

Coexistence in preferential attachment networks

Tonći Antunović^{*} Elchanan Mossel[†] Miklós Z. Rácz[‡]

November 30, 2015

Abstract

We introduce a new model of competition on growing networks. This extends the preferential attachment model, with the key property that node choices evolve simultaneously with the network. When a new node joins the network, it chooses neighbours by preferential attachment, and selects its type based on the number of initial neighbours of each type. The model is analysed in detail, and in particular, we determine the possible proportions of the various types in the limit of large networks. An important qualitative feature we find is that, in contrast to many current theoretical models, often several competitors will coexist. This matches empirical observations in many real-world networks.

2010 Mathematics Subject Classification: Primary: 05C80; Secondary: 60C05, 60G99

1 Introduction

A major challenge in complex networks is understanding the interplay between the evolution of the network and the dynamical processes that take place on it. Many networks evolve dynamically, e.g., the citation graph grows every day with new papers being published, and friendships are created and broken every minute. The changes in network structure are closely related to the processes on these nodes, e.g., the content of a Facebook page is correlated with the friendship dynamics.

In the past fifteen years there have been many studies on processes on networks [6], e.g., epidemic spreading [20], evolutionary games [19], and information cascades [30]. However, most considered the network as fixed, and then studied the process of interest on static graphs. This static viewpoint hides the fact that the networks and the processes on them *coevolve*. Although the study of such coevolution was initiated over a decade ago [29], it is only recently beginning to be explored in greater depth (see [13, 16] and references therein), and thus many questions still remain. In particular, in the context of type adoption on networks, an important open problem is to understand the phenomenon of coexistence of competing types.

Here, we present a *simple* connection which couples the growth of a network and nodal dynamics. In particular, we focus on *type adoption dynamics*, where each node has a single type from a finite set of types. When a new node joins the network, both its connections to the existing nodes and its type are influenced by the current structure of the network. As a particular instance of such a general model, we consider the dynamics where the new node chooses its connections according to linear preferential attachment [5, 11, 10], and then chooses its type based on how many of its neighbours are of a certain type; see Fig. 1 for an illustration.

This model is of interest in many cases where preferential attachment is a good representation of the evolution of the network structure and where competition between types occurs as soon as the node joins the network. A natural example is the network of scientific papers linked by citations [26]. Often there are two opposing viewpoints on an issue, leading to competition between the perspectives. Another example is

^{*}University of California, Los Angeles; tantunovic@math.ucla.edu.

[†]University of Pennsylvania and University of California, Berkeley; mossel@wharton.upenn.edu; supported by NSF grant DMS 1106999 and by DOD ONR grant N000141110140.

[‡]University of California, Berkeley; racz@stat.berkeley.edu; supported by a UC Berkeley Graduate Fellowship, by NSF grant DMS 1106999 and by DOD ONR grant N000141110140.

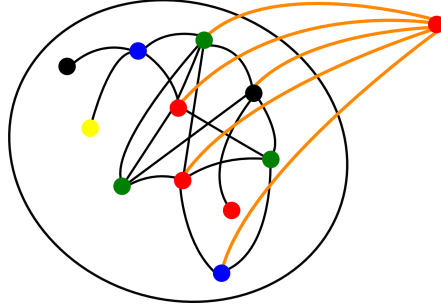


Figure 1: **Illustration of the model.** Each node in the initial graph has a colour (type) from a finite set of colours. At each time step a new node is added to the graph and connected to m existing nodes according to linear preferential attachment (here $m = 5$). When the new node joins the graph it also adopts a colour: it picks its colour according to a probability distribution which depends on the colours of its initial neighbours. See Section 1.1 for details.

that of product adoption via word-of-mouth recommendations on social networks, such as a new cell-phone user choosing a cell-phone provider based on her friends’ decisions.

A key feature of this model is the simplicity of its analysis. We explicitly calculate the possible ratios of the types in the limit of large networks. The most important property of the model is that for many settings of the parameters, none of the types dominate, which matches empirical observations in many current networks. These results thus provide a theoretical understanding of coexistence of types in preferential attachment networks. They should be compared to results on other models of competition on scale-free networks where coexistence is rarely achieved, and typically the “winner takes all” [25, 12]. (See also [1, 17] for results on related models.) The chief difference between this model and related ones in economics [2, 3], and marketing [4] is the explicit modeling of the underlying network structure.

We next describe the model and our results in more detail.

1.1 Model

For simplicity, we describe the model in the case of two types, which we refer to as red and blue colours (in the following, we use the terms “type” and “colour” interchangeably). The model naturally generalises to any number of types; see Section 3 for a description and results. The main feature of the model is that it incorporates and couples two processes: a network growing process and a type adoption process.

We consider a natural variant of the linear preferential attachment model [5, 11, 10] as the network growing process.¹ Starting from an initial graph G_0 , at each time step an additional node v is added to the graph, together with m edges connecting v to existing nodes in the graph. Each edge is chosen independently, and according to linear preferential attachment, i.e., the probability that a given edge connects v to a given existing node u is proportional to the degree of u .²

The type adoption process on the network is as follows. All nodes in the initial graph G_0 start with a type, i.e., they are either red or blue. Each additional node v receives a colour when it is added to the graph, and this colour depends on the colours of the nodes it connects to when it is added. Suppose that out of the m edges connecting the new node v to existing ones, exactly k connect to a red node. Then, conditioned on this event, v becomes red with probability p_k and blue with probability $1 - p_k$. The probabilities $p_k \in [0, 1]$, where $0 \leq k \leq m$, are parameters of the model, and can capture a wide range of behaviour. A natural choice is the *linear model*, when $p_k = k/m$ for all k . However, *nonlinear models*, when $p_k \neq k/m$ for some k , can capture diminishing and increasing returns, and even more complex behaviour.

¹The preferential attachment model was introduced by Barabási and Albert [5]; however, as pointed out by Bollobás and Riordan [11], the definition in [5] left some details of the process up to interpretation. There are several natural ways in which to define the process precisely: one option is the model introduced in [11], while here we consider a slightly different variant (the “independent model” in [10]).

²In particular, if $m > 1$ then v might connect to an existing node with multiple edges, resulting in a *multigraph*. Multiple edges between two nodes can be thought of as stronger ties between the agents represented by the nodes.

1.2 Results

We are interested in the fraction of nodes of each type. This corresponds to the fraction of scientific papers sharing a viewpoint on a given topic or to the fraction of individuals using a given company's product, i.e., the company's market share. Our main results characterise the possible limiting fractions of the colours as the size of the network goes to infinity. The results thus provide a complete phase diagram of the asymptotic behaviour of the process; see Fig. 4 for an illustration. These show that if there is negative reinforcement, then the types coexist, while if there is strong positive reinforcement, then the winner takes all. Our results precisely determine where the transition happens, and show that for most parameter values the types coexist.

To describe our results precisely, we introduce some notation. Let G_n denote the graph when n nodes have been added to the initial graph G_0 , let A_n and B_n , respectively, denote the number of red and blue nodes, respectively, in G_n , and let $a_n := \frac{A_n}{A_n+B_n}$ and $b_n := \frac{B_n}{A_n+B_n}$ denote the corresponding normalised fractions. Furthermore, let X_n (resp., Y_n) denote the sum of the degrees of red (resp., blue) nodes in G_n , and let $x_n := \frac{X_n}{X_n+Y_n}$ and $y_n := \frac{Y_n}{X_n+Y_n}$ denote the normalised fractions. We are primarily interested in the asymptotic proportion of red and blue nodes, i.e., in the limits $\lim_{n \rightarrow \infty} a_n$ and $\lim_{n \rightarrow \infty} b_n = 1 - \lim_{n \rightarrow \infty} a_n$.

As we shall see, a key role in the asymptotic behaviour of the process is played by the polynomial

$$(1) \quad P(z) = \frac{1}{2} \sum_{k=0}^m \binom{m}{k} z^k (1-z)^{m-k} \left(p_k - \frac{k}{m} \right),$$

and in particular its zero set, denoted by $Z_P := \{z \in [0, 1] : P(z) = 0\}$. This is because, as we shall show, $\{a_n\}_{n \geq 0}$ behaves approximately like the solution to a stochastic version of the ordinary differential equation (ODE) $dz/dt = P(z)$, and thus intuitively the trajectory of $\{a_n\}_{n \geq 0}$ should approximate the trajectory $\{z(t)\}_{t \geq 0}$ of this ODE.

The following two theorems confirm this intuition. There is an important distinction between the *linear model* and *nonlinear models*, which is due to the fact that in the linear model the polynomial P is identically zero and thus $Z_P = [0, 1]$, while in nonlinear models the zero set Z_P is a finite set.

Theorem 1.1 (Linear model). *Suppose that $p_k = k/m$ for all $0 \leq k \leq m$, and that $X_0, Y_0 > 0$. Then a_n converges almost surely as $n \rightarrow \infty$. Furthermore, the limiting distribution of $a := \lim_{n \rightarrow \infty} a_n$ has full support on the interval $[0, 1]$, has no atoms (i.e., for every $z \in [0, 1]$, $\mathbb{P}(a = z) = 0$), and depends only on X_0, Y_0 , and m .*

See Figure 2 for empirical histograms in the linear model with various initial parameters and various values of m , which show that a wide variety of limiting behaviours is possible.

Theorem 1.2 (Nonlinear models). *Suppose that $p_k \neq k/m$ for some $0 \leq k \leq m$, and that $X_0, Y_0 > 0$. Then a_n converges almost surely as $n \rightarrow \infty$. Furthermore, the limit is a point in the finite set Z_P .*

In nonlinear models we thus know that the asymptotic proportion of red nodes is contained in the finite zero set Z_P . But which points $z \in Z_P$ arise as the limiting proportion with positive probability? This depends on the behaviour of the polynomial around the zero $z \in Z_P$. Intuitively, since $\{a_n\}_{n \geq 0}$ is a stochastic system, we expect that stable trajectories of the ODE $dz/dt = P(z)$ should appear, but unstable trajectories should not. This intuition is confirmed and formalized in the following three theorems.

