A Dynamic Competitive Analysis of Content Production and Link Formation of Internet Content Developers

Liye Ma

Carnegie Mellon University

Job Market Paper[1]

September 2010

PRELIMINARY VERSION – PLEASE DO NOT CIRCULATE OR CITE WITHOUT AUTHOR'S PERMISSION

[1] I am indebted to Baohong Sun and Kannan Srinivasan for their guidance. I benefited from discussions with Kinshuk Jerath, David Krackhardt, Yaroslav Kryukov, and Joachim Vosgerau. I also thank seminar participants at Carnegie Mellon University for their helpful comment. All errors are mine.

A Dynamic Competitive Analysis of Content Production and Link Formation of Internet Content Developers

Abstract

The emergence of hundreds of revenue sharing content websites has greatly contributed to the proliferation of Internet social media. Content at these websites is supplied by external independent developers, whom the websites attract through revenue sharing. This leads to a competition among developers, as each tries to attract viewership to her own content. A feature recently introduced at many sites, namely allowing developers to link to one another, leads to even richer interactions in this competition, and its impact on content production and overall website viewership is little understood. In this study, we develop a dynamic oligopoly model for the competition among content developers at a website. Each developer produces content and forms links to maximize her discounted viewership net of cost of actions, and their strategic interaction is characterized as a Markov-perfect equilibrium. Applying the two-step estimator of Bajari, Benkard, and Levin (2007) to the data obtained from a popular Internet product review site, we investigate the following issues: (1) why and how do developers form links? (2) Will linking encourage or discourage content production? (3) What market structure will emerge? (4) Will linking increase or decrease the overall website viewership? We find that reciprocal links are naturally encouraged by a promote-the-promoter effect. This in turn induces developers with more content to strategically initiate links to invite reciprocation. In addition, we find that link formation affects the incentive to produce content – developers with more content and unfavorable network positions are encouraged to produce, while developers in the opposite states are discouraged. Furthermore, the current linking policy may impede competition by giving competitive advantage to a subgroup of content developers, and our simulation suggests that limiting links could increase overall viewership by 17%. Our study is among the first to examine the interdependence between online link formation and contention production in a dynamic and competitive setting.

Keywords: Internet content, social media, network, producers, dynamic game, empirical IO

1. Introduction

Content is the lifeblood of Internet marketing. The emergence of hundreds of revenue sharing content websites has greatly contributed to the recent proliferation of social media. A wide range of content vital for online business and consumer activities is provided at these websites: product reviews at Epinions.com facilitate online retailing; video clips at Youtube.com generate advertising revenue; articles at Fool.com attract subscribers, etc. Millions of viewers visit these websites on a monthly basis, making them a major component of Internet business (Table 1). Such websites typically generate revenue through advertising or sales referral. Consequently, their success depends crucially on the amount of viewership traffic they can attract.[2]

A key characteristic of such revenue sharing content sites is the *democratization of content*: instead of hiring employees to create content, companies operate these websites as platforms where external, independent developers come to supply content. Since the success of the websites depends crucially on the viewership their content attracts, the websites must encourage the independent content developers, or producers, to produce actively.[3] To encourage content production, website companies typically share revenue with each producer based on the viewership her content attracts. Interestingly, this creates an intra-website competition among the independent producers, as each seeks to maximize *the viewership of her own content*, and when viewers come to the website and choose among different producers' content, producers effectively compete against each other for viewership. To attract viewership, producers naturally

[2] Display advertising fee can be charged on a pay-per-impression basis, with rates quoted in "cost per milli", which is the fee for every thousand times the advertisement is viewed, or on a pay-per-click basis, where a fee is charged every time an advertising link is clicked. Sales referral commission is often charged on a pay-per-action basis, where a content site is paid based on the sales it helps e-commerce sites generate by directing viewers to those sites. The amount of viewership traffic is the key to all these revenue models.

[3] Both "developer" and "producer" are widely accepted terms in the industry, and they are used interchangeably in this study.

need to actively produce content – the more content a producer provides, the more likely a viewer will find what she needs from that producer, and the higher her viewership.[4]

[Insert Table 1 About Here]

Making the competition more intriguing is another feature that is being increasingly introduced to such websites: *inter-producer linking*. As Table 1 shows, most such sites now allow producers to create links pointing to other producers at the site. Links may vary by name, such as trust, favorite, follow, etc, but all serve as a form of endorsement of the target by the source, and make the target's content easily accessible from the source's. Such links together form a *producer network* that evolves over time. Since Internet viewers often navigate through links to view content, and search engines also rely on the link structure to rank search results, where a producer is positioned in this network significantly influences the viewership of her content. In general, the more incoming links a producer has, and the better positions the sources of the links have, the better is her position in the network (Brin and Page 1998). This is because incoming links drive viewership traffic to a producer's content, and a producer with more and better incoming links also gets preferential placement when search engine displays search results.

The introduction of inter-producer linking leads to several intriguing questions. Marketing research on content and linking is still at the early stage. Existing research has shown, in a static and analytical setting, that linking can promote the position of the target, and meanwhile enhance the content of the source – a viewer may visit a producer even if she does not have the desired content, if she can point to another producer who does (Mayzlin and Yoganarasimhan 2008, Katona and Sarvary 2008).[5] However, questions related to link formation in a dynamic context

[4] Other factors also matter, such as the quality and diversity of content, and will be accounted for in this study.
[5] This refers to the extension on reference links in Katona and Sarvary (2008). The main model of that paper focuses on advertising links which are price mediated, which does not apply to the situations in our study, as the links among content producers at these sites are not bought and sold but established by the sources on volition.

and the interaction of linking and content production decisions largely remain open. For example, how do producers form links over time, how do producers adjust their production decisions under the presence of linking, and how does one respond to others' decisions? More importantly, from the perspective of the website, would allowing producers to link encourage or discourage content production, and would it increase or decrease the overall viewership at the website? The objective of the websites introducing the linking feature is certainly to encourage production and increase traffic. But to find out whether this objective is met, we need a detailed understanding of how content producers interact with one another as they compete for viewership. Considering this, we address the following questions in our study: (1) What drives a producer's linking decisions over time, and when and to whom would she link to? (2) Will the ability to form links encourage or discourage a producer to produce content, and how does this impact vary across producers? (3) What market structure will emerge from this competition through content production and link formation under a given linking policy design at a website? (4) Finally, what is the overall effect of linking on the viewership at the website level, and should the website company regulate linking? Since these websites rely on the producers producing content to attract viewers, yet they can only incentivize but cannot control those producers, answers to the above questions are crucial to help the website companies understand content producers' decision process, draw implications from it, and improve their platform design.

In this study, we model the competition among content producers at a website as a dynamic game. In our model, each producer chooses her actions (produce content and link to other producers) over time to maximize her payoff – discounted viewership net of costs incurred in producing content and forming links. Producers adopt Markov strategies, and such strategies together constitute a Markov-perfect equilibrium, or MPE (Maskin and Tirole 1988, Ericson and

Pakes 1995). The equilibrium characterizes the dynamic interactions among content producers and the tradeoffs they face. In making her decisions, a content producer balances the cost and benefit of her actions, both immediate and in future, and accounts for the strategic reactions from other producers, as one's actions can change the competitive positions of others. We estimate the model using the two-step estimator recently developed by Bajari, Benkard, and Levin (2007). Applying the model and estimation approach to a dataset obtained from a popular Internet product review website, we estimate the viewership demand and cost functions, and analyze the driving forces of producers' decisions and their implications.

Our study leads to several findings. We first demonstrate that link formation is a dynamic strategic decision. We show that the nature of the competition encourages reciprocity – linking to someone who already links back – due to a *promote-the-promoter* effect. In the dynamic context, this tendency towards reciprocity further encourages certain producers to *strategically initiate non-reciprocal links* in anticipation of the reciprocation from targets, which increases viewership in future through improved position brought about by incoming links. We find that a producer with higher content volume is more likely to strategically initiate such links to "invite" reciprocation. Next, we find the dynamic effect of linking can *either encourage or discourage content production*, depending on the situations of the producers: to obtain and in anticipation of future rewards through incoming links, a producer will produce the most content when she has high content volume but low network position. Meanwhile, the prospect of linking discourages a producer with low content volume but high network position from producing content, as she expects her relative network position to diminish over time.

Furthermore, our analysis suggests that the current linking design overall could impede competition. We find that although both more content and higher network position lead to higher viewership, only the latter leads to higher net benefit once cost is accounted for. Thus potential advantage from having more content is mostly competed away, yet significant competitive advantage is accrued to better network position. That a subgroup of producers enjoys sustainable advantage over others may soften the competition, and lead to inefficiency from the website's perspective. This is confirmed in our simulation, which suggests that alleviating the imbalance through reducing links could lead to higher overall viewership at a website.

We contribute to the literature by jointly modeling content production and link formation decisions, investigating their inter-dependence in a dynamic setting, and evaluating the impact of linking when both decisions are determined endogenously. Existing studies have analyzed the impact of commerce network on firm profits (Stephen and Toubia 2009) without explicitly modeling the formation process of such network, and modeled the formation of content networks on the web in a static setting where content is exogenously given (Katona and Sarvary 2008). Our study extends the literature by analyzing how linking and content production decisions interact with each other, and we evaluate the impact of linking on website viewership when its effect on content production is accounted for. Furthermore, by studying the decision process and competition in a dynamic context, we show how inter-temporal tradeoffs and the strategic interactions among producers drive decisions over time, which cannot be shown in a static framework, such as the strategic invitation of reciprocal links and the content production in anticipation of incoming links from other producers. We also contribute to the literature by providing a rational economic framework for empirically analyzing the formation of links in a dynamic strategic setting. Our empirical findings provide much needed recommendations to industry managers.

The rest of the paper is organized as follows. In section 2 we review relevant literature. We then develop the dynamic game model in section 3. Following that, we discuss in section 4 the approach used for estimating this model. Section 5 discusses the empirical application, where we explain the data used in our study, analyze the result, and discuss the simulation. Finally, we conclude in section 6.

2. Relevant Literature

Our work is related to the broad literature on Internet content and on economic networks. Marketing researchers have shown great interest in Internet content, specifically on product reviews and online word-of-mouth (WOM). Chevalier and Mayzlin (2006) investigate the effect of online book reviews on sales, and find that improvement in reviews leads to higher relative sales. Godes and Mayzlin (2004) find that the dispersion of conversation in online communities has explanatory power on TV ratings. Chintagunta et al. (2010) find the valence of online reviews influence the box-office sales of movies. While the effect of online product reviews has been studied frequently, relatively less attention has been paid to the supply of such reviews, especially when they are supplied as information goods with profit incentive. Supply-side structural models have generally only recently gained attention in marketing (Srinivasan 2006), and our work fills in this gap in the case of Internet content.

Our work is also related to the formation of economic networks and their impacts. A rich literature exists on the formation of social and economic networks. For example, Bala and Goyal (2000) develop a non-cooperative game model to study linking decisions. Jackson (2004) gives an extensive survey on network formation literature with emphasis on stability and efficiency. Most studies use certain general value functions arising from network; while given the wide variety of networks, it is reasonable to expect that the benefit of the network, and its formation in turn, be situation specific. Two studies in marketing focus on the creation of links online. Mayzlin and Yoganarasimhan (2008) investigates why an author of an Internet blog may link to another competing blog, even though doing so effectively promotes her rival. They show that the ability to link to information is valuable to readers in addition to the ability to produce the information – if the blog does not have the information, readers will still appreciate a link to another blog that does. The borrowed content effect in our study models this effect. Katona and Sarvary (2008) study the formation of links among content sites as a non-cooperative game, where links are created either for paid advertising or for reference effect in the extended model. In both studies, the content at the websites is treated as exogenous. In contrast, Stephen and Toubia (2009) study the effect of online commercial networks. They find that allowing online retailers to link to one another creates economic value, and such value comes from improved accessibility. The study focuses on the effect of the network and does not explicitly address its formation process. Our study contributes to the literature by jointly studying both network formation and content production decisions and highlighting their interaction effect in a dynamic setting.

