Strategic diffusion in networks through contagion☆

Théophile T. Azomahou[a,c], Daniel C. Opolot[b,*]

[a] *Maastricht University, School of Business and Economics, Tongersestraat 53, 6211 LM Maastricht*
[b] *AIFMRM, University of Cape Town, Private Bag X3, Rondebosch, 7701 Cape Town*
[c] *University of Clermont Auvergne and CERDI, 65 Boulevard F. Mitterrand, 63009 Clermont-Ferrand CEDEX 1*

Abstract

Social networks provide a platform through which firms and governments can strategically diffuse products and desirable practices. This paper studies the process of strategic diffusion in the presence of network externalities using an evolutionary game theory framework. We show that firms and governments can exploit the notion of contagion and tendency of decision makers to experiment on choices to strategically diffuse products and practices. A contagious product requires a small set of initial adopters to trigger diffusion to the wider network. Thus, a firm can reduce costs associated with targeting and advertising by strategically making her product contagious. We also show that the expected waiting time until a contagious choice is adopted by the entire network is independent of the population size. This implies that in large networks, even if the level of experimentation is very low, the diffusion process does not get trapped indefinitely in a suboptimal equilibrium.

Keywords: Strategic diffusion, networks, contagion, expected waiting time
JEL: D8, C73

1. Introduction

The diffusion of products and practices through social interactions has been recognized for the past five decades.[1] The growth of internet and online social networks have increased interest in how firms and governments can harness the power of social networking to promote private and social goals. Recent research has examined how firms and governments can target specific individuals in a network to trigger diffusion of a product or behaviour to a larger proportion of the population. However, the existing literature on targeting ignores the tendency of agents to experiment or make mistakes in their choices of products and practices. How does experimentation affect diffusion? Can firms and governments exploit agents' tendency to experiment to reduce the potential costs to targeting?

Although the existing literature provides important intuition on the conditions for successful diffusion, the results are based on a binary $\{0, 1\}$ choice diffusion process on large random networks. Considering

*Corresponding author:
Email address: daniel.opolot@uct.ac.za (Daniel C. Opolot)

[1] See for example Coleman et al. (1966) on the diffusion of medical innovations, Foster and Rosenzweig (1995); Conley and Udry (2010) on diffusion of agricultural products and practices, Reingen et al. (1984); Godes and Mayzlin (2004) on diffusion of brand choices by consumers.

large random networks (where the population size is assumed to be infinitely large) permits the analyst to focus on the effect of aggregated network variables such as density, average and variance of neighbourhood sizes. These measures do not however capture local variations within the network, which are important in situations where a firm wishes to diffuse her product to specific sub-regions and not to the entire network. What alternative network measures could firms or governments use in this case?

This paper addresses these questions through an evolutionary game theory framework. We suppose that there is a finite set of agents who interact through a social network. Agents repeatedly revise their choices/actions over time, and their choices exhibit network externalities. Network externalities exist in adoption of information technologies, fashion products, social and institutional norms, and political actions, and are characterized by inertia in that nobody adopts if no one else has adopted.

We consider an evolutionary process where the5 probability of choosing a given option is separable into components capturing social influence and individual experimentation. Including experimentation in the decision making is a standard assumption in evolutionary game theory and evolutionary economics. It is akin to the models of *bounded rationality*: the notion that although individuals aim to make optimal choices, their decisions are frequently subject to mistakes. In the well-studied Bass and related diffusion models, experimentation is interpreted as innovation so that the society can be divided into two categories based on the timing of adoption: *innovators* as early adopters, and *imitators* as late adopters. The social influence component of choice probabilities captures network externalities. We assume that agents *myopically* best respond to the distribution of their neighbours' choices. Assuming myopia in the decision making process is consistent with the notion of bounded rationality in that agents are not only prone to mistakes but are also incapable of keeping memory of the entire history of play.

There are several key differences between our model and existing models of diffusion in the literature. Firstly, in contrast to the Bass and related models of diffusion for first-time purchases (see Mahajan et al. (1990) and Rogers (2003) for surveys), we model a diffusion process where agents repeatedly replace the old, spoilt or no longer preferred products with new ones of the same type/brand, or with different competing products. The role that the process of experimentation plays in triggering network-wide diffusion is thus relevant in our model and less so in the models of first-time purchases. Secondly, in contrast to the mean field models of diffusion (e.g. López-Pintado (2008), Jackson and Yariv (2007),Galeotti and Goyal (2009) and Kreindler and Young (2014)) that assume infinitely large and homogeneously mixed population, we study a stochastic diffusion process on finite networks. Due to the stochastic nature of our model, the long-run probabilities are a suitable measure for characterizing the diffusion process. Moreover, unlike mean field diffusion processes that generate smooth diffusion curves, finite-population diffusion processes exhibit threshold dynamics. Our model is thus useful to firms and governments that care about the long-run survival and not just first-time purchases of their products and practices.

Our analysis proceeds by first examining the effect of the level of experimentation on long-run probabilities. We find that long-run probabilities exhibit threshold dynamics with respect to the experimentation parameter. For any two configurations say $\mathbf{x}$ and $\mathbf{y}$ describing a profile of choices in the population, there exists an optimal level of experimentation above which a switch in the relationship between long-run probabilities of $\mathbf{x}$ and $\mathbf{y}$ occurs. Above the optimal level of experimentation, best response dynamics starts to dominate the tendency to experiment, and network externalities become pronounced.

Next, we show how firms and governments can exploit the process of *contagion* to diffuse their products and practices. Contagion refers to the process where a choice spreads to a larger subregion or entire network through best responds dynamics once it has been adopted by a sufficiently small number of agents, here referred to as *initial adopters*. The properties that determine the feasibility of contagion are: (i) a unique property of the network called *contagion threshold* and (ii) a property of payoffs called the *relative payoff gain* between any pair of choices. Contagion occurs whenever the relative payoff gain is less than the contagion threshold. We show that contagious choices are played with the highest probability in the long-run. Given a network and its contagion threshold, a firm aiming to diffuse a product can then strategically adjust relative payoff gains by either increasing the relative level of innovation on their product or reducing its relative cost.

We then show that if a choice that is contagious relative to others exists, the diffusion process is fast. That is, the expected waiting time from any configuration to a configuration where all agents choose a contagious choice is independent of the population size. The implication is that in large networks, even if the level of experimentation is very low, the diffusion process does not get trapped indefinitely in an equilibrium where agents choose a non-contagious option.

Overall, our results show that a firm caring about the long-run survival of her product can rely, at least in part, on experimentation by consumers to reduce potential costs to targeting and/or advertising. Contagion, by definition, implies that the set of initial adopters is very small. A firm can thus also reduce the potential costs of targeting by strategically making her product contagious.

Our work is related to a large literature across disciplines on how firms and governments can harness the power of social networks to promote products and desirable practices. It includes Richardson and Domingos (2002); Kempe et al. (2003); Chen et al. (2009) in computer science, Kirby and Marsden (2006); Kiss and Bichler (2008) in marketing, and Kelly et al. (1991); Valente et al. (2003) in public health. A closely related literature in economics is that of diffusion in the presence of network externalities (Morris, 2000; López-Pintado, 2008; Jackson and Yariv, 2007; Sundararajan, 2007; Galeotti et al., 2010; Galeotti and Goyal, 2009). With the exception of Morris (2000) these papers study binary choice diffusion processes on random networks, and the differences with our paper are discussed above. Morris (2000) studies conditions for pairwise contagion in deterministic networks, and our paper generalizes the notion of pairwise contagion and shows how firms can strategically employ it to diffuse products.

Our paper is also related to evolutionary game theory literature that studies convergence rates. For example Ellison (1993, 2000) shows that learning is fast in some families of networks, Montanari and Saberi (2010); Young (2011) show that diffusion is fast in networks made up of cohesive subgroups, and

Kreindler and Young (2013, 2014) show that learning is fast if the level of experimentation is sufficiently large. We show that learning is fast in the presence of a contagious choice.

The remainder of the paper is organized as follows. In Section 2, we introduce a model of diffusion in the presence of network externalities. Section 3 characterizes long-run probabilities, and shows how they vary with the parameter of experimentation. Section 4 discusses the notion of contagion and shows how it can be used to diffuse a specific choice. Section 5 examines the relationship between contagion and expected waiting time, and a conclusion is offered in Section 6. All lengthy proofs are relegated to the Appendix.

2. The model

2.1. Actions, network and payoffs

We study the diffusion of products and behaviour with strategic complementarities in social networks. We focus on situations where choices exhibit coordination properties. Let $A = \{a_1, \cdots, a_j, \cdots, a_m\}$ be the action or choice set, and $U^{m\times m}$ is an $m \times m$ symmetric payoff matrix whereby $u(a_j, a_l)$ is a payoff to choice a_j when an opponent chooses a_l. If choices exhibit coordination property, then for each $a_j \in A$, $u(a_j, a_j) > u(a_j, a_l)$ for all $a_l \neq a_j$; that is, coordinating on a given choice has higher payoff than miscoordination.

Let $N = \{1, \cdots, i, \cdots, n\}$ denote a set of agents who interact through a social network represented by an $n \times n$ interaction matrix W. The elements $w_{ik} \in [0, 1]$ of W represent the weight or level of importance that agent i attaches to the choices of her neighbour k. Let small bold letters e.g. $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$,$\cdots$ denote vectors or configurations of choices in the population. Given N, let $\mathbf{X}_n$, or simply $\mathbf{X}$ where no confusion arises, be a set of all possible configurations. The cardinality of $\mathbf{X}_n$ is m^n.We consider the case of linear payoffs. That is, the payoff to i for choosing a_j given configuration $\mathbf{x}$ is

$$U_i(a_j, \mathbf{x}) = \sum_{k=1}^{n} w_{ik} u(a_j, x^k), \tag{1}$$

where x^k is the k^{th} coordinate of $\mathbf{x}$. Let $w_i = \sum_{k=1}^{n} w_{ik}$ be the total direct influence of other agents on i, and define $w_i(a_j; \mathbf{x})$ as the total direct influence on i by agents choosing a_j in configuration $\mathbf{x}$. That is,

$$w_i(a_j; \mathbf{x}) = \left\{ \sum_{k=1}^{n} w_{ik} : \; x^k = a_j \right\}$$

Then (1) can equivalently be written as

$$U_i(a_j, \mathbf{x}) = \sum_{a_l \in A} w_i(a_l; \mathbf{x}) u(a_j, a_l), \tag{2}$$

We then refer to a triple $(N, A, U^{m\times m})$ as a coordination game.

2.2. Dynamics

Given payoff and interaction structures, we consider a diffusion process where agents revise their choices over time. We model the diffusion of products where agents can replace the old, spoilt or no longer preferred products with new ones of the same type/brand, or with different competing products. This is opposed to the Bass diffusion model and related models for first-time purchases (Bass, 1969). But as in the Bass and related diffusion models, we assume that the probability that an agent chooses a given option is a function of social influence and individual experimentation.

Let $\mathbf{x}_t$ be the configuration at period t where $t = 1, 2, \cdots$. The probability $\mathbb{P}_i(a_j; \mathbf{x}_t)$ that agent i chooses a_j in the next period given the current configuration $\mathbf{x}_t$, is given by

$$\mathbb{P}_i(a_j; \mathbf{x}_t) = \frac{1}{m} \exp(-\beta) + (1 - \exp(-\beta)) BR_i(a_j; \mathbf{x}_t) \tag{3}$$

The first term $\frac{1}{m} \exp(-\beta)$ captures individual experimentation on choices and is independent of the current configuration $\mathbf{x}_t$. More specifically, an agent decides to experiment with a probability $\varepsilon = \exp(-\beta)$, common to all agents. If an agent experiments, she chooses any option with a uniform probability $\frac{1}{m}$. The second term, $BR_i(a_j; \mathbf{x}_t)$ captures social influence of other agents on i, with $1 - \varepsilon = 1 - \exp(-\beta)$ as a measure of the strength of social influence. We consider the case where agents best respond to their neighbours' choices. Let $BR(\mathbf{x}_t)$ be the set of choices that are a best response to the configuration $\mathbf{x}_t$. That is

$$BR(\mathbf{x}_t) = \{a_j : U_i(a_j, \mathbf{x}_t) \geq U_i(a_l, \mathbf{x}_t), a_j, a_l \in A\}.$$

Then $BR_i(a_j; \mathbf{x}_t)$ is the probability that i chooses a_j through best response given configuration $\mathbf{x}_t$. That is, if $a_j \in BR(\mathbf{x}_t)$ and b is the cardinality of $BR(\mathbf{x}_t)$, then $BR_i(a_j; \mathbf{x}_t) = \frac{1}{b}$, and $BR_i(a_j; \mathbf{x}_t) = 0$ if $a_j \notin BR(\mathbf{x}_t)$. In probabilistic terms, $1 - \varepsilon$ is then interpreted as the probability that an agent follows a best response behaviour.

