Social Media as an Innovation – the Case of Twitter *

by

Huaxia Rui[†] and Andrew Whinston[‡]

*We thank Thomas A. Weber for his careful review and useful comments. We are also thankful for the comments of participants in the 20th Workshop on Information Systems and Economics (2009). Existing errors are our sole responsibility.

†University of Texas at Austin, McCombs School of Business, Austin, TX 78741, phone: 512-471-6533, e-mail: ruihuaxia@mail.utexas.edu

‡University of Texas at Austin, McCombs School of Business, Austin, TX 78741, phone: 512-471-7962, fax: 512-471-0587, e-mail: abw@uts.cc.utexas.edu

Abstract

We model a social media environment as an economy where consumption of content means supply of attention and consumption of attention is through supply of content. In such an economy with large population, there exists an attention wage and the community will be segmented. Under certain conditions, users either produce or consume content, meaning a perfect division of labor. The macro-level content consumption and production is stable, suggesting the sustainability of social media. Our theory is supported by data from Twitter and suggests that the key innovation of social media is recognizing and connecting people's need for information and attention.

Keywords: social media, user-generated content, Twitter, innovation

1 Introduction

> What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention, and a need to allocate that attention efficiently among the overabundance of information sources that might consume it. (Simon, 1971)

The abundance of free content on the Internet and the assistance of online search engines have dramatically transformed the way people acquire information. Nowadays, with fast Internet access, cheap devices, and innovative services such as Twitter, YouTube, and Wikipedia, an ever-growing proportion of the content on the Internet is generated by ordinary Internet users. For example, Wikipedia has more than 13 million articles as a result of contributions by volunteers. According to YouTube Fact Sheet, people are uploading hundreds of thousands of videos daily. On Twitter, a micro-blogging site that allows its users to broadcast short messages to their followers, people can find opinions and information on a broad range of topics posted by their peers almost in real time. Despite the differences in the content format, they are each produced by a large number of ordinary Internet users, rather than by only a few publishers and television networks, as in the traditional media. This ongoing shift to social media means that user-generated content is now playing an unprecedented role in people's life and will probably completely alter the way content is generated in our society in the future. [1]

The fascinating phenomenon of social media poses interesting and important research questions. For example, although it is quite reasonable that users consume others' content because they get benefits from content consumption, it is not yet clear what motivates people to contribute content. As is often reported, only a small proportion of users actually contribute content, while most users are "inactive" in the sense that they do not contribute

[1] User-generated content also has potential value to companies. See for example a recent Wall Street Journal article "Follow the Tweets" by Rui, Whinston, and Winkler, published on November 30, 2009.

any content at all. For example, on Wikipedia, the top 15% of the most prolific editors account for 90% of Wikipedia's edits. In a recent article on the blog of Harvard Business School, Bill Heil and Mikolaj Piskorski found that the top 10% of prolific Twitter users accounted for more than 90% of the tweets, and they suggest that Twitter resembles more of a one-way, one-to-many publishing service. [2] A natural question to ask is if this "seemingly" unbalanced structure is sustainable and one may wonder whether social media is a revolution or merely another fad? This leads to the more fundamental question of what is the unique innovation of social media. Is it about technological advancement like fast chips and better algorithm, or something else? We argue in this paper through both theoretical modeling and empirical study that the key innovation of social media is recognizing and connecting people's need for information and attention. People have a natural need for information that drives them to search for content on the Internet. The more content available, the better such need will be satisfied. [3] On the other hand, attention from others is also extremely valuable and is greatly appreciated by many people. For example, by getting attention, people get publicity, vanity, or ego gratification from peer recognition. There are also various ways for people to monetize the attention they get from others. However, people are often intrinsicly or extrinsicly heterogeneous in their preferences towards content (information) and attention. Moreover, the productivity of generating valuable content also varies among the population. Some people are in a much better position than others in terms of producing useful and interesting content because of their knowledge, profession, etc. Just like many financial innovations that recognized the heterogeneity of people's attitudes towards different risk and thereby created the marketplace for people to exchange these risk, social media services like Twitter have successfully created the marketplace for people to exchange content and attention.

[2] http://blogs.hbr.org/cs/2009/06/new_twitter_research_men_follo.html

[3] Nowadays, search cost is low because of many powerful search engines that are freely available.

To better understand our theory, we conceptualize a social media environment as an economy where consumption of content means supply of attention and consumption of attention is possible only through the supply of content. Both the supply of content and the supply of attention involve investment of time, hence time is money in this economy. Users in the economy interact with each other through their allocations of time into content consumption and content production. To make this rigorous, we develop a game-theoretical model to capture and study this interaction among users, and then use a massive collection of data from Twitter to test four hypotheses that are based on the theory.

More specifically, we model each user as a rational agent who faces a limited endowment of time and tries to optimally allocate her time to maximize her utility from the consumption of content and attention. Their decisions are intertwined since a user's utility from consumption depends on others' production decisions while a user's utility from production depends on others' consumption decisions. We explore the equilibrium outcome of this game by characterizing how users choose their roles in the community and consequently how the community is segmented into content consumers, content producers, and content prosumers (i.e. users who engage in both content production and consumption). We find that as the community size grows large enough, there exists a unique constant for the community which we interpret as the community attention wage for producing content. Each user compares their two individual reservation wages with the community wage to determine whether they want to be a content producer, consumer, or prosumer. Interestingly, we also identify the existence of a special equilibrium, called the partition equilibrium in our term, which occurs when users' utility functions satisfy a certain linearity condition. In the partition equilibrium, users self-select themselves into either the group of content consumers or the group of content producers. These results suggest that there is a tendency for specialization of users' utility maximization strategies. They also shed some new light on our understanding of the role of those so-called "inactive" users in many social media sites. To understand the sustainability of the social media, we extend the model to the dynamic setting where users also choose endowments of time based on their opportunity costs. We find that the system that characterizes the content demand and supply at the macro-level has a non-trivial, asymptotically stable equilibrium point suggesting the sustainablity of social media supported by user-generated content.

To empirically verify our theory, we use data collected from twitter.com to test four hypotheses. Our first hypothesis says that users who value attention more tend to produce content more frequently. We support this hypothesis with nearly 3 million Twitter user profiles. The second hypothesis says that more capable Twitter users (in terms of producing content) produce content more frequently which is again consistent with the data. Our third and fourth hypotheses are based on an econometric model characterizing the distribution of users' frequencies of generating content and are also supported by the data.

The paper is organized as follows. Section 2 reviews the relevant literature. Section 3 presents our basic theoretical model. In Section 4, we focus on the partition equilibrium and discuss the macro-level stability property of the partition equilibrium. In Section 5, we propose four hypotheses and test them with the data from Twitter. We conclude in Section 6 with a discussion of our findings, their practical implications, limitations and suggestions for future work.

2 Relevant Literature

There is a diverse literature related to our research questions, among which we will mainly introduce three streams of literature: economics, management, and computer science.

From a traditional economics point of view, contributing content in an online community is like the private production of public goods. Producing content is a cooperative behavior, while consuming content without contributing content is a non-cooperative behavior. The puzzle is why so many users actively spend their valuable time providing content instead of free-riding on others' contribution, which would eventually lead to the collapse of the social media. Unlike the offline world, where formal enforcement mechanisms like contracts and agreements can be made, in the online world, people are only loosely connected and formal means of enforcement are just not feasible. Economists have also studied informal enforcement mechanisms, including personal enforcement and community enforcement. Personal enforcement involves retaliation to the non-cooperative agent by the victim who plays a cooperative strategy. However, in most large social media sites, users seldom have such kind of long-term relationships. On the other hand, the community enforcement mechanism offers another explanation as to why people cooperate (Kandori 1992). Roughly speaking, cooperation may be sustained in the community because people fear that if they stop cooperating, non-cooperative behavior will "spread" like a disease to others, and this contagious process will eventually bring down the community from which they all benefit. Although such cooperative equilibrium seems to fit the context of online communities, it is quite fragile in the sense that small noise may cause the complete breakdown of cooperation in the community. These days, large social media sites have tens of millions of users. It will be extremely difficult, if not impossible, to sustain this type of cooperative equilibrium.

Rather than assuming that contribution is purely a cooperative behavior that only benefits others, some economists and sociologists take another approach by arguing that contributing itself benefits the contributor. In the literature of economics of gift and charity, researchers suggest that people have a taste for giving. For example, Andreoni argues that "egoists" and "impure altruists" not only care about supplying public good but also experience a "warm glow" from having "done their bit" (Andreoni 1989). Roberts, Hann and Slaughter categorize three factors that could motivate users to contribute to open source software (OSS) development: [4] intrinsic, extrinsic, and internalized extrinsic factors (Roberts

[4]OSS could be viewed as a special type of user-generated content. Developers spend time developing

et al. 2006). Intrinsic factors refers to the satisfaction and enjoyment obtained from creating and contributing (Shah 2006) while extrinsic factors refers to the incentive provided by external environments, including organization rewards (Bock et al. 2005), career opportunities (Shah 2006, Jeppesen, 2006), etc. Internalized extrinsic factors refers to extrinsic motivations that are self-regulated instead of directly imposed by external environments, like reputation (Wasko 2005) and status seeking (Robert 2006). Lerner, Pathak, and Tirole (2006) argue that open source software developers may get some short- or long-run benefits. For example, a programmer may find intrinsic pleasure, get ego gratification from peer recognition, attract potential future employers etc. Whether the intrinsic factors, extrinsic factors, or internalized extrinsic factors motivate a user to contribute, it is reasonable to assume that, in most cases, more attention will lead to higher value to users who contribute. This is also consistent with Lerner, Pathak, and Tirole's suggestion that the more visible the performance to the relevant audience (peers, labor market, and venture capital community), the stronger such benefits will be. In this paper, we do not go into the underlying psychological and sociological mechanisms of why people contribute content in online communities. Rather, we assume that these underlying mechanisms manifest themselves through people's seek of attention from others.

