A Dynamics for Advertising on Networks

L. Elisa Celis[1], Mina Dalirrooyfard[2], and Nisheeth K. Vishnoi[1]
[1] École Polytechnique Fédérale de Lausanne (EPFL)
[2] Massachusetts Institute of Technology (MIT)

Abstract. We study the following question facing businesses in the world of online advertising: how should an advertising budget be spent when there are competing products? Broadly, there are two primary modes of advertising: *i*) the equivalent of billboards in the real-world and (search or display) ads online that convert a percentage of the population that sees them, and *ii*) social campaigns where the goal is to select a set of initial adopters who influence others to buy via their social network. Prior work towards the above question has largely focused on developing models to understand the effect of one mode or the other. We present a stochastic dynamics to model advertising in social networks that allows both and incorporates the three primary forces at work in such advertising campaigns: (1) the type of campaign – which can combine buying ads and seed selection, (2) the topology of the social network, and (3) the relative quality of the competing products. This model allows us to study the evolution of market share of multiple products with different qualities competing for the same set of users, and the effect that different advertising campaigns can have on the market share. We present theoretical results to understand the long-term behavior of the parameters on the market share and complement them with empirical results that give us insights about the, harder to mathematically understand, short-term behavior of the model.

1 Introduction

Online advertising is now a $1.5 Trillion industry and, on social networks alone, it accounts for over $23 Billion worldwide [22]. This is currently 13.9% of all digital ad spending, and 70% of marketers will spend more on social media advertisements in the coming year [30]. Comparatively, spending on television advertisements is approximately $39 Billion [21]. In fact, the market is now so large that outside companies have arisen as consultants in this space; IBM alone spent over $100 million dollars just to develop their advertising consulting business in 2014 [29].

It is believed that social influence has a powerful effect on customer decisions [1, 11, 3], and leveraging its power has been an important aspect of many advertising campaigns. In formal studies, this has primarily been addressed by finding the optimal set of *seeds*, i.e., initial adopters who are given the product for free, with the goal of maximizing the extent or speed at which the product spreads throughout the social network [15, 14]. The success of such campaigns has been heralded; see [25]. More traditional forms of advertising focus on buying ads, either online (keyword or banner ads) or offline (in billboards, magazines or TV). The goal of these ads is to maximize the probability that a user who sees the ad switches to that product. Studies have focused on targeting the right group of people to view the ad, or improving the ad's appeal [13, 18].

These two advertising strategies are often related and the social influence can magnify the effect of traditional advertising. Indeed, once a user has a product, they influence their *neighbors* regardless of whether they converted by being seeded, seeing an advertisement, or were themselves influence by another. In order to study these effects quantitatively, one needs a dynamics that can capture the spread and competition of products via social influence when both types of advertisements are at play. Such a model, minimalistically, should capture:
(1) the kind of campaigns (a) via seed selection and (b) via traditional ads,
(2) the mechanism of social influence and competition amongst products, and
(3) the quality (which can take into account the price) of products.
While there is a rich body of prior works where mathematical models have been developed and analyzed for subsets of parameters above, to the best of our knowledge, there is no formal study that incorporates all of the parameters (1)-(3) above; see Section 1.2 for a discussion.

1.1 Our Contributions

Our main contributions are a mathematical model to facilitate the study of the effect of parameters (1)-(3) above on the market share, a set of technical results that allow us to understand the long-term behavior of the model, and a set of complementary empirical results that give us insights about the model in the short-term (where it seems difficult to analyze the model rigorously). Our model is inspired by viewing the competition among different products for the same market base as an evolutionary dynamics on a network and realizing that the various advertising parameters such as quality, traditional advertising and the spread of influence can be captured as parameters such as *fitness*, *mutation* and *selection* in this setting; see also [20, 17, 5, 28].

Our Model. Consider the setting in which there are m products and each person uses exactly one product $i \in [m]$ at every time step; i.e., the products compete for the user base.[3] Each time step is a pre-determined time period during which an individual has an opportunity to switch to a different product; depending on the domain, the length of this time period could vary from minutes (e.g., web browsers) to years (e.g., cars). The main quantity we are interested in is the evolution of the *market share*, i.e., the fraction of people using each product.

Quality of a product. We let a_i be a positive number such that, given the option of all products $1, \ldots, m$ and no outside influence, a user selects option i with probability proportional to a_i. Hence, the a_is capture the relative quality (which can take the price into account) of product i compared to other products; we refer to this as the product's *fitness*. The owner of product i can potentially increase a_i by improving their product's quality.

Social network and competition. The influence network is captured by a weighted, directed graph $G = (V, E, w)$ where each user is a node $u \in V$, and a directed edge $uv \in E$ represents the fact that u has influence on v. Let $n = |V|$ denote

[3] We can think of one products as the "null" choice – i.e., no product is selected.

the number of nodes or users. The weights $w : E \mapsto \mathbb{R}_{\geq 0}$ quantify the amount of influence u has on v. If we let $S_i(t)$ be the set of vertices who are using product i at time t, then the probability that a node v decides to use product i at time $t+1$ due to social influence is proportional to $\sum_{uv \in E, u \in S_i(t)} w(uv)a_i$. In other words, node v will *select* a product based on which products her neighbors use, the quality of those products, and the amount of influence the neighbors have on her. We expect a node to be more easily influenced by a neighbor using a good product than a bad one – folding the fitness into the influence step captures the competition between products in this way.