We assume that agents revise their choices simultaneously at discrete intervals $t = 1, 2, \cdots$. Let $P_\beta(\mathbf{x}, \mathbf{y}) = \mathbb{P}(\mathbf{x}_{t+1} = \mathbf{y};\ \mathbf{x}_t = \mathbf{x})$ be the probability that configuration $\mathbf{x}$ is followed by configuration $\mathbf{y}$ after a single period of iteration, where the subscript β represents the parameter of experimentation defined above. Simultaneous revision of choices implies that

$$P_\beta(\mathbf{x}, \mathbf{y}) = \mathbb{P}(\mathbf{x}_{t+1} = \mathbf{y};\ \mathbf{x}_t = \mathbf{x}) = \prod_{i=1}^{n} \mathbb{P}_i \left(y^i; \mathbf{x}_t = \mathbf{x} \right) \tag{4}$$

The dynamics in (4) follows a Markov chain on the configuration space $\mathbf{X}_n$. We denote by P_β for the Markov transition matrix with $P_\beta(\mathbf{x}, \mathbf{y})$, for all $\mathbf{x}, \mathbf{y} \in \mathbf{X}_n$, as its elements. Our analysis of the diffusion process defined above will then rely on the well-known properties of Markov chains. With slight abuse of meaning, we throughout the paper refer to the quadruple $(N, A, U^{m \times m}, W)$ as a *coordination game diffusion process.*

We proceed by first examining the long-run behaviour of the diffusion process, and in particular, the limit configurations and their long-run probabilities. Relying on properties of long-run probabilities, we show how a firm or planner can diffuse her product or desired behaviour to a larger fraction of the population through contagion. The process of contagion is directly influenced by network and payoff structures. We then show that products that satisfy the property of contagion diffuse fast. Our approach is different from the mean field models of diffusion on random networks, where agents make choices sequentially (López-Pintado, 2006, 2008; Jackson and Yariv, 2007; Kreindler and Young, 2013, 2014). These models produce a smooth diffusion curve originally studied in Bass (1969). When agents follow a simultaneous revision protocol, the diffusion process exhibits threshold dynamics. Moreover, we study diffusion processes where agents repeatedly revise their choices over time. Examining long-run probabilities is thus a more appropriate way to study the diffusion process defined in this paper. Our results are then useful to firms that care about their long-run survival in the market.

3. Long-run behaviour

3.1. Limit configurations of the Markov chain

Before examining the long-run behaviour of P_β for $\beta < \infty$, it helps to first examine the properties of the special case where $\beta = \infty$. Let $P = P_\infty$. The diffusion process with transition matrix P corresponds to the situation where agents do not experiment and only make choices based on best response evaluations. The equilibrium configurations of P are equivalent to its limit or absorbing configurations. Let $\mathbf{L}$ be a set of all limit configurations of P, and L_j for a typical element of $\mathbf{L}$; we also write $\mathbf{a}_l$ for a limit configuration where all agents coordinate on choice a_l. The composition of $\mathbf{L}$ depends on the payoff and network structures. Generally, it consists of all monomorphic configurations $\mathbf{a}_l$ for $l = 1, \cdots, m$. Depending on the network structure, $\mathbf{L}$ may also contain configurations where choices co-exists in the population. The following example illustrates how such configurations may arise.

Example 1: Consider payoffs for a bilateral interaction between the row and column players depicted in Table 1. The underlying payoffs exhibit coordination properties, and the bilateral interaction is a coordination game. This game has three Nash equilibria: Two of which are pure strategy Nash equilibria where players coordinate on choices a_1 and a_2, and the third is a mixed-strategy Nash equilibrium in which both the row and column players choose a_1 with probability $\phi_{21} = \frac{3}{5}$ and a_2 with probability $\phi_{12} = \frac{2}{5}$.[2]

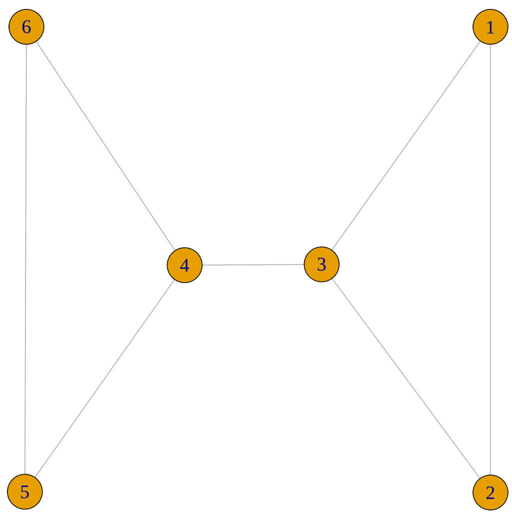
[2]This follows from the fact that if ϕ_{12} and $\phi_{21} = 1 - \phi_{12}$ are the probabilities with which the column player chooses a_2 and a_1 respectively so that the row player's payoff from choosing a_1 and a_2 respectively are $\phi_{12}u(a_1, a_2) + (1 - \phi_{12})u(a_1, a_1)$ and $\phi_{12}u(a_2, a_2) + (1 - \phi_{12})u(a_2, a_1)$, then the mixed strategy equilibrium is a solution to $\phi_{12}u(a_1, a_2) + (1 - \phi_{12})u(a_1, a_1) = \phi_{12}u(a_2, a_2) + (1 - \phi_{12})u(a_2, a_1)$, which yields

$$\phi_{12} = \frac{u(a_1, a_1) - u(a_2, a_1)}{u(a_1, a_1) - u(a_2, a_1) + u(a_2, a_2) - u(a_1, a_2)} = \frac{2}{5}$$

$$\phi_{21} = \frac{u(a_2, a_2) - u(a_1, a_2)}{u(a_1, a_1) - u(a_2, a_1) + u(a_2, a_2) - u(a_1, a_2)} = \frac{3}{5}$$

	a_1	a_2
a_1	$4, 4$	$0, 2$
a_2	$2, 0$	$3, 3$

Table 1: A 2-player 2-choice coordination game.



$$W = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 1/3 & 1/3 & 0 & 1/3 & 0 & 0 \\ 0 & 0 & 1/3 & 0 & 1/3 & 1/3 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1/2 & 1/2 & 0 \end{bmatrix}$$

Figure 1: An example of a network with cohesive subgroups, and the respective interaction matrix W on the right hand side.

For any interaction structure, the monomorphic configurations $\mathbf{a}_1 = (a_1, a_1, a_1, a_1, a_1, a_1)$ and $\mathbf{a}_2 = (a_2, a_2, a_2, a_2, a_2, a_2)$ are limit configurations of P. Now consider the interaction structure in Figure 1 with the interaction matrix W as defined. For this network, and payoff in Table1, in addition to limit configurations $\mathbf{a}_1$ and $\mathbf{a}_2$, the configurations $\mathbf{a} = (a_1, a_1, a_1, a_2, a_2, a_2)$ and $\mathbf{b} = (a_2, a_2, a_2, a_1, a_1, a_1)$ where choices co-exist are also limit configurations of P. To see why, first note that the quantities ϕ_{12} and ϕ_{21} are the minimum proportion of neighbours an agent requires to adopt a_2 and a_1 respectively for her to do likewise.

Similarly, $\lceil \phi_{12} k_i \rceil$ and $\lceil \phi_{21} k_i \rceil$, where $\lceil x \rceil$ is the smallest integer larger than x, are the numbers of neighbours agent i requires to choose a_2 and a_1 respectively for i to do likewise. For the game in Table 1 and network in Figure 1, it follows that agents $\{1, 2, 5, 6\}$ who each have degree $k = 2$, will switch from a_1 to a_2 if at least $\lceil \frac{4}{5} \rceil = 1$ of their neighbours choose a_2 and from a_2 to a_1 if $\lceil \frac{6}{5} \rceil = 2$ of their neighbours choose a_1. The bottleneck however occurs with agents $\{3, 4\}$ who each require at least $\lceil \frac{6}{5} \rceil = 2$ of their neighbours to choose a_2 for them to do likewise. This in turn makes the configurations $\mathbf{a} = (a_1, a_1, a_1, a_2, a_2, a_2)$ and $\mathbf{b} = (a_2, a_2, a_2, a_1, a_1, a_1)$ Nash equilibria and absorbing configurations of P.

3.2. Long-run probabilities

The long-run behaviour of a Markov chain is defined in terms of its *stationary distribution*. The stationary distribution describes the fractional amount of time the process spends in each configuration in the long-run. Or equivalently, the probability with which each configuration is visited in the long-run. Formally, let $\mathbf{q}_0$ be an m^n-row vector representing an initial distribution of a Markov chain. For example if the chain starts from a configuration where all agents coordinate on a_j, then $\mathbf{q}_0$ is a vector of all zeros except a one in configuration $\mathbf{a}_j$. After t iterations, the distribution is $\mathbf{q}_t = \mathbf{q}_0 P_\beta^t$. The stationary distribution π_β is then an m^n-vector defined as $\pi_\beta = \lim_{t\to\infty} \mathbf{q}_0 P_\beta^t$.

Now, consider the case when $\beta = \infty$. The corresponding transition matrix P can be partitioned into block matrices whose number is equal to the number of limit configurations. Each block matrix describes the transitions within configuration of the *basin of attraction* of a limit configuration. A basin of attraction of limit configuration L_j, here denoted by $D(L_j)$, is the set of all configurations from which the Markov chain can converge to L_j through best response dynamics; that is, in the absence of experimentation. Clearly, if π is the stationary distribution of P, then for each subset $D(L_j)$, $\sum_{\mathbf{x}\in D(L_j)} \pi(\mathbf{x}) = 1$. Since the partitions $D(L_j)$ of the configuration space $\mathbf{X}_n$ are disjoint, that is $D(L_j)\cap D(L_l) = \emptyset$ for all $L_j, L_l \in \mathbf{L}$, the dynamics within each $D(L_j)$ can thus be studied independently. The direct implication is that the long-run behaviour of P depends on the starting configuration. And once the process reaches a limit configuration, which may include configurations where choices co-exist, it gets trapped in it indefinitely.

The tendency of agents to experiment or make errors in their choices, however, ensures that in the long-run, every configuration is reached with a positive probability; in this situation, $\pi_\beta(\mathbf{x}) > 0$ for all $\mathbf{x} \in \mathbf{X}_n$. Our goal in this section is to examine how the parameter of experimentation, and the interaction between payoffs and network structure affect long-run probabilities. After which, we examine strategies a firm or planner can use to increase the long-run probability of their product or desired behaviour being adopted. The first step in the following analysis is a discussion of methods for characterizing the stationary distribution of a Markov chain. The following definitions and notations are used in the analysis that follows.

Definition 1. *Let $g \subset \mathbf{X}\times\mathbf{X}$ be any oriented graph defined within the configuration space $\mathbf{X}$. Then for a subset $W \subset \mathbf{X}$ and its complement $\bar{W}$, we denote by $G(W)$ a set of all oriented graphs satisfying two conditions: (i) no arrows start from W and exactly one arrow starts from each configuration outside of W, (ii) each $g \in G(W)$ has no loops.*

From Definition 1, if W is a singleton set, say $W = \{\mathbf{x}\}$, then $G(\{\mathbf{x}\})$ is a set of all spanning trees of $\mathbf{x}$, also known as $\mathbf{x}$-trees. Consider the case where $\mathbf{X} = \{\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}, \mathbf{e}, \mathbf{f}, \mathbf{g}, \mathbf{h}\}$; Figure 2 presents two examples of $\mathbf{g}$-trees.

Let $d(L_j)$ be the cardinality of $D(L_j)$. For each $D(L_l)$, let $\partial L_l = D(L_l)\backslash L_l$, the basin of attraction of L_l with L_l excluded. Each graph $g \in G(\{\mathbf{x}\})$ can be partitioned into subgraphs $g(D(L))$ and $g(\partial L, L)$

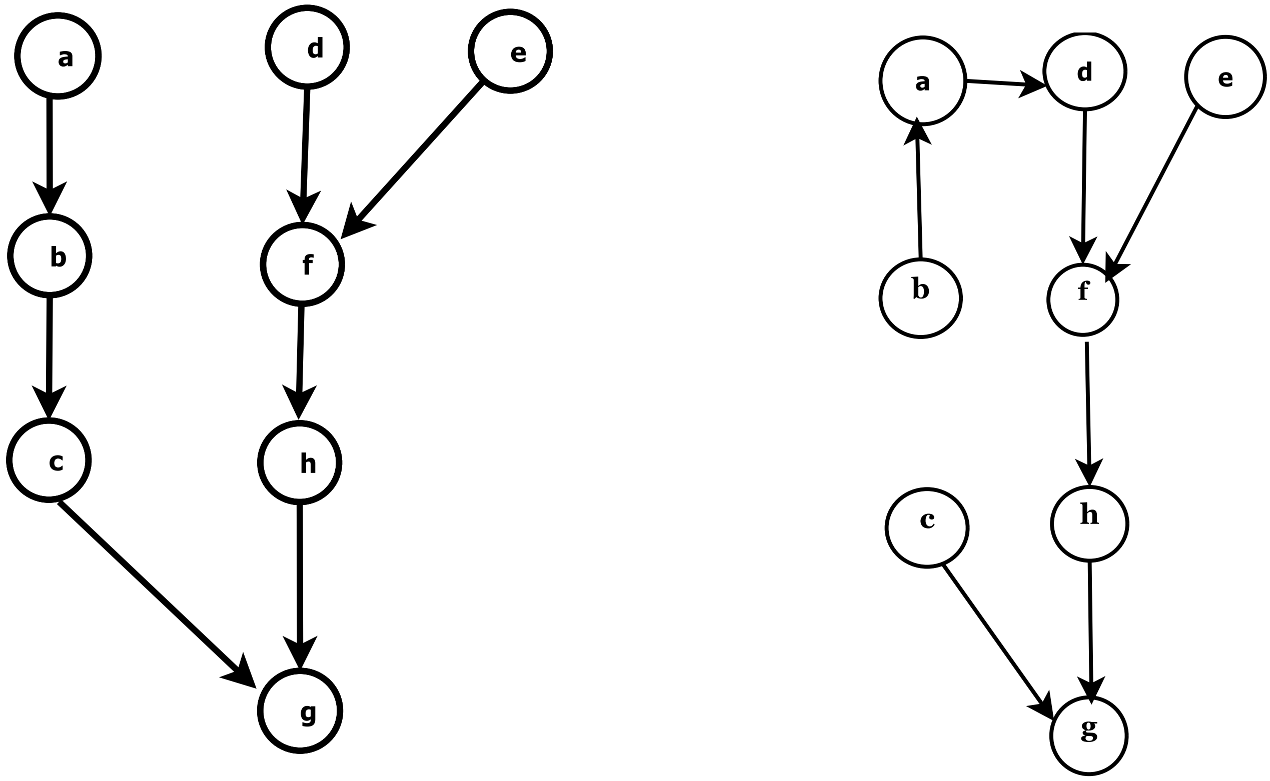


Figure 2: Examples of $\mathbf{g}$-trees, that is $G(\{\mathbf{g}\})$ graphs and in which the configuration space $\mathbf{X} = \{\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}, \mathbf{e}, \mathbf{f}, \mathbf{g}, \mathbf{h}\}$.

defined as follows. For some $L_j \in \mathbf{L}$ and corresponding subset of configurations $D(L_j)$, $g(D(L_j))$ is a set of subgraphs of g consisting of transitions between either the configurations in $D(L_j)$ or from a configuration in $D(L_j)$ to another not in $D(L_j)$, and that each of these transitions involves at least one experimentation. The subgraph $g(\partial L_j, L_j)$ on the other hand consists of transitions of g among configurations of $D(L_j)$ that involve best response. For each $g \in G(\{\mathbf{x}\})$, let $r_g(L_j)$ be the total number of experimentations involved in the transitions of sub-graph $g(D(L_j)) \in g$ and let $r_g(\partial L_j, L_j)$ be the cardinality of $g(\partial L_l, L_j)$.

