Stochastic dynamics and Edmonds' algorithm

Jonathan Newton* and William H. Sandholm†

November 25, 2019

Abstract

Recently, there has been a revival of interest in cyclic decompositions of stochastic dynamics. These decompositions consider the behavior of dynamics over the short, medium and long run, aggregating cycles of behavior into progressively larger cycles, eventually encompassing the entire state space. We show that these decompositions are equivalent to the aggregative stage of Edmonds' algorithm and that this equivalence can be used to recover well-known results in the literature.

1. Introduction

It is known that, under a variety of behavioral dynamics, including variants of best response dynamics, behavior in a population can take a long time to converge to equilibrium (Ellison, 1993).[1] This is not universal, and indeed it has been shown that in several situations convergence is relatively rapid. These situations included specific interaction structures (Ellison, 1993, 2000; Young, 2011), processes with various forms of inertia (Norman, 2009; Arieli and Young, 2016), matching problems (Newton and Sawa, 2015), and specific payoff parameters under both individualistic (Arieli et al., 2019) and coalitional (Newton and Angus, 2015) dynamics.

Of course, in other settings, rapid convergence may fail to occur. In such cases, the dynamics of behavior in the short and medium run become important. Recently, some economists have turned their attention to studying such behavior. In particular, Cui and

*Institute of Economic Research, University of Kyoto, Yoshida-honmachi, Sakyo-ku, Kyoto 606-8501. e-mail: `newton@kier.kyoto-u.ac.jp`; website: `jonathannewton.net`.

†Department of Economics, University of Wisconsin, 1180 Observatory Drive, Madison, WI 53706. e-mail: `whs@ssc.wisc.edu`; website: `www.ssc.wisc.edu/~whs`.

[1] The meaning of "a long time" can either be interpreted qualitatively or mathematically, an example of the latter being that convergence time increases (exponentially, polynomially, linearly) in population size. The equilibria considered in such models tend to be some subset of Nash equilibria with good stability properties.

Zhai (2010); Levine and Modica (2016) consider the cyclic decomposition approach of Freidlin and Wentzell (1984). This approach describes short run behavior as cycles (of which a single absorbing state is a special case) that can be combined into larger cycles that describe medium run behavior. These cycles can in turn be combined into cycles that describe long run behavior.

In the current paper, we show that these cyclic decompositions are equivalent to the aggregative stage of the famous Chiu-Liu-Edmond's (CLE) algorithm (Chu and Liu, 1965; Edmonds, 1967). The CLE algorithm solves (in polynomial time) the problem of finding minimal (or maximal) spanning trees of weighted directed graphs. When the vertices of a graph correspond to states in a Markov chain and the edge weights measure transition probabilities between these states, solutions to this problem predict long run behavior (Freidlin and Wentzell, 1984), a connection that was later introduced to economics (Young, 1993; Kandori et al., 1993) and applied to many different economic problems (see Sandholm, 2010; Newton, 2018, for an overview). Our result further illuminates this connection: just as long run behavior corresponds to the solutions of minimum cost spanning tree problems, short and medium run behavior corresponds to the algorithm that is commonly used to solve such problems.

The usefulness of our result is immediately apparent. Many implementations of the CLE algorithm exist. Applying such code to a dynamic process such as those considered here, the interim iterations of the algorithm describe the short and, subsequently, medium run behavior of the process, with the final output describing long run behavior. In fact, the disaggregative stage of the CLE algorithm that gives minimal spanning trees can also be shown to give long run stable states. Hence, these two concepts are equivalent and the minimal spanning tree characterization of long run stability can be regarded as a direct implication of the CLE algorithm. Furthermore, the disaggregation of cyclic decompositions leads to a natural structure on the cycles, previously noted by Catoni (1999) for the particular cyclic decompositions considered in Freidlin and Wentzell (1984). This structure permits precise estimation of (the order of magnitude of) probability flows out of cycles and, consequently, the stationary distribution of the process.

The paper is organized as follows. Section 2 gives the model, definitions and some preliminary results. Section 3 contains the main analysis and results. All proofs are given in the Appendix.

2. Model

2.1 Primitives

Let $\{P^\eta\}_\eta$, $\eta \geq 0$ be a parameterized collection of Markov transition matrices on the finite state space $\mathcal{X}$. That is, for $x, y \in \mathcal{X}$, P^η_{xy} denotes the probability of a transition from state x to state y at parameter value η. Assume that the collection is *weakly regular* (Sandholm, 2010). That is, P^η is continuous in η, the process is irreducible for strictly positive η, and if, for some x, $y \neq x$, $P^0_{xy} = 0$ and $P^{\hat{\eta}}_{xy} > 0$ for some $\hat{\eta} > 0$, then $\{P^\eta_{xy}\}_\eta$ satisfies

$$(1) \qquad P^\eta_{xy} = \exp\left(-\frac{1}{\eta}\left(k + o(1)\right)\right) \qquad \text{for some } k \geq 0,$$

where $o(1)$ represents a term that approaches zero as η approaches zero.

2.2 Composite states

For $\eta > 0$, irreducibility of the process implies the existence of a unique invariant probability distribution, μ^η. Assume that the limit $\mu^0 = \lim_{\eta \to 0} \mu^\eta$ exists. A set of states $\alpha \subseteq \mathcal{X}$ shall be referred to as a *composite*. Let μ^η_α denote the mass placed on α by μ^η. If $\mu^0_{\{x\}} > 0$, we say that x is *stochastically stable* (Foster and Young, 1990). For $\eta > 0$, we denote the share of the invariant probability mass on composite α that flows each period to another composite β by $\bar{P}^\eta_{\alpha\beta}$. That is,

$$(2) \qquad \bar{P}^\eta_{\alpha\beta} = \frac{1}{\mu^\eta_\alpha} \sum_{\substack{x \in \alpha \\ y \in \beta}} \mu^\eta_{\{x\}} P^\eta_{xy}.$$

Note that when α and β are singleton sets, $\alpha = \{x\}$ and $\beta = \{y\}$, we have that $\bar{P}^\eta_{\alpha\beta} = P^\eta_{xy}$.

We define a cost function

$$C(\cdot,\cdot) : (\text{Powerset}(\mathcal{X}) \setminus \{\varnothing\}) \times (\text{Powerset}(\mathcal{X}) \setminus \{\varnothing\}) \to \mathbb{R}_+ \cup \{\infty\}$$

that will measure the order of magnitude of probability flows between composites. If $\bar{P}^{\hat{\eta}}_{\alpha\beta} > 0$ for some $\hat{\eta} > 0$, then define

$$(3) \qquad C(\alpha, \beta) = \lim_{\eta \to 0} -\eta \log \bar{P}^\eta_{\alpha\beta}$$

and if $\bar{P}^\eta_{\alpha\beta} = 0$ for all η, then let $C(\alpha, \beta) = \infty$. Cost functions measure the order of magnitude of transition probabilities for low values of η. Transitions with a high cost are less likely than transitions with a low cost. Similarly define the stationary distribution decay rate as

$$(4) \qquad r_\alpha = \lim_{\eta \to 0} -\eta \log \mu_\alpha^\eta,$$

which measures the order of magnitude of invariant probabilities for low values of η.

2.3 Cycles

For a given partition $\mathcal{P}_\ell$ of $\mathcal{X}$ with at least two elements, define a *least cost transition correspondence* $\sigma_\ell \colon \mathcal{P}_\ell \rightrightarrows \mathcal{P}_\ell$ and a function $\underline{C}$ that gives the cost of such least cost transitions. For each composite $\alpha \in \mathcal{P}_\ell$, we define

$$(5) \qquad \sigma_\ell(\alpha) = \operatorname*{argmin}_{\beta \in \mathcal{P}_\ell \smallsetminus \{\alpha\}} C(\alpha, \beta) \qquad \text{and} \qquad \underline{C}(\alpha) = \min_{\beta \in \mathcal{P}_\ell \smallsetminus \{\alpha\}} C(\alpha, \beta).$$

Note that there is no ℓ subscript on $\underline{C}$. This is because the quantity $\underline{C}(\alpha)$ will prove to be independent of the partition structure of $\mathcal{X} \smallsetminus \alpha$.

A *cycle* in $\mathcal{P}_\ell$ is a set $\Gamma_\ell \subseteq \mathcal{P}_\ell$ such that $\Gamma_\ell = \bigcup_{m=0}^{\bar{m}} \{\alpha^m\}$ for some sequence $\alpha^0, \ldots, \alpha^{\bar{m}}$ that satisfies $\alpha^1 \in \sigma_\ell(\alpha^0)$, $\alpha^2 \in \sigma_\ell(\alpha^1)$, $\ldots$, $\alpha^0 \in \sigma_\ell(\alpha^{\bar{m}})$. Note that the sequence $\alpha^0, \ldots, \alpha^{\bar{m}}$ may contain repeated elements. Let $\mathbf{\Gamma}_\ell$ be the set of all cycles in $\mathcal{P}_\ell$.

