Contagion, network cohesion and long run stability in evolutionary games

Daniel C. Opolot

School of Economics, University of Cape Town

Abstract

This paper studies how the network structure affects the long-run equilibria emerging in coordination games when agents are myopic best responders. Our analysis builds-on the properties of the process of contagion. We demonstrate that when contagion is feasible, the *network diameter*, a measure of the cohesiveness of the whole network, determines the uniqueness of long-run equilibria. The *maximum group cohesion* is one of the network measures that determines the feasibility of contagion. We show that for cyclic networks where each player has the same number of neighbours, there exists a threshold network diameter above which strategies in the smallest *iterated p-best response set*, for p equal to the maximum group cohesion, are uniquely stochastically stable. We discuss how these results can be extended to evolutionary dynamics in arbitrary networks using different network measures that determine the feasibility of contagion.

JEL codes: C72, C73, D83, D85.

Keywords: Evolutionary dynamics, networks, stochastic stability, iterated p-best response sets.

1. Introduction

Many game theoretic models have multiple equilibria. A well-known example is the coordination game which arises when players benefit from coordinating their activities by making the same decisions.[1] When a game contains multiple equilibria, there is no obvious way of knowing which equilibrium players will settle on. This creates challenges for researchers working on policy-related questions and experimentalists aiming to test the underlying behavioural assumptions of game theoretic models.

[*] *Postal address*: Private Bag X3, Rondebosch, 7701 Cape Town
Email address: opodanchris@gmail.com

[1] Coordination games find many applications in economics and other social sciences. A few examples include the choice of common technology and legal standards, modelling of social norms, speculative currency attacks, and political actions such as protests and tactical voting.

The models of social learning (i.e. where individuals adjust their behaviour over time by observing and learning from others' actions) with persistent randomness have become useful tools for selecting among multiple equilibria. In these models, although all outcomes are likely, in the long run, some outcomes are more likely than others (Foster and Young, 1990; Kandori et al., 1993; Young, 1993). Outcomes that retain a positive probability of being played in the long-run as the amount of noise vanishes are said to be *stochastically stable*. Stochastic stability as a solution concept selects fewer and, in some cases, unique outcomes. However, recent studies show that the selected outcomes depend on the interaction structure (Alós-Ferrer and Weidenholzer, 2007; Peski et al., 2010).[2] The mechanisms through which the interaction structure determines stochastically stable outcomes are not well-understood.

This paper shows that *contagion* – the spread of choices from a small group of players to the whole population – is one of the mechanisms that determines the robustness of stochastic stability to the interaction structure. We consider a stochastic evolutionary model of best response with mutations (BRM) proposed by Young (1993) and Ellison (2000). In this model, players revise their strategies over time by choosing best responses to their opponents' strategies and with a fixed small probability, they randomly choose strategies that are not best responses. These random trembles are justified in the literature as either pure experimentation or bounded rationality on the part of players. To demonstrate how contagion matters, we focus on evolutionary dynamics in symmetric coordination games played on *regular cyclic networks* (i.e. cyclic networks where each player has the same number of neighbours). When the number of neighbours is odd, cyclic networks exhibit an asymmetry that can be used to compare different network structures.

Contagion affects evolutionary dynamics, and hence, stochastically stable outcomes, in three ways. First, when contagion is feasible, a subset of strategies can spread from a small group of players to the whole network through best response dynamics. *Group cohesion* is one of the network measures that determines when and which strategies can spread contagiously. A group of players in a given network is η-cohesive if every player in that group has at least proportion η of her interactions within the group. The maximum group cohesion, $\eta(G)$, of a network, G, is the maximum cohesiveness of any group of players. This definition of group cohesion ensures that a *p-best response set* of strategies of a coordination game (i.e. strategies that are best

[2] One of the implications of these findings is that the modeller must keep track of the identity of agents when computing stochastically stable outcomes. Unfortunately, the time complexity of stochastic stability algorithms increases exponentially with the population size. It is thus important to identify some properties of the interaction structure that can be used to predict stochastically stable outcomes.

responses to any mixture that places on them a mass of at least p) will spread contagiously from any c-cohesive group of players, for $c \geq \frac{1}{2}$, to the whole network whenever p is less or equal to $1 - \eta(G)$.

Second, when p is less or equal to $1 - \eta(G)$, strategies in the *smallest iterated p-best response set*, A^* (i.e. a set of strategies that remains after iteratively applying the principle of a p-best response set) spread from any c-cohesive group of players, for $c \geq \frac{1}{2}$, to the whole network through the process of *step-by-step contagion* (i.e. iterative application of contagion).[3] Step-by-step contagion ensures that the *net cost*, measured in terms of the number of mutations, of reaching an absorbing set of states containing only strategies in A^* is bounded from above by the size of the smallest c-cohesive group.

Third, contagion ensures that the cost of leaving the basin of attraction of an absorbing set of states containing only strategies in A^* depends on the *cohesiveness* of the network. *Network cohesion*, in contrast to group cohesion, measures the reachability between any pair of players (i.e. the shortest distances between pairs of players). The *network diameter*, which is the maximum-shortest distance between any pair of players, sufficiently captures the cohesiveness of the whole network. We show that whenever contagion is feasible, there exists a threshold network diameter above which the cost of leaving the basin of attraction of an absorbing set of states containing only strategies in A^* is larger than the net cost of reaching it.

These three observations then imply that, for regular cyclic networks, there exists a threshold network diameter above which strategies in the smallest iterated p-best response set, for $p \leq 1 - \eta(G)$, are uniquely stochastically stable.[4] When these conditions do not hold, it is easy to construct an example of a coordination game and network where strategies that are not in the smallest iterated p-best response set (e.g. a payoff dominant strategy) are stochastically stable.

Although we focus on regular cyclic networks, the principles underlying the relationship between contagion and stochastic stability extend to arbitrary networks. In Section 6, we present three network measures that determine the feasibility of contagion in *strongly connected* arbi-

[3] Specifically, a set of strategies A^T is step-by-step contagious if, starting from any state $\mathbf{x}$, there exists a sequence of strategy sets $A^1, A^2, \cdots, A^T$ with corresponding absorbing sets $\mathbf{A}^1, \mathbf{A}^2, \cdots, \mathbf{A}^T$ containing only strategies in $A^1, A^2, \cdots, A^T$ respectively, such that strategies in A^1 spread contagiously starting from $\mathbf{x}$; strategies in A^2 spread contagiously starting from any state in $\mathbf{A}^1$; strategies in A^3 spread contagiously starting from any state in $\mathbf{A}^2$; $\cdots$; and strategies in A^T spread contagiously starting from any state in $\mathbf{A}^{T-1}$.

[4] This conclusion follows from Ellison (2000, Theorem 2), which states that if the net cost of reaching the basin of attraction of an absorbing set of states containing only strategies in some set A^* is strictly smaller than the cost of leaving it, then strategies in A^* are uniquely stochastically stable.

trary networks (i.e. arbitrary networks where every two players are connected by some path): *the maximum degree*, which is the largest number of direct neighbours of any player; the *neighbourhood contagion threshold*, defined as some number α such that for all players at any distance from the *closed 1st neighbourhood* of any player (i.e. a set of direct neighbours of any player i, with i included), at least proportion α of their neighbours are closer to that 1st neighbourhood; the *subgroup contagion threshold*, defined as some number α such that for all players at any distance from any c-cohesive group of players, for $c \geq \frac{1}{2}$, at least proportion α of their neighbours are closer to that group. For all these three network measures, we demonstrate that when contagion is feasible, the network diameter determines when strategies in the smallest iterated p-best response set are uniquely stochastically stable.

Except for Alós-Ferrer and Weidenholzer (2007), many papers on evolutionary dynamics with best response focus on examining conditions under which a p-dominant strategy (i.e. a strategy that is a best response to any mixture that places on it a mass of at least p) is stochastically stable.[5] Young (1993) and Maruta (1997) show that a $\frac{1}{2}$-dominant strategy is stochastically stable in complete networks (i.e. where each player interacts with every other player); Blume (1995), Young (1998), Lee and Valentinyi (2000) and Lee et al. (2003) show that a risk-dominant strategy (i.e. a $\frac{1}{2}$-dominant strategy of a 2×2 coordination game) is uniquely stochastically stable in a 2-dimensional grid network; Ellison (1993, 2000), Weidenholzer (2012) and Jiang and Weidenholzer (2017) show that a $\frac{1}{2}$-dominant strategy is uniquely stochastically stable in regular cyclic networks where each player has $2k$ neighbours; Alós-Ferrer and Weidenholzer (2008) show that a $\frac{1}{\Delta(G)}$-dominant strategy is uniquely stochastically stable in network G with $\Delta(G)$ as the maximum degree; Peski et al. (2010) shows that a p-dominant strategy is uniquely stochastically stable in network G with $\delta_0(G)$ as the smallest odd degree if $p \leq \frac{1}{2}\left(1 - \frac{1}{\delta_0(G)}\right)$; and Opolot (2018) shows that strategies in a p-best response set are uniquely stochastically stable in network G with $\bar{\eta}(G)$ as the contagion threshold if $p \leq \bar{\eta}(G)$. A p-dominant strategy and a p-best response set are special cases of the smallest iterated p-best response set. The present paper thus generalizes most of the results from the aforementioned papers. Moreover, our analysis provides a framework for broadly predicting stochastically stable outcomes.

Alós-Ferrer and Weidenholzer (2007) consider a BRM model in 3×3 coordination games

[5] There is a parallel literature that studies long-run stability in evolutionary models of network formation (i.e. where players can re-wire their connections to maximize payoffs) (Robson and Vega-Redondo, 1996; Jackson and Watts, 2002; Goyal and Vega-Redondo, 2005; Alós-Ferrer and Weidenholzer, 2008; Man, 2012; Staudigl and Weidenholzer, 2014; Bilancini and Boncinelli, 2020). These papers find that a payoff dominant strategy is stochastically stable.

played on a cyclic network where each player has two neighbours. They show that a *globally pairwise risk-dominant strategy* (i.e. a strategy that is pairwise risk-dominant relative to every other strategy) of a 3×3 symmetric coordination game is stochastically stable if it satisfies the partial bandwagon property (PBP). A pair of strategies $\{a_j, a_l\}$ satisfies the PBP if the best responses to any mixture that place all its mass on $\{a_j, a_l\}$ is within $\{a_j, a_l\}$. We demonstrate in Section 4 that if a globally pairwise risk-dominant strategy satisfies the PBP, then it is the smallest iterated $\frac{1}{2}$-best response equilibrium of a 3×3 symmetric coordination game. Thus, our results partly extend the findings in Alós-Ferrer and Weidenholzer (2007) to all cyclic networks and symmetric coordination games with more than 3 strategies.

Finally, this paper is related to the literature on contagion in networks. Morris (2000) considers a model of best response dynamics on unbounded networks and shows that a p-dominant strategy of a 2×2 coordination game is contagious if p is less or equal to the contagion threshold. Alós-Ferrer and Weidenholzer (2008) consider a model of imitation dynamics and show that, with some restrictions on the information structure, a Pareto-dominant equilibrium of a 2×2 coordination game is contagious in strongly connected networks. Oyama and Takahashi (2015) consider a model of best response dynamics on unbounded networks and first establish conditions under which the risk-dominant and Pareto-dominant strategies of a 3×3 coordination game containing a dominated strategy are contagious; they then proceed to compare networks in terms of their power of inducing contagion in general supermodular games. Azomahou and Opolot (2018) consider a model of best response dynamics in finite networks and show that a p-dominant strategy is contagious whenever p is less or equal to the (neighbourhood) contagion threshold. Our analysis builds on the concepts of contagion developed in these papers and uses it to establish conditions under which the smallest iterated p-best response set is uniquely stochastically stable in a BRM model.

The remainder of the paper is organized as follows. Section 2 outlines an evolutionary model of best response with mutations. Section 3 presents the concepts of contagion and network cohesion and discusses how they affect evolutionary dynamics. The main results are presented in Sections 4. Section 5 discusses the main results and how they are related to the literature, and Section 6 discusses potential ways of extending our results to arbitrary networks. Concluding remarks are offered in Section 7 and lengthy proofs are contained in the Appendix.

2. A model of stochastic evolution

We consider a finite set of players, $N = \{1, 2, \cdots, i, \cdots, n\}$, connected through a network, where each player plays a given coordination game with her neighbours.

2.1. Network structure

We aim to show how the interplay between contagion and network cohesion determines the unique stochastically stable outcomes in evolutionary models. For this purpose, it is enough to focus on regular but not necessarily symmetric networks. A network is defined by a graph, $G(N, E)$, where E is the set of edges connecting players in N. For each pair of players $i, j \in N$, an edge, $i \to j$, from i to j implies that i observes j's actions. Let N_i denote the set of i's neighbours, and write n_i for the cardinality of N_i, also commonly known as the *degree* of i.

We consider *undirected cyclic networks* where each player has k neighbours (we discuss how our results can be extended to arbitrary networks in Section 6). A network is undirected if the existence of an edge $i \to j$, implies the existence of a reverse edge $j \to i$. An undirected cyclic network, $G_k(n)$, with degree k and size n has the following properties:

(i) when k is an even number, each $i \in N$ has $\frac{k}{2}$ adjacently placed neighbours to the left and $\frac{k}{2}$ to the right;

(ii) when k is odd, each $i \in N$ has either $\lceil k/2 \rceil$ adjacently placed neighbours to the left and $\lfloor k/2 \rfloor$ neighbours to the right, or vice versa, where $\lceil x \rceil$ is the least integer greater than or equal to x and $\lfloor x \rfloor$ is the greatest integer less than or equal to x.

We write $\mathcal{G}_k$, for $k = 2, 3, \cdots, n-1$, for the set of all undirected cyclic networks where each player has degree k.

2.2. Coordination games

For each pairwise interaction between two neighbours, we consider a *symmetric strict coordination game*. A coordination game consists pf two components, a set of pure strategies, $A = \{a_1, \cdots, a_j, \cdots, a_m\}$, and a payoff matrix M. Each element, m_{jl}, of M is a payoff to a player playing strategy a_j against an opponent playing strategy a_l. Let Σ be the set of all mixed strategies over A so that, for any $\sigma \in \Sigma$, $\sigma(a_j)$ is the mass that σ places on a_j. We consider linear payoffs, where the payoff of a pure strategy a_j against a mixture σ is $U(a_j \mid \sigma) = \sum_{a_k \in A} \sigma(a_l) m_{jl}$.

A mapping $U : \Sigma \to \mathbb{R}^m$, where m is the size of the strategy set and $U(\sigma) = (U(a_1 \mid \sigma), \cdots, U(a_m \mid \sigma))$, is a coordination game if $m_{jj} \geq m_{jl}$ for all $a_j, a_l \in A$ and $a_l \neq a_j$. It is a symmetric coordination game if m_{jl}, for all $a_j, a_l \in A$, is identical for both players. We focus on (symmetric) *strict coordination games* where $m_{jj} > m_{jl}$ for all $a_j, a_l \in A$ and $a_l \neq a_j$. For each $\sigma \in \Sigma$, we write $BR(\sigma)$ for the set of pure strategy best responses to σ. That is,

$$BR(\sigma) = \{a_j \in A \mid U(a_j \mid \sigma) \geq U(a_l \mid \sigma)\, \forall a_l \in A\}.$$

2.3. Unperturbed evolutionary process

We consider an evolutionary process where players simultaneously revise their strategies over time by choosing best responses to strategy distributions in their neighbourhoods. Let $\mathbf{x}$ denote the profile of strategies, and let x^i be the strategy i plays in profile $\mathbf{x}$. Each strategy profile is a state of an evolutionary process, and we denote the set of all states by $\mathbf{X}$. For each $\mathbf{x} \in \mathbf{X}$, let $\sigma_i(a_l; \mathbf{x})$ be the proportion of i's neighbours playing strategy a_l in profile $\mathbf{x}$, and let $\sigma_i(\mathbf{x}) = (\sigma_i(a_1; \mathbf{x}), \cdots, \sigma_i(a_m; \mathbf{x}))$ be the distribution of strategies in i's neighbourhood under strategy profile $\mathbf{x}$. The payoff of a_j against a mixture $\sigma_i(\mathbf{x})$ is then $U\Big(a_j \mid \sigma_i(\mathbf{x})\Big) = \sum_{a_l \in A} \sigma_i(a_l; \mathbf{x}) m_{jl}$, and $BR(\sigma_i(\mathbf{x}))$ is the set of best responses to $\sigma_i(\mathbf{x})$.

Let $\mathbf{x}(t)$ be the strategy profile at time t, and $x^i(t)$ the respective ith strategy in profile $\mathbf{x}(t)$. We consider a *myopic best response dynamics*, where at time $t+1$ player i chooses $x^i(t+1) \in BR(\sigma_i(\mathbf{x}(t)))$ with probability $\frac{1}{|BR(\sigma_i(\mathbf{x}(t)))|}$, where $|S|$ is the cardinality of set S. The assumption of myopia is standard in the literature of evolutionary game theory and it captures the idea that economic agents are incapable of keeping track of the entire history of play and performing complex evaluations associated with forward-looking decision making. This behavioural assumption is a departure from the traditional assumption of a rational forward looking economic agent.

This dynamic process leads to a finite Markov chain on the state space $\mathbf{X}$. The associated transition matrix, here denoted by P, is *homogeneous* since we assume that the network and payoff matrix do not change over time. Given P, we write $P(\mathbf{x}, \mathbf{y})$ for the probability of transiting from state $\mathbf{x}$ to $\mathbf{y}$ in a single step. We refer to the quadruple (U, N, G, P) as the *unperturbed evolutionary process* on network G.

For an unperturbed process (U, N, G, P), a set $W \subseteq \mathbf{X}$ is *absorbing* if once entered, is never exited.[6] If W is a singleton set, then it is an *absorbing state*, and if it is non-singleton, then it is an *absorbing cycle*. For example, if $W = \{\mathbf{x}, \mathbf{y}\}$ forms an absorbing cycle, then $P(\mathbf{x}, \mathbf{y}) = 1$ and $P(\mathbf{y}, \mathbf{x}) = 1$, so that once (U, N, G, P) enters W, it cycles between $\mathbf{x}$ and $\mathbf{y}$ indefinitely. We write $\mathbf{A}$ for the set of all absorbing sets of (U, N, G, P).

2.4. Perturbed evolutionary process

Following the literature the model is completed by adding the possibility of rare mutations to players' choices. We consider a model of best response with mutations (BRM) introduced in

[6] That is, $W \subseteq \mathbf{X}$ is an *absorbing set* of (U, N, G, P) if for all $\mathbf{y} \in W$, the probability $\mathbb{P}\left(\mathbf{x}(t+1) \in W \mid \mathbf{x}(t) = \mathbf{y}\right) = 1$, and that for all $\mathbf{y}, \mathbf{z} \in W$, there exists $\tau > 0$ such that $P\left(\mathbf{x}(t+\tau) = \mathbf{z} \mid \mathbf{x}(t) = \mathbf{y}\right) > 0$.

Young (1993), Kandori et al. (1993) and Ellison (2000). In this framework, players experiment and choose strategies that are not best responses at random with a fixed small probability ε, independent across players and across time. That is, with probability $1-\varepsilon$ a player chooses a best response and with probability ε chooses a strategy at random from a uniform distribution over A. These random trembles are justified as either experimentation by players (Newton, 2012), or, when combined with myopia, as bounded rationality on the part of players.

Specifically, let $c(\mathbf{x}, \mathbf{y})$ be the number of mutations involved in direct transition from $\mathbf{x}$ to $\mathbf{y}$. That is, the number of players whose strategies in state $\mathbf{y}$ are different from those played in $\mathbf{x}$, and that their choices in $\mathbf{y}$ are not best responses to $\mathbf{x}$. Let $|BR(\sigma_i(\mathbf{x}))| = b_i(\mathbf{x})$. Then the probability, $P_\varepsilon(\mathbf{x}, \mathbf{y})$, that the perturbed process transits from state $\mathbf{x}$ to $\mathbf{y}$ in a single period is given by[7]

$$P_\varepsilon(\mathbf{x}, \mathbf{y}) = \left(\frac{\varepsilon}{m}\right)^{c(\mathbf{x},\mathbf{y})} \prod_{i=1}^{n-c(\mathbf{x},\mathbf{y})} \left(1 - \frac{m - b_i(\mathbf{x})}{m b_i(\mathbf{x})}\varepsilon\right) \tag{1}$$

For the remainder of the paper, we write P_ε for a homogeneous transition matrix with transition probabilities described by (1), and refer to (U, N, G, P_ε) as the *perturbed evolutionary process*, or an *evolutionary process of best response with mutations* (BRM).