Traditional advertising. We allow users to switch products independently of the social influence as in the previous paragraph, e.g., after seeing a billboard ad. We let Q^v_{ij} be the probability that node v using product j spontaneously converts, or *mutates* to product i. For mathematical convenience, we assume that each of these $m \times m$ mutation matrices $Q^v > 0$ and is the same (denoted Q) for each v. Note that Q need not be a symmetric matrix. A company could increase its spending on traditional advertising to increase Q_{ij} for $j \neq i$.

Seed selection. Finally, the owner of a product can select a *seed set* $S \subseteq V$ of people to whom they give the product for free in the beginning of the process, effectively forcing their conversion. The users are under no obligation to continue with this product in future time steps.

The problem. When allocating a budget, a company with product i should then evaluate the tradeoffs between increasing a_i (i.e., improving the product), increasing $Q_{i\cdot}$ (i.e., increasing ads and hence mutations to itself), or increasing $|S|$ (i.e., getting more initial adopters).

As in prior work, we assume the influence network is fixed, and hence a company cannot modify it to its benefit. We also assume that network can be seeded only at the first time step. These aspects combine to form a stochastic dynamical system that can be viewed as a Markov chain over the state space $\{1, 2, \ldots, m\}^n$. Our Markov chain can be described in several ways without changing the limiting distribution. For instance, one could consider a description where the traditional advertising competes with the social influence; for example, at each time step, with some probability a node spontaneously changes its state from i to j, and with some probability it is influenced by a neighboring node. We discuss possible shortcomings and extensions of our model in Section 4.

Theoretical Results. One of the main difficulties that arises in analyzing our model is the fact that the a_is are not all the same. We start by noting that when $Q_{ij} \in (0, 1)$ for all i, j and the in-degree of each node in the influence network is at least one, our process forms an ergodic Markov chain (even if the network is not connected) and, thus, has a unique stationary distribution. As a consequence, it is clear that the selection of the seed set S has no effect in the long-term since the process converges to this stationary distribution regardless of the initial state. However, the *network structure may still play a role in determining the shape of the steady state distribution.* Allowing some Q_{ij}s to be zero might give rise to absorbing states in our Markov chain; this would also allow for the possibility of the starting state affecting the steady state. See Section 4 for a discussion.

Towards computing this unique steady state, it can be shown that closed form solutions for the steady state do not exist except in the most trivial of cases. Even for a very simple settings of our model (e.g., unweighted, undirected graphs with $Q_{ij} = 0$ for $i \neq j$, a result of [5] can be used to show that computing the steady state exactly is #P hard. This leaves us with two alternatives: (1) derive weaker, but asymptotically good, bounds on the steady state of the stochastic process analytically. Or (2) deploy the Markov Chain Monte Carlo (MCMC) framework to get samples from close to the steady state in order to compute the required statistics – here, it becomes important to prove that the mixing time of the underlying Markov chain is fast.

Deterministic approximation. Towards (1), we study a *deterministic* dynamical system that can be viewed as a mean-field approximation to the stochastic dynamics. Roughly, this process has the same set of parameters, but instead of a single product, it maintains a probability distribution at each node that indicates its preference among the products. The nodes update these probabilities deterministically taking into account the influence of its neighbors and the fitnesses and the matrix Q; see Equation (1) and Lemma 1. The advantage of working with this deterministic process is that we can precisely characterize its steady state (see Theorem 1). Further, computing this deterministic steady state turns out to be an eigenvector problem for an $m \times m$ matrix. As a simple consequence, for the two-product case, one can even obtain a formula for the steady state in closed form for which it is clear that the steady state will primarily consist of the product with the highest quality.

Concentration. Our theoretical results indicate that, despite the possibility of correlations, the market shares in the steady state of our stochastic dynamics are likely to be *concentrated* around those predicted by the deterministic dynamics (see Theorem 2). The quality of the concentration depends on the number of nodes in the network and the minimum in-degree of the network (the higher these numbers are, the better the concentration). Thus, when the network's size and degree are large, the deterministic process could be taken as a first-order approximation to the stochastic process.

Mixing time. Towards (2), we show that for all graphs with large enough minimum degree (roughly $\log n$ – see Theorem 3), the Markov chain underlying our stochastic dynamics mixes fast. Key to the proof turns out to be the deterministic dynamics mentioned above. We show how the geometry of the unique fixed point of this deterministic dynamics can also allow us to construct a contractive coupling to prove rapid mixing.

To summarize, our model is amenable to a rigorous analysis and, importantly, our theoretical results suggest algorithms with provable bounds to estimate the statistics such as market share from the steady state. Further, the time to convergence to steady state for a particular set of parameters for our model being fast in many cases implies that the advertising strategies are efficient and can have the desired outcome in a reasonable time – after all, a strategy that needs 100 years to attain 99% of the market share is not useful. Extending our results to all networks (removing the minimum degree condition) seems quite challenging.

