and when the link efficiency δ (i.e the degree to which consumers find following a link as preferable as reading content directly on a page) is higher. The link target, on the other hand, is worse off as more consumers stay at the link source's page (higher ρ). Surprisingly, the link target is sometimes better off with a higher δ, that is, when consumers find following links as attractive as reading content directly. Higher δ implies higher utility from visiting the link source which implies more consumers attracted away from the outside alternative. When ρ is low, the link target receives most of the revenue of these additional visitors.

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 $k_2 = k > 0$ respectively. Throughout

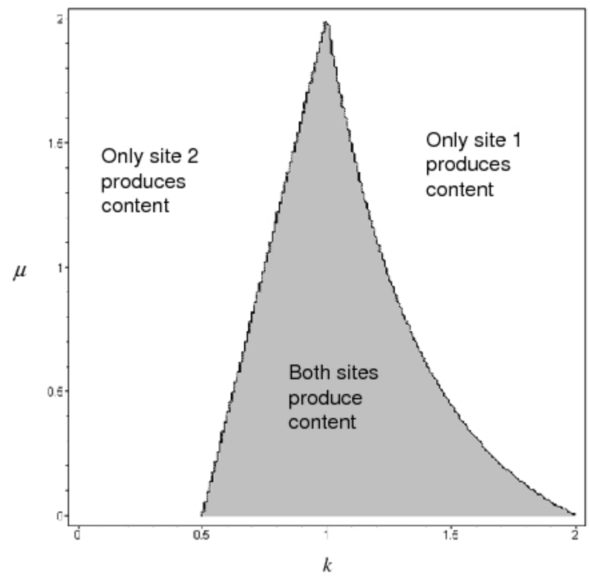


Figure 4: Equilibrium regions when sites have heterogeneous costs and there is no possibility to form links.

the section we further assume that consumers experience no disutility from following links ($\delta = 1$) and focus our attention to the more interesting set of parameters k , μ and ρ.

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

$$c_1^{NL},\ c_2^{NL} = \begin{cases} 0 & , \frac{1}{2k} - \mu + \frac{\sqrt{4k\mu+1}}{2k} & \text{if } k < \frac{2+\mu}{4} \\ \frac{(2k-1)(c_2^{NL}+\mu)-\mu}{2-k} & , \frac{6-k(2\mu(k+1)+3)+\sqrt{(2-k)^2(4\mu(k+1)+9)}}{2(k+1)^2} & \text{if } \frac{2+\mu}{4} \leq k \leq \frac{-1+\sqrt{1+4\mu}}{\mu} \\ \frac{1}{2} - \mu + \frac{\sqrt{4\mu+1}}{2} & , 0 & \text{if } k > \frac{-1+\sqrt{1+4\mu}}{\mu} \end{cases} \tag{7}$$

In the special case $\mu = 0$ the above expression simplifies to:

$$c_1^{NL},\ c_2^{NL} = \begin{cases} 0 & , \frac{1}{k} & \text{if } k < \frac{1}{2} \\ \frac{3(2k-1)}{(k+1)^2} & , \frac{3(2-k)}{(k+1)^2} & \text{if } \frac{1}{2} \leq k \leq 2 \\ 1 & , 0 & \text{if } k > 2 \end{cases}$$

and better conveys the key properties of the solution: The most striking property of this setting is that there is no room for a second competitor who is substantially weaker (less efficient) than the most efficient content producer. For example, when $\mu = 0$, if the cost parameter ratio between the high and low cost producer rises above 2, only the low cost producer can profitably enter the market, producing the monopoly content level that corresponds to its cost parameter. When the two sites are not too unevenly matched in terms of their content production costs then both produce positive content quality and capture positive market share. In the latter case sites overinvest in content as discussed in the previous section.

Figure 4 depicts the equilibrium regions that correspond to (7). Observe that, the stronger the outside option μ, the narrower the range of cost parameters for which both competitors are viable.

The ability to place links to the other site's content becomes even more important in such a setting as it now allows inefficient sites that would otherwise not be viable to enter the market and generate positive profits.

Proposition 5. *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 equilibrum exists where sites do not establish links in equilibrium. Their content levels are then given by (7).*