For any $g \in G(\{\mathbf{x}\})$, define the probability $P_\beta(g)$ as

$$P_\beta(g) = \prod_{(\mathbf{z},\mathbf{y})\in g} P_\beta(\mathbf{z}, \mathbf{y})$$

The quantity $P_\beta(g)$ captures the likelihood of reaching $\mathbf{x}$ from every other configuration through the paths contained in g. Let $g_{\min}(\mathbf{x}) = \underset{g\in G(\{\mathbf{x}\})}{\operatorname{argmax}} P_\beta(g)$ be the $\mathbf{x}$-tree with the largest likelihood $P_\beta(g)$, where we write $g_{\min}(\mathbf{x})$ to imply an $\mathbf{x}$-tree with the shortest path, which corresponds to largest likelihood. Let $r_{\mathbf{x}}(\partial L_j, L_j)$ and $r_{\mathbf{x}}(L_j)$ be the corresponding values of $r_g(\partial L_j, L_j)$ and $r_g(L_j)$ respectively when $g = g_{\min}(\mathbf{x})$.

Lemma 1. *For a coordination game diffusion process $(N, A, U^{m\times m}, W)$, let $\mathbf{x}, \mathbf{y} \in \mathbf{X}_n$ be any pair of configurations. The ratio of stationary distributions of $\mathbf{x}$ and $\mathbf{y}$ is bounded by*

$$\frac{\pi_\beta(\mathbf{x})}{\pi_\beta(\mathbf{y})} = \exp\left[-\sum_{L_j\in\mathbf{L}} \Big(\big(r_{\mathbf{x}}(L_j) - r_{\mathbf{y}}(L_j)\big)\big(\beta_m - \beta'_m\big)\Big) \pm \kappa\right] \tag{5}$$

where $\beta_m = \beta - \ln m^{-1}$, $\beta'_m = -n \ln \left[1 - e^{-\beta}\left(1 - \frac{1}{m}\right)\right]$, *and* κ *is a uniform constant.*

Proof. See Appendix A.1

Lemma 1 provides bounds for the ratio of stationary distributions of any pair of configurations. It uses the highest probability graphs argument developed in Freidlin and Wentzell (1984). The quantities $r_{\mathbf{x}}(L_j)$ for each $L_j \in \mathbf{L}$ is the number of experimentations required to exit the basin of attraction $D(L_j)$ of L_j if evolution from L_j to $\mathbf{x}$ follows the minimum $\mathbf{z}$-tree. The quantity $C_{\mathbf{x}}(\beta) = -\sum_{L_j \in \mathbf{L}} r_{\mathbf{x}}(L_j)(\beta_m - \beta'_m)$, is the total cost of the minimum $\mathbf{x}$-tree, and $\mathcal{E}_{\mathbf{x}} = \sum_{L_j \in \mathbf{L}} r_{\mathbf{x}}(L_j)$ is the total number of experimentations along the minimum $\mathbf{x}$-tree. The larger the cost of reaching $\mathbf{x}$, the smaller the associated long-run probability.

For any configuration $\mathbf{x}$, the cost $C_{\mathbf{x}}(\beta)$ is a non-monotonic function of experimentation parameter β. It is separable into $-\sum_{L_j \in \mathbf{L}} r_{\mathbf{x}}(L_j)\beta_m$, the cost of reaching $\mathbf{x}$ through experimentation, and $\sum_{L_j \in \mathbf{L}} r_{\mathbf{x}}(L_j)\beta'_m$, the cost of reaching $\mathbf{x}$ through best response dynamics. At the maximum level of experimentation, when $\beta = 0$, the cost of reaching $\mathbf{x}$ through best response dominates and is equal to $n \ln(m) \sum_{L_j \in \mathbf{L}} r_{\mathbf{x}}(L_j)$, while that through experimentation is $-\ln(m) \sum_{L_j \in \mathbf{L}} r_{\mathbf{x}}(L_j)$. The former increase with the total number of experimentations $\mathcal{E}_{\mathbf{x}}$ and the population size n, while the latter decreases with $\mathcal{E}_{\mathbf{x}}$. The positive relationship between $n \ln(m) \sum_{L_j \in \mathbf{L}} r_{\mathbf{x}}(L_j)$ and $\mathcal{E}_{\mathbf{x}}$, and n, at small values of β, is because agents make decision entirely by experimentation so that the larger the number of experimentations required to reach $\mathbf{x}$, the smaller the cost; and the larger the population size n, the higher the likelihood of ending up with $\mathcal{E}_{\mathbf{x}}$ experimentations.

As the level of experimentation decreases, there exists an optimal value of β denoted by β^*, so that for any $\beta > \beta^*$, the cost of reaching $\mathbf{x}$ through experimentation starts to dominate; that is, $C_{\mathbf{x}}(\beta) < 0$. The optimal β is solution to the equation $\beta_m = \beta'_m$, where β'_m is a increasing function of n. Figure 3 plots β^* with n for $m = 4$, showing that it increases logarithmically.

Lemma 1 states that the ratio of stationary distributions of any two configurations is an exponential function of the difference of the total costs of their minimum graphs. Let $\Phi_\beta(\mathbf{x}, \mathbf{y})$ denote the respective cost difference between $\mathbf{x}$ and $\mathbf{y}$; that is

$$\Phi_\beta(\mathbf{x}, \mathbf{y}) = -(\beta_m - \beta'_m) \sum_{L_j \in \mathbf{L}} \left((r_{\mathbf{x}}(L_j) - r_{\mathbf{y}}(L_j))\right)$$

If $\Phi_\beta(\mathbf{x}, \mathbf{y}) < 0$, then the total cost of reaching $\mathbf{x}$ is higher (in a negative sense) than the cost of reaching $\mathbf{y}$, and vice versa when $\Phi_\beta(\mathbf{x}, \mathbf{y}) > 0$. Consequently, when $\Phi_\beta(\mathbf{x}, \mathbf{y}) < 0$, the long-run probability of configuration $\mathbf{x}$ is less than the long-run probability of $\mathbf{y}$. The corollary below then follows from Lemma 1.

Corollary 1. *For a coordination game diffusion process* $(N, A, U^{m \times m}, W)$*, let* $C_{\mathbf{x}}(\beta)$ *and* $C_{\mathbf{y}}(\beta)$ *be the respective costs associated with the minimum graphs of configurations* $\mathbf{x}, \mathbf{y} \in \mathbf{X}_n$*. There exists a* $\beta^* > 0$ *so that for all* $\beta > \beta^*$*, if* $C_{\mathbf{x}}(\beta) < C_{\mathbf{y}}(\beta)$*, then* $\pi_\beta(\mathbf{x}) < \pi_\beta(\mathbf{y})$

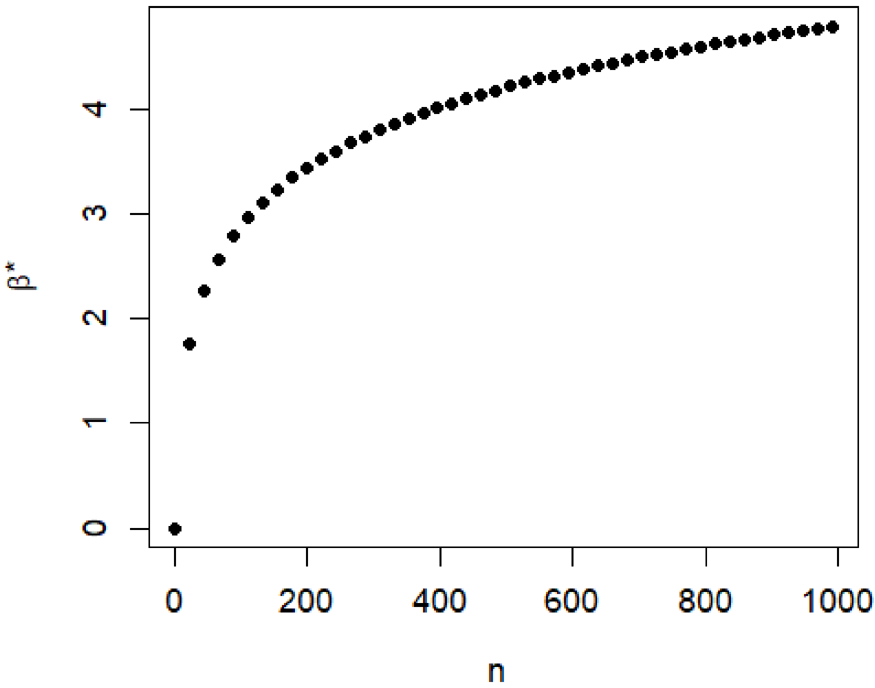


Figure 3: A plot of β^* against n, for $m = 4$.

The direct implication of Corollary 1 is that any economic agent (e.g. a firm or a planner) aiming to diffuse a product or behaviour exhibiting coordination properties needs to have knowledge of two aspects of the diffusion process. First, the level of experimentation by the decision makers. In this case it suffices to know whether $\beta < \beta^*$ or $\beta > \beta^*$. Second, given the level of experimentation, it is necessary to have knowledge or a good estimate of $\mathcal{E}_{\mathbf{x}}$, the total number of experimentations involved in the shortest paths to other configurations, especially for configurations in which a majority of agents choose the competing product or behaviour. If agents are rational, but in a bounded sense (which is a common assumption in evolutionary game theory), so that $\beta > \beta^*$, then it suffices to have only knowledge of $\mathcal{E}_{\mathbf{x}}$. The quantities $\mathcal{E}_{\mathbf{x}}$ for all $\mathbf{x}$ are computable from the interaction between payoffs and the network structures. In the next section, we show how a firm or a planner can exploit the process of contagion to increase the probability of their product or behaviour relative to another.

4. Diffusion through contagion

This section examines how a firm or a planner can exploit the properties of the network and payoffs to increase the long-run probability of a product or behaviour. In particular we show that if the network and payoff structures are known, then a firm can exploit the process of *contagion* to diffuse their product. Contagion is defined on a pairwise level but can be generalized to multiple choices. That is, for a pair of choices $a_j, a_l \in A$, choice a_l is pairwise contagious relative to a_j ($a_l \mathrm{PRC} a_j$) if starting from a monomorphic configuration $\mathbf{a}_j$, choice a_l can spread through best response once it has been adopted by a small set of agents called *initial adopters*. In the Bass and related diffusion models, the set of initial adopters is normally referred to as innovators, where experimentation behaviour is interpreted as innovativeness.

Formally, for any pair $a_j, a_l \in A$, let $N(\mathbf{a}_j \to \mathbf{a}_l)$ be the set of initial adopters required to trigger contagion from a_j to a_l. Let $n(\mathbf{a}_j \to \mathbf{a}_l)$ be the respective cardinality. Let also $n(a_j; \mathbf{x}_t)$ be the number of agents in the population choosing option a_j when the configuration at t is $\mathbf{x}_t$.

Definition 2. *For a pair of choices $a_j, a_l \in A$, let $\beta > \beta^*$ and for n sufficiently large, let $n(\mathbf{a}_j \to \mathbf{a}_l)$ be independent of n and that $n(\mathbf{a}_j \to \mathbf{a}_l) \ll \frac{1}{2}n$. A choice a_l is said to be pairwise contagious relative to a_j ($a_l PRC a_j$) if $n(a_j; \mathbf{x}_{t+1}) > n(a_j; \mathbf{x}_t)$ for all $t \geq t'$ whenever $n(a_l; \mathbf{x}_{t'}) = n(\mathbf{a}_j \to \mathbf{a}_l)$.*

There are two conditions in the definition of contagion above. First, contagion is normally defined for best response dynamics in the absence of experimentation (see for example Morris (2000)). As such, we require $\beta > \beta^*$ so that contagion occurs if best response dynamics dominates dynamics due to experimentation. Moreover, a reasonable assumption that is always made in evolutionary game theory and economics is that agents are boundedly rational. That is, they maximise payoffs and make mistakes with a small probability. Second, the set of initial adopters must be small, and its size independent of n. If the set of initial adopters is close to $\frac{1}{2}n$ and/or grows with n, then evolution from $\mathbf{a}_j$ to $\mathbf{a}_l$ will be dominated by experimentation rather than best response.

