Sticky Content and the Structure of the Commercial Web

Scott Duke Kominers
Department of Economics, Harvard University
& Harvard Business School
Cambridge, MA 02138
kominers@fas.harvard.edu
skominers@hbs.edu

ABSTRACT

We give the first model of the effects of *sticky content*, generalizing the PageRank-inspired model of the commercial web introduced by Katona and Sarvary. In our model, websites buy and sell links on each others' pages. Consumer browsing behavior is based upon the PageRank process, but is affected by websites' respective levels of sticky content. We discuss two varieties of sticky content: *attracting content*, which induces consumers to return regularly, and *entrapping content*, which both attracts consumers and maintains consumer attention.

We characterize the effects of both forms of sticky content upon the web network structure and the distribution of utility. The set of web network equilibria is independent of the distribution of attracting content. By contrast, entrapping content does affect the equilibrium web network. However, an inverse relationship between commercial content levels and the number of outgoing links is preserved. Although attracting content is universally beneficial for websites, entrapping content is not. A website without incoming links prefers to have entrapping content, but a website with incoming links and sufficiently large PageRank prefers not to have entrapping content. We thus observe endogenous specialization of website business models: heavily trafficked sites primarily profit through sponsored outlink traffic and hence prefer to have little entrapping content; low-traffic sites primarily profit through sales of on-site content and hence prefer to entrap users.

Categories and Subject Descriptors

J.4 [**Computer Applications**]: Social and Behavioral Sciences – Economics; K.4.4 [**Computers and Society**]: Electronic Commerce

Keywords

Sticky content, Internet advertising, Network formation, PageRank

1. Introduction

Sticky content, website content which induces return traffic and holds user attention, is now ubiquitous across the *commercial world wide web*, the network of websites seeking to profit from economic exchange. Today, commercial websites such as online stores, web portals, and search engines regularly include sticky services such as weather updates, daily news headlines, webmail, chat rooms, and online games.

Despite its prevalence, sticky content has received relatively little attention in the academic literature. The management and marketing literatures have highlighted the importance of sticky content (e.g., Clarke and Flaherty [5]) and have suggested methods by which individual websites might make themselves more "sticky" (e.g., Haywood [6]). Nonetheless, these literatures have neither modeled nor rigorously discussed the micro- and macro-level effects of sticky content.[1]

In this paper, we give the first model of sticky content's effects. Our model generalizes the game-theoretic commercial web model of Katona and Sarvary [9], in which websites purchase incoming links from each other in a simultaneous game, and consumer browsing behavior, which determines revenues, is modeled upon the browsing process introduced by Brin and Page [4] for their computation of PageRank.[2] This model allows for a novel theoretical investigation of the equilibrium web network. Katona and Sarvary [9] identify several important properties of this network, such as a form of specialization of sites' revenue models, and also study extensions of their model in which sites were allowed to establish outgoing *reference links* or to be listed in search engines' indices.

We study two different types of sticky content. First, we study *attracting content*, content (such as weather updates

[1]A comprehensive study of the consequences of sticky content may have been prevented by complications inherent to macro-level studies of the internet. Indeed, both theoretical and empirical studies of the web network face myriad difficulties. For example, neither inter-site consumer browsing behavior nor the channels of web advertising are well-understood. Furthermore, the web network is constantly—and dynamically—evolving.

[2]This model of consumer browsing is well-established and empirically supported in the computer science literature (see Langville and Meyer [11] for a survey and for further references). PageRank-based generative models for the world wide web have also appeared within the operations research literature. For example, Pandurangan, Raghavan, and Upfal [14] developed a structural web model which explains an observed power law distribution of PageRank. Additionally, the model of Immorlica, Jain, and Mahdian [7] is likely PageRank-inspired.

or daily news headlines) which generates regular return traffic to specific sites. Then, we examine *entrapping content* content (such as webmail, chat rooms, or online games) which both attracts consumers and causes consumers to remain on the same site for long periods of time. We obtain two surprising results. First, relative levels of attracting sticky content do not affect the set of web network equilibria. Second, entrapping content is not universally beneficial for websites.[3] One consequence of this second result is an endogenous specialization of website business models: heavily trafficked sites primarily profit through sponsored outlink traffic and hence prefer to have little entrapping content; low-traffic sites primarily profit through sales of on-site content and hence prefer to entrap users.

1.1 Outline of the Paper. The remainder of this paper is organized as follows. In Section 2, we briefly survey the relevant literature on sticky content, network formation, and the commercial web. We then present our base model in Section 3. In Section 4, we introduce our model of attracting sticky content, and explain such content's effects upon the web network structure and the distribution of utility. Then, in Section 5, we describe and partially characterize the structure of the web network in the presence of entrapping sticky content. Finally, we endogenize sites' sticky content levels in Section 5.2 and discuss sites' incentives to develop entrapping sticky content. Our conclusion, presented in Section 6, indicates some intuitions managers may draw from our results and suggests some directions for future research. All proofs are presented in the Appendix.

2. Literature Survey

As we mentioned in the prior section, there has been little academic investigation of sticky content and its effects. The sparse attention sticky content has received has been confined to the marketing literature. Lewis and Bridger [12] highlight the importance of sticky content for retailers hoping to ensure consumer loyalty. Clarke and Flaherty [5] briefly mention the presence of entrapping sticky content on internet portal sites, but do not undertake any extended discussion of such content. Also, Clarke and Flaherty [5] do not acknowledge the existence of attracting sticky content. In perhaps the most detailed approach to sticky content currently available in the literature, Haywood [6] studies eBay's use of sticky content and marketing, presenting several motivations for a site to adopt sticky content, and identifying paradigms for "good sticky design."

Throughout this sparse literature, sticky content is often treated as purely entrapping, and is argued to be universally beneficial for websites. In contrast to this claim, we demonstrate in Section 5.2 that entrapping sticky content can be undesirable for high-traffic websites.

