Evolution of Two-Sided Markets

Ravi Kumar
Yahoo! Research
701 First Avenue
Sunnyvale, CA 94089.
ravikumar@yahoo-inc.com

Yury Lifshits
Yahoo! Research
701 First Avenue
Sunnyvale, CA 94089.
lifshits@yahoo-inc.com

Andrew Tomkins*
Google, Inc.
1600 Amphitheater Parkway
Mountain View, CA 94043.
atomkins@gmail.com

ABSTRACT

Two-sided markets arise when two different types of users may realize gains by interacting with one another through one or more platforms or mediators. We initiate a study of the evolution of such markets. We present an empirical analysis of the value accruing to members of each side of the market, based on the presence of the other side. We codify the range of value curves into a general theoretical model, characterize the equilibrium states of two-sided markets in our model, and prove that each platform will converge to one of these equilibria. We give some early experimental results of the stability of two-sided markets, and close with a theoretical treatment of the formation of different kinds of coalitions in such markets.

Categories and Subject Descriptors. H.3.m [**Information Storage and Retrieval**]: Miscellaneous

General Terms. Economics, Experimentation, Theory

Keywords. Two-sided markets, equilibrium, preferential attachment, coalitions

1. INTRODUCTION

When we consider network effects, we typically think of scenarios like adoption of the telephone or fax machine, or an instant messaging client of a particular provider, or a social networking site like Facebook or LinkedIn, in which the benefit to each new user grows as a function of the number of existing users on the platform.

In this paper, we consider instead network effects in two-sided markets. *Two-sided markets* arise when two different types of users may realize gains by interacting with one another through one or more platforms or mediators. Video game platforms such as Sony PlayStation and Microsoft Xbox compete with one another to attract users and game developers. Users prefer platforms with many games,

*This work was done while the author was at Yahoo! Research.

while developers prefer platforms with many users. Similarly, males prefer to join online dating services with many appropriate females, and vice versa. Auction sites connect buyers with sellers; the buyers select a platform that offers many products, while the sellers select a platform with many buyers. Websites for jobs, travel, professional video, real estate, automobiles, shopping, classified advertising, and so forth, are all two-sided markets. Websites for photography, user-generated video, and online question answering arguably also fall into this category as most users operate as consumers, while a few users operate as providers (of photos, videos, or answers). There is a network effect, but the value of a platform to a player of type A depends on the players of type B the platform has managed to attract. We refer to this as a *cross-side network effect*, to distinguish it from the *same-side network effect* that occurs with fax machines or instant messaging clients.

Some two-sided markets may display negative same-side network effects (buyers prefer there to be few other buyers in the market, to keep costs low; job seekers prefer there to be few other job seekers in the market to keep competition low), or may display positive same-side network effects (a gaming platform may offer multi-player online games in which a user derives value from the presence of other users on the platform). However, in this paper we focus on systems that are characterized primarily by their cross-side network effects.

A body of literature in the field of Economics has grown up around two-sided markets focused largely on questions of pricing; see [13] for an overview. In this paper, we consider characterizations of the cross-side network effects in a two-sided market, and present a general model for such markets with a characterization of their stable equilibria.

Our contributions. We begin with an empirical study of single platforms in the space of questions and answers. First, we consider a set of 0.25M questions about computer programming from Stack Overflow (`stackoverflow.com`) over a one-year period. Next, we consider a set of 23M questions over 90 weeks drawn from Yahoo! Answers, a popular general-purpose question and answer platform. We study the rate of arrival of new question-askers (questioners) as a function of the number of current answer-providers (answerers) in the system (cross-side network effect), and as a function of the number of questioners in the system (same-side network effect), and then consider also the dual process showing the arrival of new answerers into the system. We show that different sides of the market can display radically different patterns, and that one side of the market can change its growth behavior over the life of the system.

Based on these findings, we postulate a model for two-sided markets. Consider the following simple example to give a flavor of the models we will consider. A set of platforms connect users to publishers. Each user and each publisher joins exactly one of the platforms (our models will not in general require this). During each round, a new user arrives and selects a platform with probability equal to the fraction of publishers who have joined that platform. Then a new publisher joins, and selects a platform based on the fraction of users at that platform. In this simple system, the joining decision of a user is made with direct proportionality to the market share of publishers enjoyed by a given platform, and vice versa. Due to this congruence of joining behavior, in steady state, each platform converges to the same share of users as publishers. The overall system also converges to a steady-state share distribution, but the share across platforms may be quite different.

We would like to understand how this situation changes if a user selects a platform with probability proportional to an arbitrary monotonically increasing function of the share of publishers at that platform. In general, even for two platforms and the simple publisher behavior above, the situation becomes complex. There may be multiple stable equilibria including simultaneously both trivial solutions in which a platform dominates, and nontrivial solutions in which each maintains non-zero share. We present a characterization of these equilibria, and show that in our model, each platform in the market will in expectation converge to some equilibrium.

We then study the share distribution of a variety of two-sided markets using page impression data from the Yahoo! toolbar. We are able to show that over a range of two-sided markets, share of platforms tends to remain stable over a broad period of time. While our data does not provide visibility into the individual market events, this preliminary examination agrees with the predictions of our model.

Finally, we expect that a multi-platform two-sided market involving significant cross-side (or same-side) network effects must incur some inefficiency due to the partitioning of the users among the platforms. This leads us to consider situations in which two or more platforms may fruitfully form a coalition. We define a *technical coalition* as an agreement by which two platforms share their user databases, giving new entrants the perception that each of those platforms is larger than before. New entrants, however, join a single one of the two platforms. We also consider a *business coalition*, in which platforms perform a virtual merge, new entrants join the new entity, and revenues are distributed to the owners of the two original platforms according to some distribution mechanism. We show that a properly structured business coalition will always be of interest to all users in the market, while a technical coalition will be of interest only if certain share conditions are met.