Given Definition 2, we next establish conditions under which pairwise contagion is feasibility in $m \times m$ coordination games. The are two measures required in establishing conditions for contagion. The first measure is the *relative payoff-gain* ϕ_{jl} defined defined for each pair of choice $a_j, a_l \in A$ as

$$\phi_{jl} = \frac{\delta_{jl}(a_j)}{\delta_{jl}(a_j) + \delta_{lj}(a_l)}$$

where $\delta^i_{kl}(\mathbf{a}_j) = U_i(a_k, \mathbf{a}_j) - U_i(a_l, \mathbf{a}_j) = u(a_k, a_j) - u(a_l, a_j) = \delta_{kl}(a_j)$, is identical for all agents. In 2×2 coordination games, ϕ_{jl}, defined for a pair of choices a_j and a_l, is the minimum proportion of neighbours each agent requires to choose a_l for her to do likewise given that the remaining proportion choose a_j.[3]

[3]The relation follows from considering the best response to $\mathbf{a}_j$ for distributions where $w_i(a_l; \mathbf{a}_j)$ is the total weight that i attaches to her neighbours choosing option a_l and $w_i - w_i(a_l; \mathbf{a}_j)$ choose option a_j, and zero otherwise. For such a distribution, agent i will choose a_l if

$$(w_i - w_i(a_l; \mathbf{a}_j))u(a_l, a_j) + w_i(a_l; \mathbf{a}_j)u(a_l, a_l) \geq (w_i - w_i(a_l; \mathbf{a}_j))u(a_j, a_j) + w_i(a_l; \mathbf{a}_j)u(a_j, a_l)$$

Dividing by w_i and solving for $\frac{w_i(a_l; \mathbf{a}_j)}{w_i}$ yields

$$\frac{w_i(a_l; \mathbf{a}_j)}{w_i} \geq \frac{u(a_j, a_j) - u(a_l, a_j)}{u(a_j, a_j) - u(a_l, a_j) + u(a_l, a_l) - u(a_j, a_l)} = \frac{\delta_{jl}(a_j)}{\delta_{jl}(a_j) + \delta_{lj}(a_l)}$$

Hence agent i will choose a_l whenever at least

$$\phi_{jl} = \frac{\delta_{jl}(a_j)}{\delta_{jl}(a_j) + \delta_{lj}(a_l)}$$

of her neighbours choose a_l and the rest choose a_j.

The second measure is the *contagion threshold* $\eta(G)$ of a network, defined as follows. Let $S \subset N$ be a subgroup of agents belonging to the same arbitrarily chosen neighbourhood. That is, for each pair $i, j \in S$, either i and j are directly linked, or there exists a path $i = i_1 \to i_2 \to \cdots \to i_k \to i_{k+1} \to \cdots \to i_K = j$ and/or vice versa, whereby all $i_1, i_2, \cdots, i_k, \cdots, i_K \in S$. Let $\bar{S} = N \backslash S$ be the complement of S. Write $S(a_l)$ to imply that all agents in S choose a_l and write $\bar{S}^{[p,S]}(a_j)$ for agents in $\bar{S}$ choosing a_j for whom at least proportion p of their interactions are with agents in S.

Now, consider a sequence $\{S_\tau(a_l)\}_{\tau \geq 1}$ whereby $S_\tau(a_l) \subset S_{\tau+1}(a_l)$. There exists a complementary sequence $\{\bar{S}_\tau(a_j)\}_{\tau \geq 1}$ whereby $\bar{S}_\tau(a_j) \supset \bar{S}_{\tau+1}(a_j)$. And since $S \cup \bar{S} = N$, it follows that $S_{\tau+1}(a_l) \backslash S_\tau(a_l) = \bar{S}_\tau(a_j) \backslash \bar{S}_{\tau+1}(a_j)$. Given sequences $\{S_\tau(a_l)\}_{\tau \geq 1}$ and $\{\bar{S}_\tau(a_j)\}_{\tau \geq 1}$, there exists a *maximum* p, called the *contagion threshold* of the underlying network, so that for a corresponding sequence $\{\bar{S}_\tau^{[p,S]}(a_j)\}_{\tau \geq 1}$, a_l is a best response for each $i \in \bar{S}_\tau^{[p,S]}(a_j)$. Consequently, for each $\tau \geq 1$, $S_{\tau+1}(a_l) \backslash S_\tau(a_l) = \bar{S}_\tau(a_j) \backslash \bar{S}_{\tau+1}(a_j) = \bar{S}_\tau^{[p,S]}(a_j)$.

The contagion threshold is a unique property of a network. The above definition however also applies to subregions of a network. Morris (2000) provides a characterization of contagion thresholds for various families of deterministic networks (see also Lelarge (2012) for a related characterization for random networks).

Example 2: Consider a network depicted in Figure 4. Consider the cohesive subgroup $S_1 = \{1, 2, 3, 4, 5\}$, which is a cyclic sub-network where each agent has two neighbours to the left and right. This sub-network has a contagion threshold of $\frac{1}{2}$. This is because any choice a_l will spread through best response from configuration $\mathbf{a}_j$ only if a_l is a best to any distribution where at least half of an agent's interactions are with neighbours choosing a_l and the rest choosing a_j.

For the entire network of Figure 4, however, the contagion threshold is $\frac{1}{3}$. Consider the case where contagion of choice a_l starts from say agents 2 and 3 of subgroup S_1. If each agent requires at least proportion $\frac{1}{2}$ to choose a_l for them to do likewise, then contagion stops with agent 7. At this stage, agent 8 has only one neighbour (agent 7) choosing a_l and the rest (agents 9 and 12) choosing a_j, which is less than the half required for any agent to switch to a_l. Choice a_l will thus spread to the entire network only if agent 8 requires only one out of three neighbours (a proportion of $\frac{1}{3}$) to switch to a_l for her to do likewise. A similar argument applies to agent 5 should contagion start from subgroup $S_2 = \{8, 9, 10, 11, 12\}$; hence the contagion threshold is $\frac{1}{3}$.

Given a network G, the contagion threshold $\eta(G)$ and relative payoff gains interactively determine when contagion is feasible. For a 2×2 coordination game, the necessary and sufficient condition for pairwise contagion of a_l relative to a_j is $\phi_{jl} \leq \eta(G)$. The following lemma generalizes this condition for contagion in 2×2 coordination games to $m \times m$ coordination games.

Lemma 2. *For an $m \times m$ coordination game, define the relative payoff gain for a pair of choices a_l and a_j as*

$$\eta_{jl} = \max_{a_k \neq a_l} \left\{ \frac{\delta_{kl}(a_j)}{\delta_{kl}(a_j) + \delta_{lk}(a_l)} \right\}$$

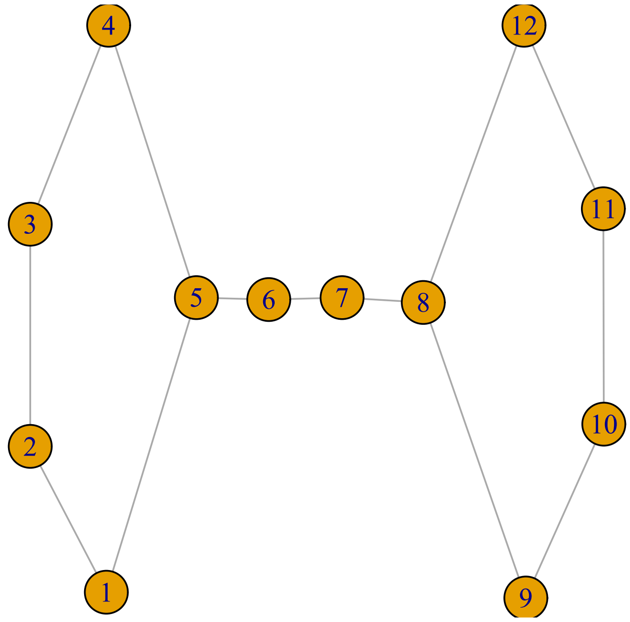


Figure 4: A network with cohesive subgroups.

Then for an arbitrary network G with contagion threshold $\eta(G)$, pairwise contagion of choice a_l relative to a_j is feasible if $\eta_{jl} \leq \eta(G)$ and $\eta_{jl} = \phi_{jl}$.

Proof. See Appendix A.2

Lemma 2 describes a necessary and sufficient condition for pair-wise contagion between any pair of choices. The necessary condition is that the relative payoff-gain η_{jl} must be less or equal to the contagion threshold. It is however sufficient for $\eta_{jl} = \phi_{jl}$. If $\eta_{jl} > \phi_{jl}$, then there exists a distribution in which proportion q of neighbours choose a_l and $1-q$ choose a_j, for some $\phi_{jl} < q < \eta_{jl}$ such that another choice $a_k \notin \{a_l, a_j\}$, is a best response.

The relative payoff gain ϕ_{jl} between options a_j and a_l contains relevant information regarding the relative levels of innovation and cost between the two options. Firstly, by fixing the payoff $u(a_j, a_j)$ of coordinating on choice a_j, the payoff of coordinating on a_l can be rewritten as $u(a_l, a_l) = u(a_j, a_j) + \alpha$, where α measures the level of innovation on product a_l relative to product a_j. Secondly, notice that the payoff $u(a_l, a_j)$ of miscoordination, that is choosing a_l when all neighbours choose a_j, is equivalent to the *net private benefit* of choosing a_l, excluding its social benefit. The net private benefit is equal to intrinsic benefit minus the cost associated with a given choice; that is $u(a_l, a_j) = h(a_l) - c(a_l)$, where $h(a_l)$ is the intrinsic benefit and $c(a_l)$ is the cost of choosing a_l. By fixing $u(a_j, a_l)$, we can rewriting $u(a_l, a_j) = u(a_j, a_l) + \lambda$ so that λ is a measure of extra net private benefit of a_l relative to a_j. A firm or planner can then increase λ by reducing the cost of option a_l. The relative payoff gain ϕ_{jl} can then be rewritten as

$$\phi_{jl} = \frac{[u(a_j, a_j) - u(a_j, a_l)] - \lambda}{2[u(a_j, a_j) - u(a_j, a_l)] - \lambda + \alpha} \tag{6}$$

From (6), we see that given a network G and its associated contagion threshold, a firm has two possible actions through which she can make her product contagious relative to another: by increasing the social benefit through innovation, and/or increasing the private benefit through cost reduction.

A natural extension of pairwise contagion is *path-wise contagion.* An option a_l is path-wise contagious relative to a_j (a_lPTCa_j) if there exists a path $a_j = a_{j_1} \rightarrow a_{j_2} \rightarrow \cdots \rightarrow a_{j_\tau} \rightarrow \cdots \rightarrow a_T = a_l$ along which $a_{j_{\tau+1}}$ is pairwise contagious relative to a_{j_τ}. The following proposition establishes the relationship between contagion and long-run probabilities.

Proposition 1. *For a coordination game diffusion process $(N, A, U^{m\times m}, W)$, let $\beta > \beta^*$. If $a_l PRC a_j$ or $a_l PTC a_j$ and no other $a_k \neq a_j, a_l$ is pairwise contagious relative to a_l, then $\pi_\beta(\mathbf{a}_l) > \pi_\beta(\mathbf{a}_j)$.*

Proof. See Appendix A.3

Proposition 1 shows that the long-run probability of a product or behaviour can be increased relative to another through contagion. By increasing the level of innovation on their product or reducing its cost, a firm can reduce its product's relative payoff gain, making it pairwise contagious relative to another. Consider the following natural extensions of the definitions of relative pairwise and path-wise contagion.

Definition 3. *Given an arbitrary network with contagion threshold $\eta(G)$ and a choice set $A = \{a_1, \cdots, a_j, \cdots, a_m\}$, choice a_l is said to be globally contagious if it is pair-wise contagious relative to all $a_j \neq a_l$.*

Definition 4. *Given an arbitrary network with contagion threshold $\eta(G)$ and a choice set $A = \{a_1, \cdots, a_j, \cdots, a_m\}$, choice a_l is said to be path-wise contagious if there exists a directed path from every choice $a_j \neq a_l$ to a_l along which pair-wise contagion is feasible.*

In relation to Lemma 2, choice a_l is globally contagious if $\eta_{jl} \leq \eta(G)$ and that $\eta_{jl} = \phi_{jl}$ for all $a_j \neq a_l$. It is path-wise contagious if there exists a path $a_j = a_{j_1} \rightarrow \cdots \rightarrow a_{j_\tau} \cdots \rightarrow a_{j_T} = a_l$ from each $a_j \neq a_l$, along which $\eta_{j_\tau j_{\tau+1}} \leq \eta(G)$ and that $\eta_{j_\tau j_{\tau+1}} = \phi_{j_\tau j_{\tau+1}}$. The following corollary is immediate extension of Proposition 1

Corollary 2. *For a strategic diffusion process with m-choices exhibiting coordination properties, let $\beta > \beta^*$. If $\mathbf{a}_l$ is globally or path-wise contagious then $\pi_\beta(\mathbf{a}_l) > \pi_\beta(L_j)$ for all $L_j \neq \mathbf{a}_l$.*

Corollary 2 states that products that are globally or path-wise contagious are chosen with highest probability in the long-run. A firm may find it difficult if not impossible to control path-wise contagion since it requires making other products that are not under her control to be pairwise contagious relative to each other. Global contagion is however under a firm's control, since it requires her product to be pairwise contagious relative to every other. The necessary information a firm needs is the topology of the network structure and relative payoff gains of her product to others; such information can be obtained through surveys.

The results of Proposition 1 and Corollary 2 are derived for monomorphic configuration where all agents coordinate on the same choice. The results therefore apply to situations where a choice say a_l spreads to the entire network through contagion once the set of initial adopters switch to a_l. But as discussed above, contagion can be defined for subregions of the network whenever limits configurations with coexisting choices are present.

Example 3: Consider a 2×2-choice coordination game with payoffs depicted in Table 1. The respective payoff gains are $\eta_{12} = \phi_{12} = \frac{2}{5}$ and $\eta_{21} = \psi_{21} = \frac{3}{5}$. Consider the network depicted in Figure 5, where the contagion threshold for the entire network $\eta(G) = \frac{1}{3}$. Since both $\phi_{12} > \eta(G)$ and $\phi_{21} > \eta(G)$, none of the choices can spread through contagion to the entire network. The infeasibility of contagion for the entire network then implies that there must exist limit configurations in which choices a_1 and a_2 coexist. Indeed, there are two additional limit configurations to $\mathbf{a}_1$ and $\mathbf{a}_2$, $L_1 = (a_1, a_1, a_1, a_1, a_1, a_2, a_2, a_2, a_2, a_2, a_2)$ and $L_2 = (a_2, a_2, a_2, a_2, a_2, a_2, a_1, a_1, a_1, a_1, a_1)$.