Although there has been substantial work on general endogenous network formation, most models of this literature do not appear effective for the web framework. The general network formation models of Bala and Goyal [3] and Slikker and van den Nouweland [16], for example, focus on the cost of establishing links and assume that players are identical.[4] Neither of these requirements is appropriate for the study of the web in the presence of sticky content, as website content in this setting is heterogeneous and the cost of establishing a link depends upon the content of the linking site.[5]

Immorlica, Jain, and Mahdian [7] have provided one effective model of endogenous network formation in a web setting. They study the design of site hyperlink structure, treating this structure as a network formed endogenously within a web site. They use a model of browsing dynamics similar to that of Katona and Sarvary [9] in which site users follow a random walk across links. However, because their analysis is confined to intra-site dynamics, Immorlica, Jain, and Mahdian [7] do not capture site reentry, and assume away the possibility of cycles in the network graph. Neither of these simplifications are appropriate for site-level studies of the web.

As detailed by Katona and Sarvary [9], work on the web link network structure relates to the broad literature on advertising, but does not fit well within preexisting advertising frameworks.[6] Recently, however, there have been studies of web link networks. For example, Mayzlin and Yoganarasimhan [13] model how bloggers develop reference links to each others' pages. Somewhat similarly, Stephen and Toubia [17] analyze the economic value of web network links, and Sgroi [15] employs a web graph model to study the stability of different web network configurations. These studies, however, have not addressed sticky content.

Additionally, the Katona and Sarvary [9] consumer browsing model is one of the few applications of the PageRank model within the social science literature. Our work further extends this framework, introducing a generalization which is new and is unique to the social science literature.

3. The Base Model

Our model generalizes that of Katona and Sarvary [9], hence we use the notation of Katona and Sarvary [9] whenever possible. The web is represented as a directed graph G with node set $V(G)$ and edge set $E(G)$. Each node $i \in V(G)$ corresponds to a website and each edge $(i \to j) \in E(G)$ represents to a link from site i to site j. For two websites $i, j \in V(G)$ we write $i \to j$ (resp. $i \not\to j$) if there is (resp. is not) a link between i and j. We also write $d_i^{\text{out}} := \text{outdeg}(i)$ for $i \in V(G)$; this parameter represents the number of links emanating from website i.

3.1 Websites. As in Katona and Sarvary [9], we restrict our attention to the analysis of *websites* (rather than *webpages*). This unit of analysis is convenient for the study of commercial networks, as each unit represents a single agent. This approach is also appropriate in the study of the web's network structure: incoming links typically point to the main page of a site, while outgoing links may emanate from any page. Since sticky content is typically attached to a site, the site-level approach remains valid in our study.

Each website i is assumed to have an exogenously given commercial content parameter $c_i \in [0, 1]$ and a sticky content parameter $s_i \geq 0$. We will endogenize the sticky content

[3]As we show in Proposition 7, a website with incoming links and sufficiently large PageRank prefers not to have entrapping content.

[4]This latter restriction—that players are identical—is pervasive throughout the literature on network formation (see Jackson [8] for a survey).

[5]As such, this work draws conclusions which do not describe the web network architecture. For example, the analysis of Bala and Goyal [3] indicates that the equilibrium network architecture should be either a "wheel" or a "star."

[6]Bagwell [2] gives a comprehensive survey of the advertising literature; Katona and Sarvary [9] explain why their model of web network formation as a link-purchasing game does not map directly into preexisting work.

parameter in Section 5.2, but for now we will treat it as exogenous. We assume that each site pays a fixed operating cost (normalized to 0) and an additional cost C each time a commercial sale is completed. For example, the fixed cost could represent the cost of setting up a webpage, while the per-visitor cost could represent supply or shipping costs.

There is a market for links between sites. Each site i offers to sell links at a per-click price $q_i \geq 0$. We assume that $q_i = q(c_i)$ is increasing in the commercial content parameter c_i.[7] As we will discuss in Section 6.1.1, the effects of sticky content on outgoing link prices are unclear. Consequently, we make no assumptions regarding the interaction of q_i and the sticky content parameter s_i.

3.2 Consumer Behavior. The consumer browsing process is modeled as a random walk across the web. As such, our model ignores the strategic behavior of consumers. This is clearly a simplification, as then site quality and reputation do not directly affect browsing behavior. Nonetheless, random-walk consumer traffic models of the type we use have been demonstrated to be good proxies for site traffic and quality (see Langville and Meyer [11]). Thus, although our model does not directly account for strategic elements of consumer browsing behavior, its predictions regarding site traffic are still realistic.

Following Katona and Sarvary [9], we use a model of consumer browsing inspired by that used by Brin and Page [4] in their computation of PageRank. However, we refine the browsing model slightly to account for the presence of sticky content. We assume $n := |V(G)|$ sites and a unit mass of consumers initially distributed according to a distribution $r^{(0)} := (r_1^{(0)}, \ldots, r_n^{(0)})$, with $r_i > 0$ for all i and $\sum_{i=1}^n r_i^{(0)} = 1$.[8] Consumers browse randomly in a sequence of stages $t = 0, 1, \ldots$. For simplicity, we assume that all consumers "log on" to the internet at the beginning of stage 0 and never "log off." This simplification seems appropriate, since we are only concerned with the steady-state traffic levels on each site.[9] In stage t, a consumer currently at site i will either remain at or follow a link from site i with probability δ ($0 < \delta < 1$). With probability $1 - \delta$, she will jump to a random website, choosing a given site i with probability $r_i^{(0)}$.[10]

We let $r^{(t)}$ be the distribution of consumers on websites at stage t of this process. If a well-defined limit distribution $r := (r_1, \ldots, r_n) = \lim_{t\to\infty} r^{(t)}$ results, then this distribution represents the expected number of visitors at each site. In the following sections, we will flesh out our models of sticky content and justify the convergence of the browsing process in our setting. For now, however, we assume the existence of the limit distribution r and use it to compute site' revenues.