Theorem 1.3 (Nonlinear models, stable equilibria). *Suppose that $p_k \neq k/m$ for some $0 \leq k \leq m$, and that $X_0, Y_0 > 0$. Suppose that $z \in Z_P \cap (0, 1)$ is such that there exists an $\varepsilon > 0$ such that $P > 0$ on $(z - \varepsilon, z)$ and $P < 0$ on $(z, z + \varepsilon)$. Then $\mathbb{P}(\lim_{n \rightarrow \infty} a_n = z) > 0$, i.e., a_n converges to z with positive probability. Similarly, if $0 \in Z_P$ and $P < 0$ on $(0, \varepsilon)$, or if $1 \in Z_P$ and $P > 0$ on $(1 - \varepsilon, 1)$, then there is a positive probability of convergence of a_n to 0 or 1, respectively.*

Theorem 1.4 (Nonlinear models, unstable equilibria). *Suppose that $p_k \neq k/m$ for some $0 \leq k \leq m$, and that $X_0, Y_0 > 0$. Suppose that $z \in Z_P \cap (0, 1)$ is such that there exists an $\varepsilon > 0$ such that $P < 0$ on $(z - \varepsilon, z)$ and $P > 0$ on $(z, z + \varepsilon)$. Then $\mathbb{P}(\lim_{n \rightarrow \infty} a_n = z) = 0$. Similarly, if $0 \in Z_P$ and $P > 0$ on $(0, \varepsilon)$, or if $1 \in Z_P$ and $P < 0$ on $(1 - \varepsilon, 1)$, then the probability of convergence of a_n to 0 or 1, respectively, is zero.*

Theorem 1.5 (Nonlinear models, touchpoints). *Suppose that $p_k \neq k/m$ for some $0 \leq k \leq m$, and that $X_0, Y_0 > 0$. Suppose that $z \in Z_P \cap (0, 1)$ is such that there exists an $\varepsilon > 0$ such that P is either strictly positive or strictly negative on the union of the intervals $(z - \varepsilon, z)$ and $(z, z + \varepsilon)$. Then $\mathbb{P}(\lim_{n \rightarrow \infty} a_n = z) > 0$.*

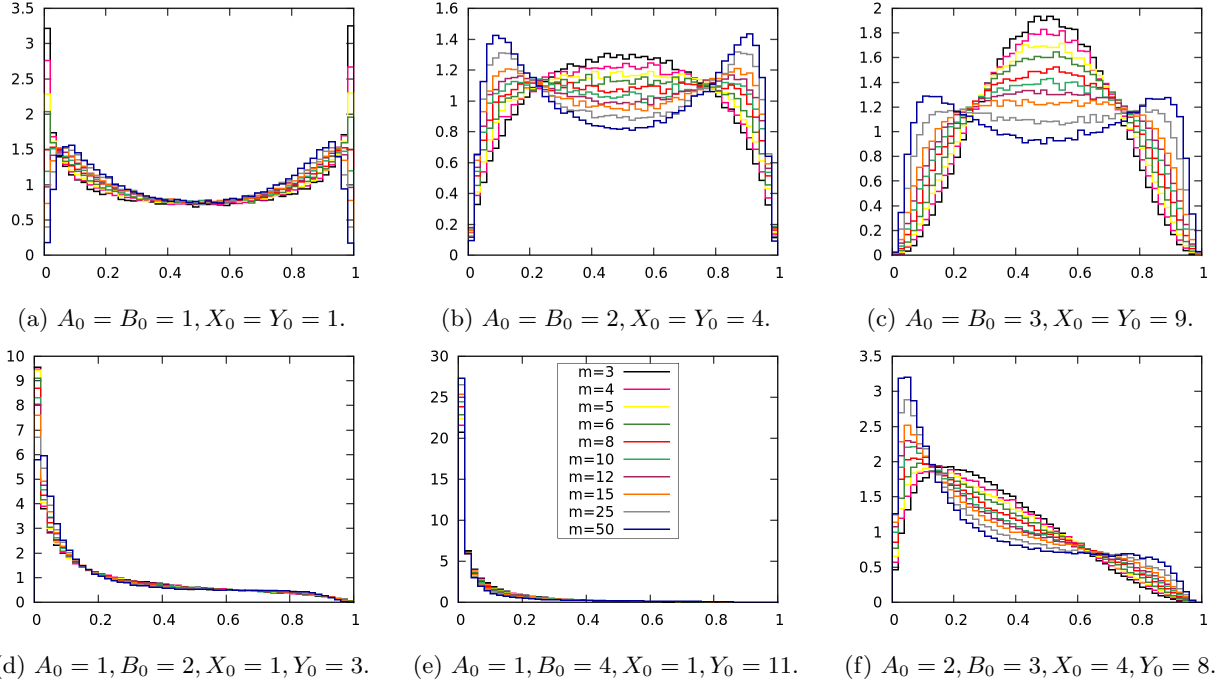


Figure 2: **Empirical histograms of a_n in the linear model** for $n = 10^5$, from 2×10^5 simulations. Each subfigure has different initial parameters (see subcaptions), and in each case empirical histograms for ten different values of m are plotted. See Fig. 2e for the key to all plots.

See Figure 3 for an illustration of the polynomial P for various values of the parameters $\{p_k\}_{0 \leq k \leq m}$, and what the various limiting proportions can be in each case.

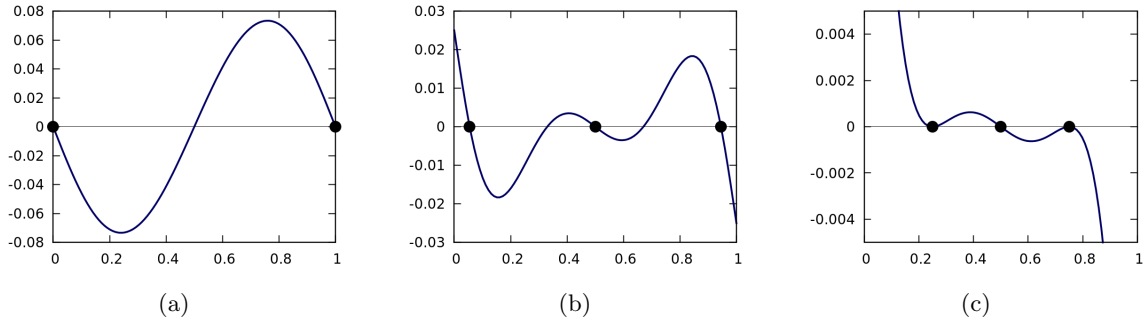


Figure 3: **Examples of the polynomial P and possible limiting proportions.** In each case there is no bias towards either colour, i.e., $p_k + p_{m-k} = 1$ for all $0 \leq k \leq m$. (a) Majority choice: $p_k = 1$ if $k > m/2$ and $p_k = 0$ otherwise ($m = 5$ in the figure). The possible limits are 0 and 1, i.e., the winner takes all. (b) The parameters here are: $m = 9, p_5 = p_6 = 0.5, p_7 = p_8 = p_9 = 0.95$. Such an example is plausible if the strength of the signal from the neighbours matters: if $3 \leq k \leq 6$, then the signal towards either colour is weak, so just flip a fair coin to choose, but if $0 \leq k \leq 2$ or $7 \leq k \leq 9$ then there is a strong signal towards one of the colours, so pick that colour with probability close to 1. In this example the possible limits are: $z_1 \approx 0.055, z_2 = 0.5, z_3 \approx 0.945$, and there are also two zeros of P which cannot be limits. (c) This is an example where P has touchpoints. The parameters are: $m = 6, p_4 = 1031/1710, p_5 = p_6 = 35/38$, and the two touchpoints are at $z_1 = 1/4$ and $z_2 = 3/4$. Both of these, as well as $z_3 = 1/2$, can be limits.

The theorems above provide a complete phase diagram of the asymptotic behaviour of the process in the case of two types. To illustrate this, see Figure 4, which shows phase diagrams for $m = 3$ and $m = 4$ when

there is no bias towards either colour, i.e., when $p_k + p_{m-k} = 1$ for all $0 \leq k \leq m$. This condition implies that $P(z) = -P(1-z)$ and so $1/2 \in Z_P$, but $1/2$ need not be a limit point (see Fig. 3).

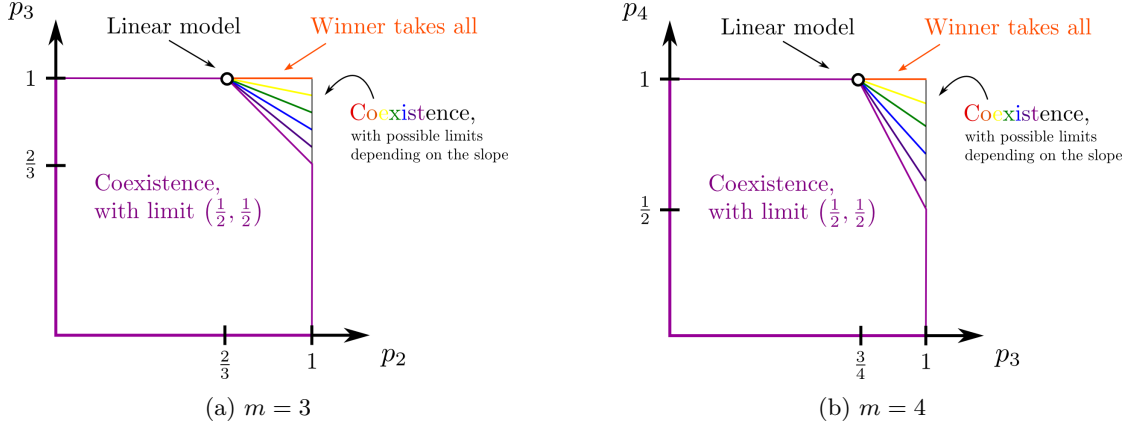


Figure 4: **Phase diagrams when there is no bias towards either colour/type**, i.e., when $p_k + p_{m-k} = 1$ for all $0 \leq k \leq m$. Let $q_k := p_k - k/m$. (a) If $q_2 < 0$ or if $q_2 + q_3 \leq 0$ and $q_3 < 0$, then $\lim_{n \rightarrow \infty} a_n = 1/2$, i.e., in this case the network is split evenly among the two types in the limit. The linear model is the case of $q_2 = q_3 = 0$. Finally, if $q_2 + q_3 > 0$ and $q_2 > 0$, then let $\alpha = -\frac{q_3}{q_2} \in [0, 1]$; the possible limits of a_n are then $\frac{1}{2} \pm \frac{1}{2} \sqrt{\frac{3-3\alpha}{3+\alpha}}$. In particular, when $\alpha = 0$ then the winner takes all, and if $\alpha \in (0, 1)$, then the two types coexist in the limit. (b) This is similar to (a). Here, if $2q_3 + q_4 > 0$ and $q_4 > 0$, then let $\beta = -\frac{q_4}{q_3} \in [0, 2]$; the possible limits of a_n are then $\frac{1}{2} \pm \frac{1}{2} \sqrt{\frac{2-\beta}{2+\beta}}$.

Coexistence. In particular, the theorems above show that in many cases the two colours coexist in the limit. Indeed, since $P(0) = \frac{1}{2}p_0$ and $P(1) = \frac{1}{2}(p_m - 1)$, $p_0 = 0$ or $p_m = 1$ is necessary for one of the colours to asymptotically take over the network. Whenever $p_0 > 0$ and $p_m < 1$ the two colours coexist in the limit, and thus the model provides a theoretical understanding of coexistence in preferential attachment networks.

A natural extension of the model is to consider more than two colours. For clarity of presentation, we postpone the discussion of this until later: see Section 3 for a description of the model with many colours and the corresponding results and conjectures.

It is also natural to consider variants of the model with different underlying network evolution models. For instance, consider the affine preferential attachment model, where a given edge connects the new node v to a given existing node u with probability proportional to the degree of u plus some fixed constant c which is greater than -1 . This gives rise to a random graph with power-law degree distribution with exponent $3 + c$. The linear preferential attachment model is the special case with $c = 0$. The type adoption process with affine preferential attachment is slightly more involved, so to keep the exposition simple we do not go into details. However, the same techniques apply, and we get the same results as above for $c \geq 0$. For $c \in (-1, 0)$ the behavior is more delicate, and we leave it as an exercise to the reader to flesh out the details. See Section 4 for further discussion on the affine preferential attachment model. Another natural variant is to consider uniform attachment—this is even simpler than preferential attachment, and the same results apply. We leave the understanding of other network evolution models for future work.

1.3 Outline of the paper

First, in Section 2 we prove the results described in Section 1.2. Then in Section 3 we study the case of three or more types, and finally we conclude with open questions and directions for future research in Section 4.