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

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

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

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

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

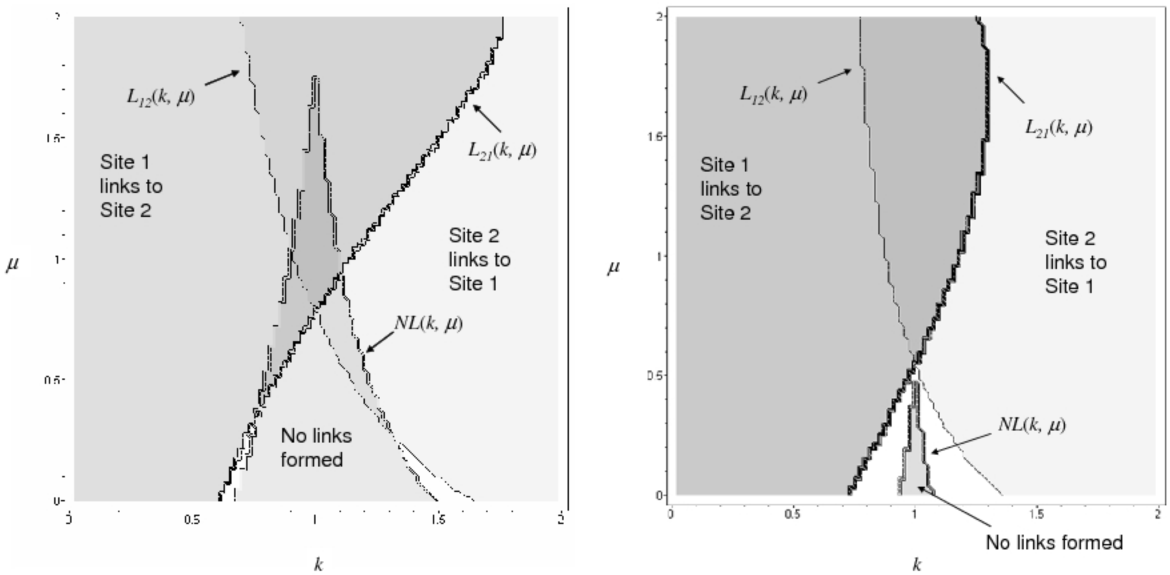
$$c_1 = c_S = \frac{\rho c_T}{2c_T + \mu}$$

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

Figure 5 depicts the parameter regions where each of the above equilibria become possible. The following points are worth noting:

- The parameter region where the no-link equilibrium is sustainable shrinks as ρ increases. A higher ρ implies a higher ability of the link source to benefit from the content at the link target, and thus higher relative attractiveness of linking vs. head-on content competition.

- 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.



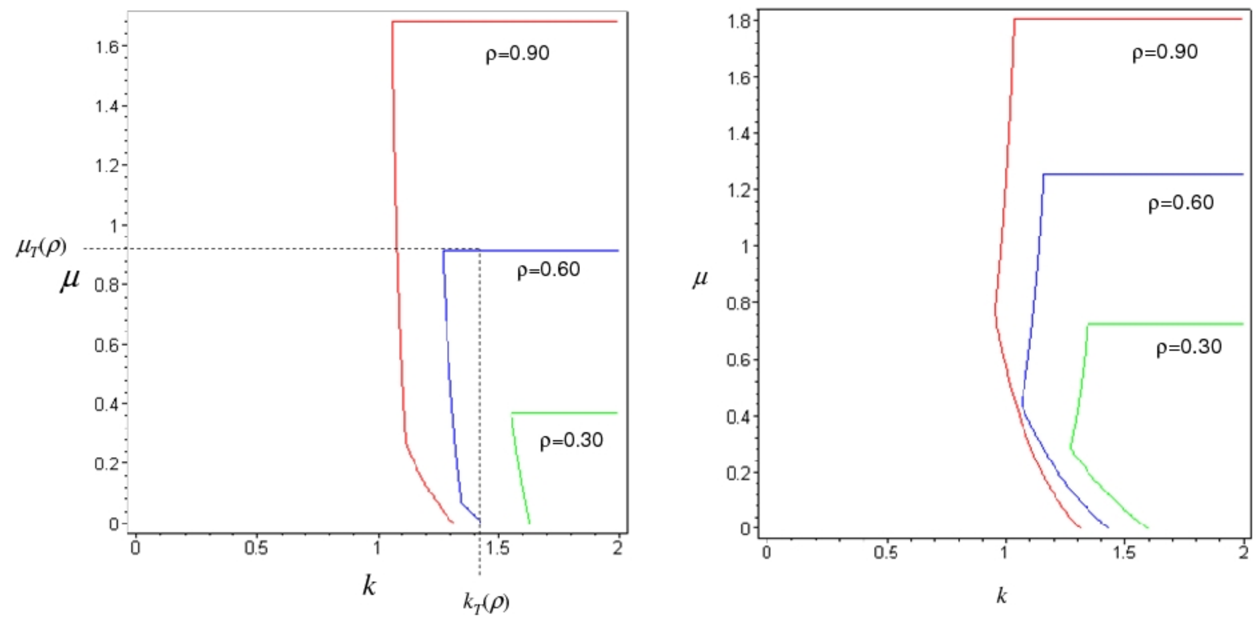
(a) ρ=0.25 (b) ρ=0.75

Figure 5: Equilibrium regions when sites have heterogeneous costs and links are possible. The white regions are regions where no pure equilibria exist.

- Somewhat surprisingly, we find that, when the cost differential between the two sites is not very large, an equilibrium exists where the *more efficient* site places a link to the less efficient site. For example, we see that there are equilibria where site 1 links to site 2 when $k > 1$ or where site 2 links to site 1 when $k < 1$. The parameter range where this behavior is feasible is broader the lower the ρ. This interesting outcome is related to the intuition behind link formation in the case of two homogeneous sites. Recall that, for small ρ, the expectation of becoming a link target enables a site to invest in substantially better content than without links, because it now expects to receive additional revenue both from additional anchor traffic as well as from additional link traffic. This additional investment attracts traffic away from the outside option μ, further strengthening the incentives to invest. This situation entails complementarities that enable sites that are somewhat less efficient than their competitors to aggressively invest in content to the point where their competitors will find it preferable to link to them rather than compete head-on.