We assume that, for every round a consumer spend at site i (including the round in which the consumer arrives at site i for the first time), that consumer purchases a product from site i with probability 1. The site's net revenue from commercial content sales is therefore proportional to $r_i(c_i - C)$. Note that this model captures a setting in which consumers buy products from site i with probability $0 < z_i < 1$, as we could set $\tilde{c}_i := z_i c_i$, and replace c_i in our model by $\tilde{c}_i$ (so that the commercial content revenue equation takes the form $r_i(\tilde{c}_i - C) = r_i(z_i c_i - C)$).

4. Attracting Sticky Content

Attracting sticky content is sticky content which increases individuals' likelihoods of returning to sites. To represent this behavior, we assume that the presence of attracting sticky content on website i increases the likelihood that a consumer will start browsing from website i. Practically, this means that many consumers choose i as their "homepage." Once at any site i, however, consumers are equally likely to stay at i or to follow any of the d_i^{out} links out of i.

This sort of consumer behavior might occur, for example, when sites increase their stickiness by listing local weather conditions. A consumer might adopt such a site as her homepage, as it is useful to receive regular updates regarding the local weather. However, a consumer is unlikely to spend an inordinate amount of time on such a site—after learning the weather conditions, she will continue her web traversal.

We model this process by assuming that the starting distribution $r^{(0)}$ is a function of attracting sticky content levels:

$$r^{(0)} = \left(\frac{s_1}{S}, \ldots, \frac{s_n}{S}\right), \tag{1}$$

where $S = \sum_{i=1}^n s_i$.[11] Thus, consumers are more likely to start their browsing process at sites with higher levels of sticky content. Movement of consumers across sites is defined by the transition probability matrix

$$M := (M_{ij})_{1\leq i,j\leq n} = \begin{cases} \frac{1}{d_i^{\text{out}}+1} & i \to j \\ 0 & i \not\to j, \end{cases}$$

so that once at a site i an individual is equally likely to remain at i or to follow any individual link out of i.[12]

The transition matrix M does not directly incorporate commercial content levels. Nonetheless, our Proposition 2 indicates that traffic towards high-quality sites arises endogenously: sites with high commercial content levels purchase the most inlinks, hence they draw substantial traffic.

With these definitions, we may write

$$r^{(t+1)} = \delta \cdot r^{(t)} \cdot M + (1 - \delta) \cdot r^{(0)}. \tag{2}$$

Since $r^{(0)}$ is a probability distribution and each row of M contains at least one nonzero element, the convergence of $r^{(t)}$ is guaranteed by the following result from the theory of Markov chains (see [10]).

[7] This is consistent with a result of Katona and Sarvary [9] which shows that in a model with endogenous price setting, $q_i = q(c_i)$ is increasing in commercial content levels.

[8] This is a departure from Katona and Sarvary [9], who assumed that $r^{(0)} = (\frac{1}{n}, \ldots, \frac{1}{n})$. As we will detail in our models of sticky content's effects on consumer browsing behavior, we allow any choice of $r^{(0)}$ which is a probability distribution.

[9] Our results are unchanged if we allow users to log on and off in each stage, so long as at most measure 0 of consumers log on or off at a time.

[10] This $1 - \delta$ "reset probability" represents the user opening a new web browser or typing a new URL into her browser window.

[11] We also assume that the s_i are such that the Markov chain discussed in (2) irreducible; this may in some cases force $s_i > 0$ for particular i.

[12] Note that the traditional PageRank computation ignores the presence of sponsored search links and only counts unsponsored ("reference") links towards a site's PageRank. We diverge from this convention because the "PageRank" in our model is not intended as a measurement of site quality—rather, it measures steady-state consumer traffic levels.

LEMMA 1. *If $r^{(t)}$ is a probability distribution for all t, then the sequence $\{r^{(t)}\}_{t=0}^{\infty}$ is convergent as $t \to \infty$.*

We may now examine the structure of the web network in the presence of attracting sticky content. By the global balance equations of Markov chains, the limit distribution $r = (r_1, \ldots, r_n)$ must satisfy

$$r_i = (1-\delta)\frac{s_i}{S} + \delta\left(\frac{r_i}{d_i^{\text{out}}+1} + \frac{r_{i1}}{d_{i1}^{\text{out}}+1} + \cdots \frac{r_{ik}}{d_{ik}^{\text{out}}+1}\right),$$

where $i1, \ldots, ik$ are the sites linking to site i. Since r_i is the entry traffic on site i, the number of visitors clicking on each outgoing link from i is $(\delta r_i)/(d_i^{\text{out}}+1)$. Thus, the total price p_i of an advertising link from site i is

$$p_i = q_i \frac{\delta r_i}{d_i^{\text{out}}+1}.$$

In this setting, the total utility of site i is

$$u_i = \underbrace{r_i(c_i - C)}_{\text{sale of content}} + \underbrace{p_i d_i^{\text{out}}}_{\text{sale of outlinks}} - \underbrace{\sum_{j \to i} p_j}_{\text{purchase of inlinks}}. \qquad (3)$$

4.1 The Network Equilibrium. Katona and Sarvary [9] show that there exists at least one Nash equilibrium in a simultaneous link-purchasing game with objective function (3). Furthermore, their Proposition 1 shows that in all of these equilibria:

1. The out-degree is a weakly decreasing function of content: for any sites $i, j \in V(G)$ with $c_i < c_j$ we have $d_i^{\text{out}} \geq d_j^{\text{out}}$.
2. If $c_i \neq c_j$ for all $i, j \in V(G)$, then in-degree and limit traffic levels are increasing functions of content.

When only attracting sticky content is present, we obtain the following surprising result regarding the equilibria of the simultaneous link-purchasing game.