2 Proofs

This section contains the proofs of our main results described in Section 1.2, and is structured as follows. First, in Section 2.1 we show how the asymptotic behaviour of $\{a_n\}_{n \geq 0}$ is the same as that of the sum-of-degrees process $\{x_n\}_{n \geq 0}$, which is more convenient to study, as it is a Markov process. Then in Section 2.2 we study the linear model and prove Theorem 1.1. Next in Section 2.3 we recall results from the theory of stochastic approximation processes, and finally in Section 2.4 we prove our results concerning nonlinear models.

2.1 Reduction to the sum-of-degrees process

To understand the process $\{A_n\}_{n \geq 0}$ (and thus the normalised process $\{a_n\}_{n \geq 0}$), it is more convenient to study the time evolution of the sum of the degrees of each type. The reason for this is that the process $\{A_n\}_{n \geq 0}$ is not a Markov process, but the joint process $\{(A_n, X_n)\}_{n \geq 0}$ is indeed Markov. It evolves as follows. Given (A_n, X_n) , u_{n+1} is drawn from the binomial distribution with parameters m and x_n . Subsequently, I_{n+1} is drawn from the Bernoulli distribution with parameter $p_{u_{n+1}}$. We then have

$$(2) \quad A_{n+1} = A_n + I_{n+1},$$

$$(3) \quad X_{n+1} = X_n + u_{n+1} + mI_{n+1}.$$

The following lemma tells us that in order to understand the asymptotic behaviour of $\{a_n\}_{n \geq 0}$, it is enough to understand the asymptotic behaviour of $\{x_n\}_{n \geq 0}$. Consequently, in the following we analyse the latter, as this is a Markov process.

Lemma 2.1. *Suppose $\{x_n\}_{n \geq 0}$ converges almost surely (a.s.) and let $x := \lim_{n \rightarrow \infty} x_n$ denote the limit. If $P(x) = 0$ a.s., then $\{a_n\}_{n \geq 0}$ converges a.s. as well, and $\lim_{n \rightarrow \infty} a_n = x$ a.s.*

Proof. Let \mathcal{F}_n denote the filtration of the process until time n . Given \mathcal{F}_n , the probability that the node added at time $n+1$ is red is

$$\begin{aligned} \mathbb{P}(A_{n+1} - A_n = 1 | \mathcal{F}_n) &= \sum_{k=0}^m \binom{m}{k} x_n^k (1-x_n)^{m-k} p_k = x_n + \sum_{k=0}^m \binom{m}{k} x_n^k (1-x_n)^{m-k} q_k \\ &= x_n + 2P(x_n) =: f(x_n), \end{aligned}$$

where $q_k = p_k - k/m$. Thus $\mathbb{E}(A_{n+1} - A_n | \mathcal{F}_n) = f(x_n)$. Define $M_n = A_n - A_0 - \sum_{i=0}^{n-1} f(x_i)$, with initial condition $M_0 = 0$. The previous calculation tells us that $\{M_n\}_{n \geq 0}$ is a martingale with respect to the filtration \mathcal{F}_n . Moreover, this martingale has bounded increments since $M_{i+1} - M_i = A_{i+1} - A_i - f(x_i) \in [-1, 1]$, and thus $\lim_{n \rightarrow \infty} M_n/n = 0$ a.s.

Let $x := \lim_{n \rightarrow \infty} x_n$. Since $P(x) = 0$, we have $f(x) = x$. Since f is continuous, we have $\lim_{n \rightarrow \infty} f(x_n) = f(x) = x$ a.s., and thus the Cesàro mean of the sequence $\{f(x_n)\}_{n \geq 0}$ also converges to the same limit: $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(x_i) = x$ a.s. The claim then follows from the fact that $M_n/n = \frac{A_n}{n} - \frac{A_0}{n} - \frac{1}{n} \sum_{i=0}^{n-1} f(x_i)$ and $\lim_{n \rightarrow \infty} (a_n - \frac{A_n}{n}) = 0$. \square

2.2 Linear model

Proof of Theorem 1.1. In the linear model, when $p_k = \frac{k}{m}$ for all $k = 0, 1, \dots, m$, we have that $P \equiv 0$, and thus $\mathbb{E}(X_{n+1} - X_n | \mathcal{F}_n) = 2mx_n$. Since $x_{n+1} - x_n = \frac{X_{n+1} - X_n - 2mx_n}{S_{n+1}}$, it follows that $\mathbb{E}(x_{n+1} - x_n | \mathcal{F}_n) = 0$, i.e., $\{x_n\}_{n \geq 0}$ is a martingale. Since it is also bounded, it converges almost surely. Lemma 2.1 then implies that $\{a_n\}_{n \geq 0}$ converges a.s. as well, and $\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} x_n$ a.s.

We use a variance argument to show that the distribution of $x := \lim_{n \rightarrow \infty} x_n$ has full support on $[0, 1]$.

First note that $(x_{n+1} - x_n)^2 = \left(\frac{X_{n+1} - X_n - 2mx_n}{S_0 + 2m(n+1)} \right)^2 \leq \frac{1}{(n+1)^2}$, and consequently for any n_0 we have

$$(4) \quad \mathbb{E} \left((x - x_{n_0})^2 \middle| \mathcal{F}_{n_0} \right) = \lim_{n \rightarrow \infty} \mathbb{E} \left((x_n - x_{n_0})^2 \middle| \mathcal{F}_{n_0} \right) = \sum_{j=n_0}^{\infty} \mathbb{E} \left((x_{j+1} - x_j)^2 \middle| \mathcal{F}_{n_0} \right) \leq \sum_{j=n_0}^{\infty} \frac{1}{(j+1)^2} \leq \frac{1}{n_0}.$$

Now let $(r, r + \varepsilon) \subset (0, 1)$ be any fixed interval. Our goal is to show that $\mathbb{P}(x \in (r, r + \varepsilon)) > 0$. Let n_0 be an integer such that $n_0 \geq \frac{18}{\varepsilon^2}$ and $\mathbb{P}(x_{n_0} \in (r + \frac{\varepsilon}{3}, r + \frac{2\varepsilon}{3})) > 0$ (this is possible since for large enough n_0 there exists a sequence of events such that $x_{n_0} \in (r + \frac{\varepsilon}{3}, r + \frac{2\varepsilon}{3})$). Now condition on this event; (4) implies that

$$\mathbb{E} \left((x - x_{n_0})^2 \mid x_{n_0} \in \left(r + \frac{\varepsilon}{3}, r + \frac{2\varepsilon}{3} \right) \right) \leq \frac{1}{n_0} \leq \frac{\varepsilon^2}{18},$$

which in turn implies that $\mathbb{P}(|x - x_{n_0}| \leq \frac{\varepsilon}{3} \mid x_{n_0} \in (r + \frac{\varepsilon}{3}, r + \frac{2\varepsilon}{3})) \geq \frac{1}{2}$. We conclude that

$$\mathbb{P}(x \in (r, r + \varepsilon)) \geq \mathbb{P} \left(|x - x_{n_0}| \leq \frac{\varepsilon}{3} \mid x_{n_0} \in \left(r + \frac{\varepsilon}{3}, r + \frac{2\varepsilon}{3} \right) \right) \mathbb{P} \left(x_{n_0} \in \left(r + \frac{\varepsilon}{3}, r + \frac{2\varepsilon}{3} \right) \right) > 0.$$

Finally, showing that the distribution of x has no atoms can be done by adapting arguments by Pemantle [21]. First, let us describe how the process $\{x_n\}_{n \geq 0}$ is related to time-dependent Pólya urn processes that Pemantle studies in [21].

Time-dependent Pólya urn processes are generalisations of the classical Pólya urn process, where the number of balls added to the urn is allowed to vary with time. Although $\{x_n\}_{n \geq 0}$ is not a time-dependent Pólya urn process, the following slight modification of the preferential attachment process does give a time-dependent Pólya urn process. When adding a new node v to the graph $G_n = (V_n, E_n)$, add its m neighbours one by one, and after adding each neighbour, *update* the degree of the neighbour. Let \tilde{X}_n denote the sum of the degrees of red nodes at time n in this model. Consider also a time-dependent Pólya urn process $\{Z_n\}_{n \geq 0}$ where at times $t \neq 0 \pmod m$ a single ball is added to the urn, and at times $t = 0 \pmod m$ the number of balls added to the urn is $m + 1$. It can be seen that if $\tilde{X}_0 = Z_0$, then \tilde{X}_n and Z_{mn} have the same distribution. Thus Pemantle's results [21, Theorem 3, Theorem 4] apply directly and show that the distribution of $\lim_{n \rightarrow \infty} \tilde{x}_n$ (this limit exists a.s.) has no atoms.

Since our setting is close to Pemantle's original setting, we only sketch the proof that the distribution of x has no atoms, and leave the details to the reader.

To show that the distribution of x has no atoms on $(0, 1)$, we can adapt the variance arguments of [21, Theorem 3]. Fix $r \in (0, 1)$. Suppose on the contrary that $\mathbb{P}(x = r) > 0$. Then for every $\varepsilon > 0$ there exists n_0 and some event $\mathcal{A} \in \mathcal{F}_{n_0}$ having positive probability such that $\mathbb{P}(x_n \rightarrow r \mid \mathcal{A}) \geq 1 - \varepsilon$; in fact, n_0 can be as large as desired. Define $c := \frac{r(1-r)^{m/2}}{10 \times 2^{m/2}}$ and let $N = \max \left\{ \frac{S_0}{m}, \frac{2}{c^2 \min\{r, 1-r\}} \right\}$. One can then show, via variance arguments, the following two inequalities. First, for every $n \geq N$,

$$\mathbb{P} \left(\sup_{k \geq n} |x_k - r| \geq \frac{c}{\sqrt{n}} \mid \mathcal{F}_n \right) \geq \frac{1}{2}.$$

Second, defining $\mathcal{B} = \left\{ |x_n - r| \geq \frac{c}{\sqrt{n}} \right\}$, we have that for every $n \geq N$,

$$\mathbb{P} \left(\inf_{k \geq n} |x_k - r| \geq \frac{c}{2\sqrt{n}} \mid \mathcal{F}_n, \mathcal{B} \right) \geq \frac{c^2}{16}.$$

Putting these together we have that for every $n \geq N$, the probability given \mathcal{F}_n is at least $\frac{c^2}{32}$ that some x_{n+k} will be at least $\frac{c}{\sqrt{n}}$ away from r and no subsequent x_{n+k+l} will ever return to the interval $\left[r - \frac{c}{2\sqrt{n}}, r + \frac{c}{2\sqrt{n}} \right]$. This contradicts our initial assumption and so $\mathbb{P}(x = r) = 0$.

To show that the distribution of x has no atoms at 0 and 1, we can adapt the arguments of [21, Theorem 4]. The main idea is a domination argument. Let $\{v_n\}_{n \geq 0}$ be the Pólya urn process where at each time step $2m$ balls are added to the urn, and let $v_0 = x_0$. Then the distribution of x_n can be dominated by the distribution of v_n , in the sense that $\mathbb{E}(h(x_n)) \leq \mathbb{E}(h(v_n))$ for every continuous bounded convex function h . In other words, x_n is smaller than v_n in the convex order [28]. Since the limiting distribution of $\{v_n\}_{n \geq 0}$ is a beta distribution, which does not have an atom at zero, one can then take $h_\varepsilon(x) := \max\{0, 2 - x/\varepsilon\}$ and let $\varepsilon \rightarrow 0$ to conclude that the distribution of x cannot have an atom at zero either. We refer the reader to [21, Theorem 4] for more details. See also the proof of Theorem 1.4 for the endpoints in Section 2.4. \square

2.3 Stochastic approximation processes

The key observation in the analysis of the asymptotic behaviour of $\{x_n\}_{n \geq 0}$ is that it is a stochastic approximation process. Stochastic approximation was introduced in 1951 by Robbins and Monro [27], whose goal was to approximate the root of an unknown function via evaluation queries that are necessarily noisy. There has been much follow-up research, see, e.g., the monograph by Nevelson and Hasminskii [18]. The setup of stochastic approximation arises naturally in the study of Pólya urn processes; see the survey [24] for details. In particular, we use results of Hill, Lane and Sudderth [14], who studied generalised (nonlinear) Pólya urn processes, and we also use subsequent refinements by Pemantle [22, 23]. We state the main theorems here and refer to the original papers for more details; see also the survey [24]. Stochastic approximation results in higher dimensions will be discussed in Section 3.