Figure 6 depicts the parameter regions where link equilibria result in higher profits and higher content for the link target (higher relative to equilibria where no links are possible).[6] It is interesting to contrast these to the corresponding results of the previous section. In settings with homogeneous sites, when linking is sustainable, it *always* results in higher profits for the link target, because it reduces the inefficiency of content overinvestment when the two (evenly matched) sites compete head-on. When sites are heterogeneous this result only holds either (a) when the two sites do not

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



(a) Regions of lower profit

(b) Regions of lower content

Figure 6: 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.

have large differences in their cost parameters, or (b) when the link source is substantially less efficient than the list target but μ is sufficiently large compared to ρ. Specifically, observe that in Figure 6 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.

We will formally derive a version of this result in Section 5.1 when discussing the impact of aggregators. 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 enter the market, free-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 and, thus, offers utility δc_T to consumers. 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 $\frac{(1+\delta)c_T}{(1+\delta)c_T+\mu} \geq \frac{c_T}{c_T+\mu}$, with the inequality strict for $\mu > 0$. By making it easier to access good content the entry of additional content nodes, thus, increases the attractiveness of the entire content ecosystem and attracts traffic (i.e. audience and/or time spent) away from the outside alternative μ. If ρ is small compared to μ, most of the additional traffic that flows to the link source clicks on its link and eventually 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.

In summary, linking has qualitatively different implications when sites are roughly equally efficient than when sites have substantial disparities in their ability to produce content. In the former case, it serves an economically and (if ρ not too high) socially beneficial role. It helps reduce inefficiencies that arise from overinvestment in content and, as long as ρ is not too high, allows the site that serves as the link target to produce better content than in the no link case, content that, through linking, becomes available to all consumers. This makes the entire content ecosystem more attractive to consumers. In contrast, when one site is substantially less efficient than the other, linking has an ambivalent effect. On the one hand it allows the less efficient competitor, who in a setting without links would have low or zero market share, to free ride on the more efficient site's content and capture a higher market share. On the other hand, by increasing the number of access points to good content, links can make the entire content ecosystem more attractive to consumers. This, in turn, can attract traffic away from outside alternatives into the content ecosystem. Depending on how much of this additional traffic clicks through to the link target, the link target can end up better off or worse off relative to a setting where there are no links and no inefficient competitors.

This last tradeoff 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 an infinite cost of content production and, thus, are limited to attracting visitors by linking to content creators. These sites are usually called aggregators and many argue that their presence hurts content creators, by attracting away traffic without creating any value. Others, however, argue that aggregators create value by offering visitors an improved browsing experience, which results in an audience expansion that benefits content creators as well.

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 them.

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. The analysis of the previous section implies that such a site will produce almost no content and will place a link to the highest content site. 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 a lower content and links to the other (site 1).

In both scenarios the aggregator links to content of quality c_1 and, thus, offers utility δc_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 $\frac{(1+\delta)c_1+c_2}{(1+\delta)c_1+c_2+\mu} \geq \frac{c_1+c_2}{c_1+c_2+\mu}$, with the inequality strict for $\mu > 0$. By making it easier to access good content aggregators, thus, increase the attractiveness of the entire content ecosystem and attract traffic (i.e. audience and/or time spent) away from the outside alternative. That is, in settings with outside alternatives the entry of aggregators increases the total traffic that flows into the content ecosystem.

The impact of aggregators on individual content site traffic is two-fold. On the one hand, we have $\frac{c_i}{(1+\delta)c_1+c_2+\mu} < \frac{c_i}{c_1+c_2+\mu}$: 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 its original traffic 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. (It is easy to see that the low content site is always worse off.) 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. The following proposition formalizes the above discussion:

Proposition 6. *(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 total profits that sites 1 and 2 make are higher in the presence of an aggregator iff

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

The preceding analysis helps put the arguments of both the proponents and opponents of content aggregators in perspective. Aggregator opponents are right in that the entry of an aggregator into a content ecosystem "steals" anchor traffic from every incumbent content node. 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. To the extent that aggregators retain most of the revenue of their anchor traffic, they are, indeed, socially harmful. In most cases, however, aggregators send a fraction $1 - \rho$ of their anchor traffic (this traffic includes both the traffic they "stole" from incumbent sites as well as additional traffic they attracted from outside alternatives) off to the targets of the links they contain. If the fraction of the clickthrough traffic is substantial, the net impact of aggregators is positive for the best content sites: aggregators increase the total traffic that flows into the content ecosystem and direct most of it to the best content sites.

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 6 predicts that incumbent sites are better off in the presence of aggregators if and only if $\rho < 0$, i.e. never.