We are interested in characterizing the long run behaviour of (U, N, G, P_ε). Since (U, N, G, P_ε) is a Markov chain, its long-run behaviour is described by the *invariant distribution* – the probability distribution over the state space that describes the long run average time spent in each state. Let $\Sigma(\mathbf{X})$ be the set of all probability distributions over $\mathbf{X}$. The invariant distribution of (U, N, G, P_ε), $\pi_\varepsilon = \lim_{t\to\infty} \nu P_\varepsilon^t$, for any $\nu \in \Sigma(\mathbf{X})$, exists and is unique. This is because, for any fixed $\varepsilon > 0$, each state of (U, N, G, P_ε) is reachable from every other state (i.e. (U, N, G, P_ε) is irreducible and ergodic).

The long stable states of (U, N, G, P_ε) are those that maximize π_ε. But since computing π_ε is difficult, the standard approach in the literature of evolutionary games is to focus on the *limit invariant distribution* π^* defined by $\pi^* = \lim_{\varepsilon\to 0} \pi_\varepsilon$. The motivation for this is that π^*, which is easier to compute, provides an approximation to π_ε when ε is small. The limit invariant distribution exists and the set of states in the support of π^*, $\{\mathbf{x} \in \mathbf{X} |\ \pi^* > 0\}$, is called a *stochastically stable set* (Young, 1993; Ellison, 2000). The stochastically stable set of

[7]This follows because the probability that a player chooses a strategy that is not a best response is $\frac{\varepsilon}{m}$. That is, with probability ε a player mutates, and when a mutation occurs, a player chooses uniformly randomly from the set of strategies. Thus, the probability that $c(\mathbf{x}, \mathbf{y})$ players simultaneously mutate to play strategies chosen in $\mathbf{y}$ is $\left(\frac{\varepsilon}{m}\right)^{c(\mathbf{x},\mathbf{y})}$. The complementary number of players, $n - c(\mathbf{x}, \mathbf{y})$, follow best response, and hence, choose strategies in $\mathbf{y}$ with probability $\frac{\varepsilon}{m} + (1-\varepsilon)\frac{1}{b_i(\mathbf{x})} = 1 - \frac{m-b_i(\mathbf{x})}{m b_i(\mathbf{x})}\varepsilon$.

(U, N, G, P_ε) will be contained in $\mathbf{A}$, the set of all absorbing sets of (U, N, G, P) (Young, 1993). We refer to the set of strategies played in a stochastically stable set as *stochastically stable strategies*.

To compute the stochastically stable set of (U, N, G, P_ε), we use the *radius-coradius* approach introduced by Ellison (2000). Let $D(W)$ be the basin of attraction of $W \subseteq \mathbf{A}$, the set of initial states from which (U, N, G, P) converges to W with probability one. That is,

$$D(W) = \left\{ \mathbf{y} \in \mathbf{X} \mid \mathbb{P} \left(\exists t' \text{ such that } \mathbf{x}(t) \in W \; \forall t > t' \mid \mathbf{x}(0) = \mathbf{y} \right) = 1 \right\}$$

The radius of a basin of attraction of W, $R(W)$, is the minimum number of mutations needed to exit $D(W)$ when (U, N, G, P) starts from any $\mathbf{x} \in W$. Specifically, define a *path* from W to some $Z \neq W$ as the finite sequence of distinct states $(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_T)$ with $\mathbf{x}_1 \in W$ and $\mathbf{x}_T \in Z$, and that $\mathbf{x}_\tau \notin Z$ for $2 \leq \tau \leq T-1$. Let $S(W, Z)$ be the set of all paths from W to Z. The cost $c(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_T)$ of the path $(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_T)$ is defined as

$$c(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_T) = \sum_{\tau=1}^{T-1} c(\mathbf{x}_\tau, \mathbf{x}_{\tau+1})$$

where $c(\mathbf{x}_\tau, \mathbf{x}_{\tau+1})$ is defined in (1). Let $C(W, Z) = \min_{(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_T) \in S(W,Z)} c(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_T)$, which is the cost of a minimum path from W to Z. Then

$$R(W) = \min_{(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_T) \in S(W, \mathbf{X} - D(W))} c(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_T) = C(W, \mathbf{X} - D(W)).$$

The *coradius* of the basin of attraction of W, $CR(W)$, is the cost of reaching $D(W)$ from any other state. That is, $CR(W) = \max_{\mathbf{x} \notin D(W)} \min_{(\mathbf{x}_1, \cdots, \mathbf{x}_T) \in S(\mathbf{x}, D(W))} c(\mathbf{x}_1, \cdots, \mathbf{x}_T)$.

Let $W_1, W_2, \cdots, W_H$, with $W_H \subseteq W$, be the sequence of absorbing sets through which the path $(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_T) \in S(\mathbf{x}, W)$ passes consecutively, where $\mathbf{x} = \mathbf{x}_1 \in D(W_1)$ and $W_h \not\subseteq W$ for $h < H$. The modified cost of $(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_T)$ is defined as

$$C^*(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_T) = c(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_T) - \sum_{l=2}^{H-1} R(W_h) \tag{2}$$

The modified cost of the minimum path in $S(\mathbf{x}, W)$ is defined as

$$C^*(\mathbf{x}, W) = \min_{(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_T) \in S(\mathbf{x}, W)} C^*(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_T) \tag{3}$$

The *modified coradius* of W is defined as $CR^*(W) = \max_{\mathbf{x} \notin W} C^*(\mathbf{x}, W)$. A set $W \subseteq \mathbf{X}$ is uniquely stochastically stable if either $R(W) > CR(W)$ or $R(W) > CR^*(W)$ (Ellison, 2000, Theorems 1 & 2).

3. How contagion and network cohesion matter

We aim to demonstrate how contagion affects long-run stable outcomes of the evolutionary framework described above. A set of strategies $A' \subset A$ is contagious on a given network if they can spread through best response dynamics from a small group of players to the whole network. Formally, contagion is defined through best response sequences.

Definition 1. *A sequence of strategy profiles* $\{\mathbf{x}_t\}_{t=0}^{\bar{t}}$ *of* (U, N, G, P)*, for some* $\bar{t} \geq 2$*, is a best response sequence if it satisfies the following properties:* (i) *for all* $t \geq 1$*, there exists at least one* $i \in N$ *such that* $x_t^i \neq x_{t-1}^i$*;* (ii) *if* $x_t^i \neq x_{t-1}^i$*, then* $x_t^i \in BR(\sigma_i(\mathbf{x}_{t-1}))$*.*[8]

Property (i) of Definition 1 requires that at least one player must switch a strategy at each period – this follows because we consider a dynamic process with a simultaneous revision protocol. Property (ii) requires players to switch strategies through best response dynamics.

Definition 2. *Let* (U, N, G, P) *start from some* $\mathbf{x} \in \mathbf{X}$*. Strategies in* $A' \subset A$ *spread contagiously on* G *from a subgroup of players,* $I(A'; \mathbf{x}) \subset N$*, if there exists some* $\bar{t} \geq 2$ *such that every best response sequence* $\{\mathbf{x}_t\}_{t=0}^{\bar{t}}$ *with* $\mathbf{x}_0 = \mathbf{x}$ *and* $x_1^i \in A'$ *for all* $i \in I(A'; \mathbf{x})$ *satisfies* $x_{\bar{t}}^i \in A'$ *for all* $i \in N$*.*

The set $I(A'; \mathbf{x})$ is a set of initial adopters of strategies in A' given that (U, N, G, P) starts from $\mathbf{x}$, and we write $\mu(A'; \mathbf{x})$ for its cardinality. Let $\mu^*(A'; \mathbf{x})$ be the smallest possible value of $\mu(A'; \mathbf{x})$ from which strategies in A' can spread contagiously starting from $\mathbf{x}$. That is, $\mu^*(A'; \mathbf{x})$ is the minimum number of mutations needed to trigger evolution of (U, N, G, P) from $\mathbf{x}$ to some state in $\mathbf{A}'$, a set of absorbing sets containing only strategies in A'. We derive upper bounds of $\mu^*(A'; \mathbf{x})$ for $\mathcal{G}_k$ networks below.

Definition 3. *Strategies in* $A' \subset A$ *are contagious on* G *if, starting from every* $\mathbf{x} \notin \mathbf{A}'$*, they can spread contagiously from a group of* $\mu^*(A'; \mathbf{x}) < \frac{n}{2}$ *players.*

There are several node-based and aggregate network measures that determine when and which strategies can spread contagiously. An example of a node-based measure that determines the feasibility of contagion is the largest degree of G, $\Delta(G)$. Let strategies in $A' \subset A$ be best responses when played by at least proportion p of neighbours. Then they can spread contagiously on any *strongly connected* network $G(n)$ with $n \geq 5$ (i.e. a network where every two players are connected by some path) if $p \leq 1/\Delta(G)$. This is because if $p \leq 1/\Delta(G)$, then

[8] This definition is similar to Oyama and Takahashi (2015, Definition 1) but different in that we consider simultaneous best response dynamics in finite networks.

for each $i \in N$ with corresponding degree n_i, strategies in A' are best responses when played by $\lceil pn_i \rceil = \lceil n_i/\Delta(G) \rceil = 1$ neighbour. Thus, strategies in A' spread contagiously from any two adjacently placed players.[9]

Here, we consider a network measure, *the maximum group cohesion*, that captures the overall structure of the network. For some $Z \subset N$ and $i \in Z$ in network G, define $\eta_i(Z, G)$ as the proportion of i's neighbours in Z. That is,

$$\eta_i(Z, G) = \frac{|N_i \cap Z|}{n_i}$$

A group $Z \subset N$ of players is α-cohesive if $\eta(Z, G) = \min_{i \in Z} \eta_i(Z, G) \geq \alpha$. The maximum group cohesion in G is then $\eta(G) = \max_{Z \subset N} \eta(Z, G)$. For any $G_k \in \mathcal{G}_k$ with k even, $\eta(G_k) = \frac{1}{2}$. Specifically, every subgroup consisting of $\frac{k}{2} + 1$ adjacently placed players is $\frac{1}{2}$-cohesive. For any $G_k \in \mathcal{G}_k$ with k odd, $\eta(G_k) = \frac{\lceil k/2 \rceil}{k}$. Specifically, there are groups of either $\lceil k/2 \rceil + 1$ or $k + 1$ adjacently placed players that are $\frac{\lceil k/2 \rceil}{k}$-cohesive. Note that since $\lceil k/2 \rceil = \frac{k}{2}$ when k is even, it is correct to say that all $\mathcal{G}_k$ networks have a maximum group cohesion of $\frac{\lceil k/2 \rceil}{k}$.

The necessary condition for strategies in A' (i.e. strategies that are best responses when played by at least proportion p of neighbours) to be contagious on network G is for p to be less of equal to $1 - \eta(G)$. To see why, notice that when $p > 1 - \eta(G)$ strategies in A' cannot spread contagiously from other regions of the network to an $\eta(G)$-cohesive group of players. That is, let Z be an $\eta(G)$-cohesive group of players and let $p > 1 - \eta(G)$. If all players in $N \backslash Z$ play strategies in A' and players in Z play strategies in $A \backslash A'$, then strategies in A' are not best responses to all $i \in Z$ because each $i \in Z$ has at most $1 - \eta(G)$ neighbours in $N \backslash Z$. However, if $p \leq 1 - \eta(G)$, then strategies in A' can spread to Z, although not guaranteed. The following definition and lemma formalize this notion for $\mathcal{G}_k$ networks.

For any nonempty subset $A' \subseteq A$ and any $\sigma \in \Sigma$, let $\sigma_{A'} = \sum_{a_j \in A'} \sigma(a_j)$ be the mass that σ places on strategies in A'. Given a set I of adjacently placed players in some $G_k \in \mathcal{G}_k$, we write $N_{I_1}(s)$ for a set of players in $N \backslash I$ with at least s neighbours in I; and for all $r = 2, 3, \cdots, d_I$, $N_{I_r}(s)$ is the set of players in $N \backslash \{I \cup N_{I_1}(s) \cup \cdots \cup N_{I_{r-1}}(s)\}$ with at least s neighbours in $N_{I_{r-1}}(s)$, where d_I is the value of r at which $N_{I_r}(s) \neq \emptyset$ but $N_{I_{r+1}}(s) = \emptyset$. Let $n_{I_r}(s)$ be the cardinality of $N_{I_r}(s)$.

Definition 4. *A nonempty subset of strategies $A' \subseteq A$ is a p-best response set of a symmetric strict coordination game U if for all $\sigma \in \Sigma$ with $\sigma_{A'} \geq p$, $BR(\sigma) \subseteq A'$.*

[9] That is, starting from any state $\mathbf{x} \notin \mathbf{A'}$, if any two adjacently placed players play strategies in A' at some $t \geq 1$, then they play strategies in A' from t onward; followed by their direct neighbours at $t + 1$; then their neighbours' neighbours; and so on, until the entire network plays strategies in A'.

In analogy to local interactions, $A' \subseteq A$ is a p-best response set if whenever at least proportion p of a player's neighbours play strategies in A', all of a player's best responses are themselves in A'.

Lemma 1. *Let $A' \subset A$ be a p-best response set of U. Then strategies in A' spread contagiously on any $G_k \in \mathcal{G}_k$ from a group of $\lceil pk \rceil + 1$ adjacently placed players if $p \leq 1 - \eta(G_k) = \frac{\lfloor k/2 \rfloor}{k}$.*

Proof. Consider a scenario where $(U, N, \mathcal{G}_k, P)$ starts from any $\mathbf{x} \notin \mathbf{A}'$ and at $t = 1$, let $\lceil pk \rceil + 1$ adjacently placed players in $I(A'; \mathbf{x}) = \{j, j+1, \cdots, j + \lceil pk \rceil\}$ play strategies in A'. Then, at $t = 2$, all players in I play strategies in A' because each $i \in I$ has at least $\lceil pk \rceil \geq p$ neighbours play strategies in A'. Similarly, all players in $N_{I_1}(s)$, for $s = \lceil pk \rceil$, play strategies in A' at $t = 2$. At $t = 3$, all players in $I \cup N_{I_1}(s) \cup N_{I_2}(s)$ play strategies in A'. This process continues until the whole network eventually plays strategies in A'. Thus, strategies in A' spread contagiously from $\lceil pk \rceil + 1$ adjacently placed players. $\square$

We can extend the above definition of contagion to *step-by-step contagion* as follows.

Definition 5. *Strategies in $A' \subset A$ are step-by-step contagious on G if, starting from any $\mathbf{x} \notin \mathbf{A}'$, there exists a sequence of strategy sets $A^0, A^1, \cdots, A^T$, with $A' = A^T$, and respective absorbing sets $\mathbf{A}^0, \mathbf{A}^1, \cdots, \mathbf{A}^T$, such that strategies in A^1 spread contagiously from $\mu^*(A^1; \mathbf{x}) < \frac{n}{2}$ players; and for $1 \leq \tau \leq T - 1$, strategies in $A^{\tau+1}$ spread contagiously from $\mu^*(A^{\tau+1}; \mathbf{x}_\tau) < \frac{n}{2}$ players, for all $\mathbf{x}_\tau \in \mathbf{A}^\tau$.*

Consider a partition of A into $A^0 \supseteq A^1 \supseteq \cdots \supseteq A^T$ so that, for $\tau = 0, 1, \cdots, T-1$, strategies in $A^{\tau+1}$ are best responses when played by at least proportion p of neighbours and proportion $1 - p$ play strategies in $A^\tau \backslash A^{\tau+1} = \bar{A}^\tau$. Under this partitioning of A, set A^T is referred to as an *iterated p-best response set* (Tercieux, 2006). Formally, given a nonempty set $A' \subseteq A$ containing $m' \leq m$ strategies, let $\Sigma_{A'} = \{\sigma \in \Sigma \mid \sigma_{A'} = 1\}$ be the set of probability distributions over A that place a total mass of 1 to A'. Recall the definition of the original symmetric strict coordination game as a mapping $U : \Sigma \to \mathbb{R}^m$. We define $U\mid_{A'}: \Sigma_{A'} \to \mathbb{R}^{m'}$ as the restricted version of U where players may only choose strategies in A'.

Definition 6. *A nonempty set of strategies $A^T \subseteq A$ is an iterated p-best response set of U if for some $T \geq 1$, there exists a sequence $A^0, A^1, \cdots, A^T$ with $A = A^0 \supseteq A^1 \supseteq \cdots \supseteq A^T$ such that A^τ is a p-best response set in $U\mid_{A^{\tau-1}}$, for each $\tau = 1, \cdots, T$.*

A strategy, say a_l, is an iterated p-dominant equilibrium of U if $A^T = \{a_l\}$. Oyama et al. (2015, Proposition 2) show that for any $p \leq \frac{1}{2}$, any coordination game has a *smallest iterated*

	a_1	a_2	a_3
a_1	10	0	0
a_2	9.1	11	10
a_3	0	9	13

Figure 1: A 3×3 symmetric coordination game with strategy a_3 as a globally pairwise risk dominant equilibrium.

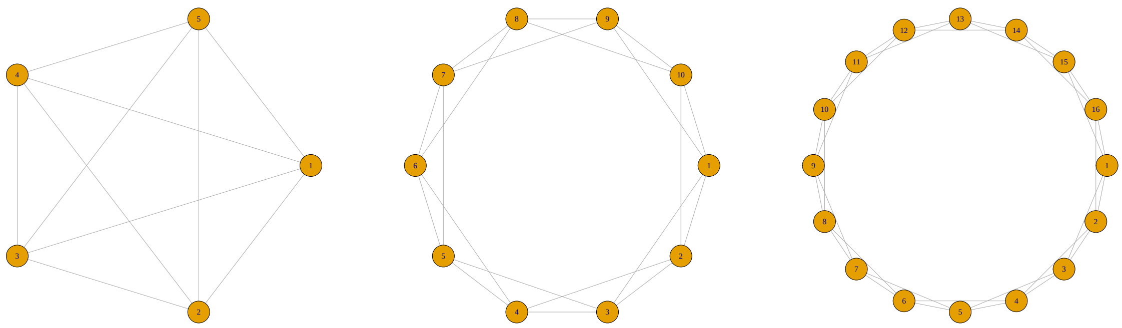
p-best response set. The sequence $A^0 \supseteq A^1 \supseteq \cdots \supseteq A^T$ is an *associated sequence* of A^T. For the coordination game in Figure 1, strategy a_3 is the smallest iterated p-best response equilibrium, for $p > \frac{2}{5}$. This is because, for $A^1 = \{a_2, a_3\}$, the best response to all $\sigma \in \Sigma$ with $\sigma_{A^1} > \frac{0.9}{10.9}$ is in A^1; and a_3 is a best response to all $\sigma \in \Sigma_{A^1}$ with $\sigma(a_3) > \frac{2}{5}$.

Lemma 2. *Let $A^T \subseteq A$ be the smallest iterated p-best response set of U. Then strategies in A^T are step-by-step contagious on any $G_k \in \mathcal{G}_k$ if $p \leq 1 - \eta(G_k) = \frac{\lfloor k/2 \rfloor}{k}$. Along every sequence of absorbing sets $\mathbf{A}^\tau, \mathbf{A}^{\tau+1}, \cdots, \mathbf{A}^T$ traversed by any path from some $\mathbf{x} \notin \mathbf{A}^T$ to $\mathbf{A}^T$, $\mu^*(A^{\tau+1}; \mathbf{x}_\tau) \leq \lceil pk \rceil + 1$.*

The proof of Lemma 2 follows from Lemma 1. That is, starting from any $\mathbf{x} \in D(\mathbf{A}^\tau)$, strategies in $A^{\tau+1}$ spread contagiously on $G_k \in \mathcal{G}_k$ if $p \leq 1 - \eta(G_k)$. Applying this notion iteratively proves Lemma 2.