Let $\{Z_n\}_{n \geq 0}$ be a stochastic process in \mathbb{R} adapted to a filtration $\{\mathcal{F}_n\}$. Suppose that it satisfies

$$(5) \quad Z_{n+1} - Z_n = \frac{1}{n} (F(Z_n) + \xi_{n+1} + R_n),$$

where $F : \mathbb{R} \rightarrow \mathbb{R}$, $\mathbb{E}(\xi_{n+1} | \mathcal{F}_n) = 0$, and the remainder terms $R_n \in \mathcal{F}_n$ go to zero and also satisfy $\sum_{n=1}^{\infty} n^{-1} |R_n| < \infty$ almost surely. Such a process is known as a *stochastic approximation process*.

Intuitively, trajectories of a stochastic approximation process $\{Z_n\}_{n \geq 0}$ should approximate the trajectories $\{Z(t)\}_{t \geq 0}$ of the corresponding ODE $dZ/dt = F(Z)$. Moreover, since $\{Z_n\}_{n \geq 0}$ is a stochastic system, we expect that stable trajectories of the ODE should appear, but unstable trajectories should not. This intuition is confirmed and formalized in the following statements (quoted from the survey [24]); for proofs and more details see the papers cited above.

Theorem 2.2 (Convergence to the zero set of F). *Suppose that $\{Z_n\}$ is a stochastic approximation process and that $\mathbb{E}(\xi_{n+1}^2 | \mathcal{F}_n) \leq K$ for some finite constant K . If F is bounded and continuous, then Z_n converges almost surely to the zero set of F .*

Theorem 2.3 (Convergence to stable equilibria). *Suppose that $\{Z_n\}$ is a stochastic approximation process with a bounded and continuous F , and that $\mathbb{E}(\xi_{n+1}^2 | \mathcal{F}_n) \leq K$ for some finite constant K . Suppose that there is a point z and an $\varepsilon > 0$ with $F(z) = 0$, $F > 0$ on $(z - \varepsilon, z)$ and $F < 0$ on $(z, z + \varepsilon)$. Then $\mathbb{P}(Z_n \rightarrow z) > 0$. Similarly, when $F : [0, 1] \rightarrow \mathbb{R}$, if $F(0) = 0$ and $F < 0$ on $(0, \varepsilon)$ or if $F(1) = 0$ and $F > 0$ on $(1 - \varepsilon, 1)$, then there is a positive probability of convergence to 0 or 1, respectively.*

Theorem 2.4 (Nonconvergence to unstable equilibria). *Suppose that $\{Z_n\}$ is a stochastic approximation process with a bounded and continuous F . Suppose that there is a point $z \in (0, 1)$ and an $\varepsilon > 0$ with $F(z) = 0$, $F < 0$ on $(z - \varepsilon, z)$ and $F > 0$ on $(z, z + \varepsilon)$. Let $x^+ = \max\{x, 0\}$ and $x^- = -\min\{x, 0\}$ denote the positive and negative parts of x , respectively. Suppose further that when $Z_n \in (z - \varepsilon, z + \varepsilon)$ then $\mathbb{E}(\xi_{n+1}^+ | \mathcal{F}_n)$ and $\mathbb{E}(\xi_{n+1}^- | \mathcal{F}_n)$ are bounded above and below by positive constants depending only on ε . Then $\mathbb{P}(Z_n \rightarrow z) = 0$.*

Pemantle studied the case of touchpoints for generalised (nonlinear) Pólya urn processes in [23]. His proof extends to the following result.

Theorem 2.5 (Convergence to touchpoints). *Suppose that $\{Z_n\}$ is a stochastic approximation process with a bounded and continuously differentiable F , and that $|\xi_n| \leq K$ a.s. for some finite constant K . Suppose that $z \in Z_P$ is a touchpoint, i.e., there exists an $\varepsilon > 0$ such that either $F > 0$ on $(z - \varepsilon, z) \cup (z, z + \varepsilon)$ or $F < 0$ on $(z - \varepsilon, z) \cup (z, z + \varepsilon)$. Then $\mathbb{P}(Z_n \rightarrow z) > 0$.*

2.4 Nonlinear models

We first show that $\{x_n\}_{n \geq 0}$ is a stochastic approximation process (i.e., that it satisfies (5)) with the function P as in (1). Subsequently we show how this implies our results in Section 1.2 using the results described in Section 2.3.

Lemma 2.6. *The process $\{x_n\}_{n \geq 0}$ is a stochastic approximation process with the function $F = P$ as in (1). Furthermore, the noise term ξ_n is bounded: $|\xi_n| \leq 2$ for all $n \geq 1$.*

Proof. From (3) we have that the conditional expectation of $X_{n+1} - X_n$ is:

$$\mathbb{E}(X_{n+1} - X_n | \mathcal{F}_n) = \sum_{k=0}^m \binom{m}{k} x_n^k (1 - x_n)^{m-k} (k + mp_k) = 2mx_n + 2mP(x_n).$$

One can check that $x_{n+1} - x_n = \frac{X_{n+1} - X_n - 2mx_n}{S_{n+1}}$ and consequently $\mathbb{E}(x_{n+1} - x_n | \mathcal{F}_n) = \frac{2m}{S_{n+1}}P(x_n)$, with P as in (1). We can then write $\{x_n\}_{n \geq 0}$ as a stochastic approximation process as claimed in the statement of the lemma, i.e., we can write

$$x_{n+1} - x_n = \frac{1}{n} (P(x_n) + \xi_{n+1} + R_n)$$

with appropriately defined ξ_{n+1} and R_n . Define ξ_{n+1} as

$$(6) \quad \xi_{n+1} = n(x_{n+1} - x_n - \mathbb{E}(x_{n+1} - x_n | \mathcal{F}_n)).$$

The remainder term R_n can then be written as

$$R_n = -\frac{S_0 + 2m}{S_0 + 2m(n+1)}P(x_n).$$

Clearly $R_n \in \mathcal{F}_n$. Let us now show that $\sum_{n=1}^{\infty} n^{-1} |R_n| < \infty$. A crude bound on P is $|P(t)| \leq \frac{1}{2} \sum_{k=0}^m \binom{m}{k} |p_k - k/m| t^k (1-t)^{m-k} \leq \frac{1}{2} \sum_{k=0}^m \binom{m}{k} t^k (1-t)^{m-k} = \frac{1}{2}$. Therefore $|R_n| \leq \frac{1}{2} \frac{S_0 + 2m}{S_0 + 2m(n+1)}$, so indeed we have $\sum_{n=1}^{\infty} n^{-1} |R_n| < \infty$.

Finally, to bound the noise term, notice that $|x_{n+1} - x_n| = \left| \frac{X_{n+1} - X_n - 2mx_n}{S_0 + 2m(n+1)} \right| \leq \frac{2m}{2m(n+1)} = \frac{1}{n+1}$. Then using (6) and the triangle inequality, we get that $|\xi_n| \leq 2$. \square

The results in Section 1.2 now follow. First, note that Lemma 2.1 implies that it is enough to show the claims in Theorems 1.2, 1.3, 1.4, and 1.5 for the process $\{x_n\}_{n \geq 0}$ (instead of for the process $\{a_n\}_{n \geq 0}$).

Proof of Theorem 1.2. This follows directly from Lemma 2.6 and Theorem 2.2. \square

Proof of Theorem 1.3. This follows directly from Lemma 2.6 and Theorem 2.3. \square

The proof of Theorem 1.4 is more involved. This is in line with related work in the literature, where conditions for nonconvergence to unstable equilibria are more difficult to find than similar results for convergence to stable equilibria (see [24] for a discussion). Recall the proof of Theorem 1.1, where we showed that the limiting distribution in the linear model has no atoms: we used a variance argument for points in $(0, 1)$, and a domination argument for the endpoints 0 and 1. Our proof of Theorem 1.4 follows similar lines.

We first proceed by proving Theorem 1.4 for points $z \in (0, 1) \cap Z_P$. Intuitively, the process has sufficient noise which prevents it from converging to z . The following lemma is key to bounding the noise of the process from below. Its proof is simple when $p_0 < 1$ and $p_m > 0$; the proof is only lengthy because it deals with the special cases when $p_0 = 1$ or $p_m = 0$.

Lemma 2.7. *Suppose that the parameters $\{p_k\}_{0 \leq k \leq m}$ do not fall into one of the following three cases: (a) $p_k = 0$ for all $0 \leq k \leq m$; (b) $p_k = 1$ for all $0 \leq k \leq m$; (c) $m = 1$, $p_0 = 1$, $p_1 = 0$. Suppose that $z \in (0, 1) \cap Z_P$. Then there exist integers k_1 and k_2 such that $k_1 < 2mz < k_2$ and, if $x_n \in (\delta, 1 - \delta)$ for some $\delta > 0$, then the probabilities $\mathbb{P}(X_{n+1} - X_n = k_1 | \mathcal{F}_n)$ and $\mathbb{P}(X_{n+1} - X_n = k_2 | \mathcal{F}_n)$ are bounded away from zero by a positive function of δ and the parameters $\{p_k\}_{0 \leq k \leq m}$.*

Proof. In the following we always assume that $x_n \in (\delta, 1 - \delta)$. If $p_0 < 1$ then we can choose $k_1 = 0$ since $\mathbb{P}(X_{n+1} - X_n = 0 | \mathcal{F}_n) = (1 - x_n)^m (1 - p_0) \geq \delta^m (1 - p_0)$. Similarly, if $p_m > 0$ then we can choose $k_2 = 2m$, since $\mathbb{P}(X_{n+1} - X_n = 2m | \mathcal{F}_n) = x_n^m p_m \geq \delta^m p_m$. The rest of the proof deals with the cases when either $p_0 = 1$ or $p_m = 0$.

First consider the case when $p_0 > 0$ and $p_1 = p_2 = \dots = p_m = 0$. In this case $P(s) = \frac{1}{2} [p_0(1-s)^m - s]$, which is decreasing in $[0, 1]$, so it has a single zero in $(0, 1)$. In fact, $P(1/2) < 0$, so the single zero of P in $(0, 1)$ is in $(0, 1/2)$, and thus we can take $k_2 = m$. If $p_0 < 1$ then we can take $k_1 = 0$ as described above.

Finally, if $p_0 = 1$ and $m > 2$, then we can take $k_1 = 1$. This is because the zero of P in $(0, 1)$ is in $(\frac{1}{2m}, \frac{1}{2})$, which follows from the fact that $P(\frac{1}{2m}) > 0$. The case when $p_0 = p_1 = \dots = p_{m-1} = 1$ and $p_m < 1$ follows similarly.