Empirical Results. While the theoretical results above give an indication of the asymptotic market share, sometimes we may be interested in the short-term value of an advertising campaign.[4] We study this regime in Section 3, where we conduct empirical studies in order to understand the effect of the model parameters in the short-term. For each experiment, we isolate a single parameter, either the network, the seed set, the mutation parameters, or the product quality, and strive to evaluate its effect on the market share.

Networks. Despite differences in origin, size and properties of the three real-world networks we consider, we observe that the market share of a product over time converges relatively fast; see Figure 1(a). While the networks we study have relatively low degree (and hence our theoretical results do not apply), we believe that this fast convergence is because the diameter of the networks, as in most social networks, is small. Proving that small diameter suffices for fast convergence would be an interesting direction for future work. Furthermore, we observe that on all networks, for our simulation parameters, the population convergence to having almost all of its mass on the best product. As we will see in later experiments, the model is extremely sensitive to the gap in quality between the best and second-best product. Perhaps when the qualities are (nearly) the same, the effect of the network in the steady state would be more clear empirically. Despite these similarities, it is clear that the model converges faster on some networks than on others, and an exploration of the short-term market share remains important.

Parameters $S, \mathbf{a}$ and Q. While the choice of which seeds are selected for S affect the market share in the short-term, the improvement is roughly linear in the size of the seed set (see Figure 1(b)). In contrast, the improvement in market share is a sigmoid in a_1 and $Q_{1\cdot}$ with the inflection point in the range of realistic values (see Figure 1). Hence, this suggests that there are thresholds for a_1 and $Q_{1\cdot}$, such that (if our budget allows) we should ensure to cross. The threshold for a_1 is simply the maximum quality of a competing product – intuitively, *having the best product ensures that the steady state is in our favor.* As long as a_1 is larger than the other $a_\cdot$, we find that small increases in $Q_{1\cdot}$ seem to have the largest positive effect in the market share in the short-term. In essence, the $Q_{1\cdot}$ function as a way to continuously generate seeds (as opposed to selecting them only once with S). This suggests that as long as we have the best product, *in the short-term, increasing $Q_{1\cdot}$, e.g., by improving or increasing the number of ads, is more important both than selecting seeds and improving product quality.* Overall, our results lead to the following qualitative insights:

- the initial seed set has a minimal effect on the limiting market share,
- a new product must have the highest quality in order to gain significant market share, and
- improving traditional advertisement can be more effective than improving seed set selection.

[4] For example, for products such as cars, time steps may be on the order of years. Hence, we may be interested in a constant number of time steps, which is less than the fastest mixing time we could hope for.

1.2 Related Work

There is a large literature on optimal advertising strategies in order to maximize the adoption of goods (see [26] for a survey). These works focus on optimizing the ads and product quality; e.g., [13, 18]. In parallel, another long line of work on local interaction with regard to social influence and the adoption of products (see [19] for a survey). Such a work often uses stochastic models (known varyingly as diffusion or cascades) and dynamical systems exist (see [7] and [4] respectively). Models for understanding social influence in networks often take the form of some kind of a *threshold rule*, as first proposed by [10, 24], and many variations have been studied. Towards this, theoretical and empirical studies have focused on the problems of finding either the optimal size of, or the optimal seeds in, the set S (e.g., [14, 12]). For instance, in an important piece of work, [15] proved that the problem of *which* seeds to select, given a size constraint, is NP-hard and also provide greedy approximation algorithms for this problem.

A key contribution of our work is a model which allows us to optimal advertising and in the presence of social influence, thus bridging these two literatures. To the best of our knowledge, very few works have tackled this challenging problem. In two notable exceptions study a monopolist firm (i.e., the a single product) being sold to a network of individuals that can influence each other. Here, the quality of the product can be captured by a level of *effort* [9] or *price* [9] set by the firm, and the product spreads across a (random, infinite) network via a dynamical system which depends on the degree distribution of the vertices. In contrast, our formulation allows us to consider arbitrary and finite networks in which multiple products compete for market share.

Our stochastic model draws from two different models that arose in the study of asexual finite populations; specifically that of [6] and [17] (see also [20] for a general reference on evolutionary models). The first model underlies the study of evolution of viruses and has been used [27] to inform drug design despite their apparent simplicity. There is no explicit network in the first model, although it is equivalent to our model in a complete unweighted graph with self-loops if all $Q_{ij} = q$ for $i \neq j$ and some fixed parameter q. The second model studies network models of evolution and there is recent rigorous work [5], but without mutations which are crucial to our setting. Prior to our work on network models with mutations, the only rigorous studies we are aware of were for the complete graph case; initiated by [6] and followed by [28]. The network structure makes the analysis significantly harder and raise many interesting questions. Our terminology (*fitness*, *mutation*, and *selection*) is emphasized to draw these parallels to the informed reader. The use of deterministic dynamics or mean-field approximations to study stochastic processes has a rich history in the probability literature [23, 31, 2], however, non-asymptotic results such as our concentration result are rare. Also, the use of such a deterministic dynamics to bound the mixing time of a stochastic process is quite new and, to the best of our knowledge in two different lines of works; see [28, 8] and the discussions therein.

