Convergence rates for stochastic evolution in networks

Théophile T. Azomahou

Maastricht University—UNU-MERIT, 6211 TC Maastricht, The Netherlands

Daniel C. Opolot*

Maastricht University—UNU-MERIT, 6211 TC Maastricht, The Netherlands

This Version—September 2014

Abstract: Stochastic evolutionary models are credited for their ability to derive sharp predictions about equilibrium selection when multiplicity exists. A major criticisms of these models concerns the convergence rates to long-run stable sets and to stationarity. In particular, there are cases in which the expected waiting times increase exponentially with the population size. Under local interactions however, the expected waiting times are shown to be finitely bounded even when the population size is arbitrarily large. This paper provides a detailed characterization of convergence rates in networks using two measures; *expected waiting time* and *convergence time* as measures for the transition times between subsets of states and to stationarity or selection. We provide tighter bounds for each for positive rather than limit noise levels. Three factors are key in influencing the convergence rates, the payoff gains, network topology and noise level. The network topology and payoff gains interactively determine whether or not a strategy is *globally contagious* and the convergence rates are a function of this property. When global contagion is feasible, the expected waiting time from any other subset of states to the long-run stable set is shorter for highly than sparsely connected networks. The reverse is true for convergence time. Networks for which global contagion is infeasible lead to shorter convergence time.

Keywords: Stochastic evolution, networks, expected waiting time, convergence time.

JEL classification: C73, D80.

1. Introduction

Evolutionary game theory provides a framework for modeling repeated interactions and experimentation among economic agents over time. The basic idea of evolutionary game dynamics is that individuals are repeatedly matched to play a given game, and over time each learns their opponents' play and subsequently adjust their behavior. The adjustment process is such that players are assumed to best respond myopically to opponents' strategies, but their assessments are subject to random shocks. The major contributions of evolutionary game theory to game theory are

*Corresponding author: Opolot D. C., Maastricht University—UNU-MERIT, Keizer Karelplein 19, 6211 TC Maastricht, The Netherlands, (opolot@merit.unu.edu), Tel. +31 433884440, Fax +31 433884499.

equilibrium prediction or selection (by identifying long-run or stochastically stable outcomes), and explaining how Nash Equilibrium comes about through simple adaptive behavioral rules such as bounded-rationality and myopia.

One of the major discussions in stochastic evolutionary literature concerns the rate at which equilibrium selection occurs. The computation of stochastically stable sets involves taking the limits of noise. But as pointed out by Ellison (1993) and Sandholm (2010), assuming vanishing noise levels also implies that the convergence time to the long-run stationary distribution (hence selection) becomes exceptionally long. More specifically, once the process gets trapped in one of the quasi-stable states (possibly a socially undesirable one), the waiting time to exit its *basin of attraction* becomes arbitrarily large for vanishing noise. There are however cases in which the waiting times are shown to be finite and even sometimes to be independent of the population size. This is particularly the case when agents are matched to play with only a subset of others in the population. Robson and Vega-Redondo (1996) shows that the convergence rates are high when players are matched in a round robin tournament mechanism, where each player confronts each other player exactly once. Ellison (1993), Sandholm (2001), Montanari and Saberi (2010) show that under local interaction (networks), equilibrium selection is fast. Recent work by Kreindler and Young (2013) demonstrates that even under global interactions, it is possible to derive conditions on the payoff for the given noise levels and vise versa for which selection is fast (occurs in a finite time).

The present paper aims to provide a detailed analysis of convergence rates for general stochastic evolutionary game dynamics in networks. The two convergence rate measures that we characterize explicitly are the *expected waiting time* and *convergence or selection time*. The expected waiting time measures how long it takes the learning process to exit a given domain or subset of states, and in particular we will be talking about exiting a given (or subset of) basin of attraction. It is therefore a measure of medium-run dynamics of the process. The convergence time measures how fast the process attains its long-run invariant distribution, or equivalently the time scales for which equilibrium selections results should be considered relevant.

Our characterization of expected waiting times utilizes methods for the concept of *hitting time*. Hitting time is a classical concept in the study of finite Markov chains and processes. The decay rates for hitting times in the case of Markov chains with rare transitions resulting from small perturbations (as is the case in this paper) is characterized in detail by Freidlin and Wentzell (1984) and Catoni (1999) using techniques from large deviations theory. Beggs (2005) shows how the results and concepts developed in Freidlin and Wentzell (1984) and Catoni (1999) can be used to characterize expected waiting times and equilibrium selection in stochastic evolutionary models. Beggs (2005) analysis improves on the earlier results by Ellison (1993) on expected waiting times. Though these papers provide good insights on the nature of hitting times in stochastic evolutionary processes, their analysis focuses on generic characterizations. Here, we focus on expected waiting times for stochastic evolution in networks. To be able to fully characterize the effects of various network topologies, we adopt a simpler model of mistakes, in which the probability of making a mistake $\mathscr{P} = \frac{1}{m}\exp(-\beta)$, identical for all players, strategies and states. The parameter m is the size of the action set and β is the parameter of noise, such that the larger β the smaller the probability

of mistakes. For β equal to infinity, we have the case of best-response dynamics.

The characterization of convergence time that we provide utilizes methods from the concept of mixing time. Mixing time is also a traditional concept in finite Markov chains and processes. It is a measure of the time it takes the chain to attain its invariant distribution. Since long-run or stochastically stable sets solely depend on the properties of the invariant distribution, mixing time can thus be employed to characterize the time to selection. The structural bounds for mixing time of finite Markov chains have been established in the literature (e.g Diaconis and Stroock (1991) and Sinclair (1992)). Its lower and upper bounds are a function of the second largest eigenvalue of the associated transition matrix. Diaconis and Stroock (1991) and Sinclair (1992) have established geometric bounds for the second eigenvalue of the Markov chain transition matrices. These measure is however computationally demanding for large state space chains. They are also less intuitive in the case of stochastic evolution in which the interest is to understand how the underlying game, learning rules and network topology affect the convergence time. Here, we provide bounds for convergence time in terms of the readily computable and fairly understood concept in the literature of stochastic evolution; the maximum expected waiting time. The properties we derive for expected waiting time therefore extend directly to the characterization of convergence time.

By explicitly characterizing bounds for convergence rates, we are able to derive deeper insights into the role of topologies in stochastic evolution in networks. We establish convergence rates results for families of deterministic and random networks. We find the key factor of the evolutionary process to be whether or not a strategy is *globally contagious*. For a given network topology, an action is said to be globally contagious if its payoff gain relative to all other actions is such that it requires only a small fraction of initial adopters to trigger a global cascade to the whole population. Globally contagious actions are then those with maximum payoff gains and thus form the long-run stable states. In 2×2 matrix games for example, if global contagion is feasible then the risk-dominant action is the globally contagious one. For given payoff gains, the more sparsely connected a network is the more likely that an action with the highest pain off gain is globally contagious. For a given network topology and payoff gains, if global contagion is not feasible then we have the case of step–by–step evolution.

When global contagion is feasible, then for various families of networks the expected waiting times from any other closed communication class to the long-run stable closed communication class are independent of the population size. This result is not particularly surprising as similar findings were obtained by Ellison (1993) for k-nearest neighbor interactions. The main difference with Ellison (1993) however is that we provide tighter bounds for expected waiting times independently of the assumption of vanishing mutation rates. Secondly, we establish this result for a wider range of networks beyond the k-nearest neighbor interactions. More generally, for networks of the same family say n-dimensions nearest neighbor interactions or n-dimensions l-max distance interactions, whenever global contagion is feasible the expected waiting time for highly connected network is shorter than for sparsely connected networks. This result is not true otherwise. It is also infeasible to derive such conclusions from analysis in the derivation of expected waiting times heavily relies on the assumption of vanishing noise.

In the case of convergence times, whenever global contagion is feasible, sparsely connected

networks tend to have longer convergence time than highly connected networks. In fact, under these conditions the convergence time for global interactions is shorter than for local interactions. This is not necessarily true whenever global contagion is not feasible. The underlying reason for this finding is that convergence time is an increasing function of maximum expected waiting time. The maximum expected waiting time corresponds to the expected waiting time from the long-run stable communication class to the least stable communication class. The maximum expected waiting time for local interactions in which global contagion is feasible is generally greater than for global interactions.

Whenever global contagion is not feasible, the evolutionary process exhibits a relatively different behavior hence convergence rates as compared to the case in which it is feasible. Infeasibility of global contagion implies that there exists intermediate close communication classes induced by the network topology. These intermediate closed communication classes correspond to states in which at least two *cohesive* or *close-knit* subgroups (definitions according to Morris (2000) and Young (2011)) of players play different actions. Morris (2000); Lee, Szeidl, and Valentinyi (2003); Young (2011) and Montanari and Saberi (2010) show that the expected waiting times to the long-run stable closed communication class is finite. An evolutionary process with such intermediate closed communication classes also relates to the notion of step-by-step evolution according to Ellison (2000), in which it is demonstrated that under step-by-step evolution the selection process is relatively fast. Here, we formalize these notions for evolutionary processes in networks using bounds we derive for expected waiting times and convergence times. Indeed, we find that the convergence time (time to selection) is shorter when the network topology and payoff gains collectively induce step-by-step evolution than even the case of global interactions.

Finally, this paper is also related to others in the literature that characterize evolutionary processes in random networks using mean field best-response dynamics (e.g. Sandholm (2001); Jackson and Yariv (2007); Alós-Ferrer and Weidenholzer (2008); Lelarge (2012) and Kreindler and Young (2013, 2014)). Most of these literature, with the exception of Kreindler and Young (2013, 2014), focuses on long-run stability rather than convergence rates as we do here. Kreindler and Young (2013, 2014) derive upper bounds for expected waiting times to long-run stable states and show that it is independent of the population size as long as the payoff gain is sufficiently large and the noise level is not too low. Though the analysis we present derives some results that are similar qualitatively, there are several differences in both methodology and quantitative results. Kreindler and Young (2013, 2014) focus on sequential dynamics rather than simultaneous as we do, which implies differences in analytical methods and in turn quantitative results. For example, we find the the structure of the network to be relevant in determining convergence rates, unlike in Kreindler and Young (2013, 2014) in which it does not. Moreover, our analysis extends to $m \times m$ matrix games and includes convergence times.

The remainder of the paper is organized as follows. In section 2 we introduce the general framework of stochastic evolutionary dynamics in networks. Section 3 introduces expected waiting times, providing its general treatment and computation. Similarly, Section 4 introduces convergence time. Sections 5 and 6 focus on the characterizations of convergence rates for families of deterministic and random networks when global contagion is feasible. Section 7 provides a characterization of the

case in which global contagion is not feasible.

2. The model

We consider a general class of $N-$person $m \times m$-matrix repeated games. A set of agents $\mathcal{N} = \{1, \cdots, i, \cdots, N\}$ are matched repeatedly to play a normal form game Γ. The matching rule determines whether agents interact with the entire population (global) or a subset of others (social network). This in turn also determines the information set of players, whether they observe the entire distribution of strategies in the population or in their local neighborhood. In the case in which agents are matched with a subset of others, the network of interactions is modeled in graph theoretic manner. Let $G(\mathcal{N}, E)$ be a graph with $\mathcal{N}$ vertices representing the number of agents and E edges linking different pairs of agents, such that a graph g_{ij} defines the connection between i and j. If $g_{ij} = 1$ then a directed link exists from i to j, and zero implies otherwise. $G(\mathcal{N}, E)$ is thus a directed network describing the relationship of any one agent with every other agent in the population. The *adjacency matrix* $\mathscr{A}$ of $G(\mathcal{N}, E)$ is defined as an $N \times N$ matrix with entries being the elements of g_{ij}. The *neighborhood* of agent i, $\mathcal{N}_i$, is defined as $\mathcal{N}_i = \{j \in \mathcal{N} | g_{ij} = 1\}$, and gives the set of players to which i is linked to. The cardinality $\#\mathcal{N}_i = k_i$, is the *degree* of i.

The strategy set $X = a, b, c, \cdots$ is assumed to be identical for all agents and has cardinality m. We denote the action chosen by each $i \in \mathcal{N}$ at period t by $x_{i,t}$ and $\mathbf{x}_t = (x_{1,t}, \cdots, x_{n,t})$ is the strategy profile at t. Each $\mathbf{x}_t \in \mathbf{X} = X^N$ will also be referred to as the *population state* or simply the *state* of the learning process at t.

For each game Γ, let $U \in \mathbb{R}^{m \times m}$ be the associated payoff matrix such that $U(a, b)$ is the base payoff to an agent playing a when his opponent plays b. A typical example is the pure coordination game in Table 1. Given the information set of agent i, that is the observed distribution of strategies $\mathbf{x}$ either of the entire population or of i's neighborhood, the expected payoff to i from playing action a when his opponents profile is $\mathbf{x}_{-i}$ is then

$$U_i(a, \mathbf{x}_{-i}) = \sum_{j \in \mathcal{N}_i} J_{ij} U(a, x^j), \tag{1}$$

where x^j is the j^{th} coordinate of $\mathbf{x}$. The parameter J_{ij} takes on values in the closed interval $[0, 1]$. For example if agents consider average payoff over their neighbors' strategies, then $J_{ij} = \frac{1}{k_i}$ for all $j \in \mathcal{N}_i$.

		player j	
		a	b
player i	a	$U(a,a)$	$U(a,b)$
	b	$U(b,a)$	$U(b,b)$

Table 1: Payoff structure for the symmetric pure coordination game. The payoffs are for the row-player.

2.1. *Dynamics and learning scheme*

There are two main components that completely describe the learning process; the *revision protocol* and *choice probabilities*. Here, we consider the simultaneous revision protocol. When a player receives a revision opportunity, he evaluates the current expected payoff to each of the available pure strategies. A player's information set is the distribution of strategies of a subset of other players in the neighborhood. Each player's objective is to select strategies that he evaluates as best, this assessment is however subject to random shocks. We adopt a relatively simpler model of mistakes (hence choice probabilities) to be able to fully capture the effect of the topology of networks on convergence rates. We assume the probability of making a mistake to be $\mathscr{P} = \frac{1}{m}\exp(-\beta)$, identical for all players, strategies and states. The parameter m is the size of the action set and β is the parameter of noise, such that the larger β the smaller the probability of mistakes. For β equal to infinity, we have the case of best-response dynamics. The choice probabilities are thus of the form,

$$\mathbb{P}_i(a, \mathbf{x}) = (1-\varepsilon)BR_i(a, \mathbf{x}) + \frac{1}{m}\varepsilon \tag{2}$$

where $\varepsilon = \exp(-\beta)$ and $BR_i(a, \mathbf{x})$ is the probability that i plays action a under best-reply dynamics given the population state $\mathbf{x}$. That is

$$BR_i(a, \mathbf{x}) = \begin{cases} 1 & \text{if } a \in \arg\max_{b \in X} U_i(b, \mathbf{x}_{-i}) \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

This model of mistakes is similar to that in Freidlin and Wentzell (1984), Kandori, Mailath, and Rob (1993), Young (1993), Catoni (1999) and Beggs (2005) among others. The above dynamics defines a Markov chain on a finite state space $\mathbf{X}$ whose transition matrix P_ε is defined by transition probabilities $P_\varepsilon(\mathbf{x}, \mathbf{y})$,

$$P_\varepsilon(\mathbf{x}, \mathbf{y}) = \prod_{i=1}^{n} \mathbb{P}_i(y^i, \mathbf{x}) \tag{4}$$

2.2. *Closed communication classes and basins of attraction*

The perturbed Markov chain $(\mathbf{X}, P_\varepsilon)$ has a unique invariant (stationary) distribution $\pi_\varepsilon = \lim_{t\to\infty} \mathbf{q}_0 P_\varepsilon^t$, where $\mathbf{q}_t$ is the vector of probability mass functions at period t. The stationary distribution describes the amount of time the process spends in each state in the long-run. The *closed communication classes* of $(\mathbf{X}, P_\varepsilon)$ are also the closed communication classes of the identical process $(\mathbf{X}, P)$ without mistakes.[1] In the coordination game of Table 1 above for example, the limit sets include the singleton sets in which all players play strategy A and in which they all play strategy B. Generally, the limit sets of $(\mathbf{X}, P)$ can include sets that are cycles and those in which players use different strategies. When the interactions are governed by a social network , the number of closed communication classes is enhanced. In particular, there will exists singleton closed communication

[1] A set $\Omega \subset \mathbf{X}$ is a closed communication class of $(\mathbf{X}, P)$ if $\forall \mathbf{y} \in \Omega$, $\mathbb{P}(\mathbf{x}_{t+1} \in \Omega | \mathbf{x}_t = \mathbf{y}) = 1$, and that $\forall \mathbf{y}, \mathbf{z} \in \Omega$, there exists a $\tau > 0$ such that $\mathbb{P}(\mathbf{x}_{t+\tau} = \mathbf{z} | \mathbf{x}_t = \mathbf{y}) > 0$

classes in which strategies co-exists and different *cohesive* subgroups adopt different strategies. The closed communication classes of $(\mathbf{X}, P)$ that results in the long-run depends on the initial state of the process. We denote the typical limit set of $(\mathbf{X}, P)$ and hence recurrent class of $(\mathbf{X}, P_\varepsilon)$ by Ω, and by $\mathbf{\Omega}$ for the set of all such subsets

The *basin of attraction* $\partial\Omega$ of a closed communication class Ω is defined as $\partial\Omega = \{\mathbf{x} \in \mathbf{X} | \mathbb{P}(\exists T \text{ s.t } \mathbf{x}_t \in \Omega \ \forall \ t > T | \mathbf{x}_0 = \mathbf{x}) = 1\}$. That is, the set of states from which the chain without noise converges to Ω. We also write $\tilde{\Omega} = \partial\Omega \cup \Omega$ and will repeated refer to the process $(\mathbf{\Omega}, P_\varepsilon)$ as a chain with transitions between closed communication classes and $(\tilde{\mathbf{\Omega}}, \tilde{P}_\varepsilon)$ as the chain with transitions between elements $\tilde{\Omega}$. The chains $(\mathbf{\Omega}, P_\varepsilon)$ and $(\tilde{\mathbf{\Omega}}, \tilde{P}_\varepsilon)$ have the same dimensions. We also write $\mathbb{P}(\Omega, \partial\Omega')$ for the transition probability from some state in the closed communication class Ω to some state in the boundary of the basin of attraction $\partial\Omega'$.