The literature on the economics of Peer-to-Peer (P2P) networks is also related to this paper because, broadly speaking, sharing resources on P2P networks is analogous to contributing content in online communities. Researchers have long been discussing the "free-riding" problem of P2P networks. Various incentive mechanisms have been proposed to tackle this problem. In a recent paper, Feldman et al. (2006) developed a modeling framework that takes users' generosity into account. Their paper focused very specifically on P2P networks and their suggestion that free-riding could be sustainable in equilibrium is very illuminating

valuable open-source software and users spend time using these software and obtaining value from their usage.

and reinforces our result on user segmentation in equilibrium. Also, the "generosity-driven" view is analogous to our "attention-driven" view. However, the users' decision problem in their model is rather simplistic compared with our more comprehensive one. In this sense, our model significantly extends their work.

In a recent experimental study on the communication structure of virtual communities, Sohn and Leckenby (2007) compared the performance of two different communication structures—one based on a public electronic bulletin board (group-generalized exchange, G_{EX}) and the other based on the interpersonal networks (network-generalized exchange, N_{EX}, e.g., blogs, Twitter). They found that the latter is more effective in the sense that more contributions were made. Their explanation is that the network-generalized exchange structure enhances contribution efficacy because in N_{EX} each user's contributions are presented separately from those of others. Our interpretation for this is that in G_{EX}, even though the community as a whole gets attention, individual users do not get much attention, which severely limits users' motivation to contribute. On the other hand, in N_{EX}, contribution can attract attention directly to the contributor because content is accessed at each user's individual homepage rather than the common pool.

There is also a growing interest in the computer science literature on user-generated content. For example, Huberman et al. (2008) showed through an analysis of a massive data set from YouTube that the productivity exhibited in crowdsourcing exhibits a strong positive dependence on attention, measured by the number of downloads. They found that lack of attention leads to a decrease in the number of videos uploaded and the consequent drop in productivity. In another recent paper, Guo et al. (2009) empirically studied the patterns of user content generation in three online communities, including a blog system, a social bookmark sharing network, and a question-answering social network. They found that the rank order distribution of user posting follows stretched exponential distribution, which is quite close to our finding of exponential distribution of user contribution rates in

Section 5. The major difference is that we derived the exponential distribution from our theoretical model, with simplifying assumptions on the distribution of users' heterogeneity. The theoretical foundation of our empirical study differentiates our work from that literature.

3 The Model

3.1 Model Setup

There are n users in an online community where n is large. In this section, we assume that each user spends a fixed amount of time in the community; we call this the time budget and denote by T_i for user i with $T_i \in [\underline{T}, \overline{T}], 0 < \underline{T} < \overline{T} < \infty$. There are two ways for a user to spend time in the community: consuming content or producing content. We use r_i to denote the proportion of time user i spends on consuming content produced by other users and w_i the proportion of the time user i spends on producing content; thus $r_i + w_i = 1, r_i \geq 0, w_i \geq 0$. User i will produce $S_i = q_i T_i w_i$ amount of content, where q_i is the productivity of user i in producing valuable content, $q_i \in [\underline{q}, \overline{q}]$, $0 < \underline{q} < \overline{q} < \infty$, and w_i is the decision variable for user i.

We denote the total amount of content produced in the community by

$$S = \sum_{k=1}^{n} q_k T_k w_k. \tag{1}$$

We also use S_{-i} and S_{-ij} to denote the total amount of content produced by everyone except user i and the total amount of content produced by everyone except user i, j respectively.

A user obtains utility from two sources. First, since a user gets information or pleasure from the content, she obtains utility from consuming the content. The amount of utility a user can get by consuming content depends not only on the time she devotes to it but also on the amount of content available to her. Hence, we model this part of utility as follows:

$$u_i^r = \psi(T_i r_i)\phi(S_{-i}), \tag{2}$$

where $\psi(0) = 0, \psi' > 0, \psi'' \leq 0$, and $\phi(0) = 0, \phi' > 0, \lim_{x\to\infty} \phi(x) < \infty$. The concavity assumption of ψ means decreasing marginal utility from consuming content. The monotonicity assumption of ϕ means that the larger the amount of content available, the more utility a user can get from consumption for each time unit. If there is no content available, then a user can't get any value from consuming so that the utility is zero ($\phi(0) = 0$). Although an increase of content available to user i will lead to an increase of utility given $T_i r_i$, we believe such effect is bounded as S_{-i} goes to infinity so that $\lim_{x\to\infty} \phi(x) < \infty$. The rationale behind this is that a user can only get a limited amount of utility for each unit of time, regardless of how much content is available to her.

The second way for a user to obtain utility is by producing content. A user may enjoy publicity so that the more attention she gets from others who consume her content, the higher her utility is. To model this, we assume that the utility a user gets from content production is proportional to the total attention she gets from all other users, which we measure by the total time others spend on consuming her content. Intuitively, a user with more content should have a higher chance of getting attention from another user. To make things simple, we assume the amount of attention user i can get from user j is proportional to the content produced by user i, which is:

$$a_i^j = \frac{q_i T_i w_i}{S_{-j}} T_j r_j.$$

In the above, the denominator is the total amount of content available to user j. So how much attention a user can get depends on her relative standing in terms of content in the community. Summing up over j gives the total attention user i gets:

$$a_i = \sum_{j \neq i} \frac{q_i T_i w_i}{S_{-j}} T_j r_j.$$

We use $\alpha_i \in [\underline{\alpha}, \overline{\alpha}]$ to account for users' heterogeneity in terms of their preferences for attention relative to that of content consumption where $0 \leq \underline{\alpha} \leq \overline{\alpha}$. The larger α is, the more a user values attention. User i's utility function is then defined as follows:

$$u_i = \psi(T_i r_i)\phi(S_{-i}) + \alpha_i \cdot \sum_{j \neq i} \frac{q_i T_i w_i}{S_{-j}} T_j r_j. \quad (3)$$

The structure of the static game is described as follows. First, each user is endowed with (α, q, T), and the population endowment $(\alpha_i, q_i, T_i), i = 1, \cdots, n$ is common knowledge. Then, each user decides the proportion of time she will spend on content consumption (r_i) and content production (w_i) to maximize her utility.

The equilibrium concept we use is Nash equilibrium, and we are only interested in pure-strategy Nash equilibrium, where each user chooses $w_i \in [0, 1]$. The utility maximization problem for user i, $i = 1, 2, \cdots, n$, is formally written as:

$$\max_{0 \leq w_i \leq 1} u_i = \psi(T_i(1 - w_i))\phi(S_{-i}) + \alpha_i \cdot \sum_{j \neq i} \left(1 - \frac{S_{-ij}}{S_{-ij} + q_i T_i w_i}\right) T_j r_j, \quad (4)$$

which is a complex problem in general because the solution depends on all other users' decisions. Particularly, we should be aware that in equilibrium, $r_j, w_j, 0 \leq j \leq n$ are all functions of α_i, q_i and w_i. However, when the community size is large enough, certain properties of the equilibrium could be characterized. Although our results hold under general distribution of (α, q, T), for ease of illustration, we assume from now on that the distribution of (α, q, T) is absolute continuous in its support.

3.2 Homogeneous Users

In the simplest case where users are homogeneous in the sense that $(\alpha_i, q_i, T_i) = (\alpha_0, q_0, T_0)$, we would expect each user of the community to serve both as a content producer and content consumer. Indeed, such kind of equilibrium always exists. Assume $w_j = w_0, r_j = 1 - w_0$, $j = 1, \cdots, i-1, i+1, \cdots, n$; then user i's utility function simplifies to:

$$u_i = \psi(T_0(1-w_i))\phi\left((n-1)T_0 w_0 q_0\right) + \alpha_0(n-1)T_0(1-w_0)\frac{w_i}{(n-2)w_0 + w_i}.$$

The first-order condition is

$$-\psi'(T_0(1-w_i))\phi\left((n-1)T_0 w_0 q_0\right) + \alpha_0(n-1)(1-w_0)\frac{(n-2)w_0}{((n-2)w_0 + w_i)^2} = 0.$$

We need to examine whether w_0 is the solution to the above equation. Substituting $w_i = w_0$ into the equation, we have:

$$-\psi'(T_0(1-w_0))\phi\left((n-1)T_0 w_0 q_0\right) + \alpha_0 \frac{n-2}{n-1}\frac{1-w_0}{w_0} = 0.$$

Notice that the left-hand side is negative if $w_0 = 1$ and positive if $w_0 \to 0$ as long as $\alpha_0 > 0$. So there always exists $w_0 \in (0, 1)$ satisfying the above equation, which means each user's first order condition is satisfied. Therefore, in a homogeneous community, as long as the common α is positive, there exists a symmetric equilibrium where everyone spends time in both content consumption and content production.

3.3 Heterogeneous Users

It makes more sense to assume that community users are heterogeneous in terms of both productivity (q_i) and motivation (α_i). A natural question to ask is how the consumption and production of content will be organized in such a heterogeneous community. For example, is there an equilibrium where everyone in the community serves as both content producer and content consumer, in a way closer to our intuitive understanding of the social media? Or on the other extreme, is there an equilibrium where users self-select themselves into either content consumers or producers but not both, in a way resembling the feature of the traditional media? It turns out that the answer lies somewhere in between. While the division of labor is inevitable, the content consumption and content production must be balanced, i.e., the consumption time and production time must be of the same order. This is in clear contrast to the traditional media where only a handful of newspaper and television broadcast content to the massive audience.

To characterize the equilibrium, we start by categorizing users in the community. Users could be classified into three groups in any equilibrium: those who only consume content (i.e., $I_C = \{i : w_i = 0\}$), those who only produce content (i.e., $I_P = \{i : w_i = 1\}$), and those who both consume and produce content (i.e., $I_M = \{i : 0 < w_i < 1\}$). We call user i a consumer if $i \in I_C$, a producer if $i \in I_P$, and a prosumer if $i \in I_M$. Denote n_C, n_P, n_M to be the number of users in each group correspondingly with $n_C + n_P + n_M = n$. Also, we define $T^C = \sum_{i \in I_C} T_i$, which is the total amount of time spent by consumers, and denote $T^C_{-i} = T^C - T_i$ as the total amount of time spent by all consumers except user i. The following lemma partially characterizes the equilibrium structure of a heterogeneous community.