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 have very low 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 that constitute their main added value. In our setting the establishment of a single link can be justified on the basis of various arguments. For example, one can assume that consumers have limited cognitive budget and are only willing to follow one link at most. Or that there are screen size constraints that limit the number of possible links (this argument is particularly pertinent in mobile device environments). Or that link formation incurs a small cost so it is not rational to add links that do not add to the aggregator's expected utility to consumers (recall that this utility is the maximum of a site's own content and the content of all link targets).

Under the above assumptions, aggregator entry induces competition among the two incumbent sites for the aggregator's link and its associated traffic. 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 randomizes 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 almost certainly find the top site. 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 as well.[7]

We will now perform an analysis similar to that of Section 4.2 in this new setting. Our main objective is to explore how the aggregator's search parameter s affects 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 $k_i = 1$. We also set $\delta = 1$ so that we can focus on the role of parameters μ and ρ.

5.2.1 No links allowed

First, we look at the case when there is no link between the two incumbents. 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:

$$\pi_i = \frac{c_i^s}{c_i^s + c_j^s} \cdot \frac{c_i + (1-\rho)c_i}{2c_i + c_j + \mu} c_i + \frac{c_j^s}{c_i^s + c_j^s} \cdot \frac{c_i}{c_i + 2c_j + \mu} c_i - \frac{1}{2} c_i^2. \tag{8}$$

We determine the symmetric equilibrium.

Lemma 7. *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*

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

[7] 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.

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 overinvesting in content. 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 of content and profits.

Proposition 8. *The symmetric equilibrium of Lemma 7 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 increasing (decreasing) in* ρ. *For intermediate 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, leading to higher content and lower profits. The rather complex effect of the aggregator's clickthrough 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.

5.2.2 Links allowed

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 each to 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 *each of* site S and 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:[8]

$$\pi_S = \frac{c_T}{3c_T+\mu}\rho c_S - \frac{1}{2}c_S^2 \qquad \pi_T = \frac{c_T + 2(1-\rho)c_T}{3c_T+\mu}c_T - \frac{1}{2}c_T^2 \tag{10}$$

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

1. *If* $\mu \leq \overline{NL}(\rho, s)$*, an equilibrum exists where sites do not establish links in equilibrium. Their content levels are then given by (9).*
2. *If* $\mu \geq \underline{L}(\rho, s)$*, an equilibrium exists where site S links to site T and*

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

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

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

The above result is analogous to Proposition 2, 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 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 7 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 1). 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 gets squeezed, each content site finds it increasingly

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

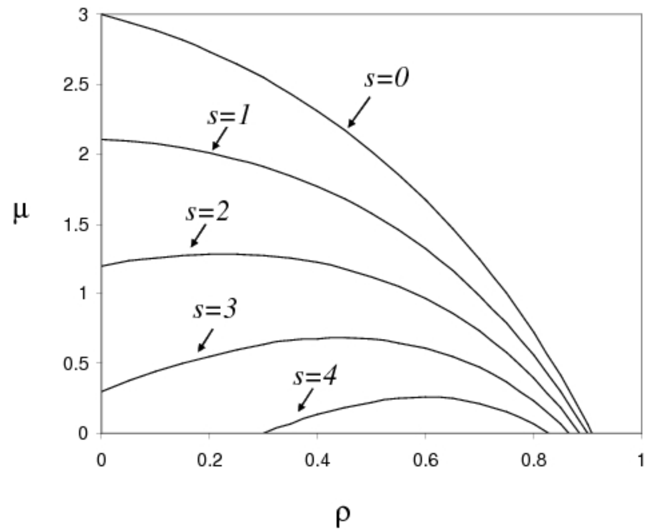


Figure 7: As the aggregator' search parameter s increases, the parameter region where no-link equilibria are sustainable (the area below the curves) shrinks.

attractive to deviate from the equilibrium, reduce its own content production and simply place a link to the other site.

Figure 8 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 1). 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 placing a link to another site (site 1), site 2 signals to the aggregator that the other site has better content. The aggregator then always links to site 1. For low ρ, site 1 then receives link traffic from both site 2 and the aggregator. Expecting all this traffic, site 1 is able to invest in high quality content that site 2 finds difficult to surpass if it were to deviate from the link equilibrium. However, by deviating from the link equilibrium site 2 withholds the 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.

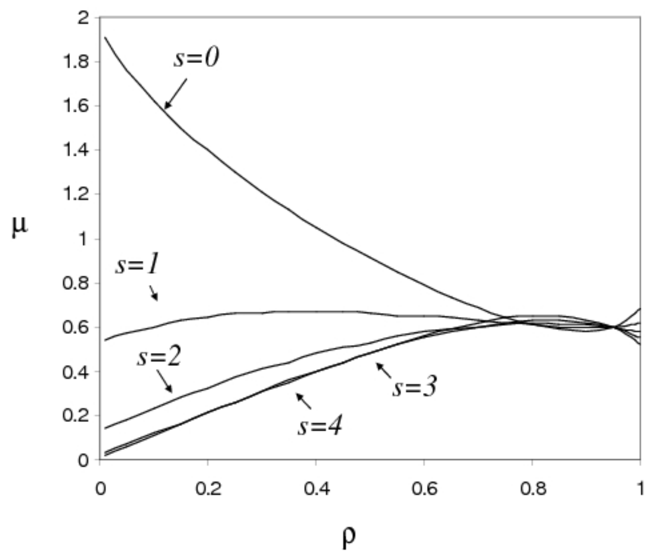


Figure 8: For low ρ, as the aggregator's search parameter s increases, the parameter region where link equilibria are sustainable (the area above the curves) expands.