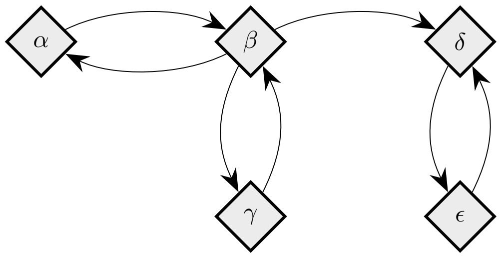
Lemma 2.1. *If α^m, α^n are elements of some cycle Γ_ℓ in $\mathcal{P}_\ell$, then*

$$r_{\alpha^m} + \underline{C}(\alpha^m) = r_{\alpha^n} + \underline{C}(\alpha^n).$$

To see the intuition behind Lemma 2.1, consider a cycle with an associated sequence $\alpha^0, \ldots, \alpha^{\bar{m}}$. At a stationary distribution, (i) the flow of probability from α^0 to α^1 cannot be greater than the total flow of probability out of α^1; (ii) the order of magnitude of the total flow from α^1 equals the order of magnitude of the largest flows from α^1, one of which is to α^2. Therefore, the order of magnitude of the total flow from α^0 to α^1 is no greater than the order of magnitude of the total flow from α^1 to α^2. Writing in terms of decay rates, recalling that larger decay rates imply a smaller order of magnitude, and iterating the above argument, we have

$$r_{\alpha^0} + \underline{C}(\alpha^0) \geq r_{\alpha^1} + \underline{C}(\alpha^1) \geq \ldots \geq r_{\alpha^{\bar{m}}} + \underline{C}(\alpha^{\bar{m}}) \geq r_{\alpha^0} + \underline{C}(\alpha^0),$$

which gives the result. Hence the probability flows around any given cycle in $\mathcal{P}_\ell$ are all of the same order of magnitude, even if there are many such cycles with possibly shared



$$\mathbf{\Gamma}_l = \Big\{\{\alpha,\beta\},\{\beta,\gamma\},\{\delta,\epsilon\},\{\alpha,\beta,\gamma\}\Big\}$$
$$\mathbf{\Gamma}_l^C = \Big\{\{\delta,\epsilon\}\Big\}$$
$$\mathbf{\Gamma}_l^S = \Big\{\{\alpha,\beta\},\{\beta,\gamma\},\{\delta,\epsilon\}\Big\}$$

Figure 1: **Cycles.** For given least cost correspondence, the set of cycles, closed cycles and simple cycles.

elements.

2.4 Cyclic decompositions

A nested sequence of partitions $\mathcal{P}_0,\ldots,\mathcal{P}_L$, $\mathcal{P}_0 = \bigcup_{x\in X}\{\{x\}\}$ and $\mathcal{P}_L = \{X\}$, is a *cyclic decomposition* if, for all $\alpha \notin \mathcal{P}_\ell$, $\alpha \in \mathcal{P}_{\ell+1}$, we have that α is the union of the elements of some cycle Γ_ℓ in $\mathcal{P}_\ell$. We say that Γ_ℓ is *consolidated* to give α. If we define the *birthday* of composite α as $d(\alpha) = \min\{\ell\colon \alpha \in \mathcal{P}_\ell\}$, and let $\Pi(\alpha) = \{\pi \in \mathcal{P}_{d(\alpha)-1}\colon \pi \subset \alpha\}$ be the set of *pieces* of α, meaning the composites merged to form α on its birthday, then for all $\alpha \in \mathcal{P}_\ell$, $\ell \geq 1$, $|\alpha| \geq 2$, we have that $\Pi(\alpha) = \Gamma_{d(\alpha)-1}$ for some $\Gamma_{d(\alpha)-1} \in \mathbf{\Gamma}_{d(\alpha)-1}$.

A cycle Γ_ℓ in $\mathcal{P}_\ell$ is *closed* if it is closed under σ_ℓ in that $\sigma_\ell(\Gamma_\ell) \subseteq \Gamma_\ell$ (see Figure 1). Note that any two closed cycles must be disjoint. Let $\mathbf{\Gamma}_\ell^C$ be the set of closed cycles in $\mathcal{P}_\ell$.

Definition. *An FW decomposition is a cyclic decomposition $\mathcal{P}_0^{FW},\ldots,\mathcal{P}_L^{FW}$ such that, for $\ell = 0,\ldots,L-1$, every cycle in $\mathbf{\Gamma}_\ell^C$ is consolidated to give $\mathcal{P}_{\ell+1}^{FW}$.*

Lemma 2.2. *An FW decomposition exists and is unique.*

These decompositions are those considered in Freidlin and Wentzell (1984) and in later simplifications and extensions of their analysis (Catoni, 1999; Cui and Zhai, 2010; Levine and Modica, 2016).

A cycle $\mathcal{P}_\ell$ is *simple* if it satisfies the definition of a cycle for some sequence $\alpha^0,\ldots,\alpha^m$ with no repeated elements (see Figure 1). Let $\mathbf{\Gamma}_\ell^S$ be the set of simple cycles in $\mathcal{P}_\ell$. Freidlin and Wentzell (1984, p.180) effectively assume that $\sigma_\ell(\cdot)$ is always a singleton. Under this assumption, every cycle is both closed and simple. This assumption is not made in Catoni (1999); Cui and Zhai (2010); Levine and Modica (2016). Note that in models with uniform mutations such as those of Kandori et al. (1993) and Young (1993) it is rarely the case that $\sigma_0(\cdot)$ is a singleton.

Definition. *A CLE decomposition is a cyclic decomposition* $\mathcal{P}_0^{Ed}, \dots, \mathcal{P}_L^{Ed}$ *such that, for* $\ell = 0, \dots, L-1$, *precisely one cycle* $\Gamma_\ell \in \mathbf{\Gamma}_\ell^S$ *is consolidated to give* $\mathcal{P}_{\ell+1}^{Ed}$.

Lemma 2.3. *A CLE decomposition exists.*

A CLE decomposition consolidates one simple cycle at each stage. The reason we call this a CLE decomposition is that we will show that such a decomposition is identical to the steps followed by the famous Chu-Liu/Edmonds' Algorithm (Edmonds, 1967; Chu and Liu, 1965).

3. Analysis

Take a cyclic decomposition $\mathcal{P}_0, \dots, \mathcal{P}_L$. At any value of η, the probability mass on any composite $\alpha \in \mathcal{P}_\ell$ under the invariant measure must be distributed amongst its pieces. That is, $\mu_\alpha^\eta = \sum_{\alpha^m \in \Pi(\alpha)} \mu_{\alpha^m}^\eta$. This implies the following relationship in the limit as $\eta \to 0$.

Lemma 3.1. *If* $\alpha \in \mathcal{P}_\ell$, $|\alpha| > 1$, *then* $r_\alpha = \min_{\alpha^m \in \Pi(\alpha)} r_{\alpha^m}$.

Using Lemmas 2.1 and 3.1, we can show the relationship between the cost function on composites in $\mathcal{P}_\ell$ and the cost function on composites in $\mathcal{P}_{\ell-1}$.

Lemma 3.2. *If* $\omega, \alpha \in \mathcal{P}_\ell$, $d(\omega) < \ell$, $d(\alpha) = \ell$, $\Pi(\alpha) = \{\alpha^0, \dots, \alpha^{\bar{m}}\}$, *then*

$$C(\omega, \alpha) = \min_{\alpha^m \in \Pi(\alpha)} C(\omega, \alpha^m), \tag{6}$$

$$C(\alpha, \omega) = \min_{\alpha^m \in \Pi(\alpha)} \left(\max_{\alpha^n \in \Pi(\alpha)} \underline{C}(\alpha^n) - \underline{C}(\alpha^m) + C(\alpha^m, \omega) \right), \tag{7}$$

This shows that when building a cyclic decomposition, we can ignore raw transition probabilities except at the first step when we calculate the values of the cost function on $\mathcal{P}_0$.

The intuition for (6) is that the order of magnitude of the probability flow from ω to α equals the order of magnitude of the largest probability flow from ω to any of the pieces of α.

The intuition for (7) is more subtle. Consider any $\alpha^m \in \Pi(\alpha)$. Lemma 3.1 implies that the order of magnitude of the probability mass on α^m as a proportion of the probability mass on α is given by $r_{\alpha^m} - \min_{\alpha^n \in \Pi(\alpha)} r_{\alpha^n}$. By Lemma 2.1, this quantity equals $\max_{\alpha^n \in \Pi(\alpha)} \underline{C}(\alpha^n) - \underline{C}(\alpha^m)$, the first two terms of (7). The final term in (7) gives the share of the probability mass on α^m that flows to ω. The sum of the three terms then gives the order of magnitude of the share of the probability mass on α that flows to ω from α^m. Considering all possible $\alpha^m \in \Pi(\alpha)$, we obtain (7).

Note that (6) and (7) do not directly tell us the value for $C(\alpha, \beta)$ when α and β are both newly created composites, $d(\alpha) = d(\beta) = \ell$. However, we show in Lemma A.2 in Appendix A that if $\Pi(\alpha)$ and $\Pi(\beta)$ are cycles and one of the cycles is consolidated, then the other cycle remains a cycle in the new partition and can be consolidated at the next step. Therefore, $C(\alpha, \beta)$ is given by applying (6) and (7) consecutively. Using this ability to consider decompositions as sequential consolidations of a single cycle at each step, we obtain the following theorem.

Theorem 3.3. *For any given cyclic decomposition $\mathcal{P}_0, \ldots, \mathcal{P}_{L^*}$, there exists a CLE decomposition $\mathcal{P}_0^{Ed}, \ldots, \mathcal{P}_L^{Ed}$ of which $\mathcal{P}_0, \ldots, \mathcal{P}_{L^*}$ is a subsequence.*

3.1 Edmonds' Algorithm

Consider a weighted, directed graph on the vertex set $\mathcal{P}_\ell$. Let the weight of a directed edge from from α to β be given by $C(\alpha, \beta)$.