2 Theoretical Results

In this section we formally state our theoretical results which concern the behavior of the market share in the long-term. At the expense of slight repetition, we begin by stating the stochastic dynamics formally and introducing the corresponding random variables. Subsequently, we describe the deterministic dynamics that will help us approximate the steady state behavior of our stochastic dynamics for large enough networks. This is followed by a proof of the approximation result. Finally, we present our result on the mixing time of the stochastic process. Due to space constraints, we simply sketch the main ideas; complete proofs appear the full version of this paper.

2.1 Preliminaries and the Stochastic Process

For a vertex v, let $N_{\text{in}}[v]$ denote the set of edges coming in to v. Let F be an $m \times m$ diagonal matrix where $F_{ii} = a_i$ and $F_{ij} = 0$ for $i \neq j$. Recall that Q_{ij} denotes the probability of type j mutating to type i. Q is column-stochastic: if 1_m denotes the all 1 vector of dimension m, $1_m^\top Q = 1_m$.

At each time $t \geq 0$, each node in the graph has a type in $\{1, \ldots, m\}$. We denote the type of vertex $v \in V$ at time t by the random variable $X_v^{(t)}$. In this notation, given $(X_u^{(t)})_{u \in V}$, our stochastic dynamics can be mathematically thought of as the following three steps: 1) Each vertex u replaces the type $X_u^{(t)}$ by $a_{X_u^{(t)}}$ many copies of the same type. 2) Each vertex v looks at the set of vertices which point to it, i.e., $N_{\text{in}}[v]$ and *selects* who to copy in the following way. For each $u \in N_{\text{in}}[v]$, first it further multiplies each type currently residing at u by a factor of $w(uv)$. Subsequently, it samples a type from the multi-set union of the populations residing at each $u \in N_{\text{in}}[v]$ uniformly at random and independently for each v. For a vertex v, call this chosen type $Z_v^{(t+1)}$. 3) $Z_v^{(t+1)}$ independently mutates for each v according to the matrix Q. That is, a type j mutates to a type i with probability Q_{ij}. The resulting type at vertex v is denoted by $X_v^{(t+1)}$. As remarked earlier, there are other ways to describe the model that does not change the stochastic properties; for instance, the order of the second and the third steps are interchangeable or can be combined.

For convenience we work with integral weights and fitnesses, however, note that the matrix F can be scaled by an arbitrary constant without changing either process; similarly, the weights of incoming edges to any node can be scaled by an arbitrary constant. As long as the scaling is by constants, the theorems and their proofs continue to hold as such. For the analysis, by replacing a weighted edge with multiple edges, we may assume without loss of generality that the graph is unweighted. Hence, subsequently, we think of $N_{\text{in}}[v]$ as a multi-set and when we talk of its cardinality, it is the cardinality of the multi-set. We let $\delta = \min_v |N_{\text{in}}[v]|$.

We start by noting that when $Q > 0$ and $\delta \geq 1$, this stochastic process has a unique stationary distribution which we denote by π. Note, that this π will in general depend on the network structure. Thus, we can study the time to stationarity or the mixing time of the Markov chain. Recall that the *mixing time* $t_{\text{mix}}(\varepsilon)$ is defined as the smallest time such that for any starting state, the

distribution of the state $X^{(t)}$ at time t is within total variation distance ε of π. For concreteness, we use $t_{\text{mix}}(1/4)$. It is well known that $t_{\text{mix}}(\varepsilon) \leq t_{\text{mix}}(1/4) \log 1/\epsilon$.

2.2 The Deterministic Dynamical System

We now present the deterministic counterpart to our stochastic model and argue how it functions as its first order approximation. Here, at each time $t \geq 0$, each node in the graph has a probability distribution over the set $\{1, \ldots, m\}$. We denote this distribution for a vertex $v \in V$ at time t by the vector $p_v^{(t)} \in \Delta_m$ where $\Delta_m = \{x \in \mathbb{R}^m,\ x \geq 0,\ \sum_{i=1}^m x_i = 1\}$. Given $(p_v^{(t)})_{v \in V}$, we now describe how to generate $p_v^{(t+1)}$. In the first step, each vertex v multiplies each coordinate of $p_v^{(t)}$ be the corresponding fitness to obtain $Fp_v^{(t)}$. Note that this is no longer a probability vector. Then, each type present at each vertex, mutates according to the matrix Q resulting in an intermediate population $QFp_v^{(t)}$. Finally, each vertex v looks at the set of vertices which point to it, i.e., $N_{\text{in}}[v]$ and updates its distribution over the types to $p_v^{(t+1)}$ by taking the weighted average of the intermediate probability distributions from among $u \in N_{\text{in}}[v]$ and subsequently normalizing it to be a probability distribution. Formally,

$$p_v^{(t+1)} = \frac{\sum_{u \in N_{\text{in}}[v]} QFp_u^{(t)}}{\sum_{j=1}^m \sum_{u \in N_{\text{in}}[v]} QFp_u^{(t)}(j)}. \tag{1}$$

Let $P^{(t)}$ denote the $m \times n$ matrix where the u-th column is the vector $p_u^{(t)}$. Thus, we can think of the deterministic process as implicitly specifying a dynamical system $f : \Delta_m^n \mapsto \Delta_m^n$ defined by the rule $P^{(t+1)} = f(P^{(t)})$. We show that starting from any initial point, the dynamical system converges to a unique limit P which has the property that each column is the same; thus, rather surprisingly, the correlations induced by Equation (1) disappear with time and the network has no effect in the long-term behavior of this dynamics.