PROPOSITION 2. *When sticky content is attracting (and so only affects the starting vector $r^{(0)}$) and prices q_i are fixed, the set of network equilibria is independent of the starting vector $r^{(0)}$.*

The model of Katona and Sarvary [9] is the special case of our model in which each site i has sticky content level $s_i \equiv s$, for some constant $s > 0$, hence it follows from Proposition 2 that the network equilibria in the presence of attracting sticky content satisfy the properties of the equilibria in the model Katona and Sarvary [9]. In particular, at least one network equilibrium exists.

The structure of the web is therefore robust to individuals' choices of homepages. Indeed, attracting sticky content has no effect on web network structure so long as prices are held fixed. However, levels of attracting sticky content do affect the *ex post* distribution of utility: the stickier sites will sell more commercial content and will receive more outgoing traffic than will less sticky sites. Consequently, the marginal benefit of sticky content is increasing as a function of the commercial content parameter.

5. Entrapping Sticky Content

Although the model of the prior section appropriately models many forms of sticky content, it is not all-encompassing. Some forms of sticky content not only impact individuals' starting decisions but also distract consumers from their explorations of the web. For example, webmail and internet game services may entrap consumers, causing them to remain on the same site for long periods of time. In this section, we model such *entrapping sticky content.*

Extending the model of Section 4, we allow the entrapping sticky content of a site i to impact the probability that a consumer will remain on i. We continue to parametrize the starting state $r^{(0)}$ as in (1), but now use the transition matrix

$$M := (M_{ij})_{1 \leq i,j \leq n} = \begin{cases} \frac{s_i}{d_i^{\text{out}}+s_i} & i = j \\ \frac{1}{d_i^{\text{out}}+s_i} & i \to j \\ 0 & i \not\to j. \end{cases}$$

Under the Markov process (2), the distribution $r^{(t)}$ is still a probability distribution at each stage, so long as $s_i > 0$ for all sites i. We henceforth assume $s_i > 0$ for all i, so that we may find (by Lemma 1) a limiting distribution $r = (r_1, \ldots, r_n) = \lim_{t\to\infty} r^{(t)}$ as before.[13] This distribution satisfies

$$r_i = (1-\delta)\frac{s_i}{S} + \delta\left(\frac{r_i s_i}{d_i^{\text{out}}+s_i} + \frac{r_{i1}}{d_{i1}^{\text{out}}+s_{i1}} + \cdots \frac{r_{ik}}{d_{ik}^{\text{out}}+s_{ik}}\right),$$

where as before $i1, \ldots, ik$ are the sites linking to site i. The total price p_i of an advertising link from site i is

$$p_i = q_i \frac{\delta r_i}{d_i^{\text{out}}+s_i}$$

and the utility of site i is once again given by (3).

5.1 The Network Equilibrium. It is clear that the distribution of sticky content does impact the equilibrium network structure in this model, so that no analogue of Proposition 2 holds in the presence of entrapping sticky content. In this section, we examine how entrapping sticky content affects the equilibria of the simultaneous link-purchasing game. We give a general characterization result and then discuss limiting cases.

PROPOSITION 3. *When sticky content is entrapping, there is at least one Nash equilibrium in the simultaneous link-purchasing game. In any such equilibrium, the out-degree is a weakly decreasing function of content: for any sites $i, j \in V(G)$ with $c_i < c_j$ we have $d_i^{\text{out}} \geq d_j^{\text{out}}$.*

This result ensures the existence of an equilibrium in the presence of sticky content and partially recovers the structure of the network obtained in the model of Katona and Sarvary [9] and in the prior section.

We now address limiting cases of the entrapping sticky content model, observing two corollaries which follow directly from the proof of Proposition 3. The first such result indicates what happens when the stickiness of some site i approaches 0.

COROLLARY 4. *In the limit as $s_i \to 0$ for a site i, the web graph approaches a network in which $j \to i$ if and only if*

$$c_i - C + \delta q_i \geq q_j,$$

for all sites $j \neq i$.

[13] If $s_i \equiv 1$ for all sites i, then we recover the model of Katona and Sarvary [9]. However, no other cases of the attracting sticky content model can be recovered in the setting of entrapping content.

Corollary 4 is intuitive—if consumers surf the web without actually spending time on site i, then i will buy a link from site j if and only if the profits expected from an inlink $j \to i$ are positive.

We may also examine the limiting case in which sites are arbitrarily sticky. As a site i becomes arbitrarily sticky, its outlink price has no effect on net profits. Thus, we obtain a second corollary.

COROLLARY 5. *In the limit as $s_i \to \infty$ for a site i, the web graph approaches a network in which $j \to i$ if and only if*

$$c_i - C \geq (1 - \delta) q_j,$$

for all sites $j \neq i$.

If a site i is sufficiently sticky, then a consumer reaching site i will likely remain at i until she decides to start a new traversal. An outlink from such a site i has little value, since it will rarely be clicked. The parameter δ mediates this effect, as it controls the likelihood that a consumer will follow an outlink before initiating a new traversal. In the extreme case in which $\delta = 1$, consumers at site j only leave via outlinks, hence the value of outlinks from j are valuable even in the presence of large amounts of entrapping content.

5.2 Endogenous Sticky Content Levels. In the prior sections, we treated sticky content levels as exogenous. Now, we relax this assumption so that we may examine the cross-complementarity between a site's commercial content and its sticky content. We need only examine this when sticky content is entrapping, since the results of Section 4.1 show that attracting sticky content is uniformly desirable for commercial websites. Throughout this section, then, sticky content is always assumed to be entrapping.