Our work draws from the rich literature on empirical industrial organizations from the methodology perspective. Specifically, we adopt the concept of Markov perfect equilibrium, or MPE (Maskin and Tirole 1988, Ericson and Pakes 1995, Maskin and Tirole 2001), for modeling dynamic oligopolistic competitions. Early estimation methods for MPE (Pakes and McGuire 1994, Pakes and McGuire 2001) extend the nested fixed point approach (Rust 1987) to explicitly compute equilibrium strategies. But the high dimensionality of typical dynamic competition models restricts the use of such methods to games with only few players. Recent advancement leads to several two-step estimators (Aguirregabiria and Mira 2007, Bajari, Benkard and Levin

2007, Pakes, Ostrovsky and Berry 2007) which extend the conditional choice probabilities approach (Hotz and Miller 1993). Such two step estimators bypass explicit computation of equilibrium by calculating continuation values through forward simulation, and by doing so enable the estimation of dynamic games with many players. Ackerberg et al. (2007) provides a comprehensive survey of these estimation methodologies. We implement the estimator developed in Bajari, Benkard and Levin (2007), hereafter BBL. The BBL estimator has been used for studies in industrial organizations (e.g. Ryan 2009), and has been adopted in marketing literature recently (Yao and Mela 2010).

3. Model

We discuss the model in this section. To prepare for the model, we begin with a brief summary of the key elements of the *industry setup*. We consider a content *website* on the Internet. *Viewers* come to the website to view content, which is produced by external, independent *content producers*, whom the website attracts through revenue sharing.

Each content producer seeks to maximize the viewership *of her own content* over time. In addition to *producing content*, a producer can *create links* pointing to other producers. Since viewers can easily follow a link to navigate to the target producer's content from the source producer's, a link benefits the *target* producer by putting her in a good position to receive viewership traffic. Furthermore, when viewers search for a specific topic and the content from multiple producers matches that search criteria, the search engine ranks the search results based on the linking structure, where producers with more incoming links and links from other producers with good positions receive preferential placement. Links thus again help the *targets* through this *positional benefit*. For the *source* of a link, the benefit is to *enhance content*, as a producer who links to other producers makes it convenient for viewers to find the content they want, and will be favored by viewers.

This industry setup leads to a *competition* among content producers, since each producer cares about her own viewership only, and viewers choose the content from multiple producers.[6] To attract viewership effectively, each producer must make her production and linking decisions while taking into account her own situation, other producers' situations, and the strategic response to her actions by other producers. She also needs to balance current and future benefits. Such considerations lead to interesting *dynamic interactions*. For example, more content attracts higher viewership, but producing content also incurs a cost. Depending on a producer's position, this cost-benefit tradeoff may or may not justify production. However, having more content may also attract links from other producers, which improves her position later on. This additional benefit could make content production worthwhile, even if it does not attract much immediate viewership. Such dynamic interactions among maximizing agents call for a dynamic oligopoly model, which we use in this study.

In our model, there are J independent content producers competing for viewership. Time is discrete and is indexed by t, $t = 1,2,...$. In each time period, each producer decides whether to produce content and whether to link to other producers. In the following subsections, we first describe the viewership demand market that clears in each time period given producers' content states and the link structure. We then discuss producers' dynamic content production and link formation decisions, and how content and link structure evolve according to such decisions.

[6] For example, a viewer may search for a topic, and read only the top two articles on the list retrieved by the search engine. In this case, each producer wants her content placed in the top two positions, and is competing against other producers for that.

Finally, we explain the dynamic competition and the equilibrium concept, and discuss the tradeoffs faced by producers which shape their strategies.

3.1 Viewership Demand

There are M consumers, or viewers, in each period.[7] Each viewer chooses to view the content of one content producer among the J producers at the website, or chooses to go to an external website, i.e. the *outside option*. This viewership constitutes the demand for producers' content. We adopt a logit demand model, which has been widely used in modeling oligopolistic competitions (e.g. Berry 1994, Berry et al 1995, Dube et al 2009), to characterize viewership demand in this per-period market.[8] The discrete-choice framework of the logit demand model reflects the competitive nature of the viewership demand, i.e. viewership of one producer's content may come the cost of another's. A viewer i's latent utility from reading the content of producer j in period t is:

$$(1) \qquad u_{i,j,t} = \begin{cases} \bar{u}_{i,j,t} + \varepsilon_{i,j,t} = f(C_{j,t}, P_{j,t}, C^b_{j,t}; \beta_i) + g(Q_j, Q^b_{j,t}; \gamma) + \varepsilon_{i,j,t} & j = 1..J \\ 0 + \varepsilon_{i,0,t} & j = 0 \end{cases}$$

In equation (1), $\bar{u}_{i,j,t} = f(C_{j,t}, P_{j,t}, C^b_{j,t}; \beta_i) + g(Q_j, Q^b_{j,t}; \gamma)$ is the deterministic component of the utility. $C_{j,t}$ is the content *quantity* of producer j at time t, $P_{j,t}$ is a numeric measure of her *network position*, and Q_j is a vector of quality variables of the producer that remains constant

[7] The terms "viewer" and "reader" are used interchangeably in this study.

[8] The logit demand model is based on a discrete-choice framework, yet it is possible that a reader may read multiple articles of a producer in a period, e.g., reading the product reviews of different products, or the content of several producers. An in-depth modeling of such behavior requires detailed clickstream data of readers which we unfortunately do not have. Instead, we treat each pageview as one single viewer in our model (that is, if a viewer reads three product review articles in the period, it is counted as three viewers in the model). This reduced-form treatment of readership demand can be improved by explicitly modeling a viewer's navigation behavior, which we leave for future research as richer data become available.

over time.[9] Furthermore, $C^b_{j,t}$ measures the total quantity of *borrowed content*, i.e. content derived from linking to other producers. Similarly, $Q^b_{j,t}$ measures the average quality of the producers being linked to. These measures are explained in detail later when we discuss producer actions and the network structure. The function $f(.;\beta_i)$ specifies how content, network position, and borrowed content enter into the utility function, with β_i as the parameter. Since viewer navigation behavior is not explicitly modeled, we estimate multiple specifications of functional forms for $f(.;\beta_i)$, with the best specification chosen through model selection. The function $g(.;\gamma)$ captures the quality differentiation among producers. Quality is used mainly for control purpose in our study, so we adopt a linear specification with γ as the parameter: $g(Q_j, Q^b_j;\gamma) = (Q_j, Q^b_j)^t \gamma$.

The relative attractiveness of a producer is determined by the amount of content she has, i.e. the content quantity, the location of the producer in the network, i.e. the network position, and the quality of the producer. Furthermore, the attractiveness of a producer is also influenced by the content of the other producers she links to. Intuitively, the more content a producer has, the more viewership she would receive, as viewers are more likely to find the content they want. Similarly, the more prominent a producer's position in the network, the higher viewership demand she would receive, as her content will receive more preferential placement by the search engine, and more viewers may be directed to her content when they navigate through the links. Borrowed content should further enhance a producer's attractiveness due to the convenience benefit it affords the viewers. We expect these to be reflected from the parameter vector β_i in

[9] In our model, we treat quality as a characteristic of the producers instead of content. This assumes away potential variation of quality across different content produced by the same producer. This is a reasonable assumption in the context of our study, since the quality of individual content is not observed before a viewer decides to view the content.

accordance with the specific functional form. For example, we expect all coefficients to be positive if factors enter the utility function linearly.

Finally $\varepsilon_{i,j,t}$ is an i.i.d random component which follows the type I Extreme Value distribution, resulting in the familiar logit probability of viewer i choosing producer j at time t:

(2) $$\Pr_{i,j,t} = \frac{\exp\{\bar{u}_{i,j,t}\}}{1+\sum_{j'=1}^{J}\exp\{\bar{u}_{i,j',t}\}}$$

Note that this viewership model is a reduced form one, and assumes away any explicit state-dependence on viewer's side. In reality, a viewer's behavior in one period may be influenced by her past behaviors, e.g. she becomes a routine follower of a content producer. In our model, this dependence can come indirectly through the persistence of a producer's state: a product review of an obsolete product produced earlier may be of no value now, but it attracted viewers at that time, some of whom then continues to visit the producer's page, and this is reflected in the utility function where a cumulative measure of content is used.[10]

Viewers may have different navigation patterns and content requirements, which results in different relative emphasis placed on different components in the utility function.[11] This heterogeneity is captured using a *latent class* approach (Kamakura and Russell 1989). That is, we assume there are N segments of viewers, each characterized by its own set of coefficients, $\{\beta_n\}_{n=1..N}$, and portion of each type is denoted as λ_n, so that $\sum_{n=1}^{N}\lambda_n = 1$.

[10] Since the emphasis of our study is on producer's production and linking behavior, structurally modeling viewer's persistence over time adds great complexity to the model but might not provide much added value. It also requires detailed viewer navigation data. We leave the joint structural modeling of producer and consumer behavior for future research.

[11] In the case of a sequence of page views, certain page views may be related more to the page content (e.g. following a topic search) while others may be related more to network positions (e.g., navigating through links or using a search engine that accounts for network positions). The heterogeneity also captures this effect, since a viewer in the model actually corresponds to a viewer-page view pair in the real world, as discussed earlier.

3.2 Content Producer

In any time period, a content producer j is characterized by a collection of variables: $\{C_{j,t}, P_{j,t}, C^b_{j,t}, Q_j, Q^b_{j,t}\}$. Content, network position, and borrowed content all evolve over time according to the actions of both producer j and other producers. A producer can take two types of actions, *content production* and *link formation*. We discuss these actions below and how the variables evolve according to these actions.

3.2.1 Content Production

A producer's content quantity, $C_{j,t}$, is determined solely by her own production decisions over time. In each period, a producer decides whether to produce additional content to add to her webpage – write another product review, break another news story, create another analytical report, etc – and if yes, the amount of content to produce. We denote this action by producer j at time t as $a^p_{j,t}$, where the superscript p indicate it is the production decision. Specifically,

$$a^p_{j,t} = \begin{cases} 0 & \text{do not produce content} \\ k & \text{produce } k \text{ units of content, } k \in \{1,2...\} \end{cases} \tag{3}$$

In the equation, k represents the number of units of content produced. Each unit of content may correspond to an article in the real world, thus the action is discrete.

Producing content increases the content quantity at a producer's webpage, $C_{j,t}$. Meanwhile, there is an opposite, *depreciation*, force at work: a product review will become less needed as the reviewed product becomes obsolete; a news story will become non-news after a few days; an analytical report will become less relevant as the situation expires, etc. Similar to existing literature modeling capacity depreciation (e.g. Besanko and Doraszelski 2004), we assume that the producer's content at a website depreciates with a certain ratio over time. Combining the effects of production and depreciation, the content quantity at a producer's webpage evolves as:

$$C_{j,t} = \delta C_{j,t-1} + a^{p}_{j,t} \tag{4}$$

In equation (4), $\delta \in (0,1)$ is the depreciation rate of the content. The smaller the value of δ is, the faster is the depreciation.

Producing content is a costly activity. We denote the cost of producing k units of content by producer j as $c^{prod}(k, X_j;\phi)$, with $c^{prod}(0, X_j;\phi) = 0$, i.e. the producer incurs no cost if she does not produce content. X_j is a vector of characteristics of producer j that may affect cost, and ϕ is a vector of parameters for the production cost function. The production cost is expected to be an increasing function of k, the units of content produced. The exact functional form of $c^{prod}(.)$ used in this study is specified in section 5 where we discuss the empirical application.

3.2.2 Link Formation

In each time period, a producer may also create a link pointing to another producer, assuming one to that producer does not already exist.[12] We denote this action by producer j at time t as $a^{l}_{j,t}$, where the superscript l indicate it is the linking decision. Specifically:[13]

$$a^{l}_{j,t} = \begin{cases} 0 & \text{do not create link} \\ j' & \text{create a link to producer } j',\ j' \in \{1..J\},\ j' \neq j \end{cases} \tag{5}$$

[12] Links are at producer level instead of content level, e.g. from producer *A* to *B* instead of a specific article of producer *A* to that of producer *B*.
[13] In our model, we consider the case where only creation but not removal of links is allowed. This is consistent with the dataset used in the empirical application. In real-world settings, certain websites allow link removal, while others do not. It is straightforward to extend our model to allow link removal. Also, we assume that a producer can create only one link in a period. This assumption is also made based on the dataset used in this study, and it is also straightforward to change it to allow a producer to create multiple links in a period.

Link formation may also be a costly activity. To form a link, a producer needs to spend time specifying so at the website. We denote the cost of creating a link by producer j as $c^{link}(j', X_j; \psi)$. The cost may vary according to the target of the link. For example, if reciprocity has intrinsic value, the producer will incur higher cost creating a non-reciprocal link, i.e. links to a producer j' when j' already links back at her, than creating a non-reciprocal one. Similar to production cost, ψ is the vector of parameters for the linking cost function. The exact functional form of $c^{link}(.)$ used in this study is specified in section 5.

3.2.3 Producer Network and Network Position

The links created by all producers together form a *producer network*, which is formally represented as a directed graph. Each node in the graph corresponds to a producer, and an edge exists if the producer corresponding to the source node has a link pointing to the producer corresponding to the destination node. The network evolves as producers create links over time. The network at time period t is denoted as G_t.