Lemma 1. *In any equilibrium of the game, the three groups are characterized as follows*

$$
\begin{aligned}
i \in I_C &\iff \alpha_i q_i \le h_C(i) = \frac{S_{-i}\phi(S_{-i})\psi'(T_i)}{T^C_{-i} + S_{-i}\sum_{j \in I_m, j \neq i} \frac{T_j r_j}{S_{-ij}}}, \\
i \in I_P &\iff \alpha_i q_i \ge h_P(i) = \frac{(S_{-i} + q_i T_i)^2}{S_{-i} T^C_{-i} + (S_{-i} + q_i T_i)^2 \sum_{j \in I_M, j \neq i} \frac{S_{-ij}}{(S_{-ij} + q_i T_i)^2} T_j r_j} \phi(S_{-i})\psi'(0), \\
i \in I_M &\iff h_C(i) < \alpha_i q_i < h_P(i).
\end{aligned}
$$

Notice that the above lemma does not fully characterize each group because the right-hand side of each inequality depends on the decisions of each user in the community. However, it does suggest that the product αq measures to some extent the willingness of a user to produce content.

The next lemma is a technical result that we use in Proposition 1, but the intuition is also not difficult to understand. Basically, it says that as the community size grows, the total attention (total consumption time) can not increase too fast compared with the increase of the amount of content. In others words, the total consumption time and total production time must be balanced.

Lemma 2. *In any equilibrium*[5]*:*

$$\lim_{n\to\infty} S = \infty, \quad \lim_{n\to\infty} \frac{T^C_{-i}}{S^\beta_{-i}} = 0, \quad \lim_{n\to\infty} \sum_{j\in I_M, j\neq i} \frac{T_j r_j}{S^\beta_{-ij}} = 0, \quad , \forall i, j = 1, 2, \cdots, n, \forall \beta > 1.$$

Based on Lemma 1 and Lemma 2, we are able to characterize the asymptotic properties of $h_P(i)$ and $h_C(i)$, the thresholds that determine whether a user chooses to become a consumer, a producer, or a prosumer. The following proposition claims that there will be one common h_P for all users, while $h_C(i)$ only depends on h_P and T_i as $n \to \infty$.

Proposition 1.

$$\lim_{n\to\infty} h_P(i) = h_P, \quad \lim_{n\to\infty} h_C(i) = h_P \frac{\psi'(T_i)}{\psi'(0)} = h_C(T_i),$$

where h_P is a constant. With infinitely large population, user i becomes a

- *producer if $W > W_i^H$,*
- *prosumer if $W_i^H > W > W_i^L$, and*

[5]We exclude one type of pathological equilibrium, where $\lim_{n\to\infty} \frac{n_M}{n} \neq 0$, but $\lim_{n\to\infty} w_i = 0, \forall i \in I_M$.

- *consumer if* $W < W_i^L$.

where $W = 1/h_P$ *is interpreted as the community attention wage and*

$$W_i^H = \frac{1}{\alpha_i q_i}, \quad W_i^L = \frac{\psi'(T_i)}{\alpha_i q_i \psi'(0)}$$

are interpreted as user i's high reservation attention wage and low reservation attention wage respectively.

Notice that W could roughly be interpreted as the ratio of attention supply to content supply, which serves as an indicator of the community wage for producing content in the "attention economy". Each user has two reservation wage levels: the high reservation wage W_i^H and the low reservation wage W_i^L. If the community wage is higher than her high reservation wage, she chooses to become a producer. If the community wage is lower than her high reservation wage but higher than her low reservation wage level, she chooses to become a prosumer. If the community wage is lower than her low reservation wage level, she chooses to become a consumer.

Based on Proposition 1, we could depict how users in the community are divided into producers, consumers, and prosumers as is shown in the $\alpha - q$ plane in Figure 1. In particular, users with $\alpha_i q_i > h_P$ become producers; users with $\alpha_i q_i < h_P \frac{\psi'(\overline{T})}{\psi'(0)}$ become consumers; users with $h_P \frac{\psi'(\underline{T})}{\psi'(0)} < \alpha_i q_i < h_P$ become prosumers; users with $h_P \frac{\psi'(\overline{T})}{\psi'(0)} < \alpha_i q_i < \frac{\psi'(\underline{T})}{\psi'(0)}$, depending on their T_i, either become consumers or prosumers.

Notice that the above result is analogous to the result in a recent paper by Katona and Sarvary (2007) in which they studied the network structure of the commercial World Wide Web. They found that there is a specialization across sites in revenue models: high content sites tend to earn revenue from the sales of content while low content sites from the sales of traffic. In our case, highly motivated and more capable users obtain utility from content production while users who are not highly motivated and less capable obtain utility from

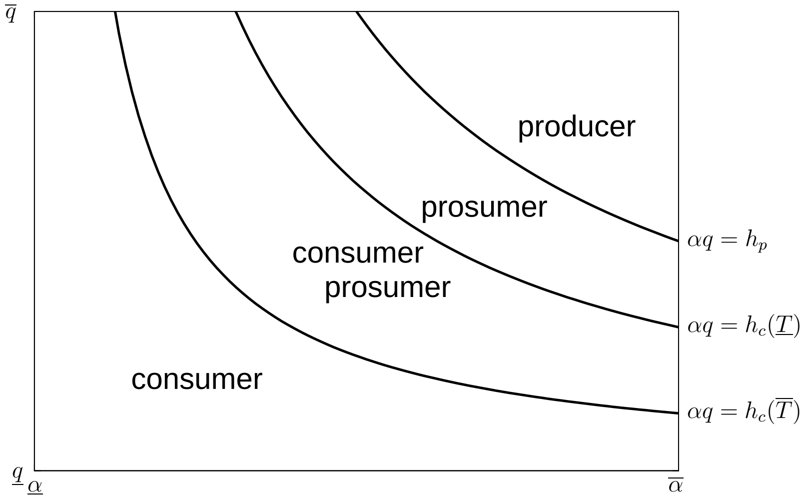

Figure 1: The Segmentation of the Social Media Community

content consumption.

From Lemma 2 and the proof of Lemma 1, we also have the following result.

Corollary 1. *If the community size is large enough, in any Nash equilibrium, (1)* $T_i = T_k, q_i = q_k, \alpha_i > \alpha_k \Rightarrow w_i > w_k$ *(2)* $T_i = T_k, \alpha_i = \alpha_k, q_i > q_k \Rightarrow w_i > w_k$.

The above result claims that in equilibrium, users with higher α or higher q devote a larger proportion of time to content production, holding everything else equal, which is consistent with our intuition.

4 The Partition Equilibrium

4.1 Existence of the Partition Equilibrium

Based on Proposition 1, when $\psi(r_i)$ is a linear function, $W_i^H = W_i^L$, that is, the two reservation wage levels of user i coincide, which means she either chooses to become a content producer or a content consumer. More precisely, under absolute continuous distribution of (α, q), the proportion of those users with $h_C(i) < \alpha q < h_P(i)$ shrinks to zero as $h_C(i)$ and $h_P(i)$ become closer and closer. So we have the following result.

Proposition 2. *With absolute continuous distribution of* (α, q, T)*,* $\lim_{n\to\infty} \frac{n_M}{n} = 0$ *if and only if* $\psi'' = 0$ *(i.e., if* $\psi(\cdot)$ *is linear).*

When $\psi(\cdot)$ is a linear function, we call the equilibrium where $W_i^H = W_i^L, \forall i = 1, 2, \cdots, n$ the partition equilibrium, where the community is simply divided into two groups: content producers and content consumers. This kind of equilibrium is of particular interest because it offers one possible explanation for the often observed phenomenon that in many large social media sites a small proportion of the users accounts for the majority of the content. According to Proposition 2, in the partition equilibrium, there will be a clear division of labor. Content producers often are seen as very active users, and content consumers are often labeled as the "inactive" majority. Even though some may argue that content consumers are free riders of the community, our model suggests that they are as important as content producers to sustain the community because the attention they provide is the main drive for producers to keep contributing.

It should be noted that we do not require n to actually be infinity to have the partition equilibrium. All we need is that n is large enough so that $h_C(i)$ and $h_P(i)$ are close enough to h_P for any i, in which case, $\alpha_i q_i$ falls between $h_C(i)$ and $h_P(i)$ with probability close to zero. With absolute continuous distribution of (α, q), partition equilibrium occurs with probability approaching 1 as n becomes larger and larger.

The intuition of Proposition 2 is as follows. A user's allocation of time on content consumption and content production depends on the marginal utility she can get from the two alternatives. When the community is so large that the content available to her is infinitely abundant, the marginal utility of spending time on consumption approaches to some constant for all users because in such a "content-rich" community, content availability is not a concern for a user and the only determinant of her utility from consumption is her time devoted in consumption. On the other hand, the marginal utility from spending time on content production varies across users because they have different preferences for attention (α) and different productivity (q). In the limit where users' marginal utility from consumption becomes the same, those who have marginal utility from production higher than this constant become producers and those who have marginal utility from production lower than this constant become consumers. The abundance of content in the community makes its users homogeneous in their preferences for consuming content while keeping them heterogeneous in their preferences for producing content.

In this section, we always assume that $\psi(\cdot)$ is linear so that we could focus on the partition equilibrium. Since ψ is linear, we denote $\psi'(T) = \tau > 0, \forall T \geq 0$ for convenience, which means $\psi(T) = \tau T$. Also, we normalize $\lim_{S\to\infty} \phi(S) = \frac{1}{\tau}$ so that $\psi'(T) \lim_{S\to\infty} \phi(S) = 1, \forall T > 0$. This is without loss of generality since we could scale α for the whole community to adjust $\phi(S)$.

The next result shows how the threshold is determined by the population distribution of user profiles (α, q, T).

Corollary 2. *In the partition equilibrium, (1) content consumers obtain utility* $u_i^C = \psi(T_i)\phi(S)$*,* $i \in I_C$*; (2) content producers obtain utility* $u_i^P = \alpha_i q_i T_i \frac{T_C}{S}$*,* $i \in I_P$*; (3) the threshold* $h = \lim_{n\to\infty} h_P(i) = \lim_{n\to\infty} h_C(i)$ *is determined by the following equation:*

$$h = \frac{\sum_{\alpha_i q_i > h} q_i T_i}{\sum_{\alpha_i q_i < h} T_i} \tag{5}$$

Next, we study how a change of α or q in the population affects the content generation in the community. We say the population shifts up in α if $\alpha_i' \geq \alpha_i$ and $\alpha_i \geq \alpha_j \Rightarrow \alpha_i' \geq \alpha_j'$, where α_i' is the value of α for user i after the shift. Shifting up in q is similarly defined.