We continue to treat commercial content as exogenous and use the browsing model of Section 5. Now, however, we assume that sites may develop sticky content at a per-unit price of K. Then, the utility of site i is given by

$$\begin{aligned} u_i = \underbrace{r_i(c_i - C)}_{\text{sale of content}} + \underbrace{p_i d_i^{\text{out}}}_{\text{sale of outlinks}} \\ - \underbrace{\sum_{j \to i} p_j}_{\text{purchase of inlinks}} - \underbrace{K \cdot s_i.}_{\text{cost of sticky content}} \end{aligned} \quad (4)$$

We assume that if all sites are infinitely sticky, then no sites will purchase inlinks.[14] This is reasonable, since if all sites are sufficiently sticky then there are no benefits from purchased outlinks if $\delta < 1$—all consumers will be trapped on sticky sites until they decide to reinitiate browsing randomly.[15] By Corollary 5, this means that we must have $c_i < (1 - \delta) q_j + C$ for all sites i and j. It follows that for all sites i other than the site $i^* = \text{argmax}_{1 \leq i \leq n} c_i$ with the largest commercial quality level,

$$c_i < (1 - \delta) q_i + C. \quad (5)$$

For consistency, we assume that (5) holds for site $i = i^*$ as well. In practice, this would arise if there were even a small amount of uncertainty in the market. For example, if i^* fears the entry of a new site i^{**} with $c_{i^{**}} > c_{i^*}$, then it must set its price according to (5).

Additionally, we assume that the total sticky content level S is an exogenous constant. Although we require this assumption for reasons of tractability, it appears to be reasonable. Indeed, if one website develops and publicizes a type of sticky content, then other web face declining benefits from adding the same type of content.[16] Thus, if we think of our parameter s_i as representing the stickiness induced by a certain type of sticky content, the "total stickiness" S which can be created is bounded by a constant B. Assuming that consumers use the stickiness of the web to its maximum potential, this bound is actually achieved, $S = B$.

With these preliminaries, we may proceed with our discussion of endogenous sticky content. We suppose that the web network is previously established, so that each site i sets its sticky content level given knowledge of its incoming and outgoing links. We let s_i^* be the optimal level of sticky content for site i, assuming such a value exists. We make the following observation, which serves as a sort of benchmark.

PROPOSITION 6. *If a site i has no inlinks, then $\frac{\partial s_i^*}{\partial c_i} > 0$.*

The intution behind this result is clear. A site i with no inlinks can only profit when consumers start at site i. Therefore, site i would like for its stickiness s_i to be large. The marginal value of traffic through site i is increasing in c_i, hence the value of stickiness for site i is, as well.

For sites with inlinks, the comparative static of Proposition 6 is preserved, although its magnitude is reduced. We show this and more in our next proposition.

PROPOSITION 7. *Let $R_i := \sum_{j \to i} \frac{r_j}{S(d_j^{\text{out}} + s_j)}$. Then, s_i^* is well-defined when $R_i \leq \frac{(d_i^{out})^2}{S}$. Moreover,*

1. *we have $\frac{\partial s_i^*}{\partial c_i} > 0$ for any i such that $R_i \leq \frac{(d_i^{out})^2}{S}$, and*

2. *as $R_i \to \frac{(d_i^{out})^2}{S}$, we have that $\frac{\partial s_i^*}{\partial c_i} \to 0$.*

If $R_i < \frac{(d_i^{out})^2}{S}$ is sufficiently large, then site i would prefer not to have entrapping sticky content.[17]

The last component of this result indicates a striking difference between attracting sticky content and entrapping sticky content. While attracting sticky content is always desirable, entrapping sticky content can hurt the revenues of sufficiently popular sites. This at first appears unintuitive—why should a site ever prefer not to have any sticky content? The reason is that a site with sufficiently many inlinks stands to gain larger profits from flow traffic than from capturing consumers.[18] For an extreme example: if a site i has inlinks from every site on the internet, then any consumer who leaves site i will return within a short period of time. Thus, whereas site i derives only commercial revenues from trapped consumers, site i is assured a steady flow of

[14] By the proof of Corollary 5, this effectively occurs whenever all the stickiness levels s_i exceed some large, positive constant.

[15] The empirical computer science literature often assumes that $\delta \approx .85 < 1$, so requiring $\delta < 1$ seems appropriate.

[16] For example, the earliest webmail purveyors maintain have substantially larger numbers of consumers than do sites who only recently began offering webmail services.

[17] Combining this assertion with the fact that that $\frac{\partial s_i^*}{\partial c_i} > 0$ indicates that increasing commercial content can partially—but not completely—offset the negative effects of having excessive amounts of entrapping sticky content.

[18] This logic tacitly requires that outlink prices are sufficiently high; this is assured by (5).

both commercial and outlink revenues if it receives heavy incoming traffic and can avoid having consumers become entrapped within its pages.

We thus observe endogenous specialization of website business models. In our model, the most heavily trafficked sites want little entrapping content; these sites profit primarily through sponsored outlink traffic. By contrast, low-traffic sites profit primarily through sales of on-site content and hence prefer to entrap users. Katona and Sarvary [9] identified a similar effect, in which sites with low-quality commerical content endogenously specialize in outlink sale. Our result is substantially different from that of Katona and Sarvary [9], however, as the comparative static relevant to Proposition 7 is in a function of PageRank and sticky content levels and is independent of commercial content levels.

6. Discussion and Conclusion

Extending the model of Katona and Sarvary [9], we introduced the first model of sticky content, a form of noncommercial content prevalent on the web. We considered how attracting and entrapping sticky content may affect consumer behavior. We showed that the levels of attracting sticky content do not affect the equilibrium web network structure. However, such sticky content has substantial positive effects on *ex post* utility levels and is therefore uniformly desirable. By contrast, entrapping sticky content is actually undesirable for popular, highly trafficked sites.

As the actual level of content desired is mediated by the number of outlinks and the price of outgoing links, these results have implications for link purchase and marketing strategies. All sites should seek to add high-quality attracting sticky content. Practically, every commercial website i should seek to develop sticky content which leads consumers to select i as their homepage. Examples of such content include weather reports, news bytes, and search bars.[19] By contrast, highly trafficked sites with entrapping sticky content must sell a large number of outlinks; the number of outlinks required for such a site to be viable is an increasing function of the in-degree of the site.