The understanding of this P reduces to understanding the long-term behavior of the dynamics $g : \Delta_m \mapsto \Delta_m$ which maps a point x to $g(x) = \frac{QFx}{\|QFx\|_1}$. Since $QF > 0$, the Perron-Frobenius theorem implies that QF has a unique positive eigenvector $p \in \Delta_m$ with a positive eigenvalue λ_1. Therefore $g(\cdot)$ has a unique fixed point p in the interior of the simplex Δ_m. The Perron-Frobenius theorem also implies that for every $x \in \Delta_m$, $\lim_{t \to \infty} (QF)^t x / \lambda_1^t \to p$. In fact, we show that for all $t > \frac{4 \log \frac{1}{c\epsilon}}{\log \frac{\lambda_1}{|\lambda_2|}}$, we have $\|g^t(x) - p\|_\infty < \epsilon$ for any $x \in \Delta_m$, where c is a constant independent of x and $\lambda_1 > |\lambda_2|$ are the top two eigenvalues of QF with largest magnitudes. Using this convergence result, we prove that P, the limit of f, is the matrix $p1_n^\top$.

Theorem 1. ***(Limit of the deterministic process.)*** *Let f be the dynamical system as defined above. Then, given an $\epsilon > 0$, for all $X \in \Delta_m^n$, $\|f^t(X) - p1_n^\top\|_\infty \leq \epsilon$ for $t \geq \frac{4 \log \frac{M(m+1)}{c\delta\epsilon}}{\log \frac{\lambda_1}{|\lambda_2|}}$, for some constant c independent of X and the graph. λ_1, λ_2 are as above, $M = \sum_{i=1}^m a_i$ and $\delta = \min_v |N_{\text{in}}[v]|$.*

The difficulty in the proof of this theorem arises from the fact that f is a non-linear dynamical system acting on a matrix. The key observation is that when f is applied on a rank-one matrix, the outcome is a rank-one matrix. Consequently, we can prove the theorem above when the starting point is a rank one matrix. To prove it when the starting point is general rank matrix $\sum_i e_i v_i^\top$, (here e_i is the standard basis vector) we can write it as a sum of rank-one matrices and show that the application of f^t (f t-times) results in, roughly, a matrix of the form $\sum_i (QF)^t e_i v_i B_t^\top$, where A is a fixed positive matrix, while B_t is less nice. However, we can use the fact that $(QF)^t e_i \to p$ for all i, along with the fact that f has a bounded Lipschitz constant to complete the proof of the theorem. The proof of this theorem is deferred to the full version due to space constraints.

2.3 Relationship Between the Stochastic and Deterministic Models

We now present our results relating the deterministic and the stochastic process. We let $X^{(t)} \in \{0,1\}^{m \times n}$ be the matrix where $X^{(t)}_{iv} = 1$ if the state of vertex v is i and 0 otherwise. Thus, with a slight overload of notation (which should be clear from the context), $X^{(t)}_v$ denotes the column vector of $X^{(t)}$ corresponding to the vertex v.

Thus, $(X^{(t)} 1_n) e_i$ is the number of vertices of type i. Let $D^{(t)}$ denote the vector $\frac{1}{n} \cdot X^{(t)} 1_n$. The starting point is the following easy to verify equality which relates one step of the Markov chain to the deterministic process.

Lemma 1. $E[D^{(t+1)} \mid X^{(t)}] = \frac{1}{n} \cdot f(X^{(t)}) 1_n$.

Moreover, using Chernoff bounds, we can show that $D^{(t+1)}$ is close to $\frac{1}{n} \cdot f(X^{(t)}) 1_n$ with high probability. Ideally, we would like to show that if $X \in \{0,1\}^{m \times n}$ is sampled from the stationary distribution π of the stochastic process, then the corresponding counting vector $D = \frac{1}{n} X 1_n$ is concentrated around the vector p (as defined in Theorem 1). An obvious approach would be to argue that we can iterate this argument over t steps and take a union bound to ensure that $D^{(t)}$ remains close to the deterministic process with a high probability. However, this approach suffers from a couple of major problems: the first is that it is not clear how to iterate. It would be possible to iterate if we could ensure that $\frac{1}{n} \cdot f(X^{(t)}) 1_n$ is very close to $g\left(\frac{1}{n} X^{(t)} 1_n\right)$. This is for instance true when the underlying graph is a complete graph but does not hold in general. Even assuming that this is true for a moment, the second obstacle is that g is not necessarily a global contraction and in fact might even expand discrepancies, so that the discrepancy between the behavior of the random process and the deterministic prediction can grow exponentially with time. Thus, we can use the union bound only over at most $O(\log n)$ steps. To get around this problem, we use the fact that p is an attracting fixed point of g. Thus, we know that within $O(\log n)$ steps, g reaches a $n^{-\Theta(1)}$-neighborhood of its fixed point, and does not subsequently leave this neighborhood. Using this fact, we can essentially bootstrap the naive union bound argument described in the previous paragraph to show that starting from *any* state, the random process also reaches in $\Theta(\log n)$ steps a distribution that is concentrated on a $n^{-\Theta(1)}$ neighborhood of the fixed

point of the deterministic process. Thus, the behavior of the deterministic limit close to its fixed point turns out to be the crucial ingredient in understanding the convergence properties of the stochastic finite population process. The following theorem shows that most of the mass of the stationary distribution of the stochastic process is concentrated around the fixed point of the deterministic process when the underlying network has large minimum degree.