Corollary 3. *(1) Shifting up of* α *in the population increases both the amount of content produced and the number of producers. (2) Shifting up of* q *in the population increases the amount of content produced.*

Notice that a content producer's utility $\alpha_i q_i T_i \frac{T^C}{S}$ might not increase as q shifts up in the population. But a content consumer's utility does increase since S will increase. This suggests that producers as a whole do not have the incentive to improve productivity. The increase of productivity is driven by the competition for attention. On the other hand, the increase of α does increase the utility of content producers, as well as the utility of content consumers.

4.2 The Macro Dynamics of the Partition Equilibrium

Previously, we have taken T_i as fixed. Here we relax this assumption by allowing users to optimally choose their endowment time T_i based on their individual opportunity cost. For simplicity, we model this opportunity cost as a quadratic function $T_i^2/2\theta_i$, where θ_i reflects the heterogeneity of opportunity costs among users. Now both T_i and w_i are users' decision variables. To keep the model tractable, we assume $\psi''(\cdot) = 0$ and n is large.

The main purpose of this subsection is to show that social media supported by user-generated content is robust in the sense that the macro-level production and consumption of content is stable when there are small changes in T_i, $i = 1, \cdots, n$. The stability contrasts with the instability of the contagious equilibrium discussed in the community enforcement model (Kandori 1992). We argue from a theoretical point of view that the attention-driven perspective could be a more practical explanation for the thriving of user-generated contents. People contribute content to the community not because they fear the community will collapse as a result of their not contributing, but because they could obtain utility from the attention they will get by contributing content.

First, we characterize the equilibrium in this new game.

Proposition 3. *The Nash equilibrium of the game is characterized as follows: (1) User with* $\alpha_i q_i < h = \frac{S}{T^C}$ *chooses* $T_i = \theta_i \phi(S)\tau$ *and* $w_i = 0$*; (2) User with* $\alpha_i q_i > h = \frac{S}{T^C}$

chooses $T_i = \frac{T^C}{S}\theta_i\alpha_i q_i$ *and* $w_i = 1$*.* S *and* T^C *are determined as the solution to the following equations:*

$$\begin{cases} S = \sum_{\alpha_i q_i > \frac{S}{T^C}} \frac{T^C}{S}\alpha_i q_i^2\theta_i \\ T^C = \sum_{\alpha_i q_i < \frac{S}{T^C}} \phi(S)\tau\theta_i \end{cases} \tag{6}$$

We could interpret w_i as user i's using habit which does not change over time. On the other hand, we interpret T_i as her daily or weekly time spent in the community that may fluctuate over time around its average. Hence, we write T_i as $T_i(t)$ and Equations (6) defines a dynamic system of content production and consumption at the macro-level. The next result shows that the point (S, T^C) determined in the equilibrium identified in Proposition (3) is actually an asymptotically stable equilibrium.

Proposition 4 (Macro-Level Stability)**.** *The macro-level dynamics of content production and consumption are characterized by Equations (7)*

$$\begin{cases} T^C = g_1(S) = k_1\phi(S) \\ S = g_2(T^C) = k_2\sqrt{T^C} \end{cases}, \tag{7}$$

where $k_1 = \sum_{i\in I_C}\tau\theta_i$ *and* $k_2 = \sqrt{\sum_{i\in I_P}\alpha_i q_i^2\theta_i}$ *are constant. This dynamic system has an asymptotically stable equilibrium point* (T^{C*}, S^*) *where* $T^{C*} > 0$*,* $S^* > 0$*.*

Since $(T^C, S) = (0, 0)$ is also an asymptotically stable equilibrium point, neither of the two equilibria are globally stable. Even with some small change of T_i, as long as the user's usage habit (i.e., w_i) does not change simultaneously with T_i, each user's choices of T_i will converge to the T_i in the equilibrium since T_i is determined by the macro-level variables S, T^C, which will converge to the values in equilibrium based on the above proposition. We close this section with a comparative static result.

Corollary 4. *Increase of q or α in the population leads to more supply of content (S) and more demand of content (T^C).*

The above result implies that improvement of technological environment (e.g., faster Internet access, popularity of smart phones, camera & camcorder, new services like Flickr, YouTube, Twitter etc.) could stimulate the "attention economy" by increasing q. In the next section, we take a close look at one of these innovative services and actually test some of the results in model.

5 Empirical Test

5.1 Data Description

We collected data from twitter.com, which is an open social networking and micro-blogging service launched publicly in July 2006. It is one of the fastest-growing social network sites in 2009 and is estimated to have tens of millions users. Users can use Twitter to post and read messages known as tweets (also known as updates), which are text-based posts of up to 140 characters. A user's followers are the users who subscribe to receive the user's tweets. Users do not necessarily need to know the people they are following and vice versa. Twitter is particularly effective in connecting people's need for information and attention through this structure.

We used Twitter's open application programming interface (API) to develop a program to collect user information. Such information includes user name, location, number of updates, number of followers, number of people they are following (called friends in the API documentation), date of account creation, and a short description of himself or herself. Our data includes 3.61 million Twitter user profiles collected in 2009 and 2010.[6] For our research,

[6]The actual total number of Twitter users is never published by the company although some people estimate that it might be around 30 million in the middle of 2009.

we are mainly interested in the frequency with which new tweets or updates are posted by each user. To obtain this data, we divide the total number of updates by the total number of days since an account was created. We denote the derived variable by $UPDATERATE_i$, which we believe is a reasonable proxy of $T_i w_i$ in our model. This is because the length of a tweet is limited to 140 characters so that, on average, the time needed to write a tweet should not differ too much. However, the amount of content in each tweet depends on the capability of the producer (i.e., q_i). Some content producers post very informative or insightful tweets while some producers post mostly trivial or even spam tweets.

A Twitter user has the option of putting a link on her profile. This link is displayed on the user's Twitter homepage and can be clicked by visitors who might be interested in this particular user. We use a dummy variable $LINK$ to denote whether a Twitter user has a link or not. A Twitter user also has the option of writing a short biography which will be displayed on her Twitter homepage. The maximal number of characters allowed is 160. We capture this attribute by the normalized biography length BIO which is the total number of characters of the biography divided by 160. We also include the number of followers each user has and the number of friends each user has. [7] Table 1 summarizes the variables used in our empirical study.

Table 1: Summary of Variables

Name	Meaning	Min	Max	Mean
$UPDATERATE$	number of tweets posted per day	0	1009.9	0.67858
$LINK$	1 if the user has a link, 0 otherwise	0	1	0.25045
BIO	length of biography	0	1	0.14375
$FOLLOWERS$	number of followers	0	2986300	211.34
$FRIENDS$	number of users they are following	0	774450	190.03
$DAYS$	number of days since account creation	91	1421	264.59

[7] These two number are highly correlated because of the "following back" etiquette (i.e., a user often follows back another user who follows him).

5.2 Hypotheses

About 25% of the Twitter users put a link on their profile while the other 75% do not. These links are usually their personal blog sites or the organizations they represent. In other words, these are generally the sites they try to promote so that more people will visit. We believe that users with a link on their profile value attention more than those without a link on their profile because of users' self-selection based on how much they value attention. Those who highly value attention are more likely to put a link on their profile to actively seek attention in the form of clicks on the link. On the other hand, users who do not care as much about attention do not bother to set up a personal or organizational page and put that on their profile. This underlying self-selection process leads to the result that users with a link on their profile value attention more than those without a link on their profile. Similar arguments indicate that those users who write a lot in their short biography value attention more than those who write little or nothing for their biography.

Translated into our model, the above arguments suggest that those with link on their profile and those who write more in their biography have larger α values. On the other hand, Proposition 3 implies that users with larger α values will devote more time to content production in equilibrium if we assume α, q, and θ are distributed independently, hence, we have the following hypothesis.

Hypothesis 1. *(H1-a)Twitter users with a link on their profile have a higher* $UPDATERATE$ *than users without a link on their profile.*

(H1-b)Twitter users with larger BIO *have higher* $UPDATERATE$.

The number of followers a Twitter user has is an indicator of a user's capability of producing content because users tend to follow others whose tweets are interesting or useful to them. Analogous to the idea of PageRank, we would expect users with more followers to be more productive in the sense that they are more capable of producing content that others find interesting or useful. So we take $FOLLOWERS$ as a proxy of q in our model. Based on Proposition 3, and again by assuming that α, q, θ are distributed independently, we propose the following hypothesis.

Hypothesis 2. *Twitter users with higher $FOLLOWERS$ have higher $UPDATERATE$.*

We first do the logarithmic transformation on the dependent variable $UPDATERATE$ so that it could take negative values. Correspondingly, we take the logarithmic transformation on $FOLLOWERS$ and $FRIENDS$. We use the following regression model to test the Hypothesis 1 and Hypothesis 2.

$$\begin{aligned} \ln(UPDATERATE_i) = b_0 + b_1 \times LINK_i + b_2 \times BIO_i + b_3 \times \ln(FOLLOWERS_i) \\ + b_4 \times \ln(FRIENDS_i) + \epsilon_i. \end{aligned} \tag{8}$$

If our hypotheses are correct, the estimates of b_1, b_2 and b_3 from Equation (8) should all be positive and significant.

Our third hypothesis regards the distribution of $UPDATERATE$ among Twitter users. Assuming linearity of $\psi(\cdot)$, Proposition 2 suggests that as the community size becomes large enough, it will be partitioned into two groups: users in the bottom left region of the $\alpha - q$ plane become content consumers, users in the upper right region become content producers. So we would expect a large proportion of Twitter users to have very low $UPDATERATE$.