Lemma 2 implies that if $p \leq \frac{\lfloor k/2 \rfloor}{k}$, then the absorbing set $\mathbf{A}^T$ can be reached from any other state through the process of step-by-step contagion. We further develop this idea in Section 4 and use it to derive the conditions under which the upper bound for the coradius of $\mathbf{A}^T$ is $\lceil pk \rceil + 1$.

We now demonstrate that when contagion is feasible, the number of mutations needed to exit the basin of attraction of $\mathbf{A}^T$ (i.e. the radius of $\mathbf{A}^T$) depends on the *network cohesion* – the cohesiveness of the whole network. Network cohesion is a measure of reachability among players. Specifically, let d_{ij} be the *distance* (the *geodesic*) from i to j. That is, d_{ij} is the length of the shortest path from i to j. Let $d_i = \max_{j \neq i} d_{ij}$ be the maximum distance from i to any other player. This parameter captures the reachability of other players from i. We then define the cohesiveness of network G as the maximum reachability between any two players in G. That is, the network cohesion of G is $d(G) = \max_{i \in N} d_i$. This parameter is also referred, in graph theory, as the *diameter* of G.



(a) $G_4(5)$ with diameter = 1 (b) $G_4(10)$ with diameter = 3 (c) $G_4(16)$ with diameter = 4

Figure 2: Examples of $\mathcal{G}_4$ networks.

Lemma 3. *For any $G_k(n) \in \mathcal{G}_k$ with k even, $d(G_k(n)) = \lceil \frac{n-1}{k} \rceil$; for any $G_k(n) \in \mathcal{G}_k$ with k odd, $\lceil \frac{n-1}{k+1} \rceil \leq d(G_k(n)) \leq \lceil \frac{n-1}{k-1} \rceil$.*

Proof. Pick any $i \in N$ and let N_{i_r} be the set of players at distance r from i (i.e. the r-order neighbours of i) and let n_{i_r} be its cardinality. Then the maximum distance from i to any other player is the value of d_i that solves the equation,

$$\sum_{r=1}^{d_i} n_{i_r} = n - 1$$

Since for any $G_k(n) \in \mathcal{G}_k$, $d_i = d(G_k(n))$ for all $i \in N$, it follows that the diameter of $G_k(n)$ is the value of $d(G_k(n))$ that solves the equation,

$$\sum_{r=1}^{d(G_k(n))} n_{i_r} = n - 1, \quad \text{for any } i \in N \tag{4}$$

For any $G_k(n) \in \mathcal{G}_k$ with k even, $n_{i_r} = k$ for all $r = 1, \cdots, d_i$. This is because every player has $\frac{k}{2}$ neighbours to the left and to the right. Substituting into (4) yields the desired value of $d(G_k(n)) = \lceil \frac{n-1}{k} \rceil$

For any $G_k(n) \in \mathcal{G}_k$ with k odd, each player has either $\lfloor k/2 \rfloor$ neighbours to the left and $\lceil k/2 \rceil$ to the right or vice versa. And since $2\lfloor k/2 \rfloor = k - 1$ and $2\lceil k/2 \rceil = k + 1$, it follows that $k - 1 \leq n_{i_r} \leq k + 1$ for all $r = 1, \cdots, d_i$. Substituting into (4) yields the desired result. $\square$

Example 1. To see how network cohesion affects the radius of $\mathbf{A}^T$, consider the coordination game of Figure 1 played on $G_4(5)$, $G_4(10)$ and $G_4(16)$ networks of Figure 2. When this game is played on $\mathcal{G}_4$ networks, strategy a_1 is a best response (BR) only when played by all four neighbours; a_2 is a BR when played by one neighbour and three neighbours play a_1, and when played by three neighbours and the other plays a_3; a_2 is also a BR when one neighbour plays a_1 and three play a_3, and when one neighbour plays a_3 and three play a_1; a_3 is a BR when played by two neighbours and the other two neighbours play a_2.

It is easy to rule out $\mathbf{a_1}$ from a set of stochastically stable states in all $\mathcal{G}_4$ networks since one mutation triggers an exit from $D(\mathbf{a}_1)$, but at least four mutations are necessary for a transition from $\mathbf{a}_2$ and $\mathbf{a}_3$ to $\mathbf{a}_1$. For $G_4(5)$, $CR^*(\mathbf{a}_2) = 1 < 2 = R(\mathbf{a}_2)$, and hence, $\mathbf{a}_2$ is uniquely stochastically stable. That is, one mutation to a_1 triggers a transition from $\mathbf{a}_3$ to $\mathbf{a}_2$. However, at least two mutations are needed for a transition from $\mathbf{a}_2$ to $\mathbf{a}_3$.

For $G_4(10)$, $CR^*(\mathbf{a}_3) = CR^*(\mathbf{a}_2) = 2 = R(\mathbf{a}_2) = R(\mathbf{a}_3)$, and hence, $\mathbf{a}_2$ and $\mathbf{a}_3$ constitute the stochastically stable set. That is, starting from $\mathbf{a}_3$, if players 1 and 6 mutate to a_1 at $t = 1$, then at $t = 2$, players $\{2, 3, 4, 5, 7, 8, 9, 10\}$ switch to a_2. At $t = 3$, all players play a_2, and hence, $R(\mathbf{a}_3) = CR^*(\mathbf{a}_2) = 2$. Next, starting from $\mathbf{a}_2$, let players $\{1, 2\}$ mutate to a_3 at $t = 1$. Then from $t = 2$ onward, players $\{1, 2\}$ and $\{3, 10\}$ alternate between a_2 and a_3 (given that a_3 is a best response when played by two neighbours and the rest play a_2). Denote this absorbing cycle by W so that $R(\mathbf{a}_2) = C(\mathbf{a}_2, W) = 2$. Starting from some state in W, one mutation by either player 3 or 10 will trigger a contagious spread of a_3 to the entire network.[10] Thus, $R(W) = C(W, \mathbf{a}_3) = 1$, and $CR^*(\mathbf{a}_3) = C(\mathbf{a}_2, W) + C(W, \mathbf{a}_3) - R(W) = 2$.[11]

For $G_4(16)$, $CR^*(\mathbf{a}_3) = 2 < 3 = R(\mathbf{a}_3)$, and hence, $\mathbf{a}_3$ is uniquely stochastically stable. To see why, first note that $CR^*(\mathbf{a}_3) = 2$ follows from the steps in the derivation of $CR^*(\mathbf{a}_3)$ for $G_4(10)$ network above. Second, starting from $\mathbf{a}_3$, let two players, say $\{1, 9\}$ mutate to a_1 at $t = 2$. Then at $t = 2$, $\{2, 3, 7, 8, 10, 11, 15, 16\}$ switch to a_2 but $\{1, 4, 5, 6, 9, 12, 13, 14\}$ play a_3. At $t = 3$, $\{1, 9\}$ play a_2 and the rest play a_3. At $t = 4$, the whole network plays a_3. Thus, at least three mutations are needed to trigger an exit from $D(\mathbf{a}_3)$.

Example 1 illustrates two concepts. First, strategy a_3 is step-by-step contagious on $\mathcal{G}_4$ networks since it is the smallest iterated p-best response equilibrium for all $p > \frac{2}{5}$. Specifically, starting from any $\mathbf{x} \neq \mathbf{a}_3$, strategies in $A^1 = \{a_2, a_3\}$ spread contagiously from any $\mu^*(A^1; \mathbf{x}) = 2$ adjacently placed players. And starting from any $\mathbf{x}_1 \in \mathbf{A}^1$, $A^2 = a_3$ spreads contagiously from $\mu^*(A^2; \mathbf{x}_1) = 1$ player.

Second, the smallest iterated p-best response equilibrium (i.e. strategy a_3) becomes uniquely stochastically stable as the network cohesion decreases (i.e. as network diameter increases). The intuition, as Example 1 demonstrates, is that when the network diameter is large, exiting the basin of attraction of $\mathbf{A}^T$ requires multiple simultaneous mutations from different regions of the

[10]That is, let either player 3 or 10 mutates to a_3 at some t so that $\{1, 2, 3\}$ or $\{1, 2, 10\}$ simultaneously play a_3. Then at $t + 1$, all $\{1, 2, 3, 4, 10\}$ play a_3; at $t + 2$, all $\{1, 2, 3, 4, 5, 9, 10\}$ play a_3; and at $t + 3$, the entire network plays a_3.

[11]Note that a path from $\mathbf{a}_1$ to $\mathbf{a}_3$ has a small net cost. That is, since $R(\mathbf{a}_1) = C(\mathbf{a}_1, \mathbf{a}_2) = 1$, it follows that $C^*(\mathbf{a}_1, \mathbf{a}_2, W, \mathbf{a}_3) = C(\mathbf{a}_1, \mathbf{a}_2) + C(\mathbf{a}_2, W) + C(W, \mathbf{a}_3) - R(\mathbf{a}_2) - R(W) = 1$.

network.

Specifically, consider a scenario (in Example 1 above) where (U, N, G_4, P) starts from $\mathbf{A}^2 = \{\mathbf{a}_3\}$. We see that for network $G_4(5)$, if one player in $I = \{1\}$ mutates to a_1, all other players switch to a_2 because they are one step away from I. For network $G_4(10)$, if $I = \{1\}$ mutates to a_1, all players in $N_{I_1}(1) = \{2, 3, 9, 10\}$ switch to a_2 at $t = 2$. However, because the number of players in $N \backslash \{I \cup N_{I_1}(1)\} = \{4, 5, 6, 7, 8\}$, that all still play a_3 at $t = 2$, is sufficiently large, the whole network eventually reverts to play a_3. Thus, exiting the basin of attraction of $\mathbf{A}^2$ requires two mutations from two different regions of the network (e.g. by players $\{1, 6\}$). For $G_4(16)$, when $I = \{1, 9\}$ (i.e. two players from two different regions of the network) mutate to a_1, all players in $N_{I_2}(1) = \{4, 5, 6, 12, 13, 14\}$ stick to a_3 because they are two steps away from I. This set of players sufficiently triggers the contagious spread of a_3. Thus, exiting $D(\mathbf{A}^2)$ in network $G_4(16)$, requires three mutations from three different regions of the network.[12]

Note that $(U, N, \mathcal{G}_k, P)$ only reverts to $\mathbf{A}^2$ after a few mutations to a_1 because strategy a_3 is step-by-step contagious on $\mathcal{G}_4$ networks. However, for any $G_k \in \mathcal{G}_k$ network where a_3 is not step-by-step contagious (i.e. where $p > \frac{2}{5} > 1 - \eta(G_k)$), $(U, N, \mathcal{G}_k, P)$ need not revert to $\mathbf{A}^2$, and hence, can exit the basin of attraction of $\mathbf{A}^2$ with one mutation regardless of the network diameter. For example, if the game in Figure 1 is played on $\mathcal{G}_3$ networks where $p > \frac{2}{5} > 1 - \eta(G_3) = \frac{1}{3}$, one mutation to a_1 triggers an exit from $D(\mathbf{A}^2)$ to some absorbing state containing both a_3 and a_2. As we show in Section 4 below, contagion ensures that $(U, N, \mathcal{G}_k, P)$ does not contain absorbing states where strategies in $\bar{A}^\tau$ coexist with strategies in $A^{\tau+1}$.

Thus, when contagion is feasible, the network diameter determines the radius of $\mathbf{A}^T$, and hence, whether $\mathbf{A}^T$ is uniquely stochastically stable. When the network diameter is larger compared to T, exiting the basin of attraction of $\mathbf{A}^T$ requires simultaneous mutations from multiple regions of the network. We further develop this idea below and show that, given p, T and $\mathcal{G}_k$ networks, there exists some $G_k^* \in \mathcal{G}_k$ with threshold diameter $d(G_k^*)$ such that, for all $G_k \in \mathcal{G}_k$ with $d(G_k) \geq d(G_k^*)$, the minimum number of mutations needed to exit the basin of attraction of $\mathbf{A}^T$ is larger than $CR^*(\mathbf{A}^T)$. Since each $G_k(n) \in \mathcal{G}_k$ is uniquely defined by n, the above statement can equivalently be stated as: given p, T and $\mathcal{G}_k$ networks, there exists some n^* whereby, for all $G_k(n) \in \mathcal{G}_k$ with $n \geq n^*$, the minimum number of mutations needed to exit the basin of attraction of $\mathbf{A}^T$ is larger than $CR^*(\mathbf{A}^T)$.

[12] Note that the density of connections does not play a role in determining the number of mutations needed to exit the basin of attraction of $\mathbf{A}^2$ because all $G_k \in \mathcal{G}_k$ have the same density equal to k (e.g. all $\mathcal{G}_4$ networks have density of 4, regardless of the size of the network).

4. Stochastic stability of iterated p-best response sets

The following definitions and notations are used in the statement of Theorem 4 below. For $q \in (0,1)$, let σ^q_{jl} denote any $\sigma \in \Sigma$ with $\sigma(a_l) = q$ and $\sigma(a_j) = 1-q$. Let η_{lj} be some positive real number such that for all $q > \eta_{lj}$, $BR(\sigma^q_{lj}) = \{a_j\}$. For any triple of strategies $a_j, a_h, a_l \in A$, define a payoff parameter $\beta_{lh}(\sigma^q_{lj})$ as a proportion of a player's neighbours that must play a_j and the rest play a_l for a player to be indifferent between a_l and a_h; and for $\beta_{lh}(\sigma^q_{lj}) < q < \eta_{lj}$, $a_h \in BR(\sigma^q_{lj})$. Given a payoff matrix M, the expression for $\beta_{lh}(\sigma^q_{lj})$ is[13]

$$\beta_{lh}(\sigma^q_{lj}) = \frac{m_{ll} - m_{hl}}{m_{ll} - m_{hl} + m_{hj} - m_{lj}} \tag{5}$$

Associated with $\beta_{lh}(\sigma^q_{lj})$ are the parameters: $n_{lh}(\sigma^q_{lj})$, which is the minimum number of neighbours that should play a_j and the rest play a_l, for a_h to be a best response; $n^*_{\tau(v+1)}(\sigma^q_{\tau v})$, for $0 \le v \le \tau - 1$, which is the minimum number of neighbours that must play a strategy in $\bar{A}^v$, and the rest play strategies in A^τ, for some strategies in A^{v+1} to be best responses; $n(A^\tau)$, which is the minimum $n^*_{\tau(v+1)}(\sigma^q_{\tau v})$ over all $v \in [0, \tau-1]$; and $n^*(A)$, which is the minimum $n(A^\tau)$ over all $\tau \in [0, T-1]$. That is, letting $\mathbb{N}_+$ be a set of all non-negative integers, then:

$$n_{lh}(\sigma^q_{lj}) = \begin{cases} \lceil \beta_{lh}(\sigma^q_{lj})k \rceil & \text{if } \beta_{lh}(\sigma^q_{lj})k \notin \mathbb{N}_+ \\ \lceil \beta_{lh}(\sigma^q_{lj})k \rceil + 1 & \text{if } \beta_{lh}(\sigma^q_{lj})k \in \mathbb{N}_+ \end{cases} \tag{6}$$

$$n^*_{\tau(v+1)}(\sigma^q_{\tau v}) = \min_{a_l \in A^\tau} \left(\min_{a_j \in \bar{A}^v} \left(\min_{a_h \in A^{v+1}} n_{lh}(\sigma^q_{lj}) \right) \right) \quad \text{for } 0 \le v \le \tau - 1 \tag{7}$$

$$n(A^\tau) = \min_{v \in [0, \tau-1]} n^*_{\tau(v+1)}(\sigma^q_{\tau v}) \tag{8}$$

$$n^*(A) = \min_{\tau \in [0,T]} n(A^\tau) \tag{9}$$

Finally, let parameters $n(p,k)$ and γ^* be defined as follows.

$$n(p,k) = \begin{cases} \lceil pk \rceil & \text{if } 3\lceil pk \rceil \le 2(\lceil k/2 \rceil + 1) \\ \lceil pk \rceil + 1 & \text{if } 3\lceil pk \rceil > 2(\lceil k/2 \rceil + 1) \end{cases} \tag{10}$$

$$\gamma^* = \begin{cases} \lceil \frac{n(p,k)}{n^*(A)+1} \rceil & \text{if } \frac{n(p,k)}{n^*(A)+1} \notin \mathbb{N}_+ \\ \lceil \frac{n(p,k)}{n^*(A)+1} \rceil + 1 & \text{if } \frac{n(p,k)}{n^*(A)+1} \in \mathbb{N}_+ \end{cases} \tag{11}$$

[13]That is, $\beta_{lh}(\sigma^q_{lj})$ is the value of q at which $U\left(a_h \mid \sigma^q_{lj}\right) = U\left(a_l \mid \sigma^q_{lj}\right)$, so that for $\beta_{lh}(\sigma^q_{lj}) < q < \eta_{lj}$, $U\left(a_h \mid \sigma^q_{lj}\right) > U\left(a_l \mid \sigma^q_{lj}\right), U\left(a_j \mid \sigma^q_{lj}\right)$. The resulting expression for $\beta_{lh}(\sigma^q_{lj})$ is then, $(1-q)m_{hl} + qm_{hj} = (1-q)m_{ll} + qm_{lj}$, which yields $q = \frac{m_{ll} - m_{hl}}{m_{ll} - m_{hl} + m_{hj} - m_{lj}} = \beta_{lh}(\sigma^q_{lj})$

Theorem 4. *Let A^T be the smallest iterated p-best response set of a symmetric strict coordination game U. Then, for $p \leq \frac{\lfloor k/2 \rfloor}{k}$, A^T is uniquely stochastically stable in an evolutionary model with random mutations, $(U, N, \mathcal{G}_k, P_\varepsilon)$, if:*

(i) for $T = 1$ and $\mathcal{G}_{n-1}$ networks, $n \geq 3$

(ii) for $T = 1$ and $d(G_k) \geq 2$, $n \geq k+2$;

(iii) for all $T \geq 2$, $n \geq \gamma^*\big(\lceil pk \rceil + 3 - n^*(A)\big) + 2\gamma^*\lceil k/2 \rceil (T-1)$.

The proof of Theorem 4 follows in three steps. The first step establishes the structure of $\mathbf{A}$, the set of all absorbing sets of $(U, N, \mathcal{G}_k, P)$. Recall that $\mathbf{A}$ consists of absorbing states and absorbing cycles. We categorize the set of absorbing states into *monomorphic absorbing states* – absorbing states that contain only one strategy, denoted by $M(\mathbf{A})$, and *polymorphic absorbing states* – absorbing states that contain more than one strategy, denoted by $Q(\mathbf{A})$.

Lemma 5. *For a symmetric strict coordination game U, let $A^{\tau+1}$ be the p-best response set in $U\mid_{A^\tau}$. If $p \leq \frac{\lfloor k/2 \rfloor}{k}$, then for all $\tau = 0, 1, \cdots, T-1$, the evolutionary process $(U, N, \mathcal{G}_k, P)$:*

(i) does not possess absorbing states where strategies in $\bar{A}^\tau = A^\tau \backslash A^{\tau+1}$ coexist with strategies in $A^{\tau+1}$;

(ii) can have absorbing cycles where strategies in $\bar{A}^\tau$ coexist with strategies in $A^{\tau+1}$.

Proof. See Appendix B ∎

The proof of Lemma 5 relies on the properties of the process of contagion discussed in Section 3. First, we show that if $p \leq \frac{\lfloor k/2 \rfloor}{k}$, then starting from some state $\mathbf{x}$ where a set $I = \{j, j+1, \cdots, j + \lfloor k/2 \rfloor\}$ of $\lfloor k/2 \rfloor + 1$ adjacently placed players play strategies in $A^{\tau+1}$ and the rest play strategies in A^τ, strategies in $A^{\tau+1}$ will spread contagiously to the whole network. That is, starting from $\mathbf{x}$ at $t = 1$, all $i \in I$ play strategies in $A^{\tau+1}$ at $t = 2$ because each $i \in I$ has at least $\lfloor k/2 \rfloor \geq \lceil pk \rceil$ neighbours play strategies in $A^{\tau+1}$ and the rest play strategies in $\bar{A}^\tau$. Still at $t = 2$, all players in $N_{I_1}(s)$, for $s = \lceil pk \rceil$, play strategies in $A^{\tau+1}$ because each has at least $\lceil pk \rceil$ neighbours play strategies in $A^{\tau+1}$ and the rest play strategies in $\bar{A}^\tau$. From $t = 3$ onward, all players in $I \cup N_{I_1}(s) \cup N_{I_2}(s)$ play strategies in $A^{\tau+1}$. This iterative process continues until the entire network eventually plays strategies in $A^{\tau+1}$.