2. RELATED WORK

The economic study of two-sided markets was started by Rochet and Tirole [14] and later extended by Weyl [16]. Their work is primary focused on pricing. Lee [9] modeled two-sided markets with a few suppliers as a strategic game between platforms and suppliers. He showed that competition can lead to non-optimal equilibria. In another work, he [10] also studied how vertical integration (exclusive contracts between suppliers and platforms) affects market shares.

For certain types of products, a new customer adds value to all existing ones. This principle is called the *network effect*. In 1995 Robert Metcalfe [11] postulated that the total value of a product can grow quadratically with the number of consumers. Subsequently, Briscoe, Odlyzko, and Tilly suggested a more conservative $n \log n$ asymptotic. The detailed studies of network effect in two-sided markets was started by Parker and Van Alstyne [12] only recently. In particular, they modeled the case when it is profitable to have a zero price on one side of the market. With his recent "Linked Data" [3] initiative, Tim Berners-Lee is making another case for network effect: he postulates that connecting structured data on the Web will generate a lot of new value.

Much of the analysis of network effect in two-sided markets has employed game theory, but a growing body of work instead studies competition in such markets as a random process. In particular, the model of preferential attachment [2, 4] has been proposed to explain degree distribution in social networks, and recently Immorlica et al. [7] employed another random process to model technology diffusion in social networks.

3. MODEL

A *two-sided market* is a system of connections between three types of agents: users of the first type (consumers, males, questioners), users of the second type (suppliers, females, answerers) and intermediaries (marketplaces, platforms, portals). For simplicity, we will assume in our discussion that the two sides of the market are referred to as "blue" and "green," and will employ these terms for the remainder of the paper.

Let k be the number of platforms. At any point of time, each platform has a number of users of both types. Let $B_i(t), G_i(t)$ be the absolute audience sizes for platform $i \in [k]$ at time t. Correspondingly, let $b_i(t), g_i(t)$ be the market shares of platform i. The *market state* at time t can be described by the vector

$$M(t) = (B_1(t), \ldots, B_k(t), G_1(t), \ldots, G_k(t)).$$

In the most general setting, the evolution of the market follows the *market equation*: $M(t + 1) = \mathcal{F}(M(t))$, where $\mathcal{F}$ is a stochastic function.

Ideally, the market equation should encapsulate the following aspects:

- Cross-side network effects: if a new blue user joins some platform, then it makes the platform more attractive for the green users, and vice versa.

- Same-side network effect: if the number of green users grows, then it can add value to the platform through shared reviews or shared access tools. Note however that in some cases there might be a negative same-side network effect: the more green users in a platform, the higher is competition.

- Fitness factor: these are factors of the platform that are external and network-independent. Pricing, branding, and usability of a platform may all contribute to its fitness.

- Multi-homing: in some two-sided markets users can join exactly only one platform, whereas in others multi-homing is allowed.

- Market maturity: in an emerging market we can simplify the model by saying that new users arrive but nobody leaves. In a mature market the stream of arriving users is balanced by a stream of departing ones, and a significant number of users may switch from one platform to another.

- Platform changes: platforms can arrive, depart, or form coalitions.

- Individual properties of users: in the simplest model, one can assume that only the number of users matter in the market equation. In reality, the attractiveness of the platform is a function of the set of its current users, not just the cardinality of this set.

3.1 Attachment curves model

We model the evolution of a two-sided market as a temporal process, but the market equation given above is very general. We now define a particular type of two-sided markets about which we can make stronger statements. In our model, we will employ platform-dependent *attachment curves* $\{V_i\}$ and $\{W_i\}$ that specify the probability with which a new user of a given type (V_i for green users, W_i for blue users) will join platform i. Thus, $V_7(0.2)$ gives the likelihood that a new green user will join platform 7 if the platform has 20% market share of the blue users. Attachment curves succinctly model cross-side network effects of various forms, along with per-platform fitness effects that may manifest themselves differently for different market configurations. In the following, we will present our basic attachment curves model, and discuss how the model may be extended to cover many of the more complex interactions described above.

The market equations for our basic model are referred to as the *attachment curves equations*, defined for platform i as follows:

$$\begin{pmatrix} b_i(t+1) \\ g_i(t+1) \end{pmatrix} = (1-\epsilon_t)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} b_i(t) \\ g_i(t) \end{pmatrix} + \epsilon_t \vec{X}_{\substack{V_i(g_i(t)) \\ W_i(b_i(t))}}.$$

We now describe the components of these equations: the attachment functions V, W, the change rate ϵ_t, and the join random variable $\vec{X}$.

Attachment curves. We introduced the attachment curves above. Each function $V_i, W_i : [0,1] \to [0,1]$ is monotonically increasing. In many cases $V(0) = W(0) = 0$ and $V(1) = W(1) = 1$, meaning that no green user wants to join a platform without any blue user and every green user is eager to sign up when all of the blue users are already aboard.

Rate of market change. The sequence ϵ_t reflects proportional shrinking of market share at time t that can be due to (1) users leaving the market, (2) users switching from one platform to another, or (3) total growth of market users that makes the same absolute quantity of users represent a lesser market share. At this point, we note a couple of important special cases. When users arrive one at a time and no one switches or leaves, $\epsilon_t = \frac{1}{t+B(0)}$. In a mature market with a constant fraction of users switching during each timestep we have $\epsilon_t = \epsilon$. In extending the model to cover absolute growth rather than change in share, it may be necessary to introduce separate birth and death rates to the model, but our basic formulation combines these notions together.

Join random variable. The vector random variable $\vec{X}_{\substack{V_i(g_i(t)) \\ W_i(b_i(t))}}$ gives the fraction of blue users and green users joining platform i.

For purposes of analysis, we define a continuous variation of the attachment curves model as follows:

$$\begin{aligned} b_i'(t) &= -\epsilon_t b_i(t) + \epsilon_t V_i(g_i(t)) + \nu(g_i, b_i, t), \\ g_i'(t) &= -\epsilon_t g_i(t) + \epsilon_t W_i(b_i(t)) + \mu(g_i, b_i, t). \end{aligned}$$

Here, ν, μ are stochastic noise functions with zero expectation.