Theorem 2. ***(Concentration of the stationary distribution.)*** *There exist constants $\gamma, \beta > 0$ depending only upon m, Q and F such that if $\sigma \geq (\log n)^{-\gamma}$, and if D is the frequency vector obtained from the stationary distribution π of the stochastic process, then $\Pr[\|D - p\|_\infty \leq 2\sigma] = 1 - 1/n^\beta$ for all graphs with minimum degree $\delta \geq \Omega((\log n)^{1+\alpha})$ where $\alpha > 0$ is an arbitrarily small constant.*

Note that our empirical results (see Section 3) indicate that this should hold for all networks with small diameter and we leave it as a challenging open problem to prove this formally. We omit the proof of this theorem due to space constraints.

2.4 The Mixing time of the Stochastic Process

We now present our main result on the mixing time of our stochastic dynamics.

Theorem 3. ***(Mixing time.)*** *Let m, Q, F be fixed and let G be a graph on n vertices (for n large enough) such that that $\delta = \Omega\left((\log n)^{1+\alpha}\right)$ for any $\alpha > 0$. Then, $t_{\text{mix}}(1/4) = O(\log n)$, where m, Q, F, α are constants.*

Note that the mixing time bound is quite surprising as the size of the state space is 2^n and the mixing time roughly a $\log \log$ of the size of the state space. The proof relies on a multi-phase *coupling* argument. Establishing a coupling between two identical (but correlated) copies of a Markov chain which reduces the expected distance is a generic technique to establish mixing time bounds. At a very high level, we can demonstrate a contractive coupling when the two chains reach close enough to their steady state, which is related to the convergent point of the deterministic dynamical system. To prove that they reach close to their steady state we need to ensure that there is measure concentration in each step of the Markov process, that the number of steps is small, and that is where the minimum degree bound seems to help us. The proof is quite technical and builds on an extends the framework of [28] who proved a similar result for the case of the complete graph. We omit the proof here due to space constraints.

3 Empirical Results: Short-Term Market Share

We now consider an empirical evaluation of the effect of the model parameters on the market share in the short-term. In lieu of a real influence network, we consider three real-world social network datasets for our simulations. We run the stochastic process by considering a network in which all users have the same product, and introduce a single new product to the market (i.e., $m = 2$). We call our new product A, and let it correspond to $i = 1$. We then measure the new product's market share as a function of the various parameters S, $\mathbf{a}$ and Q. Unless specified otherwise, in the simulations we take $a_1 = 1.1$, $a_2 = 1$,

and $Q_{ij} = 0.0025$ for $i \neq j$,[5] the seed set S is a single randomly selected node in the graph, and the process is run for $T = 30$ time steps on the Facebook network described below. We average over $k = 50$ simulations; error bars depict the standard error of the mean.

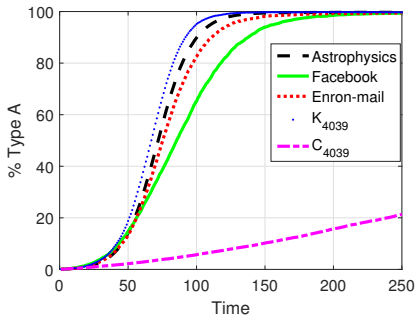
Networks. The networks we use for our simulations were collected by the Stanford Network Analysis Project and are publicly available [16]. We take the largest strongly connected component of each network for our simulations. We use a subset of the Facebook network, an ASTRO-PH collaboration network from the e-print arXiv website in which nodes are authors and there is an edge between two nodes if they are co-authors on at least one paper, and the Enron email network where nodes are email addresses internal to the Enron company and an edge represents the fact that at least one email was exchanged. We purposely select very different types of networks across which influence could propagate, and further compare against the complete graph and the cycle where the number of nodes is the same as for the Facebook graph.

Despite dramatic differences in the origin, size and connectivity of the networks, all networks appear to converge relatively fast to similar steady-states (see Figure 1(a)). There is no clear relationship between any of the graph properties and the observed mixing time, although we hypothesize that the small diameter plays a key role. Indeed, the cycle reaches the steady state at a much slower rate; exploring this further could be an interesting direction for future work. Due to space constraints, in the remainder of this section we only present the simulation results from the Facebook network dataset; the results on the other network topologies are similar and lead to the same qualitative conclusions.