From the topology of the network, a numerical measure of each producer's network position, $P_{j,t}$, can be derived. As discussed earlier, the position of a producer in the network greatly influences the amount of viewership traffic directed to her content – the more incoming links a producer's has, and from the more prominent positions those incoming links come, the more traffic will be directed to the producer. Thus, both the number of incoming links and the positions of the sources matter. The PageRank measure (Brin and Page 1998), initially adopted by Google, elegantly captures both effects. Statistically, PageRank represents the probability of reaching each web page in a network when viewers follow a random walk along the links.

PageRank is equivalent to the eigenvector centrality of a damped adjacency-graph of the network. Interestingly, a rich literature in sociology has well established the importance of eigenvector centrality in social networks (e.g. Bonacich 1987, Faust and Wasserman 1992, Wasserman and Faust 1994, Bonacich and Lloyd 2001), where higher centrality it is associated with higher prestige. Recent marketing literature (Katona & Sarvary 2008) has also adopted PageRank in characterizing the network position of players. Following these, we use the PageRank of each producer in the network as the measure of her network position:

$$P_{j,t} = PageRank_{j,t} \tag{6}$$

The computation of PageRank is explained in the Appendix. The higher the PageRank, the more prominent a producer's position is in the network. This is the network position measure that enters into the demand function as specified in equation (1).

That incoming links increase a producer's position also means a producer's own position will *reduce* when she creates a link pointing to another producer – an outgoing link increases the target's position, and since position is relative, it would also reduce that of the source. This constitutes a *strategic* cost of link formation, which must be balanced with the benefit of borrowed content.

3.2.4 Borrowed Content

When a producer j has a link to another producer j', the content of producer j' can be easily accessed when a reader is viewing producer j's content. This augments the source's content, making the producer's webpage more appealing (Katona and Sarvary 2008). This effect is captured in our model using *borrowed content*, $C^{b}_{j,t}$, which is simply the sum of the content of all other producers being linked to at the time:

(7) $$C_{j,t}^{b} = \sum_{j'=1}^{J} C_{j',t} I\{j' \neq j, j \rightarrow j'\}$$

In the equation, $I\{.\}$ is the indicator function which equals 1 if the link exists and 0 otherwise.

Similarly, the borrowed quality $Q_{j,t}^{b}$ is the average of quality measures of the producers being linked to:

(8) $$Q_{j,t}^{b} = \sum_{j'=1}^{J} Q_{j'} I\{j' \neq j, j \rightarrow j'\} / \sum_{j'=1}^{J} I\{j' \neq j, j \rightarrow j'\}$$

3.3 Dynamic Competition

The competition among content producers over time is naturally modeled as a dynamic game. The key characteristic of the competition is that actions taken by producers not only determine the current payoff, but also affect future strategic interactions. Consequently, when a producer makes content production and link formation decisions, she needs to account for not only the current benefit, but also the future benefit according to the strategic response to her actions by other producers.

In each time period, the state of the competition is fully described by a set of commonly observed state variables. Producers take actions to maximize their respective discounted payoffs. Such actions are taken based on the current state of competition and in anticipation of the strategic response. The solution concept for producer’s optimizing behavior is that of Markov-perfect equilibrium, or MPE (Ericson and Pakes 1995). In an MPE, the strategy played by each producer is a Markov strategy, where actions are fully determined by the current state, and the strategy of each producer constitutes the best response to other producers’ strategies.

3.3.1 State

The state at time period t, denoted as s_t, is the collection of the individual content states of all producers and the state of the producer network: $s_t = (s_{1,t},...,s_{J,t},G_t)$, where $s_{j,t} = \{C_{j,t}, Q_j, X_j\}$ characterizes the quantity of producer j's content in period t and the characteristics of the producer related to quality and cost, and G_t contains the topology of the producer network. Note that $s_{j,t}$ does not include $P_{j,t}$, as the position of each producer in the network is fully determined by the topology of the network, which is encoded in G_t; nor does it include $C^b_{j,t}$ or $Q^b_{j,t}$, as the borrowed content is determined jointly by the topology of the network and the content of all producers. In another word, $P_{j,t}$, $C^b_{j,t}$ and $Q^b_{j,t}$ are *derived* from the state instead of the primitives of the state.

3.3.2 Action

In each time period, producer j's action $a_{j,t} = (a^p_{j,t}, a^l_{j,t})$ is its content production and link formation decision. Let a_t denote the vector of actions taken by all producers at time t, i.e. $a_t = (a_{1,t},...,a_{J,t})$.

Consistent with extant literature (e.g. Rust 1987, BBL 2007), we assume that before choosing her action at time t, each producer j receives an action-specific private shock $\nu_{j,t}(a_{j,t})$ that is independent among producers and over time. Since in our setting the actions are discrete, this private shock is a vector where each element corresponds to a specific action that can be taken at the time. Also consistent with extant literature, we assume the private shock follows an extreme value distribution. This private shock is needed in dynamic game models to account for the variability in actions that goes beyond the observed states. The collection of action-specific private shocks across all producers is denoted as $\nu_t = (\nu_{1,t},...,\nu_{J,t})$.

3.3.3 Payoff

In each time period, according to the viewership market demand and producer actions, producer j 's current-period payoff is:

$$(9) \qquad \pi_j(a_t, s_t, \nu_{j,t}) = mr\sum_{n=1}^{N} M\lambda_n \Pr_{n,j}(s_t) - c^{prod}(a^p_{j,t}, X_j;\phi) - c^{link}(a^l_{j,t}, X_j;\psi) + \nu_{j,t}(a_{j,t})$$

In equation (9), mr is the marginal benefit associated with each viewer visit, and $M\lambda_n$ is the number of viewers in segment n. In each period, the payoff of producer j is the benefit of viewership demand net of any cost associated with the action taken by the producer.

Each producer is concerned not just with the payoff of the current period, but also the overall payoff over time. The total discounted payoff to producer j at time t, which the producer seeks to maximize, is:

$$(10) \qquad E[\sum_{\tau=t}^{\infty} \beta^{\tau-t} \pi_j(a_\tau, s_\tau, \nu_{j,\tau}) \mid s_t]$$

In equation (10), $\beta \in [0,1)$ is the discount factor. The expectation is over the private shock, producers' actions in the current period, as well as future states, actions, and private shocks. As is shown clearly in the equation, the payoff to a producer depends on not only her own actions, but also the actions of other producers. This leads to strategic interactions which are characterized using an MPE.

3.3.4 Strategy and Equilibrium

We assume all producers follow Markov strategies. A Markov strategy profile σ of the dynamic game is the collection of the strategies of all producers: $\sigma = (\sigma_1, \sigma_2, ..., \sigma_J)$ where σ_j is the strategy played by producer j which depends on the state and the private shock, $\sigma_j : S \times \nu_j \mapsto A_j$, where S is the set of all states, ν_j is the set of private shocks and A_j is the set of all actions producer j can take.

Given a strategy profile, a producer's value function is the expected discounted payoff given the state, integrated over private shocks. It can be written recursively as follows:

(11) $$V_j(s;\sigma) = E_\nu[\pi_j(\sigma(s,\nu), s, \nu_j) + \beta \int V_j(s';\sigma) dP(s'|\sigma(s,\nu), s) \mid s]$$

When choosing a strategy, a producer needs to take into account not only the current state, but also other producers' strategies. Following convention in literature, we use σ_{-j} to denote the strategies played by all producers other than producer j. A producer's optimization problem is:

(12) $$V_j(s;\sigma_{-j}) = \max_{\sigma_j(s,\nu)} \{E_\nu[\pi_j((\sigma_j(s,\nu), \sigma_{-j}(s,\nu)), s, \nu_j) + \beta \int V_j(s';\sigma_{-j}) dP(s'|(\sigma_j(s,\nu), \sigma_{-j}(s,\nu)), s) \mid s]\}$$

The strategy which is the solution to equation (12) for producer j is the best response of the producer to others' strategies. An MPE is a strategy profile $\sigma^* = (\sigma_1^*, \sigma_2^*, ..., \sigma_J^*)$ where each producer's strategy is the best response to other producers' strategies. That is, in an MPE, when holding the strategies of other producers unchanged, no producer can increase its own expected payoff by unilaterally deviating to another strategy:

(13) $$V_j(s;\sigma_j^*, \sigma_{-j}^*) \geq V_j(s;\sigma_j, \sigma_{-j}^*), \forall s, \sigma_j$$

With observations of viewership demand and producer actions according to the states over time, we can estimate the parameters for the viewership demand model and the dynamic structural parameters, i.e. cost parameters, using the optimality condition implied by the equilibrium, which we discuss in detail in section 4.

3.4 Inter-temporal Tradeoffs

We now qualitatively discuss the tradeoffs content producers face in their production and linking decisions which are incorporated in the model. When deciding whether to produce content, producers obviously face a tradeoff between the cost incurred in producing content and the viewership such content attracts over time. Furthermore, there are several tradeoffs induced by linking, which lead to interesting interactions among producers. To begin with, when linking to another producer, a producer faces the tradeoff between borrowing the content of another producer and lower network position arising from promoting her competitor. Depending on how much the borrowed content will help and how severely the link will reduce her own network position, the producer may or may not find it worthwhile to form a link. Interestingly, when we take this tradeoff a step further, to consider not only whether to form a link but also whom to link to, we can see this tradeoff provides a simple explanation to a well known phenomenon in networks: the tendency to form *reciprocal* links. Reciprocity can be explained by social norm in sociology literature (Gouldner 1960), and through reward and punishment schemes in repeated games (Axelrod and Hamilton 1981). In the setting of our study, however, reciprocity may arise naturally from the consideration of network position. To see this, recall that the source's network position positively influences the target's. Suppose producer *A* wants to create a link, and producer *B* already has a link to producer *A* while producer *C* does not. Then if *A* links to *B*, thereby improving *B*'s position, the enhanced position of *B* will be partially carried over to *A*.

Whereas if *A* links to *C*, who is not *A*'s source, then *A* will not get this indirect benefit. Other things equal, this *promote-the-promoter* effect would favor reciprocal links over non-reciprocal ones.[14] That is, it is better to promote one's own promoter instead of another producer.

More tradeoffs come into play when we consider the interactions among producers over time. When making linking decisions, since a reciprocal links are naturally encouraged, a forward-looking producer may intentionally create a non-reciprocal link, if she expects that the producer she links to would reciprocate in the near future. That is, a producer may strategically create a link to "invite" reciprocation. The tradeoff she faces in this decision is between lower network position now and higher network position later on, if and when the target reciprocates. Furthermore, the prospect of linking may also encourage or discourage content production. A producer may be encouraged to produce more content than she otherwise would, if she expects that by producing more content, she can attract incoming links from other producers later on. The tradeoff she faces in this decision is between the cost of producing content now and better network position later on when she receives incoming links. At the same time, if a producer expects her competitors to receive incoming links, which diminishes her relative network position over time, she may produce less than she otherwise would. All these tradeoffs play a central role in determining content production and link formation decisions, and lead to the equilibrium strategy adopted by content producers.

4. Estimation

Our estimation requires that the content production and link formation decisions of all producers over a number of time periods are observed, so is the per-period viewership of each producer's

[14] A Monte-Carlo simulation using random graphs will easily show that, on average, the reduction in network position through creating a reciprocal link is less than that through creating a non-reciprocal link.

content in multiple time periods. The parameters to be estimated are the segment-specific viewership demand coefficients and the sizes of the segments, the quality coefficients, the content depreciation rate, the marginal benefit to the producer per reader visit, and the cost parameters of content production and link formation, as summarized below:

$$Param = (\{\beta_n, \lambda_n\}_{n=1..N}, \gamma, \delta, mr, \phi, \psi)$$

The marginal benefit and the cost parameters are not jointly identified. Considering this, we normalize $mr = 0.001$ for identification, which implies that the unit of account for cost is the marginal benefit per thousand views.[15] The first half of the parameters, $(\{\beta_n, \lambda_n\}_{n=1..N}, \gamma, \delta)$, are the parameters governing the viewership market in each period. The identification rests on the cross-sectional and inter-temporal variation of the content and network of producers, together with the corresponding variation of viewership. The second half of the parameters, (ϕ, ψ), are the dynamic structural parameters that together with the viewership demand parameters govern the dynamic competition, the identification and estimation of which rest on the optimality condition of the equilibrium.

