Media, Aggregators and the Link Economy:
A Strategic Network Formation Perspective on the Future of Content

Chrysanthos Dellarocas
Boston University
School of Management
dell@bu.edu

Zsolt Katona
Haas School of Business
University of California, Berkeley
zskatona@haas.berkeley.edu

William Rand
Robert H. Smith School of Business
University of Maryland
wrand@umd.edu

Abstract. The debate between content aggregators and content producers is raging and may soon reach the courts, or even the U.S. Congress. The appropriation of advertising and other revenues associated with site traffic is a key point of contention. The future of journalism and the future of copyright are at stake. This important debate has so far involved CEOs, bloggers, consultants and lawyers but almost no academics. We fill this gap by proposing what we believe is the first rigorous attempt to model the micro and macro level implications of hyper-linking and content aggregation for content industries and consumers. Our aim is to provide a scientific framework that can help assess the implications of the diverse strategies and policy interventions that are currently on the table. To conquer the inherent complexity of the settings we study, we follow a three-pronged approach that consists of game-theoretic modeling, agent-based simulation and data analysis. Our first round of modeling has uncovered some key insights and parameters whose value qualitatively affects the nature of the predicted outcomes and their normative implications. Our continued work places an emphasis on developing methods for empirically estimating some of the key parameters that our initial modeling work has uncovered.

Keywords: content industries, news industries, aggregators, online advertising, link economy

Content industries are in turmoil. Traditional content creators, such as newspapers, are facing stiff competition from “amateur” substitutes, such as blogs, as well as from new players, such as search engines and other types of aggregators, that attract traffic and revenues by hosting collections of links to others’ original content.

As traditional content organizations are seeing their business models collapse, some have vocally denounced aggregators, demanding payment for incoming links and lobbying for copyright law reform [1]. Others point out that without links there is no traffic and that content organizations must adapt and learn to play by the new rules of what is often referred to as the link economy [2]. Additional proposals include micropayments for content, novel revenue sharing schemes between aggregators and content providers, the transformation of content organizations into platforms that work with freelance content creators and the formation of cooperative aggregators, jointly owned by coalitions of media organizations, to compete with the likes of Google and the Huffington Post [3]. Key aggregator executives, such as Google’s Eric Schmidt, assert that it is to their interest to see content creators survive since, without good content, links are worthless [4]. The debate is complex, lively and often emotional. It may soon reach the courts or even the U.S. Congress. So far it has involved well-known CEOs, journalists, bloggers, consultants and lawyers but, interestingly, very few academics. Conspicuously missing is a scientific basis for assessing the numerous arguments and counter-arguments currently on the table and for guiding the industry forward.

The objective of this work is to fill this gap by developing a rigorous modeling foundation that captures the changes taking place in the world of content, helps the various stakeholders assess the implications of the diverse proposals on what to do next and understand what strategies work and which don’t within the context of each.

The key elements of our approach consist of:

I. modeling content industries as dynamic networks of nodes (content sites, aggregators, blogs, etc.) that make interdependent strategic decisions on how to allocate effort between content creation and link formation

II. recognizing that the formation of links presupposes costly search whose effectiveness depends on every other node’s content and link formation decisions

III. explicitly modeling users and their news consumption practices

We explore the following broad sets of questions:
1. how organizations can best combine content creation, search and linking to third-party content to optimize traffic and revenue in dynamic settings

2. what are the properties (quality, diversity, equity, etc.) of content networks that emerge from such strategic processes and how they affect the payoffs of the various stakeholders (content organizations, consumers, society as a whole)

3. how does the presence of aggregators affect (1) and (2)

4. how do the various “new business models for news” currently being proposed (e.g. micropayments, revenue sharing, coalition formation, etc.) affect (1), (2) and (3)

We attack the above questions using a combination of game theoretic modeling, agent-based simulation and empirical validation. We use analytical methods to construct baseline models and gain initial insights. We complement these with an agent-based simulation framework (already implemented; using Netlogo [5]) to capture the full complexity of our settings we aim to study and thus gain more precise predictions and insights. As is the case with many complex systems, some predictions of our models qualitatively depend on the values of key parameters. We use empirical methods to estimate these parameters and validate our theoretical predictions.

Our models assume that there is a cost to content creation that is related to its quality, and a cost associated with searching for content (for the purpose of linking to it) that is also related to the quality of the results. Furthermore, we model consumer news consumption as a multi-armed bandit problem [6]: consumers begin their regular (e.g. daily) news consumption from an anchor node (e.g. a source of news they have come to like and trust) and then follow whatever links are posted on that node. Consumers prefer to stick with the same anchor node but occasionally “try out” different nodes and switch to them if they happen to give them access to better content. We assume that content nodes gather income from every visitor who consumes their content, and that they also gather income from visitors who were directed to their node via a link; however, we posit that there is some income that is lost as a result of each link that a user has to traverse before arriving at a destination node. This loss can be due to factors such as: (a) that some aggregator users do not actually click through to the final node, or (b) that there is a finite amount of online advertising money out there and since advertisers advertise on both aggregator nodes and content nodes, links necessarily devalue the amount of income a content creator can receive [4]. A key focus of our ongoing work is to empirically deepen our understanding of consumer news and advertising consumption processes, especially in terms of how they are affected by the presence of aggregators.