Our basic model is focused on *share effects* as we use the fraction of users of a certain type on the platform as an input to the attachment function. One could also introduce $B(t)$, $G(t)$, and t as parameters of the attachment function, allowing different behavior at different stages of the market's development.

3.2 Connections to other models

We now describe the connections between the attachment curves model and a variety of related models.

Preferential attachment. In the classic preferential attachment model [2], newcomers connect to agents in the network with probability proportional to their current degrees. This is modeled as follows:

$$V_i(x) = W_i(x) = x.$$

Notice that variations of the preferential attachment model can be encoded by suitably modifying the attachment curves: for a general preference rule, $V_i(x) \propto x^\alpha$ and $W_i(x) \propto x^\alpha$ or for incorporating fitness as in [4], $V_i(x) \propto \min(1, x^\alpha + f_i)$ and $W(B) \propto \min(1, x^\alpha + f_i')$, and so on.

Social influence. In a broad sense many network effects can be classified as "social influence". Here we consider models in which an agent joins a platform (adopts a technology) if the number of her friends already using this technology is above some threshold. This phenomenon motivates many models of influencers, propagation of epidemics, and technology adoption [8, 1, 7]. In the case of two-sided markets, we assume that there is an underlying graph of "valuable connections" between agents of two types. An agent will join a platform if it can serve as an intermediary for a sufficient number of her valuable connections. Let us show how this informal principle can be encoded in the attachment curves framework. Consider a mature market with G green users and B blue users, and assume each user has p valuable connections, of which a platform must provide at least $q \le p$ in order to attract the user. Assume that the valuable connections of an incoming user are selected at random from the G green users in the market. Then the probability that an incoming user will join a specific market is given by the following attachment function:

$$V(g_i) = \frac{\sum_{j=q}^{p} \binom{j}{g_i G}\binom{(p-j)}{(1-g_i)G}}{\binom{p}{G}}.$$

The function W is defined analogously.

Early vs. late adopters. Adoption of a new technology or an exchange platform will often proceed in waves: a group

of early adopters first joins the platform, attracting a second larger group of early majority users, who in turn grant confidence to a late majority, and so forth. This model can be formalized by a "staircase" attachment curve. E.g.,

$$V_i(x) = \begin{cases} 0 & x \in [0, 0.25) \\ 0.5 & x \in [0.25, 0.75) \\ 1 & x \in [0.75, 1] \end{cases}$$

The interpretation is that blue users will not join until the system has 25% of all green users, then suddenly half of all blue users are willing to sign up. When the fraction of green users exceeds 75%, the remaining 50% of blue users also join.

Remark on Ceyhan–Mousavi–Saberi model. Very recently Ceyhan, Mousavi, and Saberi present a generic model for network effect [15]. They model a newcomer's platform selection by the following stochastic equation:

$$i_{\text{join}} = \arg\max_i A_i(M(t)) + \nu_i, \tag{1}$$

where A_i is an attractiveness function based on current user-bases of platforms and ν_i is a stochastic noise function.

Attachment curves model has a few important differences from (1). First, we consider two-sided markets and thus, have a system of two equations. As shown in our experimental section below, blue-to-green network effects can differ significantly from green-to-blue effects.

Second, instead of using the $\arg\max$ function, we use Bernoulli random variables. This is an important distinction. There are many cases in which two users will have very different utilities from platforms and thus, it is not natural to model their choice as maximization of a universal utility plus noise. Consider a practical scenario example with two insurance networks of doctors, one of which has two-thirds of the market share and the other has the remaining one-thirds. Let every user simply join the platform that hosts its closest doctor. In this case, two-thirds of the population will join the smaller network. As we see, postulating that selection probabilities are proportional to some attractiveness function is very natural in this case. Preferential attachment models also use probabilities proportional to the attractiveness approach (as opposed to using $\arg\max$). Thus, the two models may be appropriate in different circumstances.

Finally, the attachment curves model allows study of the evolution of a single platform. We can start a random "chicken-and-egg" process from $0, 0$ market position and see how far the platform will grow guided by the given attachment curves and the V, W functions.

4. SINGLE-PLATFORM VIEW

4.1 A study

In this section, we study the attachment curves of a typical two-sided market, namely, online *question-answering* applications. In these applications, the two sides of the market are the *questioners*, i.e., the users who pose questions to the system, and *answerers*, i.e., the users who answer posed questions. Our analysis is based on two large such portals: Yahoo! Answers (`answers.yahoo.com`) and the relatively new company Stack Overflow (`stackoverflow.com`). We first describe the two datasets we use in our experiments.

Yahoo! Answers is a question-answering portal that has been operational since 2005. The scope of questions is general-purpose and each question is tagged by the questioner with

pre-defined categories ranging from Arts to Entertainment to Politics. For the purposes of our analysis, we obtained the entire data from June 26, 2005 to April 20, 2007, spanning about 90 weeks and about 23 million questions. The data is of the following schema: for each question, it contains the questioner's id, the question's category, the time when the question was asked, and for each answer to the question, the answerer's id, and the time when the answer was provided. Using the ids and timestamps, it is easy to extract the audience sizes for both sides of the market, i.e., the number of unique questioners $Q(t)$ and answerers $A(t)$ at any time t. Figure 1 shows the growth of the audience size (i.e., unique number of questioners and answerers) for Yahoo! Answers.
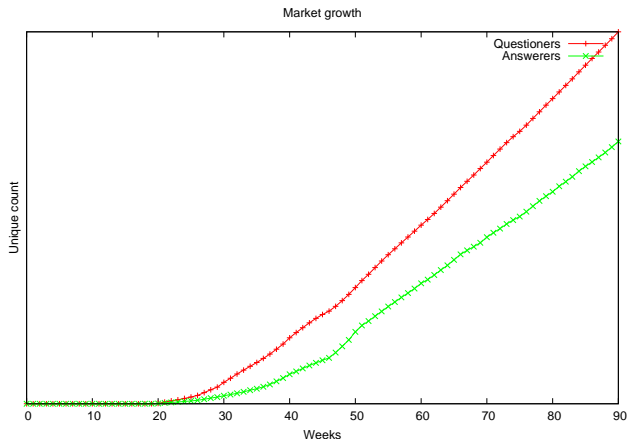

Figure 1: Audience growth for Yahoo! Answers.