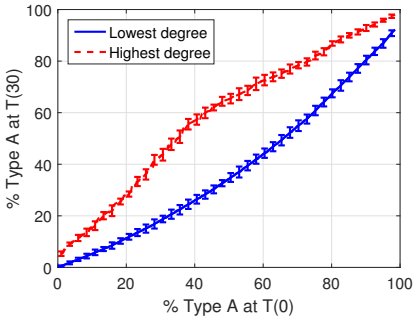
Seed Sets. As observed in Section 2, if $Q > 0$, the seed set does not have an effect on the market share in the steady state as the chain is ergodic. However, it could affect the process in the short-term. Hence, we study the effect of the seed set by varying the number of nodes it contains, and which set of nodes are selected. Several empirical studies have shown that a good heuristic for maximizing influence is to select the highest-degree nodes, while poorly performing heuristics include selecting the lowest-degree nodes (see, e.g., [12]). Hence, for a seed set of size k, we compare two possible sets: the set that contains the k highest degree vertices, and the set that contains the k lowest degree vertices, breaking ties arbitrarily.

In Figure 1(b) we observe that, roughly, increasing the number of seeds in the set only improves the market share linearly; indeed, in order to capture 50% of the market after $T = 30$ time steps, even using high-degree seeds, one would have had to start by seeding 30% of the population. In contrast, as we will see below, there are nonlinear gains observed by increasing either a_1 or Q_{12}. This suggests that, even in the short-term, increasing the number of seeds may not be the best approach. Furthermore, while optimizing the selection of seeds is indeed beneficial, it does not appear to lead to significant gains, simply altering the constant of the linear improvement rather than changing its nature.

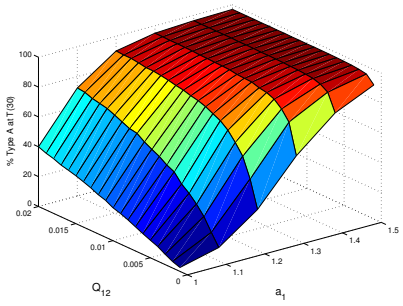
[5] The choice of parameters is inspired by the parameters we get when a fitting our model to real world datasets; see the full version of the paper.



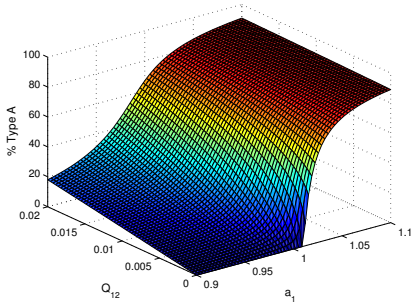
(a) The effect of the network on the con-



(b) The effect of seed selection on the



(c) The effect of a_1 and Q_{12} on the market share after $T = 30$ time steps.



(d) The effect of a_1 and Q_{12} on the market share in the limit.

Fig. 1. The effect of the various model parameters on the market share. Unless varied in the given simulation, we take $a_1 = 1.1$, $a_2 = 1$, $Q_{ij} = 0.0025$ for all $i \neq j$, and we use the Facebook network.

Product Fitness and Mutation. We then vary both Q_{12} and a_1 and measure the market share both after $T = 30$ time steps and in the steady state (see Figure 1). Here, $a_2 = 1$ and $Q_{21} = 0.01$; i.e., our competitors have an extremely aggressive and successful advertising campaign that spontaneously converts nodes. We first observe that Q_{12} has almost no effect on the steady state of the distribution. However, in the short-term, the choice of Q_{12} can result in attaining anything from 0 to the maximal market share within 30 time steps. We observe that $Q_{12} > .001$ suffices to reach close to the steady state distribution, and below this range the transition is a sharp sigmoid; e.g., increasing Q_{12} from 0.0025 to 0.005 when $a_1 = 1.1$ increases the market share by over 30%. Furthermore, we observe that even for very small differences in quality, e.g., $a_1 > 1.05 > 1 = a_2$, which corresponds to users preferring product A approximately 51.2% of the time, product A will capture almost 100% of the market in the limit. However, a_1 needs to be approximately 2.5, which corresponds to being preferred approximately 71.4% of the time, in order to achieve the same market share in the short-term ($T = 30$) when $Q_{12} = 0.001$. Again, the improvement in market share as a function of a_1 is a sigmoid, making the improvement in marketshare superlinear in a_1 when it is less than 50%. These observations suggest that it is likely to be more beneficial to improve the fitness or the advertising as opposed to the seed set in order to improve market share.

4 Conclusion and Future Work

The main conceptual contribution of this paper is to introduce a new model to study the role and interaction of the key forces of interest in online advertising. Technically, we show that our model, not only goes beyond the state-of-the-art in capturing the relevant parameters, but can also be analyzed rigorously for a large class of graphs. Combined with our empirical and numerical results on the short-term behavior of advertising campaigns on the market share, we are led to interesting predictions and take-homes.

An obvious generalization of our model would allow for *node-specific* parameters $\mathbf{a}^v$ and Q_{ij}^v. The different fitnesses/mutations can capture distinct user preferences, or targeted advertising campaigns that only display online ads to certain users. In particular, this may mean that some or all of the Q_{ij}^vs are 0 or tend to 0; this would leads to additional mathematical challenges as the Markov chain may no longer be ergodic and hence multiple stationary distributions would arise. As with other models that do not incorporate mutations [5], we expect that the initial seed set can have a larger effect in such settings as the process may converge to different steady states.