We have solved versions of the above model in which a group of content creators is “invaded” by a group of aggregators, the latter being modeled as nodes that have a high cost of content production and a low cost of search. Our key results can be summarized as follows:

1. By providing consumers with easy access to more quality content than any single content node can produce, aggregators supplant traditional content producers as the preferred anchor node from which most consumers begin their news consumption sessions.

2. By reducing consumers’ search costs of accessing good content, aggregators induce an increase in the cumulative amount of content that consumers consume. This, in turn, drives more traffic to content producers, giving them incentives to produce higher quality content.

3. By appropriating part of the benefit of site traffic (see discussion in previous paragraph), aggregators reduce the per-visitor profits of content producers. This, in turn, reduces their incentives to produce good content.

The cumulative prediction of our model depends on the interplay of these three forces. We use the symbol $\alpha$ to represent the fraction of aggregators in the population of nodes, $\beta$ to denote the increase in content consumption due to the presence of aggregators and $\gamma$ to denote the per-visitor revenue penalty that a content node suffers when a user reaches that content via an aggregator link and not by visiting directly the content node.

If $\beta$ is large relative to $\gamma$, a win-win scenario ensues: Consumers consume more and better content, content nodes produce better content and generate higher profits and aggregators are profitable. However, as $\gamma$ grows, aggregators appropriate a larger slice of the site traffic revenue. There is a threshold of $\gamma$ above which the revenue losses experienced by content producers become larger than the revenue gains due to additional traffic. Above that threshold, content producers have diminishing revenues and diminishing incentives to produce good content. Interestingly, however, lower quality content also reduces consumers’ appetite for content and, thus, aggregator profits. Consistent with Eric Schmidt’s thinking [4] our model shows that the incentives of aggregators and content producers are more closely aligned than most people in the industry are willing to acknowledge.
The impact of the fraction of aggregators ($\alpha$) is subtle: If $\alpha$ is too small then aggregators are not significant enough to substantially increase the population’s appetite for content. If $\alpha$ is too large (which would be an extreme scenario in settings where aggregators prove to be substantially more profitable than content producers), then there is not enough original content and everyone loses. Just as in the case of chemical catalysts or some pairs of symbiotic biological species, there exists a “golden ratio” of aggregators and content producers that optimizes social welfare; welfare suffers if the system moves away from this ratio in either direction.

As is obvious from the above discussion, the relative values of $\alpha$, $\beta$, $\gamma$ are key to the qualitative nature of the model’s predictions as well as to its normative implications. For example, large $\beta$ and small $\gamma$ leads to a situation where aggregators are net social benefactors, whereas small $\beta$ and large $\gamma$ leads to a situation where the presence of aggregators damages both content producers and consumers.

Our continuing work in this area will develop sound empirical estimates of these three key parameters and, more broadly, more elaborate data-driven models of the consumer-level processes of news and associated advertising consumption. We have initial ideas on how to go about this task. For example, $\alpha$ can be assessed by comparing click-through patterns of users who begin their news consumption session in a content node vs. those who use aggregators as anchors. We could assess $\beta$ by observing click-through patterns on an aggregator website such as Google News to determine how often a reader actually clicks through to the underlying website, and by observing the sponsored link click-through behavior of users when visiting aggregator vs. content nodes. Parameter $\gamma$ can be estimated by observing the various types of nodes involved in the coverage of particular news topics.

We hope that our work will allow us to enrich the current debate on the future of content with a scientific framework that will help content producers, aggregators, and policy makers better assess the economic and social implications of the various new business models and policy interventions that are being considered and perhaps, even inspire entirely new ideas that will help move this important set of industries forward.

(The status of this work is preliminary; however we expect to have a working paper available by the end of February.)

References

1. The current debate between the Associated Press and aggregators (including Google) is perhaps the best example. The debate is ongoing and there are several good online articles and postings available about it.
2. Blogger and CUNY Journalism Professor Jeff Jarvis is widely credited with coining the term “link economy”. Most of his essays appear on his blog http://www.buzzmachine.com
3. Several of these ideas are summarized in Arnon Mishkin’s influential blog posting The Fallacy of the Link Economy (http://paidcontent.org/article/419-the-fallacy-of-the-link-economy/)