We wish to analyze the attachment curves for these platforms. Recall that an attachment curve shows how the current volume of green users at a platform will influence the probability of a newly-arriving blue selecting that platform. Thus, the attachment curve may be viewed as the relationship between $G_i(t)$ and $\Delta B_i(t)$. Depending on the formulation, the behavior of a newly-arriving blue may be determined by the share of green users at the platform, or the absolute number of green users at the platform. In this experiment, we have detailed information about two platforms, but we don't have information about the entire market. Thus, we study the relationship between absolute number of blue users and change in number of green users. So we consider $\Delta Q(t) = Q(t+1) - Q(t)$ vs. $A(t)$ and $\Delta A(t) = A(t+1) - A(t)$ vs. $Q(t)$. Figure 2 shows the attachment curves.

Before discussing this data, we present our second example. Stack Overflow is a young portal, operational for the past year. The scope of questions is restricted to programming and software development, and each question is tagged by the questioner with short free text such as c#, perl, iphone, etc. For the purposes of our analysis, we obtained all the question-answer data from July 31, 2008 to July 31, 2009, spanning a year and about 0.25 million questions. The data is of the following schema: for each question, the questioner's id, the tags attached to the question, the time when the question was asked, and for each answer to the question, the answerer's id, and the time when the answer was given. As before, we can study the audience growth and attachment curves for this market. Figure 3 shows the growth and attachment curves for Stack Overflow.
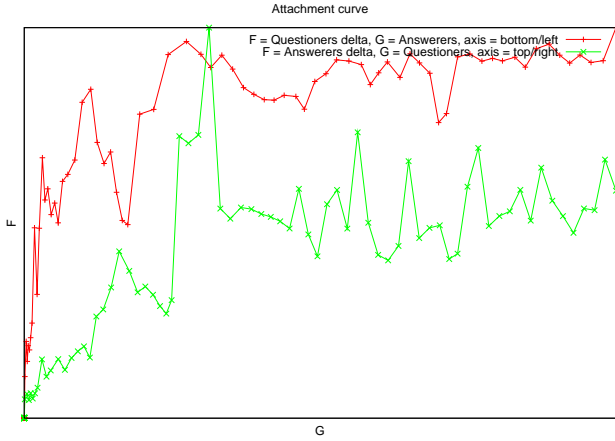
Figure 2: Attachment curves for Yahoo! Answers.

There are several takeaways from this analysis of the attachment curves.

(1) The attachment curves for both Yahoo! Answers and Stack Overflow are monotonically increasing or constant (up to perhaps large local variations). This justifies the monotonicity assumption in the model. Also, our model's ability to incorporate any attachment curve is quite crucial since from the experiments it is clear that they can be intricate.

(2) The attachment curves for Stack Overflow show a clear difference: $\Delta A(t)$ vs. $Q(t)$ is mostly flat whereas $\Delta Q(t)$ is increasing (linearly) as a function of $A(t)$. Thus, the questioners display a strong cross-side network effect, finding benefit in the presence of larger numbers of answerers (more on this below), while the answerers do not display such a pattern. This shows that the attachment curves for some platforms could be asymmetric. This asymmetry leads to a quadratic growth in questioners (seen from the system $Q'(t) \propto A(t)$ and $A'(t) \propto$ constant), while answerers grow only linearly. In fact, even for submarkets in Stack Overflow, obtained by selecting questions with tags "perl" and "iphone," their corresponding attachment curves still show marked differences between the two sides of the market (Figure 4). The attachment curves for Yahoo! Answers are much less asymmetric in comparison.

(3) It is important to clarify the relationship between $\Delta Q(t)$, $\Delta A(t)$, $Q(t)$, and $A(t)$. In particular, since both sides of the market are growing rapidly, how do we know that the network effect is cross-side rather than same-side? If there is a cross-side network effect at work, questioners will grow at a rate proportional to the number of answerers, but if this occurs during a time period in which questioners and answerers are growing in lockstep, then we will also observe a strong non-causal correlation between the number of questioners and the growth in the number of questioners. To tease apart this distinction, we observe that, while the two groups are both growing rapidly, there are periods during which one accelerates faster than the other. We model the underlying network effect as "traditional" preferential attachment, and fit $\Delta Q(t)$ with respect to both $A(t)$ and $Q(t)$ for the stable portion (last 26 weeks) of the curves in Figure 3. The R^2 value of the former is 0.93 whereas of the latter is 0.89, showing that $\Delta Q(t)$ is better predicted by $A(t)$ than by $Q(t)$ itself. From this observation, we have
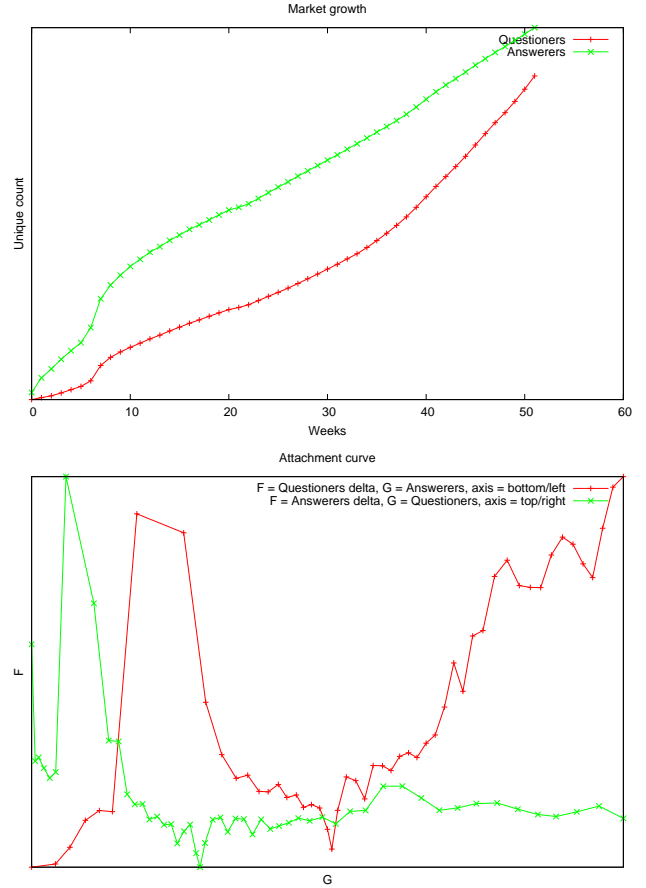