Second, we show that if $p \leq \frac{\lfloor k/2 \rfloor}{k}$, then $(U, N, \mathcal{G}_k, P)$ does not contain absorbing states where strategies in $\bar{A}^\tau$ and $A^{\tau+1}$ coexist within subgroups of $\lfloor k/2 \rfloor + 1$ adjacently placed players. Specifically, each state where players in I play both strategies in $\bar{A}^\tau$ and $A^{\tau+1}$ is either transient or belongs to an absorbing cycle. Thus, when $p \leq \frac{\lfloor k/2 \rfloor}{k}$, strategies in $\bar{A}^\tau$ cannot co-exists with strategies in $A^{\tau+1}$ in an absorbing state.

To prove Lemma 5 (ii), it is sufficient to provide an example of an absorbing cycle where strategies in $\bar{A}^\tau$ coexist with strategies in $A^{\tau+1}$ in $\mathcal{G}_k$ networks. The simplest example is $\mathcal{G}_2$ networks, whereby, any two states where adjacent players alternate between strategies in $\bar{A}^\tau$ and $A^{\tau+1}$ form an absorbing cycle.

Lemma 5 (i) implies that, when $p \leq \frac{\lfloor k/2 \rfloor}{k}$, absorbing states can be categorized into absorbing states containing only strategies in $\bar{A}^\tau$, for all $\tau = 0, 1, \cdots, T$. That is, since there are no absorbing states where strategies in $\bar{A}^\tau$ co-exist with strategies in $A^{\tau+1}$, for all $\tau = 0, 1, \cdots, T$, sets $M(\mathbf{A})$ and $Q(\mathbf{A})$ can be categorized into $M(\bar{\mathbf{A}}^\tau)$ and $Q(\bar{\mathbf{A}}^\tau)$, for $\tau = 0, 1, \cdots, T$, which are the sets of monomorphic and polymorphic absorbing states containing only strategies in $\bar{A}^\tau$ respectively. Thus, $\mathbf{A}$ can be expressed as

$$\mathbf{A} \equiv \bigcup_{\tau=0}^{T} \left(M(\bar{\mathbf{A}}^\tau) \cup Q(\bar{\mathbf{A}}^\tau) \right) \bigcup L(\mathbf{A}) \equiv \bigcup_{\tau=0}^{T} \left(M(\bar{\mathbf{A}}^\tau) \cup Q(\bar{\mathbf{A}}^\tau) \cup L(\mathbf{A}^\tau) \right) \tag{12}$$

where $L(\mathbf{A})$ is the set of all absorbing cycles; and $L(\mathbf{A}^\tau)$ is the set of all absorbing cycles containing either only strategies in $\bar{A}^\tau$ or strategies in $\bar{A}^\tau$ together with some strategies in $A^{\tau+1}$. We let $\bar{\mathbf{A}}^\tau = M(\bar{\mathbf{A}}^\tau) \cup Q(\bar{\mathbf{A}}^\tau) \cup L(\mathbf{A}^\tau)$ and write $\mathbf{A}^\tau$ for a set of all absorbing sets containing only strategies in A^τ, that is,

$$\mathbf{A}^\tau \equiv \bigcup_{v=\tau}^{T} \left(M(\bar{\mathbf{A}}^v) \cup Q(\bar{\mathbf{A}}^v) \cup L(\mathbf{A}^v) \right) \tag{13}$$

The second step of the proof of Theorem 4 is a derivation of the upper bound of the radius of $\mathbf{A}^T$ (i.e the minimum number of mutations needed to trigger an exit from $D(\mathbf{A}^T)$). Firstly, when $T = 1$ and $p \leq \frac{\lfloor k/2 \rfloor}{k}$, $R(\mathbf{A}^T) \geq \lceil k/2 \rceil + 1$ for any $G_k(n) \in \mathcal{G}_k$ with $n \geq 3$. To see why, consider a scenario where $(U, N, \mathcal{G}_k, P)$ starts from some $\mathbf{x} \in \mathbf{A}^1$. If $\lceil k/2 \rceil$ players mutate to strategies in $\bar{A}^0$ at $t = 1$, $(U, N, \mathcal{G}_k, P)$ reverts to a state in $D(\mathbf{A}^1)$ at $t = 2$, and eventually to some state in $\mathbf{A}^1$. This is because when $p \leq \frac{\lfloor k/2 \rfloor}{k}$, strategies in $\bar{A}^0$ are best responses only when played by at least $\lceil k/2 \rceil + 1$ neighbours. Thus, the number of mutations needed to trigger an exit from $D(\mathbf{A}^T)$ is at least $\lceil k/2 \rceil + 1$.

Secondly, if $T = 1$ and $p \leq \frac{\lfloor k/2 \rfloor}{k}$, then, for any $G_k(n) \in \mathcal{G}_k$ with $k \geq 3$ and $d(G_k) \geq 2$, $R(\mathbf{A}^1) \geq \lceil k/2 \rceil + 2$ whenever $n \geq k + 2$. To see why, consider a scenario where $\lceil k/2 \rceil + 1$ players mutate to strategies in $\bar{A}^0$ at $t = 1$. If these $\lceil k/2 \rceil + 1$ players are adjacently placed so that $I(\bar{A}^0; \mathbf{x}) = \{j, j+1, \cdots, j + \lceil k/2 \rceil\}$, then no player has $\lceil k/2 \rceil + 1$ neighbours in I, and hence the evolutionary process reverts to $\mathbf{A}^1$ at $t = 2$. However, if, for some $j > \lfloor \frac{\lceil k/2 \rceil + 1}{2} \rfloor$, the elements of $I(\bar{A}^0; \mathbf{x})$ are arranged in the following manner:

$$I(\bar{A}^0; \mathbf{x}) = \left\{ j - \left\lfloor \frac{\lceil k/2 \rceil + 1}{2} \right\rfloor, \cdots, j-1, j+1, \cdots, j + \left\lceil \frac{\lceil k/2 \rceil + 1}{2} \right\rceil \right\},$$

then at most two players in $N \backslash I$ have $\lceil k/2 \rceil + 1$ neighbours in I. For this scenario, at $t = 2$, two players play strategies in $\bar{A}^0$ and the rest play strategies in A^1. Since $\lceil k/2 \rceil + 1 > 2$ when $k \geq 3$, it follows that at the end of period $t = 2$, all players have at least $\lfloor k/2 \rfloor \geq \lceil pk \rceil$ neighbours play strategies in A^1. Thus, at $t = 3$, all players play strategies in A^1, and hence, at least $\lceil k/2 \rceil + 2$ mutations are needed to trigger an exit from $D(\mathbf{A}^1)$.

When $T \geq 2$, the evolutionary process can exit $D(\mathbf{A}^T)$ with less or equal to $\lceil pk \rceil$ mutations to strategies in some $\bar{A}^\tau$, for $0 \leq \tau \leq T-2$. As discussed in Section 3, this scenario happens when the network diameter is small compared to T. We show that given k, p, and T, there exists a value of n, n^* (corresponding to some threshold diameter $d(G_k(n^*))$), below which $(U, N, \mathcal{G}_k, P)$ can exit $D(\mathbf{A}^T)$ with μ mutations, for $n(A^T) \leq \mu \leq \lceil pk \rceil + 1$, to some strategies in $A \backslash A^T$. Above n^*, at least $\mu + 1$ mutations are needed to trigger an exit from $D(\mathbf{A}^T)$. The following steps present the intuition behind this result and the details are presented in Appendix Appendix C.

Let $(U, N, \mathcal{G}_k, P)$ start from some $\mathbf{x} \in \mathbf{A}^T$. At $t = 1$, let μ adjacently placed players in $I = \{j, j+1, \cdots, j+\mu-1\}$, where $n^*_{T(\tau+1)}(\sigma^q_{T\tau}) \leq \mu \leq \lceil pk \rceil + 1$, mutate to strategies in some $\bar{A}^\tau$, for $0 \leq \tau \leq T-2$. Consider a special case where the r-order neighbourhoods of I do not overlap. That is, for $r = 1, 2, \cdots, d_I$, players in $N_{I_r}(s_r)$, for $s_r = n^*_{T(\tau+r)}(\sigma^q_{T(\tau+r-1)})$, do not have direct neighbours in $N_{I_{r-2}}(s_{r-2})$. Then the following scenario can unfold from $t = 2$ onward:

$t = 2$	all players in $I \cup N_{I_1}(s_1)$, for $s_1 = n^*_{T(\tau+1)}(\sigma^q_{T\tau})$, play strategies in $A^{\tau+1}$ (this follows from the definition of $n^*_{T(\tau+1)}(\sigma^q_{T\tau})$); all players in $N \backslash \{I \cup N_{I_1}(s_1)\}$ play strategies in A^T.
$t = 3$	all players in $I \cup N_{I_1}(s_1)$ play strategies in $A^{\tau+1}$; players in $N_{I_2}(s_2)$, for $s_2 = n^*_{T(\tau+2)}(\sigma^q_{T(\tau+1)})$, play strategies in $A^{\tau+2}$; and players in $N \backslash \{I \cup N_{I_1}(s_1) \cup N_{I_2}(s_2)\}$ play strategies in A^T.
$t = 4$	all $i \in I \cup N_{I_1}(s_1)$ play strategies in $A^{\tau+1}$; all $i \in N_{I_2}(s_2)$ play strategies in $A^{\tau+2}$; all $i \in N_{I_3}(s_3)$, for $s_3 = n^*_{T(\tau+3)}(\sigma^q_{T(\tau+2)})$, play strategies in $A^{\tau+3}$; all $i \in N \backslash \{I \cup N_{I_1}(s_1) \cup N_{I_2}(s_2) \cup N_{I_3}(s_3)\}$ play strategies in A^T.
– –	– –
$t = T - \tau$	all $i \in I \cup N_{I_1}(s_1)$ play strategies in $A^{\tau+1}$; all $i \in N_{I_2}(s_2)$ play strategies in $A^{\tau+2}$; all $i \in N_{I_3}(s_3)$ play strategies in $A^{\tau+3}$; $\cdots$; all $i \in N_{I_{T-\tau-1}}(s_{T-\tau-1})$, for $s_{T-\tau-1} = n^*_{T(T-1)}(\sigma^q_{T(T-2)})$, play strategies in A^{T-1}; and all $i \in N \backslash \{I \cup N_{I_1}(s_1) \cup \cdots \cup N_{I_{T-\tau-1}}(s_{T-\tau-1})\}$ play strategies in A^T.

For this scenario, we see that after $t = T - \tau$ iterations, at least $n - \left(\mu + \sum_{r=1}^{T-\tau-1} n_{I_r}(s_r)\right)$

players play strategies in A^T. Let z, Z, and ϕ be defined as follows:

$$z = n - \left(\mu + \sum_{r=1}^{T-\tau-1} n_{I_r}(s_r)\right)$$

$$Z = N \backslash \{I \cup N_{I_1}(s_1) \cup \cdots \cup N_{I_{T-\tau-1}}(s_{T-\tau-1})\}$$

$$\phi = \mu + \sum_{r=1}^{T-\tau} n_{I_r}(s_r)$$

From the strategy configuration at $t = T - \tau$, if $z \leq \lceil pk \rceil$, then, at $t = T - \tau + 1$, all $i \in Z$ can switch to strategies in $\bar{A}^{T-1}$ if a sufficiently large number of players in $N_{I_{T-\tau-1}}(s_{T-\tau-1})$ play strategies in $\bar{A}^{T-1}$. For this scenario, the evolutionary process $(U, N, \mathcal{G}_k, P)$ can converge to an absorbing set containing either only strategies in $A \backslash A^T$, or strategies in A^T together with some strategies in $A \backslash A^T$. When this happens, $(U, N, \mathcal{G}_k, P)$ exits the basin of attraction of $\mathbf{A}^T$ with μ mutations to strategies in $\bar{A}^\tau$.

However, if $z \geq \lceil pk \rceil + 1$, then, from $t = T - \tau + 1$ onward, all players in Z play strategies in A^T because each $j \in Z$ has at least $\lceil pk \rceil$ neighbours playing strategies in A^T and the rest playing strategies in A^{T-1}. For this scenario, the above evolutionary process will eventually converge to some state in $\mathbf{A}^T$. The intuition is as follows.

Let $N^1_{I_r}[N_{I_{r+1}}]$ be a set of players in $N_{I_r}(s_r)$ with at least $\lceil pk \rceil$ neighbours in $N_{I_{r+1}}(s_{r+1})$; $N^2_{I_r}[N_{I_{r+1}}]$ is a set of players in $N_{I_r}(s_r)$ with at least $\lceil pk \rceil$ neighbours in $N_{I_{r+1}}(s_{r+1}) \cup N^1_{I_r}[N_{I_{r+1}}]$; and more generally, $N^v_{I_r}[N_{I_{r+1}}]$ is a set of players in $N_{I_r}(s_r)$ with at least $\lceil pk \rceil$ neighbours in $N_{I_{r+1}}(s_{r+1}) \cup N^1_{I_r}[N_{I_{r+1}}] \cup \cdots \cup N^{v-1}_{I_r}[N_{I_{r+1}}]$.

Then, starting from $t = 4$, the following iterative process unfolds at the background of the evolutionary process described above. At $t = 4$, players in $N^1_{I_1}[N_{I_2}]$ play strategies in $A^{\tau+2}$; at $t = 5$, players in $N^1_{I_1}[N_{I_2}] \cup N^2_{I_1}[N_{I_2}]$ play strategies in $A^{\tau+2}$; players in $N^1_{I_2}[N_{I_3}]$ play strategies in $A^{\tau+3}$. This iterative process continues until some $t = t_1 \geq 5$ where all players in $N_{I_1}(s_1)$ play strategies in $A^{\tau+2}$ and all players in $N_{I_2}(s_2)$ play strategies in $A^{\tau+3}$. At $t = t_1 + 1$, all players in I with at least $\lceil pk \rceil$ neighbours in $N_{I_1}(s_1)$ play strategies in $A^{\tau+2}$; all players in $N^1_{I_1}[N_{I_2}]$ play strategies in $A^{\tau+3}$. At some $t = t_1 + t_2$, all players in $N_{I_3}(s_3)$ play strategies in $A^{\tau+4}$, so that at $t = t_1 + t_2 + 1$, all players in $N^1_{I_2}[N_{I_3}]$ play strategies in $A^{\tau+4}$; and so on. Eventually, at some $t^* > T - \tau + 2$, the entire network plays strategies in A^T.

Thus, when $z = n - \phi \geq \lceil pk \rceil + 1$, at least $\mu + 1$ mutations to strategies in $\bar{A}^\tau$ are necessary to trigger an exit from $D(\mathbf{A}^T)$. Now, starting from some $\mathbf{x} \in \mathbf{A}^T$, let two sets of players, $I_1 = \{1, 2, \cdots, \mu - 1\}$ and $I_2 = \{\frac{n}{2} + 1, \frac{n}{2} + 2, \cdots, \frac{n}{2} + \mu\}$, each of size μ, mutate to strategies in $\bar{A}^\tau$. Then, following the same steps above, at least $n - 2\phi$ players play strategies in A^T after $t = T - \tau$ iterations. If $n - 2\phi \geq 2(\lceil pk \rceil + 1)$, then $(U, N, \mathcal{G}_k, P)$ will converge to some state in

$\mathbf{A}^T$, so that at least $2(\mu+1)$ mutations to strategies in $\bar{A}^\tau$ are needed to trigger an exit from $D(\mathbf{A}^T)$. More generally, if $n - \gamma\phi \geq \gamma(\lceil pk + 1)\rceil$, at least $\gamma(\mu+1)$ mutations to strategies in $\bar{A}^\tau$ are needed to trigger an exit from $D(\mathbf{A}^T)$.

To derive the minimum number of mutations to any strategies in $A\backslash A^T$ that is needed to trigger an exit from $D(\mathbf{A}^T)$, consider the worst scenario, which corresponds to the smallest possible value of $n - \gamma\phi$. Firstly, we show in Appendix Appendix A.2 that $n_{I_r}(s_r) \leq 2\lceil k/2\rceil - 2(s_r - 1)$, for $r = 1, \cdots, d_I$. Secondly, since $s_r = n^*_{T(\tau+r)}(\sigma^q_{T(\tau+r-1)})$, it follows from (8) that, for all $r = 1, \cdots, d_I$, $1 \leq r \leq T - \tau$ and $0 \leq \tau \leq T-1$, $s_r \geq n(A^T)$. The smallest possible value of $n - \gamma\phi$ is then obtained when $\tau = 0$, $n_{I_r}(s_r) = 2\lceil k/2\rceil - 2(s_r - 1)$, $s_r = n(A^T)$. Thus, at least $\gamma(\mu+1)$ mutations are needed to trigger an exit from $D(\mathbf{A}^T)$ whenever

$$\begin{aligned}
&\min(n - \gamma\phi) \geq \gamma\left(\lceil pk\rceil + 1\right) \\
&\Rightarrow n - \gamma\left(\mu + \sum_{r=1}^{T-1}\Big(2\lceil k/2\rceil - 2(n(A^T) - 1)\Big)\right) \geq \gamma\left(\lceil pk\rceil + 1\right) \\
&\Rightarrow n - \gamma\Big(\mu + \Big(2\lceil k/2\rceil - 2(n(A^T) - 1)\Big)(T-1)\Big) \geq \gamma\left(\lceil pk\rceil + 1\right) \\
&\Rightarrow n \geq \gamma\left(\mu + \lceil pk\rceil + 1\right) + \gamma\Big(2\lceil k/2\rceil - 2(n(A^T) - 1)\Big)(T-1)
\end{aligned} \tag{14}$$

Since $\mu \geq n(A^T) \geq n^*(A)$, it follows that at least $\gamma(n^*(A)+1)$ mutations are needed to trigger an exit from $D(\mathbf{A}^T)$, that is, $R(\mathbf{A}^T) \geq \gamma(n^*(A)+1)$, whenever

$$n \geq \gamma\left(n^*(A) + \lceil pk\rceil + 1\right) + \gamma\left(2\lceil k/2\rceil - 2(n^*(A) - 1)\right)(T-1) \tag{15}$$

The following example helps to further illustrate the intuition behind this iterative process.

Example 2. Consider an evolutionary process $(U, N, \mathcal{G}_k, P)$ on some $G_2 \in \mathcal{G}_2$ where $A = \{a_1, a_2, a_3, a_4\}$ and U has the following properties: a_4 is the smallest iterated $\frac{1}{2}$-best response equilibrium; the associated sequence of $A^3 = \{a_4\}$ is $\bar{A}^0 = \{a_1\}$, $\bar{A}^1 = \{a_2\}$, $\bar{A}^2 = \{a_3\}$, $A^3 = \{a_4\}$; for all $0 \leq \tau \leq 2$, $n^*_{3(\tau+1)}(\sigma^q_{3\tau}) = 1$. Let $(U, N, \mathcal{G}_k, P)$ start from $\mathbf{A}^3 = \{\mathbf{a}_4\}$, where $\mathbf{a}_4$ is a monomorphic absorbing state containing strategy a_4. At $t = 1$, let one player, $I = \{n\}$, mutate to a strategy in $\bar{A}^0$. Then the strategy configurations evolve from $t = 2$ onward as follows:

$t = 2$ | player n reverts to strategy $a_4 \in A^1$;
for $s_1 = n^*_{31}(\sigma^q_{30}) = 1$, players in $N_{I_1}(s_1) = \{n-1, 1\}$ play a strategy in A^1. This follows by definition of $n^*_{31}(\sigma^q_{30})$;
players in $N\backslash\{I \cup N_{I_1}(s_1)\}$ play a_4.