Nonetheless, for a site with sufficiently many outgoing links, the marginal benefit of entrapping sticky content is always increasing in the quality of the site's commercial content. Thus, sites such as search engines and online comparison tools which maintain huge numbers of outlinks and high quality commercial content stand to benefit from all forms of stickiness.[20] Such sites should invite consumers to create accounts, so as to ensure consumer loyalty. They should also present "specials," to encourage consumers both return regularly and to browse deeper into the site's pages.[21]

6.1 Directions for Future Study. Like the model of Katona and Sarvary [9], our model is limited in its approach to consumer behavior. Although we have generalized and extended the PageRank-based model introduced by Katona and Sarvary [9], we have still maintained two problematic underlying assumptions: consumer search is homogeneous and random. However, this simplification is forced in part by the state of empirical knowledge, as inter-site consumer browsing behavior is not yet empirically well-understood. Clearly, it would be desirable to assess the implications of our model on real-world data, and to conduct more general empirical studies of consumer browsing patterns. Additionally, a few logical extensions to our model seem apparent: for example, attracting sticky content might draw consumer traffic to specific outlinks, or consumer browsing could entail search-costs similar to those employed in the model of Athey and Ellison [1] for sponsored search listings.

Also, our model assumes away competitive dynamics between sites. These dynamics do, indeed, have effects on the provision of sticky content. However, this limitation may not be material, as competitive dynamics appear to primarily affect the network structure. Instead, we feel that it would be better for future work to address several extensions to our framework, which we describe below.

6.1.1 Effects on Price Levels. It follows from the discussion in Section 5 that the effects of entrapping sticky content on per-click price levels q_i are unclear.[22] A sticky website i will attract more consumers beginning web traversals. However, if site i entraps its consumers, then its outgoing traffic levels decrease. These two symptoms of entrapping sticky content respectively increase and decrease the the optimal link price q_i. It would be interesting to examine the optimal price choice in this framework, as well as a joint optimization of price and sticky content. Such an exploration might give further information about the network structure, along the lines of Proposition 1(ii) of Katona and Sarvary [9].

6.1.2 Reference Links. Katona and Sarvary [9] address an extension to their model which introduces *reference links*, costless outlinks which sites create to increase their own content values. The addition of reference links substantially complicates the framework, but Katona and Sarvary [9] are still able to draw conclusions about network structure under stronger assumptions.

Our Proposition 7 shows that the presence of inlinks can reduce a site's desire to develop entrapping sticky content. Thus, it is likely that the presence of reference links to a site i will decrease the sticky content level desired by i. This is not entirely clear, however, as if site i also adopts reference links, then these links may devalue the outgoing traffic emanating from i.

6.1.3 Network Dynamics. Neither our model nor that of Katona and Sarvary [9] addresses the presence of dynamics in the evolution of the world wide web. This appears to be a serious omission, as web sites assuredly manage both their content levels and their advertising links dynamically, with attention to other sites' actions. Although a fully dynamic model of web network formation may be out of reach, even a sequential-game model of endogenous sticky content and link formation would improve our understanding of results such as Proposition 7.

[19] This explains why common homepage sites (such as browser and ISP "web portals") have moved to include search bars served by major search engines. In some sense, this also explains why many consumers choose to set their homepages to the sites of popular search engines.

[20] Although the commercial content of a search engine is not purchased by the consumers of the search features, its sale rate is proportional to consumer traffic, so we may think of it within our framework.

[21] An example of this latter behavior appears on `PriceWatch.com`, which maintains a "Tech Specials Going On Now!" box on its homepage. This box is updated at least as often as the site is reloaded.

[22] By contrast, it is nearly immediate that the optimal link price should be increasing in attracting sticky content levels.

7. Acknowledgements

The author was partly supported by a grant from the Harvard College Program for Research in Science and Engineering and by a National Science Foundation Graduate Research Fellowship. He is especially grateful to Susan Athey for supervising the work and for her commentary and support. He also thanks Zachary Abel, Pablo Azar, Peter Coles, Erik Demaine, Edward Glaeser, Andrea Hawksley, David Parkes, Sven Seuken, Andrei Shleifer, several anonymous referees, and participants in the Harvard EconCS Research Workshop for their helpful comments and suggestions.

8. References

[1] Athey, S., and Ellison, G. Position auctions with consumer search. Mimeo, Harvard University, 2008.

[2] Bagwell, K. The economic analysis of advertising. In *Handbook of Industrial Organization*, M. Armstrong and R. Porter, Eds., vol. 3. Elsevier, 2007.

[3] Bala, V., and Goyal, S. A noncooperative model of network formation. *Econometrica 68* (2000), 1181–1229.

[4] Brin, S., and Page, L. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems 30* (1998).

[5] Clarke III, I., and Flaherty, T. B. Web-based B2B portals. *Industrial Marketing Management 32* (2003).

[6] Haywood, A. J. Online auctions: User experience insights from eBay. Chimera Working Paper 2006-10, 2006.

[7] Immorlica, N., Jain, K., and Mahdian, M. Game-theoretic aspects of designing hyperlink structures. In *Proceedings of the 2nd International Workshop on Internet and Network Economics (WINE 2006)*, vol. 4286 of *Lecture Notes in Computer Science*. 2006.

[8] Jackson, M. O. A survey of models of network formation: Stability and efficiency. In *Group Formation in Economics: Networks, Clubs and Coalitions*, G. Demange and M. Wooders, Eds. Cambridge University Press, 2004.

[9] Katona, Z., and Sarvary, M. Network formation and the structure of the commercial world wide web. *Marketing Science 27* (2008), 764–778.

[10] Langville, A. N., and Meyer, C. D. Deeper inside PageRank. *Internet Mathematics 1* (2004), 335Ű–400.

[11] Langville, A. N., and Meyer, C. D. *Google's PageRank and Beyond: The Science of Search Engine Rankings*. Princeton University Press, 2006.