Definition. *A (directed) spanning tree with root $\alpha \in \mathcal{P}_\ell$ is an acyclic directed graph on $\mathcal{P}_\ell$ in which every vertex except α has precisely one outgoing edge.*

The algorithm of Chu and Liu (1965) and Edmonds (1967) can be used to find spanning trees that have minimum or maximum sum of edge weights.[2] The algorithm has two stages, an aggregative stage in which simple cycles in the graph are consolidated and a disaggregative stage in which this consolidation is reversed and a spanning tree constructed. We shall follow the steps for constructing a spanning tree with a minimum sum of edge weights.

3.1.1 Aggregation

Start with the complete, weighted, directed graph G_0 on $\mathcal{P}_0^{Ed}$ that has edge weights given by $C(\{x\}, \{y\})$ for each $\{x\}, \{y\} \in \mathcal{P}_0^{Ed}$. At step ℓ, $\ell \geq 1$, of the aggregative stage of the algorithm, the graph $G_{\ell-1}$ on $\mathcal{P}_{\ell-1}^{Ed}$ is used to construct a complete, weighted, directed graph G_ℓ on a new partition $\mathcal{P}_\ell^{Ed}$ of $\mathcal{X}$.

To do this, the algorithm considers some subgraph $\bar{G}_{\ell-1}$ of $G_{\ell-1}$ on $\mathcal{P}_{\ell-1}^{Ed}$ that, for each vertex, includes a single outgoing edge from amongst its outgoing edges with minimal weight. By construction, $\bar{G}_{\ell-1}$ must include at least one *graphical simple cycle*, a sequence

[2] The cited papers find spanning arborescences (every vertex except α has precisely one *incoming* edge), which is the same problem following a transformation.

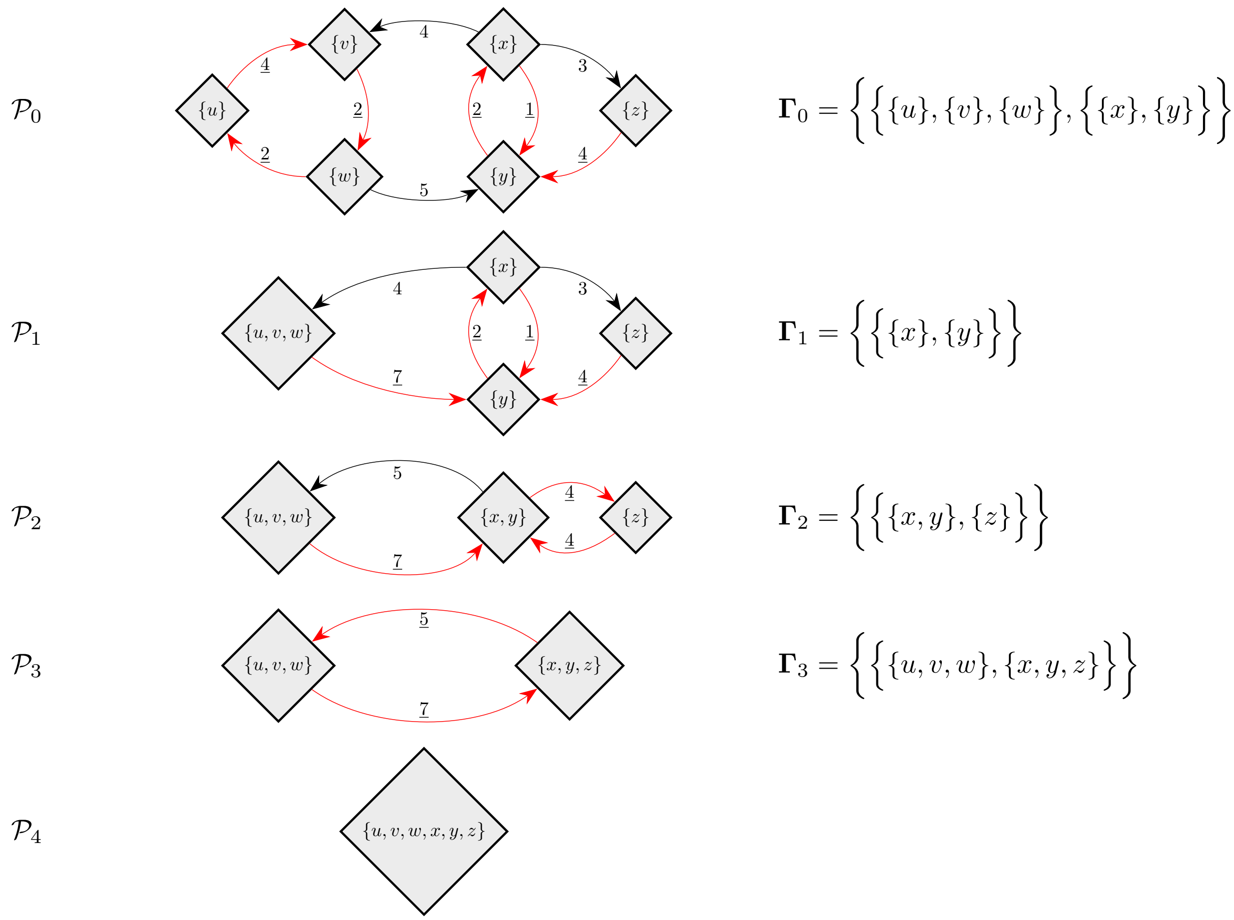


Figure 2: **Aggregation of states in a CLE decomposition/Edmonds' algorithm.** This diagram illustrates cyclic decomposition, moving step by step from $\mathcal{P}_0$ to $\mathcal{P}_4$. The cost function $C(\cdot, \cdot)$ is given by edge weights, with missing edges denoting transitions with infinite cost. The values of $C(\cdot, \cdot)$ on the initial partition $\mathcal{P}_0$ are assumed. Costs for partitions $\mathcal{P}_\ell$, $\ell > 0$, are calculated as described in the text. Least cost transitions are denoted by an underlined cost and a red arrow, so that a subset of elements in a partition is a cycle if is the vertex set for a (graphical) simple cycle of such edges in the diagram. At each step, one simple cycle is consolidated. For example, in moving from $\mathcal{P}_0$ to $\mathcal{P}_1$, cycle $\{\{u\}, \{v\}, \{w\}\} \subset \mathcal{P}_0$ is consolidated to form $\{u, v, w\} \in \mathcal{P}_1$, with the costs of transitions to and from $\{u, v, w\}$ given by expressions (6) and (7) in the text. For example, $C(\{u, v, w\}, \{y\}) = \min\{4 - 4 + \infty, 4 - 2 + \infty, 4 - 2 + 5\} = 7$.

$\beta^0, \beta^1, \ldots, \beta^{\bar{n}}$ of non-repeated vertices, with edges from β^n to β^{n+1} for $n = 0, \ldots, \bar{n} - 1$, and an edge from $\beta^{\bar{n}}$ to β^0.

One graphical simple cycle is then consolidated to a single vertex and new edge weights are calculated.[3] For $\ell = 0$, as edge weights are given by $C(\cdot, \cdot)$, graphical simple cycles clearly correspond to the simple cycles defined earlier in the paper. Furthermore, the values that the algorithm gives for the new edge weights following the consolidation of a graphical simple cycle (see, e.g. p.1398 of Chu and Liu, 1965) are exactly the values of the cost function given in (6) and (7). Thus, at every stage in the aggregation, the set of graphical simple cycles in the algorithm corresponds to the set of simple cycles in a CLE decomposition, and the edge weights calculated by the algorithm correspond to the cost functions calculated by a CLE decomposition.

So CLE decompositions are the aggregative stage of Edmonds' algorithm. Theorem 3.3 then tells us that all cyclic decompositions, including FW decompositions, are reduced form versions of the aggregative stage of Edmonds' algorithm. An example of such a decomposition/aggregration is given in Figure 2.

3.1.2 Disaggregation

After the aggregative stage of the algorithm creates a CLE decomposition $\mathcal{P}_0^{Ed}, \ldots, \mathcal{P}_L^{Ed}$, the disaggregative stage of the algorithm (illustrated in Figure 3) proceeds as follows.

Assume we have constructed spanning trees $T_L, T_{L-1}, \ldots, T_\ell$ on the partitions/vertex sets $\mathcal{P}_L^{Ed}, \mathcal{P}_{L-1}^{Ed}, \ldots, \mathcal{P}_\ell^{Ed}$. We construct a graph on $\mathcal{P}_{\ell-1}^{Ed}$, starting with T_ℓ on $\mathcal{P}_\ell^{Ed}$. Take $\alpha_\ell \in \mathcal{P}_\ell^{Ed}$ that was consolidated in moving from $\mathcal{P}_{\ell-1}^{Ed}$ to $\mathcal{P}_\ell^{Ed}$ at the aggregative stage of the algorithm. Recall that $\mathcal{P}_{\ell-1}^{Ed} = \mathcal{P}_\ell^{Ed} \setminus \{\alpha_\ell\} \cup \Pi(\alpha_\ell)$.

(1) Firstly, for any edges in T_ℓ that do not include α_ℓ, include a corresponding edge in $T_{\ell-1}$ between the same vertices.

(2) Secondly, note that $\Pi(\alpha_\ell)$, by definition of a CLE decomposition, is a simple cycle in $\mathcal{P}_{\ell-1}^{Ed}$. Thus we can write $\Pi(\alpha_\ell) = \{\alpha_{\ell-1}^0, \ldots, \alpha_{\ell-1}^{\bar{m}}\}$ for $\alpha_{\ell-1}^1 \in \sigma_{\ell-1}(\alpha_{\ell-1}^0), \ldots, \alpha_{\ell-1}^{\bar{m}} \in \sigma_{\ell-1}(\alpha_{\ell-1}^{\bar{m}-1}), \alpha_{\ell-1}^0 \in \sigma_{\ell-1}(\alpha_{\ell-1}^{\bar{m}})$. Add corresponding directed edges to $T_{\ell-1}$, that is from $\alpha_{\ell-1}^0$ to $\alpha_{\ell-1}^1$, …, $\alpha_{\ell-1}^{\bar{m}-1}$ to $\alpha_{\ell-1}^{\bar{m}}$ and $\alpha_{\ell-1}^{\bar{m}}$ to $\alpha_{\ell-1}^0$. The vertices $\Pi(\alpha_\ell)$ and these edges form a graphical simple cycle.