Now we can assume that there exist $1 \leq i \leq m$ and $0 \leq j \leq m-1$ such that $p_i > 0$ and $p_j < 1$. This implies that $\mathbb{P}(X_{n+1} - X_n = j | \mathcal{F}_n) \geq \delta^m (1 - p_j) > 0$ and $\mathbb{P}(X_{n+1} - X_n = m + i | \mathcal{F}_n) \geq \delta^m p_i > 0$. Thus if $z = 1/2$ then we can take $k_1 = j$ and $k_2 = m + i$. If $0 < z < 1/2$ then we can again take $k_2 = m + i$, and we just need to show the existence of an appropriate k_1 . Assume by contradiction that there does not exist an appropriate k_1 , i.e., for all $\ell < 2mz$, $p_\ell = 1$. Then we have

$$P(s) \geq \frac{1}{2} \left[\sum_{0 \leq k < 2mz} \binom{m}{k} s^k (1-s)^{m-k} - s \right] = \frac{1}{2} \left[1 - s - \sum_{2mz \leq k \leq m} \binom{m}{k} s^k (1-s)^{m-k} \right].$$

By Markov's inequality for a binomial random variable, this latter sum evaluated at z is at most $1/2$, and since $z < 1/2$, we must have $P(z) > 0$, which is a contradiction. The case of $1/2 < z < 1$ is similar. \square

Proof of Theorem 1.4 for $z \in (0, 1)$. This follows from Lemma 2.6 and Theorem 2.4. The only condition of Theorem 2.4 that needs to be checked additionally is that $\mathbb{E}(\xi_{n+1}^+ | \mathcal{F}_n)$ and $\mathbb{E}(\xi_{n+1}^- | \mathcal{F}_n)$ are bounded away from zero by positive numbers when $x_n \in (z - \varepsilon, z + \varepsilon)$ for small enough $\varepsilon > 0$; this can be done using Lemma 2.7. In the special cases (a), (b), and (c) described in Lemma 2.7, the statement of Theorem 1.4 is vacuously true, since in each case the polynomial P has no zeros at which it is increasing. Thus we may assume that we are not in these special cases, and we can use Lemma 2.7. Recall that

$$\xi_{n+1} = \frac{n}{S_{n+1}} \{X_{n+1} - X_n - 2m(x_n + P(x_n))\}.$$

Define $\tilde{\varepsilon} := \frac{1}{2} \min\{2mz - k_1, k_2 - 2mz\}$, where k_1 and k_2 are given by Lemma 2.7, and let $\varepsilon > 0$ be small enough such that whenever $x_n \in (z - \varepsilon, z + \varepsilon)$, necessarily $2m(x_n + P(x_n)) \in (2mz - \tilde{\varepsilon}, 2mz + \tilde{\varepsilon})$. If $x_n \in (z - \varepsilon, z + \varepsilon)$ then we have $\mathbb{E}(\xi_{n+1}^+ | \mathcal{F}_n) \geq \frac{n}{S_{n+1}} \tilde{\varepsilon} \mathbb{P}(X_{n+1} - X_n = k_2 | \mathcal{F}_n)$, where $\lim_{n \rightarrow \infty} \frac{n}{S_{n+1}} = \frac{1}{2m}$, and by Lemma 2.7 the probability $\mathbb{P}(X_{n+1} - X_n = k_2 | \mathcal{F}_n)$ is bounded from below by a positive function of z , ε , and the parameters $\{p_k\}_{0 \leq k \leq m}$. We can similarly bound $\mathbb{E}(\xi_{n+1}^- | \mathcal{F}_n)$ from below. \square

We next prove Theorem 1.4 for the endpoints 0 and 1. The main idea of the proof is to compare the behaviour near the endpoints of our process of interest to that of a standard Pólya urn process where $2m$ balls are added at each time step. In order to formalize this, we make use of several different stochastic orders; we refer to [28] for an overview of these. We proceed by defining these stochastic orders and stating a few results on them, before proving Theorem 1.4.

Definition 1 (Stochastic orders). *Let X and Y be random variables.*

We say that X is smaller than Y in the usual stochastic order (denoted by $X \leq_{\text{st}} Y$) if $\mathbb{E}(\phi(X)) \leq \mathbb{E}(\phi(Y))$ for all increasing continuous functions $\phi: \mathbb{R} \rightarrow \mathbb{R}$ for which these expectations exist.

We say that X is smaller than Y in the convex order (denoted by $X \leq_{\text{cx}} Y$) if $\mathbb{E}(\phi(X)) \leq \mathbb{E}(\phi(Y))$ for all continuous convex functions $\phi: \mathbb{R} \rightarrow \mathbb{R}$ for which these expectations exist.

We say that X is smaller than Y in the increasing convex order (denoted by $X \leq_{\text{icx}} Y$) if $\mathbb{E}(\phi(X)) \leq \mathbb{E}(\phi(Y))$ for all increasing continuous convex functions $\phi: \mathbb{R} \rightarrow \mathbb{R}$ for which these expectations exist.

Lemma 2.8. *Two random variables X and Y satisfy $X \leq_{\text{icx}} Y$ if and only if there is a random variable Z such that $X \leq_{\text{st}} Z \leq_{\text{cx}} Y$.*

Proof. See [28, Theorem 4.A.6. (a)]. \square

Lemma 2.9. *Let X and Y be two random variables with cumulative distribution functions F and G , respectively, and bounded supports. Suppose that $\mathbb{E}(X) \leq \mathbb{E}(Y)$, and also that if $t_1 < t_2$ and $G(t_1) < F(t_1)$ then $G(t_2) \leq F(t_2)$. Then $X \leq_{\text{icx}} Y$.*

Proof. See [28, Theorem 4.A.22. (b)]. \square

Lemma 2.10. *Consider the standard Pólya urn process where $2m$ balls are added at each time step. Let x_n^1 and x_n^2 be the proportions of red balls at the n^{th} step of two realizations of this process. If $x_n^1 \leq_{\text{cx}} x_n^2$, then $x_{n+1}^1 \leq_{\text{cx}} x_{n+1}^2$, i.e., the Pólya urn process preserves dominance in the convex order.*

Proof. See Proposition 1 in [21], in particular equation (13). \square

Proof of Theorem 1.4 for the endpoints. We prove nonconvergence to 1 when $P(1) = 0$ and $P < 0$ on $(1 - \varepsilon, 1)$ for some $\varepsilon > 0$; the proof for the other endpoint is analogous. In the following fix $0 < \varepsilon < 1/m$.

The main idea of the proof is to compare the behaviour near 1 of our process of interest to that of a standard Pólya urn process where $2m$ balls are added at each time step. To aid in this comparison we also introduce an auxiliary process which is a combination of these two. We begin by describing these processes.

Our process of interest is $\{X_n\}_{n \geq 0}$, together with its normalised process $\{x_n\}_{n \geq 0}$. Let $\{\bar{X}_n\}_{n \geq 0}$ denote the process of the number of red balls in a standard Pólya urn process where $2m$ balls are added at each time step and the initial conditions are the same as those for the process $\{X_n\}_{n \geq 0}$, i.e., $\bar{X}_0 = X_0$. Let $\{\bar{x}_n\}_{n \geq 0}$ denote the normalised process, i.e., $\bar{x}_n = \frac{\bar{X}_n}{S_0 + 2mn}$. Let $\{\tilde{X}_n\}_{n \geq 0}$ denote the auxiliary process, with initial condition $\tilde{X}_0 = X_0$, and let $\{\tilde{x}_n\}_{n \geq 0}$ denote the normalised process, i.e., $\tilde{x}_n = \frac{\tilde{X}_n}{S_0 + 2mn}$. We define this auxiliary process as follows. For $1 - \varepsilon < x \leq 1$, given $\tilde{x}_n = x$ let \tilde{X}_{n+1} have the same distribution as X_{n+1} given $x_n = x$. For $x \leq 1 - \varepsilon$, let

$$\begin{aligned} \mathbb{P}\left(\tilde{X}_{n+1} = \tilde{X}_n \mid \tilde{x}_n = x\right) &= 1 - x \\ \text{and } \mathbb{P}\left(\tilde{X}_{n+1} = \tilde{X}_n + 2m \mid \tilde{x}_n = x\right) &= x. \end{aligned}$$

In other words, when $\tilde{x}_n > 1 - \varepsilon$ then evolve the auxiliary process according to our process of interest, and when $\tilde{x}_n \leq 1 - \varepsilon$ then evolve it as a Pólya urn process.

We first show that it suffices to prove the claim for the auxiliary process, i.e., it suffices to show that $\mathbb{P}(\lim_{n \rightarrow \infty} \tilde{x}_n = 1) = 0$. Define the following events:

$$A_n := \left\{ \lim_{k \rightarrow \infty} x_k = 1, x_k > 1 - \varepsilon \text{ for all } k \geq n \right\}, \quad \tilde{A}_n := \left\{ \lim_{k \rightarrow \infty} \tilde{x}_k = 1, \tilde{x}_k > 1 - \varepsilon \text{ for all } k \geq n \right\}.$$

If $\mathbb{P}(\lim_{n \rightarrow \infty} x_n = 1) > 0$, then there exists $n_0 < \infty$ such that $\mathbb{P}(A_{n_0}) > 0$. In particular, there exists $y_0 \in (1 - \varepsilon, 1)$ such that both probabilities $\mathbb{P}(x_{n_0} \geq y_0)$ and $\mathbb{P}(A_{n_0} \mid x_{n_0} = y_0)$ are positive. In fact, we claim that $\mathbb{P}(A_{n_0} \mid x_{n_0} = y)$ is positive for all $y_0 \leq y < 1$.

To see this, consider two realizations of our process, $\{X_n^1\}_{n \geq 0}$ and $\{X_n^2\}_{n \geq 0}$, together with the normalised processes $\{x_n^1\}_{n \geq 0}$ and $\{x_n^2\}_{n \geq 0}$. Given $1 - \varepsilon < x_n^1 < x_n^2 < 1$, we can couple X_{n+1}^1 and X_{n+1}^2 such that for any $0 \leq k \leq 2m$, $X_{n+1}^1 - X_n^1 = k$ implies that either $X_{n+1}^2 - X_n^2 = k$ or $X_{n+1}^2 - X_n^2 = 2m$. This is possible due to two facts. First, since $\varepsilon < 1/m$, on the interval $(1 - \varepsilon, 1)$ the function $x \mapsto x^m$ is increasing, while for $0 \leq k < m$, the functions $x \mapsto x^k(1 - x)^{m-k}$ are decreasing. Consequently

$$\mathbb{P}(\text{Bin}(m, x_n^1) = m) < \mathbb{P}(\text{Bin}(m, x_n^2) = m)$$

and for $0 \leq k < m$,

$$\mathbb{P}(\text{Bin}(m, x_n^1) = k) > \mathbb{P}(\text{Bin}(m, x_n^2) = k),$$

where $\text{Bin}(m, x)$ denotes a binomial random variable with parameters m and x . Second, $P(1) = 0$ implies that $p_m = 1$. Repeated application of this coupling shows that for any $1 - \varepsilon < y^1 < y^2 < 1$ we have

$$\mathbb{P}(A_n \mid x_n = y^1) \leq \mathbb{P}(A_n \mid x_n = y^2);$$

in particular, we have that

$$\mathbb{P}(A_{n_0} \mid x_{n_0} = y) \geq \mathbb{P}(A_{n_0} \mid x_{n_0} = y_0)$$

for all $y \geq y_0$.