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 low fractions of traffic 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.

In summary, the presence of aggregators that selectively place links to a subset of available content increases competition among content sites. If sites do not have the ability to place links to each other this forces them to overinvest in content, which reduces their profits. The ability to place links to other content sites can be a way out of such costly head-on competition. At the same time, however, links reveal to the aggregator who has the best content. As a result, when aggregators are not good at figuring out high quality content independently of the link structure, their presence can makes sites more reluctant to place links to each other. As aggregators get better at figuring out high quality content independently of links, the quality signaling aspects of linking become less relevant and the presence of aggregators makes sites more eager to form link networks so as to avoid head-on competition. As discussed in Section 4, when aggregators send sufficiently large fractions of traffic to their link targets (i.e. when ρ is not very large) such link networks are generally good for consumers (the link target is able to invest in higher content than in the no-link case and this content becomes accessible to everyone) and result in higher link target as well as total profits relative to no-link equilibria.

6 Extension: 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 what we assumed so far when analyzing the effects of links on content investments and profits. However, there are technical means for the target to refuse a link or to indicate that it does not desire to be linked to. Examples include news sites that do not display articles to visitors who do not reach such articles via the site's front page. Other sites do not allow search engines to crawl their content, effectively refusing to receive links from a search engine. It is not surprising that such practices are becoming prevalent in light of complaints of content producers about unfair linking practices.[9] It is, thus, worthwhile to assess their economic implications.

In this section we explore how our results change in settings where link formation requires the agreement of both the source and the target. Our main results show that, whereas the ability to refuse links can give efficient sites the power to keep inefficient competitors out of the market, it can also result in situations where the link source is forced to unnecessarily over-invest in content that does not benefit itself or consumers. Furthermore, the increased competition among content sites that is induced by the presence of aggregators makes it less likely that content nodes will refuse links coming from aggregators.

6.1 Two sites

Here, we modify the basic two homogeneous player model presented in Section 4.2 and allow link targets to veto (refuse) a link after it has been created by the link source. The timing is therefore the following. Sites first make their decisions on content investments. Then they decide whether they want to link to another site. Finally, each link target decides whether it allows or refuses the link.[10]

As one would expect, a site refuses a link from its peer if the link source's content is low relative to the link target. Simple calculations show that a link from site i to site j will be accepted by by the target as long as

$$c_i \geq \frac{\delta(\rho c_j - \mu(1-\rho))}{1+(1-\rho)\delta}.$$

Otherwise, the link will be refused by the target and sites will rely on their own content. When sites are heterogeneous, the above argument implies that targets will refuse links from inefficient competitors. When the competing sites are homogeneous, the following result summarizes how this affects content and linking decisions and how it changes the results of Proposition 2.

[9]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/

[10]Our results do not change if we reverse the order of link creation by the source and the refusal decision by the target as long as these decisions are made after the content investments have been settled.

Proposition 10. *There exist thresholds* $\overline{SCL}(\delta,\rho) \geq \underline{CL}(\delta,\rho) \geq \underline{L}(\delta,\rho)$ *and* $\overline{NCL}(\delta,\rho) \geq \overline{NL}(\delta,\rho)$ *such that:*

1. *If* $k\mu \leq \overline{NCL}(\delta,\rho)$ *then sites do not establish links in equilibrium and* $c_i^* = c_j^* = c_{NL}$.
2. *If* $k\mu \geq \overline{SCL}(\delta,\rho)$ *then there are two asymmetric equilibria where one site links to the other and* $c_i^* = c_S \leq c_j^* = c_T$*, which is identical to the case without possible link refusal.*
3. *If* $\overline{SCL}(\delta,\rho) \geq k\mu \geq \underline{L}(\delta,\rho)$ *then there are two asymmetric equilibria where one site links to the other and*
$$c_i^* = \frac{\delta(\rho c_T - \mu(1-\rho))}{1+(1-\rho)\delta} \geq c_S, \quad c_j^* = c_T$$
4. *There is no equilibrium in pure strategies otherwise.*

Figure 9 illustrates that the results are similar to the case when links cannot be refused (Figure 1) but the ability to refuse links makes linking less likely. The parameter region in which linking is feasible shrinks, especially for high values of ρ. This is not surprising, since link targets are hurt the most when link sources are able to retain a high proprotion of the traffic they attract. The region of the no-link equilibrium, on the other hand, expands, and these forces result in a potential reduction in profits due to the increased competition without links. Nonetheless, for a wide range of parameters, linking is still feasible. If μ is high enough this happens because both the target and the source are better off due to the reduced competition and increased audience of link equilibria.