3. Expected waiting time

In this section, we provide bounds for expected waiting times between closed communication classes. The results we establish in this section are also crucially relevant for the analysis of convergence times in the next section. The characterization of expected waiting times utilizes the concepts of hitting time for finite Markov chains. In particular, we make us of Lemma 3.4 in Freidlin and Wentzell (1984, Chapter 6) and Lemma 4 in Catoni (1999). Before stating the Lemma the following notations and definitions are necessary.

Given the Markov chain $(\mathbf{X}, P_\varepsilon)$, let $W \subset \mathbf{X}$ be a subset of $\mathbf{X}$ and $W^c = \mathbf{X} \setminus W$ its complement. For any oriented graph $g \subset \mathbf{X} \times \mathbf{X}$ and any $\mathbf{x} \in \mathbf{X}$, write $g(\mathbf{x}) = \{\mathbf{y} : (\mathbf{x}, \mathbf{y}) \in g\}$ denote the immediate successors of $\mathbf{x}$. More generally $g^n(\mathbf{x}) = \bigcup_{\mathbf{y} \in g^{n-1}(\mathbf{y})} g(\mathbf{x})$.

Definition 1: *Let $G(W)$ be the set of oriented graphs $g \subset \mathbf{X} \times \mathbf{X}$ satisfying*

1. *for any $\mathbf{x} \in \mathbf{X}$, $\#g(\mathbf{x}) = \mathbf{1}_{W^c}$ (that is no arrows start from W and exactly on arrow starts from each state outside of W),*
2. *for any $\mathbf{x} \in \mathbf{X}$, $\mathbf{x} \notin \mathcal{O}_g(\mathbf{x})$, where $\mathcal{O}_g(\mathbf{x}) = \bigcup_{n=1}^{+\infty} g^n(\mathbf{x})$ is the set of points that can be reached from $\mathbf{x}$; the orbit of $\mathbf{x}$ under g (that is g has no loops).*

Definition 2: *For any $\mathbf{x} \in \mathbf{X}$ and $\mathbf{y} \in W$, write*

$$G_{\mathbf{x},\mathbf{y}}(W) = \begin{cases} \{g \in G(W) : \mathbf{y} \notin \mathcal{O}_g(\mathbf{x})\} & \text{if } \mathbf{x} \in W^c \\ G(W) & \text{if } \mathbf{x} = \mathbf{y} \\ \emptyset & \text{if } \mathbf{x} \in W \setminus \{\mathbf{y}\} \end{cases}$$

That is $G_{\mathbf{x},\mathbf{y}}(W)$ is the set of graphs $g \in G(W)$ linking $\mathbf{x}$ to $\mathbf{y}$. The following Lemma is proved in Freidlin and Wentzell (1984) and in Catoni (1999).

Lemma 1: *Let $T(W) = \inf\{t \geq 0 : \mathbf{x}_t \in W\}$ be the first time the chain reaches W. For any $W \neq \emptyset$, any $\mathbf{x} \in W^c$,*

$$\mathbb{E}[T(W) | \mathbf{x}_0 = \mathbf{x}] = \left(\sum_{\mathbf{y} \in W^c} \sum_{g \in G_{\mathbf{x},\mathbf{y}}(W \cup \{\mathbf{y}\})} P_\varepsilon(g) \right) \left(\sum_{g \in G(W)} P_\varepsilon(g) \right)^{-1} \tag{5}$$

where $\mathbb{E}$ stands for expectation and $P_\varepsilon(g) = \prod_{(\mathbf{y},\mathbf{z}) \in g} P_\varepsilon(\mathbf{y}, \mathbf{z})$.

For $\mathbf{x} \neq \mathbf{y}$, the graph $G_{\mathbf{x},\mathbf{y}}(W \cup \{\mathbf{y}\})$ is the set of all $G(W)$ graphs that link $\mathbf{x}$ to $\mathbf{y}$. Implies that if W^c is a singleton set, then the component $\sum_{\mathbf{y}\in W^c} \sum_{g\in G_{\mathbf{x},\mathbf{y}}(W\cup\{\mathbf{y}\})} P_\varepsilon(g)$ on the right hand side of (5) is equal to one.

The above definitions and Lemma 1 directly extend to the Markov chains $(\mathbf{\Omega}, P_\varepsilon)$ and $(\tilde{\mathbf{\Omega}}, \tilde{P}_\varepsilon)$. That is for $(\mathbf{\Omega}, P_\varepsilon)$, $P_\varepsilon(g) = \prod_{(\Omega,\Omega')\in g} P_\varepsilon(\Omega, \Omega')$ in (5) above, and for $(\tilde{\mathbf{\Omega}}, \tilde{P}_\varepsilon)$

$$\mathbb{E}[T(W)|\ \tilde{\Omega}_0 = \tilde{\Omega}] = \left(\sum_{\tilde{\Omega}'\in W^c} \sum_{g\in G_{\tilde{\Omega},\tilde{\Omega}'}(W\cup\{\tilde{\Omega}'\})} \tilde{P}_\varepsilon(g) \right) \left(\sum_{g\in G(W)} \tilde{P}_\varepsilon(g) \right)^{-1} \tag{6}$$

for any $W \neq \emptyset$, any $\tilde{\Omega}' \in W^c$, where $\tilde{P}_\varepsilon(g) = \prod_{(\tilde{\Omega},\tilde{\Omega}')\in g} \tilde{P}_\varepsilon(\tilde{\Omega}, \tilde{\Omega}')$.

A complete characterization of the expected waiting times therefore requires establishing the structure of the probabilities $P_\varepsilon(\Omega, \Omega')$. First, it is convenient to split the transition $\Omega \to \Omega'$ into two transitions, from $\Omega \to \partial\Omega'$ and $\partial\Omega' \to \Omega'$. Such that

$$P_\varepsilon(\Omega, \Omega') = \mathbb{P}(\Omega, \partial\Omega')\mathbb{P}(\partial\Omega', \Omega') \tag{7}$$

We refer to the probabilities $\mathbb{P}(\Omega, \partial\Omega')$ as *exit probabilities*, in the sense of exiting the basin of attraction of Ω, and $\mathbb{P}(\partial\Omega', \Omega')$ as *contagion probabilities* for the reason that will become clear below.

3.1. Exit probabilities

The characterization of exit probabilities on the properties of the chain without noise. More specifically, on the properties of the basins of attractions of closed communication classes. We define two quantities relating any pair of closed communication classes say Ω and Ω'. The first quantity is $R(\Omega, \partial\Omega')$, the number of mistakes required for the transition from some state $\mathbf{x}$ in Ω to another $\mathbf{y}$ in $\partial\Omega'$ to occur under the zero noise chain. Here, R stands for *radius*. The normalized version of R is $r(\Omega, \partial\Omega') = \frac{1}{N} R(\Omega, \partial\Omega')$. The second quantity is $C(\Omega, \partial\Omega')$, the *cost* of the transition from some state $\mathbf{x}$ in Ω to another $\mathbf{y}$ in $\partial\Omega'$ under the chain with noise. The cost is formally given by

$$C(\Omega, \partial\Omega') = -R(\Omega, \partial\Omega') \ln \mathscr{P} = -R(\Omega, \partial\Omega') \left(\ln m^{-1} - \beta \right) \tag{8}$$

We also write $c(\Omega, \partial\Omega') = \frac{1}{N} C(\Omega, \partial\Omega')$ for the normalized cost.

Lemma 2: *For any pair Ω and Ω' of closed communication classes,*

$$\mathbb{P}(\Omega, \partial\Omega') \approx K_N \exp\left\{-N \left(c(\Omega, \partial\Omega') + F(r(\Omega, \partial\Omega'))\right)\right\} \tag{9}$$

where $F(r) = r \ln r - (1 - r) \ln(1 - r)$.

Proof. See Appendix A.1 □

3.2. Contagion probabilities

This section is devoted to deriving bounds for contagion probabilities in terms of the parameters of the network topology and the underlying game. The characterization is based on the notion

that once the chain enters the boundaries of the basin of attraction, it acquires a quasi-stationary distribution over the state space of the given basin of attraction. The quasi-stationary distribution attained places most weight on the corresponding closed communication class. The contagion rate within a basin of attraction is precisely the rate at which the chain attains its quasi-stationary distribution. The following notations and definitions are useful in stating the result.

Since $\mathbb{P}(\partial\Omega, \Omega)$ increases with the *contagion rate* $\mathcal{R}(\partial\Omega, \Omega)$ from $\partial\Omega \to \Omega$, we can thus write $\mathbb{P}(\partial\Omega, \Omega) = \mathcal{R}(\partial\Omega, \Omega)^{K_\Omega}$, where $0 \leq K_\Omega \leq 1$ depend on the properties of the process within $\tilde{\Omega}$. We refer to $\mathcal{R}(\partial\Omega, \Omega)$ as the contagion rate since the dynamics $\partial\Omega \to \Omega$ is governed by best-response, hence behaves as though the strategy is spreading contagiously. If an action played in any state say $\mathbf{x}$ in Ω is contagious, then K_Ω will be close to one. It is close to zero otherwise. We begin by characterizing the contagion rate in terms of the parameters of the network structure.

Given the normalized adjacency matrix $\mathscr{A}$ associated with the network of interactions, let $\rho(\mathscr{A}) = (\lambda_1, \cdots, \lambda_N)$ be its eigenvalue spectrum ordered in such away that $\lambda_1 = 1 \geq \lambda_2 \geq \cdots \geq \lambda_N$. Denote by Σ_ε for a player's individual transition matrix given an opponent's actions. That is let $\mathbb{P}(x_j|x_i)$ be the probability that a given player plays action $x_j \in X$ given that his opponent is playing $x_i \in X$ in the current period. Then Σ_ε is given by

$$\Sigma_\varepsilon = \begin{pmatrix} \mathbb{P}(x_1|x_1) & \mathbb{P}(x_2|x_1) & \cdots & \mathbb{P}(x_m|x_1) \\ \mathbb{P}(x_1|x_2) & \mathbb{P}(x_2|x_2) & \cdots & \mathbb{P}(x_m|x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{P}(x_1|x_m) & \mathbb{P}(x_2|x_m) & \cdots & \mathbb{P}(x_m|x_m) \end{pmatrix} \tag{10}$$

Let also $\rho(\Sigma_\varepsilon) = (\vartheta_1, \cdots, \vartheta_m)$ be the eigenvalue spectrum of Σ_ε. The contagion rate is formally defined as follows.

Definition 3: *For a given basin of attraction say $\tilde{\Omega}$ with respective closed communication class Ω, let $\nu_{\tilde{\Omega}}$ be the quasi-stationary distribution of the chain within $\tilde{\Omega}$. Then the contagion rate $\mathcal{R}(\partial\Omega, \Omega)$ within $\tilde{\Omega}$ is defined as*

$$\mathcal{R}(\partial\Omega, \Omega) = 1 - \limsup_{t \to T_{\tilde{\Omega}}} \left|\left| P_\varepsilon^t \mathbf{q}_{\tilde{\Omega}} - \nu_{\tilde{\Omega}} \right|\right|^{\frac{1}{t}}$$

where $T_{\tilde{\Omega}}$ is the period at which $\nu_{\tilde{\Omega}}$ is attained once the chain is in $\tilde{\Omega}$, and $\mathbf{q}_{\tilde{\Omega}}$ is the PMF that places most weight on the initial state of the chain in $\tilde{\Omega}$. The following Proposition provides bounds for contagion rates.

Proposition 1: *The contagion rate $\mathcal{R}(\partial\Omega, \Omega)$ within a basin of attraction has the following bounds*

$$1 - \lambda_2(G) \leq \mathcal{R}(\partial\Omega, \Omega) \leq 1 - \vartheta_m \lambda_2(G)$$

where $\lambda_2(G)$ is the second largest eigenvalue of the normalized adjacency matrix of network G.

Proof. See Appendix A.2 □

The eigenvalues of Σ_ε, ϑ_m are functions of the underlying payoff structure and the parameter β. Such that for large values of β, ϑ_m tend to one.The quantity $1 - \lambda_2(G)$ is also generally referred to

as the *spectral gap* of G. A bound on λ_2 can be established through its relationship with the graph conductance $\phi(G)$. We give examples for specific graphs below.

EXAMPLE: The following relation between second the eigenvalue of a network graph G, $\lambda_2(G)$ and $\phi(G)$ can be derived from Cheeger inequality: See Appendix A.3 for more detail concerning the relation plus derivations for the following examples.

$$\lambda_2(G) \leq 1 - \frac{\phi(G)^2}{2}. \tag{11}$$

(i) Complete network (G_{com}), a network structure in which every vertex is connected to every other vertex: $\lambda_2(G_{com}) \leq \frac{7}{8}$.

(ii) $1-D$ cyclic network (G_{cyc}), a network in which vertices are arranged in a circle and every vertex is connected to two other neighboring vertices: $\lambda_2(G_{cyc}) \leq \frac{N^2-2}{N^2}$.

(iii) $2D$ $N \times N$ lattice network (G_{2D}), a lattice structure constructed with periodic boundary conditions such that each agent is connected to 4 neighbors: $\lambda_2(G_{2D}) \leq \frac{16N^2-1}{16N^2}$.

(iv) Random d-regular network (G_{d-r}), a network structure in which each of the N vertices is connected to d other vertices chosen at random: $\lambda_2(G_{d-r}) \leq \frac{7}{8}$.

(v) Newman's small world network (G_{nsw}), a network structure in which the mean shortest-path between nodes increases sufficiently slowly (logarithmically) as a function of the number of nodes in the network: $\lambda_2(G_{nsw}) = 1 - \mathcal{O}\left(\frac{c}{(\ln N)^2}\right)$, where c is a constant.

Proof. See Appendix A.3 □

In general, densely connected and random networks have higher second largest eigenvalues compared to sparsely connected network. The direct implication of this result, together with the finding on expected waiting time, is that diffusion of strategic behavior or choices is faster in networks made up of cohesive subgroups. These cohesive subgroups should be small enough to favor shorter expected waiting time but large enough and sufficiently connected to favor higher diffusion rate within them.

Given the bounds for $\mathcal{R}(\partial\Omega, \Omega)$, we can then express the contagion probabilities for β sufficiently large as

$$\mathbb{P}(\partial\Omega, \Omega) \approx \exp\left\{K_\Omega \ln(1 - \lambda_2(G))\right\}, \tag{12}$$

and together with (9), equality (7) becomes

$$P_\varepsilon(\Omega, \Omega') \approx K_N \exp\left\{-N\left(c(\Omega, \partial\Omega') + F(r(\Omega, \partial\Omega'))\right) + \gamma(\Omega')\right\} \tag{13}$$

where $\gamma(\Omega') = K_{\Omega'} \ln(1 - \lambda_2(G))$.

THEOREM 1: *For any $W \neq \emptyset$, any $\Omega \in W$*

$$\mathbb{E}[T(W^c)|\ \mathbf{x}_0 \in \Omega] \leq K_T \exp\left\{\min_{g \in G(W^c)} \psi(g) - \min_{\Omega' \in W} \min_{g \in G_{\Omega,\Omega'}(W \cup \{\Omega'\})} \psi g\right\} \tag{14}$$

where $\psi(g) = N(c(g) + F(r(g))) - \gamma(g)$, $c(g) = \sum_{(\Omega,\Omega') \in g} c(\Omega, \partial\Omega')$, $\gamma(g) = \sum_{(\Omega,\Omega') \in g} \gamma(\Omega')$ and $F(r(g)) = \sum_{(\Omega,\Omega') \in g} F(r(\Omega, \partial\Omega'))$.

Proof. See Appendix A.4 □

Through the analysis that follows we let $M_W(\Omega) = \min_{\Omega' \in W} \min_{g \in G_{\Omega,\Omega'}(W \cup \{\Omega'\})} \psi(g)$, $M_W = \min_{g \in G(W^c)} \psi(g)$ and $H_W(\Omega) = M_W - M_W(\Omega)$. From (14), for finite and relatively small values of β, two factors influence the expected waiting time for evolution in networks. The first is the radii (resistances) of closed communication classes, which are determined by the underlying payoff and network topology. The second factor is the spectral property of the underlying network. For a given model of mistakes and equal basins of attraction, highly connected networks exhibit lower expected waiting times than sparsely connected networks. This result follows from the discussion about second largest eigenvalues of graphs above. We elaborate on this point further below for specific families of networks. As β becomes exceedingly large, the effect of the spectral gap of the underlying network on the expected waiting time becomes negligible, that is

$$\lim_{\beta \to \infty} \frac{1}{\beta} \ln \mathbb{E}(T(W^c) | \, \mathbf{x}_0 \in \Omega) = N \left(\min_{g \in G(W^c)} r(g) - \min_{\Omega' \in W} \min_{g \in G_{\Omega,\Omega'}(W \cup \{\Omega'\})} r(g) \right)$$
$$= \min_{g \in G(W^c)} R(g) - \min_{\Omega' \in W} \min_{g \in G_{\Omega,\Omega'}(W \cup \{\Omega'\})} R(g)$$

where $R(g) = \sum_{(\Omega,\Omega') \in g} R(\Omega, \partial\Omega')$. This is then equivalent to limit results in Beggs (2005). A more detailed examination of the expected waiting times for various families of networks is performed in sections that follow. Here, we provide a simple example demonstrating its computation.