Now consider the auxiliary process. For one, we have $\mathbb{P}(\tilde{x}_{n_0} \geq y_0) > 0$. Moreover, if $x_{n_0} = \tilde{x}_{n_0}$, on the event A_{n_0} we can couple the processes $\{x_n\}_{n \geq n_0}$ and $\{\tilde{x}_n\}_{n \geq n_0}$ so that $x_n = \tilde{x}_n$ for all $n \geq n_0$, which shows that

$$\mathbb{P}(\tilde{A}_{n_0} \mid \tilde{x}_{n_0} = y) \geq \mathbb{P}(A_{n_0} \mid x_{n_0} = y_0) > 0$$

for all $y \geq y_0$. In particular, this shows that $\mathbb{P}(\lim_{n \rightarrow \infty} x_n = 1) > 0$ implies that $\mathbb{P}(\lim_{n \rightarrow \infty} \tilde{x}_n = 1) > 0$. Thus it suffices to show that $\mathbb{P}(\lim_{n \rightarrow \infty} \tilde{x}_n = 1) = 0$.

We claim that $\tilde{x}_n \leq_{\text{icx}} \bar{x}_n$ implies that $\mathbb{P}(\lim_{n \rightarrow \infty} \tilde{x}_n = 1) = 0$. To see this, for $\delta > 0$ define the function $g_\delta : [0, 1] \rightarrow [0, 2]$ by $g_\delta(x) = \max\{0, 2 - 1/\delta + x/\delta\}$. This is an increasing continuous convex function, and so $\tilde{x}_n \leq_{\text{icx}} \bar{x}_n$ implies that

$$(7) \quad \mathbb{P}(\tilde{x}_n > 1 - \delta) \leq \mathbb{E}(g_\delta(\tilde{x}_n)) \leq \mathbb{E}(g_\delta(\bar{x}_n)) \leq 2\mathbb{P}(\bar{x}_n > 1 - 2\delta).$$

We know that the limiting distribution of \bar{x}_n is a beta distribution, and thus

$$\lim_{\delta \rightarrow 0} \lim_{n \rightarrow \infty} \mathbb{P}(\bar{x}_n > 1 - 2\delta) = 0.$$

By (7) this then implies that $\mathbb{P}(\lim_{n \rightarrow \infty} \tilde{x}_n = 1) = 0$.

We prove $\tilde{X}_n \leq_{\text{icx}} \bar{X}_n$ (which is equivalent to $\tilde{x}_n \leq_{\text{icx}} \bar{x}_n$) by induction on n ; for $n = 0$ this is immediate since the initial conditions agree. Fix now a positive integer n , and consider a random variable X which attains integer values in the interval $[X_0, S_0 + 2mn]$, and let $x = \frac{X}{S_0 + 2mn}$. Denote by \bar{X} a random variable with distribution $\mathbb{P}(\bar{X} = X + 2m \mid X) = x$ and $\mathbb{P}(\bar{X} = X \mid X) = 1 - x$. Similarly, let \tilde{X} denote a random variable with distribution the same as that of \tilde{X}_{n+1} conditioned on $\tilde{X}_n = X$. Following Pemantle [21], the induction step follows from the following two claims: (i) $\tilde{X} \leq_{\text{icx}} \bar{X}$, and (ii) $X \leq_{\text{icx}} Y$ implies that $\bar{X} \leq_{\text{icx}} \bar{Y}$.

First, it is enough to show that for any fixed r , conditioned on $x = r$ we have $\tilde{X} \leq_{\text{icx}} \bar{X}$; one can then integrate out the conditioning to get (i). We show this by checking the conditions of Lemma 2.9. First, when $r \leq 1 - \varepsilon$ we have $\mathbb{E}(\tilde{X} \mid x = r) = \mathbb{E}(\bar{X} \mid x = r)$ by the definition of the auxiliary process. If $r > 1 - \varepsilon$ then we have $\mathbb{E}(\bar{X} \mid x = r) = r(S_0 + 2mn) + 2mr$, while $\mathbb{E}(\tilde{X} \mid x = r) = r(S_0 + 2mn) + 2m(r + P(r))$. Since $r > 1 - \varepsilon$, $P(r) < 0$, and thus $\mathbb{E}(\tilde{X} \mid x = r) < \mathbb{E}(\bar{X} \mid x = r)$. This shows that $\mathbb{E}(\tilde{X} \mid x = r) \leq \mathbb{E}(\bar{X} \mid x = r)$. The other condition of Lemma 2.9 holds automatically due to the fact that conditioned on $X = \ell$, the distribution of \bar{X} is supported on the two values $\{\ell, \ell + 2m\}$, while the support of the distribution of \tilde{X} is contained in the interval $[\ell, \ell + 2m]$.

In view of Lemmas 2.8 and 2.10, to show (ii) it is enough to show that $X \leq_{\text{st}} Y$ implies $\bar{X} \leq_{\text{st}} \bar{Y}$, i.e., that for any increasing function ϕ we have $\mathbb{E}(\phi(\bar{X})) \leq \mathbb{E}(\phi(\bar{Y}))$. By conditioning on X and Y , we have $\mathbb{E}(\phi(\bar{X})) = \mathbb{E}(\bar{\phi}(X))$ and $\mathbb{E}(\phi(\bar{Y})) = \mathbb{E}(\bar{\phi}(Y))$, where $\bar{\phi}(t) := \phi(t)(1 - \alpha t) + \phi(t + 2m)\alpha t$, where $\alpha = (S_0 + 2mn)^{-1}$ and t is such that $0 \leq \alpha t \leq 1$. Since $X \leq_{\text{st}} Y$, we only need to show that $\bar{\phi}$ is increasing. Indeed, if $t_1 < t_2$ then

$$\begin{aligned} \bar{\phi}(t_2) - \bar{\phi}(t_1) &= (\phi(t_2 + 2m) - \phi(t_1 + 2m))\alpha t_1 + (\phi(t_2) - \phi(t_1))(1 - \alpha t_1) \\ &\quad + (\phi(t_2 + 2m) - \phi(t_2))\alpha(t_2 - t_1), \end{aligned}$$

which is nonnegative, since all of the terms on the right hand side are nonnegative. □

Proof of Theorem 1.5. This follows directly from Lemma 2.6 and Theorem 2.5. □

3 Many colours/types

It is both natural and important to study competition between more than two colours/types. The model naturally extends in this direction, and in this section we present our results regarding $N \geq 3$ competing types. In the following, vectors will be denoted using boldface, subscripts typically correspond to time and superscripts correspond to the indices of types. Furthermore, denote by Δ^N the probability simplex in \mathbb{R}^N .

The natural extension of the model to multiple competing types is as follows. At time zero, there is a graph G_0 , where each node is of exactly one of the N types. At each timestep a new node is added to the

graph, and is connected to m nodes of the original graph according to linear preferential attachment. The types of these m neighbours induce a vector of types \mathbf{u} , where u^i is the number of neighbours of type i . The type of the new node is then determined according to a random draw from the distribution $p_{\mathbf{u}} = \{p_{\mathbf{u}}^i\}_{i \in [N]}$. The probabilities $\{p_{\mathbf{u}}^i\}_{\mathbf{u}, i}$ are parameters of the model.

As in the case of two types, our primary interest is in the fraction of nodes of each type. Let A_n^i denote the number of nodes of type i at time n , and let $\mathbf{A}_n = (A_n^1, \dots, A_n^N)$ denote the resulting vector of types. Let \mathbf{a}_n denote the normalised vector of types, such that $\sum_{i=1}^N a_n^i = 1$. Furthermore, let X_n^i denote the sum of the degrees of type i nodes at time n , let $\mathbf{X}_n = (X_n^1, \dots, X_n^N)$ denote the resulting vector of degrees, and let \mathbf{x}_n be the normalised vector of degrees, such that $\sum_{i=1}^N x_n^i = 1$.

As in the $N = 2$ case, there is a clear distinction between the *linear model*, when $p_{\mathbf{u}}^i = \frac{u^i}{m}$ for all \mathbf{u} and $i \in [N]$, and *nonlinear models*, when there exist \mathbf{u} and $i \in [N]$ such that $p_{\mathbf{u}}^i \neq \frac{u^i}{m}$. In fact, the linear model for $N \geq 3$ types reduces to the linear model for two types. This is because in the linear model, if we want to study the evolution of the size of type i , then we can group all other types into a single “mega-type”, denoted by $-i$, and run the process with two types: type i and “mega-type” $-i$. Due to linearity, the original process with N types and the process with type i and “mega-type” $-i$ can be coupled such that the evolution of type i is identical in the two processes. Consequently, in the linear model all the results of the $N = 2$ case apply. In particular, we have the following theorem.

Theorem 3.1 (Linear model). *Assume that $p_{\mathbf{u}}^i = \frac{u^i}{m}$ for all \mathbf{u} and $i \in [N]$, and that $X_0^i > 0$ for all $i \in [N]$. Then \mathbf{a}_n converges almost surely, and the limiting distribution has full support on Δ^N , and no atoms. \square*

In nonlinear models, as we will see later, a key role in the asymptotic behaviour of the process $\{\mathbf{a}_n\}_{n \geq 0}$ is played by the vector field

$$(8) \quad \mathbf{P}(\mathbf{y}) = \frac{1}{2} \sum_{i=1}^N \sum_{\mathbf{u}} \binom{m}{\mathbf{u}} (\mathbf{y})^{\mathbf{u}} \left[p_{\mathbf{u}}^i - \frac{u^i}{m} \right] \boldsymbol{\delta}^i,$$

where $\binom{m}{\mathbf{u}} = \frac{m!}{u^1! \dots u^N!}$ denotes the multinomial coefficient, $(\mathbf{y})^{\mathbf{u}} = (y^1)^{u^1} (y^2)^{u^2} \dots (y^N)^{u^N}$, and $\boldsymbol{\delta}^i$ is the N -dimensional unit vector whose i^{th} coordinate is 1 and all other coordinates are 0. Let us denote the zero set of this vector field on the probability simplex by $Z_{\mathbf{P}} := \{\mathbf{y} \in \Delta^N : \mathbf{P}(\mathbf{y}) = \mathbf{0}\}$; this will be important later.

The behaviour of the process in the general nonlinear model with multiple types is involved, and its complete theoretical analysis is as of yet out of our reach. Nonetheless, based on partial theoretical results, we conjecture the following asymptotic behaviour, which is similar to that in the case of two types.

Conjecture 3.2 (Nonlinear models). *Assume that there exist \mathbf{u} and $i \in [N]$ such that $p_{\mathbf{u}}^i \neq \frac{u^i}{m}$, and that $X_0^i > 0$ for all $i \in [N]$. Then \mathbf{a}_n converges almost surely and the limit is a point in the zero set $Z_{\mathbf{P}}$.*

In the rest of this section we describe theoretical progress towards this conjecture. As in the case of two competing types, the problem can be cast in a (multidimensional) stochastic approximation framework.