Estimating dynamic games is challenging due to "curse of dimensionality" – the state space has high dimensionality as it incorporates the states of all players. Early estimation methods (e.g. Pakes and Mcguire 1994) rely on explicitly solving for equilibrium through value-function iteration, and have limited scalability. Recently developed two-step estimators call for estimating as many structural parameters offline as possible, and bypassing the computation of equilibrium when estimating the dynamic structural parameters. Our estimation is implemented using one such two-step estimator as specified in BBL (2007). BBL approaches the estimation task in two stages. In the first stage, we recover the equilibrium strategy of producers in reduced form, based

[15] This follows the industry standard on display advertising, where fees are quoted as cost-per-mille, or CPM, which represents the amount an advertiser needs to pay for every thousand times an advertisement is displayed to viewers.

on observed states and actions. Estimation of the equilibrium strategy, also termed the *policy function*, should strike the right balance between flexibility and data availability. A flexible functional form is desired for accurate representation of the equilibrium strategy, but it also requires more data. The second task for the first stage is to estimate the transition of states over time according to producer actions. The viewership demand will also be estimated in the first stage. In the second stage, using the knowledge of policy function, state transition, and viewership demand estimation in the first stage, we perform forward-simulation of the observed policy versus perturbed policies. As the observed policy constitutes an equilibrium, the optimality condition dictates that a producer's payoff when she plays the equilibrium strategy is no less than her payoff under an alternative perturbed strategy, while other producers still follow the equilibrium strategy. This optimality constraint forms the basis for constructing the objective function of a GMM estimator.

As is common in research on empirical dynamic games, we focus on symmetric pure strategy equilibrium. Such restriction allows us to pool data across all producers, which reduces data requirement and improves estimation efficiency.

4.1 First Stage

In the first-stage of the estimation, we recover the policy function, the state transition process, and the viewership market demand system.

4.1.1 Policy Function

In equilibrium, each producer chooses her action based on her own state as well as the states of other producers and the producer network. In the first stage of estimation, we recover this policy function, σ^*, which maps states to actions, in reduced form. BBL recommends using flexible functional forms to approximate the equilibrium policy with precision, which needs to be balanced with data availability.

Facing this tradeoff, we first transform the state space by deriving the vectors of network positions and borrowed content of all producers from the content state of each individual producer and the network structure – these are the variables that enter the utility functions directly. We then partition the transformed state space of an individual producer into quintiles along both the content dimension and the network position dimension. For each cell in this partitioned state space, we run a separate set of regressions with producer actions as dependent variables. The independent variables include the quality and cost related characteristics of the producer, the borrowed content and quality of the producer, the number of other producers in each cell of the partitioned state space, and the average quality of other producers.[16]

Since linking actions differ by destination, we distinguish the target on the following four dimensions: reciprocity, content, network position, and quality. We separate a reciprocal link from a non-reciprocal one, and for each of the other dimensions, we perform a median-split on the target: separate a high content producer (whose content quantity is above median) from a low content one (below median); separate a high network position producer from a low network position one; separate a high quality producer from a low quality one.[17] There are thus sixteen different types of linking targets, which combined with an action of no-link results in seventeen possible linking actions. We estimate each regression function using generalized linear models, with log link function for content production and logit link function for link formation.

[16] Essentially we are estimating the policy function nonparametrically on a producer's own state but parametrically on other producers' states. Ideally, the policy function should be estimated nonparametrically over the entire state space, but the high dimensionality of the state space makes this impractical, as to do so requires enormous amount of data. BBL suggests using local linear regression, which is similar to what we do here.

[17] Since quality attributes are constant over time in our model, the multi-dimensional quality measure of a producer can be reduced to a single dimensional number once the viewership demand is estimated, by weighting based on the estimated coefficients.

The set of regression functions through this estimation fully describes the strategy played by each producer in equilibrium. These policy functions form the basis for forward-simulation that is used in the second stage of the estimation to recover dynamic structural parameters.

4.1.2 State Transition

State transition probabilities are needed for performing forward-simulations in the second stage of estimation. In our model, the transition of states given the actions of all producers is deterministic – linking actions deterministically change the network structure, while production actions together with depreciation deterministically change content state. Consequently, state transition does not need to be estimated once the policy function is recovered. In the second stage forward simulation, we simply simulate producer actions based on the estimated policy function, and state transition can be calculated deterministically once the actions are simulated.[18]

4.1.3 Viewership Demand

The viewership market demand in each period can be estimated rather straightforwardly with MLE. Denote $s^m_{j,t}(s_t;\{\lambda_n,\beta_n\}^N_{n=1},\gamma)=\sum_{n=1}^{N}\lambda_n \Pr_{n,j}(s_t;\beta_n,\gamma)$ as the theoretical market share of producer j at time t given the state and the parameters, and the actual market share observed from data as $\hat{s}^m_{j,t}$.[19] Assuming that the difference $\eta_{j,t}=\log\hat{s}^m_{j,t}-\log s^m_{j,t}$ follows an i.i.d. normal distribution (Holmes 2009), the parameters $\{\{\lambda_n,\beta_n\}^N_{n=1},\gamma\}$ can be estimated using maximum likelihood.[20] The market size, i.e. the total number of viewers, M, is needed for calculating

[18] Content production is similar to investment in empirical IO, where studies also use probabilistic state transition models (e.g. Besanko and Doraszelski 2004). The difference is minor, as a tradeoff between the precision of states and the precision of state transition. Our model allows for deterministic state transition because the exact content state and the state of the producer network are used.

[19] The superscript *m* represents "market". This is to avoid confusion with the same symbol *s* that represents producer state.

[20] For the case of one viewer segment only, this is the same as the inversion suggested in Berry (1994).

market share, and is assumed to be observed.[21] The content depreciation parameter, δ, could be estimated either jointly with the other parameters of the viewership market demand equation, or separately in an offline manner.

4.2 Second Stage

We now discuss the second stage estimation of the dynamic structural parameters, i.e. cost of producing content and forming links. The key to the second stage estimation is the optimality condition of an equilibrium: given the equilibrium strategy profile $\sigma^* = (\sigma_1^*, \sigma_2^*, ..., \sigma_J^*)$, for any alternative strategy $\sigma_j^{'}$ for an arbitrary producer j and a randomly chosen state s, the equilibrium condition dictates that:

(14) $$V_j(s, \sigma_j^*, \sigma_{-j}^*; \phi, \psi) \geq V_j(s, \sigma_j^{'}, \sigma_{-j}^*; \phi, \psi)$$

Given a specific σ^*, a tuple $x = \{j, s, \sigma_j^{'}\}$ indexes one such equilibrium condition. Following BBL's notation, define

(15) $$g(x; \phi, \psi) = V_j(s, \sigma_j^*, \sigma_{-j}^*; \phi, \psi) - V_j(s, \sigma_j^{'}, \sigma_{-j}^*; \phi, \psi)$$

And define objective function

(16) $$Q(\phi, \psi) = \int (\min\{g(x; \phi, \psi), 0\})^2 dH(x)$$

where H is a distribution over the set X of the equilibrium conditions. Then the true parameter (ϕ_0, ψ_0) satisfies:

(17) $$Q(\phi_0, \psi_0) = 0 = \min_{\phi \in \Phi, \psi \in \Psi} Q(\phi, \psi)$$

The estimation is the empirical counterpart of this condition: let $\{X_k\}_{k=1}^{n_I}$ be a set of n_I randomly chosen optimality conditions. For each $X_k = \{j_k, s_k, \sigma_{j_k}^{'}\}$, we calculate the payoff of

[21] Changing the market size will change only the constant term of the estimated demand parameters.

the focal producer j_k when she follows the equilibrium strategy, $\hat{V}_{j_k}(s_k, \sigma^*_{j_k}, \sigma^*_{-j_k}; \phi, \psi)$, and that when she follows the alternative strategy, $\hat{V}_{j_k}(s_k, \sigma^{'}_{j_k}, \sigma^*_{-j_k}; \phi, \psi)$, for a proposed parameter value (ϕ, ψ). The empirical counterpart of the objective function is then

(18)
$$
\begin{aligned}
Q_n(\phi,\psi) &= \frac{1}{n_I}\sum_{k=1}^{n_I}(\min\{\hat{g}(X_k;\phi,\psi),0\})^2 \\
&= \frac{1}{n_I}\sum_{k=1}^{n_I}(\min\{(\hat{V}_{j_k}(s_k,\sigma^*_{j_k},\sigma^*_{-j_k};\phi,\psi) - \hat{V}_{j_k}(s_k,\sigma^{'}_{j_k},\sigma^*_{-j_k};\theta),\phi,\psi\})^2
\end{aligned}
$$

BBL shows that $Q_n(.)$ can be calculated through forward simulation, and the parameter that minimizes the objective function

(19)
$$(\hat{\phi},\hat{\psi}) = \underset{\phi\in\Phi,\psi\in\Psi}{\arg\min}\, Q_n(\phi,\psi)$$

is a consistent estimate of the true parameter under mild regularity conditions. This recovers the mean estimate of the parameter, while the standard error can be calculated using re-sampling of these equilibrium conditions.

5. Empirical Application

5.1 Data

Our data is obtained from a popular online product review website, which in recent years consistently attracts several million visitors on a monthly basis. A product reviewer can start writing product reviews once she creates an account at the website. The products reviewed at the website range from automobiles to toys, books, and movies, etc. Such *reviews* correspond to the *content* in our model. In addition to writing product reviews, a reviewer can also link to other reviewers by putting them into her list of trusted reviewers. Creating a link is solely at the discretion of the source reviewer, without the need for consent from the target reviewer. Such links together form a so-called *web of trust* among reviewers, and this corresponds to the *producer network* in our model. Viewers can easily navigate through the trust links to go from one reviewer's reviews to the reviews of another reviewer whom she trusts. Furthermore, product reviews written by reviewers who are trusted by many other reviewers, and trusted by reviewers who are themselves trusted by other reviewers, will receive preferential placement when viewers search the website. The position of a reviewer in this web of trust thus heavily influences the likelihood of her reviews being accessed by viewers.[22]

Although there are thousands of reviewers writing reviews at the website, in this study we focus on a small group of the most active ones, known at the website as the *top reviewers*. These top reviewers write product reviews frequently and consistently over time, and they are paid by the website based on the viewership their reviews attract. This group of elite reviewers is suitable for the model we developed earlier, as they are likely dedicated producers who are driven by profit incentive and who choose their actions strategically.[23] This small group of reviewers also is responsible for a significant share of the website viewership traffic.[24] Focusing on this group also eases the estimation of the model, as the number of players is kept at a reasonable level, and a long history of content production and link formation decisions is available for these active producers.

Our data set contains the decisions of writing reviews and creating links at the daily level, from June 2008 to March 2010. It also contains the viewership information starting from November 2009: for each four-day period starting from November 2009, the number of times

[22] In an interview, the former CTO of the company said of the trust system "… based on anecdotal evidence, those who have started using it end up completely depending on it to navigate the site."

[23] That is, as compared to other "occasional" users who write reviews infrequently, and who may be driven by other incentives such as a spontaneous desire to express one's opinion, for which a strategic framework may not be applicable.

[24] Comparing the viewership statistics of this group of reviewers with the website level statistics suggests they are responsible for about 30% of the overall viewership.

each reviewer's reviews is visited is recorded.[25] There are a total of 199 top reviewers at the site. Among them, 6 left the site during the period, and we exclude them from the data set.

[Insert Table 2 About Here]

The summary statistics are reported in Table 2. As is shown in the table, these top reviewers are highly active in writing reviews, averaging one review article per reviewer about every three days. In addition to writing reviews, they also created over two thousand links over the period, although the frequency of creating links is lower than writing reviews, with each reviewer adding a link roughly every two months. These top reviewers together attract a large audience, totaling more than six million view counts over a period of about four months. Comparing with website level traffic information, we know that this small group is responsible for about 30% of the total visit at the website, a significant share.

Based on the information available at the website, we use three variables for the *quality* factors in our model: *diversity*, *popular*, and *advisor*. These three variables and their summary statistics are described in Table 3. Together, these factors cover three important aspects which can affect viewership demand in addition to content volume and network position: diversity, popularity, and quality.