[12] Lewis, D., and Bridger, D. *The Soul of the New Consumer*. Nicholas Brealey Publishing, 2001.

[13] Mayzlin, D., and Yoganarasimhan, H. Link to success: How blogs build an audience by promoting rivals. Mimeo, Yale School of Management, 2008.

[14] Pandurangan, G., Raghavan, P., and Upfal, E. Using PageRank to characterize web structure. *Internet Mathematics 3* (2006), 1–20.

[15] Sgroi, D. Social network theory, broadband and the world wide web. Cambridge Working Paper in Economics 0603, 2006.

[16] Slikker, M., and van den Nouweland, A. Network formation models with costs for establishing links. *Review of Economic Design 5* (2000), 333–362.

[17] Stephen, A. T., and Toubia, O. Deriving value from social commerce networks. *Journal of Marketing Research*, forthcoming.

Appendix

Proof of Proposition 2. We may write the limit traffic r_i of a site i as

$$r_i = \frac{d_i^{\text{out}} + 1}{d_i^{\text{out}} + 1 - \delta} \left((1-\delta)\frac{s_i}{S} + \delta \sum_{j \to i} \frac{r_j}{d_j^{\text{out}} + 1} \right). \quad (6)$$

The utility function u_i therefore takes the form

$$u_i = \frac{d_i^{\text{out}} + 1}{d_i^{\text{out}} + 1 - \delta} \left(c_i - C + \delta q_i \frac{d_i^{\text{out}}}{d_i^{\text{out}} + 1} \right) (1-\delta)\frac{s_i}{S} + \delta \sum_{j \to i} r_j \left(\frac{\frac{d_i^{\text{out}}+1}{d_i^{\text{out}}+1-\delta} \left(c_i - C + \delta q_i \frac{d_i^{\text{out}}}{d_i^{\text{out}}+1} \right) - q_j}{d_j^{\text{out}} + 1} \right). \quad (7)$$

The first term of (7) is independent of the decision of site i. Site i therefore purchases links to maximize the second term,

$$\delta \sum_{j \to i} r_j \left(\frac{\frac{d_i^{\text{out}}+1}{d_i^{\text{out}}+1-\delta} \left(c_i - C + \delta q_i \frac{d_i^{\text{out}}}{d_i^{\text{out}}+1} \right) - q_j}{d_j^{\text{out}} + 1} \right). \quad (8)$$

However, $r_j > 0$ for all j and the term

$$\frac{\frac{d_i^{\text{out}}+1}{d_i^{\text{out}}+1-\delta} \left(c_i - C + \delta q_i \frac{d_i^{\text{out}}}{d_i^{\text{out}}+1} \right) - q_j}{d_j^{\text{out}} + 1}$$

of (8) is independent of the sticky content level s_i. The result follows.

Proof of Proposition 3. Our approach follows that used by Katona and Sarvary [9] in the proof of their Proposition 1. First, we prove that any equilibrium satisfies the claimed condition. The limit traffic r_i of site i is

$$r_i = \frac{d_i^{\text{out}} + s_i}{d_i^{\text{out}} + s_i(1-\delta)} \left((1-\delta)\frac{s_i}{S} + \delta \sum_{j \to i} \frac{r_j}{d_j^{\text{out}} + s_j} \right).$$

The utility function u_i therefore takes the form

$$u_i = \frac{d_i^{\text{out}} + s_i}{d_i^{\text{out}} + s_i(1-\delta)} \left(c_i - C + \delta q_i \frac{d_i^{\text{out}}}{d_i^{\text{out}} + s_i} \right) (1-\delta)\frac{s_i}{S} + \delta \sum_{j \to i} r_j \left(\frac{\frac{d_i^{\text{out}}+s_i}{d_i^{\text{out}}+s_i(1-\delta)} \left(c_i - C + \delta q_i \frac{d_i^{\text{out}}}{d_i^{\text{out}}+s_i} \right) - q_j}{d_j^{\text{out}} + s_j} \right). \quad (9)$$

As we found in the Proof of Proposition 2, then, the utility function u_i splits into two terms, one of which is independent of the decision of site i. Site i therefore purchases links to maximize the second term of (9). Assuming that sites buy any links to which they are indifferent, site i will buy a link from site j if and only if

$$\frac{d_i^{\text{out}} + s_i}{d_i^{\text{out}} + s_i(1-\delta)} \left(c_i - C + \delta q_i \frac{d_i^{\text{out}}}{d_i^{\text{out}} + s_i} \right) \geq q_j. \quad (10)$$

Holding fixed the actions of sites $j \neq i$, the left-hand term of (10) is a constant specific to site i.

Now, if $q_k < q_j$ for sites j and k, then in equilibrium all sites who buy a link from site j must also buy a link from site k. Since the prices $q_i = q(c_i)$ are increasing in commercial content, the result follows.

Now, to obtain an equilibrium, we extend the strategy space slightly to capture mixed-strategy equlibria of the original game. It is well-known that any game with a convex, compact strategy space and continuous payoff function which is quasi-concave in the players' own strategies has a pure-strategy Nash equilibrium. As in Katona and Sarvary [9], we extend our model to allow for "partial links," making the strategy space continuous: a site may establish a partial link with weight $0 < w \leq 1$, paying fraction w of the cost of a full link and receiving a w proportion of traffic. The payoff to site i of a weight-w inlink from site j is given by

$$\delta w \left(\frac{\frac{d_i^{\text{out}}+s_i}{d_i^{\text{out}}+s_i(1-\delta)} \left(c_i - C + \delta q_i \frac{d_i^{\text{out}}}{d_i^{\text{out}}+s_i} \right) - q_j}{d_j^{\text{out}} + s_j(1-\delta)} \right) \cdot \left((1-\delta)\frac{s_j}{S} + \delta \sum_{k \to j} \frac{r_k}{d_k^{\text{out}} + s_k} \right).$$