If we consider only the sub-network G_1 made up of subgroup $S_1 = \{1, 2, 3, 4, 5\}$, this sub-network has contagion threshold of $\eta(G_1) = \frac{1}{2}$. Choice a_2 can thus spread through contagion in this sub-network; that is a_2 is pairwise contagious relative to a_1 under G_1. Starting from $\mathbf{a}_1$, the diffusion process can thus evolve through contagion to limit configuration L_1. Proposition 1 then states that if $\beta > \beta^*$, then $\pi_\beta(\mathbf{a}_1) < \pi_\beta(L_1) = \pi_\beta(L_2)$. To see how, recall that $\pi_\beta(\mathbf{a}_1) < \pi_\beta(L_1)$ if the cost difference $\Phi_\beta(\mathbf{a}_1, L_1) < 0$, where

$$\Phi_\beta(\mathbf{a}_1, L_1) = -(\beta_m - \beta_m') \sum_{L_j \in \mathbf{L}} \left((r_{\mathbf{a}_1}(L_j) - r_{L_1}(L_j)) \right)$$

For $\beta > \beta^*$, it is then sufficient to show that $r(\mathbf{a}_1) > r(L_1)$ where $r(\mathbf{a}_1) = \sum_{L_j \in \mathbf{L}} r_{\mathbf{a}_1}(L_j)$ and $r(L_1) = \sum_{L_j \in \mathbf{L}} r_{L_1}(L_j)$.

From the relationship between payoffs in Table 1 and Figure 5, we find for the maximum probability $\mathbf{a}_1$-tree that: $r_{\mathbf{a}_1}(\mathbf{a}_2) = 4$, $r_{\mathbf{a}_1}(L_1) = 5$, $r_{\mathbf{a}_1}(L_1) = 5$ and $r_{\mathbf{a}_1}(\mathbf{a}_1) = 0$ since 4 experimentations are required to exit the basin of attraction (BOA) of $\mathbf{a}_2$ to the BOA of L_1, 5 experimentations to exit the BOA of L_1 to the BOA of $\mathbf{a}_1$, and 5 experimentations to exit the BOA of L_2 to the BOA of $\mathbf{a}_1$. In total $r(\mathbf{a}_1) = 14$.

For the maximum probability L_1-tree, we find that: $r_{L_1}(L_2) = 1$, $r_{L_1}(\mathbf{a}_2) = 4$, $r_{L_1}(\mathbf{a}_1) = 1$ and $r_{L_1}(L_1) = 0$ since 1 experimentation is required to exit the BOA of L_2 to the BOA of $\mathbf{a}_2$, 4 experimentations are required to exit the basin of attraction (BOA) of $\mathbf{a}_2$ to the BOA of L_1, and 1 experimentation is required to exit the basin of attraction (BOA) of $\mathbf{a}_1$ to the BOA of L_1. In total, $r(L_1) = 6$. Hence, $r(\mathbf{a}_1) > r(L_1)$ and $\Phi_\beta(\mathbf{a}_1, L_1) < 0$.

In concluding this section, we highlight the following two implications of our results. First, in relation to targeting individuals in the network, (i) firms and governments can exploit decision makers' proneness to mistakes. When agents experiment, they switch to new choices even if they are not the optimal options, which in turn reduces the potential costs of targeting specific agents in the network. Recall that the reason for targeting is to offset the diffusion process from its equilibrium so as to trigger diffusion of a new product or practice. (ii) contagion implies that the number of initial adopters is very small even for a very large population. By strategically making her product contagious, a firm can reduce the cost of targeting and advertising.

Second, the analysis in this section assumed that $\beta > \beta^*$. This assumption is consistent with the

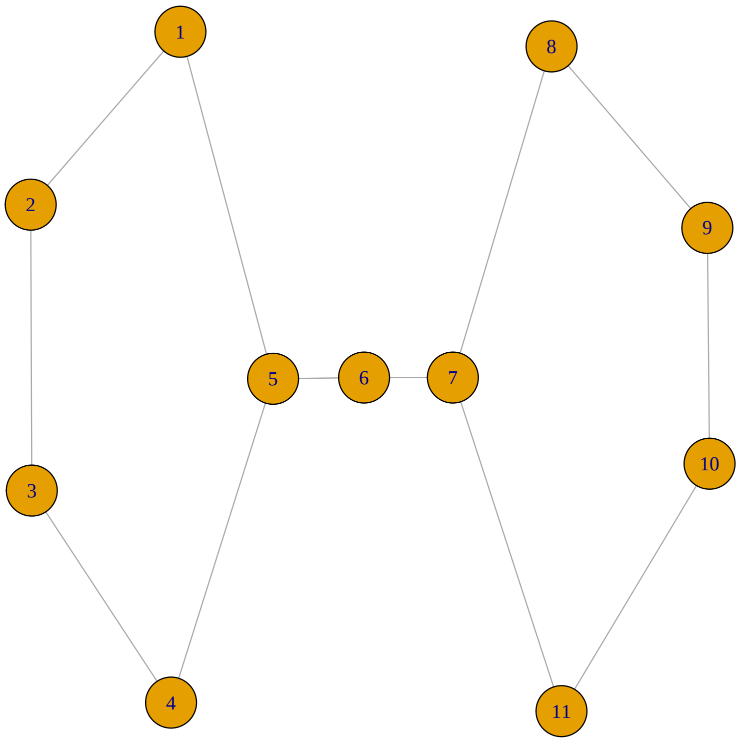


Figure 5: A network with cohesive subgroups.

standard assumptions in evolutionary game theory and evolutionary economics where agents are assumed to be boundedly rational. It is thus not a far-fetched assumption even from a point of view of observable behaviour of consumers and evolution of human behaviour; consumers do not makes choices entirely randomly even if they may make mistakes, neither does human behaviour evolve entirely randomly. Unlike relative payoff gains however, the level of experimentation that agents adopt is not under a firms' or planner's control. Provided that $\beta > \beta^*$, our results above hold. And even more so when the level of experimentation is very small (i.e. β is very large). The problem that arises in situations where β is very large is that the diffusion process becomes very slow. That is, once the process gets trapped in one of the basins of attraction, it takes very long to exit it if agents rarely experiment. It particularly becomes problematic if the expected waiting time to exit a basin of attraction is an increasing function of n, so that in large networks, the diffusion process may never converge to its long-run probabilities. In the next section, we show that if there exists a choice that is globally or pairwise contagious, the expected waiting time to the configuration with the largest long-run probability is independent of n.

5. Contagion and expected waiting time

This section establishes the relationship between the expected waiting time to the configuration with the largest long-run probability, and the property of contagion. We show that if the maximum long-run probability configuration consists of agents choosing a globally or path-wise contagious option, then the expected waiting time to such a configuration is independent of the population size when β is large. The direct implication of this result is that even in large networks and with low levels of experimentation, the diffusion process converges fast to the configurations with the largest long-run probabilities. The problem of slow diffusion does not arise in situations where the level of experimentation is sufficiently high. Kreindler and Young (2013) and Kreindler and Young (2014) indeed show that when β is sufficiently small. Here, we examine the case where β is very large (i.e. $\beta \to \infty$) so that β'_m as defined above, is negligibly small compared to β_m. The expected waiting time is formally defined as follows.

Definition 5. *Let $W \subset \mathbf{X}$ be a subset of the configuration space and $\bar{W}$ its complement. Define $T(W) = \inf\{t \geq 0 \,|\mathbf{x}_t \in W\}$ to be the first time W is reached. The expected waiting time from some configuration $\mathbf{x} \in \bar{W}$ to W is then defined as $\mathbb{E}\left[T(W) \,|\mathbf{x}_0 = \mathbf{x}\right]$.*

Let L_* be the configuration with maximum long-run probability. We aim to show that diffusion is *fast* when L_* is either globally or pairwise contagious. That is, there exists a function $F(\beta)$ that is independent of n so that $\mathbb{E}\left[T(L_*) \,|\mathbf{x}_0 = \mathbf{x}\right] \leq F(\beta)$ for any initial configurations $\mathbf{x}_0$.

Define $r_*(L_j) = \min_{\mathbf{x}\in\mathbf{X}} r_{\mathbf{x}}(L_j)$, to be the minimum number of experimentations required to exit $D(L_j)$, the BOA of L_j, also generally referred to the *radius* of $D(L_j)$. Clear, for $\beta > \beta^*$, if L_* is the limit configuration with the maximum long-run probability, then

$$r(L_*) = \sum_{L_j \in \mathbf{L}} r_{L_*}(L_j) = \sum_{L_j \in \mathbf{L}} r_*(L_j)$$

Let $r_* = \max_{L_j\in\mathbf{L}} r_*(L_j)$ and $\mathbb{E}[T(L_*)] = \max_{\mathbf{x}\in\mathbf{X}} \mathbb{E}[T(L_*) \mid \mathbf{x}_0 = \mathbf{x}]$. The following proposition relates the expected waiting time to the limit configuration with maximum long-run probability $\mathbb{E}[T(L_*)]$, to r_*, the maximum-minimum number of experimentations required to exit a BOA.

Proposition 2. *For a coordination game diffusion process $(N, A, U^{m\times m}, W)$, if L_* is the configuration with the maximum long-run probability, then*

$$\lim_{\beta\to\infty} \frac{\ln \mathbb{E}[T(L_*)]}{\beta} = r_* \tag{7}$$

Proof. See Appendix A.4

Proposition 2 shows that the expected waiting time to reach the configuration with maximum long-run probability from any other configuration takes the form

$$\mathbb{E}[T(L_*)] \leq \exp\left[\beta r_* + f(m, n, \beta)\right] \tag{8}$$

The function $f(m, n, \beta)$ increases with m and n and decreases with β. For $\beta \to \infty$, $f(m, n, \beta) \to 0$. The quantity r_* is a maximum radius of a basin of attraction as defined above. If L_* consists of agents coordinating on a choice that is globally or path-wise contagious, then for each $L_j \in \mathbf{L}$, $r_*(L_j) \ll \frac{1}{2}n$, small and independent of n. This directly implies that $r_* \ll \frac{1}{2}n$, small and independent of n. Hence for β very large, the expected waiting time to the configuration with maximum long-run probability is independent of n if a globally or path-wise contagious choice exists.

Compared to existing results on convergence rates of evolutionary processes such as Ellison (1993), Young (2011), Kreindler and Young (2013) and Kreindler and Young (2014), the result in Proposition 2 is driven more by contagion and less by noise. Kreindler and Young (2014) in particular also find that learning is fast in networks, but they consider a 2×2 coordination game with random sampling and with deterministic dynamics. Moreover, they define fast learning as the case in which noise is large to the extent that only one unique equilibrium exists. On a contrary, Proposition 2 shows that under appropriate conditions learning is also fast in stochastic evolutionary processes of $m \times m$ coordination games. Young (2011) shows that learning is fast in stochastic evolutionary processes with sufficiently large noise provided the network is made up of cohesive subgroups. The underlying reason is that when agents interact in small enough cohesive subgroups, limit configurations induced by the network exist. Such intermediate limit configurations require fewer number of mutations to exit. This argument is observable in Proposition 2 and Example 3. The former states that the expected waiting time is an exponential function of the largest radius of a basin of attraction, while the latter demonstrated how cohesive subgroups reduce the radii of basins of attractions of monomorphic limit configurations. Hence, the results in Young (2011) extends to situations with small noise.

Montanari and Saberi (2010) provide bounds for expected waiting times in a stochastic evolutionary process with sequential dynamics. Just as in Proposition 2 above, they also consider the case of small noise. They however derive bounds for expected waiting times as increasing functions of the population size. This means that as n grows, the expected waiting times become infinitely large and hence learning is slow. Proposition 2 establishes conditions under which learning is fast in situations with small noise.

The results of Proposition 2 are also related to Morris (2000), Sandholm (2001) and Oyama et al. (2015). Since our definitions of global and path-wise contagion are extensions of the notion of pair-wise contagion according to Morris (2000), Proposition 2 then directly implies that learning is fast in the 2×2 symmetric game in Morris (2000). Sandholm (2001) and Oyama et al. (2015) show that $\frac{1}{k}$-dominant choices are almost globally stable in a random interaction model where agents best-respond to a sample of k agents selected from the population. An interaction structure where agents randomly sample k others in the population has a contagion threshold $\eta(G) = \frac{1}{k}$. Lemma 2 and Definition 3 together imply that a $\frac{1}{k}$-dominant is globally contagious in such an interaction structure. Proposition 2 thus implies that learning is fast in the evolutionary game models of Sandholm (2001) and Oyama et al. (2015).

6. Concluding remarks

We have studied the diffusion of products and practices that exhibit network externalities using evolutionary game theory framework. Evolutionary game theory captures many realistic aspects of individual decision processes, and most notably, the tendency to experiment or make mistakes on optimal choices, and myopia – the inability to remember the entire history of play in complex social interactions. We examined how the level of experimentation by decision makers affects diffusion, showing that it induces threshold dynamics on long-run probability of a given choice being adopted.

We demonstrated how firms and governments can harness the power of social networks to strategically diffuse products and practices. In particular, we show how the process of contagion can be exploited to increase the likelihood of a product or practice being adopted. By exploiting agents' tendency to experiment and the notion of contagion, a firm can reduce potential costs associated with strategic targeting or advertising.

Our analysis focused on decision processes where agents make choices through best response, and we defined contagion as spreading through best response. One direction in which our model can be extended is considering situations where agents make decisions through imitation rather than best response. There are cases in evolutionary game theory where imitation and best response lead to different long-run outcomes (see for example Alós-Ferrer and Weidenholzer (2008)). The notion of contagion could then be defined as spreading through imitation.

Appendix A. Appendix

Appendix A.1. Proof of Lemma 1

We use the following results from Freidlin and Wentzell (1984) to prove Lemma 1.