To derive the distribution of $UPDATERATE$, we again use the setup in Section 4.3 by allowing users to choose T_i. From Proposition 3, we know $T_i = \frac{T^C}{S}\theta_i\alpha_i q_i, i \in I_P$. Hence, producers with the same $\alpha q\theta$ value should spend the same amount of time producing content, resulting in the same $UPDATERATE$. We need assumptions on the distribution of (α, q, θ) to derive the distribution of $UPDATERATE$. For simplicity, we assume that the distribution of θ_i is independent of that of (α_i, q_i), and (α_i, q_i) is uniformly distributed in the region $(\underline{\alpha}, \overline{\alpha}) \times (\underline{q}, \overline{q})$. Hence, the number of content producers who spend time T, $T_2 < T < T_1$ producing content is proportional to the shadow area in Figure 2, which is computed as follows:

$$A(T_1, T_2) = \int_{T_1/\bar{q}}^{\bar{\alpha}} (\bar{q} - \frac{T_1}{x})dx - \int_{T_2/\bar{q}}^{\bar{\alpha}} (\bar{q} - \frac{T_2}{x})dx \tag{9}$$

$$= (T_2 - T_1)(1 + \ln\bar{\alpha} + \ln\bar{q}) + T_1 \ln T_1 - T_2 \ln T_2, \tag{10}$$

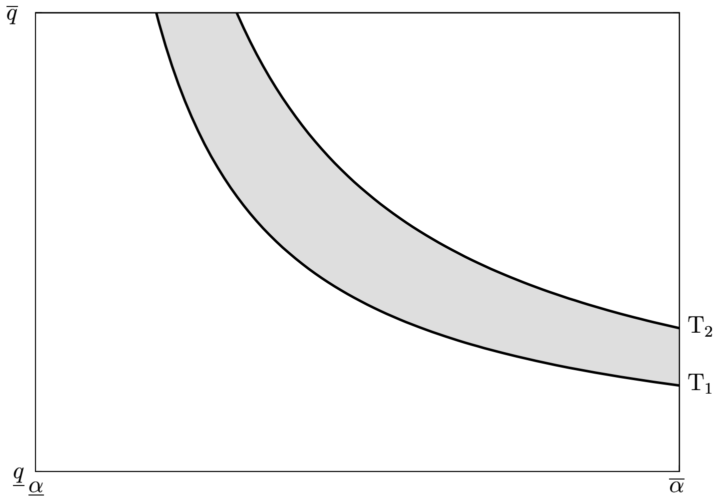

Figure 2: Hypothesis 3

$$\lim_{T_2 \to T_1} \frac{A(T_1, T_2)}{T_2 - T_1} = 1 + ln\bar{\alpha} + ln\bar{q} + \lim_{T_2 \to T_1} \frac{T_1 lnT_1 - T_2 lnT_1 + T_2 lnT_1 - T_2 lnT_2}{T_2 - T_1} \tag{11}$$

$$= ln\bar{\alpha} + ln\bar{q} - lnT_1. \tag{12}$$

So the following econometric model should characterize the distribution of $UPDATERATE$ among producers.

$$USERCOUNT_k = a_0 + a_1 \times ln(RATE_k), \tag{13}$$

where $USERCOUNT_k$ is the number of users whose $UPDATERATE$ is in the interval:

$$[RATE_k, RATE_{k+1}].$$

We use the following cutoff points to categorize all the Twitter users in our sample:

$$RATE_0 = 0, RATE_{k+1} = RATE_k + 0.01, k = 0, 1, \cdots, 499.$$

Hence, we have the following hypothesis.

Hypothesis 3. *The coefficient of $ln(RATE_k)$ in Equation (13) is negative and significant.*

Even though "Twitter resembles more of a one-way, one-to-many publishing service more than a two-way, peer-to-peer communication network" [8], Twitter's function as a communication tool among friends is still significant. This would suggest that there will be a lot of noise in the above regression model because the observed $UPDATERATE$ reflects both content produced for attention and content produced for communication. Such effect is particularly salient when $UPDATERATE$ is small. However, it is reasonable to assume that as $UPDATERATE$ becomes larger, the impact of such noise will diminish.

Based on the above analysis, we conjecture that Equation (13) should fit the data better when we exclude small values of $RATE_k$.

For ease of illustration, we denote the sample with all 500 points as $SAMPLE_0$, the subsample of $\{(USERCOUNT_k, RATE_{k+1}), k = 1, 2, \cdots, 499\}$ as $SAMPLE_1$, the subsample of $\{(USERCOUNT_k, RATE_{k+1}), k = 2, \cdots, 499\}$ as $SAMPLE_2$ and so on. We also denote R^2_m as the R-squared of the regression of Equation (13) with $SAMPLE_m$. We consider R-squared as a reasonable measure of how well Equation (13) fits the data. Hence, we have the following hypothesis.

[8]http://blogs.hbr.org/cs/2009/06/new_twitter_research_men_follo.html

Hypothesis 4. *R_m^2 is increasing in m.*

5.3 Results

We have collected 3,615,972 profiles of Twitter users who joined Twitter at least 91 days before we collected their profiles. We impose this 91-day restriction to get reasonably accurate estimate of $UPDATERATE$. For Hypothesis 1 and Hypothesis 2, we remove those who have 0 friends, or 0 followers, or 0 updates so that we can do the logarithmic transformation on the data. After this, there are 2,891,298 user profiles left for us to test (H1) and (H2). Table 2 shows the estimation results of Equation (8).

The coefficients of $LINK$, BIO, and $\ln(FOLLOWERS)$ are all significantly positive, which supports (H1) and (H2). We notice that $\ln(FRIENDS)$ is also significantly positive, although the effect is much weaker compared with other explanatory variables. This is reasonable since the more people a Twitter user is following, the more likely they will have conversation on Twitter, which leads to more updates.

Table 2: OLS Estimation Results of Hypothesis 1 & 2

Variable	coefficient	t-value
$LINK$	0.29195***	116.71
BIO	0.34199***	76.762
$\ln(FOLLOWERS)$	0.69024***	620.81
$\ln(FRIENDS)$	0.027822***	23.435
$Constant$	-4.7516***	1881.4
R^2	0.42842	-
Observations	2,891,298	-

Figure 3 shows the first 50 points of $SAMPLE_0$ in the form of a histogram. The horizontal axis is $UPDATERATE \times 100$ and the vertical axis is the number of users. So each bin represents the users with $UPDATERATE$ falling in the range of the bin. It is clear from Figure 3 that content consumers constitute a large proportion of the Twitter community.

The model estimation of Equation (13) using $SAMPLE_1$ is shown in Table 3, which

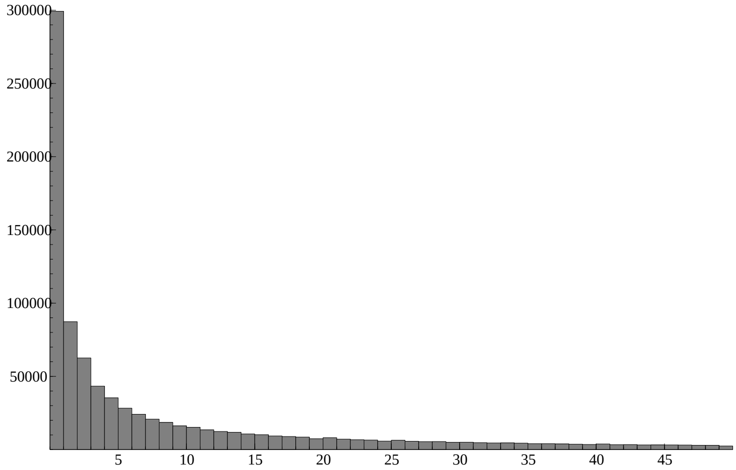

Figure 3: Histogram of Updates By Sample Twitter User

clearly supports Hypothesis 3. The R-squared is reasonable. [9]

Table 3: OLS Estimation Results of Hypothesis 3

Variable	coefficient	t-value
$\ln(UPDATERATE)$	-4224.4***	6.0098
Constant	4300.0***	4.9947
R^2	0.473	-
Observations	499	-

Finally, we use the same sample of Twitter users to construct $SAMPLE_k, k = 2, 3, \cdots, 100$ and estimate Equation (13). As a comparison, we also use 100 randomly generated subsamples from $SAMPLE_1$ with size from 400 to 499 to run regression. Figure 4 shows how the R-squared changes as we use different subsamples to estimate.

As we can see, for the randomly selected subsamples, there is no clear relationship between the R-squared and the size. However, for the R-squared of $SAMPLE_k, k = 0, 1, \cdots, 99$, the trend is very clear. First, by switching from $SAMPLE_0$ to $SAMPLE_1$, the R-squared increases from 0.266 to 0.473, which results from the fact that the content consumers are excluded from the samples. As m keeps increasing, the R-squared also increases, which

[9]If we use $SAMPLE_m$ with larger m, the R-squared increases significantly and reaches values higher than 0.9.

supports Hypothesis 4.

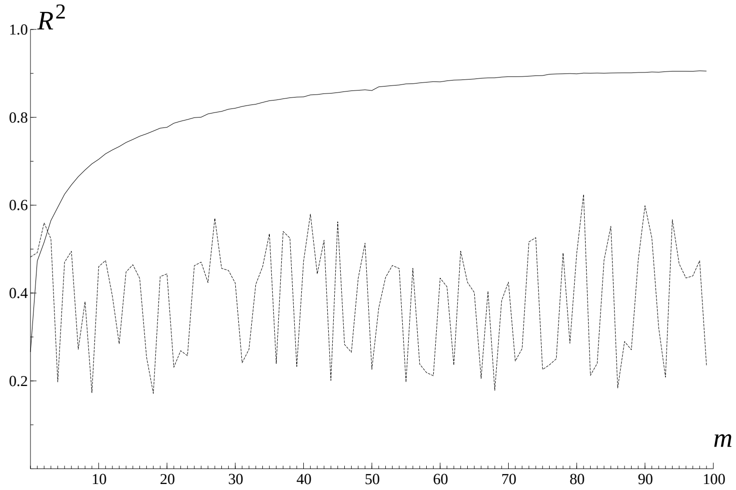

Figure 4: R^2 of $SAMPLE_m$

6 Conclusion and Limitations

What is the key innovation of social media services like Twitter? Inspired by the famous saying by Herbert Alexander Simon on the relationship between information and attention, we conceptualize a social media environment as an economy where people supply content to get attention and "purchase" content through the supply of attention. Although the underlying mechanisms of how users obtain utility from attention might vary from economic incentives to psychological and sociological motivations (e.g., status seeking, social connection), we treat them abstractly as users' taste for attention. To further study the feature of this attention economy and better understand the phenomena observed in many social media sites, we developed a game-theoretical model and studied the interaction among the players in the economy with the additional assumption of large population.