3.3. *Example*

Consider a 3×3 matrix coordination game in which there are three singleton closed communication classes $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$. Figure 1 summarizes the transitions among closed communication classes where the weight on each arrow is the quantity $\psi(\Omega, \Omega') = N\left(c(\Omega, \partial\Omega') + F(r(\Omega, \partial\Omega'))\right) - \gamma(\Omega')$.

Corollary 1: *Given $\psi(\Omega, \Omega')$ for each pair (Ω, Ω') in Figure 1,*

(i) if $W = \{\mathbf{a}, \mathbf{b}\}$, then $\mathbb{E}(T(W^c) | \, \mathbf{x}_0 = \mathbf{a}) \leq K_T \exp(3.33N)$,

(ii) if $W = \{\mathbf{a}, \mathbf{c}\}$, then $\mathbb{E}(T(W^c) | \, \mathbf{x}_0 = \mathbf{c}) \leq K_T \exp(4.96N)$.

Proof. When $W = \{\mathbf{a}, \mathbf{b}\}$, the least cost graph linking W to $\mathbf{c}$ is $\mathbf{a} \to \mathbf{b} \to \mathbf{c}$. Implying that $M_W = 3.58$. The graph that minimizes $M_W(\mathbf{a})$ is $\mathbf{a} \to \mathbf{b}$, hence $M_W(\mathbf{a}) = 0.25$. Implying that $H_W(\mathbf{a}) = M_W - M_W(\mathbf{a}) = 3.33$.

When $W = \{\mathbf{a}, \mathbf{c}\}$, we have $\{\mathbf{a} \to \mathbf{b}, \mathbf{c} \to \mathbf{b}\}$ to be the least cost graph linking W to $\mathbf{b}$. Hence $M_W = 3.58$. The graph $\mathbf{a} \to \mathbf{b}$ minimizes $M_W(\mathbf{c})$, hence $M_W(\mathbf{c}) = 0.25$. Implying that $H_W(\mathbf{c}) = 4.96$. The respective expected waiting times directly follow by substituting for $H_W(\Omega)$. □

4. Convergence time

In this section, we derive bounds for the convergence time as a measure of how fast (slow) equilibrium selection occurs. We employ the concept of mixing time for finite Markov chains, which is a fairly well documented concept. Mixing times can be defined in terms of various distance

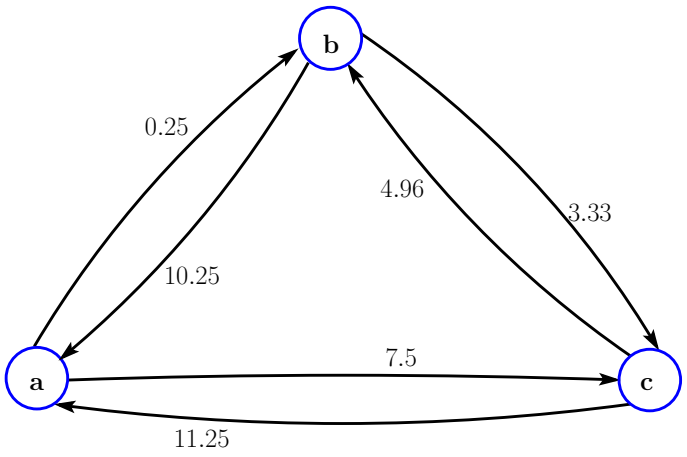

Figure 1: The weight on each arrow is the quantity $N\left(c(\Omega, \partial\Omega') + F(r(\Omega, \partial\Omega'))\right) - \gamma(\Omega')$

measures such as L^p-norms, but here we opt for the most intuitive distance measure; the *total variation distance*. Total variation distance is a probabilistic measure of the difference between two distribution. Formally, let ν and π be two probability distributions on the state space $\mathbf{X}$. Then the total variation distance is defined by

$$||\nu - \pi||_{TV} = \max_{\chi \subset \mathbf{X}} |\nu(\chi) - \pi(\chi)|.$$

That is it is the maximum difference between the probabilities assigned to a single event by the two distributions. Total variation distance can be equivalently expressed as,

$$||\nu - \pi||_{TV} = \frac{1}{2} \sum_{\mathbf{x} \in \mathbf{X}} |\nu(\mathbf{x}) - \pi(\mathbf{x})|. \tag{15}$$

The validity of relation (15) can be demonstrated geometrically using Figure 2. First notice that both ν and π being probabilistic measures implies that $\nu(\chi^*)$ is the area under the curve ν but to the left of the vertical dashed line $\mathbf{x} = \chi^*$. Since for both distributions the total area under the curve is one, it follows that the area of regions A and B are equal. For these two specific distributions (of Figure 2), the total variation distance between them is thus the area of the region A, which by symmetry equals the area of region B. The quantity $\sum_{\mathbf{x} \in \mathbf{X}} |\nu(\mathbf{x}) - \pi(\mathbf{x})|$ in (15) is the area of both regions A and B, hence the factor of $\frac{1}{2}$ in the equality. Relation (15) is essentially true for all forms of probabilistic distributions.

The convergence time $T_c(P_\varepsilon)$ of the process P_ε is then formally defined as,

$$T_c(P_\varepsilon) = \min\left\{t : \left|\left|\mathbf{q}_0 P_\varepsilon^{t'} - \pi\right|\right| \leq \varpi; \forall t' > t\right\}, \tag{16}$$

where $\mathbf{q}_0$ is the initial distribution. We also write $\mathbf{q}_t$ for the distribution at t. For a given initial distribution, the convergence time is therefore the time it takes the process to get close to its stationary distribution.

The usual approach in the study of mixing times is to first derive the lower and upper bounds in terms of the second largest eigenvalue of the transition matrix of the chain. The analysis then focuses of establishing the properties of the second eigenvalue, which can be done in several ways.

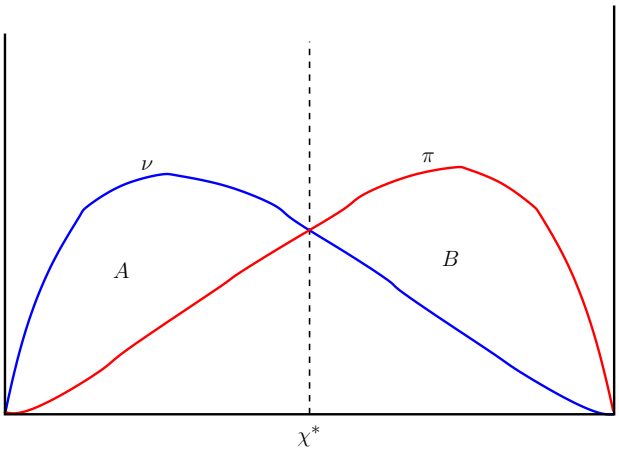

Figure 2: The total variation distance between distributions ν and π.

One technique is by making use of the well known Cheeger's inequality (Alon, 1986; Aldous, 1987; Sinclair and Jerrum, 1989; Mohar, 1989). That is if μ_2 is the second largest eigenvalue of the P_ε, then

$$1 - 2\phi(P_\varepsilon) \leq \mu_2 \leq 1 - \frac{\phi(P_\varepsilon)^2}{2} \tag{17}$$

where $\phi(P_\varepsilon)$, the conductance of P_ε, is given by $\phi(P_\varepsilon) = \min_{\substack{W \subset \mathbf{X} \\ \sum_{\mathbf{x}\in W} \pi(\mathbf{x}) \leq \frac{1}{2}}} \phi(P_\varepsilon, W)$. And,

$$\phi(P_\varepsilon, W) = \frac{1}{\sum\limits_{\mathbf{x}\in W} \pi(\mathbf{x})} \sum_{\mathbf{x}\in W} \sum_{\mathbf{y}\in W^c} \pi(\mathbf{x}) P_\varepsilon(\mathbf{x}, \mathbf{y}). \tag{18}$$

The second technique involves the use of the concept of multicommodity flows. The details of both of these characterizations can be found in Diaconis and Stroock (1991) and Sinclair (1992). These characterizations are however not readily computable for large state space Markov chains. Moreover, they are not directly intuitive with respect to how the underlying game and network affect convergence rates. Here, we express the upper bounds for convergence time in terms of the *maximum expected waiting time*. This enables us to directly apply some of the well known results on expected waiting times for stochastic evolutionary processes, together with what we established in Section 3 above, to the analysis of convergence time. It also ensures that we can focus on the properties of the reduced Markov chain $(\tilde{\mathbf{\Omega}}, \tilde{P}_\varepsilon)$ rather than the original cumbersome chain. The following theorem establishes the upper bound for the convergence time.

THEOREM 2: *For any $W \neq \emptyset$, any $\tilde{\Omega} \in W$ and $\tilde{\Omega}' \in W^c$, let $\mathscr{T} = \max_{\tilde{\Omega},\tilde{\Omega}'} \mathbb{E}[T(\tilde{\Omega}')|\, \mathbf{x}_0 \in \tilde{\Omega}]$. Let also $\tilde{\pi}$ be the stationary distribution of $\tilde{P}_\varepsilon$. Then the convergence time T_c is bounded from above by*

$$T_c \leq \frac{K_c}{2\tilde{\pi}_{\min}^2} \left(\ln \tilde{\pi}_{\min}^{-1} + \ln(2\varpi)^{-1} \right) \mathscr{T} \tag{19}$$

where K_c is some constant, $\tilde{\pi}_{\min} = \min_{\tilde{\Omega}\in\tilde{\mathbf{\Omega}}} \tilde{\pi}(\tilde{\Omega})$ and $\tilde{\pi}_{\max} = \max_{\tilde{\Omega}\in\tilde{\mathbf{\Omega}}} \tilde{\pi}(\tilde{\Omega})$.

Proof. A detailed proof is provided in Appendix A.5 whose sketch is as follows. The proof focuses on the dynamics of the reduced chain $(\tilde{\mathbf{\Omega}}, \tilde{P}_\varepsilon)$. The first step involves showing that

$$T_c \leq \frac{1}{\ln \tilde{\mu}_2}\left(\ln \tilde{\pi}_{\min}^{-1} + \ln(2\varpi)^{-1}\right) \approx \frac{1}{1-\tilde{\mu}_2}\left(\ln \tilde{\pi}_{\min}^{-1} + \ln(2\varpi)^{-1}\right) \tag{20}$$

where $\tilde{\mu}_2$ is the second largest eigenvalue of the $\tilde{P}_\varepsilon$. The approximation is valid for values of $\tilde{\mu}_2$ close to one, which is true for the type of dynamics we considered in this paper (a consequence of small perturbations). Inequality (20) appears in the literature of finite Markov chains in closely related forms (e.g. Diaconis and Stroock (1991)).

The next step of the proof is then to show that

$$\frac{1}{1-\tilde{\mu}_2} \leq \frac{K_c}{2\tilde{\pi}_{\max}^2}\mathscr{T} \tag{21}$$

The right hand side of (21) follows from a set of Theorems from Bremaud (1999) and Takacs (2006). More specifically, as a corollary of Takacs (2006, Theorem 1.2 and Proposition 3.1), the expected waiting time between any two given subsets say $\tilde{\Omega}$ and $\tilde{\Omega}'$ is given by

$$\mathbb{E}(T(\tilde{\Omega}')|\, \mathbf{x}_0 \in \tilde{\Omega}) = \sum_{k=2}^{\#\tilde{\mathbf{\Omega}}} \frac{1}{1-\tilde{\mu}_k}\left(\tilde{v}_{\tilde{\Omega}'}^2(k) - \tilde{v}_{\tilde{\Omega}}(k)\tilde{v}_{\tilde{\Omega}'}(k)\right) \tag{22}$$

where $\tilde{\mathbf{v}}(k)$ is the eigenvector corresponding to $\tilde{\mu}_k$, such that $\tilde{v}_l(k)$ is the lth element of $\tilde{\mathbf{v}}(k)$. We thus have that

$$\mathscr{T} = \max_{\tilde{\Omega},\tilde{\Omega}'} \mathbb{E}(T(\tilde{\Omega}')|\, \mathbf{x}_0 \in \tilde{\Omega}) = \max_{\tilde{\Omega},\tilde{\Omega}'} \sum_{k=2}^{\#\tilde{\mathbf{\Omega}}} \frac{1}{1-\tilde{\mu}_k}\tilde{v}_{\tilde{\Omega}'}(k)\left(\tilde{v}_{\tilde{\Omega}'}(k) - \tilde{v}_{\tilde{\Omega}}(k)\right) \tag{23}$$

We can then exploit the properties of the eigenvectors of finite state Markov chains. The eigenvector corresponding to the first eigenvalues (first eigenvector) is an all one vector. The other remaining eigenvectors can be used to detect (or partition the state space into) almost invariant subsets (e.g. Froyland (2005) and Billings and Schwartz (2008)). This concept is familiar in the literature of cluster detection. More generally, the second eigenvector can be used to partition the chain into two almost invariant subsets, the third eigenvector separates the chain into three almost invariant subsets, and so forth. The eigenvector corresponding to the $\#\mathbf{\Omega}$th eigenvalue of P_ε and $\tilde{P}_\varepsilon$ therefore partitions the chain into $\#\mathbf{\Omega}$ almost invariant subsets, which are basically the respective basins of attraction.

The detection process proceeds by identifying the minima (regions of negative values) and maxima (peaks and regions of positive values) of the eigenvectors. The second eigenvector therefore has one maximum and one minimum. Since the maximum of the right hand side of (23) occurs when the differences $\tilde{v}_{\tilde{\Omega}'}(k) - \tilde{v}_{\tilde{\Omega}}(k)$ are maximum, we can thus approximate $\max_{\tilde{\Omega},\tilde{\Omega}'}\left(\tilde{v}_{\tilde{\Omega}'}(k) - \tilde{v}_{\tilde{\Omega}}(k)\right) \approx 2\max_{\tilde{\Omega}\in\tilde{\mathbf{\Omega}}} \tilde{v}_{\tilde{\Omega}}(k)$ for all $k \geq 2$. Note that each point of $\tilde{\mathbf{\Omega}}$ for which $\tilde{\mathbf{v}}(k)$ attains its minima and maxima are also those at which the stationary distribution $\tilde{\pi}$ attains its maxima. It follows that there exist constants K_k for $k \geq 2$ such that $\max_{\tilde{\Omega}\in\tilde{\mathbf{\Omega}}} |\tilde{v}_{\tilde{\Omega}}(2)| = K_2 \max_{W\subset\tilde{\mathbf{\Omega}}} \tilde{\pi}(W), \cdots,$ $\max_{\tilde{\Omega}\in\tilde{\mathbf{\Omega}}} |\tilde{v}_{\tilde{\Omega}}(\#\tilde{\mathbf{\Omega}})| = K_{\#\tilde{\mathbf{\Omega}}} \max_{\tilde{\Omega}\in\tilde{\mathbf{\Omega}}} \tilde{\pi}(\tilde{\Omega})$. Since $\max_{W\subset\tilde{\mathbf{\Omega}}} \tilde{\pi}(W) \geq \cdots \geq \max_{\tilde{\Omega}\in\tilde{\mathbf{\Omega}}} \tilde{\pi}(\tilde{\Omega}) = \tilde{\pi}_{\max}$, we then have that

$$\mathscr{T} = \max_{\tilde{\Omega},\tilde{\Omega}'} \sum_{k=2}^{\#\tilde{\mathbf{\Omega}}} \frac{1}{1-\tilde{\mu}_k}\tilde{v}_{\tilde{\Omega}'}(k)\left(\tilde{v}_{\tilde{\Omega}'}(k) - \tilde{v}_{\tilde{\Omega}}(k)\right) \geq 2K_{\#\tilde{\mathbf{\Omega}}}\tilde{\pi}_{\max}^2 \sum_{k=2}^{\#\tilde{\mathbf{\Omega}}} \frac{1}{1-\tilde{\mu}_k} \geq 2K_{\#\tilde{\mathbf{\Omega}}}\tilde{\pi}_{\max}^2 \frac{1}{1-\tilde{\mu}_2} \tag{24}$$

Substituting for $\mathscr{T}$ in (20) then completes the proof, where $K_c = \frac{1}{K_{\#\tilde{\mathbf{\Omega}}}}$. □

Theorem 2 establishes an upper bound for the convergence time in terms of the measure of the reduced chain, the minimum and maximum values of the stationary distribution and the maximum expected waiting time. An explicit characterization of the stationary distributions for finite state Markov chains is performed by Freidlin and Wentzell (1984) and Catoni (1999). More generally, for each $\tilde{\Omega} \in \tilde{\mathbf{\Omega}}$,

$$\tilde{\pi}(\tilde{\Omega}) = \left(\sum_{g \in G(\{\tilde{\Omega}\})} \tilde{P}_{\varepsilon}(g) \right) \left(\sum_{\tilde{\Omega}' \in \tilde{\mathbf{\Omega}}} \sum_{g \in G(\{\tilde{\Omega}'\})} \tilde{P}_{\varepsilon}(g) \right)^{-1} \tag{25}$$

where $\tilde{P}_{\varepsilon}(g) = \prod_{(\tilde{\Omega},\tilde{\Omega}') \in g} \tilde{P}_{\varepsilon}(\tilde{\Omega}, \tilde{\Omega}')$. From the definition of $G(W)$ graphs above, $G(\{\tilde{\Omega}\})$ is the set of all spanning trees in which there exists a unique directed path from every $\tilde{\Omega}' \neq \tilde{\Omega}$ to $\tilde{\Omega}$, also known as $\tilde{\Omega}$-trees. The probability $\tilde{P}_{\varepsilon}(\tilde{\Omega}, \tilde{\Omega}') = \mathbb{P}(\partial\tilde{\Omega}, \tilde{\Omega})\mathbb{P}(\tilde{\Omega}, \partial\tilde{\Omega}')$, and by making use of (12) and (9),

$$\tilde{P}_{\varepsilon}(\tilde{\Omega}, \tilde{\Omega}') = K_N \exp\left\{-N\left(c(\Omega, \partial\Omega') + F(r(\Omega, \partial\Omega'))\right) + \gamma(\Omega)\right\},$$

such that

$$\frac{1}{\tilde{\pi}(\tilde{\Omega})} \leq K_{\pi} \exp\left\{\varphi(\tilde{\Omega})\right\} \tag{26}$$

where

$$\varphi(\tilde{\Omega}) = \min_{g \in G(\{\tilde{\Omega}\})} \left(N[c(g) + F(r(g))] - \gamma(g)\right) - \min_{\tilde{\Omega}' \in \tilde{\mathbf{\Omega}}} \min_{g \in G(\{\tilde{\Omega}'\})} \left(N[c(g) + F(r(g))] - \gamma(g)\right)$$

and K_{π} is a constant.