It follows that the utility function u_i is quasi-concave. Since the extended strategy space is compact and convex, at least one pure-strategy Nash equilibrium of the extended game exists. Furthermore, in this equilibrium each site establishes at most one partial link.[23]

Proof of Proposition 6. If site i has no inlinks then its traffic arises only through consumers starting at i. Then,

$$r_i = \frac{d_i^{\text{out}} + s_i}{d_i^{\text{out}} + s_i(1-\delta)}(1-\delta)\frac{s_i}{S}$$

and the utility function u_i is given by

$$u_i = \frac{d_i^{\text{out}} + s_i}{d_i^{\text{out}} + s_i(1-\delta)} \left(c_i - C + \delta q_i \frac{d_i^{\text{out}}}{d_i^{\text{out}} + s_i} \right) (1-\delta)\frac{s_i}{S} - K s_i.$$

We may then compute explicitly that

$$\frac{\partial^2 u_i}{\partial c_i \partial s_i} = \frac{1}{S}\left[1 - \delta \left(\frac{d_i^{\text{out}}}{d_i^{\text{out}} + s_i(1-\delta)} \right)^2 \right], \tag{11}$$

$$\frac{\partial^2 u_i}{\partial s_i^2} = -\frac{2(1-\delta)\delta(d_i^{\text{out}})^2(C - c_i + (1-\delta)q_i)}{S(d_i^{\text{out}} + s_i(1-\delta))^3}. \tag{12}$$

By assumption (5), we know that $c_i < (1-\delta)q_i + C$, so (12) is strictly negative. Since (11) is strictly positive, the result follows from the formula

$$\frac{\partial s_i^*}{\partial c_i} = -\frac{\frac{\partial^2 u_i}{\partial c_i \partial s_i}}{\frac{\partial^2 u_i}{\partial s_i^2}}.$$

Proof of Proposition 7. The first-order condition of utility with respect to sticky content is

$$\begin{aligned} &\frac{c_i - C - KS - \frac{\delta(c_i - C - (1-\delta)q_i)(d_i^{\text{out}})^2}{(d_i^{\text{out}} - \delta s_i + s_i)^2}}{S} \\ &\qquad + \delta \sum_{j \to i} r_j \frac{d_i^{\text{out}}(c_i - C - (1-\delta)q_i)}{(d_i^{\text{out}} - \delta s_i + s_i)^2 (d_j^{\text{out}} + s_j)} \\ &= \frac{c_i - C - KS - \frac{\delta(c_i - C - (1-\delta)q_i)(d_i^{\text{out}})^2}{(d_i^{\text{out}} - \delta s_i + s_i)^2}}{S} \\ &\qquad + \delta \frac{d_i^{\text{out}}(c_i - C - (1-\delta)q_i)}{(d_i^{\text{out}} - \delta s_i + s_i)^2} \sum_{j \to i} \frac{r_j}{(d_j^{\text{out}} + s_j)}. \end{aligned}$$

We first examine what happens if a site i has no inlinks. In this case, setting the first-order condition to 0 and solving yields

$$s_i^* = \frac{(\delta - 1)d_i^{\text{out}} + \frac{\sqrt{(\delta-1)^2\delta(d_i^{\text{out}})^2(C - c_i + (1-\delta)q_i)}}{\sqrt{C - c_i + KS}}}{(\delta-1)^2}.^{24} \tag{13}$$

So that s_i^* is well-defined in this case, we must have

$$C + KS > c_i, \tag{14}$$

from equations (5) and (13).[25] Taking the partial derivative of s_i^* (in (13)) with respect to c_i, we find that

$$\frac{\partial s_i^*}{\partial c_i} = \frac{\delta(d_i^{\text{out}})^2((1-\delta)q_i - KS)}{2(C - c_i + KS)\Phi}, \tag{15}$$

where

$$\Phi = \sqrt{(\delta-1)^2\delta(d_i^{\text{out}})^2(C - c_i + (1-\delta)q_i)(C - c_i + KS)}.$$

Proposition 6 shows that (15) is positive, hence

$$(1-\delta)q_i > KS \tag{16}$$

when site i has no inlinks.

Since all the constants in (14) and (16) are determined in advance of network formation, these conditions must hold for any site i, irrespective of whether i has inlinks.

Now, for a general site i, the optimal level of sticky content is given by

$$s_i^* = \frac{(\delta-1)d_i^{\text{out}} + \frac{\sqrt{(\delta-1)^2(C - c_i + (1-\delta)q_i)\delta\left((d_i^{\text{out}})^2 - R_i S\right)}}{\sqrt{C - c_i + KS}}}{(\delta-1)^2};$$

the last part of the proposition follows since this is expression is continuous in R_i and negative at $R_i = \frac{(d_i^{\text{out}})^2}{S}$. Then, we may compute the comparative static of s_i^* with respect to c_i:

$$\frac{\partial s_i^*}{\partial c_i} = \frac{\delta\left((d_i^{\text{out}})^2 - R_i S\right)((1-\delta)q_i - KS)}{2(C - c_i + KS)^{3/2}\Psi}, \tag{17}$$

where

$$\Psi = \sqrt{(\delta-1)^2\delta(C - c_i + (1-\delta)q_i)\left((d_i^{\text{out}})^2 - R_i S\right)}.$$

The expression (17) is positive when $R_i < \frac{(d_i^{\text{out}})^2}{S}$ and approaches 0 as $R_i \to \frac{(d_i^{\text{out}})^2}{S}$.

[23] A site must be indifferent about a link in order to establish a partial link, since if there is a profit increase from a partial link then there is a larger profit increase from a full link. In equilibrium, a site may be indifferent about at most one link.

[24] In determining s_i^*, we solved a quadratic equation and maintained the larger of the two roots. (The smaller root is uninteresting for our purposes, as it is always negative.)

[25] We assume that s_i^* is well-defined in the special case in which i has no inlinks.