We find that there exists a community wage for contributing content and users have two individual reservation wages: one is the reservation wage for becoming a producer (the high reservation wage) and the other is the reservation wage for becoming a consumer (the low reservation wage). If the community wage is larger than the high reservation wage, the user becomes a content producer. If the community wage is less than the low reservation wage, the user becomes a content consumer. If the community wage falls between these two reservation wages, the user becomes a prosumer. The proportion of prosumers in the community crucially depends on how concave the consumption part of the utility function $\psi(\cdot)$ is. The more concave $\psi(\cdot)$ is (i.e., the less marginal utility a user will get if she spends more time consuming content), the larger the prosumers' proportion is. In the extreme case when $\psi(\cdot)$ is linear, there are no prosumers in the community almost surely: users become either content producers or content consumers. This partition equilibrium outcome is both theoretically interesting and practically suggestive. We have shown that under the assumption of a partition equilibrium, the system characterizing the macro-level content consumption and production is stable in the sense that there is an asymptotically stable equilibrium point involving massive content production and consumption. This result gives strong support to the sustainability of social media.

To bring theory to practice, we collected from Twitter nearly 3 million user profiles that contains information about each user's frequency of contributing content and each user's individual characteristics. The four hypotheses we proposed based on our theoretical model are all strongly supported by the empirical study.

Our theoretical model and empirical study suggest that the unique innovation of social media is recognizing and connecting people's need for information and attention. A practical implication of this is that social media services should be designed in a way that facilitate its role as a marketing place that connects people's need for information and attention. For example, a decentralized content structure where content is accessed at authors' individual (and possibly customized) directory is an effective way of boosting the market for exchange of content and attention.

Although we believe our work has provided important insights into our understanding of social media and user-generated content in general, there are limitations in both the theoretical modeling and the empirical study. First, we haven't fully characterized the equilibrium in the general case when $\psi(\cdot)$ is nonlinear, although we point out the two limit conditions in Proposition 1. The model becomes intractable when we try to solve analytically each prosumer's optimal w. One way to extend our model is to run computer simulation to further explore properties of the equilibrium. In our model, we have a very simple treatment of content measurement and competition among content suppliers. It is interesting to study the content competition in the social media environment with a more comprehensive model similar to the one in Mullainathan and Shleifer's model (2005). Second, our empirical study is also limited because of the fact that both α and q are unobservable and we have to find proxy variables for them. In our current dataset, both $LINK$ and BIO are good proxy variables for α and $FOLLOWERS$ is also a reasonable proxy variable for q. However, we still need concrete theory as the basis of our choice of proxy variables for the latent variables. For example, relevant theories and experiments from psychology or sociology might be particularly useful. Along this line, it is an obvious potential research opportunity to find a better dataset from Twitter or some other social media sites. Another future research direction is to extend this paper to incorporate dynamic elements, both theoretically and empirically. Users may start out with no information at all about an online community. How do they adjust their behavior dynamically as they learn more about the online community? It would be interesting to capture this evolution through modeling, as well as through empirical studying (e.g., by constructing time-series data of user contributions).

Social media supported by user-generated content is now pervasive on the Internet. The popularity of Twitter and YouTube offers a fascinating glimpse into the economic and societal impact of social media. This paper opens up an important avenue for future research on the innovation and evolution of social media.

7 Appendix: Proofs

Proof of Lemma 1.

Proof. Define:

$$f_i(w_i) = \frac{du_i}{dw_i} = -T_i\psi'(T_i(1-w_i))\phi(S_{-i}) + \alpha_i q_i T_i \sum_{j\neq i} \frac{S_{-ij}}{(S_{-ij}+q_iT_iw_i)^2} T_j r_j.$$

It follows that $\frac{df_i(w_i)}{dw_i} < 0$. We need to examine the incentive constraints of the three groups in equilibrium.

1) Content consumers: I_C

For user $i \in I_C$, the necessary and sufficient condition for choosing $w_i = 0$ is:

$$f_i(0) = -T_i\psi'(T_i)\phi(S_{-i}) + \alpha_i q_i T_i \sum_{j\neq i} \frac{T_j r_j}{S_{-ij}} \leq 0. \tag{14}$$

Since $\frac{r_j}{S_{-ij}} = 0$ if $j \in I_P$ and $\frac{r_j}{S_{-ij}} = \frac{1}{S_{-i}}$ if $j \in I_C$, $\sum_{j\neq i} \frac{T_j r_j}{S_{-ij}} = \frac{1}{S_{-i}} \sum_{j\in I_C, j\neq i} T_j + \sum_{j\in I_M, j\neq i} \frac{T_j r_j}{S_{-ij}}$. Inequality 14 becomes to

$$\alpha_i q_i \leq \frac{S_{-i}}{T^C_{-i} + S_{-i}\sum_{j\in I_M, j\neq i}\frac{T_j r_j}{S_{-ij}}} \psi'(T_i)\phi(S_{-i}) = h_C(i)$$

2) Content producers: I_P

For user $i \in I_P$, the necessary and sufficient condition for choosing $w_i = 1$ is:

$$f_i(1) = -T_i\psi'(0)\phi(S_{-i}) + \alpha_i q_i T_i \sum_{j\neq i} \frac{S_{-ij}}{(S_{-ij}+q_iT_i)^2} T_j r_j \geq 0 \tag{15}$$

$$\sum_{j\neq i}\frac{S_{-ij}}{(S_{-ij}+q_iT_i)^2}T_jr_j=\sum_{j\in I_C,j\neq i}\frac{S_{-ij}}{(S_{-ij}+q_iT_i)^2}T_j+\sum_{j\in I_M,j\neq i}\frac{S_{-ij}}{(S_{-ij}+q_iT_i)^2}T_jr_j$$
$$=\frac{S_{-i}}{(S_{-i}+q_iT_i)^2}T^C_{-i}+\sum_{j\in I_M,j\neq i}\frac{S_{-ij}}{(S_{-ij}+q_iT_i)^2}T_jr_j$$

Inequality (15) becomes:

$$\alpha_iq_i\geq\frac{(S_{-i}+q_iT_i)^2}{S_{-i}T^C_{-i}+(S_{-i}+q_iT_i)^2\sum_{j\in I_M,j\neq i}\frac{S_{-ij}}{(S_{-ij}+q_iT_i)^2}T_jr_j}\psi'(0)\phi(S_{-i})=h_P(i)$$

Apparently, a user with $h_C(i)<\alpha_iq_i<h_P(i)$ will choose $0<w_i<1$. □

Proof of Lemma 2.

Proof. First, we notice that in equilibrium, $\lim_{n\to\infty}h_C(i)\nrightarrow\infty$ and $\lim_{n\to\infty}h_P(i)\nrightarrow 0$. From that, the fact that $\lim_{n\to\infty}S=\infty$ is straightfoward. We now prove the other two limit conditions in (1) and (2) respectively.

(1) If $n_M/n\nrightarrow 0$ as $n\to\infty$, then we immediately have $n_M\to\infty$ as $n\to\infty$. Pick $1>\delta>0,\underline{w}>0$ such that $\forall n$, at least $n_M\delta$ of those $i\in I_M$ choose $w_i>\underline{w}$.[10]

$$S_{-i}=\sum_{j\neq i}q_jT_jw_j\geq\sum_{j\in I_M,j\neq i}q_jT_jw_j\geq(n_M\delta-1)\underline{qwT}$$

$$\frac{T^C_{-i}}{S^\beta_{-i}}<\frac{\overline{T}n}{\underline{q}^\beta\underline{w}^\beta\underline{T}^\beta(n_M\delta-1)^\beta}\to 0\text{ as }n\to 0$$

If $n_M/n\to 0$ as $n\to\infty$, then we must have $T^C_{-i}\to\infty$ because otherwise $h_C(i)\to\infty$. Furthermore, $\frac{T^C_{-i}}{S_{-i}}\nrightarrow\infty$ as $n\to\infty$ because otherwise $h_P(i)\to 0$ too. Therefore, $\lim_{n\to\infty}\frac{T^C_{-i}}{S^\beta_{-i}}=0,\forall\beta>1$.

[10] Technically, it is possible that such $(\delta,\underline{w})$ does not exist. In this case, $\frac{n_M}{n}\nrightarrow 0$, but $w_i\to 0,\forall i\in I_M$. This is obviously a very unrealistic situation. Hence, we exclude this pathological case in the paper.

(2) If $n_M/n \nrightarrow 0$, then:

$$\sum_{j\in I_M, j\neq i} \frac{T_j r_j}{S_{-ij}^{\beta}} < \frac{n_M \overline{T}}{(n_M\delta - 2)^{\beta} \underline{w}^{\beta} \underline{q}^{\beta} \underline{T}^{\beta}} \rightarrow 0 \text{ as } n \rightarrow 0,$$

where $\delta, \underline{w}$ are defined in previously.

If $n_M/n \rightarrow 0$, then $n_C/n \nrightarrow 0$ because otherwise $\frac{T_{-i}^C}{S_{-i}} \rightarrow 0$, $h_C(i) \rightarrow \infty$. So we must have $\sum_{j\in I_M, j\neq i} T_j r_j < T_{-i}^C$. Now:

$$\sum_{j\in I_M, j\neq i} \frac{T_j r_j}{S_{-ij}^{\beta}} < \frac{S_{-i}^{\beta}}{(S_{-i} - \overline{q}\overline{T})^{\beta}} \frac{T_{-i}^C}{S_{-i}^{\beta}} \rightarrow 0 \text{ as } n \rightarrow 0.$$

□

Proof of Proposition 1.

Proof. We prove the proposition in three steps.