From (26) it follows that the long-run stable state or subset (corresponding to $\tilde{\pi}_{\max}$) lies in the basin of attraction $\tilde{\Omega}^*$ for which $\varphi(\tilde{\Omega}^*) = 0$. The state or subset corresponding to $\tilde{\pi}_{\min}$ is that for which $\varphi(\tilde{\Omega})$ is maximum. The following Corollary demonstrates how to compute convergence times.

Corollary 2: *Consider the an evolutionary process for which the values of the quantity* $N[c(\Omega, \partial\Omega') + F(r(\Omega, \partial\Omega'))] - \gamma(\Omega)$ *are as depicted in Figure 3. Then*

$$T_c \leq \kappa_1 \left(\kappa_2 + 4.38N\right) \exp(4.96N),$$

where $\kappa_1 = \frac{1}{2} K_c K_T K_{\pi}^2$ *and* $\kappa_2 = \ln(2K_{\pi}) - 1$ *are universal constants.*

Proof. To compute the maximum and minimum values for the stationary distributions, we need to identify and compute the values of the minimum and maximum cost $\tilde{\Omega}$-tree. From Figure 3, the least cost tree is the $\tilde{\mathbf{b}}$-tree $\{\tilde{\mathbf{a}} \to \tilde{\mathbf{b}}, \tilde{\mathbf{c}} \to \tilde{\mathbf{b}}\}$, making $\mathbf{b}$ the long-run stable state (which in this specific case also corresponds to the stochastically stable state). The respective total cost of $\tilde{\mathbf{b}}$-tree is 2.02. The maximum cost tree is the $\tilde{\mathbf{a}}$-tree with cost of 6.4, such that $\varphi(\tilde{\mathbf{a}}) = 6.4 - 2.02 = 4.38$. Hence $\tilde{\pi}_{\min}^{-1} \leq K_{\pi} \exp(4.38N)$.

The maximum expected waiting time corresponds to the hitting time of the path $W = \{\mathbf{b}, \mathbf{c}\} \to \{\mathbf{a}\}$ such that the maximum $M_W = 1.63 + 4.96 = 6.59$ and $M_W(\mathbf{c}) = 1.63$. Implying that $\mathscr{T} \leq K_T \exp(4.96N)$. Substituting for $\tilde{\pi}_{\min}^{-1}$, $\mathscr{T}$ and $\tilde{\pi}_{\min}^{-1} \leq K_{\pi}$ yields the result.

□

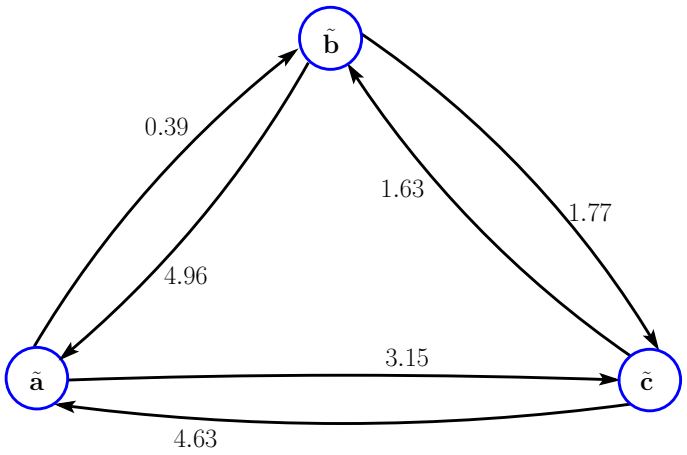

Figure 3: The weight on each arrow is the quantity $N\left(c(\Omega, \partial\Omega') + F(r(\Omega, \partial\Omega'))\right) - \gamma(\Omega)$

5. Deterministic networks

In this section, we focus on stochastic evolution in deterministic networks. We start by considering the case in which strategies can spread through contagion once threshold adoption levels are attained. An action is said to be globally contagious if it can spread to the entire population through best-response once it has been adopted by a small subgroup of players. The focus of analysis in this section is on 2×2 matrix games with the general structure in Table 1 above and the case in which players care about the average payoffs. The most relevant quantity of any given matrix game for the dynamics specified above is the relative payoff gains, which for the game in Table 1 is given by

$$\eta_{ba} = \frac{U(b,b) - U(a,b)}{U(b,b) - U(a,b) + U(a,a) - U(b,a)} \tag{27}$$

and $\eta_{ab} = 1 - \eta_{ba}$. That is η_{ba} is the fraction of neighbors that each player requires to play a for that player to do likewise and vise versa for η_{ab}. We write $\lceil m \rceil$ fo the smallest integer not less than m such that for each $i \in \mathcal{N}$ with respective degree k_i, the quantity $\lceil \eta_{ba} k_i \rceil$ is the minimum number of neighbors i requires to play a for i to do likewise.

For any given network, there exists a payoff structure and hence values of relative payoff gains η_{ab}'s for which contagion is feasible. The maximum values of relative payoffs for which contagion is feasible are called *contagion thresholds* (Morris, 2000). If η^* is the the contagion threshold for a given network topology, then any action say a for which $\eta_{ba} < \eta^*$ for all b, will be contagious. Morris (2000) provides an explicit characterization of contagion threshold for various families of networks, two of which we study below. Lelarge (2012) derives similar conditions for random networks. If the conditions for contagion are not satisfied, then the dynamic process will exhibit additional multiple closed communication classes induced by the network structure. We treat this case in a separate Section 7below.

5.1. *n-dimensions nearest neighbor interactions*

This family of networks includes as example the line and circle networks as 1-dimensional lattices, and the 2-dimension lattice of Figure 4. Morris (2000) shows the contagion threshold for such

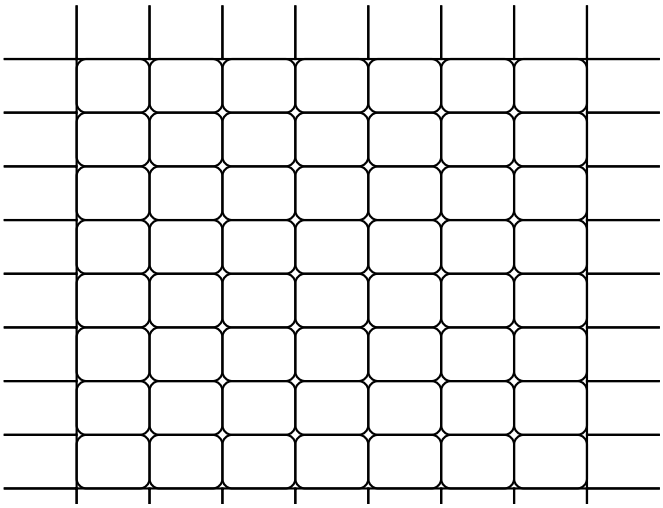

Figure 4: A 2-dimensions nearest neighbor interactions network.

networks to be $\frac{1}{2n}$. Implying that in 1-dimension lattice networks if $\eta_{ab} < \frac{1}{2}$ for all $a \neq b$, then b is a contagious action. Similarly for 2-dimensional lattice networks, if $\eta_{ab} < \frac{1}{4}$ for all $a \neq b$ then b is a contagious action. And so forth. The following proposition provides convergence rates for this family of networks.

PROPOSITION 2: *Let the underlying game be* 2×2 *matrix game of Table 1 and let the updating rule be the uniform error probabilities. Let the topology of interactions be an* n*-dimensions nearest neighbor interactions network* G_n *and that* $\eta_{ab} < \frac{1}{2n}$*. Then for sufficiently large* N*,*

(i) the expected waiting times are;

$$\mathbb{E}[T(\mathbf{b})|\mathbf{x}_0 = \mathbf{a}] \leq \frac{K_T}{2} \exp\{\beta - \ln(1 - \lambda_2(G_n))\} \tag{28}$$

$$\mathbb{E}[T(\mathbf{a})|\mathbf{x}_0 = \mathbf{b}] \leq K_T \exp\left\{N(\beta - \ln 2^{-1})\right\} \tag{29}$$

(ii) the convergence time is given by

$$T_c \leq \kappa_1 \left(\kappa_2 + (N-1)(\beta - \ln 2^{-1}) - \ln(1 - \lambda_2(G_n))\right) \exp\left\{N(\beta - \ln 2^{-1})\right\}, \tag{30}$$

where $\kappa_1 = K_c K_T K_\pi^2$ *and* $\kappa_2 = \ln(2K_\pi) - 1$ *are universal constants.*

Proof. For each n-dimensions nearest neighbor interactions network G_n if an action say b is contagious in which case $\eta_{ab} < \frac{1}{2n}$, then each player i requires only one of their neighbor to play b for i to do likewise. Implying that a transition from $\mathbf{a}$ to the boundaries of the basin of attraction $\partial\mathbf{b}$ of $\mathbf{b}$ can be triggered by one player. Such that $r(\mathbf{a}, \partial\mathbf{b}) = \frac{1}{N}$. Similarly, the transition $\mathbf{b} \to \partial\mathbf{a}$ requires all players to simultaneously switched to playing a, since each player requires at least $\frac{1}{2}n$ of their neighbors to play a if they are to do likewise. Hence $r(\mathbf{a}, \partial\mathbf{b}) = 1$.

For $m = 2$, $\mathscr{P} = \frac{1}{2}\exp(-\beta)$ such that $\ln(\mathscr{P}) = -\beta + \ln 2^{-1}$. We then have the expected waiting times to be,

$$\begin{aligned}\mathbb{E}[T(\mathbf{b})|\mathbf{x}_0 = \mathbf{a}] &\leq K_T \exp\left\{N\left(\frac{1}{N}\beta - \frac{1}{N}\ln 2^{-1} + F(\frac{1}{N})\right) - \ln(1 - \lambda_2(G_n))\right\} \\ &= \frac{K_T}{2}\exp\{\beta - \ln(1 - \lambda_2(G_n))\} \end{aligned} \tag{31}$$

where we have used $K_{\mathbf{b}} = 1$ since b is globally contagious.

$$\mathbb{E}[T(\mathbf{a})|\mathbf{x}_0 = \mathbf{b}] \leq K_T \exp\left\{N\left(\beta - \ln 2^{-1} + F(1)\right)\right\} = K_T \exp\left\{N(\beta - \ln 2^{-1})\right\} \tag{32}$$

where $K_{\mathbf{a}} = 0$ since a is not globally contagious. In the case of convergence time, we need compute the minimum and maximum values of the stationary distribution for subsets $\tilde{\Omega}$, and the maximum expected waiting time. From the above expressions of expected waiting, if β is sufficiently large (small noise) then $\mathscr{T} = \mathbb{E}[T(\mathbf{a})|\mathbf{x}_0 = \mathbf{b}] \leq K_T \exp\{N(\beta - \ln 2^{-1})\}$.

For the transition $\partial\mathbf{a} \to \mathbf{a}$ we have $\mathbb{P}(\partial\mathbf{a}, \mathbf{a}) \approx 1$ since all players have to simultaneously switch to play a for the transition $\mathbf{b} \to \partial\mathbf{a}$ to occur. We then have

$$\begin{aligned} \tilde{P}_\varepsilon(\tilde{a}, \tilde{b}) &\approx \mathbb{P}(\partial\mathbf{a}, \mathbf{a})\mathbb{P}(\mathbf{a}, \partial\mathbf{b}) \approx \mathbb{P}(\mathbf{a}, \partial\mathbf{b}) \\ &\leq \exp\left\{-N\left(\frac{1}{N}\beta - \frac{1}{N}\ln 2^{-1} + F(\frac{1}{N})\right)\right\} = K_p \exp\left\{-(\beta - \ln 2^{-1})\right\}. \end{aligned} \tag{33}$$

$$\begin{aligned} \tilde{P}_\varepsilon(\tilde{b}, \tilde{a}) &\approx \mathbb{P}(\partial\mathbf{b}, \mathbf{b})\mathbb{P}(\mathbf{b}, \partial\mathbf{a}) \\ &\leq \exp\left\{-N(\beta - \ln 2^{-1}) + K_{\mathbf{b}} \ln(1 - \lambda_2(G_n))\right\}. \end{aligned} \tag{34}$$

where we can let $K_{\mathbf{b}} = 1$ as b is globally contagious. Since for β sufficiently large $\beta - \ln 2^{-1} < (N(\beta - \ln 2^{-1}) - K_{\mathbf{b}} \ln(1 - \lambda_2(G_n)))$, the long-run stable basin of attraction is then $\tilde{\mathbf{b}}$ and hence the long-run stable state is $\mathbf{b}$. Implying that $\varphi(\tilde{\mathbf{b}}) = 0$ and $\varphi(\tilde{\mathbf{a}}) = N(\beta - \ln 2^{-1}) - \ln(1 - \lambda_2(G_n)) - (\beta - \ln 2^{-1}) = (N-1)(\beta - \ln 2^{-1}) - \ln(1 - \lambda_2(G_n))$. Such that

$$\frac{1}{\tilde{\pi}_{\max}} \equiv \frac{1}{\tilde{\pi}(\tilde{b})} \leq K_\pi \tag{35}$$

$$\frac{1}{\tilde{\pi}_{\min}} \equiv \frac{1}{\tilde{\pi}(\tilde{a})} \leq K_\pi \exp\left\{(N-1)(\beta - \ln 2^{-1}) - \ln(1 - \lambda_2(G_n))\right\} \tag{36}$$

Hence $\ln \tilde{\pi}_{\min}^{-1} = \ln K_\pi + (N-1)(\beta - \ln 2^{-1}) - \ln(1 - \lambda_2(G_n))$. Substituting for $\mathscr{T}$, $\tilde{\pi}_{\min}$ and $\tilde{\pi}_{\max}$ into the expression for T_c yields the result. □

Proposition 2 demonstrates how the network topology plays a critical role in shaping convergence rates. Unlike in the case of global interactions, local interactions reduce the expected waiting time to and increase the exit time from the long-run stable set. This finding is not entirely surprising as it has also been shown for the case of 1-dimension nearest neighbor interactions network by Ellison (1993). Proposition 2 however provides tighter bounds for general n-dimensions nearest neighbor interactions family of networks and does not rely on the assumption of vanishing noise for the proof as in Ellison (1993).

Proposition 2 also provides an upper bound for the convergence time for n-dimensions nearest neighbor interactions family of networks. From (30), it follows that for any to networks in this family say G_n and $G_{n'}$ with $n > n'$, if the relative payoff gains for both networks is such that $\eta_{ab} < \eta^*$ then the convergence time for G_n is greater than that for $G_{n'}$. This argument derives from the fact that when $n > n'$ then $\lambda_2(G_n) > \lambda_2(G_{n'})$. The result underscores the importance of high connectivity in enhancing high convergence rates. This however is not necessarily the case when $\eta_{ab} > \eta^*$, in which case global contagion is not feasible. We treat this special case as that of step-by-step evolution in Section 7 below.

5.2. *n-dimensions l-max distance interactions*

The second family of networks we consider is the n-dimensions l-max distance interactions. That is given an n-dimensional lattice network, each player interacts with all players who are within the distance of l steps away in all directions. See Figure 5 for the case of $n = 2$ and $l = 1$. Morris (2000) provides general expressions for contagion thresholds for this family of networks. We focus on the case of $n = 2$ for various values of l. In this special case the contagion thresholds are given by $\eta^* = \frac{l(2l+1)}{(2l+1)^2-1}$ for $l = 1, 2, 3, \cdots$. It follows by inspection that if $\eta_{ab} = \eta^*$ then

$$r(\mathbf{a}, \partial\mathbf{b}) = \frac{2l(2l+1)}{N} \quad \text{for } l = 1, 2, \cdots \tag{37}$$

See for example the case of $n = 2$ and $l = 1$ in Figure 5 where $r(\mathbf{a}, \partial\mathbf{b}) = \frac{6}{N}$. Relation (37) is valid so long as $\eta_{ab} = \eta^*$ and that $l < \frac{1}{2}\sqrt{N}$. The quantity $\sqrt{N}$ is the length and width of the 2-dimensional lattice. It follows that for $\eta_{ab} < \eta^*$, $r(\mathbf{a}, \partial\mathbf{b}) < \frac{2l(2l+1)}{N}$. The following proposition is for the case when $\eta_{ab} = \eta^*$.

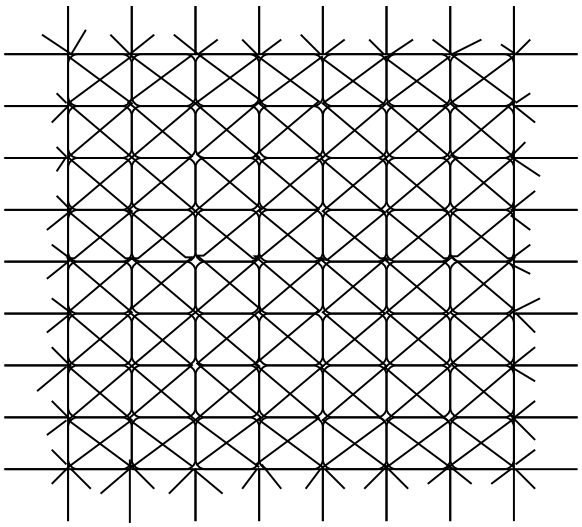

Figure 5: The 2-dimensions 1-max distance interactions.