Lemma 3. *(Freidlin and Wentzell, 1984, Lemma 3.1). Given a diffusion process P_β, the stationary distribution $\pi_\beta(\mathbf{x})$ of some configuration $\mathbf{x} \in \mathbf{X}_n$ is given by*

$$\pi_\beta(\mathbf{x}) = \left(\sum_{g \in G(\{\mathbf{x}\})} P_\beta(g) \right) \left(\sum_{\mathbf{y} \in \mathbf{X}_n} \sum_{g \in G(\{\mathbf{y}\})} P_\beta(g) \right)^{-1} \tag{A.1}$$

where the total probability $P_\beta(g)$ associated with each graph g is $P_\beta(g) = \prod_{(\mathbf{z},\mathbf{y}) \in g} P_\beta(\mathbf{z}, \mathbf{y})$.

Given the partition of $g \in G(\{\mathbf{x}\})$ into $g \equiv \bigcup_{L_j \in \mathbf{L}} \{g(D(L_j)) \cup g(\partial L_j, L_j)\}$, each $P_\beta(g)$ can be rewritten as

$$P_\beta(g) = \prod_{L_j \in \mathbf{L}} P_\beta(g(D(L_j))) P_\beta(g(\partial L_j, L_j)) \tag{A.2}$$

where $P_\beta(g(D(L_j))) = \prod_{(\mathbf{y},\mathbf{z}) \in g(D(L_j))} P_\beta(\mathbf{y}, \mathbf{z})$ and $P_\beta(g(\partial L_j, L_j))) = \prod_{(\mathbf{y},\mathbf{z}) \in g(\partial L_j, L_j)} P_\beta(\mathbf{y}, \mathbf{z})$. The following lemma follows from above definitions.

Lemma 4. *Given $g \in G(\{\mathbf{x}\})$, let $r_g(L_j)$ be the number of experimentations involved in the transitions of sub-graph $g(D(L_j)) \in g$ and let $r_g(\partial L_j, L_j)$ be the cardinality of $g(\partial L_l, L_j)$. Then*

$$P_\beta(g) = \exp \left[- \sum_{L_j \in \mathbf{L}} \left(r_g(L_j) \beta_m + r_g(\partial L_j, L_j) \beta'_m \right) \right] \tag{A.3}$$

where $\beta_m = \beta - \ln m^{-1}$ and $\beta'_m = -n \ln \left[1 - e^{-\beta} \left(1 - \frac{1}{m} \right) \right]$.

Proof. The first part of the proof follows from the fact that the probabilities of experimentation are identical for all agents and configurations, that is $\frac{1}{m}\exp(-\beta)$, it follows that

$$P_\beta(g(D(L_j))) = \left(\frac{1}{m}\exp(-\beta)\right)^{r_g(L_j)} = \exp\left(r_g(L_j)\left[\ln m^{-1} - \beta\right]\right) \tag{A.4}$$

For the probability $P_\beta(g(\partial L_j, L_j))$, since each transition in $g(\partial L_j, L_j)$ is dominated by best-response dynamics, then from (4) each transition $\mathbf{x} \to \mathbf{y}$ in $g(\partial L_j, L_j)$ has the probability $[(1-e^{-\beta}) + \frac{1}{m}e^{-\beta}]^n = [1-e^{-\beta}(1-\frac{1}{m})]^n$, where the power of n results from the assumption of simultaneous decision process. That is, y^i for each i is a best response to $\mathbf{x}$ so that $\mathbf{y}$ consists of n simultaneous best responses to $\mathbf{x}$. We have also assumed without loss of generality that $r = 1$. Let $r_g(\partial L_j, L_j)$ be the cardinality of $g(\partial L_j, L_j)$ under graph g, then

$$P_\beta(g(\partial L_j, L_j)) = \left(1 - e^{-\beta}(1-\frac{1}{m})\right)^{nr_g(\partial L_j, L_j)} = \exp\left(nr_g(\partial L_j, L_j)\ln\left[1 - e^{-\beta}\left(1 - \frac{1}{m}\right)\right]\right) \tag{A.5}$$

For notational convenience, let $\beta_m = \beta - \ln m^{-1}$ and $\beta'_m = -n\ln\left[1 - e^{-\beta}\left(1-\frac{1}{m}\right)\right]$. It follows from (A.4) and (A.5) that

$$P_\beta(g) = \exp\left[-\sum_{L_j \in \mathbf{L}}\left(r_g(L_j)\beta_m + r_g(\partial L_j, L_j)\beta'_m\right)\right] \tag{A.6}$$

□

Let $Z(\beta)$ denote the normalization factor for π_β, that is

$$Z(\beta) = \sum_{\mathbf{y}\in\mathbf{X}_n}\sum_{g\in G(\{\mathbf{y}\})} P_\beta(g)$$

The normalization factor $Z(\beta)$ is identical for all configurations $\mathbf{x} \in \mathbf{X}_n$. Let $K = \#G(\{\mathbf{x}\})$ be the cardinality of $G(\{\mathbf{x}\})$. Similarly, K is identical for all $\mathbf{x} \in \mathbf{X}_n$. Given the uniformity of $Z(\beta)$ and K, $\pi_\beta(\mathbf{x})$ is then bounded by

$$Z^{-1}(\beta)\max_{g\in G(\{\mathbf{x}\})} P_\beta(g) \le \pi_\beta(\mathbf{x}) \le Z^{-1}(\beta)K\max_{g\in G(\{\mathbf{x}\})} P_\beta(g) \tag{A.7}$$

Let $g_{\min}(\mathbf{x}) = \operatorname*{argmax}_{g\in G(\{\mathbf{x}\})} P_\beta(g)$ be the $\mathbf{x}$-tree with the maximum probability $P_\beta(g)$. Then

$$P_\beta(g_{\min}(\mathbf{x})) = \max_{g\in G(\{\mathbf{x}\})} P_\beta(g) = \exp\left[-\sum_{L_j\in\mathbf{L}}\left(r_{\mathbf{x}}(L_j)\beta_m + r_{\mathbf{x}}(\partial L_j, L_j)\beta'_m\right)\right] \tag{A.8}$$

It remains to show that along $g_{\min}(\mathbf{x})$, the quantity $r_{\mathbf{x}}(\partial L_j, L_j) = d(L_j) - r_{\mathbf{x}}(L_j)$ for each L_j, where $d(L_j)$ is the cardinality of $D(L_j)$. Note that for any $\beta > 0$, if a_j is a best response to $\mathbf{x}$ then $(1-\exp(-\beta))BR_i(a_j;\mathbf{x}) > 0$, which implies that the probability $(1-\exp(-\beta))BR_i(a_j;\mathbf{x}) + \frac{1}{m}\exp(-\beta)$ of choosing a_j given $\mathbf{x}$ is higher than the probability $\frac{1}{m}\exp(-\beta)$ of choosing any other $a_l \notin BR(\mathbf{x})$ through experimentation. The direct implication is that when $g = g_{\min}(\mathbf{x})$, the subgraph $g(D(L_j))$ for each L_j

consists of $r_{\mathbf{x}}(L_j)$ transitions where only one agent experiments and the rest choose a best response option, so that (see the proof of Lemma 4 above)

$$P_\beta(g(D(L_j))) = \exp\left(r_{\mathbf{x}}(L_j)\left[\ln m^{-1} - \beta\right]\right) \tag{A.9}$$

$$P_\beta(g(\partial L_j, L_j)) = \exp\left(n(d(L_j) - r_{\mathbf{x}}(L_j))\ln\left[1 - e^{-\beta}\left(1 - \frac{1}{m}\right)\right]\right) \tag{A.10}$$

The second alternative for the composition of $g(D(L_j))$ is a single direct transition where $r_{\mathbf{x}}(L_j)$ agents simultaneously experiment to choose a non best response choice, so that $r_{\mathbf{x}}(\partial L_j, L_j) = d(L_j) - 1$. In this case, the total probability $P_\beta(g(D(L_j)))$ of subgraph $g(D(L_j))$ is identical to that in (A.9), but $P_\beta(g(\partial L_j, L_j))$ is less than that in (A.10). When this argument is extended to all $L_j \in \mathbf{L}$, the conclusion is that the likelihood of reaching $\mathbf{x}$ through the second $\mathbf{x}$-tree is smaller than that through $g = g_{\min}(\mathbf{x})$. Any other $\mathbf{x}$-tree different from these two alternatives will imply that the composition $g(D(L_j))$ is so that $r_g(L_j) > r_{\mathbf{x}}(L_j)$ and/or $r_{\mathbf{x}}(\partial L_j, L_j) > d(L_j) - r_{\mathbf{x}}(L_j)$, for each $L_j \in \mathbf{L}$. This in turn implies that $P_\beta(g(D(L_j)))$ and $P_\beta(g(\partial L_j, L_j))$ are greater than those in (A.9) and (A.10) respectively.

It follows from (A.7) that

$$K^{-1}\exp\left[-\sum_{L_j\in\mathbf{L}}\left([r_{\mathbf{x}}(L_j) - r_{\mathbf{y}}(L_j)]\beta_m + [r_{\mathbf{x}}(\partial L_j, L_j) - r_{\mathbf{y}}(\partial L_j, L_j)]\beta_m'\right)\right] \leq \frac{\pi_\beta(\mathbf{x})}{\pi_\beta(\mathbf{y})}$$
$$\leq K\exp\left[-\sum_{L_j\in\mathbf{L}}\left([r_{\mathbf{x}}(L_j) - r_{\mathbf{y}}(L_j)]\beta_m + [r_{\mathbf{x}}(\partial L_j, L_j) - r_{\mathbf{y}}(\partial L_j, L_j)]\beta_m'\right)\right]$$

which can be rewritten as

$$\frac{\pi_\beta(\mathbf{x})}{\pi_\beta(\mathbf{y})} = \exp\left[-\sum_{L_j\in\mathbf{L}}\left([r_{\mathbf{x}}(L_j) - r_{\mathbf{y}}(L_j)]\beta_m + [r_{\mathbf{x}}(\partial L_j, L_j) - r_{\mathbf{y}}(\partial L_j, L_j)]\beta_m'\right) \pm \kappa\right] \tag{A.11}$$

where $\kappa = \ln(K)$. Substituting for $r_{\mathbf{x}}(\partial L_j, L_j) = d(L_j) - r_{\mathbf{x}}(L_j)$ in (A.11) yields

$$\frac{\pi_\beta(\mathbf{x})}{\pi_\beta(\mathbf{y})} = \exp\left[-\sum_{L_j\in\mathbf{L}}\left((r_{\mathbf{x}}(L_j) - r_{\mathbf{y}}(L_j))(\beta_m - \beta_m')\right) \pm \kappa\right] \tag{A.12}$$

Appendix A.2. Proof of Lemma 2

The value of the relative payoff gain η_{jl} follows from considering the best response set of $\mathbf{a}_j$ for distributions that assign probability q to choice a_l, probability $1-q$ to a_j and zero to all other choices, where q is here interpreted as the proportion of an agent's neighbours choosing a_l. It then follows that an agent will play a_l if

$$(1-q)u(a_l, a_j) + qu(a_l, a_l) \geq (1-q)u(a_k, a_j) + qu(a_k, a_l) \quad \text{for all } a_k \neq a_l, \tag{A.13}$$

We then have,

$$q \geq \frac{u(a_k, a_j) - u(a_l, a_j)}{u(a_k, a_j) - u(a_l, a_j) + u(a_l, a_l) - u(a_k, a_l)} = \frac{\delta_{kl}(a_j)}{\delta_{kl}(a_j) + \delta_{lk}(a_l)} \quad \text{for all } a_k \neq a_l$$

where as before $\delta_{kl}(a_j) = u_i(a_k, a_j) - u_i(a_l, a_j)$. Hence an agent will play a_l whenever at least

$$\eta_{jl} = \max_{a_k \neq a_l} \left\{ \frac{\delta_{kl}(a_j)}{\delta_{kl}(a_j) + \delta_{lk}(a_l)} \right\}$$

of his neighbours play a_l and the rest play a_j.

For an arbitrary network with contagion threshold $\eta(G)$, Morris (2000) shows that the necessary condition for pairwise contagion of a_l relative to a_j is for $\eta(G) \geq \eta_{jl}$. It remains to show that, first, any distribution that assigns probability p to any choice $a_k \neq a_l$ and $1-p$ to a_j does not support pair-wise contagion between a_j and a_l. Clearly if choice a_l is a best response to the distribution in which a proportion p of a player's neighbours play a_k and $1-p$ play a_j, then starting from $\mathbf{a}_j$, each player requires at least a fraction p of their neighbours to mutate to a_k for them to switch from a_j to a_l. Such a process is thus driven by mutations and not best response dynamics as required for contagion. Secondly, if $\eta_{jl} \neq \phi_{jl}$, then there exists a distribution in which proportion q of neighbours play a_l and $1-q$ play a_j, for some $\phi_{jl} < q < \eta_{jl}$ such that another choice $a_k \neq a_l$ is a best response. Hence in transiting from $\mathbf{a}_j$ to $\mathbf{a}_l$, the process goes through some states in which choice $a_k \neq a_l$ is played; such transitions are by definition not contagion.

Appendix A.3. Proof of Proposition 1

Recall that $\pi_\beta(\mathbf{a}_l) > \pi_\beta(\mathbf{a}_j)$ if the cost difference $\Phi_\beta(\mathbf{a}_j, \mathbf{a}_l) < 0$, where

$$\Phi_\beta(\mathbf{a}_j, \mathbf{a}_l) = -(\beta_m - \beta'_m) \sum_{L_j \in \mathbf{L}} \Big((r_{\mathbf{a}_j}(L_j) - r_{\mathbf{a}_l}(L_j)) \Big)$$

For $\beta > \beta^*$ it is then sufficient to show that $r(\mathbf{a}_j) > r(\mathbf{a}_l)$ where $r(\mathbf{a}_j) = \sum_{L_j \in \mathbf{L}} r_{\mathbf{a}_j}(L_j)$ and $r(\mathbf{a}_l) = \sum_{L_j \in \mathbf{L}} r_{\mathbf{a}_l}(L_j)$. Rewrite $r(\mathbf{a}_j)$ and $r(\mathbf{a}_l)$ as follows

$$r(\mathbf{a}_j) = \sum_{L_j \neq \mathbf{a}_j, \mathbf{a}_l} r_{\mathbf{a}_j}(L_j) + r_{\mathbf{a}_j}(\mathbf{a}_l) \tag{A.14}$$

$$r(\mathbf{a}_l) = \sum_{L_j \neq \mathbf{a}_j, \mathbf{a}_l} r_{\mathbf{a}_l}(L_j) + r_{\mathbf{a}_l}(\mathbf{a}_j) \tag{A.15}$$

Equations (A.14) and (A.15) follow from the fact that $r_{\mathbf{a}_j}(\mathbf{a}_j) = 0$, and similarly $r_{\mathbf{a}_l}(\mathbf{a}_l) = 0$. If a_l is pairwise contagious relative to a_j, then by definition $\eta_{jl} = \psi_{jl} \leq \eta(G)$. It also implies that $r_{\mathbf{a}_l}(\mathbf{a}_j) = n(\mathbf{a}_j \to \mathbf{a}_j)$, where $n(\mathbf{a}_j \to \mathbf{a}_j)$ is the number of initial adopters (equivalently, the number of experimentations) of a_l required to trigger contagion from $\mathbf{a}_j$ to $\mathbf{a}_l$. By definition of contagion, $n(\mathbf{a}_j \to \mathbf{a}_j) \ll \frac{1}{2}n$, small and independent of n.