1) First we show

$$\lim_{n\rightarrow\infty} \frac{h_C(i)}{h_P(i)} = \frac{\psi'(T_i)}{\psi'(0)}.$$

Using the expressions of $h_C(i)$ and $h_P(i)$ in Lemma 1, we have:

$$\frac{h_C(i)}{h_P(i)} = \frac{\frac{S_{-i} T_{-i}^C}{(S_{-i}+q_i T_i)^2} + \sum_{j\in I_M, j\neq i} \frac{S_{-ij}}{(S_{-ij}+q_i T_i)^2} T_j r_j}{\frac{T_{-i}^C}{S_{-i}} + \sum_{j\in I_M, j\neq i} \frac{T_j r_j}{S_{-ij}}} \frac{\psi'(T_i)}{\psi'(0)}.$$

Then

$$\frac{T_{-i}^C}{S_{-i}} - \frac{S_{-i} T_{-i}^C}{(S_{-i}+q_i T_i)^2} = T_{-i}^C \frac{2S_{-i} + q_i T_i}{S_{-i}(S_{-i}+q_i T_i)^2} q_i T_i < 2 q_i T_i \frac{T_{-i}^C}{S_{-i}^2} \rightarrow 0 \text{ as } n \rightarrow 0,$$

and

$$\sum_{j\in I_M, j\neq i} \frac{T_j r_j}{S_{-ij}} - \sum_{j\in I_M, j\neq i} \frac{S_{-ij}}{(S_{-i}+q_i T_i)^2} T_j r_j = \sum_{j\in I_M, j\neq i} T_j r_j \frac{q_i T_i (2S_{-ij} + q_i T_i)}{S_{-ij}(S_{-ij}+q_i T_i)^2}$$
$$< 2q_i T_i \sum_{j\in I_M, j\neq i} \frac{T_j r_j}{S_{-ij}(S_{-ij}+q_i T_i)}$$
$$< 2q_i T_i \sum_{j\in I_M, j\neq i} \frac{T_j r_j}{S_{-ij}^2} \to 0 \text{ as } n \to \infty.$$

Because $h_C(i) \nrightarrow \infty$ in equilibrium, $\frac{T_{-i}^C}{S_{-i}} + \sum_{j\in I_M, j\neq i} \frac{T_j r_j}{S_{-ij}} \nrightarrow 0$ as $n \to \infty$. So

$$\lim_{n\to\infty} \frac{h_C(i)}{h_P(i)} = \lim_{n\to\infty} h(i) = \frac{\psi'(T_i)}{\psi'(0)}. \tag{16}$$

2) Second, we show

$$\lim_{n\to\infty} \frac{h_C(i)}{h_C(k)} = \frac{\psi'(T_i)}{\psi'(T_k)}, \forall i, k,$$

$$\left| \sum_{j\in I_m, j\neq i} \frac{T_j r_j}{S_{-ij}} - \sum_{j\in I_m, j\neq k} \frac{T_j r_j}{S_{-kj}} \right| = \sum_{j\in I_M, j\neq i,k} \left| \frac{q_k T_k w_k - q_i T_i w_i}{S_{-ij} S_{-kj}} \right| T_j r_j + \frac{|T_k r_k - T_i r_i|}{S_{-ik}}$$
$$\leq |q_k T_k w_k - q_i T_i w_i| \max\Big(\sum_{j\in I_M, j\neq i} \frac{T_j r_j}{S_{-ij}^2}, \sum_{j\in I_M, j\neq k} \frac{T_j r_j}{S_{-kj}^2}\Big) + \frac{|T_k r_k - T_i r_i|}{S_{-ik}} \to 0 \text{ as } n \to \infty$$

It's easy to show that $\frac{T_{-i}^C}{S_{-i}} - \frac{T_{-k}^C}{S_{-k}} \to 0$ as $n \to \infty$.
Since $h_C(i) \nrightarrow \infty$, $\frac{T_{-i}^C}{S_{-i}} + \sum_{j\in I_m, j\neq i} \frac{T_j r_j}{S_{-ij}} \nrightarrow 0$. Hence:

$$\frac{h_C(i)}{h_C(k)} = \frac{\frac{T_{-k}^C}{S_{-k}} + \sum_{j\in I_m, j\neq k} \frac{T_j r_j}{S_{-kj}}}{\frac{T_{-i}^C}{S_{-i}} + \sum_{j\in I_m, j\neq i} \frac{T_j r_j}{S_{-ij}}} \frac{\psi'(T_i)\phi(S_{-i})}{\psi'(T_k)\phi(S_{-k})} \to \frac{\psi'(T_i)}{\psi'(T_k)}.$$

3) From 1) and 2), we immediately have

$$\lim_{n\to\infty}\frac{h_P(i)}{h_P(k)}=\lim_{n\to\infty}\frac{h_C(i)}{h_C(k)}\frac{h_C(k)/h_P(k)}{h_C(i)/h_P(i)}=\frac{\psi'(T_i)}{\psi'(T_k)}\frac{\psi'(T_k)/\psi'(0)}{\psi'(T_i)/\psi'(0)}=1,\forall i,k.$$

Since $h_P(k)\nrightarrow 0,\forall k$, $h_P(i)-h_P(k)\to 0,\forall i,k$. Denote $\lim_{n\to\infty}h_P(i)=h_P$, then

$$\lim_{n\to\infty}h_C(i)=\frac{\psi'(T_i)}{\psi'(0)}h_P.$$

□

Proof of Corollary 1.

Proof. By Proposition 1, we need to examine only the case when w_i and w_k are the interior solutions of the first-order conditions of user i and user k's utility maximization problem. From the proof of Lemma (1), we have:

$$\psi'(T_i(1-w_i))\phi(S_{-i})=\alpha_i q_i\sum_{j\neq i}\frac{S_{-ij}}{(S_{-ij}+q_iT_iw_i)^2}T_jr_j$$

$$\psi'(T_k(1-w_k))\phi(S_{-k})=\alpha_k q_k\sum_{j\neq k}\frac{S_{-kj}}{(S_{-kj}+q_kT_iw_k)^2}T_jr_j$$

$$\frac{\psi'(T_i(1-w_i))}{\psi'(T_k(1-w_k))}=\frac{\phi(S_{-k})}{\phi(S_{-i})}\frac{\alpha_i}{\alpha_k}\frac{\sum_{j\neq i}\frac{S_{-ij}q_i}{(S_{-ij}+q_iT_iw_i)^2}T_jr_j}{\sum_{j\neq k}\frac{S_{-kj}q_k}{(S_{-kj}+q_kT_iw_k)^2}T_jr_j}$$

(1) If $T_i=T_k,q_i=q_k,\alpha_i\geq\alpha_k$, then, using Lemma 2, one can show that:

$$\lim_{n\to\infty}\frac{\sum_{j\neq i}\frac{S_{-ij}q_i}{(S_{-ij}+q_iT_iw_i)^2}T_jr_j}{\sum_{j\neq k}\frac{S_{-kj}q_k}{(S_{-kj}+q_kT_iw_k)^2}T_jr_j}=1.$$

Hence,

$$\lim_{n\to\infty}\frac{\psi'(T_i(1-w_i))}{\psi'(T_k(1-w_k))}=\frac{\alpha_i}{\alpha_k}\geq 1$$

which implies that $w_i \geq w_k$.

(2) If $T_i = T_k, \alpha_i = \alpha_k, q_i \geq q_k$, then using Lemma 2, one can show that:

$$\lim_{n\to\infty} \frac{\sum_{j\neq i} \frac{S_{-ij} q_i}{(S_{-ij}+q_i T_i w_i)^2} T_j r_j}{\sum_{j\neq k} \frac{S_{-kj} q_k}{(S_{-kj}+q_k T_i w_k)^2} T_j r_j} = \frac{q_i}{q_k}.$$

Hence,

$$\lim_{n\to\infty} \frac{\psi'(T_i(1-w_i))}{\psi'(T_k(1-w_k))} = \frac{q_i}{q_k} \geq 1,$$

which implies that $w_i \geq w_k$. □

Proof of Proposition 2.

Proof. The "if" part is clear from Proposition 1 and is explained in the paper. The "only if" part could be similarly proved. Suppose $\lim_{n\to\infty} \frac{n_M}{n} = 0$; then we must have $\lim_{n\to\infty} \frac{n_P}{n} > 0$ because otherwise Lemma 2 will be violated. Hence,

$$\lim_{n\to\infty} \sum_{j\in I_m} \frac{T_j r_j}{S_{-ij}} = 0, \quad \lim_{n\to\infty} \sum_{j\in I_M, j\neq i} \frac{S_{-ij}}{(S_{-ij}+q_i T_i)^2} T_j r_j = 0$$

which implies that

$$\lim_{n\to\infty} h_C(i) = \frac{S}{T^C} \psi'(T_i) \lim_{n\to\infty} \phi(S), \quad \lim_{n\to\infty} h_P(i) = \frac{S}{T^C} \psi'(0) \lim_{n\to\infty} \phi(S).$$

With continuous distribution of (α, q, T), if $\psi''(\cdot) < 0$, then there is always a positive proportion of users who have $h_C(i) < \alpha_i q_i < h_P(i)$ as $n \to \infty$, and those users choose $r_i > 0, w_i > 0$. Hence $\frac{n_M}{n} \nrightarrow 0$, which is a contradiction. □

Proof of Corollary 2.

Proof. In a partition equilibrium, the utility function simplifies to:

$$u_i = \psi(T_i(1-w_i))\phi(S_{-i}) + \alpha_i q_i T_i w_i \frac{T_C}{S}.$$

since $r_j = 1, w_j = 0$ if $j \in I_C$ and $r_j = 0, w_j = 1$ if $j \in I_P$.

If $i \in I_C$, then $w_i = 0$ and $S_{-i} = S$ which implies $u_i^C = \psi(T_i)\phi(S)$. On the other hand, if $i \in I_P$, $u_i^P = \alpha_i q_i T_i \frac{T_C}{S}$.

From Proposition 1, we know in a partition equilibrium that

$$\lim_{n\to\infty} h_C(i) = \lim_{n\to\infty} h_P(i) = \frac{S}{T^C}\psi'(0) \lim_{S\to\infty} \phi(S) = \frac{S}{T^C} = h$$

and that $S = \sum_{\alpha_i q_i > h} q_i T_i$, $T^C = \sum_{\alpha_i q_i < h} T_i$; thus, h is determined by:

$$h = \frac{\sum_{\alpha_i q_i > h} q_i T_i}{\sum_{\alpha_i q_i < h} T_i}, \tag{17}$$

which always has a solution in $(0, \overline{\alpha q})$. ☐

Proof of Corollary 3.