PROPOSITION 3: *Let the underlying game be 2×2 matrix game of Table 1 and let the updating rule be the uniform error probabilities. If the topology of interactions is the 2-dimensions l-max distance interactions $G_{2,l}$ and that $\eta_{ab} = \eta^*$ then:*

(i) the expected waiting times are;

$$\mathbb{E}[T(\mathbf{b})|\mathbf{x}_0 = \mathbf{a}] \leq K_T \exp\{H(\mathbf{b})\} \tag{38}$$

where $H(\mathbf{b}) = 2l(2l+1)(\beta - \ln 2^{-1}) + NF\left(\frac{2l(2l+1)}{N}\right) - \ln(1 - \lambda_2(G_{2,l}))$.

$$\mathbb{E}[T(\mathbf{a})|\mathbf{x}_0 = \mathbf{b}] \leq K_T \exp\left\{N(\beta - \ln 2^{-1})\right\} \tag{39}$$

(ii) the convergence time is given by

$$T_c \leq \kappa_1 \left(\kappa_2 + \varphi(\tilde{\mathbf{a}})\right) \exp\left\{N(\beta - \ln 2^{-1})\right\}, \tag{40}$$

where $\varphi(\tilde{\mathbf{a}}) = \left(N - 2l(2l+1)\right)(\beta - \ln 2^{-1}) - NF\left(\frac{2l(2l+1)}{N}\right) - \ln(1 - \lambda_2(G_{2,l}))$.

Proof. First note that when $\eta_{ab} \geq \eta^*$, $r(\mathbf{b}, \partial\mathbf{a}) \approx 1$ for all $l = 1, 2, \cdots$. By following the same steps as in the proof of Proposition 2, we have that

$$\mathbb{E}[T(\mathbf{b})|\mathbf{x}_0 = \mathbf{a}] \leq K_T \exp\left\{N\left(\frac{2l(2l+1)}{N}(\beta - \ln 2^{-1}) + F\left(\frac{2l(2l+1)}{N}\right)\right) - \ln(1-\lambda_2(G_n))\right\}$$

$$= K_T \exp\left\{2l(2l+1)(\beta - \ln 2^{-1}) + NF\left(\frac{2l(2l+1)}{N}\right) - \ln(1-\lambda_2(G_{2,l}))\right\} \tag{41}$$

$$\mathbb{E}[T(\mathbf{a})|\mathbf{x}_0 = \mathbf{b}] \leq K_T \exp\left\{N\left(\beta - \ln 2^{-1} + F(1)\right)\right\} = K_T \exp\left\{N(\beta - \ln 2^{-1})\right\} \tag{42}$$

Similarly, for $\eta_{ab} = \eta^*$, $\mathbf{b}$ is the long-run stable state and hence $\frac{1}{\tilde{\pi}_{\max}} \equiv \frac{1}{\tilde{\pi}(\tilde{b})} \leq K_\pi$, and

$$\frac{1}{\tilde{\pi}_{\min}} \equiv \frac{1}{\tilde{\pi}(\tilde{a})} \leq K_\pi \exp\{\varphi(\tilde{\mathbf{a}})\}$$

where

$$\varphi(\tilde{\mathbf{a}}) = \left(N(\beta - \ln 2^{-1}) - \ln(1-\lambda_2(G_{2,l}))\right) - \left(2l(2l+1)(\beta - \ln 2^{-1}) + NF\left(\frac{2l(2l+1)}{N}\right)\right)$$

$$= (N - 2l(2l+1))(\beta - \ln 2^{-1}) - NF\left(\frac{2l(2l+1)}{N}\right) - \ln(1-\lambda_2(G_{2,l}))$$

Substituting for $\frac{1}{\tilde{\pi}_{\min}}$ and $\frac{1}{\tilde{\pi}_{\min}}$ as in the proof of Proposition 2 yields the desired result. □

The n-dimensions l-max distance interactions networks present a good example of effect of progressively increasing information sets of players. That is starting from players observing only actions of the other players one step away, then two steps away, and so forth. As l (information set) increases the expected waiting time to the long-run stable state increases, but the convergence time decreases. As l increases the network and hence the convergence rates tend to that of global interactions. It is easy to show the convergence time for global interactions when $\eta_{ab} < \eta_{ba}$ to be

$$T_c \leq \kappa_1\left(\kappa_2 + (1-2\eta_{ab})(N-1)\left(\beta - \ln 2^{-1}\right)\right)\exp\left\{N\left(\eta_{ba}(\beta - \ln 2^{-1}) + F(\eta_{ba})\right)\right\}, \tag{43}$$

which by comparison is less than the convergence time for 2-dimensions l-max distance interactions in (40) whenever $l < \frac{1}{2}\sqrt{N}$. In fact, the convergence time for global interactions in (43) is smaller than that for n-dimensions nearest neighbor interactions network (cf. in Proposition 2) as well. The general result is then that the convergence time for under global interactions is less than that under local interactions whenever global contagion is feasible.

6. Random networks

Most real world networks assume complex structures. Several authors have proposed various ways to capture the properties of such networks e.g. clustering coefficient and degree distribution. For our purpose, the degree distribution is the most suitable for capturing the effect of network topology on the convergence rates. The degree distribution $P = \{p(k)\}_{k\geq 0}$ describes the number of agents in the population with a given degree. We denote by $\langle k \rangle$ for average degree. The following definitions related to the degree distribution are useful for the results that follow.

Definition 4: *Let G and G' be two networks with respective degree distributions P and P'.*

(i) A degree distribution P is said to first order stochastically dominate (FOSD) P' if $\sum_{k=0}^{K} p(k) \leq \sum_{k=0}^{K} p'(k)$ for $1 \leq K \leq \infty$. Or for any non-decreasing function $f : \mathbb{R} \to \mathbb{R}$,

$$\sum_{k=0}^{\infty} f(k)p'(k) \leq \sum_{k=0}^{\infty} f(k)p(k).$$

(ii) A degree distribution P is said to second order stochastically dominate (SOSD) P' if for any non-decreasing concave function $f : \mathbb{R} \to \mathbb{R}$, $\sum_{k=0}^{\infty} f(k)p'(k) \leq \sum_{k=0}^{\infty} f(k)p(k)$.

(ii) A degree distribution P is said to be a mean preserving spread (MPS) of P' if P' SOSD P and that both have the same mean.

The following proposition provides the relationship between degree distributions and convergence rates.

PROPOSITION 4: *Let G and G' be two networks with respective degree distributions P and P' and let the underlying game be 2×2 matrix game of Table 1 with $\eta_{ab} < \frac{1}{2}$. We write $\mathbb{E}_G$ for the expected value under network topology G and $T_c(G)$ for convergence time under G*

(i) if P FOSD (or SOSD) P', then $\mathbb{E}_G[T(\mathbf{b})|\mathbf{x}_0 = \mathbf{a}] \geq \mathbb{E}_{G'}[T(\mathbf{b})|\mathbf{x}_0 = \mathbf{a}]$

(ii) if P MPS P' or P' FOSD P, then $T_c(G) \leq T_c(G')$

Proof. (i). For a given value of $\eta_{ab} < \frac{1}{2}$, the quantity $r(\mathbf{a}, \partial\mathbf{b})$ is an increasing function of network connectivity. This is clearly evidenced in the analysis of deterministic networks of Section 5. For example under 1-dimensions nearest neighbor interactions network $r(\mathbf{a}, \partial\mathbf{b}) = \frac{1}{N}$ while under global interactions $r(\mathbf{a}, \partial\mathbf{b}) \approx \eta_{ab}$. We thus write $r(\mathbf{a}, \partial\mathbf{b}) := f(\langle k \rangle)$, where f is an non-decreasing function in $\langle k \rangle$. Since P FOSD (or SOSD) P' implies that $\langle k \rangle \geq \langle k' \rangle$, it follows that $r_G(\mathbf{a}, \partial\mathbf{b})$ under G is greater than $r_{G'}(\mathbf{a}, \partial\mathbf{b})$ under G'. Since $\mathscr{P}$ is identical in both cases, $\mathbb{E}_G[T(\mathbf{b})|\mathbf{x}_0 = \mathbf{a}] \geq \mathbb{E}_{G'}[T(\mathbf{b})|\mathbf{x}_0 = \mathbf{a}]$.

(ii). When $\eta_{ab} < \frac{1}{2}$, all players for whom $\lceil \eta_{ba} k_i \rceil \geq k_i$ require all their neighbors to switch to a for them to do likewise. Denote by $\mathcal{N}_{ba}(G)$ for the subset of players for whom $\lceil \eta_{ba} k_i \rceil \geq k_i$ given the network G. That is

$$\mathcal{N}_{ba}(G) = \{i \in \mathcal{N} : \lceil \eta_{ba} k_i \rceil \geq k_i\}.$$

Let $\bar{k}_a$ be an integer such that all i for whom $k_i \leq \bar{k}_a$ belong to $\mathcal{N}_{ba}(G)$. Write $N_{ba}(G)$ for the cardinality of $\mathcal{N}_{ba}(G)$, that is $N_{ba}(G) = N \sum_{k=0}^{\bar{k}_a} p(k)$. For a given network, hence $\bar{k}_a$, $N_{ba}(G)$ is a non-increasing function of average degree $\langle k \rangle$. That is $N_{ba}(G) := f(\langle k \rangle)$ where f is non-increasing in $\langle k \rangle$. For P FOSD (or SOSD) P', $N_{ba}(G) \leq N_{ba}(G')$, and since $r_G(\mathbf{b}, \partial\mathbf{a})$ is an increasing function of $N_{ba}(G)$ then

$$r_G(\mathbf{b}, \partial\mathbf{a}) \leq r_{G'}(\mathbf{b}, \partial\mathbf{a}).$$

It then follows that for a given $\mathscr{P}$ hence β, $\mathbb{E}_G[T(\mathbf{a})|\mathbf{x}_0 = \mathbf{b}] \leq \mathbb{E}_{G'}[T(\mathbf{a})|\mathbf{x}_0 = \mathbf{b}]$. Note also that for a given network say G, $\mathbb{E}_G[T(\mathbf{a})|\mathbf{x}_0 = \mathbf{b}] \geq \mathbb{E}_G[T(\mathbf{b})|\mathbf{x}_0 = \mathbf{a}]$ whenever $\eta_{ab} < \frac{1}{2}$. Such that the maximum expected waiting time for G and G' are $\mathbb{E}_G[T(\mathbf{a})|\mathbf{x}_0 = \mathbf{b}]$ and $\mathbb{E}_{G'}[T(\mathbf{a})|\mathbf{x}_0 = \mathbf{b}]$ respectively. Consequently $\mathscr{T}_G \leq \mathscr{T}_{G'}$.

We then need to establish the relationships for the minimum and maximum values of the stationary distributions. First note that since $\tilde{\mathbf{b}}$ is the long-run stable subset and hence $\mathbf{b}$ is the long-run stable state, $\frac{1}{\tilde{\pi}_{\max}} \equiv \frac{1}{\tilde{\pi}(\tilde{b})} \leq K_\pi$ for both networks. We then have that for each network,

$$\frac{1}{\tilde{\pi}_{\min}} \equiv \frac{1}{\tilde{\pi}(\tilde{a})} \leq K_\pi \exp\{\varphi_G(\tilde{\mathbf{a}})\}$$

where

$$\begin{aligned}\varphi_G(\tilde{\mathbf{a}}) &= \Big[Nc_G(\mathbf{b}, \partial\mathbf{a}) + NF(r_G(\mathbf{b}, \partial\mathbf{a})) - \ln(1-\lambda_2(G))\Big] - \Big[Nc_G(\mathbf{a}, \partial\mathbf{b}) + NF(r_G(\mathbf{a}, \partial\mathbf{b}))\\ &\quad - K_{\mathbf{a}} \ln(1-\lambda_2(G))\Big]\\ &= N\Big[r_G(\mathbf{b}, \partial\mathbf{a}) - r_G(\mathbf{a}, \partial\mathbf{b})\Big](\beta - \ln 2^{-1}) + N\Big[F(r_G(\mathbf{b}, \partial\mathbf{a})) - F(r_G(\mathbf{a}, \partial\mathbf{b}))\Big]\\ &\quad - \Big[(1-K_{\mathbf{a}})\ln(1-\lambda_2(G))\Big]\end{aligned}$$

where we have again taken $K_{\mathbf{b}} = 1$, and $K_{\mathbf{a}} < 1$. Similarly,

$$\begin{aligned}\varphi_{G'}(\tilde{\mathbf{a}}) &= N\Big[r_{G'}(\mathbf{b}, \partial\mathbf{a}) - r_{G'}(\mathbf{a}, \partial\mathbf{b})\Big](\beta - \ln 2^{-1}) + N\Big[F(r_{G'}(\mathbf{b}, \partial\mathbf{a})) - F(r_{G'}(\mathbf{a}, \partial\mathbf{b}))\Big]\\ &\quad - \Big[(1-K_{\mathbf{a}})\ln(1-\lambda_2(G'))\Big]\end{aligned}$$

Since $r_G(\mathbf{b}, \partial\mathbf{a}) \leq r_{G'}(\mathbf{b}, \partial\mathbf{a})$ and $r_G(\mathbf{a}, \partial\mathbf{b}) \geq r_{G'}(\mathbf{a}, \partial\mathbf{b})$, it follows that

$$r_G(\mathbf{b}, \partial\mathbf{a}) - r_G(\mathbf{a}, \partial\mathbf{b}) \leq r_{G'}(\mathbf{b}, \partial\mathbf{a}) - r_{G'}(\mathbf{a}, \partial\mathbf{b})$$

Similarly, since P MPS P' or P' FOSD P implies that $(1-\lambda_2(G')) \leq (1-\lambda_2(G))$ and hence $-\ln(1-\lambda_2(G')) \geq -\ln(1-\lambda_2(G))$. It then follows that $\varphi_G(\tilde{\mathbf{a}}) \leq \varphi_{G'}(\tilde{\mathbf{a}})$, consequently $T_c(G) \leq T_c(G')$

□

The result in Proposition 4 supplements that for deterministic networks above. That is, the expected waiting times to the long-run stable set is shorter for sparsely connected than highly connected networks. Similarly, the convergence time is longer for sparsely connected than highly connected networks. The method of characterization is however complementary. Several authors have characterized the effect of network topologies on individual behavior and payoff in terms of degree distributions and using stochastic dominance relations (e.g Jackson and Yariv (2007)). Proposition 4 provides similar characteristic analysis but for convergence rates, such that once the degree distributions of any two networks is known, then a comparative analysis of convergence rates can be performed.

7. Step-by-step evolution

When global contagion is not feasible, that is when $\eta_{ab} > \eta^*$ for any pair of actions $a \neq b$, the network enhances the number of closed communication classes. Take for example the case of 2-dimensions nearest neighbor interactions network, if both $\eta_{ab} > \eta^*$ and $\eta_{ba} > \eta^*$ then global

contagion is not feasible. Under this condition, if $\eta_{ab} < \eta_{ba}$ then all states in which one or more enclaves of four players all play b become closed communication classes. To get from an all a state to an all b state, the chain goes through these intermediate closed communication classes induced by the network topology. It may seem at first that the presence of such intermediate states will lead to slow convergence rates when compared to making direct transition say from $\mathbf{a} \to \mathbf{b}$. On the contrary however, these intermediate states do speed-up evolution between closed communication classes induced by the payoff structure. This phenomenon was referred to as *step-by-step evolution* by Ellison (2000). Below, we use a specific example to illustrate the convergence rates in step-by-step evolution.

We consider a set of eight agents whose interactions are governed by the network topology in Figure 6, and playing a pure coordination game in Table 1. The network can be divided into two cohesive subgroups $g_1 = \{1, 2, 3, 4, 5\}$ and $g_2 = \{6, 7, 8, 9, 10\}$. A player i belongs to a given cohesive subgroup if and only if at least half of i's interactions are with members of that subgroup. The contagion threshold for this network is $\frac{1}{3}$. That is if $\eta_{ab} \leq \frac{1}{3}$ then starting from a, global contagion to be occurs when at least one player (from any of the subgroups) plays b Hence for $\eta_{ab} \leq \frac{1}{3}$, $r(\mathbf{a}, \partial\mathbf{b}) = \frac{1}{10}$. When $\frac{1}{3} < \eta_{ab} < \frac{1}{2}$, global contagion is no longer feasible, though local contagion (within cohesive subgroups) can occur. We use this situation to demonstrate the rates of convergence under step-by-step evolution. For $\frac{1}{3} < \eta_{ab} < \frac{1}{2}$, there are two additional closed communication classes to $\mathbf{a}$ and $\mathbf{b}$; $\mathbf{a}_1$ and $\mathbf{a}_2$ in which all members of g_1 play a while those in g_2 all play b and vise versa for $\mathbf{a}_2$.