Figure 3: Growth and attachment curves for Stack Overflow.

some confidence that our initial assumption that cross-side network effect is a more likely explanation is supported by the data. As we stated earlier, $\Delta A(t)$ is mostly constant but same-side network effect marginally dominates the cross-side network effect. For Yahoo! Answers, both $\Delta Q(t)$ and $\Delta A(t)$ are mostly constant.

4.2 Equilibrium analysis

In this section, we consider continuous time market equations for one platform in the mature market case (hence we omit the subscript i):

$$b'(t) = \epsilon V(g(t)) - \epsilon b(t) + \nu(V(g(t)), b(t)),$$
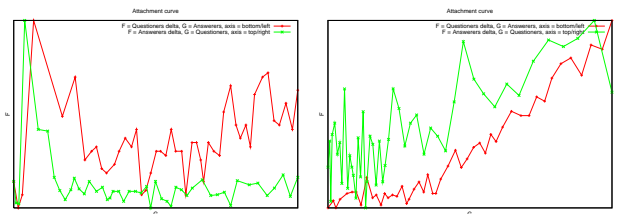$$g'(t) = \epsilon W(b(t)) - \epsilon g(t) + \mu(V(b(t)), g(t)).$$

Figure 4: Growth and attachment curves for two submarkets in Stack Overflow: "perl" and "iphone".

Since ν, μ have zero expectation, the evolution in expectation is the solution to the following system of first-order ordinary differential equations:

$$
\begin{aligned}
b'(t) &= \epsilon V(g(t)) - \epsilon b(t), \\
g'(t) &= \epsilon W(b(t)) - \epsilon g(t).
\end{aligned} \tag{2}
$$

We assume V, W are continuous and monotonically increasing on $[0, 1]$. Hence, they are Lipschitz continuous. Then, by the Picard–Lindelöf theorem [6], we obtain the following.

THEOREM 1 (UNIQUENESS). *For every starting position $b(0), g(0)$, there is a unique solution to (2).*

We now define the notion of equilibrium points.

DEFINITION 2 (EQUILIBRIUM POINT). *A pair of market shares (b, g) is an* equilibrium point *if the market equations (2) imply $b' = g' = 0$.*

Thus, the set of equilibrium points is a set of solutions for the $\{V(g) = b; W(b) = g;\}$ system of equations. Equilibrium points can be nicely visualized as follows. Take the unit square and plot the function V and the inverse of W, which is also a monotonically increasing function. All the intersection points are the equilibrium states. Figure 5 gives an example. One should read the x-axis as the share of blue users and the y-axis as the share of green users. The green attachment curve shows the change in the share of green users as a function of the share of blue users; the blue attachment curve is the transpose of the function showing change in the share of blue users as a function of the share of green users. One may verify that a point located on one of the arrows will in fact have a trajectory towards the marked equilibrium points. For any point in the unit square representing a configuration of the system, the point will move vertically towards the green curve, and horizontally towards the blue curve, during the next (infinitesimal) evolution of the system.
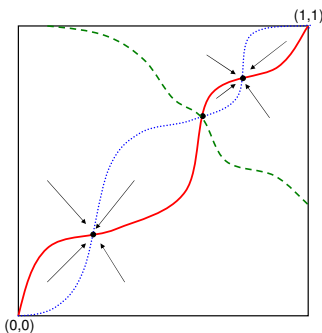

Figure 5: Representation of equilibrium points.

DEFINITION 3 (DOMINANCE AND WEAK DOMINANCE). *A market position (b, g) dominates* another market position (b^*, g^*)*, denoted $(b, g) \succ (b^*, g^*)$, if $b > b^*$ and $g > g^*$.*

Likewise, $(b, g) \succeq (b^*, g^*)$ if $b \geq b^*$ and $g \geq g^*$ (weak dominance).

LEMMA 4 (DOMINANCE). *Let $(b, g) \succ (b^*, g^*)$, and let u, u^* be the expected evolution trajectories with starting conditions (b, g) and (b^*, g^*). Then for any time t, $u(t) \succeq u^*(t)$.*

PROOF. Assume the contrary, i.e., there is a t such that $u(t)$ does not weakly dominate $u^*(t)$. Let t_0 be the infimum of such time points. By the continuity of u, u^*, and since weak dominance holds at t, weak dominance also holds at t_0. We now do a case analysis.

If both $b(t_0) = b^*(t_0)$ and $g(t_0) = g^*(t_0)$, then by the uniqueness theorem starting from t_0, both trajectories will coincide and the proof follows. If both $b(t_0) > b^*(t_0)$ and $g(t_0) > g^*(t_0)$, since our functions are Lipschitz continuous, there is a $\delta > 0$ such that the trajectory u dominates the trajectory u^* in the interval $[t_0, t_0 + \delta]$, contradicting the definition of t_0.