$t=3$	player n plays a strategy in A^1 since all her neighbours play strategies in A^1 at $t=2$;
	players in $N_{I_1}(s_1)$ play $a_4 \in A^1$ since all their neighbours play a_4 at $t=2$;
	for $s_2 = n^*_{32}(\sigma^q_{31}) = 1$, players in $N_{I_2}(s_2) = \{n-2, 2\}$ play strategies in A^2;
	players in $N\backslash\{I \cup N_{I_1}(s_1) \cup N_{I_2}(s_2)\}$ play a_4.
$t=4$	n plays $a_4 \in A^1$ since all her neighbours play a_4 at $t=3$;
	players in $N_{I_1}(s_1)$ play strategies in A^2. This is because, at $t=3$, each $j \in N_{I_1}(s_1)$ has $\lceil k/2 \rceil = 1$ neighbour (i.e. player n) play a strategy in A^1 and the other (i.e. a neighbour in $N_{I_2}(s_2)$) play a strategy in A^2. Thus, by definition of a $\frac{1}{2}$-best response set, strategies in A^2 are best responses to all $j \in N_{I_1}(s_1)$;
	players in $N_{I_2}(s_2)$ play strategies in a_4 because all their neighbours play a_4 at $t=3$;
	for $s_3 = n^*_{33}(\sigma^q_{32}) = 1$, players in $N_{I_3}(s_3) = \{n-3, 3\}$ play a strategy in $A^3 = \{a_4\}$;
	players in $N\backslash\{I \cup N_{I_1}(s_1) \cup N_{I_2}(s_2) \cup N_{I_3}(s_3)\}$ play $A^3 = \{a_4\}$.

Consider a scenario where, at $t=3$, player n plays $a_2 \in A^1$ and players in $N_{I_2}(s_2)$ play $a_3 \in A^2$. If, after $t=3$ iterations, $n - \Big(1 + n_{I_1}(s_1) + n_{I_2}(s_2)\Big) = \lceil pk \rceil = 1$, then at $t=4$, player n plays a_4; $N_{I_1}(s_1)$ play a_3; $N_{I_2}(s_2)$ play a_4; and $n_{I_3}(s_3)$ play a_3. Thus, $(U, N, \mathcal{G}_k, P)$ converges to an absorbing cycle where players in I, $N_{I_r}(s_r)$ and $N_{I_{r+1}}(s_{r+1})$, for $r = 1, 2, 3$, alternate between a_3 and a_4. For this scenario $(U, N, \mathcal{G}_k, P)$ exits $D(\mathbf{A}^3)$ with only one mutation to $\bar{A}^0$.

However, if $n - \Big(1 + n_{I_1}(s_1) + n_{I_2}(s_2)\Big) \geq \lceil pk \rceil + 1 = 2$, then:

$t=5$	n plays a strategy in A^2;
	players in $N_{I_1}(s_1)$ play strategies in $a_4 \in A^2$;
	players in $N_{I_2}(s_2)$ play a strategy in a_4. This is because, at $t=4$, each $j \in N_{I_2}(s_2)$ has at most one neighbour play a strategy in $\bar{A}^2 = \{a_3\}$ and the rest play a_4;
	players in $N\backslash\{I \cup N_{I_1}(s_1) \cup N_{I_2}(s_2)\}$ play a_4.
$t=5$	All players revert to strategy a_4.

Thus, when $z = n - \Big(\mu + \sum_{r=1}^{2} n_{I_r}(s_r)\Big) = n - 5 \geq \lceil pk \rceil + 1 = 2$ (i.e. $n \geq 7$), at least $\mu + 1 = 2$ mutations are needed to trigger an exit from $D(\mathbf{A}^3)$. Substituting for $T = 3$, $n^*(A) = 1$, $\lceil pk \rceil = 1$, $\gamma = 1$ and $\lceil k/2 \rceil = 1$ into (15) yields the same minimum value of n at which at least 2 mutations are necessary to trigger an exit from $D(\mathbf{A}^3)$.

In deriving the minimum value of n in (15), we assumed that there is no overlap between the r-order neighbourhoods of I. More generally, however, the overlap between the r-order neighbourhoods of I affect the evolution of strategy configurations from $t=4$ onward. That is, if, for some values of $r \in [3, d_I]$, some players in $N_{I_r}(s_r)$ have direct neighbours in $N_{I_{r-2}}(s_{r-2})$, then, in the above iterative process, not all players in $N_{I_r}(s_r)$ switch to strategies in $A^{\tau+r}$ at $t = r + 1$. The conditions in the following lemma account for these overlaps (the proof is presented in Appendix Appendix C).

Lemma 6. *For a symmetric strict coordination game U, let A^T be the smallest iterated p-best response set with $p \le \frac{\lfloor k/2 \rfloor}{k}$. Then, for an evolutionary process $(U, N, \mathcal{G}_k, P)$ on $\mathcal{G}_k$ networks, $R(\mathbf{A}^T) \ge \gamma(n^*(A) + 1)$ whenever*

$$n \ge \gamma\Big(\lceil pk \rceil + 3 - n^*(A)\Big) + 2\gamma\lceil k/2 \rceil(T-1) \tag{16}$$

The third step of the proof of Theorem 4 is a derivation of the upper bound for the coradius of $\mathbf{A}^T$. The following lemma shows that the modified coradius of $\mathbf{A}^T$ is bounded from above by $n(p,k)$ (a detailed proof is presented in Appendix Appendix D).

Lemma 7. *For a symmetric strict coordination game U, let A^T be the smallest iterated p-best response set with $p \le \frac{\lfloor k/2 \rfloor}{k}$. For an evolutionary process $(U, N, \mathcal{G}_k, P)$ on $\mathcal{G}_k$ networks:*

(i) if $T = 1$, then $CR^*(\mathbf{A}^1) \le \lceil p(n-1) \rceil$ for all $\mathcal{G}_{n-1}$ networks;

(ii) if (16) holds, and $\gamma(n^*(A) + 1) > n(p,k)$, then $CR^*(\mathbf{A}^T) \le n(p,k)$.

Lemma 7 (i) follows because, for $\mathcal{G}_{n-1}$ networks, if $T = 1$ and $p \le \frac{\lfloor k/2 \rfloor}{k} = \frac{\lfloor (n-1)/2 \rfloor}{n-1}$, then, starting from any $\mathbf{x} \notin \mathbf{A}^1$, $\lceil p(n-1) \rceil$ mutations to strategies in A^1 sufficiently trigger evolution to $\mathbf{A}^1$. To see why, notice that after $\lceil p(n-1) \rceil$ players mutate to strategies in A^1 at $t = 1$, the remaining $n - \lceil p(n-1) \rceil$ all switch to A^1 at $t = 2$. Since $n - \lceil p(n-1) \rceil \ge \lceil (n-1)/2 \rceil + 1 > \lfloor (n-1)/2 \rfloor$ (i.e. the number of players playing A^1 at $t = 2$ is greater than $\lfloor (n-1)/2 \rfloor$),[14] each player has at least $\lfloor (n-1)/2 \rfloor$ neighbours playing strategies in A^1 at $t = 2$ and the rest playing strategies in $\bar{A}^0$. By definition of p-best response sets, strategies in A^1 are best responses to all players at $t = 3$. Thus, $\lceil p(n-1) \rceil$ mutations sufficiently trigger evolution from any $\mathbf{x} \notin \mathbf{A}^1$ to $\mathbf{A}^1$, so that $CR^*(\mathbf{A}^1) \le \lceil p(n-1) \rceil$.

The proof of Lemma 7 (ii) follows in three steps. Firstly, we demonstrate in Appendix Appendix A.1 that the cost of any path $(\mathbf{x}, \mathbf{x}_1, \cdots, \mathbf{x}_Z) \in S(\mathbf{x}, \mathbf{A}^T)$, where $\mathbf{x}_Z \in \mathbf{A}^T$, from $\mathbf{x} \notin \mathbf{A}^T$ to $\mathbf{A}^T$, can be expressed as the sum of costs of direct paths between absorbing sets that $(\mathbf{x}, \mathbf{x}_1, \cdots, \mathbf{x}_Z)$ traverses. That is, starting from $\mathbf{x} \in D(W)$, let $W_1, \cdots, W_H$, where $W_H \subseteq \mathbf{A}^T$ but $W_h \notin \mathbf{A}^T$ for all $h < H$, be a sequence of absorbing sets through which $(\mathbf{x}, \mathbf{x}_1, \cdots, \mathbf{x}_Z)$ passes consecutively. Denote this path of absorbing sets by $(\mathbf{x}; W, W_1, \cdots, W_H)$ and let $\Gamma(\mathbf{x}, \mathbf{A}^T)$ be a

[14]This follows because when $p \le \frac{\lfloor k/2 \rfloor}{k}$, then for $\mathcal{G}_{n-1}$ networks, $n - \lceil p(n-1) \rceil \ge n - \lfloor (n-1)/2 \rfloor = 1 + (n-1) - \lfloor (n-1)/2 \rfloor = \lceil (n-1)/2 \rceil + 1$.

set of all such paths. Let $C_D(W_h, W_{h+1})$ be the cost of the minimum direct path from $D(W_h)$ to $D(W_{h+1})$ (direct in the sense that it does not pass through the basin of attraction of another absorbing set different from W_h and W_{h+1}). We show in Appendix Appendix A.1 that the modified cost of a minimum path in $S(\mathbf{x}, \mathbf{A}^T)$ is

$$C^*(\mathbf{x}, \mathbf{A}^T) = \min_{(\mathbf{x}; W, W_1, \cdots, W_H) \in \Gamma(\mathbf{x}, \mathbf{A}^T)} \left(C_D(W, W_1) + \sum_{h=1}^{H-1} \Big(C_D(W_h, W_{h+1}) - R(W_h) \Big) \right) \tag{17}$$

Secondly, we show that, for any $W \subseteq \bar{\mathbf{A}}^\tau$, there exists at least one $W' \subseteq \mathbf{A}^\tau$ with $C_D(W, W') \leq n(p, k)$. Specifically, we show that, starting from some $\mathbf{x} \in D(W)$, where $W \subseteq \bar{\mathbf{A}}^\tau$, if $\lceil pk \rceil$ is sufficiently small so that $3\lceil pk \rceil \leq 2(\lceil k/2 \rceil + 1)$, then at most $\lceil pk \rceil$ mutations to strategies in $A^{\tau+1}$ trigger an exit from $D(W)$ to some absorbing set $W' \subseteq \mathbf{A}^\tau$. However, when $3\lceil pk \rceil > 2(\lceil k/2 \rceil + 1)$, $\lceil pk \rceil + 1$ mutations are sufficient.

Thirdly, we show that, if (16) holds, then starting from some $\mathbf{x} \in D(W)$, where $W \subseteq \bar{\mathbf{A}}^\tau$, at least $\gamma(n^*(A) + 1)$ mutations to strategies in $A \backslash A^\tau$ are needed to trigger an exit from $D(W)$ to any absorbing set $W'' \subseteq \bigcup_{v=0}^{\tau-1} \Big(M(\bar{\mathbf{A}}^v) \cup Q(\bar{\mathbf{A}}^v) \cup L(\mathbf{A}^v) \Big)$, so that $C_D(W, W'') \geq \gamma(n^*(A) + 1)$. The proof of this result follows the same steps in the proof of Lemma 6.

Now, let $\gamma(n^*(A) + 1) > n(p, k)$. Then, for any $W \subseteq \bar{\mathbf{A}}^\tau$, there exists some $W' \subseteq \mathbf{A}^\tau$ for which $R(W) = C_D(W, W')$. Thus, if (16) holds, then starting from any $\mathbf{x} \in D(W)$, where $W \subseteq \bar{\mathbf{A}}^\tau$, there exists a sequence of absorbing sets $W_1, \cdots, W_H$, with $W_1 \subseteq \mathbf{A}^\tau$ and $W_H \subseteq \mathbf{A}^T$, along which $R(W) = C_D(W, W_1)$ and $R(W_h) = C_D(W_h, W_{h+1})$, for all $h = 1, \cdots, H-1$. The coradius of $\mathbf{A}^T$ is then:

$$CR^*(\mathbf{A}^T) = \max_{\mathbf{x} \in \mathbf{X} \backslash D(\mathbf{A}^T)} C^*(\mathbf{x}, \mathbf{A}^T) = \max_{\tau \in [0, T-1]} \max_{W \subseteq \bar{\mathbf{A}}^\tau} \min_{W_1 \subseteq \mathbf{A}^\tau} C_D(W, W_1) \tag{18}$$

Since $C_D(W, W_1) \leq n(p, k)$ for all $W \subseteq \bar{\mathbf{A}}^\tau$ with corresponding $W_1 \subseteq \mathbf{A}^\tau$, and for all $\tau \in [0, T-1]$, it follows that when (16) holds and $\gamma(n^*(A) + 1) > n(p, k)$, $CR^*(\mathbf{A}^T) \leq n(p, k)$.

Finally, invoking Ellison (2000, Theorem 2), when $T = 1$, set A^1 is uniquely stochastically stable in all $\mathcal{G}_{n-1}$ networks whenever $n \geq 3$. This is because when these conditions hold, $CR^*(\mathbf{A}^1) \leq \lceil p(n-1) \rceil < \lceil p(n-1) \rceil + 1 \leq R(\mathbf{A}^1)$. When $T = 1$ and $d(G_k) \geq 2$, A^1 is uniquely stochastically stable whenever $n \geq k + 2$. Firstly, for $\mathcal{G}_2$ networks, this follows because $CR^*(\mathbf{A}^1) = \lceil pk \rceil < \lceil k/2 \rceil + 1 \leq R(\mathbf{A}^1)$. For $\mathcal{G}_k$ networks with $k \geq 3$, this follows because $CR^*(\mathbf{A}^1) \leq n(p, k) < \lceil k/2 \rceil + 2 \leq R(\mathbf{A}^1)$.

When $T \geq 2$, set A^T is uniquely stochastically stable if $CR^*(\mathbf{A}^T) \leq n(p, k) < \gamma(n^*(A) + 1) \leq R(\mathbf{A}^T)$ and (16) holds. This condition can be restated as follows. Define γ^* as

$$\gamma^* = \begin{cases} \lceil \frac{n(p,k)}{n^*(A)+1} \rceil & \text{if } \frac{n(p,k)}{n^*(A)+1} \notin \mathbb{N}_+ \\ \lceil \frac{n(p,k)}{n^*(A)+1} \rceil + 1 & \text{if } \frac{n(p,k)}{n^*(A)+1} \in \mathbb{N}_+ \end{cases} \tag{19}$$

Then, for $T \geq 2$, A^T is uniquely stochastically stable if

$$n \geq \gamma^*\Big(\lceil pk \rceil + 3 - n^*(A)\Big) + 2\gamma^*\lceil k/2 \rceil (T-1) \tag{20}$$

5. Discussion and relation to the literature

Theorem 4 establishes the conditions under which the smallest iterated p-best response set is the uniquely stochastically stable set of a BRM model on $\mathcal{G}_k$ networks. It states that when $p \leq \eta(G_k)$, there exists a minimum value of n, $n^* := n^*(p, k, T, \gamma^*)$, such that strategies in A^T are uniquely stochastically stable in all $G_k(n) \in \mathcal{G}_k$ with $n \geq n^*$. This threshold value of n corresponds to the threshold diameter, $d(G_k(n^*))$, where, for all $G_k(n) \in \mathcal{G}_k$ with $n \geq n^*$, $d(G_k(n)) \geq d(G_k(n^*))$. As discussed in Section 3, the intuition is that when $p \leq \eta(G_k)$, strategies in A^T are step-by-step contagious on $\mathcal{G}_k$ networks. Step-by-step contagion then implies that the minimum number of mutations needed to exit the basin of attraction of $\mathbf{A}^T$ (i.e. the radius of $\mathbf{A}^T$) increases with the network diameter. The threshold diameter at which the radius of $\mathbf{A}^T$ is greater than the coradius of $\mathbf{A}^T$ is $d(G_k(n^*))$.

Given p and k, the threshold population size is an increasing function of T and the parameter $\gamma^* := \gamma^*(p, k, n^*(A))$. That is,

$$n^*(p,k,T,\gamma^*) = \begin{cases} 3 & \text{for } T = 1 \text{ and } \mathcal{G}_{n-1} \text{ networks} \\ k+2 & \text{for } T = 1 \text{ and } d(G_k) \geq 2 \\ \gamma^*\Big(\lceil pk \rceil + 3 - n^*(A)\Big) + 2\gamma^*\lceil k/2 \rceil (T-1) & \text{for } T \geq 2 \end{cases} \tag{21}$$

When T is large, the value of n^*, and hence, $d(G_k(n^*))$, that ensures that at least $\gamma^*(n^*(A)+1)$ mutations from γ^* different regions of the network are needed to trigger an exit from $D(\mathbf{A}^T)$, is also large. Parameter γ^* is a ratio of $n(p,k)$, which is the coradius of $\mathbf{A}^T$ (i.e. the net cost of reaching $\mathbf{A}^T$ from any other state), to $n^*(A)$, which is the smallest possible cost of leaving $D(\mathbf{A}^T)$. Given k, parameter $n^*(A)$ is computed from the payoff matrix of the underlying coordination game. If $n^*(A)$ is small relative to $n(p,k)$, then the network must have a larger diameter to ensure that exiting $D(\mathbf{A}^T)$ requires $\gamma^*(n^*(A)+1)$ simultaneous mutations from γ^* different regions of the network.

When the conditions in Theorem 4 are not satisfied, it is easy to construct an example where strategies that do not constitute the smallest iterated p-best response set (e.g. a strategy that is a payoff dominant equilibrium) are stochastically stable. For the coordination game of Figure 1, for example, it is now clear from Theorem 4 why a_3 is uniquely stochastically stable in network

	a_1	a_2	a_3
a_1	10	0	0
a_2	9.1	14	10
a_3	0	12	13

Figure 3: A 3×3 symmetric coordination game with strategy a_3 as the smallest iterated p-best response equilibrium and a_2 is a payoff dominant equilibrium.

$G_4(16)$, but both a_2 and a_3 are stochastically stable in $G_4(10)$, and a_2 is uniquely stochastically stable in network $G_4(5)$. For this game, a_3 is the smallest iterated p-best response equilibrium, for $p > \frac{2}{5}$, so that $\lceil pk \rceil = \lceil \frac{2 \times 4}{5} \rceil = 2$. Since $3\lceil pk \rceil = 2(\lceil k/2 \rceil + 1)$ when $p > \frac{2}{5}$ and $k = 4$, it follows from (10) that $n(p,k) = 2$. Note also that, for this game, $n^*(A) = n(A^2) = 1$, so that $\gamma^* = 2$. Substituting into (21) yields $n^* = 16$. Thus, strategy a_3 is uniquely stochastically stable in network $G_4(16)$ because this network satisfies both $p \leq 1 - \eta(G_4(16))$ and the threshold value of n. However, the number of players in networks $G_4(5)$ and $G_4(10)$ is less than the threshold, n^*.

For an example where a payoff dominant equilibrium, and not the smallest iterated p-best response set, is uniquely stochastically stable, consider the coordination game in Figure 3. This game has similar properties as the game in Figure 1 in that strategy a_3 is the smallest iterated p-best response equilibrium, for $p > \frac{2}{5}$, and in $\mathcal{G}_4$ networks, $n^*(A) = n(a_3) = 1$, $n(p,k) = 2$. For this game, strategy a_2, which is the payoff dominant equilibrium, is uniquely stochastically stable in all $\mathcal{G}_{n-1}$ and $\mathcal{G}_3$ networks. For the latter, it is because $p > \frac{2}{5} > \eta(G_3) = \frac{1}{3}$ for all $\mathcal{G}_3$ networks, and for the former, it is because $\mathcal{G}_{n-1}$ networks have a diameter less than the threshold diameter.