[3] This follows Edmonds (1967). The presentation of Chu and Liu (1965) consolidates multiple simple cycles at each step. This creates multiple edges between newly created composites. This is not a problem, however, as one need only consider the minimum of these edges, which is what is obtained by consecutive application of (6) and (7).

(3) Next, account for edges that were directed to or from α_ℓ in T_ℓ. Note that any such edge is associated with some $\alpha^m_{\ell-1} \in \Pi(\alpha_\ell)$ that solved the minimization problem in (6) or (7) when $\Pi(\alpha_\ell)$ was consolidated to make α_ℓ at the aggregative stage of the algorithm. Replace each edge to or from α_ℓ in T_ℓ with an edge to or from the appropriate $\alpha^m_{\ell-1}$ in $T_{\ell-1}$.

(4a) If α_ℓ was not the root of T_ℓ, then some $\alpha^m_{\ell-1} \in \Pi(\alpha_\ell)$ will now have two outgoing edges. One of these edges will be to another element of $\Pi(\alpha_\ell)$. Delete this edge. The resulting $T_{\ell-1}$ is a spanning tree on $\mathcal{P}^{Ed}_{\ell-1}$.

(4b) If α_ℓ is the root of T_ℓ, then choose an element of $\Pi(\alpha_\ell)$ whose outgoing edge has the largest edge weight. That is, choose an element that solves $\max_{\xi \in \Pi(\alpha_\ell)} \underline{C}(\xi)$. Let $\alpha^*_{\ell-1} \in \Pi(\alpha_\ell)$ be the chosen element. Delete the outgoing edge from $\alpha^*_{\ell-1}$ in $T_{\ell-1}$. The resulting $T_{\ell-1}$ is a spanning tree on $\mathcal{P}^{Ed}_{\ell-1}$.

Iterating the above, the algorithm eventually obtains a spanning tree T_0 on $\mathcal{P}^{Ed}_0$. It was shown by Chu and Liu (1965) and Edmonds (1967) that T_0 has minimal sum of edge weights amongst all spanning trees on $\mathcal{P}^{Ed}_0$.

Note that if $\alpha^*_{\ell-1}$ is the root of $T_{\ell-1}$, then $T_{\ell-2}, T_{\ell-3}, \ldots, T_0$ will all be rooted at some $\beta \subseteq \alpha^*_\ell$. For $\xi \in \Pi(\alpha_\ell)$ such that $\underline{C}(\xi) < \underline{C}(\alpha^*_{\ell-1})$, we have, by Lemma 2.1, that $r_{\alpha^*_{\ell-1}} < r_\xi$. Therefore, $\mu^\eta_\xi \to 0$ as $\eta \to 0$. This logic again applies when $\alpha^*_{\ell-1}$ is disaggregated. Continuing in this manner, we obtain the following theorem.

Theorem 3.4. *$r_\alpha = 0$ if and only if there exists a spanning tree on $\mathcal{P}_0$ with root $\{x\} \subseteq \alpha$ that has minimal sum of edge weights amongst all spanning trees on $\mathcal{P}_0$.*

Readers may recognize this as the tree characterization of stochastically stable states (Freidlin and Wentzell, 1984; Young, 1993), a result that underpins a vast literature on evolutionary game theory in economics. The decomposition based proof discussed above is notably transparent in illustrating the forces that drive this result.

Note that if $r_\alpha > 0$, then $\mu^\eta_\alpha \to 0$ at an exponential rate as $\eta \to 0$. However, for $r_\alpha = 0$, it may still be the case that $\mu^\eta_\alpha \to 0$ at a subexponential rate. If we strengthen the assumption of weakly regular Markov chains and replace expression (1) by

$$P^\eta_{xy} = \Big(a_{xy} + o(1)\Big) \exp\left(-\frac{1}{\eta}k\right) \qquad \text{for some } a_{xy} > 0 \text{ and some } k > 0, \tag{8}$$

then we have the class of regular Markov chains (Young, 1993). In this case, $r_\alpha > 0$ if and only if $\mu^\eta_\alpha \to 0$ as $\eta \to 0$. This last statement about regular Markov chains is usually proved by direct reference to the classic Markov chain tree theorem (see, e.g. Lemma 3.1 in Freidlin and Wentzell, 1984). Interestingly, it cannot be proved by a decomposition argument. As mentioned subsequent to Lemma 3.2, cyclic decompositions jettison information on probabilities other than the exponential decay costs captured by the cost function. Specifically, regardless of whether P^{η}_{xy} is given by (1) or (8), we have that $C(\{x\},\{y\}) = k$.

3.2 The invariant measure for all states

Take a cyclic decomposition $\mathcal{P}_0, \ldots, \mathcal{P}_L$. As μ^{η} is a probability measure on $\mathcal{X}$, it must be that $\mu^{\eta}_{\mathcal{X}} = 1$ for all η. Therefore $r_{\mathcal{X}} = 0$. If we write $\alpha_L = \mathcal{X} \in \mathcal{P}_L$, and let α^*_{L-1} solve $\max_{\xi \in \Pi(\alpha_L)} \underline{C}(\xi)$, then Lemma 2.1 together with Lemma 3.1 implies that $r_{\alpha^*_{L-1}} = 0$. Lemma 2.1 then further implies that for all $\alpha_{L-1} \in \Pi(\alpha_L)$, we have $r_{\alpha_{L-1}} = \underline{C}(\alpha^*_{L-1}) - \underline{C}(\alpha_{L-1})$. This logic continues as the decomposition is disaggregated and we obtain the following[4]

Theorem 3.5. *Given any cyclic decomposition $\mathcal{P}_0, \ldots, \mathcal{P}_L$, let $\{x\} = \beta^0, \ldots, \beta^{\bar{n}} = \mathcal{X}$ be the sequence (with no repetitions) of all composites in the decomposition that have x as an element. Then*

$$r_{\{x\}} = \sum_{n=1}^{\bar{n}} \max_{\xi \in \Pi(\beta^n)} \underline{C}(\xi) - \underline{C}(\beta^{n-1}). \tag{9}$$

Probably the easiest way to understand the intuition behind Theorem 3.5 is via the tree structure of cyclic decompositions. This tree structure was noted by Catoni (1999) for FW decompositions, but naturally applies to all cyclic decompositions. All this means is that every composite in a decomposition, starting from $\mathcal{P}_L = \{\mathcal{X}\}$, branches into its constituent pieces. At each branching in the tree structure, values of r can be calculated by pairwise comparison of values of $\underline{C}(\cdot)$.

Consider Figure 4, in which we depict the tree structure for the decomposition in Figure 2. At the top of Figure 4, we have composite $\{u, v, w, x, y, z\}$. As this includes the entire state space it must be that $r_{\{u,v,w,x,y,z\}} = 0$. This composite is formed of the cycle of two elements $\{u, v, w\}$ and $\{x, y, z\}$. As $\underline{C}(\{u, v, w\}) > \underline{C}(\{x, y, z\})$, it follows from Lemma 3.1 that $0 = r_{\{u,v,w\}} < r_{\{x,y,z\}}$. This logic continues down the diagram. Simply follow the diagram from the top, choosing the maximum value of $\underline{C}(\cdot)$ at every step. In the example under consideration, this procedure implies that $\{u\}$ is the only composite in $\mathcal{P}_0$ for which $r = 0$, and that therefore $\mu^0_{\{u\}} = 1$.

A similar, though less precise, approach that uses pairwise comparison is the following, that uses a rough lower bound on probability flows into $\{x\}$ and compares it to probability

[4] Theorem 3.5 here is equivalent to Theorem 4.2 of Cui and Zhai (2010) and Theorem 10 of Levine and Modica (2016).