Therefore, we only need to address the case with one equality and one inequality. Without loss of generality, let $b(t_0) = b^*(t_0)$ and $g(t_0) > g^*(t_0)$. By Lipschitz continuity, for some interval of length δ, we have $g(t) > g^*(t)$ for $t \in [t_0, t_0 + \delta]$; we will prove that $b(t) \geq b^*(t)$ as well. The market equation (2) implies the following:

$$
\begin{aligned}
(b - b^*)(0) &= 0, \\
(b - b^*)'(t) &= -\epsilon(b - b^*)(t) + \epsilon(V(g(t)) - V(g^*(t))).
\end{aligned}
$$

By the monotonicity of attachment, the last term stays nonnegative in the $[t_0, t_0 + \delta]$ interval. Thus, equivalently,

$$
\begin{aligned}
(b - b^*)(0) &= 0, \\
(b - b^*)'(t) + \epsilon(b - b^*)(t) &= p(t),
\end{aligned}
$$

where $p(t) = \epsilon(V(g(t)) - V(g^*(t))) \geq 0$. This is a first-order linear ordinary differential equation, whose explicit solution can be written as

$$
(b - b^*)(t) = e^{-a(t)} \left(\int p(t) e^{a(t)} dt + C \right), \text{ with}
$$

$$
a(t) = \int \epsilon dt.
$$

The initial condition implies $C = 0$. Since $p(t) \geq 0$, it follows that $(b - b^*)(t) \geq 0$, for $t \in [t_0, t_0 + \delta]$. This contradicts the definition of t_0. $\square$

THEOREM 5 (CONVERGENCE TO EQUILIBRIUM). *The expected evolution in a mature market always converges to equilibrium state.*

PROOF. Since the unit square is compact, the expected evolution trajectory $u(t)$ must have at least one accumulation point. We do a case analysis on the number and dominance characteristics of the accumulation points.

Case 1. Suppose there is only one accumulation point u_0. Then the whole trajectory converges to u_0, since otherwise, one can find a second accumulation point for the subsequence of positions bounded away from u_0. Since the full trajectory converges to u_0, the market equation (2) implies a zero derivative at u_0. Indeed, suppose the derivative is $v \neq 0$. Then, by Lipschitz continuity, there is a $\delta > 0$ such that infinitely many times in the future, points on the expected evolution trajectory will be of the form

$u_0 + v\delta + O(\delta^2)$. This contradicts the convergence to u_0. Therefore, if there is only one accumulation point, the trajectory converges to this point, which is an equilibrium state.

Case 2. Suppose there are two accumulation points u^*, u^{**} such that $u^* \succ u^{**}$. We call all points that dominate $u^*/3 + 2u^{**}/3$ to be the *surplus* area, and all points dominated by $2u^*/3 + u^{**}/3$ to be the *deficit* area. There is some interval $I_1 = [t - \gamma, t + \gamma]$ during which the evolution trajectory u stays in the deficit area. Since u^{**} is a accumulation point, there is a second interval $I_2 = [t+\delta-\gamma, t+\delta+\gamma]$ in which the trajectory u stays in the surplus area. Applying dominance to any pair of time points in intervals I_1 and I_2, we obtain a third time interval $I_3 = [t+2\delta-2\gamma, t+2\delta+2\gamma]$, in which u must stay in the surplus area. Iterating this argument ℓ times, we get the interval $I_{\ell+2} = [t + 2^\ell\delta - \gamma, t + 2^\ell\delta + \gamma]$, in which the trajectory u stays in the surplus area. Thus in the future there will be arbitrary long periods during which the trajectory stays in the surplus area.

Now, let t^*, δ^* be some time points when $u(t^*)$ is in the surplus area but $u(t^* + \delta^*)$ is in the deficit area. Then by the dominance lemma, $u(t^* + 2\delta^*) \prec u(t^* + \delta^*)$ will be in the deficit area as well. Repeating this argument ℓ times, we have that $u(t^* + \ell\delta^*)$ will be in the deficit area for any positive integer ℓ.

These two statements (arbitrary long stays in the surplus area and arithmetic progression of stays in the deficit area) cannot hold simultaneously. This completes the argument for this case.

Case 3. Finally, assume there is more than one accumulation point, but no such point dominates another. Let u^* be a accumulation point that minimizes $b - g$ and u^{**} be the one that maximizes $b - g$; these exist since the set of accumulation points is compact. Note that u^*, u^{**} are unique and different, for otherwise we have a pair of dominating accumulation points, or only one accumulation point.

By the market equation (2), b's derivative at u^{**} is strictly smaller than at u^*. Indeed, $b^* < b^{**}$ and $g^* > g^{**}$, thus $-\epsilon b^* > -\epsilon b^*$ and $\epsilon V(g^*) \geq \epsilon V(g^{**})$ by monotonicity. Thus, either u^* or u^{**} has a non-zero derivative. Without loss of generality, let $v \neq 0$ be the market derivative at u^*.

Since we have a Lipschitz continuous ordinary differential equation without t in it, and since u^* is a accumulation point for our trajectory $u(t)$, there exists $\delta > 0$ such that for any $t \in [-\delta, \delta]$, $U^*(t)$ is also a accumulation point, where U^* is a solution to our equation with the initial condition $U^*(0) = u^*$. Note, that $U^*(t) = u^* + vt + o(t)$. Now, we compare these new accumulation points to u^*. First of all, they cannot have both b and g bigger than u^* has since this will be Case 2. Likewise, they cannot have both coordinates smaller than u^*. They also cannot have b smaller, but g larger, since it contradicts the choice of u^* (minimizing $b-g$). Thus, the new accumulation points can only reside in the quadrant of points containing "bigger b" and "smaller g". However, since $v \neq 0$ either with positive t or with negative t, the U^* trajectory will depart from the only permitted quadrant. This completes the argument. $\square$

5. MULTIPLATFORM VIEW

In this section, motivated by the empirical observations in Section 4, we study the evolution of multi-platform two-sided markets with attachment curves. We begin with a small experiment to understand whether two-sided markets in practice tend to display stable share configurations, or tend to display wandering or unstable share.

5.1 Multiplatform markets: A study

To perform this experiment, we employ a subset of the Yahoo! toolbar data, which contains the list of urls visited by a user (whose id is anonymized). The period of evaluation for about one year, starting June 28, 2008.

We analyze three distinct two-sided markets, and a set of platforms within each market, as follows.

- Air travel reservation, where the providers are Orbitz, Expedia, Kayak, Travelocity, Tripadvisor, Yahoo! Travel, and Mobissimo.