Proof. The original threshold h is determined by:

$$h = \frac{\sum_{\alpha_i q_i > h} q_i T_i}{\sum_{\alpha_i q_i < h} T_i}. \tag{18}$$

Denote h' as the threshold after the shift and $\hat{h}$ as the threshold that keeps the same group of people content consumers/producers (i.e., $\sum_{\alpha_i' q_i < \hat{h}} T_i = \sum_{\alpha_i q_i < h} T_i$ or $\sum_{\alpha_i q_i' < \hat{h}} T_i = \sum_{\alpha_i q_i < h} T_i$).

If the population shifts up in α, then:

$$h' = \frac{\sum_{\alpha_i' q_i > h'} q_i T_i}{\sum_{\alpha_i' q_i < h'} T_i} \tag{19}$$

and

$$h = \frac{\sum_{\alpha_i' q_i > \hat{h}} q_i T_i}{\sum_{\alpha_i' q_i < \hat{h}} T_i}. \tag{20}$$

We must have $h' \geq h$ because otherwise the RHS of (19) is greater than (18), which leads to contradiction. From this result, we must have $h' < \hat{h}$ because otherwise the RHS of (19) is smaller than (20), which also leads to contradiction. This result implies that more users will become content producers and that more content will be generated.

If the population shifts up in q, then similarly we would have $h' \geq h$. Now suppose $S' = \sum_{\alpha_i q_i' > h'} < S = \sum_{\alpha_i q_i > h}$ (i.e., $h' \sum_{\alpha_i q_i' < h'} T_i < h \sum_{\alpha_i q_i < h} T_i$), then $\sum_{\alpha_i q_i' < h'} T_i < \sum_{\alpha_i q_i < h} T_i$, which implies $h' < h$, which is a contradiction. Hence, we must have $S' > S$ (i.e., more content will be generated after an upward shift of q). □

Proof of Proposition 3.

Proof. Given users' choice of T_i, $i = 1, 2, \cdots, n$, the partition equilibrium will be played. Denote user i's utility by u_i^C if he is a content consumer and u_i^P if he is a content producer. From Corollary 2 and our assumption of a quadratic cost function, we have:

$$u_i^C = T_i \phi(S) \tau - \frac{1}{2\theta_i} T_i^2, \; u_i^P = \frac{T_C}{S} \alpha_i q_i T_i - \frac{1}{2\theta_i} T_i^2,$$

where $S = \sum_{i \in I_P} q_i T_i$.

In the equilibrium, each user's choice of T_i maximizes her utility, so we have

$$\begin{cases} T_i^C = \phi(S) \tau \theta_i, & i \in I_C \\ T_i^P = \frac{T_C}{S} \alpha_i q_i \theta_i, & i \in I_P. \end{cases} \tag{21}$$

Summing up over $i \in I_C$ for the first equation, we get $T^C = \sum_{i \in I_C} \phi(S) \tau \theta_i$. Multiplying by q_i on both sides of the second equation and summing up over $i \in I_P$, we get $S = \sum_{i \in I_P} \frac{T^C}{S} \alpha_i q_i^2 \theta_i$. By Proposition 1, $i \in I_C$ if $\alpha_i q_i < \frac{S}{T^C}$ and $i \in I_P$ if $\alpha_i q_i > \frac{S}{T^C}$. So (S, T_C) is the solution to Equation (6). □

Lemma 3. *Under the assumptions of $\phi(0) = 0$, $\phi'(S) > 0$, and $\phi''(S) < 0$,*

$$\beta\phi(S) > \phi'(S)S, \forall S > 0, \beta \geq 1$$

Proof. Denote $f(S) = \beta\phi(S) - \phi'(S)S$, $f'(S) = \beta\phi'(S) - \phi'(S) - S\phi''(S) = (\beta - 1)\phi'(S) - S\phi''(S) > 0, \forall S > 0, \beta \geq 1$. Because $f(0) = 0$, $f(S) > 0, \forall S > 0$. □

Notes on the definition of equilibrium point:

A vector $\bar{\mathbf{x}}$ is an equilibrium point for a time-invariant dynamic system $\mathbf{x}(t+1) = \mathbf{f}(\mathbf{x}(t))$ if once the state vector is equal to $\bar{\mathbf{x}}$ it remains equal to $\bar{\mathbf{x}}$ for all future time (i.e., $\bar{\mathbf{x}} = \mathbf{f}(\bar{\mathbf{x}})$). An equilibrium point $\bar{\mathbf{x}}$ is stable if there is an $R_0 > 0$ for which the following is true: For every $R < R_0$, there is an r, $0 < r < R$, such that if $\mathbf{x}(0)$ is inside the spherical region $S(\bar{\mathbf{x}}, r)$, then $\mathbf{x}(t)$ is inside $S(\bar{\mathbf{x}}, R)$ for all $t > 0$. An equilibrium point $\bar{\mathbf{x}}$ is asymptotically stable whenever it is stable and there is an $\bar{R}_0 > 0$ such that whenever the state is initiated inside $S(\bar{\mathbf{x}}, \bar{R}_0)$, it tends to $\bar{\mathbf{x}}$ as time increases.

Proof of Proposition 4.

Proof. From the proof of Proposition 3 we know:

$$\begin{cases} T^C = \sum_{i \in I_C} \phi(S)\tau\theta_i \\ S = \sum_{i \in I_P} \frac{T^C}{S}\alpha_i q_i^2 \theta_i \end{cases} . \tag{22}$$

Denote:

$$\begin{cases} k_1 = \sum_{i \in I_C} \tau\theta_i \\ k_2 = \sqrt{\sum_{i \in I_P} \alpha_i q_i^2 \theta_i} \end{cases} .$$

Then we obtain Equations (7), which characterizes the content consumption and production at the macro-level. From Equations(7), we have $S^{*2} = k_1 k_2^2 \phi(S^*)$. Although,

$(S^*, T^{C*}) = (0,0)$ is always a solution, there is at least one solution $(S^*, T^{C*}) \neq (0,0)$ since $\phi'(0) > 0$, $\lim_{x\to\infty} \phi(x) < \infty$.

The proof of asymptotic stability of this equilibrium point is a simple application of Liapunov's indirect method:

$$\mathbf{G} = \begin{pmatrix} \frac{\partial g_1}{\partial T^C} & \frac{\partial g_1}{\partial S} \\ \frac{\partial g_1}{\partial T^C} & \frac{\partial g_1}{\partial S} \end{pmatrix} = \begin{pmatrix} 0 & k_1\phi'(S) \\ \frac{k_2}{2\sqrt{T^C}} & 0 \end{pmatrix} \tag{23}$$

The eigenvalues of $\mathbf{G}$ satisfy $|\lambda|^2 = \frac{k_1 k_2 \phi'(S)\tau}{2\sqrt{T^C}}$. We need to determine whether $|\lambda| < 1$ at the equilibrium point (T^{C*}, S^*):

$$|\lambda| < 1 \iff k_1 k_2 \phi'(S^*) < 2\frac{S^*}{k_2}, \tag{24}$$

$$\iff k_1 k_2^2 \phi'(S^*) S^* < 2S^{*2} = 2k_1 k_2^2 \phi(S^*), \tag{25}$$

$$\iff \phi'(S^*) S^* < 2\phi(S^*). \tag{26}$$

The last inequality is ensured by Lemma 3 □

References

Andreoni, James. 1989. "Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence." *Journal of Political Economy*, 97(6): 1447-1458.

Bock, Gee-Woo, Robert W. Zmud, Young-Gul Kim, and Jae-Nam Lee. 2005. "Behavioral Intention Formation in Knowledge Sharing: Examining the Roles of Extrinsic Motivators, Social-psychological Forces, and Organizational Climate." *MIS Quarterly*, 29: 87-111.

Feldman, Michal, Christos Papadimitriou, John Chuang, and Ion Stoica. 2006. "Free-Riding and Whitewashing in Peer-to-Peer Systems." *IEEE Journal on Selected Areas in Communications*, 24: 1010-1019.

Guo, Lei, Enhua Tan, Songqing Chen, Xiaodong Zhang and Yihong Zhang, 2009. "Analyzing Patterns of User Content Generation in Online Social Networks." *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*: 369-378.

Huberman, Bernardo A., Daniel M. Romero and Fang Wu. 2008. "Crowdsourcing, Attention and Productivity." *Working Paper*.

Jeppesen, Lars Bo and Lars Frederiksen. 2006. "Why Do Users Contribute to Firm-hosted User Communities? The case of Computer-controlled Music Instruments." *Organization Science*, 17(1): 45-63.

Kandori, Michihiro. 1962. "Socila Norms and Community Enforcement." *Review of Economic Studies*, 59(1): 63-80.

Katona, Zsolt and Miklos Sarvary. 2008. "Network Formation and the Structure of the Commercial World Wide Web." *Marketing Science*, 27(5): 764-778.

Lerner, Josh, Parag A. Pathak and Jean Tirole. 2006. "The Dynamics of Open-Source Contributors." *American Economic Review*, 96(2): 114-118.

Mullainathan, Sendhi, and Andrei Shleifer. 2005. "The Market for News." *American Economic Review*, 95(4): 1031-1053.

Robert, Jeffrey A., II-Horn Hann, and Sandra A. Slaughter. 2006. "Understanding the Motivations, Participation and Performance of Open Source Software Developers: a Longitudinal Study of the Apache Projects." *Management Science*, 52(7): 984-999.

Shah, Sonali K. 2006. "Motivation, Governance and the Viability of Hybrid Forms in Open Source Software Development." *Management Science*, 52(7): 1000-1014.

Simon, Herbert. 1971: Designing Organizations for an Information-Rich World. The Johns Hopkins Press.

Sohn, Dongyoung and John D. Leckenby. 2007. "A Structural Solution to Communication Dilemmas in a Virtual Community." *Journal of Communication*, 57(3): 435-449.

Wasko, McLure and Samer Faraj. 2005. "Why Should I Share? Examining Knowledge Contribution in Electronic Networks of Practice." *MIS Quarterly*, 29: 1-23.