We observe an interesting phenomenon in an intermediate region: when μ is not high enough to provide enough benefits to the target under our standard linking equilibrium, the link source will invest more in content to avoid having its link refused. What is interesting is that in this case the link source produces more content, not so that it attracts more consumers, but so that the target views it as a sufficiently credible competitor with whom it would rather coordinate (i.e. receive a link from) rather than compete head-on.[11] Naturally, this increased content level will be suboptimal given the link, thus the link source's profits will be reduced while the link target's profits will remain unchanged relative to a setting where the target cannot refuse the link. This latter point illustrates how the threat of link refusal hurts one player without necessarily helping the other one. In summary, the ability to refuse links may help potential link targets to avoid losses from being linked to, but in many cases can reduce aggregate profits.

6.2 Refusing links from an aggregator

The preceding analysis shows that a site may want to refuse a link from its competitor if the competitor has low content. However, when aggregators who place links to a subset of available content are present, sites may be reluctant to refuse aggregator links since these links might then be

[11] This is, therefore, a variant of a hold-up situation.

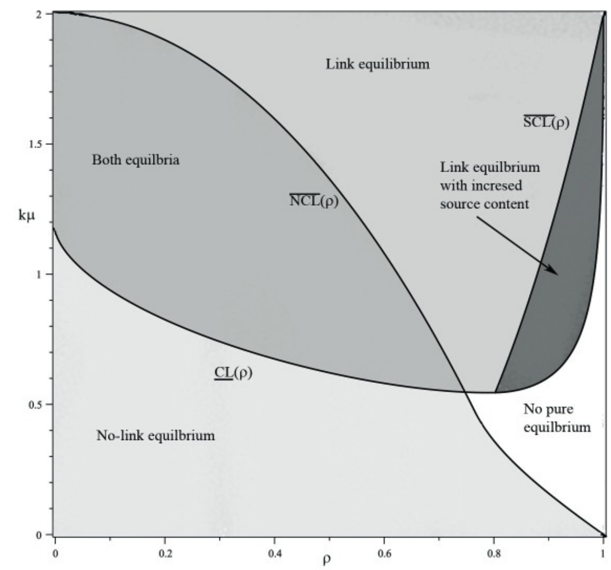


Figure 9: When links have to be approved by the target, the region where linking is sustainable shrinks, whereas the region where there is no linking expands. In the darkes shaded area linking is sustainable, but the link source invest higher than optimal in content to avoid having its link refused.

directed to their competitors. Here, 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 producers have a choice to refuse links. We assume that after the content decisions have been made, the two sites simultaneously decide whether they 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. Otherwise, if neither site accepts links, the aggregator will not be able 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.[12]We determine the subgame-perfect Nash-equilbria that are symmetric in content choice.

Proposition 11. *There always exists an equilbrium 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 equilbrium is unique iff $\rho < 1/2$. *Otherwise, there exists a second equilbrium in which both sites refuse the link from the aggregator and build a content level of* $3/8$. *Profits in the latter equilbrium are always higher.*

The results show that, when ρ is high, sites have an incentive to refuse the link from the aggregator as it attracts traffic from the content producers. However, sites have to coordinate so that they both

[12]When $\mu = 0$, the overall traffic in the system is constant, making it evident that the aggregators generate traffic at the content creator's expense. When $\mu > 0$, the results are similar, but link refusal is less likely in general.

refuse the links. If one site does allow the aggregator to link to it, the other site is under pressure to also allow the link. Otherwise the site that declines loses the possibilty to receive some (a fraction $1 - \rho$) of the traffic that was attracted by the aggregator; all these visitors will go to its competitor. 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 burden on the content sites (see Section 5.1).

The impact of accepting or refusing aggregator links on content levels has an interesting relationship with the aggregator's parameters ρ (fraction of visitors retained by aggregator) and s (aggregator's ability to discover what site has the best content). When ρ is low and s is high, content levels are lower in the link-refusal equilibrium: when links are allowed, sites compete to receive the effective aggregator's single link and this drives their content investments up. In contrast, when the aggregator is not so effective in assessing content quality (low s) and retains most of its visitors (high ρ), content investments are higher when sites refuse aggregator links. In such cases allowing links from the aggregator results in traffic being "stolen" away from the content sites, reducing their incentives to produce content.

In summary, when refusal is an option sites are always better off if they can coordinate to do it, but consumers are worse off if the sites refuse links from an efficient aggregator.

7 Concluding remarks

This paper is the first to take a comprehensive look at the economic implications of free hyperlinks in settings where content sites compete with one another 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 interesting insights:

- *Links among peer content producers can increase firm profits by reducing competition and duplicate effort.* One of the most disruptive effects of the Internet to content industries is the elimination of geographical monopolies on content creation; all content sites now directly compete for every consumer. Direct competition induces sites to overproduce what is often duplicate content. This hurts everyone's profits without substantially benefiting consumers. We find that linking allows sets of roughly equally capable sites to coordinate content production in ways that both increase aggregate profits and (if the link click-through rate is sufficiently high) consumer utility. In such equilibria one site invests in very high quality content and other sites link to it. Expecting to receive both direct and link traffic, the site that invests in content is able to produce higher quality content than in a no-link equilibrium. Through links, this content becomes available to all consumers, increasing the overall attractiveness of the content ecosystem and attracting additional consumers from outside alternatives.
- *Links only form if competition among sites is not too tough.* Links typically result in different payoffs for nodes acting as link sources and nodes acting as link targets. The nodes with

lower payoffs are tempted to abandon links and compete directly on content. We find that, when there is no outside alternative, this temptation precludes the formation of link equilibria. When there is a sufficiently strong outside alternative, sites care about their standing relative to one another but also about their standing relative to the outside alternative (which is usually strengthened through the efficiencies of linking). Thus, the presence of a strong outside alternative softens competition among sites and is a necessary condition for link equilibria to form.