[Insert Table 3 About Here]

5.2 Result – Viewership Demand

We first estimate the model of viewership demand as determined by each reviewer's content, network position, borrowed content, and quality factors. As discussed earlier, the content depreciation parameter, δ, can be estimated together with the other viewership demand

[25] The Website displays the cumulative view count at the reviewer level and the information is updated daily. However, the update is not well synchronized for all reviewers. Thus we aggregate the information into 4-day periods to eliminate the noise created by this technical issue.

parameters. In our dataset, however, the content production and link formation data covers a much longer period than the data for viewership market demand. Furthermore, the overall viewership at the website has remained fairly stable over the period for which we observe the actions. Therefore, we estimate this depreciation parameter "offline", by treating it as a discount factor and finding the value that best keeps the content quantity stable over time. We arrive at the estimate $\delta = 0.9893$ this way. The summary statistics of network positions, and those of the effective discounted content and borrowed content, both calculated according to the depreciation parameter, are reported in table 4. The market size M is set to be twice the average total visit counts at the website to allow for substitution effect among competitors' websites. Website level statistics show that there were on average 5.2 million views per month, which results in $M = 1386667$. Changing this market size will change the constant term of the utility function without affecting other parameters.[26]

[Insert Table 4 About Here]

As discussed in section 3, multiple functional forms of the function $f(.;\beta_i)$ in equation (1) need to be estimated, with the best model chosen with certain model selection criterion. This flexibility is important because our treatment of the viewership market is reduced form, so estimating multiple functional forms can give us more robust results. We estimate the following four specifications:

$$(20) \quad \begin{cases} I(Linear) & f(C_{j,t}, P_{j,t}, C^b_{j,t}; \beta_i) = \beta_{i,1} C_{j,t} + \beta_{i,2} P_{j,t} + \beta_{i,3} C^b_{j,t} \\ II(Linear - Quadratic) & f(C_{j,t}, P_{j,t}, C^b_{j,t}; \beta_i) = \beta_{i,1} C_{j,t} + \beta_{i,2} C_{j,t}{}^2 + \beta_{i,3} P_{j,t} + \beta_{i,4} P_{j,t}{}^2 + \beta_{i,5} C^b_{j,t} + \beta_{i,6} C^b_{j,t}{}^2 \\ III(Log) & f(C_{j,t}, P_{j,t}, C^b_{j,t}; \beta_i) = \beta_{i,1} \log(C_{j,t}) + \beta_{i,2} \log(P_{j,t}) + \beta_{i,3} \log(C^b_{j,t}) \\ IV(Log - Embedded) & f(C_{j,t}, P_{j,t}, C^b_{j,t}; \beta_i) = \beta_{i,1} \log(C_{j,t} + \beta_{i,3} C^b_{j,t}) + \beta_{i,2} \log(P_{j,t}) \end{cases}$$

[26] The website was established in 1999 and is at mature stage now. The website level viewership remained fairly stable over the observation period, thus we do not consider the growth of market size in this study.

Specification I is the simplest functional form that accounts for all three factors, and we expect each coefficient to be positive to reflect their positive impact on viewership demand. Specification II extends the first specification by including a quadratic term for each factor to account for potential diminishing rate of return. For example, although linking to other producers provides a convenience benefit to viewers, when there are too many such links, viewers could also feel annoyed, so the content borrowing effect could become saturated. Similarly, although having higher network position gives a producer's content favorable placement, this benefit may become saturated beyond a certain threshold, if the network position is high enough to distinguish the producer in most cases. The quadratic terms are used to capture such effects. Specification III explicitly accounts for such diminishing return by using log transformation. Finally, Specification IV also uses log transformation, but adds a weighted component of the borrowed content to the original content before applying the log.

The quality factors are included in our study mainly for control purposes, and we adopt a simple linear functional form for the quality as well as borrowed quality:

$$\begin{aligned}(21)\qquad g(Q_j, Q_j^b;\gamma) = \gamma_0 + \gamma_1 Diversity_j + \gamma_2 Popular_j + \gamma_3 Advisor_j +\\ \gamma_4 Diversity_j^b + \gamma_5 Popular_j^b + \gamma_6 Advisor_j^b\end{aligned}$$

The result of estimation is presented in Table 5 (covariates are standardized). In all four specifications I-IV, the coefficients for content, network position, and borrowed content are all positive and statistically significant. This is clear evidence that all three are important factors in determining viewership demand, where higher content volume, more prominent network position, and more borrowed content all lead to higher viewer utility and in turn higher viewership demand for the reviewer's reviews. The coefficients for the three quality factors are also all positive and statistically significant, suggesting they positively influence viewership demand. Among them, the popularity indicator has the highest impact on viewer utility.

[Insert Table 5 About Here]

Looking at specification II, we find that content and borrowed content have similar contributions to the viewer utility, while network position has higher impact than both content and borrowed content. Specification II shows the quadratic terms of borrowed content and network position both have negative signs, suggesting that diminishing return exists for both factors. The quadratic term of content is also negative. However, its magnitude is very small and it is not statistically significant. Thus there is no clear evidence of diminishing return on the content dimension. Specifications III and IV both use log transformation, where the coefficient magnitude corresponds to percentage change. Specification V is the latent class version of specification II with two segments. In both segments, both network position and borrowed content positively influence viewer utility and exhibit diminishing returns. The first segment has content coefficient larger than that in specification II. Interestingly, the second segment has a negative content coefficient, and the coefficients for network position and borrowed content are quite large. This seems to suggest that this portion of the demand is mainly driven by the position in the network and the borrowed content, but not by the producer’s own content.

Among the five competing model specifications, specification II, the Linear-Quadratic specification, has the best model fit after adjusting for number of parameters using BIC. We therefore adopt this specification as the per-period viewership demand equation for the estimation of dynamic structural parameters.

5.3 Result – Dynamic Competition

5.3.1 Policy Function

The policy function regression, which captures reviewers' writing and linking decisions, is only the intermediate step for estimating the dynamic model parameters, and the coefficients are not interpretable. Instead, we report a few patterns of producer actions based on their content and network position states.[27] Note that the policies, estimated in a reduced-form fashion, constitute the equilibrium play resulting from reviewers' dynamic competition, and encapsulate the concept of best response. In this section we simply present the observed patterns. In the subsequent section 5.4, we investigate in detail how incentives and strategic interactions lead to such actions.

Figure 1 shows the average daily content production, conditional on the reviewer's own state along the content and network position dimension. As shown in the figure, reviewers with higher content volume in general write more frequently, and it is more so for reviewers with low network positions. In fact, reviewers with high content volume but low network position write reviews most frequently. This could be unexpected at the first look – the viewership demand equation, which captures the payoff through immediate viewership, shows that reviewers with higher network positions have higher marginal benefit and thus should have higher propensity to produce content. In section 5.4, we show how this discrepancy is explained with the dynamic tradeoffs faced by reviewers.

[Insert Figures 1, 2, and 3 About Here]

Figure 2 shows the frequency of creating an outgoing link, conditional on reviewers' own state. Reviewers with higher content volume create links more frequently. Reviewers with very high network positions (5-th quintile) also create links with have higher frequencies, although not by much.

[27] Actions can be summarized according to other dimensions, too, such as quality. In this study, we focus on the two dimensions, content and network position, as they are the direct results of reviewers' review writing and link formation actions. As specified in section 4, content and network positions are each partitioned into quintiles for the policy function regression, so we report the action patterns based on the quintile partitions along these two dimensions.

Since our analysis in section 3.4 indicates that reciprocal links would be favored by reviewers, we also report the relative probability of creating a non-reciprocal link over that of a reciprocal one conditional on a reviewer's own state, as shown in Figure 3. The first to note from the figure is that reviewers of all states are much more likely to create reciprocal links than non-reciprocal ones – the ratios are all much smaller than 1. Furthermore, reviewers with higher content volume have higher relative probability to create non-reciprocal links.

Other patterns are that reviewers with more content are more likely to receive incoming links, and that reviewers with different network positions have similar likelihood of receiving incoming links as long as they have similar content, with higher network positions increasing the likelihood but only slightly. Together, these patterns summarize the decisions made by reviewers as they interact with one another in the competition, each trying to maximize her own benefit. The incentives behind these actions are analyzed in detail in section 5.4.

5.3.2 Cost Estimation

We now discuss the estimate of the dynamic structural parameters, i.e. cost parameters. To operationalize the estimation, we randomly pick 500 states from the dataset. For each state, we randomly pick one reviewer and performed two forward-simulations. In the first, all reviewers follow the equilibrium strategy according to the estimated equilibrium policy, while in the second simulation, the chosen reviewer follows a perturbed strategy. Each simulation is run for 600 periods, and repeated multiple times with the average taken. We set the discount factor to 0.9995 as our observation is at daily level, which is similar to the 0.995 often set in dynamic structural studies when data is at weekly level (e.g. Erdem & Keane 1996). We then run the minimum distance estimator to find the cost parameters which minimize the deviation from the optimality condition of equilibrium, as specified in equation (19). The standard errors of the estimates were obtained through re-sampling of the chosen state-player pairs.

For the production cost function, we adopt a linear functional form.[28] Production cost may depend on the reviewer's quality, as a reviewer needs to exert more effort to achieve higher quality. The reviewer's tenure might also influence cost, due to learning-by-doing. Considering this, we assume the unit cost of production is a linear function of the reviewer's effective quality and tenure with the website, as shown in equation (22). We also assume the cost of linking is a linear function of the reviewer's effective quality and tenure, plus an indicator of whether the link is reciprocal, as shown in equation (23). This final term is added to tease out possible intrinsic value of forming reciprocal links – an intrinsic preference for reciprocal links, aside from the consideration of how it affects viewership, would imply lower cost of forming reciprocal links that non-reciprocal links at the model primitive level, and be reflected from a negative coefficient for this final term.

(22) $$c^{prod}(a^{p}_{j,t}, X_j; \phi) = a^{p}_{j,t} \cdot (\phi_0 + \phi_1 g(Q_j, \gamma) + \phi_2 Tenure)$$ [29]

(23) $$c^{link}(a^{l}_{j,t}, X_j; \psi) = \psi_0 + \psi_1 g(Q_j, \gamma) + \psi_2 Tenure + \psi_3 I\{a^{l}_{j,t} \text{ is reciprocal}\}$$

[Insert Table 6 About Here]

The result of the estimation is reported in Table 6. The constant term for the production cost regression is 0.148 and statistically significant at .95 level. This means the cost of writing a review article is equivalent to the benefit of 148 page views, which is a reasonable number for

[28] A strictly convex cost function is often used in industrial organization literature. In our empirical application, however, it is reasonable to assume there is a unit cost for writing a review article, hence the linear form. Equilibrium condition holds as long as the market size is finite. We also estimated the quadratic specification of cost function, and the result when averaged for unit cost is similar to the linear specification. The result for quadratic cost function is available from the author upon request.

[29] In a slight abuse of notation, we use the same function symbol, *g*, to represent a reviewer's own quality effect: $g(Q_j; \gamma) = \gamma_0 + \gamma_1 Diversity_j + \gamma_2 Popular_j + \gamma_3 Advisor_j$, excluding the borrowed quality effect – cost should be determined by a reviewer's own characteristics.

unit cost estimate, as the summary statistics show that on average a review article is viewed a little over 200 times. The coefficient for reviewer quality is positive and statistically significant. This suggests that reviewers of higher quality put more effort in writing product review articles and thus incur higher cost per article written, consistent with expectation.[30] The estimate also shows that a reviewer's tenure at the website does not have a significant impact on her production cost.

The cost of linking is very close to zero, indicating that linking itself is not a high effort activity. Neither quality nor tenure is shown to have a significant effect on the cost of linking. More notable is the coefficient for the reciprocity term. The positive sign of the coefficient shows that the cost of forming a non-reciprocal link is less than that of forming a reciprocal link, although the result is not statistically significant. As discussed earlier, the existence of intrinsic value for reciprocal links would be reflected from a negative coefficient for this term, thus there is no evidence of such intrinsic value. That reciprocal links are more likely to be formed, as observed in the dataset, thus should be mainly attributed to the strategic considerations, i.e. the promote-the-promoter effect as discussed in section 3.4.

5.4 Decision Dynamics and Interdependence

Using the estimated viewership demand equations, the dynamic cost parameters, and the equilibrium policy, we now investigate the competitive dynamics in detail. To address the research questions raised for this study, we analyze three aspects of the competitive dynamics: First, we investigate the incentive to form links and how it depends on link types and reviewer

[30] A more general model is to assume that all reviewers are of the same type, and that when they write articles they can choose to write either a high or a low quality one, with the former entailing higher cost than the latter, similar for links. However, to estimate such a model requires quality information at the level of each review article, which we do not have. This is beyond the scope of our study and is left for future work. Our model can be considered as a restricted model in this broader context – each reviewer is restricted to choose a quality type and then follow it throughout the whole period.

states. Next, we analyze how linking influences content production decisions. Finally, we evaluate the net benefit accrued to reviewers at different states and the market structure that emerges from the competition.

5.4.1 Dynamics of Link Formation

The decision of whether to create a link and whom to link to is driven by both the tradeoff between borrowed content and network positions, and the dynamic interactions between reviewers. To understand the incentives to form links for reviewers at different states, we evaluate how such links impact reviewers' viewership demand.