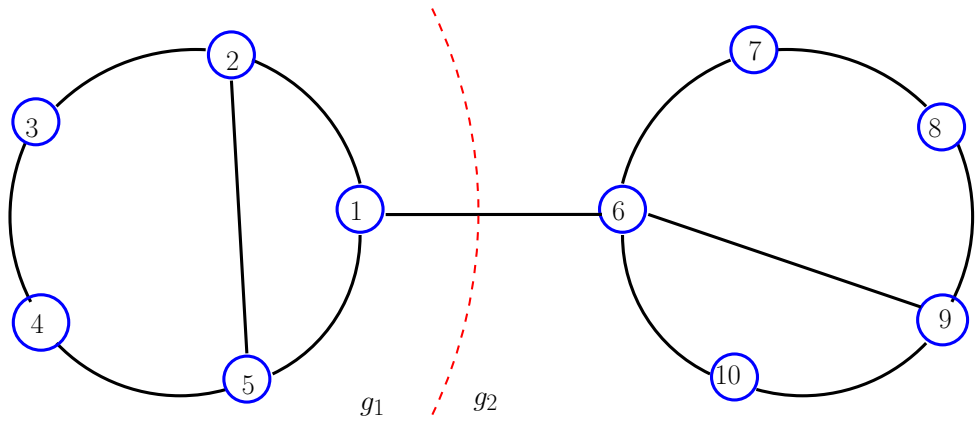

Figure 6: A general network with two cohesive subgroups

COROLLARY 3: *Let the underlying game be* 2×2 *matrix game of Table 1 with* $\frac{1}{3} < \eta_{ab} < \frac{1}{2}$. *Let the updating rule be the uniform error probabilities, and the topology of interactions be that in Figure 6. Then*

(i) the expected waiting times are;

$$\mathbb{E}[T(\mathbf{b})|\mathbf{x}_0 = \mathbf{a}] \leq K_T \exp\{2\beta - 3.6\} \tag{44}$$

$$\mathbb{E}[T(\mathbf{a})|\mathbf{x}_0 = \mathbf{b}] \leq K_T \exp\{4(\beta - 1)\} \tag{45}$$

(ii) the convergence time is given by

$$T_c \leq \kappa_1 \left(\kappa_2 + \varphi(\tilde{\mathbf{a}})\right) \exp\{4(\beta - 1)\}, \tag{46}$$

where $\varphi(\tilde{\mathbf{a}}) = 4\beta - 0.8$.

Proof. For $\frac{1}{3} < \eta_{ab} < \frac{1}{2}$, the normalized radii are as follows: $r(\mathbf{a}, \partial\mathbf{b}) = \frac{4}{10}$ (two players from g_1 and g_2 play b), $r(\mathbf{b}, \partial\mathbf{a}) = \frac{8}{10}$ (four players from g_1 and g_2 play b), $r(\mathbf{a}, \partial\mathbf{a}_1) = \frac{2}{10} = r(\mathbf{a}, \partial\mathbf{a}_2)$, $r(\mathbf{b}, \partial\mathbf{a}_1) = \frac{4}{10} = r(\mathbf{b}, \partial\mathbf{a}_2)$, $r(\mathbf{a}_1, \partial\mathbf{a}) = \frac{4}{10} = r(\mathbf{a}_2, \partial\mathbf{a})$, $r(\mathbf{a}_1, \partial\mathbf{b}) = \frac{2}{10}$, $r(\mathbf{a}_2, \partial\mathbf{b}) = \frac{1}{10}$, $r(\mathbf{a}_1, \partial\mathbf{a}_2) = \frac{6}{10}$ and $r(\mathbf{a}_2, \partial\mathbf{a}_1) = \frac{5}{10}$. Figure 7 plots the quantity $c(\Omega, \partial\Omega') + F(r(\Omega, \partial\Omega'))$ for $\beta = 5$

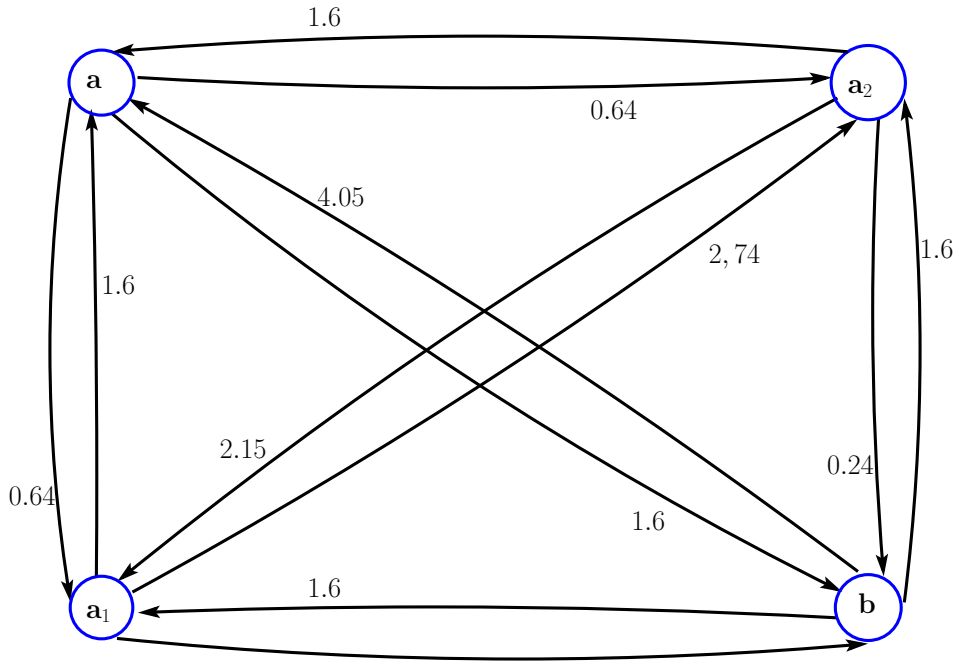

Figure 7: Each arc's weight is the quantity $c(\Omega, \partial\Omega') + F(r(\Omega, \partial\Omega'))$ for $\beta = 5$.

(i). The expected waiting time from $\mathbf{a}$ to $\mathbf{b}$ is equivalent to letting $W = \{\mathbf{a}, \mathbf{a}_1, \mathbf{a}_2\}$ and $W^c = \{\mathbf{b}\}$, then computing $\mathbb{E}[T(W^c)|\mathbf{x}_0 = \mathbf{a}]$. The least cost tree from $\mathbf{a}$ to W^c is $\{\mathbf{a} \to \mathbf{a}_2 \to \mathbf{b}, \mathbf{a}_1 \to \mathbf{b}\}$ and $\{\mathbf{a}_1 \to \mathbf{b}, \mathbf{a}_2 \to \mathbf{b}\}$ minimizes $M_W(\mathbf{a})$. Such that M_w and $M_W(\mathbf{a})$ are respectively

$$M_W = [c(\mathbf{a}, \partial\mathbf{a}_2) + F(r(\mathbf{a}, \partial\mathbf{a}_2))] + [c(\mathbf{a}_2, \partial\mathbf{b}) + F(r(\mathbf{a}_2, \partial\mathbf{b}))] + [c(\mathbf{a}_1, \partial\mathbf{b}) + F(r(\mathbf{a}_1, \partial\mathbf{b}))]$$

$$M_W(\mathbf{a}) = [c(\mathbf{a}_1, \partial\mathbf{b}) + F(r(\mathbf{a}_1, \partial\mathbf{b}))] + [c(\mathbf{a}_2, \partial\mathbf{b}) + F(r(\mathbf{a}_2, \partial\mathbf{b}))]$$

Hence,

$$H_W(\mathbf{a}) = M_W - M_W(\mathbf{a}) = c(\mathbf{a}, \partial\mathbf{a}_2) + F(r(\mathbf{a}, \partial\mathbf{a}_2)) = \frac{1}{5}\beta - 0.36$$

Implying that $\mathbb{E}[T(W^c)|\mathbf{x}_0 = \mathbf{a}] \leq K_T \exp\left\{10\left(\frac{1}{5}\beta - 0.36\right)\right\} = K_T \exp\{(2\beta - 3.6)\}$. Where the factor of 10 is the number of agents.

The expected waiting from $\mathbf{b}$ to $\mathbf{a}$ is equivalent to letting $W = \{\mathbf{b}, \mathbf{a}_1, \mathbf{a}_2\}$ and $W^c = \{\mathbf{a}\}$, then computing $\mathbb{E}[T(W^c)|\mathbf{x}_0 = \mathbf{b}]$. The least cost tree from $\mathbf{b}$ to W^c is $\{\mathbf{a}_2 \to \mathbf{b} \to \mathbf{a}_1 \to \mathbf{a}\}$ and $\{\mathbf{a}_2 \to \mathbf{b} \to \mathbf{a}_1\}$ minimizes $M_W(\mathbf{b})$. Such that M_w and $M_W(\mathbf{b})$ are respectively

$$M_W = [c(\mathbf{a}_2, \partial\mathbf{b}) + F(r(\mathbf{a}_2, \partial\mathbf{b}))] + [c(\mathbf{b}, \partial\mathbf{a}_1) + F(r(\mathbf{b}, \partial\mathbf{a}_1))] + [c(\mathbf{a}_1, \partial\mathbf{a}) + F(r(\mathbf{a}_1, \partial\mathbf{a}))]$$

$$M_W(\mathbf{a}) = [c(\mathbf{a}_2, \partial\mathbf{b}) + F(r(\mathbf{a}_2, \partial\mathbf{b}))] + [c(\mathbf{b}, \partial\mathbf{a}_1) + F(r(\mathbf{b}, \partial\mathbf{a}_1))]$$

Hence,

$$H_W(\mathbf{a}) = M_W - M_W(\mathbf{a}) = c(\mathbf{a}_1, \partial\mathbf{a}) + F(r(\mathbf{a}_1, \partial\mathbf{a})) = \frac{2}{5}\beta - 0.36$$

Implying that $\mathbb{E}[T(W^c)|\mathbf{x}_0 = \mathbf{b}] \leq K_T \exp\left\{10\left(\frac{2}{5}\beta - 0.4\right)\right\} = K_T \exp\{4\,(\beta - 1)\}$.

(ii). The maximum expected waiting time is for the transition from $\mathbf{b}$ to $\mathbf{a}$, hence $\mathscr{T} \equiv \mathbb{E}[T(W^c)|\mathbf{x}_0 = \mathbf{b}] \leq K_T \exp\{4\,(\beta - 1)\}$. Since $\mathbf{b}$ is the long-run stable state, then $\frac{1}{\tilde{\pi}_{\max}} \equiv \frac{1}{\tilde{\pi}(\tilde{b})} \leq K_\pi$. Also, $\{\mathbf{a}_2 \to \mathbf{b} \to \mathbf{a}_1 \to \mathbf{a}\}$ is the least cost tree for $G(\{\mathbf{a}\})$ and $\{\mathbf{a} \to \mathbf{a}_2 \to \mathbf{b}, \mathbf{a}_1 \to \mathbf{b}\}$ for $G(\{\mathbf{b}\})$. Such that

$$\frac{1}{\tilde{\pi}_{\min}} \equiv \frac{1}{\tilde{\pi}(\tilde{a})} \leq K_\pi \exp\{\varphi_G(\tilde{\mathbf{a}})\}$$

where

$$\begin{aligned}
\varphi_G(\tilde{\mathbf{a}}) &= 10\Big\{[c(\mathbf{a}_2, \partial\mathbf{b}) + F(r(\mathbf{a}_2, \partial\mathbf{b}))] + [c(\mathbf{b}, \partial\mathbf{a}_1) + F(r(\mathbf{b}, \partial\mathbf{a}_1))] + [c(\mathbf{a}_1, \partial\mathbf{a}) + F(r(\mathbf{a}_1, \partial\mathbf{a}))] \\
&\quad - [c(\mathbf{a}, \partial\mathbf{a}_2) + F(r(\mathbf{a}, \partial\mathbf{a}_2))] + [c(\mathbf{a}_2, \partial\mathbf{b}) + F(r(\mathbf{a}_2, \partial\mathbf{b}))] + [c(\mathbf{a}_1, \partial\mathbf{b}) + F(r(\mathbf{a}_1, \partial\mathbf{b}))]\Big\} \\
&= 10\left\{\left[\frac{4}{5}\beta - 0.8\right] - \left[\frac{2}{5}\beta - 0.72\right]\right\} \\
&= 10\left\{\frac{2}{5}\beta - 0.08\right\} \qquad (47)
\end{aligned}$$

□

Corollary 3 underscores the role played by intermediate closed communication sets induced by the topology of the network in increasing convergence rates. If the chain is to be restricted to the direct transition from $\mathbf{a}$ to $\mathbf{b}$ rather than through intermediate closed communication sets, then the expected waiting time would be

$$\mathbb{E}[T(\mathbf{b})|\mathbf{x}_0 = \mathbf{a}] \leq K_T \exp\left\{10\left(\frac{4}{10}\left(\beta - \ln 2^{-1}\right) + F\left(\frac{4}{10}\right)\right)\right\} = K_T \exp\{4(\beta - 1)\}$$

which is greater than the expected waiting time in (44) for the case with intermediate limit states. A similar argument follows for $\mathbb{E}[T(\mathbf{a})|\mathbf{x}_0 = \mathbf{b}]$.

In comparison to the case in which the 10 players interact globally, from (43) we have

$$T_c \leq \kappa_1\left(\kappa_2 + 9(1 - 2\eta_{ab})\left(\beta - \ln 2^{-1}\right)\right) \exp\left\{10\left(\eta_{ba}(\beta - \ln 2^{-1}) + F(\eta_{ba})\right)\right\}, \qquad (48)$$

Such that for $\frac{1}{3} < \eta_{ab} < \frac{1}{2}$, which is equivalent to $\frac{1}{2} < \eta_{ba} < \frac{2}{3}$, T_c in (48) is greater than T_c in (46) for β sufficiently large. Unlike in the case of local interactions in which global contagion is feasible above, for local interactions with step-by-step evolution the convergence time is shorter than under global interactions. Clearly, the presence of cohesive subgroups reduces the convergence time and this result generalizes to all networks with cohesive subgroups. The intuition behind the result is that the presence of intermediate closed communication classes reduces the resistance (sizes of basins of attraction) hence the minimum exit time for all closed communication classes involved, including the long-run stable set. This effectively reduces the expected waiting time between any pair of closed communication classes, hence the time it takes the chain to mix.

8. Concluding remarks

This paper studies extensively the convergence rates for stochastic evolution in networks. We identify the payoff gains, noise level and network topology to be the key factors. Unlike previous analysis in the literature that focused on limit noise, we place emphasis on keeping the noise levels positive. Analysis that relies on limit noise fails to capture some aspects of evolutionary dynamics in networks. The network topology and payoff gains interactively determine whether or not an action is globally contagious. Globally contagious action are those with the highest payoff gains, hence are also those that are played in the long-run.

Several insightful results derive from our analysis. First, for given payoff gains, the more sparsely connected a network is the more likely that an action with the highest payoff gain is globally contagious. When global contagion is feasible, the expected waiting times from any other closed communication class to the long-run stable closed communication class are independent of the population size. For a given family of networks, global contagion also implies that the expected waiting time for highly connected network is shorter than for sparsely connected networks. Secondly, Whenever global contagion is feasible, sparsely connected networks tend to have longer convergence time than highly connected networks, longer than even the case of global interactions. Finally, if the payoff gains and network topology induce multiple intermediate closed communication classes, that is when global contagion is infeasible, then the convergence time is faster than both under global interactions and other interaction networks in which global contagion is feasible. Possible extensions to this paper include generalizations to other models of mistakes, beyond the uniform mistakes model we have considered.

A. Appendix

A.1. *Proof of Lemma 2*

Given that the chain is in some state $\mathbf{x}$ of Ω, let a be the action that should be played by at least $R(\Omega, \partial\Omega')$ players for the transition $\Omega \to \Omega$ to occur. If we let I_i be a binary random variable taking on values one with probability $\mathscr{P}$ and zero otherwise, then

$$\mathbb{P}(\Omega, \partial\Omega') = \mathbb{P}\left(\sum_{i=1}^{N} I_i = R(\Omega, \partial\Omega')\right). \tag{A.1}$$

The expression for the right hand side of (A.1) can be established using techniques for concentration bounds. From Binomial probabilities, we have (we write R for the short form of $R(\Omega, \partial\Omega')$)

$$\mathbb{P}\left(\sum_{i=1}^{N} I_i = R(\Omega, \partial\Omega')\right) = \frac{N!}{R!(N-R)!}\mathscr{P}^R(1-\mathscr{P})^{N-R} \tag{A.2}$$

Using Sterling's approximation we have

$$\begin{aligned}
\frac{N!}{R!(N-R)!} &\approx \frac{N^N\sqrt{2\pi N}e^{-N}}{R^R\sqrt{2\pi R}e^{-R}(N-R)^{N-R}\sqrt{2\pi(N-R)}e^{-(N-R)}} \\
&= \frac{N^N}{R^R(N-R)^{N-R}}\sqrt{\frac{N}{2\pi R(N-R)}} \\
&= \sqrt{\frac{N}{2\pi R(N-R)}}\exp\left(N\ln N - R\ln R - (N-R)\ln(N-R)\right) \\
&= \sqrt{\frac{1}{2\pi r(1-r)N}}\exp\left\{-N\left(r\ln r - (1-r)\ln(1-r)\right)\right\}
\end{aligned}$$

We let $K_N = \frac{1}{\sqrt{2\pi r(1-r)N}}$, and by rewriting $\mathscr{P}^R(1-\mathscr{P})^{N-R} = \exp\left\{N\left(r\ln\mathscr{P} + (1-r)\ln(1-\mathscr{P})\right)\right\}$, we then have

$$\begin{aligned}
\mathbb{P}(\Omega, \partial\Omega') &\approx K_N \exp\left\{N\left(r\ln\mathscr{P} - r\ln r + (1-r)\frac{\ln(1-\mathscr{P})}{\ln(1-r)}\right)\right\} \\
&\approx K_N \exp\left\{N\left(r\ln\mathscr{P} - r\ln r - (1-r)\ln(1-r)\right)\right\} \\
&= K_N \exp\left\{-N\left(c(\Omega, \partial\Omega') + F(r(\Omega, \partial\Omega'))\right)\right\}
\end{aligned}$$

where $c(\Omega, \partial\Omega') = -r(\Omega, \partial\Omega')\ln\mathscr{P}$ and $F(r) = r\ln r - (1-r)\ln(1-r)$.