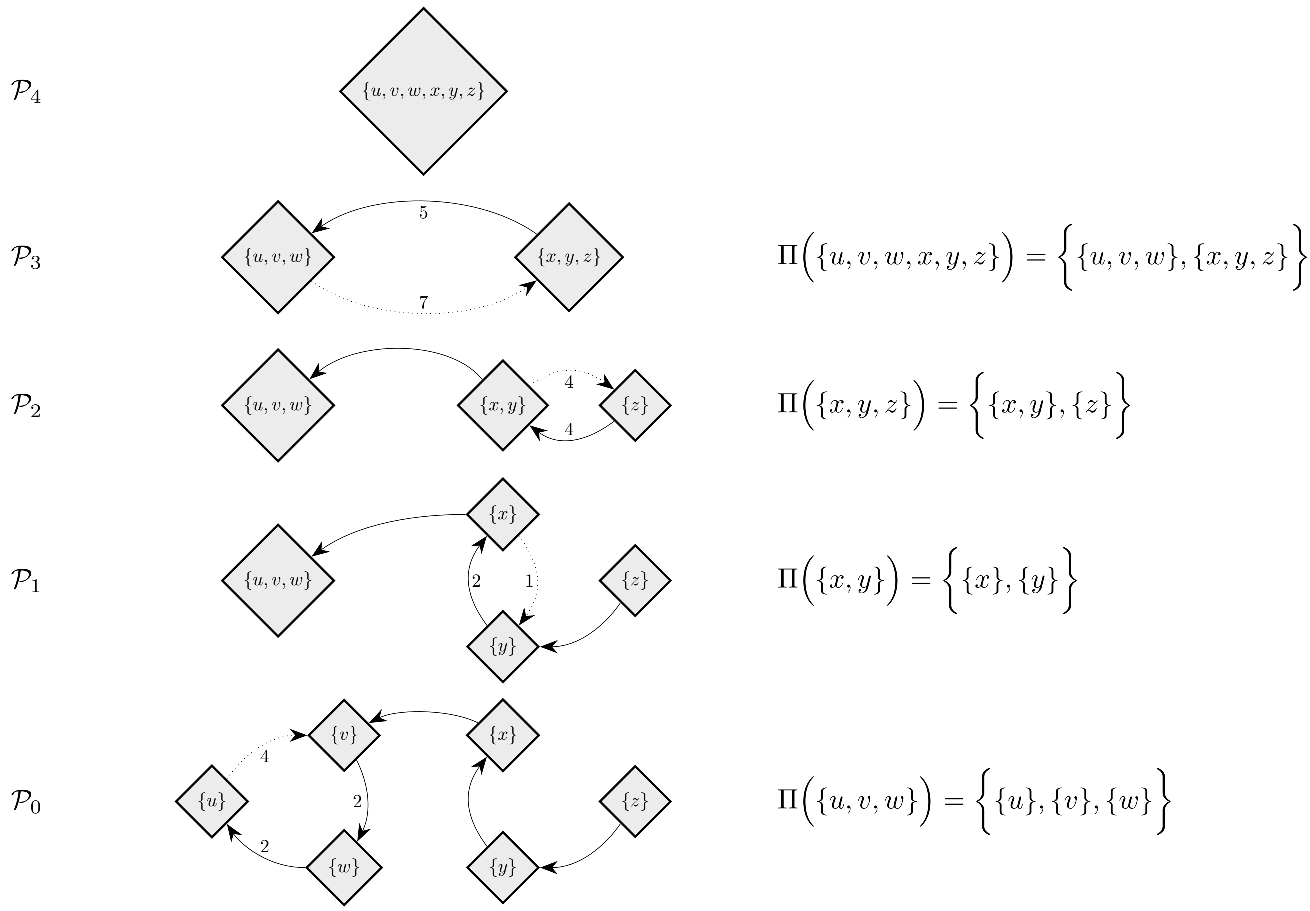


Figure 3: **Disaggregation of states under Edmonds' algorithm.** This diagram illustrates disaggregation and the construction of a spanning tree under Edmonds' algorithm, moving step by step from $\mathcal{P}_4$ to $\mathcal{P}_0$. At each step a composite is expanded into the cycle that formed it during the aggregation phase. For example, $\{x, y, z\} \in \mathcal{P}_3$ is expanded to $\{\{x, y\}, \{z\}\} \subset \mathcal{P}_2$ and edges from $\{x, y\}$ to $\{z\}$ and from $\{z\}$ to $\{x, y\}$ are added. Next, edges to or from the composite that is expanded are assigned to the elements of the cycle that solved (6) or (7) when the cycle was consolidated during the aggregation phase. For example, the edge from $\{x, y, z\}$ to $\{u, v, w\}$ is replaced by an edge from $\{x, y\}$ to $\{u, v, w\}$. Finally, an edge in the expanded cycle is deleted (denoted by a dotted line in the diagram). If the expanded composite was not the root of the tree at the previous step (e.g. $\{x, y, z\}$ in $\mathcal{P}_3$), then the element of the cycle with an outgoing edge to outside of the cycle (e.g. $\{x, y\}$ in $\{\{x, y\}, \{z\}\}$) has its outgoing edge within the cycle deleted (e.g. the edge from $\{x, y\}$ to $\{z\}$). If the expanded composite was the root of the tree at the previous step (e.g. $\{u, v, w\}$ in $\mathcal{P}_1$), then the element of the cycle with the highest least cost transition (e.g. $\{u\}$ in $\{\{u\}, \{v\}, \{w\}\}$) has its outgoing edge deleted (e.g. the edge from $\{u\}$ to $\{v\}$).

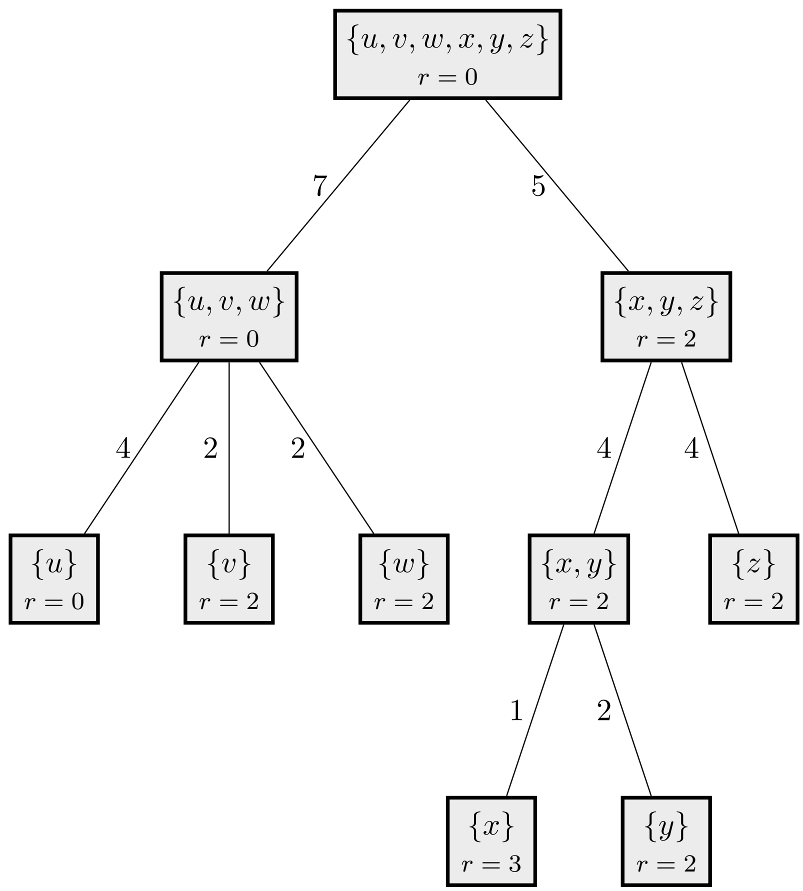


Figure 4: **The tree structure of decompositions.** The tree structure of the decomposition considered in Figure 2. Edges connect composites to their pieces below them in the diagram. Edge weights give $\underline{C}(\cdot)$ for the composite at the lower end of the edge in question. For each composite α, we give r_α which measures the order of magnitude of the probability placed on α by the invariant measure. These are calculated using the values of $\underline{C}(\cdot)$. For example, $\{x, y\}$ has $r_{\{x,y\}} = 2$, therefore by Lemma 3.1, one of its pieces has $r = 2$ and none have $r < 2$. As $\underline{C}(\{x\}) = 1 < 2 = \underline{C}(\{y\})$ and Lemma 2.1 states that $\underline{C}(\{x\}) + r_{\{x\}} = \underline{C}(\{y\}) + r_{\{y\}}$, it must be that $r_{\{x\}} = 3$ and $r_{\{y\}} = 2$.

flows out of $\{x\}$. The result can be proved directly (Ellison, 2000) or as a corollary of Theorem 3.4 (given here).

Theorem 3.6. *If, for some x, for all $y \neq x$ there exists a sequence of composites in $\mathcal{P}_0$, $\{y\} = \beta^0, \ldots, \beta^{\bar{n}} = \{x\}$, such that*

$$\sum_{n=0}^{\bar{n}-1} C(\beta^n, \beta^{n+1}) - \sum_{n=1}^{\bar{n}} \underline{C}(\beta^n) < 0, \tag{10}$$

then $\mu^0_{\{x\}} = 1$.

The proof takes any spanning tree T^y on $\mathcal{P}_0$ rooted at $\{y\}$, $y \neq x$, and constructs a spanning tree T^x rooted at $\{x\}$ by adding edges from β^n to β^{n+1}, $n = 0, \ldots, \bar{n}-1$ and deleting existing edges leaving β^n, $n = 1, \ldots, \bar{n}$. By (10), T^x has a lower sum of edge weights than T^y. Therefore, by Theorem 3.4, it must be that $\mu^0_{\{y\}} = 0$ for all $y \neq x$, and therefore $\mu^0_{\{x\}} = 1$.

Theorem 3.6 is known as the radius/modified-coradius theorem, following the terminology of Ellison (2000). Specifically, inequality (10) can be written as

$$\sum_{n=0}^{\bar{n}-1} C(\beta^n, \beta^{n+1}) - \sum_{n=1}^{\bar{n}-1} \underline{C}(\beta^n) < \underline{C}(\beta^{\bar{n}}). \tag{11}$$

The right hand side of (11) is the *radius* of x. This measures probability flows from x to any other state. The maximum (across $y \neq x$) minimum (across all sequences $\beta^0, \ldots, \beta^{\bar{n}}$) of the first term in the left hand side of (11) is the *coradius* of x. Including the second term, we have the *modified-coradius* of x. These bound (from below) probability flows from all $y \neq x$ into x. These are not always tight bounds, as we see in the example in Figure 2, where the modified-coradius of any given state in $\{u, v, w\}$ is 7, but $\underline{C}(\{x, y, z\}) = 5$, implying that true probability flows from $\mathcal{X} \setminus \{u, v, w\}$ to $\{u, v, w\}$ are significantly higher than the bound given by the modified-coradius.