The threshold values of n, and hence, of the corresponding network diameter, presented in (21) are tight. For $\mathcal{G}_{n-1}$ networks, when $n = 2$, $G_1(2)$ is a star network where all strategies are stochastically stable. Thus, the smallest possible value of n at which A^1 strategies are uniquely stochastically stable in $\mathcal{G}_{n-1}$ networks is $n = 3 = n^*$. The second threshold value of n applies to all $G_k(n) \in \mathcal{G}_k$ networks with $d(G_k(n)) \geq 2$. For these networks, $k + 2$ is the smallest possible value of n at which $d(G_k(n)) \geq 2$. For the threshold value of n when $T \geq 2$, the main source of error is the term, $2\gamma^* \lceil k/2 \rceil (T - 1)$. This is because the derivation of n^* for $T \geq 2$ implicitly assumes that all players have $\lceil k/2 \rceil$ neighbours to the left and $\lceil k/2 \rceil$ neighbours to the right. In $\mathcal{G}_k$ networks with k even, where $\lceil k/2 \rceil = \lfloor k/2 \rfloor = k/2$, the error from this term is zero. However, in $\mathcal{G}_k$ networks with k odd, the error from this term is a proportion of $T-1$.

Theorem (4) generalizes existing results on evolutionary dynamics in $\mathcal{G}_k$ networks. For example, the results in Ellison (2000) and Weidenholzer (2012), who find that $\frac{1}{2}$-dominant strategies are uniquely stochastically stable in $\mathcal{G}_k$ networks with k even, are a special case of Theorem 4 when $T = 1$ and A^1 is a singleton set. Peski et al. (2010) finds that in a BRM model, a p-dominant strategy is uniquely stochastically stable in $\mathcal{G}_k$ networks when $p \leq \frac{\lfloor k/2 \rfloor}{k}$. This is a special case of Theorem 4 when $T = 1$ and A^1 is a singleton set. Alós-Ferrer and Weidenholzer (2007) consider a BRM model in $\mathcal{G}_2$ networks and show that a globally pairwise risk dominant strategy (GPRD) of a 3×3 symmetric strict coordination game is stochastically stable if it satisfies the *partial bandwagon property* (PBP).

Strategies a_j and a_l fulfill the PBP if $BR(\sigma^q_{jl}) \subseteq \{a_j, a_l\}$ for all $q \in (0, 1)$. That is, if a player faces a profile where only a_j and a_l are played, no third strategy can be a best response. Alós-Ferrer and Weidenholzer (2007) then show that, given $A = \{a_1, a_2, a_3\}$ of a symmetric strict coordination game, if a_3 is GPRD but not $\frac{1}{2}$-dominant, and $\{a_2, a_3\}$ satisfies the PBP, then for $q = \frac{1}{2}$:

(i) a_3 is uniquely stochastically stable if $BR(\sigma^q_{12}) = a_2$ or $BR(\sigma^q_{12}) = \{a_1, a_2\}$;

(ii) a_3 is uniquely stochastically stable if $BR(\sigma^q_{21}) = a_1$ and $BR(\sigma^q_{31}) = \{a_3, a_2\}$;

(iii) a_1 and a_3 are both stochastically stable if $BR(\sigma^q_{21}) = a_1$ and $BR(\sigma^q_{31}) = a_2$.

In case (i) above, where a_3 is GPRD but not $\frac{1}{2}$-dominant, and that $BR(\sigma^q_{12}) = a_2$ or $BR(\sigma^q_{12}) = \{a_1, a_2\}$, we see that $A^2 = \{a_3\}$ is the smallest iterated $\frac{1}{2}$-best response equilibrium with $\bar{A}^0 = \{a_1\}$ and $\bar{A}^1 = \{a_2\}$. According to Theorem 4, strategy a_3 is uniquely stochastically stable in all $\mathcal{G}_k$ networks with k even. However, in cases (ii) and (iii), the smallest iterated $\frac{1}{2}$-best response set is $A^0 = A$. Thus, for these two scenarios, Theorem 4 fails to isolate a single unique stochastically stable strategy in $\mathcal{G}_k$ networks. And as evident from Alós-Ferrer and Weidenholzer (2007), any of the three strategies can belong to the stochastically stable set.

6. Potential extensions to arbitrary networks

Although the results of Theorem 4 hold primarily for $\mathcal{G}_k$ networks, the principles of contagion that underlie these results extend to arbitrary networks. There are multiple network measures that determine the feasibility of contagion. We focused on group cohesion above because it is one of the aggregate measures that captures the overall topology of the network, and it is easy to compute. Here, we discuss three different network measures that determine the feasibility of contagion and how they can be used to extend the results of Theorem 4 to arbitrary networks.

The first network measure that determines the feasibility of contagion is the largest degree of the network. As discussed in Section 3, if (U, N, G, P) starts from some configuration $\mathbf{x}$ containing either only strategies in $\bar{A}^\tau$ or strategies in $\bar{A}^\tau$ and $A^{\tau+1}$, strategies in $A^{\tau+1}$ spread contagiously to the whole network from $\mu^*(A^{\tau+1}; \mathbf{x}) = 2$ mutations whenever $p \leq 1/\Delta(G)$. Similarly, strategies in A^T are step-by-step contagious in any strongly connected network whenever $p \leq 1/\Delta(G)$.

More generally, when $p \leq 1/\Delta(G)$, the structure of $\mathbf{A}$, the set of all absorbing sets of (U, N, G, P), is identical to that in (12). That is, strategies in $A^{\tau+1}$ cannot coexist with strategies in $\bar{A}^\tau$ in an absorbing state. To see why, notice that any state where one player plays a strategy in $A^{\tau+1}$ and the rest play strategies in $\bar{A}^\tau$ is either transient or belongs to an absorbing cycle. Moreover, when any two neighbouring players play strategies in $A^{\tau+1}$ and the rest play strategies in $\bar{A}^\tau$, (U, N, G, P) will converge to an absorbing set containing only strategies in $A^{\tau+1}$. Thus, strategies in $A^{\tau+1}$ cannot coexist with strategies in $\bar{A}^\tau$ in an absorbing state. Since one mutation to $A^{\tau+1}$ triggers an exit from the basin of attraction of any $W \subseteq \bar{\mathbf{A}}^\tau$, it follows that $CR^*(\mathbf{A}^T) = 1$ whenever $p \leq 1/\Delta(G)$.

Now, consider a scenario where $n^*_{T(\tau+1)}(\sigma^q_{T\tau}) = 1$ for all $\tau = 0, 1, \cdots, T-1$. Then, starting from any $\mathbf{x} \in \mathbf{A}^T$, if one player, i, mutates to a strategy in $\bar{A}^0$ at $t = 1$, the following iterative process unfolds (where N_{i_r} is the set of players at distance r from i):

$t = 2$	$i \cup N_{i_1}$ all play strategies in A^1; $N \backslash \{i \cup N_{i_1}\}$ all play strategies in A^T.
$t = 3$	$i \cup N_{i_1}$ all play strategies in A^1; N_{i_2} all play strategies in A^2; $N \backslash \{i \cup N_{i_1} \cup N_{i_2}\}$ play strategies in A^T.
$t = 4$	$i \cup N_{i_1}$ play strategies in A^1; N_{i_2} play strategies in A^2; N_{i_3} play strategies in A^3; $N \backslash \{i \cup N_{i_1} \cup N_{i_2} \cup N_{i_3}\}$ play strategies in A^T.
$--$	$-----------------------------------$
$t = T$	$i \cup N_{i_1}$ play strategies in A^1; N_{i_2} play strategies in A^2; N_{i_3} play strategies in A^3; $\cdots$; $N_{i_{T-1}}$, play strategies in A^{T-1}; $N \backslash \{i \cup N_{i_1} \cup \cdots \cup N_{i_{T-1}}\}$ play strategies in A^T.

For this scenario, if $n - \left(1 + \sum_{r=1}^{T-1} n_{i_r}\right) \geq 2$, then players in $N \backslash \{i \cup N_{i_1} \cup \cdots \cup N_{i_{T-1}}\}$ play strategies in A^T from $t = T+1$ onward. And following the same steps in Section 4, this condition ensures that (U, N, G, P) eventually reverts to $\mathbf{A}^T$. Thus, at least two mutations (from different regions of the network) are needed to trigger an exit from the basin of attraction of $\mathbf{A}^T$ whenever $p \leq 1/\Delta(G)$ and $n - \left(1 + \sum_{r=1}^{T-1} n_{i_r}\right) \geq 2$. Note that $n - \left(1 + \sum_{r=1}^{d_i} n_{i_r}\right) = 0$,[15], and hence, when

[15] This follows because $1 + \sum_{r=1}^{d_i} n_{i_r} = n$.

$n - \left(1 + \sum_{r=1}^{T-1} n_{i_r}\right) \geq 2$, $d_i > T - 1$, or equivalently, $d_i \geq T$. Since $\min_{i \in N} d_i = \frac{d(G)}{2}$, it follows that at least two mutations are needed to trigger an exit from the basin of attraction of $\mathbf{A}^T$, (i.e. $R(\mathbf{A}^T) \geq 2$) whenever $p \leq 1/\Delta(G)$ and $\frac{d(G)}{2} \geq T$. Thus, when $p \leq 1/\Delta(G)$, strategies in A^T are uniquely stochastically stable in G if $\frac{d(G)}{2} \geq T$.

The second network measure that determines the feasibility of contagion is the *contagion threshold*. Opolot (2018) develops a measure of the contagion threshold for finite networks that is analogous to the contagion threshold for unbounded networks developed by Morris (2000). Let B_{i_r} be the closed rth *neighbourhood* of i (i.e. the set of all players within distance r from i, including i) and let b_{i_r} be the corresponding cardinality. For any $j \in N_{i_r}$, let $\eta_j^*(B_{i_{r-1}})$ be the proportion of j's neighbours in $B_{i_{r-1}}$. Then the *neighbourhood contagion threshold* of G is defined as

$$\eta^*(G) = \min_{i \in N} \min_{r \in [2, d_i]} \min_{j \in N_{i_r}} \eta_j^*(B_{i_{r-1}}).$$

From this definition of neighbourhood contagion threshold, starting from a configuration $\mathbf{x}$ containing either only strategies in $\bar{A}^\tau$ or both strategies in $\bar{A}^\tau$ and $A^{\tau+1}$, if $p \leq \eta^*(G)$, then strategies in $A^{\tau+1}$ spread contagiously from B_{i_2} of any $i \in N$ to the entire network. More generally, starting from $\mathbf{x}$, if all players in B_{i_1} of any $i \in N$ mutate to strategies in $A^{\tau+1}$, (U, N, G, P) will converge to either an absorbing set containing only strategies in $A^{\tau+1}$ or an absorbing cycle containing strategies in both $\bar{A}^\tau$ and $A^{\tau+1}$ (Opolot, 2018). This implies that (U, N, G, P) can evolve from any $\mathbf{x} \notin \mathbf{A}^T$ to $\mathbf{A}^T$ through step-by-step contagion, whereby, at most $\min_{i \in N} b_{i_1}$ mutations trigger evolution from one absorbing set to the next. Thus, when the network diameter is sufficiently large, the coradius of $\mathbf{A}^T$ is bounded from above by $\min_{i \in N} b_{i_1}$.

Similarly, starting from any $\mathbf{x} \in \mathbf{A}^T$, if players in B_{i_1} of any $i \in N$ mutate to strategies in some $\bar{A}^\tau$, for $0 \leq \tau \leq T-2$, (U, N, G, P) will eventually revert to $\mathbf{A}^T$ if the network diameter is sufficiently large. Thus, there exists some threshold value of the network diameter above which the number of mutations needed to exit the basin of attraction of $\mathbf{A}^T$ is larger than $\min_{i \in N} b_{i_1}$. Strategies in A^T are then uniquely stochastically stable in all strongly connected networks with $p \leq \eta^*(G)$ and diameter greater or equal to the threshold.

The third network measure combines the notion of group cohesion and contagion threshold. The maximum group cohesion as defined in Section 3 does not guarantee contagion in arbitrary networks. Consider an arbitrary network in Figure 4. The maximum group cohesion is $\eta(G) = \frac{3}{4}$, which corresponds to the cohesiveness of subgroup C, consisting of players $\{13, \cdots, 16\}$. Subgroup B is $\frac{2}{3}$-cohesive; subgroup D is $\frac{2}{3}$-cohesive; and subgroup E is $\frac{1}{2}$-cohesive.

Consider a scenario where $p \leq 1 - \eta(G) = \frac{1}{4}$. Then, starting from any configuration $\mathbf{x}$ containing only strategies in $\bar{A}^\tau$, strategies in $A^{\tau+1}$ spread contagiously from E, or D or $B \cup C$,

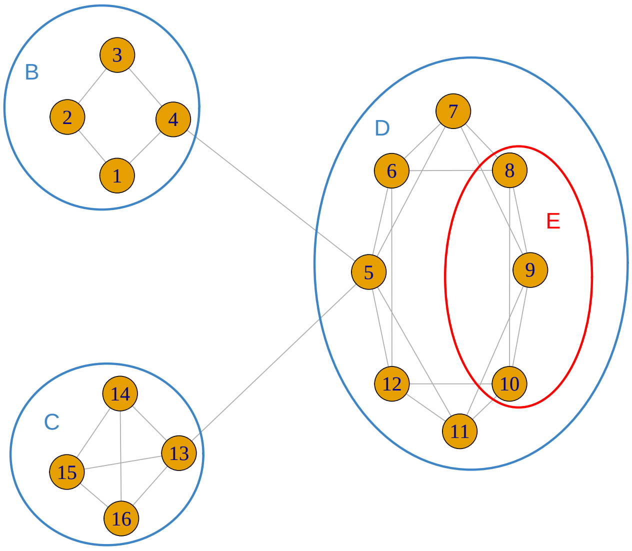


Figure 4: An arbitrary network with multiple cohesive groups: subgroup B is $\frac{2}{3}$-cohesive; subgroup C is $\frac{3}{4}$-cohesive; subgroup D is $\frac{2}{3}$-cohesive; and subgroup E is $\frac{1}{2}$-cohesive.

but not from only B and C.[16] This implies that there are absorbing states where strategies in $\bar{A}^{\tau}$ coexist with strategies in $A^{\tau+1}$, and hence, strategies in A^T are not step-by-step contagious. It also implies that, starting from any $\mathbf{x} \in \mathbf{A}^T$, if players in subgroup D mutate to strategies in some $\bar{A}^{\tau}$, for $\tau \leq T-2$, (U, N, G, P) need not revert to $\mathbf{A}^T$ regardless of the diameter of the network. Thus, the network diameter does not determine when strategies in A^T are uniquely stochastically stable.

To solve this problem, we can define a measure of contagion threshold that uses a cohesive subgroup, and not a closed 1st neighbourhoods, as a reference group. First, we identify all groups in G with group cohesion of at least $\frac{1}{2}$; let $\mathcal{Z}(G)$ denote the set of all these groups. Second, we define the *subgroup contagion threshold* as follows. For each $Z \in \mathcal{Z}(G)$, let N_{Z_1} be the set of players in $N \backslash Z$ with at least one neighbour in Z; N_{Z_2} is the set of players in $N \backslash \{Z \cup N_{Z_1}\}$ with at least one neighbour in N_{Z_1}; and more generally, N_{Z_r} is the set of players in $N \backslash \{Z \cup N_{Z_1} \cup \cdots \cup N_{Z_{r-1}}\}$ with at least one neighbour in $N_{Z_{r-1}}$. Let $d_Z(G)$ be the diameter of Z in network G, that is, the value of r at which $N_{Z_r} \neq \emptyset$ but $N_{Z_{r+1}} = \emptyset$. For any $j \in N_{Z_r}$, let $\eta'_j(N_{Z_{r-1}})$ be the proportion of j's neighbours in $N_{Z_{r-1}}$. Then the subgroup contagion threshold of G is defined as

$$\eta'(G) = \min_{Z \in \mathcal{Z}(G)} \min_{r \in [1, d_Z]} \min_{j \in N_{Z_r}} \eta'_j(N_{Z_{r-1}}) \tag{22}$$

[16] For example, starting from $\mathbf{x}$, if players in B all mutate to strategies in $A^{\tau+1}$, strategies in $A^{\tau+1}$ need not spread contagiously to D and eventually to C. This is because $p \leq \frac{1}{4}$ but player $5 \in D$ has only proportion $\frac{1}{6}$ of her neighbours in B, and hence, strategies in $A^{\tau+1}$ need not be best response to 5 even when all players in B play strategies in $A^{\tau+1}$.

The definition of the subgroup group cohesion in (22) ensures that when $p \leq \eta'(G)$, strategies in a p-best response set spread contagiously on G. Specifically, if $p \leq \eta'(G)$, then starting from some configuration $\mathbf{x}$ containing either only strategies in $\bar{A}^\tau$ or strategies in $\bar{A}^\tau$ and $A^{\tau+1}$, strategies in $A^{\tau+1}$ spread contagiously from any group of players with cohesiveness of at least $\frac{1}{2}$. For example, in the network of Figure 4, strategies in $A^{\tau+1}$ spread contagiously from $\{1, 2\}$, which is the smallest group with cohesiveness of at least $\frac{1}{2}$.

More generally, when $p \leq \eta'(G)$, the structure of $\mathbf{A}$ is identical to that in (12).[17] This implies that (U, N, G, P) can evolve from any $\mathbf{x} \notin \mathbf{A}^T$ to $\mathbf{A}^T$ through step-by-step contagion. Along a path of step-by-step contagion from $\mathbf{x} \notin \mathbf{A}^T$ to $\mathbf{A}^T$, the minimum number of mutations that trigger evolution from one absorbing set to the next is bounded from above by the size of the smallest $Z \in \mathcal{Z}(G)$. This in turn implies that, when the network diameter is sufficiently large, the coradius of $\mathbf{A}^T$ is bounded from above by the size of the smallest $Z \in \mathcal{Z}(G)$. Similarly, when $p \leq \eta'(G)$, then starting from any $\mathbf{x} \in \mathbf{A}^T$, if players in any $Z \in \mathcal{Z}(G)$ mutate to strategies in some $\bar{A}^\tau$, for $0 \leq \tau \leq T-2$, (U, N, G, P) will eventually revert to $\mathbf{A}^T$ if the network diameter is sufficiently large. Put together, these observations imply that when $p \leq \eta'(G)$, there exists a threshold network diameter such that $CR^*(\mathbf{A}^T) < R(\mathbf{A}^T)$ for all networks with diameter greater or equal to the threshold.

These three measures are just a few examples of network measures that determine the feasibility of contagion. Different measures are appropriate for different networks. Ultimately, the chosen measure must ensure that the threshold value of p below which the smallest iterated p-best response set is uniquely stochastically stable (i.e. the values of $1/\Delta(G)$, $\eta(G)$, $\eta'(G)$, and $\eta^*(G)$) is as close to $\frac{1}{2}$ as possible for a given network. For example, for the network in Figure 5a, the maximum group cohesion is a more suitable measure than the contagion threshold. But for the network in 5b, there is no difference between the maximum group cohesion and the contagion threshold, and hence, any of the two can be used.

[17]This is because, first, if $p \leq \eta'(G)$, then starting from any $\mathbf{x}$ where players in some $Z \in \mathcal{Z}(G)$ all play strategies in $A^{\tau+1}$ and all other players play strategies in A^τ, strategies in $A^{\tau+1}$ spread contagiously to the entire network so that (U, N, G, P) converges to an absorbing set containing only strategies in $A^{\tau+1}$. Second, any state where strategies in $\bar{A}^\tau$ coexist with strategies in $A^{\tau+1}$ in any $Z \in \mathcal{Z}(G)$ is either transient or belongs to an absorbing cycle. If a configuration where strategies in $\bar{A}^\tau$ coexist with strategies in $A^{\tau+1}$ in any $Z \in \mathcal{Z}(G)$ is an absorbing state, then there exists a subgroup $Z' \subset Z$ with $\min_{j \in N_{Z'_r}} \eta'_j(N_{Z'_{r-1}}) > p$, for some $r \in [1, d'_Z]$, which contradicts the condition that $p \leq \eta'(G)$.