If no other choice $a_k \neq a_l$ is pairwise contagious relative to a_l, then $n(\mathbf{a}_j \to \mathbf{a}_j) = r_{\mathbf{a}_l}(\mathbf{a}_j) < r_{\mathbf{a}_j}(\mathbf{a}_l)$. This follows from the fact that if no $a_k \neq a_l$ is PRC to a_l, in the same network where $a_l\text{PRC}a_j$, then $\eta_{lk} > \eta(G) \geq \eta_{jl}$ and hence $\eta_{lk} > \eta_{jl}$. That is, the proportion of neighbours each agent requires to choose any a_k for her to switch from a_l to a_k is greater than the proportion of neighbours each agent requires to choose a_l for her to switch from a_j to a_l. This implies that starting from $\mathbf{a}_l$ the number of experimentations to a_k required to exit the basin of attraction of $\mathbf{a}_l$ is larger than the number of experimentations required to trigger contagion from $\mathbf{a}_j$ to $\mathbf{a}_l$; and hence if $\mathbf{a}_k$ is a subsequent limit configuration to $\mathbf{a}_l$ in the maximum probability $\mathbf{a}_j$-tree, then $r_{\mathbf{a}_l}(\mathbf{a}_j) < r_k(\mathbf{a}_l) \geq r_{\mathbf{a}_j}(\mathbf{a}_l)$.

Moreover, pairwise contagion of a_l relative to a_j also implies that $n(\mathbf{a}_l \to \mathbf{a}_j) = \mathcal{O}(n)$, that is, an increasing function of n. To see why, first notice that if a_lPRCa_j then the only limit configurations associated with a_l and a_j are the monomorphic configurations $\mathbf{a}_l$ and $\mathbf{a}_j$. There is no intermediate limit configuration in which a_j and a_l coexist; that is where a proportion of say q choose option a_j and $1-q$ choice a_l. Secondly, notice that if $\eta_{jl} \leq \eta(G) \leq \frac{1}{2}$, then $\eta(G) \leq \frac{1}{2} < \eta_{lj} \leq 1$. That is, if a_l is a best response to any distribution that assigns probability $p \geq \eta_{jl}$ to a_l and $1-q$ to a_j, then starting from $\mathbf{a}_l$, a_j is a best response to any distribution that assigns probability $q \geq \eta_{lj} > 1 - \eta_{jl}$ to a_j and $1-q$ to a_l. Thus if a_lPCa_j then $\eta_{lj} > \frac{1}{2}$. That is, each agent requires more than half of their neighbours to choose a_j for them to switch from a_l to a_j. This implies that evolution from $\mathbf{a}_l \to \mathbf{a}_j$, without passing through any other limit configuration, is governed by mutations and not best response dynamics, and $n(\mathbf{a}_l \to \mathbf{a}_j; G) = \mathcal{O}(n)$; hence if $\mathbf{a}_j$ is a subsequent limit configuration to $\mathbf{a}_l$ in the maximum probability $\mathbf{a}_j$-tree, then $r_{\mathbf{a}_l}(\mathbf{a}_j) < r_{\mathbf{a}_j}(\mathbf{a}_l)$.

Now, consider the case where $\sum_{L_j \neq \mathbf{a}_j, \mathbf{a}_l} r_{\mathbf{a}_j}(L_j) = \sum_{L_j \neq \mathbf{a}_j, \mathbf{a}_l} r_{\mathbf{a}_l}(L_j)$. This is the situation in which $r(\mathbf{a}_j)$ and $r(\mathbf{a}_l)$ are differentiated by the dynamics in their basins of attractions $D(\mathbf{a}_j)$ and $D(\mathbf{a}_l)$ respectively. Since $r_{\mathbf{a}_l}(\mathbf{a}_j) < r_{\mathbf{a}_j}(\mathbf{a}_l)$, it follows that $r(\mathbf{a}_j) > r(\mathbf{a}_l)$ and hence $\Phi_\beta(\mathbf{a}_j, \mathbf{a}_l) < 0$. Clearly, this situation represents that maximum probability $\mathbf{a}_l$-tree. Any other $\mathbf{a}_l$-tree where $\sum_{L_j \neq \mathbf{a}_j, \mathbf{a}_l} r_{\mathbf{a}_j}(L_j) \neq \sum_{L_j \neq \mathbf{a}_j, \mathbf{a}_l} r_{\mathbf{a}_l}(L_j)$ must have $\sum_{L_j \neq \mathbf{a}_j, \mathbf{a}_l} r_{\mathbf{a}_j}(L_j) < \sum_{L_j \neq \mathbf{a}_j, \mathbf{a}_l} r_{\mathbf{a}_l}(L_j)$.

A similar argument applies to the case of path-wise contagion. Recall that if a_lPTCa_j, then there exists a path $a_j = a_{j_1} \to a_{j_2} \to \cdots \to a_{j_\tau} \to \cdots \to a_T = a_l$ along which $a_{\tau+1}$PRCa_τ. Just as in (A.14) and (A.15), we can rewrite $r(\mathbf{a}_j)$ and $r(\mathbf{a}_l)$ as

$$r(\mathbf{a}_j) = \sum_{L_j \neq \mathbf{a}_{j_1}, \cdots, \mathbf{a}_{j_\tau}, \cdots, \mathbf{a}_T} r_{\mathbf{a}_j}(L_j) + \sum_{\tau=2}^{T-1} r_{\mathbf{a}_j}(\mathbf{a}_{j_\tau}) + r_{\mathbf{a}_j}(\mathbf{a}_l) \tag{A.16}$$

$$r(\mathbf{a}_l) = \sum_{L_j \neq \mathbf{a}_{j_1}, \cdots, \mathbf{a}_{j_\tau}, \cdots, \mathbf{a}_T} r_{\mathbf{a}_l}(L_j) + \sum_{\tau=2}^{T-1} r_{\mathbf{a}_l}(\mathbf{a}_{j_\tau}) + r_{\mathbf{a}_l}(\mathbf{a}_j) \tag{A.17}$$

Pairwise contagion implies that $r_{\mathbf{a}_l}(\mathbf{a}_{j_\tau}) \ll \frac{1}{n}$, small and independent of n for $\tau = 2, \cdots, T-1$. Hence $\sum_{\tau=2}^{T-1} r_{\mathbf{a}_l}(\mathbf{a}_{j_\tau}) \leq \sum_{\tau=2}^{T-1} r_{\mathbf{a}_j}(\mathbf{a}_{j_\tau})$. Note however that even if $\sum_{\tau=2}^{T-1} r_{\mathbf{a}_l}(\mathbf{a}_{j_\tau}) = \sum_{\tau=2}^{T-1} r_{\mathbf{a}_j}(\mathbf{a}_{j_\tau})$, we showed above that $r_{\mathbf{a}_l}(\mathbf{a}_j) < r_{\mathbf{a}_j}(\mathbf{a}_l)$, which implies that $r(\mathbf{a}_j) > r(\mathbf{a}_l)$ and hence $\Phi_\beta(\mathbf{a}_j, \mathbf{a}_l) < 0$.

Appendix A.4. Proof of Proposition 2

The following definitions are used in the next steps of the proof. Given transition probabilities $P_\beta(\mathbf{x}, \mathbf{y})$, define a cost function $C(\mathbf{x}, \mathbf{y})$ as follows

$$\lim_{\beta \to \infty} \frac{-\ln P_\beta(\mathbf{x}, \mathbf{y})}{\beta} = C(\mathbf{x}, \mathbf{y}) \tag{A.18}$$

The cost function $C(\mathbf{x}, \mathbf{y})$ is equal to the number of experimentations that simultaneously occur in the transition $\mathbf{x} \to \mathbf{y}$. To see why, first notice that if a_j is not in the best response set of $\mathbf{x}$ so that $BR_i(a_j; \mathbf{x}) = 0$, then

$$\lim_{\beta\to\infty}\frac{-\ln \mathbb{P}_i(a_j;\mathbf{x}_t)}{\beta}=\lim_{\beta\to\infty}\frac{-\ln\left[\frac{1}{m}\exp(-\beta)+(1-\exp(-\beta))BR_i(a_j;\mathbf{x})\right]}{\beta}=1 \tag{A.19}$$

So if transition $\mathbf{x}\to\mathbf{y}$ involves r simultaneous mutations then

$$\lim_{\beta\to\infty}\frac{-\ln P_\beta(\mathbf{x},\mathbf{y})}{\beta}=C(\mathbf{x},\mathbf{y})=r \tag{A.20}$$

The next definition is related to $G(W)$ graphs in Definition 1 in Section 3.2.

Definition 6. *For any $\mathbf{x}\in\bar{W}$ and $\mathbf{y}\in W$ where $\mathbf{x}\neq\mathbf{y}$, $G_{\mathbf{x},\mathbf{y}}(W)$ is a set of all $G(W)$-graphs which link $\mathbf{x}$ to $\mathbf{y}$. For any two configurations $\mathbf{x},\mathbf{y}\in\bar{W}$, $G_{\mathbf{x},\mathbf{y}}(W\cup\{\mathbf{y}\})$ is the set of $G(W)$-graphs in which $\mathbf{x}$ is joined to some point $\mathbf{y}$ possibly itself and not to W, and that all other points of $\bar{W}$ are joined to either the same point or to W.*

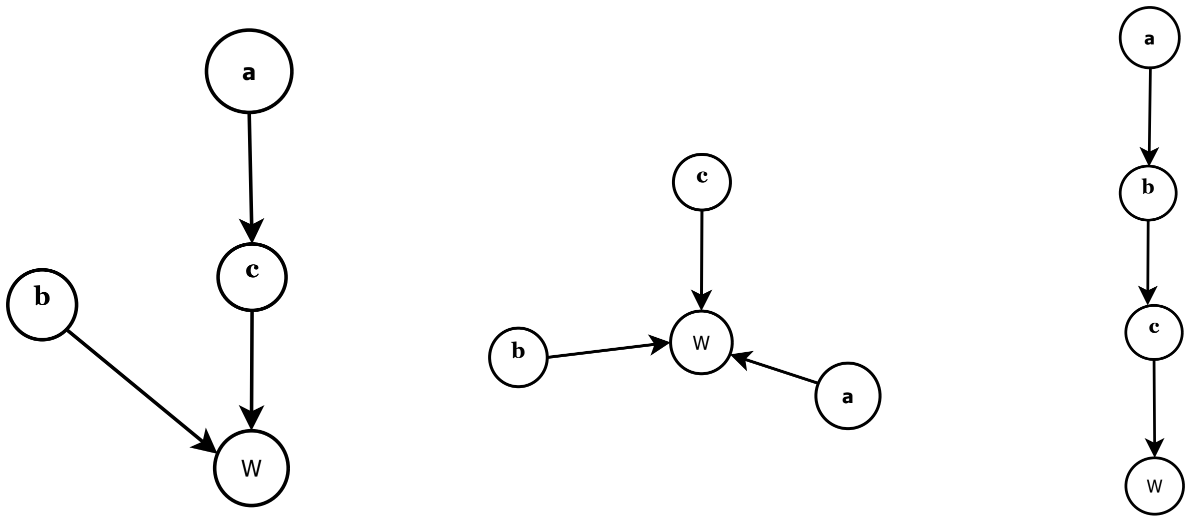


Figure A.6: Examples of $G(W)$ graphs, where $\bar{W}=\{\mathbf{a},\mathbf{b},\mathbf{c}\}$.

Consider a configuration space $\mathbf{X}=\{\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d},\mathbf{e},\mathbf{f},\mathbf{g},\mathbf{h}\}$ with examples of $\mathbf{g}$-trees depicted in Figure 2. Let $W=\{\mathbf{d},\mathbf{e},\mathbf{f},\mathbf{g},\mathbf{h}\}$, with examples of $G(W)$ graphs depicted in Figure A.6. Then examples of $G_{\mathbf{a},\mathbf{c}}(W\cup\{\mathbf{c}\})$ graphs based on $G(W)$ graphs in Figure A.6 are: $\{\mathbf{a}\to\mathbf{c},\mathbf{b}\to W\}$ for the graph on the left, $\{\mathbf{c}\to W,\mathbf{b}\to W\}$ for the middle graph, and $\{\mathbf{a}\to\mathbf{b},\mathbf{c}\to W\}$ for the graph on the right.

Let $C(g)=\sum_{(\mathbf{x},\mathbf{y})\in g}C(\mathbf{x},\mathbf{y})$. The following result is derived in Catoni (1999, Proposition 4.2).