This concludes the main body of the paper.

Appendix

A. Proofs.

Proof of Lemma 2.1. Consider a cycle $\Gamma_\ell = \bigcup_{m=0}^{\bar{m}}\{\alpha^m\}$. For $0 \le m \le \bar{m}$ and letting $\alpha^{\bar{m}+1} = \alpha^0$, at the invariant measure of the process we must have

$$\mu^\eta_{\alpha^m}\bar{P}^\eta_{\alpha^m\alpha^{m+1}} \le \sum_{\beta\in\mathcal{P}_\ell\setminus\{\alpha^{m+1}\}} \mu^\eta_{\alpha^{m+1}}\bar{P}^\eta_{\alpha^{m+1}\beta} \le \mu^\eta_{\alpha^{m+1}}\left(|\mathcal{P}_\ell|-1\right)\max_{\beta\in\mathcal{P}_\ell\setminus\{\alpha^{m+1}\}}\bar{P}^\eta_{\alpha^{m+1}\beta} \tag{12}$$

The left hand side of (12) is the flow of probability from α^m to α^{m+1}. The central term is the flow of probability from α^{m+1} to all other composites. Applying the transformation $-\eta\log(\cdot)$ to the left and right hand sides of (12) and taking limits as $\eta \to 0$, we obtain

$$r_{\alpha^m} + \underbrace{C(\alpha^m,\alpha^{m+1})}_{=\underline{C}(\alpha_m)} \ge r_{\alpha^{m+1}} + \min_{\beta\in\mathcal{P}_\ell\setminus\{\alpha^{m+1}\}} C(\alpha^{m+1},\beta) = r_{\alpha^{m+1}} + \underline{C}(\alpha^{m+1}). \tag{13}$$

As (13) holds for $n = 1,\dots,m$, we obtain

$$r_{\alpha^1} + \underline{C}(\alpha^1) \ge r_{\alpha^2} + \underline{C}(\alpha^2) \ge \ldots \ge r_{\alpha^{\bar{m}}} + \underline{C}(\alpha^{\bar{m}}) \ge r_{\alpha^1} + \underline{C}(\alpha^1), \tag{14}$$

so all of the inequalities in (14) can be strengthened to equalities, and we are done. □

Proof of Lemma 2.2. Given a partition $\mathcal{P}_\ell$ and least cost correspondence σ_ℓ, define a Markov chain on $\mathcal{P}_\ell$ with transition matrix Q given by

$$Q_{\alpha\beta} = \begin{cases} \frac{1}{|\sigma_\ell(\alpha)|}, & \text{if } \beta \in \sigma_\ell(\alpha), \\ 0, & \text{otherwise.} \end{cases} \tag{15}$$

By definition of σ_ℓ, this Markov chain has no absorbing states. However, as the state space is finite, it must have a set of recurrent classes. Any such recurrent class must have at least two elements, otherwise it would be an absorbing state. Consider one such recurrent class R. To satisfy the definition of a recurrent class, we require

(I) For all $\alpha \in R$, $\beta \notin R$, we have $Q_{\alpha\beta} = 0$.

(II) For all $\alpha,\beta \in R$, there exists a finite sequence $\alpha = \gamma^0,\dots,\gamma^n = \beta$ of elements of R such that $Q_{\gamma^m\gamma^{m+1}} > 0$ for $m = 1,\dots,n-1$.

Condition (II) implies that we can construct a sequence $\gamma^0, \gamma^1, \ldots, \gamma^{\bar{m}}, \gamma^{\bar{m}+1} = \gamma^0$ of elements of R such that $Q_{\gamma^m \gamma^{m+1}} > 0$ for $m = 0, \ldots, \bar{m}$ and $R = \bigcup_{m=0}^{\bar{m}} \gamma^m$. Conversely, the existence of such a sequence trivially implies (II). Using this observation and our definition of $Q_{\alpha\beta}$, Conditions (I) and (II) can be rewritten as

(Ia) R is closed under σ_ℓ.

(IIa) For all $\alpha, \beta \in R$, there exists a finite sequence $\gamma^0, \gamma^1, \ldots, \gamma^{\bar{m}}, \gamma^{\bar{m}+1} = \gamma^0$ of elements of R such that $\gamma^{m+1} \in \sigma_\ell(\gamma^m)$ for $m = 0, \ldots, \bar{m}$ and $R = \bigcup_{m=0}^{\bar{m}} \gamma^m$.

Conditions (Ia) and (IIa) constitute the definition of R being a closed cycle in $\mathcal{P}_\ell$. Consequently, any set of states is a recurrent class if and only if it is a closed cycle in $\mathcal{P}_\ell$.

Therefore, the set of closed cycles of $\mathcal{P}_\ell$ is uniquely determined and, consequently, so is $\mathcal{P}_{\ell+1}$. As $\mathcal{P}_{\ell+1}$ has strictly fewer elements than $\mathcal{P}_\ell$, starting from $\mathcal{P}_0 = \bigcup_{x \in \mathcal{X}} \{\{x\}\}$, the partition $\mathcal{P}_L = \{\mathcal{X}\}$ will be reached in a finite number of steps. □

Proof of Lemma 2.3. Given a partition $\mathcal{P}_\ell$ and least cost correspondence σ_ℓ, we know from the proof of Lemma 2.2 that some (closed) cycle Γ_ℓ exists. If such a cycle is simple, then we let $\Gamma_\ell^S = \Gamma_\ell$. Consider the case in which Γ_ℓ is not simple. Recall that $\Gamma_\ell = \bigcup_{m=0}^{\bar{m}} \{\alpha^m\}$ for some sequence $\alpha^0, \ldots, \alpha^{\bar{m}}$ that satisfies $\alpha^1 \in \sigma_\ell(\alpha^0)$, ..., $\alpha^1 \in \sigma_\ell(\alpha^{\bar{m}})$. Any such sequence must contain repeated elements or else Γ_ℓ would be simple.

Without loss of generality, let $\alpha^{m'} = \alpha^{m''}$ be the first repeated element within such a sequence, so that we have $\alpha^0, \ldots, \alpha^{m'}, \ldots, \alpha^{m''}, \ldots, \alpha^{\bar{m}}$. By construction, the sequence $\alpha^{m'}, \ldots, \alpha^{m''-1}$ then contains no repeated elements and $\Gamma_\ell^S = \bigcup_{m=m'}^{m''-1} \{\alpha^m\}$ satisfies the definition of a simple cycle.

Consolidate Γ_ℓ^S to obtain $\mathcal{P}_{\ell+1}$. As $\mathcal{P}_{\ell+1}$ has strictly fewer elements than $\mathcal{P}_\ell$, starting from $\mathcal{P}_0 = \bigcup_{x \in \mathcal{X}} \{\{x\}\}$, the partition $\mathcal{P}_L = \{\mathcal{X}\}$ will be reached in a finite number of steps. □

Lemma A.1. *Let $A = \{\alpha^0, \ldots, \alpha^m\}$ be a finite set of composites. If $\alpha = \bigcup_{\alpha^m \in A} \alpha^m$, then $r_\alpha = \min_{\alpha^m \in A} r_{\alpha^m}$.*

Proof. As $\alpha = \bigcup_{\alpha^m \in A} \alpha^m$, we have

$$\max_{\alpha^m \in A} \mu^\eta_{\alpha^m} \leq \mu^\eta_\alpha \leq |A| \max_{\alpha^m \in A} \mu^\eta_{\alpha^m}. \tag{16}$$

Applying the transformation $-\eta \log(\cdot)$ and taking limits as $\eta \to 0$,

$$\min_{\alpha^m \in A} r_{\alpha^m} \geq r_\alpha \geq \min_{\alpha^m \in A} r_{\alpha^m}, \tag{17}$$

thus proving the lemma. □

Proof of Lemma 3.1. As $\alpha = \bigcup_{\alpha^m \in \Pi(\alpha)} \alpha^m$, Lemma A.1 implies that $r_\alpha = \min_{\alpha^m \in \Pi(\alpha)} r_{\alpha^m}$. $\square$

Proof of Lemma 3.2. Let $\Pi(\alpha) = \{\alpha^0, \dots, \alpha^{\bar{m}}\}$. Consider the inequalities

$$\max_{\alpha^m \in \Pi(\alpha)} \underbrace{\sum_{\substack{x\in\omega \\ y\in\alpha^m}} P^\eta_{xy} \frac{\mu^\eta_{\{x\}}}{\mu^\eta_\omega}}_{=\bar{P}^\eta_{\omega\alpha^m}} \leq \underbrace{\sum_{\substack{x\in\omega \\ y\in\alpha}} P^\eta_{xy} \frac{\mu^\eta_{\{x\}}}{\mu^\eta_\omega}}_{=\bar{P}^\eta_{\omega\alpha}} \leq |\Pi(\alpha)| \max_{\alpha^m \in \Pi(\alpha)} \underbrace{\sum_{\substack{x\in\omega \\ y\in\alpha^m}} P^\eta_{xy} \frac{\mu^\eta_{\{x\}}}{\mu^\eta_\omega}}_{=\bar{P}^\eta_{\omega\alpha^m}}. \tag{18}$$

Applying the transformation $-\eta\log(\cdot)$ and taking limits as $\eta \to 0$, expression (18) becomes

$$\min_{\alpha^m \in \Pi(\alpha)} C(\omega, \alpha^m) \geq C(\omega, \alpha) \geq \min_{\alpha^m \in \Pi(\alpha)} C(\omega, \alpha^m), \tag{19}$$

thus proving expression (6) in the statement of the lemma.