We expect that the proof techniques that we developed in this paper – in particular the interplay between stochastic and deterministic dynamical systems – might be useful not only to analyze extensions and variations of our model but also in other settings. From a technical standpoint, our theoretical results strongly relied on the minimum degree assumption; an interesting open question is whether we can prove them under the condition that the diameter of the graph is small (as may be the case in many applications).

References

1. Bass, F.M.: A new product growth model for consumer durables. Management Sciences 15(5), 215–227 (1969)
2. Benaïm, M.: Dynamics of stochastic approximation algorithms. In: Seminaire de probabilites XXXIII, pp. 1–68. Springer (1999)
3. Van den Bulte, C., Joshi, Y.V.: New product diffusion with influentials and imitators. Marketing Science 26(3), 400–421 (2007)
4. Chazelle, B.: The dynamics of influence systems. In: Foundations of Computer Science (FOCS), 2012 IEEE 53rd Annual Symposium on. pp. 311–320. IEEE (2012)
5. Díaz, J., Goldberg, L.A., Mertzios, G.B., Richerby, D., Serna, M., Spirakis, P.G.: Approximating fixation probabilities in the generalized moran process. Algorithmica 69(1), 78–91 (2014)
6. Dixit, N., Srivastava, P., Vishnoi, N.K.: A finite population model of molecular evolution: Theory and computation. Journal of Computational Biology 19(10), 1176–1202 (2012)
7. Easley, D., Kleinberg, J.: Networks, crowds, and markets. Cambridge University Press (2010)
8. Efthymiou, C., Hayes, T.P., Stefankovic, D., Vigoda, E., Yin, Y.: Convergence of MCMC and loopy BP in the tree uniqueness region for the hard-core model. In: IEEE 57th Annual Symposium on Foundations of Computer Science, FOCS 2016, 9-11 October 2016, Hyatt Regency, New Brunswick, New Jersey, USA. pp. 704–713 (2016), https://doi.org/10.1109/FOCS.2016.80

9. Galeotti, A., Goyal, S.: Influencing the influencers: a theory of strategic diffusion. The RAND Journal of Economics 40(3), 509–532 (2009)
10. Granovetter, M.: Threshold models of collective behavior. American journal of sociology 83(6), 1420–1443 (1978)
11. Herr, P.M., Kardes, F.R., Kim, J.: Effects of word-of-mouth and product-attribute information on persuasion: An accessibility-diagnosticity perspective. Journal of consumer research pp. 454–462 (1991)
12. Hinz, O., Skiera, B., Barrot, C., Becker, J.U.: Seeding strategies for viral marketing: An empirical comparison. Journal of Marketing 75(6), 55–71 (2011)
13. Iyer, G., Soberman, D., Villas-Boas, J.M.: The targeting of advertising. Marketing Sci. 24(3) (2005)
14. Jain, D., Mahajan, V., Muller, E.: An approach for determining optimal product sampling for the diffusion of a new product. Journal of Product Innovation Management 12(2), 124–135 (1995)
15. Kempe, D., Kleinberg, J., Tardos, É.: Maximizing the spread of influence through a social network. In: The ninth ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 137–146. ACM (2003)
16. Leskovec, J., Krevl, A.: SNAP Datasets. http://snap.stanford.edu/data (Jun 2014)
17. Lieberman, E., Hauert, C., Nowak, M.A.: Evolutionary dynamics on graphs. Nature 433(7023) (2005)
18. Lohtia, R., Donthu, N., Hershberger, E.K.: The impact of content and design elements on banner advertising click-through rates. Journal of advertising Research 43(04), 410–418 (2003)
19. Marsden, P.V., Friedkin, N.E.: Network studies of social influence. Socio. Methods & Research 22(1) (1993)
20. Nowak, M.A.: Evolutionary dynamics. Harvard University Press (2006)
21. Palmeri, C.: Online ad spending to pass tv spots this year, consultant says (2015), http://bloom.bg/1dJ99Zf
22. Palmeri, C.: Social network ad spending to hit $23.68 billion worldwide (2015), http://bit.ly/2fng5yU
23. Pemantle, R.: When are touchpoints limits for generalized pólya urns? Proceedings of the American Mathematical Society pp. 235–243 (1991)
24. Schelling, T.C.: Micromotives and macrobehavior. WW Norton & Company (1978)
25. Scott, D.M.: The new rules of marketing & PR: How to use social media, online video, mobile applications, blogs, news releases, and viral marketing to reach buyers directly. John Wiley & Sons (2013)
26. Shy, O.: The economics of network industries. Cambridge University Press (2001)
27. Tripathi, K., Balagam, R., Vishnoi, N.K., Dixit, N.M.: Stochastic simulations suggest that HIV-1 survives close to its error threshold. PLoS Computational Biology 8(9), e1002684 (2012)
28. Vishnoi, N.K.: The speed of evolution. In: Symp. Discret. Algorithms (SODA). pp. 1590–1601 (2015)
29. Vranica, S.: Ibm pours $100 million into ad consulting (2014), http://on.wsj.com/1dxVhzI
30. Weaver, O.: How to set social advertising goals (2015), http://bit.ly/1yAseR5
31. Wormald, N.C.: Differential equations for random processes and random graphs. The annals of applied probability pp. 1217–1235 (1995)