To analyze the implication of forming links, we first quantify, using the dataset, the average change in network position through establishing an outgoing link and that through receiving an incoming link given a reviewer's state. Receiving an incoming link normally increases the reviewer's network position noticeably. Creating an outgoing link, however, reduces the network position, and a reciprocal link typically leads to smaller reduction than a non-reciprocal link as discussed earlier. We then use a subset of the data, covering the three-month period from January 2009 to March 2009, to calculate the incremental benefit of creating a link for each reviewer in each day. For creating a reciprocal link, the incremental benefit is calculated as the difference in discounted viewership between two otherwise identical scenarios except that in the second scenario the focal reviewer creates a reciprocal link to another reviewer who already links to her. Other factors are held constant in this calculation. This calculation captures the effect of creating a reciprocal link, which can be considered as a "close-loop" action.[31] For creating a non-reciprocal link, however, this calculation captures only the direct

[31] We can consider that a reciprocal link *finishes* a round of dynamic interaction – the target reviewer already has a link pointing back and will not further "respond" to the reciprocal link. Thus a "loop" is closed. In contrast, a non-reciprocal link *starts* a round of dynamic strategic interaction – the target reviewer will in subsequent periods decide whether to reciprocate. Thus a loop is opened.

effect, i.e. the tradeoff between more borrowed content and lower network position, but not the strategic aspect arising from dynamic interactions, i.e. the target reviewer may decide to reciprocate in future. To account for this dynamic interaction, we calculate the probability of a non-reciprocal link being reciprocated in future and the average days taken to receive the reciprocation, conditional on the source reviewer's state, using the equilibrium policy recovered from data. We then calculate the change in discounted viewership assuming that a reciprocal link is established with the corresponding probability and delay.

[Insert Figures 4 and 5 About Here]

The incremental benefit of creating a reciprocal link is reported in Figure 4. The result is summarized along the content dimension in quintiles. The top figure shows positive average effects for all five quintiles, suggesting that in general the benefit of more borrowed content outweighs the cost of reduced network position through forming a reciprocal link. Also, the figure shows that reviewers with more content benefit more from a reciprocal link. This is consistent with the policy function where reviewers with more content are more likely to create reciprocal links, as shown in the bottom figure of Figure 4.

The result for creating a non-reciprocal link is reported in Figure 5, also summarized along the content dimension. Creating a non-reciprocal link typically reduces network position more than does a reciprocal one. As shown in the first series of the top figure, which includes the direct effect but does not account for future reciprocation, the average incremental benefit is negative for all five reviewer quintiles, suggesting that the cost of reduced network position outweighs the benefit of more borrowed content. The incremental benefit is also significantly lower than that of forming reciprocal links. Recall that the cost estimate in section 5.3 shows no evidence of intrinsic value for reciprocal links, we know that in the context of this study, the tendency towards reciprocity is mainly explained by the comparatively favorable impacts of reciprocal links on viewership, due to the promote-the-promoter effect. This is a notable result. Sociology literature has long recognized the prominence of reciprocity in social networks, and statistical network models often consider that as model primitives. Our study provides an alternative explanation in a rational economic rather than social context, that reciprocity can be naturally favored by dynamic strategic considerations, without the need for a social explanation as model primitive.[32]

However, the first series in the top figure also shows that the more content a reviewer has, the lower her incremental benefit from forming a non-reciprocal link, yet the policy function shows that reviewers with more content are more likely to create non-reciprocal links (the bottom figure). Thus a static perspective alone does not explain the linking actions well. This discrepancy is resolved once the dynamic strategic perspective is taken into account. As the second series in the top figure shows, after accounting for future reciprocation, the incremental benefit of forming a non-reciprocal link increases significantly for all five quintiles, and reviewers with more content have higher incremental benefit. This is because a reviewer with more content is more confident to see the target reviewer reciprocate, and with shorter delay. After all, the target reviewer also can benefit from borrowed content, and when she decides to create a link, she would favor a reciprocal one to maintain her own network position, thus making the source reviewer a favorable target. This incentive to reciprocate is further enhanced when the source reviewer has more content. In essence, a reviewer is "inviting" reciprocation when creating a non-reciprocal link, in anticipation of the strategic response from the target

[32] That is, an explanation such as "people tend to form reciprocal links because by nature they like reciprocity, i.e. there is an intrinsic value to reciprocate".

reviewer, and the more content a reviewer has, the more effective this strategy is. Comparing the two scenarios clearly shows how the dynamic interactions drive reviewers' linking decisions.

5.4.2 The Impact of Linking on Content Production

We now analyze how linking influences reviewers' content production decisions. Similar to the analysis of link formation, we evaluate the incremental effect of writing reviews by reviewers at different states, accounting for the dynamic interaction effects arising from linking.

For this analysis, we use the same subset of the data as used in analyzing link formation. We begin with analyzing the direct effect of writing reviews: for each day and each reviewer, we calculate the difference in discounted viewership between two otherwise identical scenarios: in the first, the focal reviewer does not write reviews; in the second, she writes one review article, which depreciates at the estimated depreciation rate. This difference in viewership approximates the direct incremental benefit of producing one unit of content, from which the production cost is then subtracted to arrive at the net incremental benefit. The result is shown in Figure 6(A). The direct incremental benefit is much higher for reviewers with higher network positions, and is positive only for reviewers in the top two quintiles of the network position dimension. Reviewers with more content also get higher benefit, but the difference along the content dimension is not as large as along the network position dimension.

The direct benefit is only one part of the incentive behind content production. When deciding whether to produce content, a reviewer considers not only the immediate viewership, but also the future linking actions of other reviewers. For example, if a reviewer in a high content state anticipates other reviewers to link to her in the near future, which leads to higher network position, then her incentive to produce will be increased, as the additional benefit from higher network position later on adds to the direct benefit. Whereas if a reviewer in another state expects her competitors to receive more incoming links, which reduces her relative position in the network, then her incentive to produce will be lower than suggested by the direct benefit. Thus linking could significantly alter the incentive to write reviews depending on the states of reviewers.

[Insert Figure 6 About Here]

To analyze how linking influences the incentive to produce, we use the equilibrium policy to calculate the average change in network positions, arising from linking, corresponding to different reviewer states. The calculation shows that reviewers at high content and low network position states get highest average increase in network position over time, while reviewers in the opposite states see their network positions reduce later on. We then incorporate this state-dependent change in network position into the calculation of the incremental effect of writing one more review. The net incremental benefit calculated this way, reported in Figure 6(B), shows that once the prospect of linking is accounted for, content level, instead of network position, becomes the main determining factor of the incremental benefit. Reviewers in the top two quintiles of the content dimension have positive net benefit from producing content, while other reviewers have negative benefit, even for those with high network positions. In other words, a reviewer with high content but low network position writes reviews because she expects other reviewers to link to her later on, even though the immediate viewership is not much. Meanwhile, a reviewer at the opposite state finds it not worthwhile to write reviews as she foresees lower network position ahead. This is consistent with the observed policy function, which shows that content level influences the frequency of writing reviews more than network position does.

In summary, the results demonstrate a close interdependence between link formation and content production, and that linking is a major driver of reviewers' writing decisions.

Interestingly, the prospect of linking *encourages* the content production of reviewers with high content and low network positions, while *discourages* the content production of reviewers with low content and high network positions.

5.4.3 Net Benefit and Market Structure

Given the viewership demand and the cost estimates, we now analyze the net benefit accrued to reviewers at different states. To do so, we calculate the discounted net benefit over rolling six-month windows over the entire period. Net benefit is simply the viewership minus the cost of production and linking. Costs are derived from reviewers' actual decisions while viewership is inferred from reviewers' states and the estimated viewership demand equation.

[Insert Figure 7 About Here]

The result is presented in Figure 7, where we show the average net benefit on the content dimension and network position dimension. The figure shows that reviewers with higher network positions derive significantly higher benefit than reviewers with lower network positions. In contrast, however, reviewers with more content do not derive higher benefit than reviewers with less content. This may be unexpected at first look, as the demand equation shows that more content leads to higher viewership. But it is explained by the cost side: although reviewers with more content can attract higher viewership, they also incur higher cost as they write more reviews. The result shows that additional viewership demand is mostly offset by the increased cost, making a reviewer with more content no better than one with less content in terms of net benefit. This result is also reasonable when we consider the competitive effect: since content level is determined solely by a reviewer's own production decisions, were there to be significantly higher net benefit with higher content level, all reviewers would write more and in so doing, the potential advantage from higher content level would be largely competed away.

In contrast, the advantage coming from higher network positions cannot be competed away as easily. This is because, even though desirable, a reviewer cannot unilaterally increase her network position. Instead, it takes incoming links from *other* reviewers to increase that. Thus a reviewer may enjoy significant advantage from having a high network position, while other reviewers lack an effective way to counter that. Indeed, our calculation suggests that higher network positions offer significant competitive advantage and lead to higher net benefit. The competition thus results in a market where reviewers are differentiated along the network position dimension, while on the content dimension surplus is mostly competed away.

The contrast between content and network position should be taken note by companies operating such websites. As a website seeks to maximize its overall viewership, it should seek to encourage content production by creating a competitive environment internally. Any form of "sticky" competitive advantage enjoyed by a subset of producers, such as that led to by higher network positions in this context, may create imbalance in the system. This imbalance can potentially lead to differentiations that soften the competition and reduce the overall content level at the site. Consequently, the effect of linking to the overall website viewership is a matter of concern that is worth further investigation.

5.5 The Effect of Linking on Website Viewership – A Simulation

For marketing managers who operate those content websites, it is important to know whether the network among content producers increases the overall viewership at the website, and how the linking feature should be designed to generate optimal viewership outcome. If content is exogenously given, then we would expect a network superimposed among producers to increase overall viewership, as linking enhances content.[33] When content production is determined endogenously in a dynamic context, however, the overall effect of network is not at all clear. Qualitative analysis of tradeoffs also reveals forces towards both directions. On one hand, since a producer with more content is more likely to receive incoming links, linking provides an incentive for certain producers to produce more content. On the other hand, however, linking could also discourage other producers from producing content, as is shown in section 5.4. At the website level, the content enhancement effect of linking is expected to increase overall viewership. However, if there are producers with high content volume sitting at obscure positions in the network, while others occupy prominent positions yet do not have much content, then the network may hinder efficiency by not effectively directing viewer traffic to content.

Given these factors with opposite effects, the network could either increase or decrease overall viewership. A sign of concern, though, is that as shown in section 5.4, reviewers with more prominent network positions enjoy significant competitive advantage, yet competitive advantage enjoyed by a small set of players in general reduces competition intensity. This suggests that the current network may impede the competition among content producers, and that alternative policies regulating link formation may help the website improve overall viewership.

To evaluate the overall effect of network, ideally we want to compare two situations which are otherwise identical, except that in the first link creation is allowed, while in the second it is not. Similarly, we want to compare situations under alternative linking regulations, such as restricting the total number of links a developer can create, to find out which link regulation leads to best viewership outcome for the website. However, current methodological restrictions

[33] This is consistent with existing literature, which shows that network increases overall sales in an online shopping center environment (Stephen and Toubia 2009).

prohibit us from making these comparisons directly, as it requires explicitly solving for the equilibria of alternative dynamic games, which are computationally infeasible.[34]

Considering this, we resort to a "second best" approach by performing simulations which alter initial states but do not alter the existing equilibrium – since it is only the initial states that changes, while the structural parameters and the game remain the same, the existing equilibrium recovered from data still applies. To analyze through this approach whether the imbalance induced by the current network reduces viewership, we pick a state from the data, and for each reviewer, we randomly remove her outgoing links until she has no more than five outgoing links remaining.[35] After this system-wide link removal, the network becomes more sparse and balanced. We then perform two forward simulations for 60 periods, with the first starting from the original state and the second from this new state after the link removal. We compare content production, link formation, and overall viewership between these two simulations to evaluate the overall effect of the network.

[Insert Table 7 and Figure 8 About Here]

The result of the simulation is reported in Table 7. With the system wide link removal, content production, link creation, and overall viewership demand all increase significantly. The average daily viewership increases by 17.12% for the website overall, while content production and link creation both increase by more than 30%. This suggests that the current network among content producers, although providing benefit through enhancing content, also brings too much

---

[34] BBL recovers structural parameters without explicit computation of equilibrium, thus bypassing the "curse of dimensionality" issue. In policy simulation, however, any change in the "rule of the game" can potentially lead to a new equilibrium, so equilibrium must be explicitly computed. For example, if link formation is prohibited, all reviewers will adjust their strategy for writing reviews accordingly. To evaluate that change, we must compute the new equilibrium, the cost of which is prohibitive given the number of reviewers in our study. Recent methodological advancement, e.g. the concept of oblivious equilibrium developed in Weintraub, Benkard, and Van Roy (2008), can potentially solve this issue, with the drawback that the solution concept itself is an approximation. We leave the potential use of oblivious equilibrium for future work.