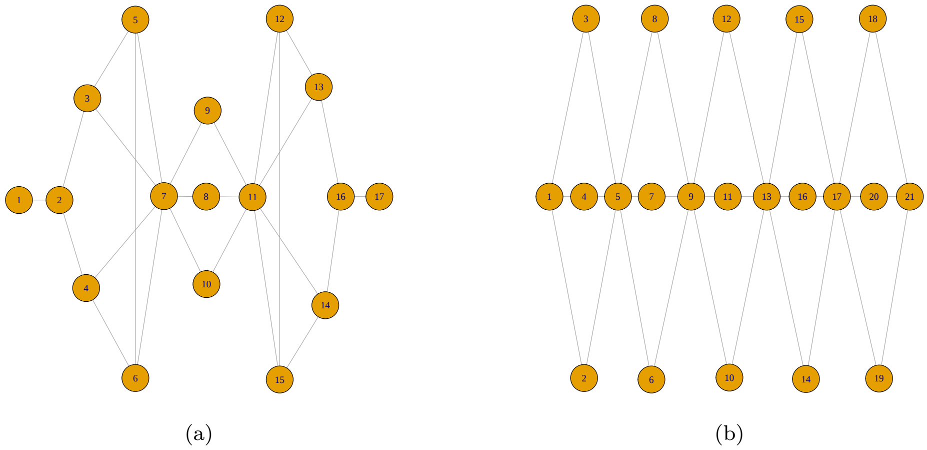


Figure 5: (a) An example of a network with $\eta(G) = \eta'(G) = \frac{3}{7}$ and $\eta^*(G) = \frac{2}{7}$. (b) A network with $\eta(G) = \eta'(G) = \eta^*(G) = \frac{1}{2}$.

7. Concluding remarks

Evolutionary game models with persistent randomness have become useful tools for equilibrium selection in games with multiple equilibria. Recent studies, however, show that the predictions of stochastic stability are not robust to the network structure. The implication of this finding is that for every evolutionary model, the modeller must keep track of the identity of the players (e.g. whom each player interacts with) when computing stochastically stable outcomes. This limits the applicability of evolutionary models since the time complexity of stochastic stability algorithms grows exponentially with the number of players. The lack of robustness of stochastic stability to the network structure also creates difficulties for experimentalists aiming to isolate and test the behavioural assumptions of evolutionary models.

One approach that can be used to establish robustness of stochastic stability to the network structure is to identify network measures that determine stochastically stable outcomes. In this paper, we adopt this approach and use the properties of the process of contagion to identify the suitable network measures. We demonstrate that the network measures that determine the feasibility of contagion also determine stochastically stable outcomes. Two of such measures are the maximum group cohesion and the network diameter, a measure of the cohesiveness of the whole network.

We first show that in regular cyclic networks, there exists a threshold value of the network diameter above which strategies in the smallest *iterated p-best response set*, for p equal to the maximum group cohesion, are uniquely stochastically stable. We then discuss alternative network measures that determine the feasibility of contagion and how they can be used to extend our results to arbitrary networks. However, the exact relationships between the suggested network measures and stochastic stability are left as avenues for future research.

Appendix A. Preliminary definitions and lemmas

Appendix A.1. Reduced form of the modified coradius

This section demonstrates that the coradius of $\mathbf{A}^T$, $CR^*(\mathbf{A}^T)$ can be rewritten as the costs of paths of absorbing sets. Recall that for any pair of sets, Z and Z', $S(Z, Z')$ is the set of all paths from Z to Z'. These paths may pass through a set of multiple absorbing sets.

Let $W_h \subset \mathbf{A}$ and $W_{h+1} \subset \mathbf{A}$ be a pair of disjoint absorbing sets with respective basins of attraction $D(W_h)$ and $D(W_{h+1})$. We define a subset $S_D(W_h, W_{h+1}) \subseteq S(W_h, W_{h+1})$ as a set of all direct paths from $D(W_h)$ and $D(W_{h+1})$. Paths in $S_D(W_h, W_{h+1})$ are direct in the sense that they do not pass through any other absorbing set so that every state along a typical $(\mathbf{x}_{h_1}, \mathbf{x}_{h_2}, \cdots, \mathbf{x}_{h_Z}) \in S_D(W_h, W_{h+1})$ is either in $D(W_h)$ and $D(W_{h+1})$, and $\mathbf{x}_{h_1} \in D(W_h)$ and $\mathbf{x}_{h_Z} \in D(W_{h+1})$. The cost of a minimum path in $S_D(W_h, W_{h+1})$ is

$$C_D(W_h, W_{h+1}) = \min_{(\mathbf{x}_{h_1}, \mathbf{x}_{h_2}, \cdots, \mathbf{x}_{h_Z}) \in S_D(W_h, W_{h+1})} c(\mathbf{x}_{h_1}, \mathbf{x}_{h_2}, \cdots, \mathbf{x}_{h_Z}) \tag{A.1}$$

where $c(\mathbf{x}_{h_1}, \mathbf{x}_{h_2}, \cdots, \mathbf{x}_{h_Z}) = \sum_{z=1}^{Z-1} c(\mathbf{x}_{h_z}, \mathbf{x}_{h_{z+1}})$.

Following preceding definitions, $S(\mathbf{x}, \mathbf{A}^T)$ is a set of all paths from $\mathbf{x} \in \mathbf{X} \backslash \mathbf{A}^T$ to $\mathbf{A}^T$. Let $W_1, W_2, \cdots, W_H$, where $\mathbf{x} \in D(W_1)$ and $W_H \subseteq \mathbf{A}^T$ but $W_h \notin \mathbf{A}^T$ for $h < H$, be a sequence of absorbing sets through which some path in $S(\mathbf{x}, \mathbf{A}^T)$ passes consecutively. Let $(\mathbf{x}; W_1, W_2, \cdots, W_{H-1}, W_H)$ be the path of these absorbing sets, starting from $\mathbf{x} \in D(W_1)$, and let $\Gamma(\mathbf{x}, \mathbf{A}^T)$ be a set of all such paths.[18]

We can then express a typical path in $S(\mathbf{x}, \mathbf{A}^T)$ as a union of direct paths between pairs $D(W_h)$ and $D(W_{h+1})$; that is, a typical path in $S(\mathbf{x}, \mathbf{A}^T)$ takes the form $\bigcup_{h=1}^{H-1} (\mathbf{x}_{h_1}, \mathbf{x}_{h_2}, \cdots, \mathbf{x}_{h_Z})$. The cost $c\left(\bigcup_{h=1}^{H-1} (\mathbf{x}_{h_1}, \mathbf{x}_{h_2}, \cdots, \mathbf{x}_{h_Z})\right)$ of some $\left(\bigcup_{h=1}^{H-1} (\mathbf{x}_{h_1}, \mathbf{x}_{h_2}, \cdots, \mathbf{x}_{h_Z})\right) \in S(\mathbf{x}, \mathbf{A}^T)$ is then

$$c\left(\cup_{h=1}^{H-1} (\mathbf{x}_{h_1}, \mathbf{x}_{h_2}, \cdots, \mathbf{x}_{h_Z})\right) = \sum_{h=1}^{H-1} \sum_{z=1}^{Z-1} c(\mathbf{x}_{h_z}, \mathbf{x}_{h_{z+1}}) = \sum_{h=1}^{H-1} c(\mathbf{x}_{h_1}, \mathbf{x}_{h_2}, \cdots, \mathbf{x}_{h_Z}) \tag{A.2}$$

From (2), the modified cost of a typical path $\left(\cup_{h=1}^{H} (\mathbf{x}_{h_1}, \mathbf{x}_{h_2}, \cdots, \mathbf{x}_{h_Z})\right) \in S(\mathbf{x}, \mathbf{A}^T)$ is given by,

$$\begin{aligned} C^*\left(\cup_{h=1}^{H} (\mathbf{x}_{h_1}, \mathbf{x}_{h_2}, \cdots, \mathbf{x}_{h_Z})\right) &= \sum_{h=1}^{H-1} c(\mathbf{x}_{h_1}, \mathbf{x}_{h_2}, \cdots, \mathbf{x}_{h_Z}) - \sum_{h=2}^{H-1} R(W_h) \\ &= c(\mathbf{x}_{1_1}, \mathbf{x}_{1_2}, \cdots, \mathbf{x}_{1_Z}) + \sum_{h=2}^{H-1} \left(c(\mathbf{x}_{h_1}, \mathbf{x}_{h_2}, \cdots, \mathbf{x}_{h_Z}) - R(W_h) \right) \end{aligned} \tag{A.3}$$

[18]The difference between $\Gamma(\mathbf{x}, \mathbf{A}^T)$ and $S(\mathbf{x}, \mathbf{A}^T)$ is that a typical path in the former consists of absorbing sets as nodes where the cost of a directed edge $W_h \to W_{h+1}$ is $C_D(W_h, W_{h+1})$, while the nodes in the latter can be any states in $\mathbf{X}$.

The modified cost of a minimum path in $S(\mathbf{x}, \mathbf{A}^T)$ is then

$$\begin{aligned}
C^*(\mathbf{x}, \mathbf{A}^T) &= \min_{\left(\bigcup_{h=1}^{H-1}(\mathbf{x}_{h_1}, \mathbf{x}_{h_2}, \cdots, \mathbf{x}_{h_Z})\right) \in S(\mathbf{x}, \mathbf{A}^T)} C^*\left(\cup_{h=1}^{H}(\mathbf{x}_{h_1}, \mathbf{x}_{h_2}, \cdots, \mathbf{x}_{h_Z})\right) \\
&= \min_{(\mathbf{x}; W_1, W_2, \cdots, W_{H-1}, W_H) \in \Gamma(\mathbf{x}, \mathbf{A}^T)} \Bigg[\min_{(\mathbf{x}_{h_1}, \mathbf{x}_{h_2}, \cdots, \mathbf{x}_{h_Z}) \in S_D(W_h, W_{h+1})} c(\mathbf{x}_{1_1}, \mathbf{x}_{1_2}, \cdots, \mathbf{x}_{1_Z}) \\
&+ \sum_{h=2}^{H-1} \left(\min_{(\mathbf{x}_{h_1}, \mathbf{x}_{h_2}, \cdots, \mathbf{x}_{h_Z}) \in S_D(W_h, W_{h+1})} c(\mathbf{x}_{h_1}, \mathbf{x}_{h_2}, \cdots, \mathbf{x}_{h_Z}) - R(W_h) \right) \Bigg] \\
&= \min_{(\mathbf{x}; W_1, W_2, \cdots, W_{H-1}, W_H) \in \Gamma(\mathbf{x}, \mathbf{A}^T)} \left(C_D(W_1, W_2) + \sum_{h=2}^{H-1} \left(C_D(W_h, W_{h+1}) - R(W_h) \right) \right) \\
&= \min_{(\mathbf{x}; W_1, W_2, \cdots, W_{H-1}, W_H) \in \Gamma(\mathbf{x}, \mathbf{A}^T)} C^*(\mathbf{x}; W_1, W_2, \cdots, W_{H-1}, W_H) \qquad \text{(A.4)}
\end{aligned}$$

Thus, to compute the minimum cost of evolving from $\mathbf{x}$ to $\mathbf{A}^T$, it is sufficient to focus on the costs of direct paths between absorbing sets through which paths in $S(\mathbf{x}, \mathbf{A}^T)$ traverse. The modified coradius $CR^*(\mathbf{A}^T)$ is then the maximum $C^*(\mathbf{x}, \mathbf{A}^T)$ over all $\mathbf{x} \in \mathbf{X} \backslash \mathbf{A}$. That is,

$$CR^*(\mathbf{A}^T) = \max_{\mathbf{x} \in \mathbf{X} \backslash \mathbf{A}} C^*(\mathbf{x}, \mathbf{A}^T)$$

Appendix A.2. τ-order neighbourhoods of $\mathcal{G}_k$ networks

Given any $G_k \in \mathcal{G}_k$, let $I = \{i, i+1, \cdots, i+\mu-1\}$ be a set of adjacently placed players in G_k, where μ is the cardinality of I. We write $N_{I_1}(s_1)$ for the set of players in $N \backslash I$ with at least s_1 neighbours in I (i.e. for each $j \in N_{I_1}(s_1)$, $\frac{|N_j \cap I|}{k} \geq s_1$, where $|N_j \cap I|$ is the cardinality of $N_j \cap I$); $N_{I_2}(s_2)$ is the set of players in $N \backslash \{I \cup N_{I_1}(s_1)\}$ with at least s_2 neighbours in $N_{I_1}(s_1)$; and more generally $N_{I_\tau}(s_\tau)$, for $0 \leq \tau \leq d_I$, is the set of players in $N \backslash \{I \cup N_{I_1}(s_1) \cup \cdots \cup N_{I_{\tau-1}}(s_{\tau-1})\}$ with at least s_τ neighbours in $N_{I_{\tau-1}}(s_{\tau-1})$. In this definition, $N_{I_0}(s_0) = I$, and d_I is the value of τ at which $N_{I_\tau}(s_\tau) \neq \emptyset$ but $N_{I_{\tau+r}}(s_{\tau+r}) = \emptyset$ for all $r \geq 1$. When the values of $s_1, s_2, \cdots, s_\tau$ are clearly defined, we refer to each $N_{I_\tau}(s_\tau)$ as the τ-order neighbours of I, and write $n_{I_\tau}(s_\tau)$ for its cardinality.

Given any $G_k \in \mathcal{G}_k$, I and $s_1, s_2, \cdots, s_\tau$, we can derive both the composition and cardinality of each $N_{I_\tau}(s_\tau)$, for $1 \leq \tau \leq d_I$. We are particularly interested in the cardinalities of $N_{I_\tau}(s_\tau)$ because they determine the evolution of $(U, N, \mathcal{G}_k, P)$ after μ simultaneous mutations. Each $N_{I_\tau}(s_\tau)$ consists of two subsets of players, $N^+_{I_\tau}(s_\tau)$ to the "right" of I, and $N^-_{I_\tau}(s)$ to the "left" of I. That is, given $I = \{i, i+1, \cdots, i+\mu-1\}$, $N^+_{I_1}(s_1) \cup N^+_{I_2}(s_2) \cup \cdots \cup N^+_{I_{d_I}}(s_{d_I})$ are players in the positive direction of I (i.e. $\mu, \mu+1, \cdots, \frac{n}{2}$) and $N^-_{I_1}(s_1) \cup N^-_{I_2}(s_2) \cup \cdots \cup N^-_{I_{d_I}}(s_{d_I})$ are players in the opposite direction (i.e. $i-1, i-2, \cdots, i-\frac{n}{2}$, where we consider $i = n$). Lemmas 8 and 9 below provide the values and bounds of the cardinalities $n^+_{I_\tau}(s_\tau)$, $n^-_{I_\tau}(s_\tau)$ and $n_{I_\tau}(s_\tau)$ of $N^+_{I_\tau}(s_\tau)$, $N^-_{I_\tau}(s_\tau)$ and $N_{I_\tau}(s_\tau)$ respectively.

Lemma 8. *For $\mathcal{G}_k$ networks with k even, the following relations hold: $n^+_{I_\tau}(s_\tau) = n^-_{I_\tau}(s_\tau) = \frac{k}{2} - (s_\tau - 1)$ and $n_{I_\tau}(s_\tau) = k - 2(s_\tau - 1)$, for $1 \le \tau \le d_I$, where $1 \le s_\tau \le n^+_{I_{\tau-1}}(s_{\tau-1})$.*

Proof. Consider a G_k network with k even. Given $I = \{i, i+1, \cdots, i+\mu-1\}$, if $\mu \le \frac{k}{2}$, then every player in $\{i+\mu, i+\mu+1, \cdots, i+\frac{k}{2}\}$ is directly connected to all players in I. This is because player i has $\frac{k}{2}$ neighbours to the right, and hence, directly connected to player $i+\frac{k}{2}$. This implies that a total of $i+\frac{k}{2}-(i+\mu-1) = \frac{k}{2}-(\mu-1)$ players to the right of I have all players in I as their direct neighbours. Additionally, player $i+\frac{k}{2}+1$ is directly connected to player $i+1$ (since she has $\frac{k}{2}$ neighbours to the left), and hence, has $\mu-1$ neighbours in I; player $i+\frac{k}{2}+2$ has $\mu-2$ neighbours in I; and more generally, player $i+\frac{k}{2}+r$, for $0 \le r \le \mu-1$, has $\mu-r$ neighbours in I. Thus, to the right of I, there are $\frac{k}{2}-(\mu-1)$ players with μ neighbours in I; $\frac{k}{2}-(\mu-2)$ players with at least $\mu-1$ neighbours in I; $\frac{k}{2}-(\mu-3)$ players with at least $\mu-2$ neighbours in I; and more generally, for $0 \le r \le \mu-1$, there are $\frac{k}{2}-(\mu-r-1)$ players to the right of I with at least $\mu-r$ neighbours in I. This implies that

$$N^+_{I_1}(\mu-r) = \left\{i+\mu, i+\mu+1, \cdots, i+\frac{k}{2}+r\right\}$$

Letting $s_1 = \mu-r$ so that $r = \mu-s_1$, and $0 \le r \le \mu-1$ implies that $1 \le s_1 \le \mu$, then:

$$N^+_{I_1}(s_1) = \left\{i+\mu, i+\mu+1, \cdots, i+\mu+\frac{k}{2}-s_1\right\} \quad \text{for } \mu \le \frac{k}{2} \text{ and } 1 \le s_1 \le \mu \tag{A.5}$$

If $\mu > \frac{k}{2}$, then player $i+\mu$ is directly connected to player $i+\mu-\frac{k}{2}$, and hence, has $\frac{k}{2}$ of her neighbours in I. Player $i+\mu+1$ is directly connected to $i+\mu-\frac{k}{2}+1$, and hence, has $\frac{k}{2}-1$ neighbours in I; player $i+\mu+2$ has $\frac{k}{2}-2$ neighbours in I; and more generally, for $0 \le r \le \frac{k}{2}-1$, player $i+\mu+r$ has $\frac{k}{2}-r$ neighbours in I. This implies that there is one player with $\frac{k}{2}$ neighbours in I; two players with at least $\frac{k}{2}-1$ neighbours in I; three players with at least $\frac{k}{2}-2$ neighbours in I; and more generally, for $0 \le r \le \frac{k}{2}-1$, there are $r+1$ players with at least $\frac{k}{2}-r$ neighbours in I. Thus,

$$N^+_{I_1}\left(\frac{k}{2}-r\right) = \{i+\mu, i+\mu+1, \cdots, i+\mu+r\}$$

Letting $s_1 = \frac{k}{2}-r$ so that $r = \frac{k}{2}-s_1$, where $0 \le r \le \frac{k}{2}-1$ implies that $1 \le s_1 \le \frac{k}{2}$, we have

$$N^+_{I_1}(s_1) = \left\{i+\mu, i+\mu+1, \cdots, i+\mu+\frac{k}{2}-s_1\right\} \quad \text{for } \mu > \frac{k}{2} \text{ and } 1 \le s_1 \le \frac{k}{2} \tag{A.6}$$

Since $N^-_{I_1}(s_1)$ and $N^+_{I_1}(s_1)$ are symmetric around I, it follows that

$$N^-_{I_1}(s_1) = \left\{i-1, i-2, \cdots, i-\frac{k}{2}+s_1-1\right\} \tag{A.7}$$

where $1 \le s_1 \le \frac{k}{2}$ when $\mu > \frac{k}{2}$, and $1 \le s_1 \le \mu$ when $\mu \le \frac{k}{2}$.

It then follows from (A.6), (A.6) and (A.7) that the cardinalities of $N^+_{I_1}(s_1)$ and $N^-_{I_1}(s_1)$ are $n^+_{I_1}(s_1) = n^-_{I_1}(s_1) = \frac{k}{2} - (s_1 - 1)$, and the corresponding cardinality of $N_{I_1}(s_1)$ is $n_{I_1}(s_1) = k - 2(s_1 - 1)$.