- *Linking can sustain market entry of inefficient players.* Head-on competition means that only reasonably efficient players can survive. By softening competition, linking allows inefficient players to remain in the market by free-riding on the content of efficient sites. If the amount of revenue that is retained in link source nodes is substantial, this represents a social cost of linking and an argument against the culture of uninhibited unilateral linking decisions. Content sites could avoid this if they could selectively refuse incoming links from substantially less efficient competitors. However, allowing link targets to refuse link also has undesirable side-effects that might lead to across-the-board efficiency losses.

- *The main benefit of aggregators to content producers comes from traffic expansion.* It is rational for aggregators to form links to the best available content. Their presence, thus, 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, aggregators then 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.

- *The presence of aggregators incurs social costs that must not be overlooked.* Since aggregators do not produce their own content, they are not sustainable in settings without links. They, therefore, represent an additional type of node that did not exist in traditional content ecosystems. Their market entry inevitably appropriates some of the revenue that would otherwise be shared among content producers. Their net effect is positive for content producers only if the traffic expansion they induce is sufficient to offset the loss of advertising revenue. Most notably, we find that aggregators are never beneficial to content producers in "closed universe" settings where there is no outside alternative and sites simply fight to extract traffic from each other.

- *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. At the same time, the increased competitive pressure brought forth by the presence of aggregators makes it more likely that content sites will form link equilibria.

Our results have implications for professional content producers, who, so far, have been somewhat reluctant to embrace the use of links to other peer sites. We demonstrate conditions under which linking to peers can be beneficial, as well as conditions under which incoming links should be refused. Our work also clarifies the multi-faceted impacts of aggregators on content ecosystems and, thus, adds important nuances to the current debate between aggregators and content producers.

One of the most striking insights of our analysis is that link equilibria often do *not* form, even though their formation can lead to higher aggregate profits and better content. This, in the view of the authors constitutes a negative side-effect of the culture of "free" links that currently pervades the web (see, for example http://www.right2link.org/). In our ongoing work we explore how associating payments with links can allow sites to share the surplus generated by linking and, thus, expand the range of settings where competing content sites agree to form mutually beneficial link equilibria.

This work is a first step towards understanding an arguably under-researched area and, as such, 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, homogeneous consumers and two or three nodes. 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. The next step of this work is to extend our results to larger networks with more realistic features (e.g. multiple topics, heterogeneous consumers, costly search for content by aggregators, etc.) and study the implications of the identified strategic interactions on the structure of the emergent content networks. Furthermore, this paper focused on free links, i.e. links that are not accompanied by economic transfers between the link source and target (e.g. organic search results). An interesting direction of future work is to extend our analysis to settings where such payments would enhance the utility for consumers, content creators, and aggregators.

References

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

Bloch, F. and Dutta, B. (2008) Communication networks with endogenous link strength. Games and Economic Behavior, forthcoming.

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

Cabrales, A., Calvo-Armegnol, A. and Zenou, Y. (2007) Social Interactions and Spillovers: Incentives, Segregations and Topology. Working Paper.

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. (2007) The Law of the Few. University of Essex, Department of Economics Discussion Paper Series No. 636, August 2007.

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

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

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).

Mayzlin, D. and Yoganarasimhan, H. (2008) Link to Success: How Blogs Build an Audience by Promoting Rivals. Working Paper, Yale School of Management.

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, Kristen, Rainie, Lee, Mitchell, Amy, Rosenstiel, Tom, and Olmstead, Kenny (2010) Understanding the Participatory News Consumer, Pew Internet & American Life Project, Research Report.

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

Appendix

Proof of Proposition 1:

Differentiating the profit function in (4) yields that the best response of site i to a content $c_j$of site j is

$$b_i(c_j) = \frac{1 - 2k\mu - 2kc_j + \sqrt{1 + 4k\mu + 4kc_j}}{2k}. \tag{11}$$

Solving for and equilbrium yields the expression in (6). It is easy to check that $c_{NL} > c_P$when $k\mu$<2. Since c_P maximizes profits, the competitive profits are suboptimal.