Now consider the inequalities

$$\max_{\alpha^m \in \Pi(\alpha)} \underbrace{\sum_{\substack{x\in\alpha^m \\ y\in\omega}} P^\eta_{xy} \frac{\mu^\eta_{\{x\}}}{\mu^\eta_\alpha}}_{=\frac{\mu^\eta_{\alpha^m}}{\mu^\eta_\alpha}\bar{P}^\eta_{\alpha^m\omega}} \leq \underbrace{\sum_{\substack{x\in\alpha \\ y\in\omega}} P^\eta_{xy} \frac{\mu^\eta_{\{x\}}}{\mu^\eta_\alpha}}_{=\bar{P}^\eta_{\alpha\omega}} \leq |\Pi(\alpha)| \max_{\alpha^m \in \Pi(\alpha)} \underbrace{\sum_{\substack{x\in\alpha^m \\ y\in\omega}} P^\eta_{xy} \frac{\mu^\eta_{\{x\}}}{\mu^\eta_\alpha}}_{=\frac{\mu^\eta_{\alpha^m}}{\mu^\eta_\alpha}\bar{P}^\eta_{\alpha^m\omega}}. \tag{20}$$

Applying the transformation $-\eta\log(\cdot)$ and taking limits as $\eta \to 0$, the central term becomes $C(\alpha, \omega)$. As $-\eta\log(|\Pi(\alpha)|) \to 0$, the left and right term converge to the same value, which we now compute.

$$\begin{aligned} -\eta\log\left(\max_{\alpha^m \in \Pi(\alpha)} \left(\frac{\mu^\eta_{\alpha^m}}{\mu^\eta_\alpha}\bar{P}^\eta_{\alpha^m\omega}\right)\right) &= \min_{\alpha^m \in \Pi(\alpha)} \left(-\eta\log\left(\frac{\mu^\eta_{\alpha^m}}{\mu^\eta_\alpha}\bar{P}^\eta_{\alpha^m\omega}\right)\right) \\ &= \min_{\alpha^m \in \Pi(\alpha)} \left(-\eta\log\mu^\eta_{\alpha^m} + \eta\log\mu^\eta_\alpha - \eta\log\bar{P}^\eta_{\alpha^m\omega}\right) \end{aligned} \tag{21}$$

Taking limits of (21) as $\eta \to 0$, we obtain

$$\begin{aligned} \min_{\alpha^m \in \Pi(\alpha)} (r_{\alpha^m} - r_\alpha + C(\alpha^m, \omega)) &= \min_{\alpha^m \in \Pi(\alpha)} \left(r_{\alpha^m} - \min_{\alpha^n \in \Pi(\alpha)} r_{\alpha^n} + C(\alpha^m, \omega)\right) \\ &= \min_{\alpha^m \in \Pi(\alpha)} \left(-\underline{C}(\alpha^m) + \max_{\alpha^n \in \Pi(\alpha)} \underline{C}(\alpha^n) + C(\alpha^m, \omega)\right), \end{aligned} \tag{22}$$

where the first equality follows from Lemma 3.1 and the second equality from the implication of Lemma 2.1 that $r_{\alpha^m} + \underline{C}(\alpha^m) = \min_{\alpha^n \in \Pi(\alpha)} r_{\alpha^n} + \max_{\alpha^n \in \Pi(\alpha)} \underline{C}(\alpha^n)$ for all $\alpha^m \in \Pi(\alpha)$.

To see that the latter holds, note that as $\Pi(\alpha)$ is a cycle in $\mathcal{P}_{\ell-1}$, Lemma 2.1 states that $r_{\alpha^m} + \underline{C}(\alpha^m)$ is constant across all $\alpha^m \in \Pi(\alpha)$. Therefore any α^m that minimizes r_{α^m} must also maximize $\underline{C}(\alpha^m)$. This proves expression (7) in the statement of the lemma. □

Lemma A.2. *If*

1. $\Gamma_k^1, \Gamma_k^2, \Gamma_k^3$ *are cycles in* $\mathcal{P}_k$.
2. $\Gamma_k^1 = \{\alpha^0, \dots, \alpha^{\bar{n}}\}$, $\Gamma_k^2 = \{\alpha^0, \dots, \alpha^{n'}\}$, $n' \leq \bar{n}$ *so that* $\Gamma_k^2 \subseteq \Gamma_k^1$.
3. $\Gamma_k^1 \cap \Gamma_k^3 = \emptyset$.
4. $\mathcal{P}_{k+1} = \mathcal{P}_k \cup \{\alpha\} \setminus \Gamma_k^2$.

then

(i) Γ_k^3 *is a cycle in* $\mathcal{P}_{k+1}$.

(ii) If $\Gamma_k^1 \neq \Gamma_k^2$, *then* $\Gamma_{k+1} = \{\alpha, \alpha^{n'+1}, \dots, \alpha^{\bar{n}}\}$ *is a cycle in* $\mathcal{P}_{k+1}$.

Proof. Consider $\omega \in \mathcal{P}_k \setminus \Gamma_k^2$.

By (6), we have that

$$C(\omega, \alpha) = \min_{n \leq n'} C(\omega, \alpha^n). \tag{23}$$

Together with the definitions of the least cost correspondence and cycles, this implies the following two facts.

Fact 1. If $\alpha^n \in \sigma_k(\omega)$ for some $n \leq n'$, then $\sigma_{k+1}(\omega) = \sigma_k(\omega) \cup \{\alpha\} \setminus \Gamma_k^2$.

Fact 2. If $\alpha^n \notin \sigma_k(\omega)$ for all $n \leq n'$, then $\sigma_{k+1}(\omega) = \sigma_k(\omega)$.

Now note that by assumptions (2) and (3) of the lemma, every $\omega \in \Gamma_k^3$ satisfies the condition that $\omega \in \mathcal{P}_k \setminus \Gamma_k^2$. Therefore Facts 1 and 2 apply and $\sigma_{k+1}(\omega)$ are such that Γ_k^3 remains a cycle in $\mathcal{P}_{k+1}$. This proves implication (i) of the lemma.

By (7), we have that

$$\begin{aligned} C(\alpha, \omega) &= \min_{\alpha^n \in \Pi(\alpha)} \left(\max_{\alpha^v \in \Pi(\alpha)} \underline{C}(\alpha^v) - \underline{C}(\alpha^n) + C(\alpha^n, \omega) \right) \\ &= \min_{n \leq n'} \left(\max_{v \leq n'} \underline{C}(\alpha^v) - \underline{C}(\alpha^n) + C(\alpha^n, \omega) \right). \end{aligned} \tag{24}$$

By definition of $\underline{C}(\cdot)$, we have that $-\underline{C}(\alpha^n)+C(\alpha^n,\omega) \geq 0$. Furthermore, $-\underline{C}(\alpha^n)+C(\alpha^n,\omega) = 0$ if and only if $\omega \in \sigma_k(\alpha^n)$. As Γ_k^1 is a cycle in $\mathcal{P}_k$, it must be that this holds for some $\alpha^n \in \Gamma_k^2$ and $\omega = \alpha^{\tilde{n}} \in \Gamma_k^1 \setminus \Gamma_k^2$. Therefore,

Fact 3. $\sigma_{k+1}(\alpha) = \cup_{n=1}^{n'} \sigma_k(\alpha^n) \setminus \Gamma_k^2$.

Consider some sequence of composites that satisfies the definition of Γ_k^1 being a cycle in $\mathcal{P}_k$. Replace every composite α^n, $n \leq n'$ in this sequence with α. Where consecutive instances of α arise as a consequence, replace these with one instance of α. Facts 1 and 2 imply that transitions along this new sequence with α as the destination are least cost transitions in $\mathcal{P}_{k+1}$. Fact 3 implies that transitions along this sequence with α as the starting point are least cost transitions in $\mathcal{P}_{k+1}$. Hence, $\Gamma_{k+1} = \{\alpha, \alpha^{n'+1}, \ldots, \alpha^{\bar{n}}\}$ is a cycle in $\mathcal{P}_{k+1}$. □

Proof of Theorem 3.3. The proof proceeds by induction.

Note that $\mathcal{P}_0^{Ed} = \mathcal{P}_0 = \bigcup_{x \in \mathcal{X}} \{\{x\}\}$.

Assume that for some k and l, we have a partial CLE decomposition $\mathcal{P}_0^{Ed}, \ldots, \mathcal{P}_k^{Ed}$ that contains $\mathcal{P}_0, \ldots, \mathcal{P}_\ell$ as a subsequence and has $\mathcal{P}_k^{Ed} = \mathcal{P}_\ell$. We complete the induction by extending $\mathcal{P}_0^{Ed}, \ldots, \mathcal{P}_k^{Ed}$ via a sequence $\mathcal{P}_k^{Ed}, \mathcal{P}_{k+1}^{Ed}, \ldots, \mathcal{P}_{k+m}^{Ed} = \mathcal{P}_{\ell+1}$.

Extending the sequence

We shall use a further induction. Assume that $\mathcal{P}_k^{Ed}$ has been extended to $\mathcal{P}_k^{Ed}, \ldots, \mathcal{P}_{k'}^{Ed}$ satisfying the following conditions. Firstly, if $\alpha \in \mathcal{P}_\ell$ and $\alpha \in \mathcal{P}_{\ell+1}$, then $\alpha \in \mathcal{P}_{k'}^{Ed}$. Secondly, if $\alpha \notin \mathcal{P}_\ell$ and $\alpha \in \mathcal{P}_{\ell+1}$, then either (i) $\alpha \in \mathcal{P}_{k'}^{Ed}$; or (ii) $\Pi(\alpha)$ is a cycle in $\mathcal{P}_{k'}^{Ed}$. These conditions are trivially satisfied for $k' = k$, as $\mathcal{P}_k^{Ed} = \mathcal{P}_\ell$. Further note that if (ii) does not apply to any α, then $\mathcal{P}_{k'}^{Ed} = \mathcal{P}_\ell$ and we are done. Otherwise, choose some α such that (ii) holds and extend $\mathcal{P}_{k'}^{Ed}$ as follows.