- Personals, where the providers are `match.com`, `eharmony.com`, Yahoo! Personals, and `plentyoffish.com`.

- Used cars, where the providers are `usedcars.com`, `cars.com`, `autotrader.com`, Edmunds, Carmax, `kbb.com`, and Ebay Motors.

Using the toolbar data, we record how many times a user visits one of the above websites each day. In this way, the market shares can be computed. (We point out the obvious caveat that this data could be biased since it is based only on the Yahoo! toolbar users.) Figure 6 shows the market shares for the three markets, aggregated for each week.

The figures show that over a one-year period in the rapidly-changing internet space, these markets seem to exhibit stable equilibria. We will revisit this observation in light of our theoretical analysis below.

5.2 Multiplatform view analysis

Let us find an instance of the attachment curves model that can explain the "constant market shares" observation discussed earlier. Let us consider a mature market with the classic attachment rule: a green user will join the platform with probability proportional to its blue market share taken to some power $\alpha \geq 0$. Let k be the number of platforms.

THEOREM 6. *If $\alpha = 1$, then any market share position of the form $b_1 = g_1, \ldots, b_k = g_k$ is an equilibrium state. If $\alpha < 1$, then the flat market share $b_1 = g_1 = \cdots = b_k = g_k = 1/k$ is the only equilibrium state. If $\alpha > 1$, then the flat market shares and market tipping (one platform controls the full market) are the only equilibrium states.*

PROOF. In equilibrium, the market equation (2) yields

$$0 = b_i'(t) = (1 - \epsilon)b_i(t) + \epsilon\frac{g_i^\alpha(t)}{\sum g_j^\alpha(t)}.$$

Thus, the equilibrium states satisfy the following system of equations:

$$g_i = \frac{b_i^\alpha(t)}{\sum b_j^\alpha(t)},$$
$$b_i = \frac{g_i^\alpha(t)}{\sum g_j^\alpha(t)}.$$

It is easy to check that the market positions from the statement of the theorem satisfy this system and hence are equilibrium states. To prove that there are no other equilibrium states, let us consider the platform with the maximum
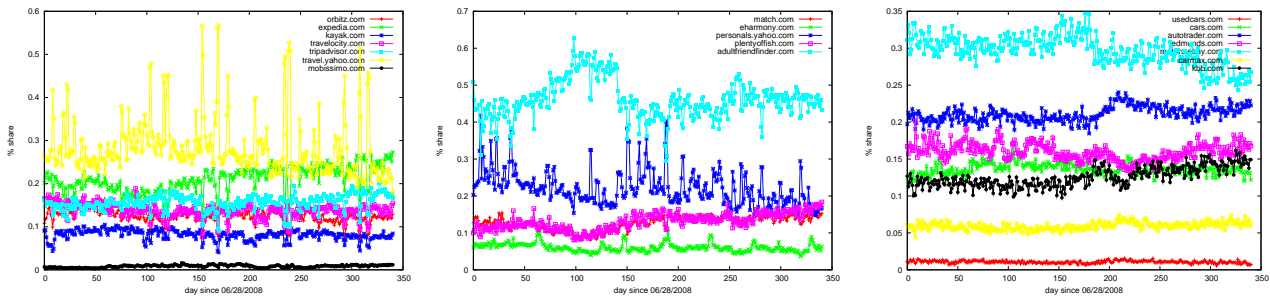
Figure 6: Relative audience shares for various multiplatform markets.

(across both blue and green) market share, and without loss of generality, let this be b_1. If $\alpha > 1$, then excluding the flat or market-tipping positions, the above system implies that $g_1 > b_1$, which contradicts the choice of b_1. If $\alpha < 1$, then excluding the flat shares position, the above system implies

$$b_1 \leq \frac{\max_j g_j^\alpha}{\sum g_j^\alpha} < \max_j g_j \leq b_1,$$

which is again a contradiction. $\square$

Suppose we start with $b_1(0) = g_1(0), \ldots b_k(0) = g_k(0)$. If $\alpha = 1$, then this state will not change since it is already in equilibrium. On the other hand, if $\alpha \neq 1$, then the market will either flatten or tip. Thus, the identity attachment curves $V(x) = W(x)$ appear to be a good model to explain the market share stability observed in Figure 6.

5.3 Coalition analysis

We use the model with identity attachment curves to analyze the potential partnerships in two-sided markets. We consider multi-platform two-sided markets, in which the value is derived from connecting the blue and green users. It seems likely that the platform providers will consider merging or otherwise unifying in order to unlock the value available by connecting the green users of one platform with the blue users of the other, and vice versa. If such a scheme generates more overall value, then perhaps an arrangement can be structured so that both platforms benefit.

Thus, it is natural to consider the formation of various types of coalitions in multi-platform two-sided markets. Here, we will analyze the situations in which it is beneficial to form coalitions. For concreteness, we will make reference to "suppliers" and "consumers" in this discussion, as an example of the green and blue users in the market.

Metcalfe [11] advocated the extreme position that the value of a network should be estimated as the square of the number of users, based on the potential connections that are enabled by the network. In this section, we follow this observation by assuming that the revenue generated by a platform is proportional to the product of the number of suppliers and the number of consumers, i.e., the revenue is proportional to the number of connections between the two sides of the market. We do not take a stance in the vociferous and ongoing debate about whether Metcalfe's law is reasonable; Briscoe, Odlyzko, and Tilly [5] advocate a more conservative revenue model of $n \log n$, which would lead to a different set of entirely reasonable conclusions. However, for simplicity, in this version we limit our attention to "Metcalfe revenue."

We study two types of coalitions: technical and business. A *technical coalition* between several platforms is an agreement to share the data they collect from suppliers and make it available to all their consumers; this is purely a data sharing agreement that presents to consumer the view that each platform in the coalition appears to have available a larger number of suppliers. A *business coalition* is a technical coalition in which any additional revenue generated due to the formation of the coalition is redistributed among the partners according to some prior agreement that was made when the coalition was formed.