The process $\{\mathbf{A}_n\}_{n \geq 0}$ is not a Markov process, and therefore we study the joint process $\{(\mathbf{A}_n, \mathbf{X}_n)\}_{n \geq 0}$, which is indeed Markov. It evolves as follows. Given $(\mathbf{A}_n, \mathbf{X}_n)$, a vector \mathbf{u}_{n+1} is drawn from the multinomial distribution with parameters m and \mathbf{x}_n . Subsequently, an index $I_{n+1} \in [N]$ is chosen from the distribution $p_{\mathbf{u}_{n+1}}$. We then have

$$(9) \quad \mathbf{A}_{n+1} = \mathbf{A}_n + \boldsymbol{\delta}^{I_{n+1}},$$

$$(10) \quad \mathbf{X}_{n+1} = \mathbf{X}_n + \mathbf{u}_{n+1} + m \boldsymbol{\delta}^{I_{n+1}}.$$

Before analysing the process $\{\mathbf{x}_n\}_{n \geq 0}$, let us show that in order to prove Conjecture 3.2 on the asymptotic behaviour of $\{\mathbf{a}_n\}_{n \geq 0}$, it is sufficient to prove a similar result on the asymptotic behaviour of $\{\mathbf{x}_n\}_{n \geq 0}$.

Lemma 3.3. *Assume that there exist \mathbf{u} and $i \in [N]$ such that $p_{\mathbf{u}}^i \neq \frac{u^i}{m}$, and that $X_0^i > 0$ for all $i \in [N]$. Assume that \mathbf{x}_n converges almost surely and the limit is a point in the zero set $Z_{\mathbf{P}}$. Then \mathbf{a}_n converges almost surely and $\lim_{n \rightarrow \infty} \mathbf{a}_n = \lim_{n \rightarrow \infty} \mathbf{x}_n \in Z_{\mathbf{P}}$.*

Proof. This is similar to the proof of Lemma 2.1. We have seen that

$$\begin{aligned}\mathbb{E}(\mathbf{A}_{n+1} - \mathbf{A}_n | \mathcal{F}_n) &= \mathbb{E}(\boldsymbol{\delta}^{I_{n+1}} | \mathcal{F}_n) = \sum_{i=1}^N \sum_{\mathbf{u}} \binom{m}{\mathbf{u}} (\mathbf{x}_n)^{\mathbf{u}} p_{\mathbf{u}}^i \boldsymbol{\delta}^i \\ &= \mathbf{x}_n + \sum_{i=1}^N \sum_{\mathbf{u}} \binom{m}{\mathbf{u}} (\mathbf{x}_n)^{\mathbf{u}} \left[p_{\mathbf{u}}^i - \frac{u^i}{m} \right] \boldsymbol{\delta}^i = \mathbf{x}_n + 2\mathbf{P}(\mathbf{x}_n) =: f(\mathbf{x}_n).\end{aligned}$$

Let $\mathbf{M}_0 = \mathbf{0}$ and define the martingale $\mathbf{M}_n := \mathbf{A}_n - \mathbf{A}_0 - \sum_{j=0}^{n-1} f(\mathbf{x}_j)$. This martingale has bounded increments, and thus $\lim_{n \rightarrow \infty} \mathbf{M}_n/n = \mathbf{0}$ a.s. By the definition of the martingale, this shows that a.s.

$$\lim_{n \rightarrow \infty} \left[\mathbf{a}_n - \frac{1}{n} \sum_{j=0}^{n-1} f(\mathbf{x}_j) \right] = \mathbf{0}.$$

Now if the limit $\lim_{n \rightarrow \infty} \mathbf{x}_n$ exists a.s., and any limit point \mathbf{x} satisfies $\mathbf{P}(\mathbf{x}) = \mathbf{0}$, then also $f(\mathbf{x}) = \mathbf{x}$, and thus the limit of the Cesàro mean of the sequence $\{f(\mathbf{x}_n)\}_{n \geq 0}$ also converges to the same limit point. This then implies that the limit $\lim_{n \rightarrow \infty} \mathbf{a}_n$ exists a.s. and is equal to $\lim_{n \rightarrow \infty} \mathbf{x}_n$. \square

The key observation in the analysis of the asymptotic behaviour of $\{\mathbf{x}_n\}_{n \geq 0}$ is that it is a stochastic approximation process. In higher dimensions, a stochastic approximation process is defined as follows. Let \mathbf{Z}_n be a stochastic process in the Euclidean space \mathbb{R}^N and adapted to a filtration $\{\mathcal{F}_n\}_{n \geq 0}$. Suppose that it satisfies

$$\mathbf{Z}_{n+1} - \mathbf{Z}_n = \frac{1}{n} (\mathbf{F}(\mathbf{Z}_n) + \boldsymbol{\xi}_{n+1} + \mathbf{R}_n),$$

where \mathbf{F} is a vector field on \mathbb{R}^N , $\mathbb{E}(\boldsymbol{\xi}_{n+1} | \mathcal{F}_n) = \mathbf{0}$, and the remainder terms $\mathbf{R}_n \in \mathcal{F}_n$ go to zero and satisfy $\sum_{n=1}^{\infty} n^{-1} \|\mathbf{R}_n\| < \infty$ a.s. Such a process is known as a stochastic approximation process.

Lemma 3.4. *The process $\{\mathbf{x}_n\}_{n \geq 0}$ is a stochastic approximation process with the vector field \mathbf{P} as in (8). Furthermore, the noise term $\boldsymbol{\xi}_n$ is bounded: $\|\boldsymbol{\xi}_n\|_1 \leq 2N$ for all $n \geq 1$.*

Proof. From (10) we have that $\mathbb{E}(\mathbf{X}_{n+1} - \mathbf{X}_n | \mathcal{F}_n) = \mathbb{E}(\mathbf{u}_{n+1} | \mathcal{F}_n) + m\mathbb{E}(\boldsymbol{\delta}^{I_{n+1}} | \mathcal{F}_n)$. Given \mathcal{F}_n , \mathbf{u}_{n+1} is multinomial with parameters m and \mathbf{x}_n , and so $\mathbb{E}(\mathbf{u}_{n+1} | \mathcal{F}_n) = m\mathbf{x}_n$. By construction, we have that

$$\mathbb{E}(\boldsymbol{\delta}^{I_{n+1}} | \mathcal{F}_n) = \sum_{i=1}^N \sum_{\mathbf{u}} \binom{m}{\mathbf{u}} (\mathbf{x}_n)^{\mathbf{u}} p_{\mathbf{u}}^i \boldsymbol{\delta}^i.$$

Let S_0 denote the sum of the degrees in G_0 , and let $S_n = S_0 + 2mn$. A simple calculation gives that $\mathbf{x}_{n+1} - \mathbf{x}_n = \frac{\mathbf{X}_{n+1} - \mathbf{X}_n - 2m\mathbf{x}_n}{S_{n+1}}$, and so we have

$$\mathbb{E}(\mathbf{x}_{n+1} - \mathbf{x}_n | \mathcal{F}_n) = \frac{m}{S_{n+1}} \sum_{i=1}^N \sum_{\mathbf{u}} \binom{m}{\mathbf{u}} (\mathbf{x}_n)^{\mathbf{u}} \left[p_{\mathbf{u}}^i - \frac{u^i}{m} \right] \boldsymbol{\delta}^i = \frac{2m}{S_{n+1}} \mathbf{P}(\mathbf{x}_n).$$

We can then write $\{\mathbf{x}_n\}_{n \geq 0}$ as a stochastic approximation process:

$$\mathbf{x}_{n+1} - \mathbf{x}_n = \frac{1}{n} [\mathbf{P}(\mathbf{x}_n) + \boldsymbol{\xi}_{n+1} + \mathbf{R}_n],$$

where

$$(11) \quad \boldsymbol{\xi}_{n+1} = n \{ \mathbf{x}_{n+1} - \mathbf{x}_n - \mathbb{E}(\mathbf{x}_{n+1} - \mathbf{x}_n | \mathcal{F}_n) \}$$

is the martingale term, and the remainder term is

$$\mathbf{R}_n = -\frac{S_0 + 2m}{S_0 + 2m(n+1)} \mathbf{P}(\mathbf{x}_n).$$

Clearly $\mathbf{R}_n \in \mathcal{F}_n$ and similarly as at the end of the proof of Lemma 2.6 one can show that $\|\mathbf{R}_n\| \leq c/n$ for some constant $c = c(N, S_0, m)$, which implies that $\sum_{n=1}^{\infty} n^{-1} \|\mathbf{R}_n\| < \infty$ a.s.

Finally, to check that $\|\boldsymbol{\xi}_n\|_1 \leq 2N$, note that

$$|x_{n+1}^i - x_n^i| = \left| \frac{X_{n+1}^i - X_n^i - 2mx_n^i}{S_{n+1}} \right| \leq \frac{2m}{2m(n+1)} = \frac{1}{n+1},$$

and then use (11). □

As in the one-dimensional case, intuitively, trajectories of a stochastic approximation process $\{\mathbf{Z}_n\}_{n \geq 0}$ should approximate the trajectories $\{\mathbf{Z}(t)\}_{t \geq 0}$ of the corresponding ODE $d\mathbf{Z}/dt = \mathbf{F}(\mathbf{Z})$. Moreover, since $\{\mathbf{Z}_n\}_{n \geq 0}$ is a stochastic system, we expect that stable trajectories of the ODE should appear, but unstable trajectories should not.

The main concept in formalizing this intuition is that of an *asymptotic pseudotrajectory*, introduced by Benaïm and Hirsch [9]. We omit the precise definition, and refer to Benaïm's lecture notes on the topic for more details [7] (see also [24, Section 2.5] for a concise summary). There are many results that give sufficient conditions for a stochastic approximation process to be an asymptotic pseudotrajectory of the corresponding ODE. In particular, [7, Proposition 4.4 and Remark 4.5] (see also [24, Theorem 2.13]), together with Lemma 3.4 and the fact that \mathbf{P} is Lipschitz, imply the following.

Corollary 3.5. *Let $\{\mathbf{x}(t)\}_{t \geq 0}$ linearly interpolate $\{\mathbf{x}_n\}_{n \geq 0}$ at nonintegral times. Then $\{\mathbf{x}(t)\}_{t \geq 0}$ is a.s. an asymptotic pseudotrajectory for the flow induced by the vector field \mathbf{P} via the ODE $d\mathbf{y}/dt = \mathbf{P}(\mathbf{y})$. □*

There are further general results about asymptotic pseudotrajectories that apply to the stochastic approximation process $\{\mathbf{x}_n\}_{n \geq 0}$, e.g., about convergence to attractors and nonconvergence to linearly unstable equilibria. However, we omit these, as we prefer to emphasise the main message of Corollary 3.5. The main point is that in order to understand the stochastic approximation process $\{\mathbf{x}_n\}_{n \geq 0}$, we need to understand the vector field \mathbf{P} , and the corresponding ODE

$$\frac{d\mathbf{y}}{dt} = \mathbf{P}(\mathbf{y}).$$

Unfortunately, understanding the behaviour of such nonlinear ODEs is a notoriously difficult subject (see, e.g., the book by Hirsch, Smale and Devaney [15]). The most successful tool in this area is Lyapunov theory (see, e.g., the recent paper [8]), and this can indeed be applied to our problem for special values of the parameters; however, it seems difficult to apply this theory to the vector field \mathbf{P} for *generic* values of the parameters $\{p_{\mathbf{u}}^i\}_{\mathbf{u}, i}$.