[35] We do not remove all links to avoid potential issues of boundary bias for the estimated policy functions.

market power to certain producers – those with high network positions – and impedes efficiency. When the field of competition is leveled, competition intensifies, with reviewers collectively producing more content, and the overall viewership at the website increases.

Figure 8 shows the simulation result in further detail along the time dimension. The viewership demand jumps immediately after the link removal, likely because reviewers with high content volumes but low network position now become more visible and attract more viewership. Over time, the demand increase moderates slightly but still holds stably above 15%. This is supported by sustained increase in content production – with a leveled playing field, the competition is intensified and reviewers collectively have higher incentive, or are forced, to produce more content. Link creation also jumps initially, but this is comparatively short-lived, as link creation falls back to the pre-removal rate after about thirty periods.

In summary, this simulation provides evidence that the current design over time leads to inefficient internal competition among reviewers. Alternative policies that regulate link formation could potentially lead to overall viewership and should be considered for experimentation at the website.

6. Discussion, Limitation and Conclusion

The advent of online social media brings about many intriguing phenomena. A prominent one is the emergence of a large number of revenue sharing content websites, which rely on external content producers to supply content and induce an internal competition for viewership among producers. The linking feature recently introduced to many websites further leads to complex and intriguing dynamic interactions among content producers. Meanwhile, the implication of linking on the overall viewership, crucial to the website platform builders, remains an open question. A detailed understanding of producers' interactions and their implications thus not only is of academic interest, but also has important managerial implications, as this phenomenon is quickly gaining momentum in the industry.

Motivated by this, we develop a dynamic oligopoly model to study the competition among content producers. In our model, producers compete against one another through producing content and forming links, and we characterize their strategic interactions using the solution concept of Markov-perfect equilibrium. We estimate the model using the data obtained from a popular product review website, leveraging the two-step estimation approach developed in Bajari, Benkard, Levin (2007), and provide a detailed analysis of the interactions among reviewers in their decision process.

Our study contributes to the literature by investigating the interactions of link formation and content production decisions, by analyzing the inter-temporal tradeoffs that drive the interactions dynamically, and by providing a rational economic framework for empirically studying the formation of networks in a dynamic strategic setting. Our study leads to several findings with managerial implications. We find that viewership demand is positively influenced by content volume and network positions, and there is a content borrowing effect through linking. We find that reciprocal links are more likely to be formed than non-reciprocal ones, and this is encouraged by the nature of the strategic interaction – a promote-the-promoter effect. This tendency towards reciprocity further induces producers with high content volume to strategically create non-reciprocal links, in anticipation of reciprocation later on which will enhance their network positions. We find that the prospect of linking encourages producers with high content volume but low network position to produce more content, yet discourages producers at opposite states. Furthermore, we find that the producers' net benefit increases with their network positions but not with their content volume, as the higher viewership from more content is offset by the higher cost incurred in producing the content. This suggests that linking may lead to inefficiency as competitive advantage is accrued to producers with high network positions. Finally, our simulation suggests that limiting the links at the website may lead to higher content production and overall viewership demand.

Managers who operate content websites can consider several alternative linking policy designs to improve efficiency. They could prohibit linking altogether by not offering the feature. This will prevent competitive advantage from being accrued to a subgroup of producers. Between completely disabling linking and not regulating linking at all, an alternative at the middle ground is to restrict the number of links each producer can form. This could alleviate the imbalance over time, while producers would also become more selective in forming links. Another alternative is to impose a time limit on links so that they expire after some time. This could make the network structure less rigid and ease the issue of unbalanced competition. Methodological restrictions limit our ability to analyze these alternative policies in detail, while industry managers could explore these and other policies through experimentation at the websites.

A few other limitations of our study can be addressed in future work. First, our study focuses on the profit motive of content producers, and we use a group of top producers for our analysis. Although most websites have a significant share of their viewership generated by a small group of elite producers, there is also a larger group of more casual content producers. This mass group of casual producers may have incentives other than profit, and a richer model is called for to study their behaviors and contributions to the business. Second, in the social media market, the line between consumers and producers is blurred. While our focus on the small group of elite producers allow us to still follow the traditional supply-side demand-side dichotomy, an exciting opportunity exists to advance the literature by investigating the dual roles the website users may play. Finally, not all content is the same, and different content may be either complements or substitutes. For feasibility reasons, our model considers all content to be of the same type, while we leave the interactions induced by different content types for future research. We also hope that, with the rapid advancement in econometrics on dynamic game estimation methodologies, we will be able to admit more heterogeneity among producers in future, and to explicitly evaluate the effects of alternative policies when they lead to different equilibrium situations.

Online content markets, and social media in general, bring much closer and more dynamic interactions among consumers, between consumers and producers, and among producers, than the traditional offline market does. With that, it also opens an exciting frontier for marketing research. Our work is an early step towards this direction, and we are confident that future research will bring further insights in this area and offer much needed managerial guidance.

References

Ackerberg, D., C.L. Benkard, S. Berry and A. Pakes, (2007), “Econometric Tools for Analyzing Market Outcomes”, Handbook of Econometrics, 6, 4171-4276

Aguirregabiria, V. and P. Mira, (2007), “Sequential Estimation of Dynamic Discrete Games”, Econometrica, 75, 1-53

Axelrod, R. and W. Hamilton, (1981), "The Evolution of Cooperation," Science, 211, 1390-96

Bajari, P., C.L. Benkard and J. Levin, (2007), “Estimating Dynamic Models of Imperfect Competition”, Econometrica, 75, 1331-1370

Bala, V. and S. Goyal, (2000), “A Non-cooperative Model of Network Formation,” Econometrica, 68, 1181–1230

Berry, S.T., (1994), “Estimating Discrete-Choice Models of Product Differentiation”, The RAND Journal of Economics, 25, 242-262

Berry, S., J. Levinsohn and A. Pakes, (1995), “Automobile Prices in Market Equilibrium”, Econometrica, 63, 841-890

Besanko D. and U. Doraszelski, (2004), “Capacity Dynamics and Endogenous Asymmetries in Firm Size”, The RAND Journal of Economics, 35, 23-49

Bonacich, P., (1987), “Power and Centrality: A Family of Measures”, The American Journal of Sociology”, 92, 1170-1182

Bonacich P. and P. Lloyd, (2001), “Eigenvector-like Measures of Centrality for Asymmetric Relations”, Social Networks, 123, 91-201

Brin, S. and L. Page, (1998), “The Anatomy of a Large-scale Hypertextual Web Search Engine”, Computer Networks and ISDN Systems, 30, 107-117

Chevalier, J.A. and D. Mayzlin, (2006), “The Effect of Word of Mouth on Sales: Online Book Reviews”, Journal of Marketing Research, Vol. XLIII, 345-354

Chintagunta, P., S. Gopinath and S. Venkataraman, (2010), “The Effects of Online User-Reviews on Movie Box-Office Performance: Accounting for Sequential Rollout and Aggregation Across Local Markets”, Marketing Science, forthcoming

Dube, J.P., G.J. Hitsch and P.E. Rossi, (2009), “Do Switching Costs Make Markets Less Competitive?” Journal of Marketing Research, Vol. XLVI, 435–445

Erdem, T. and M.P. Keane, (1996), “Decision Making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets”, Marketing Science, 15, 1-20

Ericson, R. and A. Pakes, (1995), “Markov-Perfect Industry Dynamics: A Framework for Empirical Work”, The Review of Economic Studies, 62, 53-82

Faust, K. and S. Wasserman, (1992), “Centrality and Prestige: A Review and Synthesis”, Journal of Quantitative Anthropology, 4, 23-78

Godes, D. and Mayzlin, D., (2004), “Using Online Conversations to Study Word-of-Mouth Communications”, Marketing Science, 23, 545-560

Gouldner, A.W., (1960), “The Norm of Reciprocity: A Preliminary Statement”, American Sociological Review, 25, 161-178

Holmes, T.J., (2009), “The Diffusion of Wal-Mart and Economies of Density”, Working Paper

Hotz, V.J. and R.A. Miller (1993), “Conditional Choice Probabilities and the Estimation of Dynamic Models”, The Review of Economic Studies, 60, 497-529

Jackson, M.O., (2004), “A Survey of Models of Network Formation: Stability and Efficiency,” in G. Demange and M. Wooders, eds., “Group Formation in Economics: Networks, Clubs, and Coalitions”, Cambridge University Press

Kamakura, W.A. and G.J. Russell, (1989), “A Probabilistic Choice Model for Market Segmentation and Elasticity Structure”, Journal of Marketing Research, 26, 379-390

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

Maskin, E. and J. Tirole, (1988), “A Theory of Dynamic Oligopoly: I & II”, Econometrica, 56, 549-600

Maskin, E. and J. Tirole, (2001), “Markov Perfect Equilibrium, I. Observable Actions”, Journal of Economic Theory, 100, 191-219

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

Pakes, A. and P. McGuire, (1994), “Computing Markov-Perfect Nash Equilibria: Numerical Implications of a Dynamic Differentiated Product Model”, The RAND Journal of Economics, 25, 555-589

Pakes, A. and P. McGuire, (2001), “Stochastic Algorithms, Symmetric Markov Perfect Equilibrium, and the 'Curse' of Dimensionality”, Econometrica, 69, 1261-1281

Pakes, A., M. Ostrovsky and S. Berry, (2007), “Simple Estimators for the Parameters of Discrete Dynamic Games (With Entry/Exit Examples)”, The RAND Journal of Economics, 38, 373-399

Rust, J., (1987), “Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher”, Econometrica, 55, 999-1033

Ryan, S.P., (2009), “The Costs of Environmental Regulation in a Concentrated Industry”, Working Paper

Srinivasan, K., (2006), “Invited Commentary: Empirical Analysis of Theory-Based Models in Marketing”, Marketing Science, 25, 635-637

Stephen, A. and O. Toubia, (2009), “Deriving Value from Social Commerce Networks”, The Journal of Marketing Research, forthcoming

Wasserman, S. and K. Faust, (1994), “Social Network Analysis: Methods and Applications”, Cambridge University Press

Weintraub, G.Y., C.L. Benkard and B. Van Roy, (2008), “Markov Perfect Industry Dynamics with Many Firms”, Econometrica, 76, 1375-1411

Yao, S. and C. Mela, (2010), “A Dynamic Model of Sponsored Search Advertising”, Working Paper

Appendix

A.1 PageRank

In this section we explain the detail of PageRank, the measure of network position used in our study. PageRank, first presented in Brin and Page (1998), is behind the initial Google search engine. Given a network of pages, it produces a numerical measure for each page to represent its relative importance in the network. The measure is well documented in literature, and is explained here for completeness:

Let $p_1, p_2, ... p_n$ be n nodes (web pages) which are connected by directional links. Let c_i be the out-degree of page p_i – the number of outgoing links from that page. Let d be a damping factor which value between 0 and 1. Denote A as the modified adjacency matrix for the graph of the nodes and the links, where:

(A-1) $$[A]_{ij} = \begin{cases} 1/c_i & i \to j \\ 0 & otherwise \end{cases}$$

The PageRank, denoted as PR, is a vector such that:

(A-2) $$PR = (1-d) \cdot PR + d \cdot PR \cdot A$$

The i-th element of PR is the PageRank of node p_i. A larger value indicates higher importance of the node in the network. The measure is rooted on a Markov random navigation model: assume there is a person visiting the pages; at any time, with probability d she chooses to follow an outgoing link, with link chosen randomly with equal probability when multiple outgoing links exist, and with probability $1-d$ she will jump to another page, with each page having the same probability to be the destination. The PageRank of a page is then the steady-state probability of that page being visited.

As stated in Brin and Page (1998), "*Another intuitive justification is that a page can have a high PageRank if there are many pages that point to it, or if there are some pages that point to it and have a high PageRank.*" This insight proves crucial for the success of PageRank in capturing the relative importance of web pages on the Internet, and is instrumental in our study. The PageRank is also similar to eigenvector centrality that is widely used in social network literature, where it is shown to reflect the power or prestige of a node in the network. (Let I be an $n \times n$ matrix where $[\mathrm{I}]_{ij} = 1/n$, then A-2 can be written as $PR = PR((1-d)\cdot \mathrm{I} + d \cdot A)$, i.e. the PageRank is an eigenvector of the adjacency matrix further modified by the damping factor.)