Lemma 5. *For any $W\subset\mathbf{X}$, $W\neq\emptyset$ and $\bar{W}=\mathbf{X}\backslash W$, for any $\mathbf{x},\mathbf{y}\in\bar{W}$*

$$\lim_{\beta\to\infty}\frac{\ln\mathbb{E}[T(W)|\,\mathbf{x}_0=\mathbf{x}]}{\beta}=\min_{g\in G(W)}C(g)-\min_{\mathbf{y}\in\bar{W}}\min_{g\in G_{\mathbf{x},\mathbf{y}}(W\cup\{\mathbf{y}\})}C(g) \tag{A.21}$$

Since the focus is on situations where β is very large, it suffices to consider the reduced L_j-trees describing evolution between limit configurations through experimentation, rather than $\mathbf{x}$-trees that span the entire configuration space. The reduced L_j-trees consist of directed links between limit configurations that represent the number of experimentations required to evolve from one basin of attraction (BOA) to another. That is, for each L_k, the quantity $r_{L_k}(L_j)$ for each $L_j \in \mathbf{L}$, is the number of experimentations required to exit the BOA of L_j into the BOA of another limit configuration L_l that is subsequent to L_j in the L_k-tree.

Recall the diffusion process discussed in Example 3 above, where $\mathbf{L} = \{\mathbf{a}_1, L_1, L_2, \mathbf{a}_2\}$. Figures A.7 and A.8 depict examples of reduced $\mathbf{a}_2$-trees and L_1-trees respectively. The labels on the links indicate the number of experimentations required to exit the BOA of one limit configuration into the BOA of another. For example, in the $\mathbf{a}_2$-tree on the right hand side of Figure A.7, 5 experimentations are required to exit the BOA of L_2 into the BOA of $\mathbf{a}_1$, 2 experimentations to exit the BOA of $\mathbf{a}_1$ into the BOA of $\mathbf{a}_2$, and 1 experimentation to exit the BOA of L_2 into the BOA of $\mathbf{a}_2$.

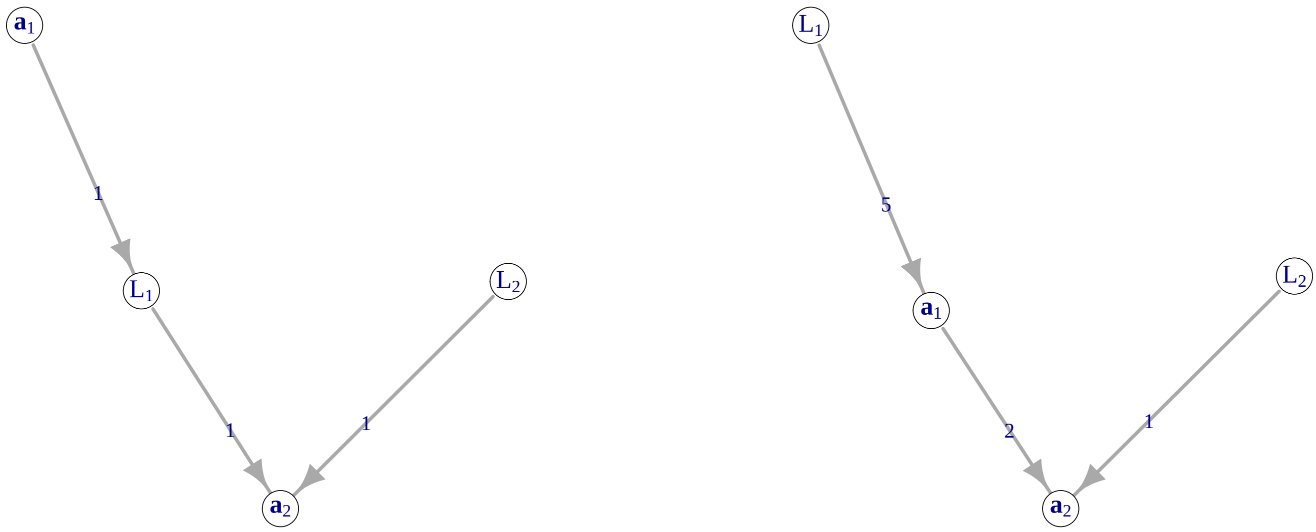


Figure A.7: Examples of reduced $\mathbf{a}_2$-trees.

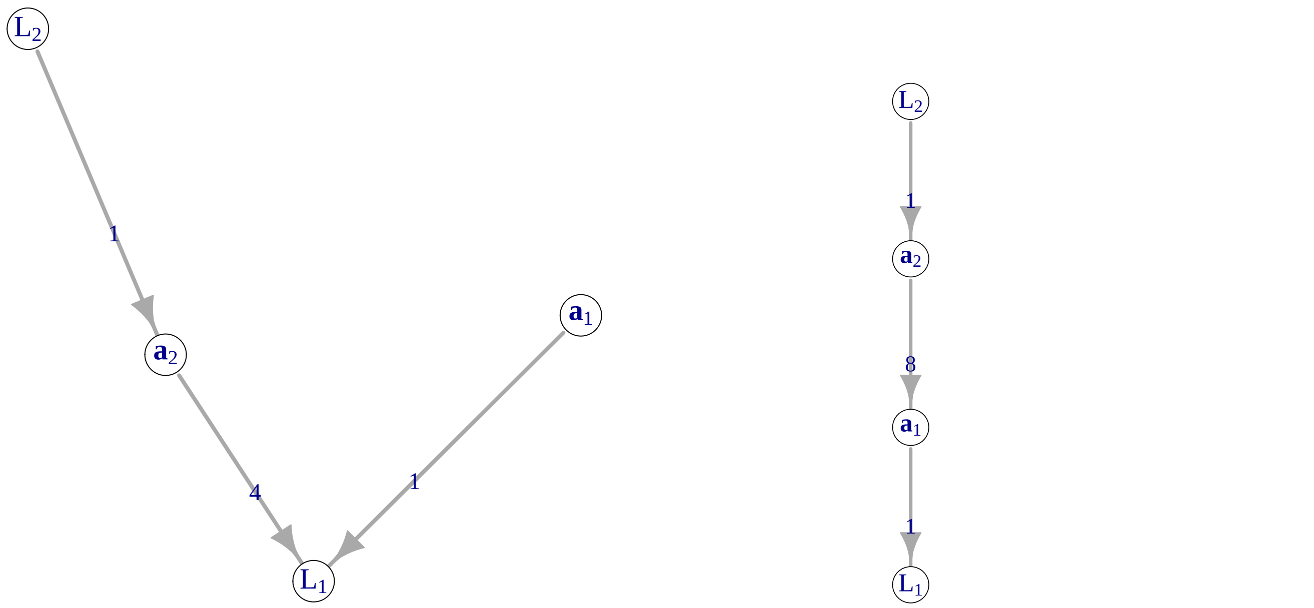


Figure A.8: Examples of reduced L_1-trees.

Define $r_*(L_j) = \min_{\mathbf{x}\in\mathbf{X}} r_{\mathbf{x}}(L_j)$, to be the minimum number of experimentations required to exit $D(L_j)$, the BOA of L_j, also generally referred to the *radius* of $D(L_j)$. Clear, for $\beta > \beta^*$, if L_* is the limit configuration with the maximum long-run probability, then

$$r(L_*) = \sum_{L_j \in \mathbf{L}} r_{L_*}(L_j) = \sum_{L_j \in \mathbf{L}} r_*(L_j)$$

Let $W = \{L_*\}$; then $\min_{g \in G(\{L_*\})} C(g) = r(L_*)$, and $\min_{L_k \in \bar{W}} \min_{g \in G_{L_j,L_k}(L_* \cup \{L_k\})} C(g) = r(L_*) - \max_{L_j \in \mathbf{L}} r_{\mathbf{x}}(L_j)$ for any $L_l \in \bar{W}$. Let $r_* = \max_{L_j \in \mathbf{L}} r_*(L_j)$ and $\mathbb{E}[T(L_*)] = \max_{\mathbf{x} \in \mathbf{X}} \mathbb{E}[T(L_*) \mid \mathbf{x}_0 = \mathbf{x}]$. Then

$$\lim_{\beta \to \infty} \frac{\ln \mathbb{E}[T(L_*)]}{\beta} = r_* \tag{A.22}$$

Acknowledgements

We are thankful for the constructive comments from Alan Kirman, Robin Cowan, Vianney Dequiedt, Jean-Jacques Herings, Arkadi Predtetchinski, Co-Pierre Georg, François Lafond, Giorgio Triulzi and Stefania Innocenti. We also gratefully acknowledge the fruitful discussions with seminar participants at The Centre d'Economie de La Sorbonne and Paris School of Economics, University of Cergy Pontoise THEMA, Maastricht Lecture Series in Economics and Cournot seminars at Bureau d'Economie Théorique et Appliquée (BETA), University of Strasbourg. This work was supported by the FERDI (Fondation pour les Etudes et Recherches sur le Développement International) and the Agence Nationale de la Recherche of the French government through the program "Investissements d'avenir ANR-10-LABX-14-01." The usual disclaimer applies.

References

References

Alós-Ferrer, C., Weidenholzer, S., 2008. Contagion and efficiency. Journal of Economic Theory 143 (1), 251–274.

Bass, F. M., 1969. A new product growth for model consumer durables. Management science 15 (5), 215–227.

Catoni, O., 1999. Simulated annealing algorithms and markov chains with rare transitions. In: Azéma, J., Émery, M., Ledoux, M., Yor, M. (Eds.), Séminaire de probabilités XXXIII. Springer, Berlin Heidelberg, pp. 69–119.

Chen, W., Wang, Y., Yang, S., 2009. Efficient influence maximization in social networks. In: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, pp. 199–208.

Coleman, J., Katz, E., Menzel, H., of Applied Social Research, C. U. B., 1966. Medical innovation: a diffusion study. Advanced Study in Sociology. Bobbs-Merrill Co.
URL https://books.google.co.za/books?id=4e1EAAAAIAAJ

Conley, T. G., Udry, C. R., March 2010. Learning about a new technology: Pineapple in ghana. American Economic Review 100 (1), 35–69.
URL http://www.aeaweb.org/articles?id=10.1257/aer.100.1.35

Ellison, G., 1993. Learning, local interaction, and coordination. Econometrica 61 (5), 1047–1071.

Ellison, G., 2000. Basins of attraction, long-run stochastic stability, and the speed of step-by-step evolution. Review of Economic Studies 67 (1), 17–45.

Foster, A. D., Rosenzweig, M. R., 1995. Learning by doing and learning from others: Human capital and technical change in agriculture. Journal of political Economy 103 (6), 1176–1209.

Freidlin, M., Wentzell, A. D., 1984. Random perturbations of dynamical systems. New York: Springer Verlag.

Galeotti, A., Goyal, S., 2009. Influencing the influencers: a theory of strategic diffusion. The RAND Journal of Economics 40 (3), 509–532.

Galeotti, A., Goyal, S., Jackson, M. O., Vega-Redondo, F., Yariv, L., 2010. Network games. Review of Economic Studies 77 (1), 218–244.

Godes, D., Mayzlin, D., 2004. Using online conversations to study word-of-mouth communication. Marketing science 23 (4), 545–560.

Jackson, M. O., Yariv, L., 2007. Diffusion of behavior and equilibrium properties in network games. The American economic review 97 (2), 92–98.

Kelly, J. A., St Lawrence, J. S., Diaz, Y. E., Stevenson, L. Y., Hauth, A. C., Brasfield, T. L., Kalichman, S. C., Smith, J. E., Andrew, M. E., 1991. Hiv risk behavior reduction following intervention with key opinion leaders of population: an experimental analysis. American journal of public health 81 (2), 168–171.

Kempe, D., Kleinberg, J., Tardos, É., 2003. Maximizing the spread of influence through a social network. In: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, pp. 137–146.

Kirby, J., Marsden, P., 2006. Connected marketing: the viral, buzz and word of mouth revolution. Elsevier.

Kiss, C., Bichler, M., 2008. Identification of influencers—measuring influence in customer networks. Decision Support Systems 46 (1), 233–253.

Kreindler, G. E., Young, H. P., 2013. Fast convergence in evolutionary equilibrium selection. Games and Economic Behavior 80, 39 – 67.

Kreindler, G. E., Young, H. P., 2014. Rapid innovation diffusion in social networks. Proceedings of the National Academy of Sciences 111 (Supplement 3), 10881–10888.

Lelarge, M., 2012. Diffusion and cascading behavior in random networks. Games and Economic Behavior 75 (2), 752–775.

López-Pintado, D., 2006. Contagion and coordination in random networks. International Journal of Game Theory 34 (3), 371–381.

López-Pintado, D., 2008. Diffusion in complex social networks. Games and Economic Behavior 62 (2), 573–590.

Mahajan, V., Müller, E., Bass, F. M., 1990. New product diffusion models in marketing: A review and directions for research. Journal of Marketing 54, 1–26.

Montanari, A., Saberi, A., 2010. The spread of innovations in social networks. Proceedings of the National Academy of Sciences.

Morris, S., 2000. Contagion. Review of Economic Studies 67 (1), 57–78.

Oyama, D., Sandholm, W. H., Tercieux, O., 2015. Sampling best response dynamics and deterministic equilibrium selection. Theoretical Economics 10 (1), 243–281.

Reingen, P. H., Foster, B. L., Brown, J. J., Seidman, S. B., 1984. Brand congruence in interpersonal relations: A social network analysis. Journal of Consumer Research 11 (3), 771–783.

Richardson, M., Domingos, P., 2002. Mining knowledge-sharing sites for viral marketing. In: Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, pp. 61–70.

Rogers, E. M., 2003. Diffusion of innovations. 5th Edition. Free Press.

Sandholm, W. H., 2001. Almost global convergence to p-dominant equilibrium. International Journal of Game Theory, 107–116.

Sundararajan, A., 2007. Local network effects and complex network structure. The BE Journal of Theoretical Economics 7 (1).

Valente, T. W., Hoffman, B. R., Ritt-Olson, A., Lichtman, K., Johnson, C. A., 2003. Effects of a social-network method for group assignment strategies on peer-led tobacco prevention programs in schools. American journal of public health 93 (11), 1837–1843.

Young, P. H., 2011. The dynamics of social innovation. Proceedings of the National Academy of Sciences 108 (4), 21285–21291.