A.2. Proof of Proposition 1

We begin by proving the following lemma.

LEMMA 3: *Let L be the number of closed communication classes and μ_j the jth eigenvalue of P_ε. Then the contagion rate within any basin of attraction Ω_l has the following lower and upper bounds*

$$1 - \mu_{L+1} \leq \mathcal{R}(\partial\Omega_l, \Omega_l) \leq 1 - \mu_{L+L}$$

Proof. The proof makes use of the spectral properties and near-complete decomposability of transition matrix P_ε. Under chain $(\mathbf{X}, P)$, the transition matrix is completely decomposable into the form

$$P = \begin{pmatrix} M_1^* & & & & \\ & \ddots & & & \\ & & M_l^* & & \\ & & & \ddots & \\ & & & & M_L^* \end{pmatrix}$$

where M_l^* for $l = 1, \cdots, L$ is a block matrix describing the transitions within each basin of attraction under $(\mathbf{X}, P)$. The rest of the undisplayed elements are zeros and L is the number of closed communication classes. All leading eigenvalues of the block matrices are ones. The transition

matrix P_ε on the other hand is near-completely decomposable into L "loosely" connected block matrices that we denote by M_l for $l = 1, \cdots, L$. Such that $P_\varepsilon = P + \epsilon P^*$, where ϵ is a small real number and P^* is an arbitrary $\#\mathbf{X}$ by $\#\mathbf{X}$ matrix. A more detailed exposition on the notion of near-complete decomposability can be found in Simon and Ando (1961). For ϵ small enough, the leading eigenvalues of the diagonal block matrices of P_ε are close to one.

Let μ_{i_l} denote the ith eigenvalue of the lth diagonal block matrix, such that $(\mu_{1_1}, \mu_{1_2}, \cdots, \mu_{1_L})$ are the largest eigenvalues in blocks 1 to L, and $(\mu_{2_1}, \mu_{2_2}, \cdots, \mu_{2_L})$ are the respective second largest eigenvalues. Index by n_l as the number of columns in diagonal block l such that the eigenvalue spectrum $\rho(P_\varepsilon)$ of P_ε can be written as $\rho(P_\varepsilon) = \left(\mu_{1_1}, \mu_{2_1}, \cdots, \mu_{n_{1_1}}, \cdots, \mu_{1_2}, \cdots, \mu_{1_l}, \cdots, \mu_{n_{l_l}}, \cdots, \mu_{1_L}, \cdots, \mu_{n_{L_L}}\right)$. The spectral decomposition of $(\mathbf{X}, P_\varepsilon)$ is then given by

$$\begin{aligned}\mathbf{q}_0 P_\varepsilon^t = \mathbf{q}_0 \mathbf{r}_{1_1} \mathbf{z}_{1_1}^T + \sum_{j=2}^{n_1} \mu_{j_1}^t \mathbf{q}_0 \mathbf{r}_{j_1} \mathbf{z}_{j_1}^T + \mu_{1_2}^t \mathbf{q}_0 \mathbf{r}_{1_2} \mathbf{z}_{1_2}^T + \sum_{j=2}^{n_2} \mu_{j_2}^t \mathbf{q}_0 \mathbf{r}_{j_2} \mathbf{z}_{j_2}^T \\ + \cdots + \mu_{1_L}^t \mathbf{q}_0 \mathbf{r}_{1_L} \mathbf{z}_{1_L}^T + \sum_{j=2}^{n_L} \mu_{j_L}^t \mathbf{q}_0 \mathbf{r}_{j_L} \mathbf{z}_{j_L}^T \end{aligned} \tag{A.3}$$

where $\mathbf{z}^T$ is the transpose of $\mathbf{z}$, and $\mathbf{r}_{j_l}$ and $\mathbf{z}_{j_l}$ are the right and left eigenvectors of μ_{j_l}.

Let $\mathbf{x}_l$ be the initial state of $(\mathbf{X}, P_\varepsilon)$ in $\tilde{\Omega}_l$ and $\mathbf{q}_{\tilde{\Omega}_l}$ be the $\#\mathbf{X}$-dimensional vector of zeros except a one at the point corresponding to the state $\mathbf{x}_l$. Let t_l be the period at which $(\mathbf{X}, P_\varepsilon)$ is in the state $\mathbf{x}_l$ and $T_{\tilde{\Omega}_l}$ the period at which it exits $\tilde{\Omega}_l$ (or equivalently the period at which it attains the quasi-stationary distribution ν_l). Then for $t_l \leq t \leq T_{\tilde{\Omega}_l}$ and all $l = 1, \cdots, L$,

$$\begin{aligned}\mathbf{q}_t &= \mu_{1_l}^t \mathbf{q}_{\tilde{\Omega}_l} \mathbf{r}_{1_l} \mathbf{z}_{1_l}^T + \sum_{j=2}^{n_l} \mu_{j_l}^t \mathbf{q}_{\tilde{\Omega}_l} \mathbf{r}_{j_l} \mathbf{z}_{j_l}^T \\ \nu_l &= \lim_{t \to T_{\tilde{\Omega}_l}} \mathbf{q}_t = \mu_{1_l}^{T_{\tilde{\Omega}_l}} \mathbf{q}_{\tilde{\Omega}_l} \mathbf{r}_{1_l} \mathbf{z}_{1_l}^T \approx \mu_{1_l}^t \mathbf{q}_{\tilde{\Omega}_l} \mathbf{r}_{1_l} \mathbf{z}_{1_l}^T\end{aligned}$$

where the approximation holds from the fact that μ_{1_l} is close to one for all l. It then follows that

$$\limsup_{t \to T_{\tilde{\Omega}_l}} \left|\left| P_\varepsilon^t \mathbf{q}_{\tilde{\Omega}_l} - \nu_{\tilde{\Omega}_l} \right|\right|^{\frac{1}{t}} = |\mu_{2_l}| \limsup_{t \to T_{\tilde{\Omega}_l}} \left|\left| \mathbf{q}_{\tilde{\Omega}_l} \mathbf{r}_{2_l} \mathbf{z}_{2_l}^T + \sum_{j=3}^{n_l} \left(\frac{\mu_{j_l}}{\mu_{2_l}}\right)^t \mathbf{q}_{\tilde{\Omega}_l} \mathbf{r}_{j_l} \mathbf{z}_{j_l}^T \right|\right|^{\frac{1}{t}} = |\mu_{2_l}|$$

Implying that $\mathcal{R}(\partial\Omega_l, \Omega_l) = |\mu_{2_l}|$ Since μ_j's are arranged in ascending order we then have that $\max_l \mu_{2_l} = |\mu_{L+1}|$ and $\min_l \mu_{2_l} = |\mu_{L+L}|$. Such that for any Ω_l,

$$1 - \mu_{L+1} \leq \mathcal{R}(\partial\Omega_l, \Omega_l) \leq 1 - \mu_{L+L}$$

□

To prove the second part of the theorem, we consider the linearization of P_ε of the form

$$\mathbf{q}_t \Psi = \mathbf{q}_0 P_\varepsilon^t \Psi = \mathbf{q}_0 \Psi \Pi_\varepsilon^t \tag{A.4}$$

where Ψ is the event matrix derived by stacking into rows all possible realizations of states of $(\mathbf{X}, P_\varepsilon)$ written in the *basis vector* form. The choice basis vector for each player $i \in N$ is a vector of zeros except a one in a position corresponding to the action i is playing. For example for a binary action

set $X = \{A, B\}$, a vector $(1, 0)$ implies that i is playing action A and $(0, 1)$ implies that i is playing action B. In the case of two players and binary action set, there are four possible realization such that

$$\Psi = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} \tag{A.5}$$

where the first row corresponds to the state in which both players play action A, and so forth. Then Π_ε is an $Nm \times Nm$ matrix defined by $\Pi_\varepsilon = \mathscr{A}^T \otimes \Sigma$, where $\otimes$ is a Kronecker product, $\mathscr{A}^T$ is the transpose of the normalized adjacency matrix $\mathscr{A}$ and Σ is the action-transition matrix defined in (10). A detailed exposition on the validity of (A.4) can be found in Asavathiratham (2001, Chapter 5). The following lemma follows directly from (A.4) and the definition of Π_ε above.

LEMMA 4: *Let $\rho(\Pi_\varepsilon) = \tilde{\mu}_1, \cdots, \tilde{\mu}_{nm}$, $\rho(\mathscr{A}) = (\lambda_1, \cdots, \lambda_N)$ and $\rho(\Sigma) = (\vartheta_1, \cdots, \vartheta_m)$ denote the eigenvalue spectra of Π_ε, $\mathscr{A}$ and Σ respectively.*

(a) If μ_1 and $\tilde{\mu}_1$ are the unique largest eigenvalues of P_ε and Π_ε respectively, then $\mu_1 = \tilde{\mu}_1 = 1$.

(b) $\rho(\Pi_\varepsilon) = (\vartheta_i \lambda_j)$ $\forall \vartheta_i \in \rho(\Sigma)$, $\lambda_j \in \rho(\mathscr{A})$.

Proof. Multiplying (A.4) by the right eigenvector $\mathbf{r}_i$ of P_ε, we have $P_\varepsilon \Psi \mathbf{r}_1 = \Psi \Pi_\varepsilon \mathbf{r}_1$. Since P_ε is a stochastic matrix, $\mu_1 = 1$, which implies that $P_\varepsilon \Psi \mathbf{r} = \Psi \mathbf{r}_1$, which is true if and only if $\Pi_\varepsilon \mathbf{r}_1 = \mathbf{r}_1$. That is $\tilde{\mu}_1 = \mu_1 = 1$. For the proof of Lemma 4 (b) see Horn and Johnson (1990, page 245, Theorem 4.2.12). □

For sufficiently small noise, $|\mu_{L+1}| = |\tilde{\mu}_{L+1}| = \lambda_2 \vartheta_1 = \lambda_2$ and $|\mu_{L+L}| = |\tilde{\mu}_{L+L}| = \lambda_2 \vartheta_m$. This completes the proof.

A.3. Proofs for Example 6

Let $G = (\mathcal{N}, E)$ be a graph or network of N vertices. Denote by $\mathcal{S}$ a subset of $\mathcal{N}$ and S its cardinality. Let $e(S, N - S)$ be the number of interactions (for a weighted network graph its is the sum of weighted interactions) between $\mathcal{S}$ and its complement $\mathcal{N} - \mathcal{S}$. Also let $d(\mathcal{S})$ denote the total degree of $\mathcal{S}$. Then the conductance of G

$$\phi(G) = \min_{\mathcal{S}, S \leq \frac{N}{2}} \frac{e(S, N - S)}{d(\mathcal{S})} \tag{A.6}$$

For regular network graphs (in which all vertices have the same degree), it is shown by (Alon and Milman, 1985) that

$$\lambda_2(G) \leq 1 - \frac{\phi(G)^2}{2}. \tag{A.7}$$

For a complete graph, since every vertex is connected to every other vertex, we have that every vertex in $\mathcal{S}$ is connected to all other vertices in $\mathcal{N} - \mathcal{S}$. This implies that $e(S, N - S) = S \times \#(\mathcal{N} - \mathcal{S}) = S \times (N - S)$, and $d(\mathcal{S}) = N \times S$ such that

$$\phi(G_{com}) = \min_{\mathcal{S}, S \leq \frac{N}{2}} \frac{S \times (N - S)}{N \times S} \geq \frac{1}{2}, \tag{A.8}$$

where the last inequality follows from the fact that $S \leq \frac{N}{2}$. We thus have that

$$\lambda_2(G_{com}) \leq \frac{7}{8} \tag{A.9}$$

In the case of a $1-D$ cyclic network, $e(S, N-S) = 2$, and $d(\mathcal{S}) = 2 \times S$ such that

$$\phi(G_{cyc}) = \min_{\mathcal{S}, S \leq \frac{N}{2}} \frac{2}{2 \times S} \geq \frac{2}{N}. \tag{A.10}$$

Hence $\lambda_2(G_{cyc}) \leq \frac{N^2-2}{N^2}$.

$2D$ network: Let the composition of $\mathcal{S}$ be chosen in such a way that the peripheral vertices (vertices at the perimeter or boundary of $\mathcal{S}$) contain approximately one edge each connecting it to the set $\mathcal{N} - \mathcal{S}$. Since it is a 2-dimensional structure there should be approximately $\sqrt{S}$ vertices forming such a boundary. This implies that $e(S, N-S) \approx \sqrt{S}$, and $d(\mathcal{S}) = 4 \times S$ such that

$$\phi(G_{2D}) = \min_{\mathcal{S}, S \leq \frac{N^2}{2}} \frac{\sqrt{S}}{4 \times S} \geq \frac{\sqrt{2}}{4N}. \tag{A.11}$$

where the last inequality follows from the fact that $\sqrt{S} \leq \sqrt{\frac{N^2}{2}}$. It follows that

$$\lambda_2(G_{2D}) \leq \frac{16N^2 - 1}{16N^2}.$$

Random d-regular network: Since for each vertex the vertices to which it is connected are chosen at random, and that the maximum size of $\mathcal{S}$ is $\frac{N}{2}$, then a typical vertex in $\mathcal{S}$ is connected to approximately $\frac{d \times (N-S)}{N}$ other vertices in $\mathcal{N} - \mathcal{S}$ such that $e(S, N-S) \approx \frac{d \times S(N-S)}{N}$. We thus have

$$\phi(G_{d-r}) = \min_{\mathcal{S}, S \leq \frac{N}{2}} \frac{\frac{d \times S(N-S)}{N}}{d \times S} \geq 1. \tag{A.12}$$

$$\lambda_2(G_{d-r}) \leq \frac{7}{8} \tag{A.13}$$

For Newman's small world networks see Durrett (2006).

A.4. Proof of Theorem 1

The proof of Theorem 1 follows from Lemma 1 and (13). Given the expressions for $P_{\varepsilon}(\Omega, \Omega')$

$$P_{\varepsilon}(g) = \prod_{(\Omega,\Omega') \in g} P_{\varepsilon}(\Omega, \Omega') \approx K_1 \exp\left\{ \sum_{(\Omega,\Omega') \in g} \left(-N \left(c(\Omega, \partial\Omega') + F(r(\Omega, \partial\Omega')) \right) + \gamma(\Omega') \right) \right\} \tag{A.14}$$

where $K_1 = (K_N)^{\#g}$. For any $W \neq \emptyset$, any $\Omega \in W$,

$$\begin{aligned}
\mathbb{E}[T(W^c) |\, \mathbf{x}_0 \in \Omega] &= \left(\sum_{\Omega' \in W} \sum_{g \in G_{\Omega,\Omega'}(W \cup \{\Omega'\})} P_{\varepsilon}(g) \right) \left(\sum_{g \in G(W)} P_{\varepsilon}(g) \right)^{-1} \\
&\leq K_T \frac{\max_{\Omega \in W} \max_{g \in G_{\Omega,\Omega'}(W \cup \{\Omega'\})} P_{\varepsilon}(g)}{\min_{g \in G(W^c)} P_{\varepsilon}(g)} \\
&= K_T \exp\left\{ \min_{g \in G(W^c)} \psi(g) - \min_{\Omega' \in W} \min_{g \in G_{\Omega,\Omega'}(W \cup \{\Omega'\})} \psi(g) \right\}
\end{aligned}$$

where K_T is some constant;

$$K_T = \frac{(\#W)(\#G_{\Omega,\Omega'}(W \cup \{\Omega'\}))}{\#G(W^c)}.$$

A.5. Proof of Theorem 2

The proof bounds mixing times of Markov chains appear in various version in the literature, see for example Levin, Peres, and Wilmer (2009). Here, we are interested in the properties of the reduced chain $(\tilde{\mathbf{\Omega}}, \tilde{P}_\varepsilon)$ with stationary distribution $\tilde{\pi}$. The formal definition of $(\tilde{\mathbf{\Omega}}, \tilde{P}_\varepsilon)$ is as follows (a generalization of the collapsed Markov chain in Aldous and Fill (1994, Chapter 2)):

$$\tilde{P}_\varepsilon(\mathbf{x}, \mathbf{y}) = P_\varepsilon(\mathbf{x}, \mathbf{y}), \tag{A.15a}$$

$$\tilde{P}_\varepsilon(\mathbf{y}, \tilde{\Omega}) = \sum_{\mathbf{x} \in \tilde{\Omega}} P_\varepsilon(\mathbf{y}, \mathbf{x}), \tag{A.15b}$$

$$\tilde{P}_\varepsilon(\tilde{\Omega}, \mathbf{y}) = \frac{1}{\pi(\tilde{\Omega})} \sum_{\mathbf{x} \in \tilde{\Omega}} \pi(\mathbf{x}) P_\varepsilon(\mathbf{x}, \mathbf{y}), \tag{A.15c}$$

$$\tilde{P}_\varepsilon(\tilde{\Omega}, \tilde{\Omega}') = \frac{1}{\pi(\tilde{\Omega})} \sum_{\mathbf{x} \in \tilde{\Omega}} \sum_{\mathbf{y} \in \tilde{\Omega}'} \pi(\mathbf{x}) P_\varepsilon(\mathbf{x}, \mathbf{y}) \tag{A.15d}$$

The following lemma is an immediate consequence of the above definition of a collapsed Markov chain.