Proof of Proposition 2:

There are two possible types of equilbrium with respect to linking: i) the one where there is no link between the two sites and they invest equally in content (c_{NL}), and ii) the 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 equilbrium. When site i links to site j, then its profit becomes

$$\pi_{i \to j} = \frac{\delta c_j \rho}{\delta c_j + c_j + \mu} c_i - \frac{k_i}{2} c_i^2.$$

Comparing these two yields that site i will link to site j iff

$$c_i \leq \frac{\delta \rho c_j (c_j + \mu)}{(1 + \delta(1 - \rho))c_j + \mu}.$$

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:

$$\pi_{j \leftarrow i} = \frac{c_j + \delta(1 - \rho)c_j}{\delta c_j + c_j + \mu} c_j - \frac{k_j}{2} c_j^2.$$

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

$$b_{i \to j}(c_j) = \frac{\delta \rho c_j}{(c_j + c_j \delta + \mu)k}$$

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 $\underline{L}(\delta, \rho)$ denote the value of μ where the above holds with equality when $k = 1$. It is easy to check that the above inequality is invariant to the values of μ and k as long as μk is fixed. Furthermore, it holds for high values of μk, yielding that the linking equilibria exists iff $\mu k \geq \underline{L}(\delta, \rho)$. Similary, let $\overline{NL}(\delta, \rho)$ denote the value of μk 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 $\mu k \leq \overline{NL}(\delta, \rho)$, completing the proof.

Proof of Corollary 3 and Corollary 4:

It is useful to start with examining the comparative statiscs with respect to $\delta$and ρ. One can check that $c_S(\delta, \rho)$ is increasing in both its variables, yielding that $\pi_S(\delta, \rho)$ is also increasing. Then it is enough to check that $c_S(1, 1) \leq c_P$,yielding $c_S(\delta, \rho) \leq c_P$for any $\delta$and ρ. Similarly, one can check that $c_T(\delta, \rho)$ is decreasing in ρ, and that $c_P \leq c_S(\delta, 1)$ for any δ. For the results depending on whether $\rho$is low or high, we check the derivates for $\rho = 0$ and $\rho = 1$and get the stated results. Since all the functions are contiuously differentiable, we get the same results for a region of small values of ρas for $\rho = 0$ and the same results for a region of high values as for $\rho = 1$.

Proof of Proposition 5:

The proof follows the exact same steps as the proof of Proposition 2. However, due to the asymmetric cost, the profits and the best responses are different for the two sites, yielding the two different tresholds for the linking equilbrium.

Proof of Proposition 6:

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 $z_3 = \delta c_1$. The expected anchor traffic of a content producer will decrease from $\frac{c_1}{2c_1+\mu}$to $\frac{c_1}{(2+\delta)c_1+\mu}$. However, they will receive some traffic through the link from the aggregator in the amount of $\left(\frac{1-\rho}{2}\right)\frac{\delta c_1}{(2+\delta)c_1+\mu}$, yielding a total traffic of $\left(1+\delta\frac{1-\rho}{2}\right)\frac{c_1}{(2+\delta)c_1+\mu}$ 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 profts, 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 determin how the amount of traffic changes at the two sites. Before the aggregator enters, sites 1 and 2 receive traffic of $\frac{(1+(1-\rho)\delta)c_1}{(1+\delta)c_1+\mu}$and $\frac{\rho\delta c_1}{(1+\delta)c_1+\mu}$, respectively. When the aggregator enters, these change to $\frac{(1+2(1-\rho)\delta)c_1}{(1+2\delta)c_1+\mu}$and $\frac{\rho\delta c_1}{(1+2\delta)c_1+\mu}$. Determining the sign of the change in traffic for the two sites and comparing the profits yields the tresholds for ρ.

Proof of Lemma 7:

We differentiate the 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 maxium. 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

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

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

Proof of Proposition 8:

Straightforward analysis of the expressions derived in Lemma *7* show the relationship between the equilibrium content profit and the model parameters.

Proof of Proposition 9:

We start by proving part 2. In the possible equilbrium where the lower quality content creator and the aggregator both link to the higher quality content site, 10 describes the payoff the source and the target. Differentiating $\pi_T$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 2, 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 similary, 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 8 for $\pi_i(c_i, c_j)$.

Proof of Proposition 10:

The proof goes along the exact same lines as that of proposition 2. One can identify the two possible equilibria: either sites do not link to each other or one site links to its competitor. However, in this case we have to take into account the possibility of link refusal and check more conditions for incentives to deviate. For the linking equilbrium to exist

$$c_S \geq \frac{\delta(\rho c_T - \mu(1-\rho))}{1 + (1-\rho)\delta}$$

has to hold. Furthermore, possible deviations from the no-linking equilbrium also have to satisfy the link-feasibilty condition. Finally, one has to identify the cases, in which linking would not be feasible under the previusly determined source content levels, but is feasible under an increased content level that satisfies the above inequality.

Proof of Proposition 11:

We start by analyzing the las stage of the game in which site decide wheter or not to refuse link from an aggregator. Since we are looking for equilbria that are symmetric in content, we can assume that aggregator would link to both sites equally likely if not refused. 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 site allowing linking is always an equilibrium of the subgame. If site j decides to refuse the link, site i has two options. If at 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 an equilbrium of the subgame. One can then determine the equilibrium content levels in the two cases using the results of proposition 2 and lemma 7.