For instance, if \mathbf{P} is a gradient, i.e., $\mathbf{P} = -\nabla V$ for some $V : \mathbb{R}^N \rightarrow \mathbb{R}$, then Corollary 3.5 and general results about asymptotic pseudotrajectories (see [7]) imply that \mathbf{x}_n converges almost surely and the limit is a point in the zero set $Z_{\mathbf{P}}$, which, by Lemma 3.3, implies that Conjecture 3.2 holds. An example of when \mathbf{P} is a gradient is when the probability of the new node adopting type i depends only on the number of type i connections, i.e., $p_{\mathbf{u}}^i = \phi(u^i)$ for some function ϕ which does not depend on i . This implies that ϕ must be of the form $\phi(z) = \alpha \frac{1}{N} + (1 - \alpha) \frac{z}{m}$ for some $0 \leq \alpha \leq 1$,³ which corresponds to a mixture of the linear model and a uniformly random choice. In this case $\mathbf{P}(\mathbf{y}) = \frac{\alpha}{2} (\frac{1}{N} \mathbf{1} - \mathbf{y})$, where $\mathbf{1} \in \mathbb{R}^N$ is the vector with all entries equal to 1, and thus when $\alpha > 0$ then \mathbf{a}_n converges a.s. to $\frac{1}{N} \mathbf{1}$.

However, for *generic* parameter values \mathbf{P} will not be a gradient. To see this, note that \mathbf{P} being a gradient implies that

$$(12) \quad \frac{\partial (\mathbf{P}(\mathbf{y}))^i}{\partial y^j} = \frac{\partial (\mathbf{P}(\mathbf{y}))^j}{\partial y^i}$$

³To see this, first observe that there are $m + 1$ free parameters when constructing such a function: $\phi(0), \phi(1), \dots, \phi(m)$. These parameters also satisfy constraints, since the probabilities have to sum to 1: $\sum_{i=1}^N p_{\mathbf{u}}^i = 1$ for every vector of types \mathbf{u} . In particular, the following m equations are linearly independent: $(N - 3)\phi(0) + \phi(0) + \phi(m - i) = 1$, for $i = 0, 1, \dots, \lfloor m/2 \rfloor$, and $(N - 3)\phi(0) + \phi(1) + \phi(i) + \phi(m - i - 1) = 1$, for $i = 1, 2, \dots, \lfloor (m - 1)/2 \rfloor$. Consequently, there can be at most 2 linearly independent solutions to these equations. The constant function $\phi(z) = 1/N$ and the linear function $\phi(z) = z/m$ are solutions, which means that any solution is a linear combination of these. The restriction of α to be in $[0, 1]$ follows from the constraint that the probabilities be nonnegative.

for every $i \neq j$. Without any restrictions, there are $(N-1) \binom{m+N-1}{N-1}$ free parameters in \mathbf{P} . The gradient condition (12) imposes an additional $\binom{N}{2}$ constraints, which will not be satisfied for generic parameter values.

4 Open problems and future directions

Our paper leaves open several interesting problems. Two immediate open questions concerning the model are the following.

- **Limiting distribution in the linear model for two types.** Our Theorem 1.1 gives us information about the limiting behaviour of $\{a_n\}_{n \geq 0}$, but it does not identify the distribution of $a := \lim_{n \rightarrow \infty} a_n$. For $m = 1$ the process $\{x_n\}_{n \geq 0}$ corresponds to a Pólya urn where whenever one draws a ball, one puts back two extra balls of the same colour. This is because when a new node joins the graph, its colour automatically becomes the colour of its initial neighbour. Thus the distribution of x —and by Lemma 2.1 the distribution of a as well—is the Beta distribution with parameters $\frac{X_0}{2}$ and $\frac{Y_0}{2}$. However, for $m > 1$ we do not know what the limiting distribution is. Note that simulations show that the limiting distribution can be bimodal; see, e.g., Figure 2b.
- **Understanding the vector field \mathbf{P} .** As discussed in Section 3, in order to understand the behaviour of the general nonlinear model in the case of multiple types—and in order to prove or disprove Conjecture 3.2—a good understanding of the vector field \mathbf{P} and the corresponding ODE $d\mathbf{y}/dt = \mathbf{P}(\mathbf{y})$ is needed. We leave this as our second open problem.

A key property of the model is its *simplicity*. However, this also means that certain aspects of real-world networks and processes influencing product adoption are simplified or not considered. It would be interesting to understand the following possible extensions of the model, and, in particular, whether anything can be said analytically in these extensions.

- **Changing preferences.** In the model once a node receives a type, that type is then fixed and cannot change over time. A possible extension of the model is to allow the type of a node to change over time. This can model changing preferences of individuals, e.g., somebody moving from one mobile phone provider to another.
- **Allowing multiple types for a single individual.** In the model a node can only have a single type. This is reasonable in many situations (e.g., an individual typically has only one mobile phone provider), but modeling other situations might require allowing nodes to simultaneously have multiple types.
- **Other network evolution models.** The preferential attachment model is a reasonable approximation of some real-world networks, and it has the advantageous property of being analytically tractable. How does the model behave under other network evolution models? Can similar results be shown analytically/experimentally? Are the results robust to small changes in the network evolution model?

In particular, the affine preferential attachment model is discussed at the end of Section 1.2. This model is more involved, because when $c \neq 0$, then $\{X_n\}_{n \geq 0}$ is not a Markov process, so one has to understand the joint evolution of $\{(A_n, X_n)\}_{n \geq 0}$. A calculation shows that $\{(a_n, x_n)\}_{n \geq 0}$ is a stochastic approximation process with vector field

$$\mathbf{F}(a, x) = \left(\frac{2m}{2m+c} (x - a) + 2P \left(\frac{2m}{2m+c} x + \frac{c}{2m+c} a \right), \frac{c}{2m+c} (a - x) + P \left(\frac{2m}{2m+c} x + \frac{c}{2m+c} a \right) \right).$$

For $c \geq 0$, the zero set of \mathbf{F} is exactly $\{(a, x) : a = x, P(x) = 0\}$, and so the same results as in Section 1.2 apply. For $c \in (-1, 0)$ the zero set of \mathbf{F} could be larger; we leave the analysis of this as an exercise to the reader.

The case of uniform attachment is even simpler than preferential attachment, since then $\{A_n\}_{n \geq 0}$ is a Markov process, and it is unnecessary to consider the sum-of-degrees process $\{X_n\}_{n \geq 0}$. The same results as in Section 1.2 apply.

We leave the study of further network evolution models for future work.

- **Other type adoption mechanisms.** The model incorporates a fairly general type adoption mechanism, but various modifications would be interesting to explore. For instance, in real life choices are often made based on the opinions of specific friends, not just based on aggregate information of one's friends.
- **Marketing.** In essence, the model describes word-of-mouth recommendations, and does not consider marketing efforts by the competing companies, such as advertising. How does incorporating marketing affect the results?

In conclusion, through a simple model we have coupled network evolution and type adoption, leading to a possible explanation of coexistence in preferential attachment networks. Exploring various modifications and extensions of this model, such as those mentioned above, will be crucial in determining the robustness of this phenomenon, and will help elucidate our understanding of these processes.

Acknowledgments

We thank Erik Bodzsár, György Korniss, Géza Meszéna, Jasmine Nirody, Nathan Ross, Allan Sly, and Isabelle Stanton for helpful discussions. We also thank Oliver Riordan and an anonymous referee for useful comments on the manuscript.

References

- [1] T. Antunović, Y. Dekel, E. Mossel, and Y. Peres. Competing first passage percolation on random regular graphs. *Arxiv preprint arXiv:1109.2575*, 2011.
- [2] W.B. Arthur. Positive Feedbacks in the Economy. *Sci. Am.*, 262(2):92–99, 1990.
- [3] W.B. Arthur. *Increasing Returns and Path Dependence in the Economy*. The University of Michigan Press, 1994.
- [4] A. Banerjee and D. Fudenberg. Word-of-mouth learning. *Game. Econ. Behav.*, 46(1):1–22, 2004.
- [5] A.L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.
- [6] A. Barrat, M. Barthélemy, and A. Vespignani. *Dynamical Processes on Complex Networks*. Cambridge University Press, 2008.
- [7] M. Benaïm. Dynamics of stochastic approximation algorithms. *Séminaire de probabilités XXXIII*, pages 1–68, 1999.
- [8] M. Benaïm, I. Benjamini, J. Chen, and Y. Lima. A generalized Pólya's urn with graph based interactions. *Random Struct. Alg.*, 46(4):614–634, 2015.
- [9] M. Benaïm and M.W. Hirsch. Asymptotic Pseudotrajectories and Chain Recurrent Flows, with Applications. *J. Dynam. Differential Equations*, 8(1):141–176, 1996.
- [10] N. Berger, C. Borgs, J.T. Chayes, and A. Saberi. Asymptotic behavior and distributional limits of preferential attachment graphs. *The Annals of Probability*, 42(1):1–40, 2014.
- [11] B. Bollobás and O. Riordan. The diameter of a scale-free random graph. *Combinatorica*, 24(1):5–34, 2004.
- [12] M. Deijfen and R. van der Hofstad. The winner takes it all. *Arxiv preprint arXiv:1306.6467*, 2013.
- [13] T. Gross and B. Blasius. Adaptive coevolutionary networks: a review. *J. R. Soc. Interface*, 5(20):259–271, 2008.

- [14] B.M. Hill, D. Lane, and W. Sudderth. A strong law for some generalized urn processes. *Ann. Probab.*, 8(2):214–226, 1980.
- [15] M.W. Hirsch, S. Smale, and R.L. Devaney. *Differential Equations, Dynamical Systems, and An Introduction to Chaos*. Academic Press, 2004.
- [16] P. Holme and J. Saramäki. Temporal networks. *Phys. Rep.*, 519(3):97–125, 2012.
- [17] M. Lelarge. Diffusion and cascading behavior in random networks. *Games and Economic Behavior*, 75(2):752–775, 2012.
- [18] M.B. Nevelson and R.Z. Hasminskii. *Stochastic Approximation and Recursive Estimation*, volume 47 of *Translations of Mathematical Monographs*. American Mathematical Society, 1976.
- [19] H. Ohtsuki, C. Hauert, E. Lieberman, and M.A. Nowak. A simple rule for the evolution of cooperation on graphs and social networks. *Nature*, 441(7092):502–505, 2006.
- [20] R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.*, 86(14):3200–3203, 2001.
- [21] R. Pemantle. A time-dependent version of Pólya’s urn. *J. Theoret. Probab.*, 3(4):627–637, 1990.
- [22] R. Pemantle. Nonconvergence to unstable points in urn models and stochastic approximations. *Ann. Probab.*, 18(2):698–712, 1990.
- [23] R. Pemantle. When are touchpoints limits for generalized Pólya urns? *Proc. Amer. Math. Soc.*, 113(1):235–243, 1991.
- [24] R. Pemantle. A survey of random processes with reinforcement. *Probab. Surv.*, 4:1–79, 2007.
- [25] B.A. Prakash, A. Beutel, R. Rosenfeld, and C. Faloutsos. Winner takes all: competing viruses or ideas on fair-play networks. In *Proc. 21st Int. Conf. World Wide Web (WWW)*, pages 1037–1046. ACM, 2012.
- [26] S. Redner. How popular is your paper? an empirical study of the citation distribution. *Eur. Phys. J. B*, 4(2):131–134, 1998.
- [27] H. Robbins and S. Monro. A stochastic approximation method. *Ann. Math. Stat.*, 22(3):400–407, 1951.
- [28] M. Shaked and J.G. Shanthikumar. *Stochastic Orders*. Springer, New York, 2007.
- [29] B. Skyrms and R. Pemantle. A dynamic model of social network formation. *Proc. Nat. Acad. Sci. U.S.A.*, 97(16):9340–9346, 2000.
- [30] D.J. Watts. A simple model of global cascades on random networks. *Proc. Nat. Acad. Sci. U.S.A.*, 99(9):5766–5771, 2002.