LEMMA 5: *Let π and $\tilde{\pi}$ be the stationary distributions of $(\mathbf{\Omega}, \tilde{P}_\varepsilon)$ and $(\tilde{\mathbf{\Omega}}, \tilde{P}_\varepsilon)$ respectively. Then for any $\tilde{\Omega} \in \tilde{\mathbf{\Omega}}$, $\tilde{\pi}(\tilde{\Omega}) = \pi(\tilde{\Omega}) = \sum_{\mathbf{x} \in \tilde{\Omega}} \pi(\mathbf{x})$*

Proof. Note that $\pi(\tilde{\Omega}) = \sum_{\mathbf{x} \in \tilde{\Omega}} \pi(\mathbf{x})$. Let $\#\mathbf{\Omega}$ and $\#\tilde{\mathbf{\Omega}}$ be the cardinalities of $\mathbf{\Omega}$ and $\tilde{\mathbf{\Omega}}$ respectively. Define an event matrix $\mathcal{E}$ as an $\#\mathbf{\Omega} \times \#\tilde{\mathbf{\Omega}}$ matrix whose entries take on a value one if a state $\mathbf{x} \in \mathbf{\Omega}$ belongs to $\tilde{\Omega} \in \tilde{\mathbf{\Omega}}$ and zero otherwise. Denote by $\mathcal{E}_{\tilde{\Omega}}$ for the $\tilde{\Omega}^{\text{th}}$ column of $\mathcal{E}$. It then follows that $\tilde{P}_\varepsilon \mathcal{E} = \mathcal{E} \tilde{P}_\varepsilon$, and that

$$\pi(\tilde{\Omega}) = \pi \mathcal{E}_{\tilde{\Omega}} \quad \forall\, \tilde{\Omega} \in \tilde{\mathbf{\Omega}}$$

Consequently, $\pi\mathcal{E} = \pi \tilde{P}_\varepsilon \mathcal{E} = \pi \mathcal{E} \tilde{P}_\varepsilon$. Implying that $\pi\mathcal{E}$ is the stationary distribution of $\tilde{P}_\varepsilon$, hence $\tilde{\pi} = \pi\mathcal{E}$. □

Since $\tilde{P}_\varepsilon$ is not symmetric, the first step of the proof is to transform $\tilde{P}_\varepsilon$ into a symmetric matrix. We can then exploit spectral properties of symmetric matrices. From reversibility of $\tilde{P}_\varepsilon$ and hence $\tilde{P}_\varepsilon$, that is $\tilde{\pi}(\tilde{\Omega}) \tilde{P}_\varepsilon(\tilde{\Omega}, \tilde{\Omega}') = \tilde{\pi}(\tilde{\Omega}') \tilde{P}_\varepsilon(\tilde{\Omega}', \tilde{\Omega})$

Recall that $\tilde{P}_\varepsilon$ is reversible. That is given the stationary distribution $\tilde{\pi}$,

$$\tilde{\pi}(\tilde{\Omega}) \tilde{P}_\varepsilon(\tilde{\Omega}, \tilde{\Omega}') = \tilde{\pi}(\tilde{\Omega}') \tilde{P}_\varepsilon(\tilde{\Omega}', \tilde{\Omega}) \qquad \forall \tilde{\Omega}, \tilde{\Omega}' \in \mathbf{\Omega} \tag{A.16}$$

We can define an equivalent symmetric matrix S such that

$$S(\tilde{\Omega}, \tilde{\Omega}') = \sqrt{\frac{\tilde{\pi}(\tilde{\Omega})}{\tilde{\pi}(\tilde{\Omega}')}} \tilde{P}_\varepsilon(\tilde{\Omega}, \tilde{\Omega}') \tag{A.17}$$

The reason for introducing S is to be able to exploit the properties of symmetric matrices, particularly the spectral theorem stating that, for symmetric matrices, there exists a set of orthonormal basis $\{\mathbf{u}_i\}_{i=1}^{|\mathbf{\Omega}|}$, such that $\mathbf{u}_i$ is an eigenfunction corresponding to the real eigenvalue $\tilde{\mu}_i$.

Now, let us denote the diagonal matrix with elements $\tilde{\pi}(\tilde{\Omega})$ by D, then we have

$$S = D^{\frac{1}{2}} \tilde{P}_\varepsilon D^{-\frac{1}{2}} \tag{A.18}$$

Let $f_i = D^{-\frac{1}{2}}\mathbf{u}_i$, (where $\mathbf{u}_i$'s are eigenfunctions of S) it follows that f_i is an eigenfunction of $\tilde{P}_\varepsilon$ corresponding to eigenvalue $\tilde{\mu}_i$, and that f_i's are orthonormal with respect to $\tilde{\pi}$. That is

$$\tilde{P}_\varepsilon f_i = \tilde{P}_\varepsilon D^{-\frac{1}{2}}\mathbf{u}_i = D^{-\frac{1}{2}}\left(D^{\frac{1}{2}}\tilde{P}_\varepsilon D^{-\frac{1}{2}}\right)\mathbf{u}_i = D^{-\frac{1}{2}} S\mathbf{u}_i = D^{-\frac{1}{2}}\tilde{\mu}_i\mathbf{u}_i = \tilde{\mu}_i f_i. \tag{A.19}$$

To prove orthonormality of the functions f_i's with respect to $\tilde{\pi}$, its useful to take note of the following definition of the *inner product*.

Let $\langle ., . \rangle$ denote the inner product on $\mathbb{R}^{\mathbf{\Omega}}$, that is

$$\langle f, g \rangle = \sum_{\tilde{\Omega}\in\mathbf{\Omega}} f(\tilde{\Omega})g(\tilde{\Omega}), \tag{A.20}$$

then we can define the inner product with respect to the distribution $\tilde{\pi}$ as

$$\langle f, g \rangle_{\tilde{\pi}} = \sum_{\tilde{\Omega}\in\mathbf{\Omega}} f(\tilde{\Omega})g(\tilde{\Omega})\tilde{\pi}(\tilde{\Omega}). \tag{A.21}$$

Let $\delta_{i,j}$ denote the Dirac delta function (that is $\delta_{i,j} = 1$ if and only if $i = j$), then

$$\delta_{i,j} = \langle \mathbf{u}_i, \mathbf{u}_j \rangle = \left\langle D^{\frac{1}{2}} f_i, D^{\frac{1}{2}} f_j \right\rangle = \sum_{\tilde{\Omega}\in\mathbf{\Omega}} \tilde{\pi}(\tilde{\Omega})^{\frac{1}{2}} f(\tilde{\Omega})\tilde{\pi}(\tilde{\Omega})^{\frac{1}{2}} g(\tilde{\Omega}) = \sum_{\tilde{\Omega}\in\mathbf{\Omega}} f(\tilde{\Omega})\tilde{\pi}(\tilde{\Omega})g(\tilde{\Omega}) = \langle f, g \rangle_{\tilde{\pi}}.$$

Note that $\tilde{P}_\varepsilon^t(\tilde{\Omega}, \tilde{\Omega}')$ is the $\tilde{\Omega}, \tilde{\Omega}'$ element of $\tilde{P}_\varepsilon^t$, implying that $\tilde{P}_\varepsilon^t(\tilde{\Omega}, \tilde{\Omega}') = (\tilde{P}_\varepsilon^t \delta_{\tilde{\Omega}'})(\tilde{\Omega})$; where $\delta_{\tilde{\Omega}'}(\tilde{\Omega})$ is a Dirac function assuming the value of unity for $\tilde{\Omega} = \tilde{\Omega}'$ and zero otherwise. Notice also that $\delta_{\tilde{\Omega}'}$ belongs to the inner product space $\mathbb{V} = (\mathbb{R}^{\mathbf{\Omega}}, \langle ., . \rangle_{\tilde{\pi}})$, and since the set $\{f_1, \cdots, f_{|\mathbf{\Omega}|}\}$ is an orthonormal basis of $\mathbb{V}$, then $\delta_{\tilde{\Omega}'}$ can be written via basis decomposition as

$$\delta_{\tilde{\Omega}'} = \sum_{i=1}^{\eta} \langle \delta_{\tilde{\Omega}'}, f_i \rangle_{\tilde{\pi}} f_i = \sum_{i=1}^{\eta} f_i(\tilde{\Omega}')\tilde{\pi}(\tilde{\Omega}')f_i \tag{A.22}$$

Substituting (A.22) and $\tilde{P}_\varepsilon^t f_i = \tilde{\mu}_i^t f_i$ gives $\tilde{P}_\varepsilon^t(\tilde{\Omega}, \tilde{\Omega}')$ as

$$\tilde{P}_\varepsilon^t(\tilde{\Omega}, \tilde{\Omega}') = \sum_{i=1}^{\eta} f_i(\tilde{\Omega}')\tilde{\pi}(\tilde{\Omega}')\tilde{\mu}_i^t f_i(\tilde{\Omega}) = \tilde{\pi}(\tilde{\Omega}') + \sum_{i=2}^{\eta} f_i(\tilde{\Omega}')\tilde{\pi}(\tilde{\Omega}')\tilde{\mu}_i^t f_i(\tilde{\Omega}) \tag{A.23}$$

Taking the absolute values of (A.23) yields

$$\begin{aligned} |\tilde{P}_\varepsilon^t(\tilde{\Omega}, \tilde{\Omega}') - \tilde{\pi}(\tilde{\Omega}')| &= \sum_{i=2}^{\eta} |f_i(\tilde{\Omega}')\tilde{\pi}(\tilde{\Omega}')f_i(\tilde{\Omega})\tilde{\mu}_i^t| \leq \sum_{i=2}^{\eta} |f_i(\tilde{\Omega}')\tilde{\pi}(\tilde{\Omega}')f_i(\tilde{\Omega})||\tilde{\mu}_2|^t \\ &\leq \tilde{\pi}(\tilde{\Omega}')\left[\sum_{i=2}^{\eta} f_i^2(\tilde{\Omega}')\sum_{i=2}^{\eta} f_i^2(\tilde{\Omega})\right]^{\frac{1}{2}} |\tilde{\mu}_2|^t \end{aligned} \tag{A.24}$$

(A.25)

where the last inequality follows from Cauchy-Schwarz inequality. Following from the definition of inner product together with that of $\delta_{\tilde{\Omega}'}(\tilde{\Omega})$ in (A.22), and relying on the orthonormality of the set f_i, we have

$$\tilde{\pi}(\tilde{\Omega}') = \langle \delta_{\tilde{\Omega}'}, \delta_{\tilde{\Omega}'} \rangle = \left\langle \sum_{i=1}^{\eta} f_i(\tilde{\Omega}')\tilde{\pi}(\tilde{\Omega}')f_i, \sum_{i=1}^{\eta} f_i(\tilde{\Omega}')\tilde{\pi}(\tilde{\Omega}')f_i \right\rangle = \tilde{\pi}(\tilde{\Omega}')^2 \sum_{i=1}^{\eta} f_i(\tilde{\Omega}')^2 \tag{A.26}$$

The implication is that $\sum_{i=2}^{\eta} f_i(\tilde{\Omega}')^2 \leq \frac{1}{\tilde{\pi}(\tilde{\Omega}')}$, and consequently

$$|\tilde{P}_{\varepsilon}^{t}(\tilde{\Omega}, \tilde{\Omega}') - \tilde{\pi}(\tilde{\Omega}')| \leq \frac{\tilde{\pi}(\tilde{\Omega}')}{\sqrt{\tilde{\pi}(\tilde{\Omega})\tilde{\pi}(\tilde{\Omega}')}} |\tilde{\mu}_2|^t \tag{A.27}$$

Such that if $\tilde{\Omega}_0$ is the initial state of $(\tilde{\mathbf{\Omega}}, \tilde{P}_{\varepsilon})$ with respective stationary distribution of $\tilde{\pi}(\tilde{\Omega}_0)$, then

$$\left|\left|\mathbf{s}_0 \tilde{P}_{\varepsilon}^{t} - \tilde{\pi}\right|\right| = \frac{1}{2} \sum_{\tilde{\Omega}' \in \tilde{\mathbf{\Omega}}} |\tilde{P}_{\varepsilon}^{t}(\tilde{\Omega}_0, \tilde{\Omega}') - \tilde{\pi}(\tilde{\Omega}')| \leq \frac{1}{2} \sum_{\tilde{\Omega}' \in \tilde{\mathbf{\Omega}}} \frac{\tilde{\pi}(\tilde{\Omega}')}{\sqrt{\tilde{\pi}(\tilde{\Omega}_0)\tilde{\pi}(\tilde{\Omega}')}} |\tilde{\mu}_2|^t \leq \frac{1}{2} \frac{1}{\tilde{\pi}_{\min}} |\tilde{\mu}_2|^t$$

where the last inequality follows from the fact that $\tilde{\pi}(\tilde{\Omega}_0)\tilde{\pi}(\tilde{\Omega}') \leq \min_{\tilde{\Omega} \in \tilde{\mathbf{\Omega}}} \tilde{\pi}(\tilde{\Omega})^2 = \tilde{\pi}_{\min}$.

For $\left|\left|\mathbf{s}_0 \tilde{P}_{\varepsilon}^{t} - \tilde{\pi}\right|\right| \leq \varpi$, then $\frac{1}{2}\frac{1}{\tilde{\pi}_{\min}} |\tilde{\mu}_2|^t \leq \varpi$. Consequently, the convergence time is

$$T_c \leq \frac{\ln(2\varpi\tilde{\pi}_{\min})}{\ln|\tilde{\mu}_2|} \tag{A.28}$$

Acknowledgments

This study has benefited from comments and suggestions of Julio Davila, Arkadi Predtetchinsk, Antoine Mandel, François Lafond, Hippolyte d'Albis, Giorgio Triulzi and Robin Cowan. 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 research was supported by UNU-MERIT. The usual disclaimer applies.

References

Aldous, D. (1987): "On the Markov chain simulation method for uniform combinatorial distributions and simulated annealing," *Probability in the Engineering and Informational Sciences*, 1(01), 33–46.

Aldous, D., and J. A. Fill (1994): "Reversible Markov Chains," .

Alon, N. (1986): "Eigenvalues and expanders," *Combinatorica*, 6(2), 83–96.

Alon, N., and V. D. Milman (1985): "λ_1, Isoperimetric inequalities for graphs, and superconcentrators," *Journal of Combinatorial Theory, Series B*, 38(1), 73 – 88.

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

Asavathiratham, C. (2001): *The influence model: a tractable representation for the dynamics of networked Markov chains*. Massachusetts Institute of Technology, Dept. of Electrical Engineering and Computer Science, Ph.D Thesis.

Beggs, A. (2005): "Waiting times and equilibrium selection," *Economic Theory*, 25(3), 599–628.

Billings, L., and I. B. Schwartz (2008): "Identifying almost invariant sets in stochastic dynamical systems," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 18(2), 023122.

Bremaud, P. (1999): *Markov chains: Gibbs fields, Monte Carlo simulation, and queues*, vol. 31. springer.

Catoni, O. (1999): "Simulated annealing algorithms and Markov chains with rare transitions," in *Séminaire de probabilités XXXIII*, ed. by J. Azéma, M. Émery, M. Ledoux, and M. Yor, pp. 69–119. Springer, Berlin Heidelberg.

Diaconis, P., and D. Stroock (1991): "Geometric bounds for eigenvalues of Markov chains," *The Annals of Applied Probability*, pp. 36–61.

Durrett, R. (2006): *Random Graph Dynamics*, Cambridge Series on Statistical and Probabilistic Mathematics. Cambridge University Press.

Ellison, G. (1993): "Learning, Local Interaction, and Coordination," *Econometrica*, 61(5), 1047–1071.

——— (2000): "Basins of Attraction, Long-Run Stochastic Stability, and the Speed of Step-by-Step Evolution," *Review of Economic Studies*, 67(1), 17–45.

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

Froyland, G. (2005): "Statistically optimal almost-invariant sets," *Physica D: Nonlinear Phenomena*, 200(3), 205–219.

Horn, R. A., and C. R. Johnson (1990): *Matrix Analysis*. Cambridge University Press.

Jackson, M. O., and L. Yariv (2007): "Diffusion of Behavior and Equilibrium Properties in Network Games," *American Economic Review*, 97(2), 92–98.

Kandori, M., G. J. Mailath, and R. Rob (1993): "Learning, Mutation, and Long Run Equilibria in Games," *Econometrica*, 61(1), 29–56.

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

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

Lee, I. H., A. Szeidl, and A. Valentinyi (2003): "Contagion and State Dependent Mutations," *The B.E. Journal of Theoretical Economics*, 3(1), 1–29.

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

Levin, D. A., Y. Peres, and E. L. Wilmer (2009): *Markov chains and mixing times*. American Mathematical Society.

Mohar, B. (1989): "Isoperimetric numbers of graphs," *Journal of Combinatorial Theory, Series B*, 47(3), 274–291.

Montanari, A., and A. Saberi (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.

Robson, A. J., and F. Vega-Redondo (1996): "Efficient Equilibrium Selection in Evolutionary Games with Random Matching," *Journal of Economic Theory*, 70(1), 65–92.

Sandholm, W. (2010): *Population Games and Evolutionary Dynamics*, Economic Learning and Social Evolution. MIT Press.

Sandholm, W. H. (2001): "Almost Global Convergence to p-Dominant Equilibrium," *International Journal of Game Theory*, pp. 107–116.

Simon, H., and A. Ando (1961): "Aggregation of Variables in Dynamic Systems," *Econometrica*, 29(2), 111–138.

Sinclair, A. (1992): "Improved Bounds for Mixing Rates of Markov Chains and Multicommodity Flow," *Combinatorics, Probability and Computing*, 1(04), 351–370.

Sinclair, A., and M. Jerrum (1989): "Approximate counting, uniform generation and rapidly mixing Markov chains," *Information and Computation*, 82(1), 93–133.

Takacs, C. (2006): "On the fundamental matrix of finite state Markov chains, its Eigensystem and its relation to hitting times," *Mathematica Pannonica*, 183, 193.

Young, H. P. (1993): "The Evolution of Conventions," *Econometrica*, 61(1), 57–84.

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