Table 1: Revenue Sharing Websites with Independent Content Producers[1]

Website	Content Type	Linking (Name)	Monthly Visitors (in millions)[3]
about.com	Advice	No	43.9
answers.yahoo.com	Questions & Answers	Yes (Fan)	43.7
associatedcontent.com[2]	General content	Yes (Favorite)	10.7
ehow.com	How-to tip	Yes (Subscription)	30.5
epinions.com	Product review	Yes (Trust)	4.1
hubpages.com	General content	Yes (Follow)	9.9
iReport.com	News report	Yes (Follow)	1.1
seekingalpha.com	Investment advice	Yes (Follow)	1.2
squidoo.com	General content	Yes (Fan)	6.6
youtube.com	Video	Yes (Subscription)	104.1

1. A list of more than 100 revenue sharing content sites can be found at http://socialmediatrader.com/resource-list-100-revenue-sharing-sites/
2. Acquired by Yahoo! in May 2010 for about $100 million.
3. Source: compete.com. June 2010

Table 2: Summary Statistics

	Mean	SD	Min	Max
Reviews Written Per Reviewer	168.14	161.07	5	704
Links Created Per Reviewer	10.31	11.19	0	67
Total View Count Per Reviewer	34375	50834	680	372188
Number of Reviewers	193	Total Reviews Written		33123
Number of Decision Days	646	Total Links Created		2039
Number of View Count Periods	28	Total View Counts		6840629
Total Non-reciprocal Links	1148	Total Reciprocal Links		891
Percent of Reciprocated Non-reciprocal Links	40.70%	Average Days Taken To Reciprocate		52.5

Table 3: Reviewer Quality Factors

Factor	Value	Description
Diversity	Integer	The number of product categories for which the reviewer writes reviews as top reviewers
Popular	Binary Indicator	The reviewer was recognized as the top 100 most popular authors before
Advisor	Binary Indicator	The reviewer is recognized as trusted source on content quality
Average Diversity		1.53
Number of "Popular" Reviewers		59
Number of "Advisor" Reviewers		112

Table 4: Summary Statistics - Content, Network Position, and Borrowed Content

	Mean	SD
Content	27.83	35.69
Network Position	5.18E-03	2.61E-03
Borrowed Content	1304.5	780.4

Table 5: Viewership Demand Estimation

Model Specification	I	II	III	IV	V	
	Linear	Linear Quadratic	Log	Log-Embedded	LatentClass	
Parameter					Segment 1	Segment 2
Content	0.2679(***)	0.3001(***)	0.1026(***)	0.1189(***)	0.5923(***)	-0.2268(.)
Content^2		-0.0041			-0.0348(*)	0.0429(*)
BorrowedContent	0.1273(.)	0.2260(**)	0.0731(**)	0.0134(.)	0.1115	1.7342(***)
BorrowedContent^2		-0.0480(**)			-0.0192	-0.2986(***)
NetworkPosition	0.3505(***)	0.8268(***)	0.5922(***)	0.7991(***)	2.112(***)	4.1225(***)
NetworkPosition^2		-0.0949(***)			-0.5463(***)	-0.4509(***)
Diversity	0.0945(***)	0.0796(***)	0.1357(***)	0.1420(***)	0.0982(***)	
Popular	0.5224(***)	0.5219(***)	0.5238(***)	0.5656(***)	0.5024(***)	
Advisor	0.0818(***)	0.0491(*)	0.0716(***)	0.0432(**)	0.0463(***)	
BorrowedDiversity	-0.0303	-0.0575	-0.1188(**)	-0.0684(***)	-0.1212(*)	
BorrowedPopular	0.1853(***)	0.1759(***)	0.1730(***)	0.1562(***)	0.2578(**)	
BorrowedAdvisor	0.1265(**)	0.1221(**)	0.1366(***)	0.1819(***)	0.0832(*)	
Constant	-10.1255(***)	-10.4231(***)	-7.7108(***)	-7.8858(***)	-10.8349(***)	
Segment Size					0.932	0.068
-LL	6930.51	6734.51	6977.46	7128.42	6692.44	
BIC	6990.82	6820.67	7037.77	7188.73	6847.52	

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1

Table 6: Dynamic Cost Parameter Estimation

	Estimate	Low 95% CI	High 95% CI
Production			
Constant	0.1481(*)	0.1326	0.1597
Quality	0.2268(*)	0.1866	0.2536
Tenure	-0.0051	-0.1156	0.0101
Link			
Constant	0.025(*)	0.0015	0.0715
Quality	-0.0646	-0.1703	0.0556
Tenure	-0.0132	-0.0226	0.0026
Reciprocal	0.1761	-0.1315	0.4755

Unit of measure: thousand page views

Table 7: Simulation – System-wide Link Removal

	Demand	Content Production	Link Formation
Average Increase Per Day	12.096	38.08	1.22
Percentage Increase	17.12%	54.40%	31.32%

Figure 1: Average Content Production by Own State

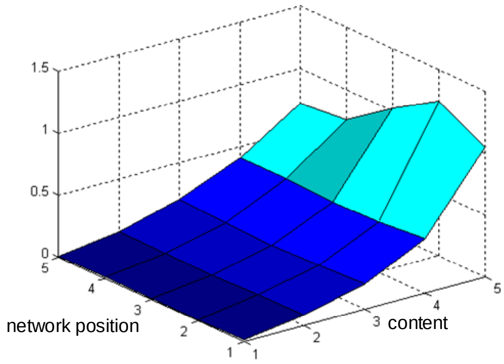


Figure 2: Probability to Form Links by Own State

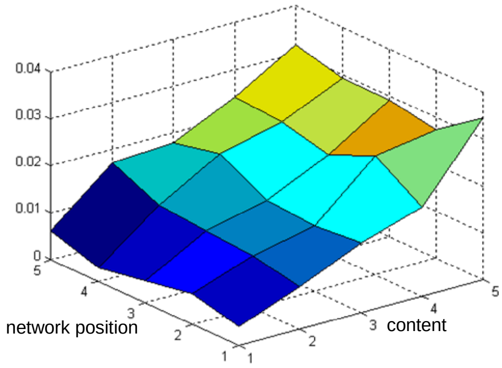


Figure 3: Relative Link Probability – Non-reciprocal over Reciprocal

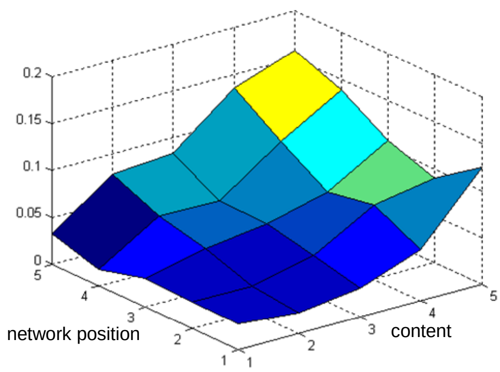
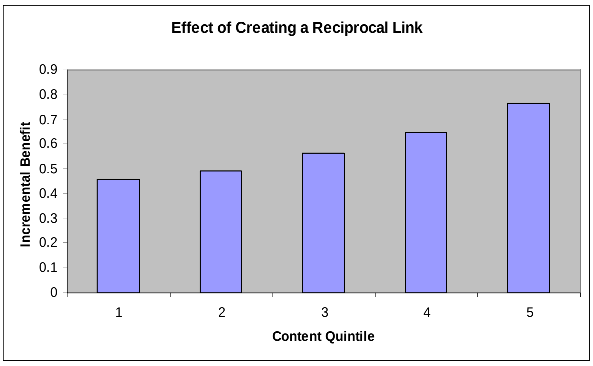


Figure 4: Effect of Creating Reciprocal Links



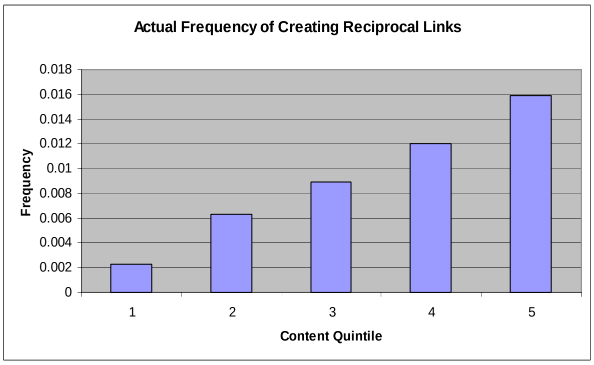
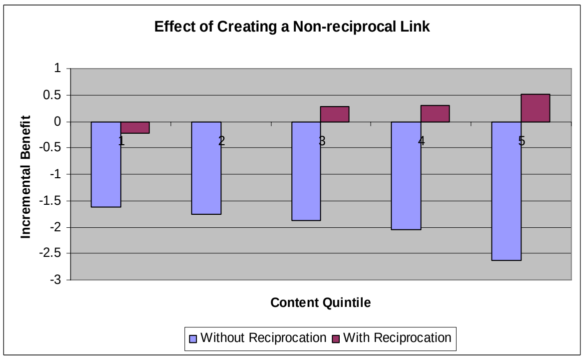


Figure 5: Effect of Creating Non-reciprocal Links



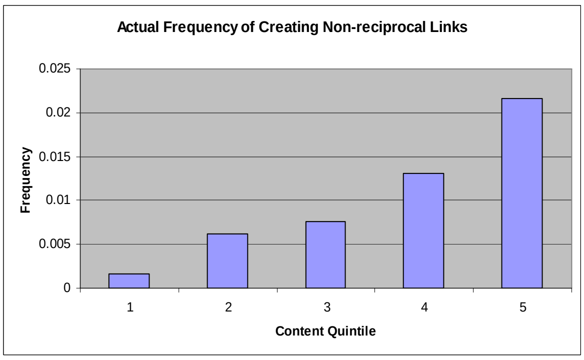
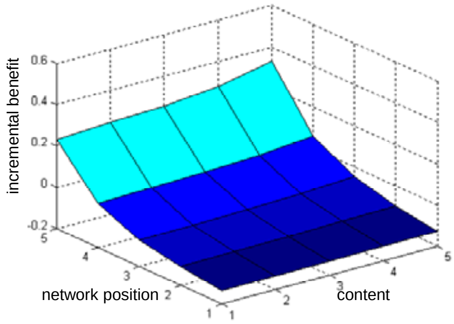
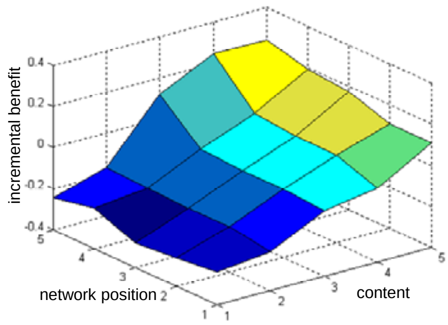


Figure 6: Incremental Benefit of Content Production by State

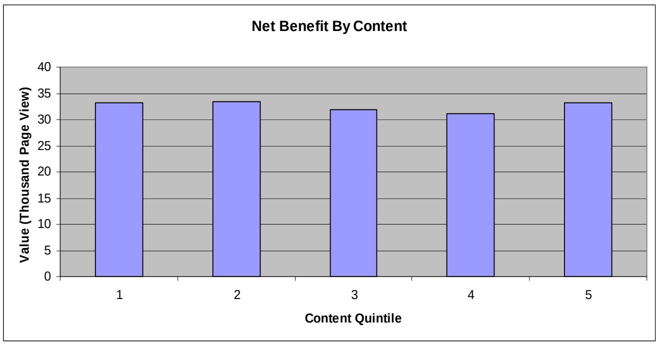


(A) not account for change in network position



(B) account for change in network position

Figure 7: Net Benefit by Content and Network Position



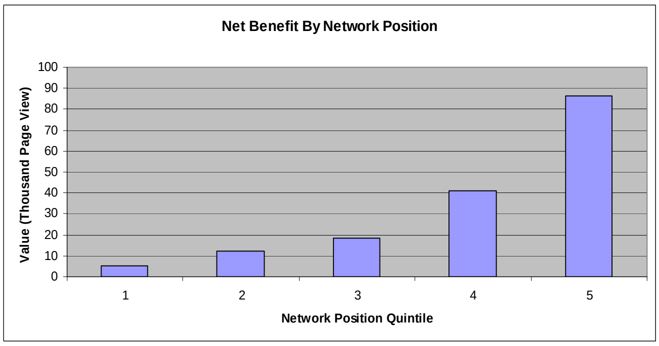


Figure 8: Simulation – System-wide Link Removal