A coalition is *beneficial* to an individual platform if it earns more money compared to operating on its own. A coalition is beneficial if it is beneficial to *all* its users. Finally, a coalition is *stable* if no sub-coalition can earn more revenue.

For the following results, we assume that both V and W are linear attachment curves. Therefore, the steady-state g_i and b_i share values have $g_i = b_i$; we will refer simply to g_i for simplicity. Let m be the total number of suppliers and n be the total number of consumers.

THEOREM 7. *In a k-platform market, if all the market shares are less than $\frac{1}{\sqrt{k}}$, then the full technical coalition is beneficial under Metcalfe's law.*

PROOF. Let $g_1, \ldots, g_k$ be the initial market shares, $g_1 + \cdots + g_k = 1$. By the structure of the market model, if the ith platform stays out of any coalitions, then its market share is expected to stay the same. Thus, by Metcalfe's law, its expected revenue is $g_i^2 mn$. After a full technical coalition is formed, from the supplier's point of view, all platforms become equally attractive since any consumer can be reached; the same is true for consumers. Thus, the market shares will move from $g_1, \ldots, g_k$ to being uniform. Now, the new revenue of any platform is mn/k, since it connects $1/k$ suppliers to all consumers as well as $1/k$ consumers to all suppliers. By our assumptions, $g_i^2 < 1/k$ and thus in the limit, all platforms earn more than prior to the coalition. For platforms with an initial market share below $1/k$, coalition both increases the revenue and the market share; thus, it is beneficial throughout the whole process. For platforms with an initial market share above $1/k$, coalition immediately increases the profits but slowly decreases the market shares. Because of that, and since in the limit coalition is beneficial, it is beneficial at all times. $\square$

It is easy to see that beneficial technical coalitions need not be closed under subset. Indeed, consider market shares proportional to $1, 1, 4, 5$. The first two platforms can form a

beneficial coalition; the first three will not (as the third platform will lose money); however all four can form a beneficial coalition again.

THEOREM 8. *For all markets, a full stable business coalition can be formed under Metcalfe's law.*

PROOF. Let $g_1, \ldots, g_k$ be the initial market shares, $g_1 + \cdots + g_k = 1$. Let the revenue sharing agreement allocate a share g_i (that is proportional to its initial market share and not to its initial revenue share) of total coalition revenue to every platform i. The total revenue of the full coalition is mn, by Metcalfe's law. Thus, the revenue-in-coalition for the ith platform is $g_i mn$. Suppose platforms $i_1, \ldots, i_\ell$ decide to drop out of full coalition to form their own alliance. By the market model, the market share of this new alliance will stay constant at the level of $g_{i_1} + \cdots + g_{i_\ell}$. Then, their total revenue will be $(g_{i_1} + \cdots + g_{i_\ell})^2 mn$, which is always smaller than the revenue of alliance users within the full coalition, i.e., $(g_1 + \cdots + g_k)^2 mn$. $\square$

Naturally, a question arises: what is the "optimal" or "fair" revenue redistribution? Consider the simplest case of two platforms with equal market share. Their standalone profits are $mn/4$ and their total in-coalition revenue is mn. Thus, the allocation of "extra profit" of $mn/2$ is an instance of the ultimatum game. This observation leads us to the following open problem: can revenue redistribution in any market be decomposed into two parts, namely, necessary payments and ultimatum payments?

6. FURTHER WORK

The key directions for further work are:

- Find a fast algorithm that, given piece-wise linear attachment curves and a starting condition, computes the equilibrium points.

- Analyze the stability of market evolution under various models of noise. Techniques from [4, 15] might be relevant here.

- Study the predictive power of the attachment curves model. How useful are the attachment curves to predict the future evolution of a two-sided market?

- Find the optimal revenue redistribution schemes for coalitions in two-sided markets.

- Study investment strategies. Knowing the attachment curves, should a platform owner invest extra effort (say, marketing campaign) to attract more green or blue users? How much should be spent? Is there a "cheap" way to create a snowball effect?

7. REFERENCES

[1] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan. Group formation in large social networks: Membership, growth, and evolution. In *Proc. 12th KDD*, pages 44–54, 2006.

[2] A. L. Barabasi and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

[3] T. Berners-Lee. Linked data. http://www.w3.org/DesignIssues/LinkedData.html.

[4] C. Borgs, J. Chayes, C. Daskalakis, and S. Roch. First to market is not everything: An analysis of preferential attachment with fitness. In *Proc. 39th STOC*, pages 135–144, 2007.

[5] B. Briscoe, A. Odlyzko, and B. Tilly. Metcalfe's law is wrong. *IEEE Spectrum*, pages 26–31, 2006.

[6] E. A. Coddington and N. Levison. *Theory of Ordinary Differential Equations*. McGraw-Hill, 1955.

[7] N. Immorlica, J. Kleinberg, M. Mahdian, and T. Wexler. The role of compatibility in the diffusion of technologies through social networks. In *Proc. 8th EC*, pages 75–83, 2007.

[8] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *Proc. 9th KDD*, pages 137–146, 2003.

[9] R. S. Lee. Competing platforms, 2008. Working paper.

[10] R. S. Lee. Vertical integration and exclusivity in platform and two-sided markets, 2009. Working paper.

[11] R. Metcalfe. Metcalfe's law: A network becomes more valuable as it reaches more users. *Infoworld*, 1995.

[12] G. Parker and M. W. V. Alstyne. Two-sided network effects: A theory of information product design. *Management Science*, 51(10), 2005.

[13] J. Rochet and J. Tirole. Two-sided markets: A progress report. *The RAND Journal of Economics*, 35(3):645–667, 2006.

[14] J.-C. Rochet and J. Tirole. Platform competition in two-sided markets. *Journal of the European Economic Association*, 1(4):990–1029, 2003.

[15] M. M. S. Ceyhan and A. Saberi. Social influence and evolution of market share. In *Proc. NetEcon*, 2009.

[16] E. G. Weyl. The price theory of two-sided markets, 2009. Working paper.