By assumption, there is some cycle, $\Gamma_{k'} = \{\alpha^1, \ldots, \alpha^{\bar{n}}\} \subseteq \mathcal{P}_{k'}^{Ed}$, that consolidates to α.

Step A

By the same argument as in the proof of Lemma 2.3, there must be some (not necessarily unique) simple cycle $\Gamma_{k'}^S \subseteq \Gamma_{k'}$. Assume, w.l.o.g., that $\Gamma_{k'}^S = \{\alpha^1, \ldots, \alpha^{n'}\}$ for some $n' \leq \bar{n}$. Consolidate $\Gamma_{k'}^S$ to composite β^1 to give the partition $\mathcal{P}_{k'+1}^{Ed}$. By Lemma A.2, other than $\Pi(\alpha)$, any other cycles in $\mathcal{P}_{k'}$ to which (ii) above applies remain cycles in $\mathcal{P}_{k'+1}^{Ed}$.

If $\beta^1 = \alpha$ then goto **End**. Else continue to **Step B**.

Step B

By Lemma A.2, $\Gamma_{k'+1} = \{\beta^1, \alpha^{n'+1}, \ldots, \alpha^{\bar{n}}\}$ is a cycle in $\mathcal{P}_{k'+1}^{Ed}$. Repeat **Step A**, inputting $\mathcal{P}_{k'+1}^{Ed}$ instead of $\mathcal{P}_{k'}^{Ed}$

End

This completes the inductive step of the second induction (**Extending the sequence**), which completes the inductive step of the first induction. ◻

Proof of Theorem 3.4. Lemma A.1 implies that $r_\alpha = \min_{x\in\alpha} r_{\{x\}}$. Therefore, $r_\alpha = 0$ if and only if there exists $x \in \alpha$ such that $r_{\{x\}} = 0$. Consider $x \in \alpha$ and a CLE decomposition $\mathcal{P}_0, \ldots, \mathcal{P}_L$. Let $\{x\} = \beta^0, \ldots, \beta^{\bar{n}} = \mathcal{X}$ be the sequence (with no repetitions) of all composites in the decomposition that have x as an element. We consider the disaggregation stage of Edmonds' algorithm as described in Section 3.1.2, in particular step (4b). This allows us to show that $r_{\{x\}} = 0$ implies that $\{x\}$ is the root of a minimal spanning tree on $\mathcal{P}_0$, whereas $r_{\{x\}} > 0$ implies that $\{x\}$ is not the root of any minimal spanning tree on $\mathcal{P}_0$.

Consider $r_{\{x\}} = 0$. Lemma 3.1 implies that $r_{\beta^n} = 0$ for $n = 0, \ldots, \bar{n}$. Hence, for $n = 1, \ldots, \bar{n}$, we have that β^{n-1} solves $\min_{\beta\in\Pi(\beta^n)} r_\beta$. Lemma 2.1 then implies that β^{n-1} solves $\max_{\beta\in\Pi(\beta^n)} \underline{C}(\beta)$. Therefore, if β^n, $n > 0$, $d(\beta^n) = \ell$ is a root of T_ℓ, then β^{n-1} may be chosen to have its outgoing edge deleted when β^n is disaggregated. When this is the case, β^{n-1} is then a root of $T_{\ell-1}$. As this argument applies for $n = \bar{n}, \bar{n}-1, \ldots, 1$, we can thereby obtain T_0 that is rooted at $\{x\}$. Chu and Liu (1965) and Edmonds (1967) proved that T_0 has minimum sum of edge weights amongst all spanning trees on $\mathcal{P}_0$.

Now consider $r_{\{x\}} > 0$. As $r_{\alpha_L} = r_{\mathcal{X}} = 0$, Lemma 3.1 implies that there exists $n > 0$ such that $r_{\beta^{n-1}} > 0$ and $r_{\beta^n} = 0$. Lemma 3.1 further implies that there exists $\beta^* \in \Pi(\beta^n)$ such that $r_{\beta^*} = 0$. Therefore, β^{n-1} does not solve $\min_{\beta\in\Pi(\beta^n)} r_\beta$. Lemma 2.1 then implies that β^{n-1} does not solve $\max_{\beta\in\Pi(\beta^n)} \underline{C}(\beta)$. Therefore, if $d(\beta^n) = \ell$ and β^n is a root of T_ℓ, then β^{n-1} will never be chosen to have its outgoing edge deleted when β^n is disaggregated. Hence β^{n-1} is never a root of $T_{\ell-1}$. Consequently, no tree T_0 with minimum sum of edge weights amongst all spanning trees on $\mathcal{P}_0$ is rooted at $\{x\}$. ◻

Proof of Theorem 3.5. For $n = 1, \ldots, \bar{n}$, by Lemma 2.1 we have

$$r_{\beta^{n-1}} + \underline{C}(\beta^{n-1}) = \min_{\xi\in\Pi(\beta^n)} r_\xi + \max_{\xi\in\Pi(\beta^n)} \underline{C}(\xi). \tag{25}$$

Rearranging (25), we obtain

$$r_{\beta^{n-1}} - \min_{\xi\in\Pi(\beta^n)} r_\xi = \max_{\xi\in\Pi(\beta^n)} \underline{C}(\xi) - \underline{C}(\beta^{n-1}). \tag{26}$$

Summing (26) over $n = 1, \ldots, \bar{n}$ and cancelling terms on the left hand side using Lemma 3.1,

$$(27)\qquad \underbrace{r_{\beta^0}}_{=r_{\{x\}} \text{ as } \beta^0=\{x\}} - \underbrace{\min_{\xi\in\Pi(\beta^{\bar{n}})} r_\xi}_{\substack{=r_{\beta^{\bar{n}}} \text{ by Lemma 2.1} \\ =0 \text{ as } \beta^{\bar{n}}=\mathcal{X}}} = \sum_{n=1}^{\bar{n}} \max_{\xi\in\Pi(\beta^n)} \underline{C}(\xi) - \underline{C}(\beta^{n-1}).$$

□

Proof of Theorem 3.6. Let x, y, $\{y\} = \beta^0, \ldots, \beta^{\bar{n}} = \{x\}$ be as in the theorem statement. Assume that $\mu^0_{\{y\}} > 0$. By Theorem 3.4, there exists a spanning tree T^y on $\mathcal{P}_0$ with minimum sum of edge weights that has $\{y\}$ as its root. Obtain a new graph T^x from T^y by deleting the existing outgoing edges from vertices $\beta^1, \ldots \beta^{\bar{n}}$ and adding edges from β^n to β^{n+1} for $n = 0, \ldots, \bar{n} - 1$. T^x is a spanning tree on $\mathcal{P}_0$ rooted at $\{x\}$ that, by (10), has lower sum of edge weights than the tree rooted at $\{y\}$. Contradiction. □

References

Arieli, I., Babichenko, Y., Peretz, R., and Young, H. P. (2019). The speed of innovation diffusion in social networks. *Econometrica*.

Arieli, I. and Young, H. P. (2016). Stochastic learning dynamics and speed of convergence in population games. *Econometrica*, 84(2):627–676.

Catoni, O. (1999). Simulated annealing algorithms and Markov chains with rare transitions. In Azéma, J., Émery, M., Ledoux, M., and Yor, M., editors, *Séminaire de Probabilités XXXIII*, pages 69–119. Springer, Berlin.

Chu, Y.-J. and Liu, T.-H. (1965). On the shortest arborescence of a directed graph. *Science Sinica*, 14:1396–1400.

Cui, Z. and Zhai, J. (2010). Escape dynamics and equilibria selection by iterative cycle decomposition. *Journal of Mathematical Economics*, 46(6):1015 – 1029.

Edmonds, J. (1967). Optimum branchings. *Journal of Research of the National Bureau of Standards*, 71:233–240.

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

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

Foster, D. P. and Young, H. P. (1990). Stochastic evolutionary game dynamics. *Theoretical Population Biology*, 38:219–232. Corrigendum, 51 (1997), 77-78.

Freidlin, M. I. and Wentzell, A. D. (1984). *Random Perturbations of Dynamical Systems*. Springer, New York.

Kandori, M., Mailath, G. J., and Rob, R. (1993). Learning, mutation, and long run equilibria in games. *Econometrica*, 61:29–56.

Levine, D. K. and Modica, S. (2016). Dynamics in stochastic evolutionary models. *Theoretical Economics*, 11(1):89–131.

Newton, J. (2018). Evolutionary game theory: A renaissance. *Games*, 9(2):31.

Newton, J. and Angus, S. D. (2015). Coalitions, tipping points and the speed of evolution. *Journal of Economic Theory*, 157(0):172 – 187.

Newton, J. and Sawa, R. (2015). A one-shot deviation principle for stability in matching problems. *Journal of Economic Theory*, 157(0):1 – 27.

Norman, T. W. (2009). Rapid evolution under inertia. *Games and Economic Behavior*, 66(2):865–879.

Sandholm, W. H. (2010). *Population Games and Evolutionary Dynamics*. MIT Press, Cambridge.

Young, H. P. (1993). The evolution of conventions. *Econometrica*, 61:57–84.

Young, H. P. (2011). The dynamics of social innovation. *Proceedings of the National Academy of Sciences*, 108 Suppl 4:21285–91.