The steps above also apply to the derivations of $N^+_{I_\tau}(s_\tau)$, $N^-_{I_\tau}(s_\tau)$ and $N_{I_\tau}(s_\tau)$, for $\tau \geq 2$. Let $N^+_{I_{\tau-1}}(s_{\tau-1}) = \{j, j+1, \cdots, j + n^+_{I_{\tau-1}}(s_{\tau-1}) - 1\}$. Note that $n^+_{I_{\tau-1}}(s_{\tau-1}) \leq \frac{k}{2}$, otherwise, $n^+_{I_{\tau-1}}(s_{\tau-1}) > \frac{k}{2}$ would imply that there exists at least one player in $N^+_{I_{\tau-1}}(s_{\tau-1})$ with more than $\frac{k}{2}$ neighbours in $N^+_{I_{\tau-2}}(s_{\tau-2})$, which contradicts the fact that in $\mathcal{G}_k$ networks with k even, each player has $\frac{k}{2}$ neighbours to the left and right. Now, player j is directly connected to $j + \frac{k}{2}$, and hence, there are $\frac{k}{2} - (n^+_{I_{\tau-1}}(s_{\tau-1}) - 1)$ players in $N^+_{I_\tau}(s_\tau)$ with $n^+_{I_{\tau-1}}(s_{\tau-1})$ neighbours in $N^+_{I_{\tau-1}}(s_{\tau-1})$. Player $j + \frac{k}{2} + 1$ is directly connected to player $j+1$, and hence, has $n^+_{I_{\tau-1}}(s_{\tau-1}) - 1$ neighbours in $N^+_{I_{\tau-1}}(s_{\tau-1})$; player $i + \frac{k}{2} + 2$ has $n^+_{I_{\tau-1}}(s_{\tau-1}) - 2$ neighbours in $N^+_{I_{\tau-1}}(s_{\tau-1})$; and more generally, for $0 \leq r \leq n^+_{I_{\tau-1}}(s_{\tau-1}) - 1$, player $i + \frac{k}{2} + r$, has $n^+_{I_{\tau-1}}(s_{\tau-1}) - r$ neighbours in $N^+_{I_{\tau-1}}(s_{\tau-1})$. Thus, there are $\frac{k}{2} - (n^+_{I_{\tau-1}}(s_{\tau-1}) - 1)$ players with $n^+_{I_\tau}(s_\tau)$ neighbours in $N^+_{I_{\tau-1}}(s_{\tau-1})$; $\frac{k}{2} - (n^+_{I_{\tau-1}}(s_{\tau-1}) - 2)$ players with at least $n^+_{I_{\tau-1}}(s_{\tau-1}) - 1$ neighbours in $N^+_{I_{\tau-1}}(s_{\tau-1})$; $\frac{k}{2} - (n^+_{I_{\tau-1}}(s_{\tau-1}) - 3)$ players with at least $n^+_{I_{\tau-1}}(s_{\tau-1}) - 2$ neighbours in $N^+_{I_{\tau-1}}(s_{\tau-1})$; and more generally, for $0 \leq r \leq n^+_{I_{\tau-1}}(s_{\tau-1}) - 1$, there are $\frac{k}{2} - (n^+_{I_{\tau-1}}(s_{\tau-1}) - r - 1)$ players in $N^+_{I_\tau}(s_\tau)$ with at least $n^+_{I_{\tau-1}}(s_{\tau-1}) - r$ neighbours in $N^+_{I_{\tau-1}}(s_{\tau-1})$. This implies that

$$N^+_{I_\tau}(n^+_{I_{\tau-1}}(s_{\tau-1}) - r) = \left\{ i + n^+_{I_{\tau-1}}(s_{\tau-1}), i + n^+_{I_{\tau-1}}(s_{\tau-1}) + 1, \cdots, i + \frac{k}{2} + r \right\}$$

Letting $s_\tau = n^+_{I_{\tau-1}}(s_{\tau-1}) - r$ so that $r = n^+_{I_{\tau-1}}(s_{\tau-1}) - s_\tau$, where $0 \leq r \leq n^+_{I_{\tau-1}}(s_{\tau-1}) - 1$ implies that $1 \leq s_\tau \leq n^+_{I_{\tau-1}}(s_{\tau-1})$, then:

$$N^+_{I_\tau}(s_\tau) = \left\{ i + n^+_{I_{\tau-1}}(s_{\tau-1}), \cdots, i + n^+_{I_{\tau-1}}(s_{\tau-1}) + \frac{k}{2} - s_\tau \right\} \quad \text{for } 1 \leq s_\tau \leq n^+_{I_{\tau-1}}(s_{\tau-1})$$

The corresponding cardinalities of $N^+_{I_\tau}(s_\tau)$, $N^-_{I_\tau}(s_\tau)$ and $N_{I_\tau}(s_\tau)$ are then respectively $n^+_{I_\tau}(s_\tau) = n^-_{I_\tau}(s_\tau) = \frac{k}{2} - (s_\tau - 1)$ and $n_{I_\tau}(s_\tau) = k - 2(s_\tau - 1)$, for $1 \leq s_\tau \leq n^+_{I_{\tau-1}}(s_{\tau-1})$. □

Lemma 9. *For $\mathcal{G}_k$ networks with k odd, the following bounds hold: $\lceil k/2 \rceil - s_\tau \leq n^+_{I_\tau}(s_\tau) \leq \lceil k/2 \rceil - (s_\tau - 1)$; $\lceil k/2 \rceil - s_\tau \leq n^-_{I_\tau}(s_\tau) \leq \lceil k/2 \rceil - (s_\tau - 1)$; and $2\lceil k/2 \rceil - 2s_\tau \leq n_{I_\tau}(s_\tau) \leq 2\lceil k/2 \rceil - 2(s_\tau - 1)$, for $1 \leq \tau \leq d_I$, where $1 \leq s_\tau \leq n^+_{I_{\tau-1}}(s_{\tau-1})$.*

Proof. First note that, by definition, the feasible number of players that make up any $G_k(n) \in \mathcal{G}_k$ is in the multiples of $k+1$. That is, for any $G_k(n) \in \mathcal{G}_k$, n takes on the values $n = h(k+1)$, for $h = 1, 2, \cdots$. For example, when $k = 3$, n takes on the values $n = 4h$, whereby, for $h = 1$, $G_3(4)$ is a complete network. Let $\mathcal{C}(k+1) = \{j, j+1, \cdots, j+k\}$ be any group of $k+1$ adjacently placed players in $G_k(n) \in \mathcal{G}_k$ with the following property: $\{j, j+1, \cdots, j+\lfloor k/2 \rfloor\}$ have $\lceil k/2 \rceil$ neighbours to the right and $\lfloor k/2 \rfloor$ neighbours to the left; and all players in $\{j+\lceil k/2 \rceil, j+\lceil k/2 \rceil+1, \cdots, j+k\}$ have $\lfloor k/2 \rfloor$ neighbours to the right and $\lceil k/2 \rceil$ neighbours to the left. Every $G_k(n) \in \mathcal{G}_k$ is made up of $\mathcal{C}(k+1)$ groups of adjacently placed players.

Let $I = \{i, i+1, \cdots, i+\mu-1\}$ be a set of μ adjacently placed players. Assume without loss of generality that all players $\{i, i+1, \cdots, i+\lfloor (\mu-1)/2 \rfloor\}$ have $\lceil k/2 \rceil$ neighbours to the right and $\lfloor k/2 \rfloor$ neighbours to the left; and all players in $\{i+\lceil (\mu-1)/2 \rceil, \cdots, j+\mu-1\}$ have $\lfloor k/2 \rfloor$ neighbours to the right and $\lceil k/2 \rceil$ neighbours to the left.

For $\mathcal{G}_k$ networks with k odd, it is sufficient to consider only scenarios where $\mu \leq \lceil k/2 \rceil$. This is because, when k is odd, $\lceil k/2 \rceil = \lfloor k/2 \rfloor + 1 \geq \lceil pk \rceil + 1 \geq CR^*(\mathbf{A}^T)$, where the second inequality follows because $p \leq \frac{\lfloor k/2 \rfloor}{k}$. Thus, if we find that $\mu \leq \lceil k/2 \rceil$ mutations are not sufficient to trigger an exit from $D(\mathbf{A}^T)$, then we conclude that at least $\lfloor k/2 \rfloor + 2$ mutations are required.

Now, first consider a scenario where $I = \{i\}$ so that $\mu = 1$. Then, for $s_1 = 1$, $n^+_{I_1}(s_1) = \lceil k/2 \rceil = \lceil k/2 \rceil - (s_1 - 1)$, $n^-_{I_1}(s_1) = \lfloor k/2 \rfloor = \lfloor k/2 \rfloor - (s_1 - 1)$ and $n_{I_1}(s_1) = n^+_{I_1}(s_1) + n^-_{I_1}(s_1) = \lceil k/2 \rceil = k - (s_1 - 1)$.

Next, consider a scenario where $2 \leq \mu \leq \lceil k/2 \rceil$. Since i has $\lceil k/2 \rceil$ neighbours to the right (i.e. i is directly connected to $i+\lceil k/2 \rceil$) and that $\mu \leq \lceil k/2 \rceil$, there are $i+\lceil k/2 \rceil-(i+\mu-1) = \lceil k/2 \rceil-(\mu-1)$ players with μ neighbours in I. Consider all players in $\{i+\lceil k/2 \rceil+1, \cdots, i+\mu+\lfloor k/2 \rfloor-1\}$. By definition of I and $\mathcal{C}(k+1)$ groups, players i, $i+\lceil k/2 \rceil$ and $i+\lceil k/2 \rceil+1$ all belong to the same $\mathcal{C}(k+1)$ group, and hence, players $i+\lceil k/2 \rceil$ and $i+\lceil k/2 \rceil+1$ have $\lceil k/2 \rceil$ neighbours to the left and $\lfloor k/2 \rfloor$ neighbours to the right. If $2 \leq \mu \leq 3$, then player $i+\lceil k/2 \rceil+1$ belongs to a different but adjacent $\mathcal{C}(k+1)$ than i so that all players in $\{i+\lceil k/2 \rceil+1, \cdots, i+\mu+\lfloor k/2 \rfloor-1\}$ have $\lfloor k/2 \rfloor$ neighbours to the left and $\lceil k/2 \rceil$ to the right. However, when $\mu \geq 4$, player $i+\lceil k/2 \rceil+1$ belongs to the same $\mathcal{C}(k+1)$ group as i, and hence, has $\lceil k/2 \rceil$ neighbours to the left and $\lfloor k/2 \rfloor$ to the right. This implies that, given I and μ, there exists some $0 \leq r' \leq \mu-1$, whereby, all players in $\{i+\lceil k/2 \rceil+1, \cdots, i+\lceil k/2 \rceil+r'\}$ have $\lceil k/2 \rceil$ neighbours to the left and $\lfloor k/2 \rfloor$ to the right, and all players in $\{i+\lceil k/2 \rceil+r'+1, \cdots, i+\mu+\lfloor k/2 \rfloor-1\}$ have $\lfloor k/2 \rfloor$ neighbours to the left and $\lceil k/2 \rceil$ to the right.

Thus, there are $\lceil k/2 \rceil-(\mu-1)$ players with μ neighbours in I; if $r' > 0$, then $\lceil k/2 \rceil-(\mu-2)$ players have at least $\mu-1$ neighbours in I; if $r' > 1$, then $\lceil k/2 \rceil-(\mu-3)$ players have at least $\mu-2$ neighbours in I; and more generally, for $0 \leq r \leq r'$, $\lceil k/2 \rceil-(\mu-r-1)$ players have at least $\mu-r$ neighbours in I. For $r > r'$, $\lceil k/2 \rceil-(\mu-r-1)$ players have at least $\mu-r-1$ neighbours in I. Using the equality $i+\mu-1+(\lceil k/2 \rceil-(\mu-r-1)) = i+\lceil k/2 \rceil+r$, we then have,

$$N^+_{I_1}(\mu-r) = \{i+\mu, i+\mu+1, \cdots, i+\lceil k/2 \rceil+r\} \quad \text{for } 0 \leq r \leq r' \leq \mu-1 \tag{A.8}$$

$$N^+_{I_1}(\mu - r - 1) = \{i+\mu, i+\mu+1, \cdots, i+\lceil k/2\rceil + r\} \quad \text{for } 0 \leq r' < r \leq \mu - 1 \tag{A.9}$$

Letting $s_1 = \mu - r$ so that, for $r \leq r'$, $r = \mu - s_1$ and $s_1' = \mu - r'$, where $0 \leq r \leq r' \leq \mu - 1$ implies that $1 \leq s_1 \leq s_1' \leq \mu$, we have

$$N^+_{I_1}(s_1) = \{i+\mu, i+\mu+1, \cdots, i+\mu+\lceil k/2\rceil - s_1\} \quad \text{for } 1 \leq s_1 \leq s_1' \leq \mu \tag{A.10}$$

If on the other hand $s_1 = \mu - r - 1$, so that, for $r > r'$, $r = \mu - s_1 - 1$, where $0 \leq r' < r \leq \mu - 1$ implies that $1 \leq s_1' < s_1 \leq \mu$, then

$$N^+_{I_1}(s_1) = \{i+\mu, i+\mu+1, \cdots, i+\mu+\lceil k/2\rceil - s_1 - 1\} \quad \text{for } 1 \leq s_1' < s_1 \leq \mu \tag{A.11}$$

By replicating the same steps above, we obtain the following compositions of $N^-_{I_1}(s_1)$:

$$N^-_{I_1}(s_1) = \{i-1, i-2, \cdots, i-\lceil k/2\rceil + s_1 - 1\} \quad \text{for } 1 \leq s_1 \leq s_1' \leq \mu \tag{A.12}$$

$$N^-_{I_1}(s_1) = \{i-1, i-2, \cdots, i-\lceil k/2\rceil + s_1\} \quad \text{for } 1 \leq s_1' < s_1 \leq \mu \tag{A.13}$$

We see from (A.10), (A.11), (A.12) and (A.13) that when $2 \leq \mu \leq \lceil k/2\rceil$, $n^+_{I_1}(s_1)$ and $n^-_{I_1}(s_1)$ take on any of the values $\lceil k/2\rceil - (s_1-1)$, $\lceil k/2\rceil - s_1$ or $\lfloor k/2\rfloor - (s_1-1)$. But for k odd, we have $\lfloor k/2\rfloor - (s_1-1) = \lfloor k/2\rfloor + 1 - s_1 = \lceil k/2\rceil - s_1$. Thus $n^+_{I_1}(s_1)$ and $n^-_{I_1}(s_1)$ are bounded from below by $\lceil k/2\rceil - s_1$ and from above by $\lceil k/2\rceil - (s_1-1)$; and $2\lceil k/2\rceil - 2s_1 \leq n_{I_1}(s_1) \leq 2\lceil k/2\rceil - 2(s_1-1)$.

These bounds generalize to $n^+_{I_\tau}(s_\tau)$, $n^-_{I_\tau}(s_\tau)$ and $n_{I_\tau}(s_\tau)$, for $2 \leq \tau \leq d_I$. Consider the following labelling of the elements of $N^+_{I_{\tau-1}}(s_{\tau-1})$:

$$N^+_{I_{\tau-1}}(s_{\tau-1}) = \{j, j+1, \cdots, j+n^+_{I_{\tau-1}}(s_{\tau-1}) - 1\}$$

We first consider a scenario where j has $\lceil k/2\rceil$ neighbours to its right. For this scenario, there are two possible arrangements for players in $\left\{j+\lceil k/2\rceil+1, \cdots, j+\lceil k/2\rceil + n^+_{I_{\tau-1}}(s_{\tau-1}) - 1\right\}$. The first arrangement is where there exists some r' so that for $1 \leq r' \leq n^+_{I_{\tau-1}}(s_{\tau-1}) - 1$, each player in $\{j+\lceil k/2\rceil+1, \cdots, j+\lceil k/2\rceil + r'\}$ has $\lceil k/2\rceil$ neighbours to her left, while each player in $\{j+\lceil k/2\rceil+r'+1, \cdots, j+\lceil k/2\rceil+n^+_{I_{\tau-1}}(s_{\tau-1})-1\}$ has $\lfloor k/2\rfloor$ neighbours to her left. The second possible arrangement is where all players in $\left\{j+\lceil k/2\rceil+1, \cdots, j+\lceil k/2\rceil+n^+_{I_{\tau-1}}(s_{\tau-1})-1\right\}$ have $\lfloor k/2\rfloor$ neighbours to their left. It is not possible to have an arrangement where players in $\{j+\lceil k/2\rceil+1, \cdots, j+\lceil k/2\rceil+r'\}$ have $\lfloor k/2\rfloor$ neighbours to their left, while players in $\{j+\lceil k/2\rceil+r'+1, \cdots, j+\lceil k/2\rceil+n^+_{I_{\tau-1}}(s_{\tau-1})-1\}$ have $\lceil k/2\rceil$ neighbours to their left. This is because, by definition of $\mathcal{G}_k$ networks with k odd, if j has $\lceil k/2\rceil$ neighbours to the right and $j+\lceil k/2\rceil+1$ has $\lfloor k/2\rfloor$ neighbours to the left, then j and $j+\lceil k/2\rceil+1$ belong to different but adjacent $\mathcal{C}(k+1)$ groups. It then follows, also by definition of $\mathcal{G}_k$ networks with k odd, that all players in $\{j+\lceil k/2\rceil+1, \cdots, j+\lceil k/2\rceil+\lceil k/2\rceil\}$ all have $\lfloor k/2\rfloor$ neighbours to their left. And since

$n^+_{I_{\tau-1}}(s_{\tau-1}) \leq \lceil k/2 \rceil$, it follows that all players in $\{j+\lceil k/2\rceil+1, \cdots, j+\lceil k/2\rceil+n^+_{I_{\tau-1}}(s_{\tau-1})-1\}$ have $\lfloor k/2 \rfloor$ neighbours to their left.

Now, consider a scenario where j has $\lceil k/2 \rceil$ neighbours to its right, and for $1 \leq r' \leq n^+_{I_{\tau-1}}(s_{\tau-1}) - 1$, each player in $\{j + \lceil k/2\rceil + 1, \cdots, j + \lceil k/2\rceil + r'\}$ has $\lceil k/2 \rceil$ neighbours to her left, while each player in $\{j+\lceil k/2\rceil+r'+1, \cdots, j+\lceil k/2\rceil+n^+_{I_{\tau-1}}(s_{\tau-1})-1\}$ has $\lfloor k/2 \rfloor$ neighbours to her left. For this scenario, a total of $j+\lceil k/2\rceil-(j+n^+_{I_{\tau-1}}(s_{\tau-1})-1) = \lceil k/2\rceil-(n^+_{I_{\tau-1}}(s_{\tau-1})-1)$ players have $n^+_{I_{\tau-1}}(s_{\tau-1})$ neighbours in $N^+_{I_{\tau-1}}(s_{\tau-1})$. Additionally, $\lceil k/2\rceil - (n^+_{I_{\tau-1}}(s_{\tau-1}) - 2)$ players have at least $n^+_{I_{\tau-1}}(s_{\tau-1}) - 1$ neighbours in $N^+_{I_{\tau-1}}(s_{\tau-1})$; $\lceil k/2\rceil - (n^+_{I_{\tau-1}}(s_{\tau-1}) - 3)$ players have at least $n^+_{I_{\tau-1}}(s_{\tau-1}) - 2$ neighbours in $N^+_{I_{\tau-1}}(s_{\tau-1})$; and more generally, for $0 \leq r \leq r'$, $\lceil k/2\rceil - (n^+_{I_{\tau-1}}(s_{\tau-1}) - r - 1)$ players have at least $n^+_{I_{\tau-1}}(s_{\tau-1}) - r$ neighbours in $N^+_{I_{\tau-1}}(s_{\tau-1})$. For $1 \leq r' < r \leq n^+_{I_{\tau-1}}(s_{\tau-1})-1$, $\lceil k/2\rceil-(n^+_{I_{\tau-1}}(s_{\tau-1})-r-1)$ players have at least $n^+_{I_{\tau-1}}(s_{\tau-1})-r-1$ neighbours in $N^+_{I_{\tau-1}}(s_{\tau-1})$. It then follows that

$$N^+_{I_\tau}(n^+_{I_{\tau-1}}(s_{\tau-1}) - r) = \Big\{j + n^+_{I_{\tau-1}}(s_{\tau-1}), \cdots, j + \lceil k/2\rceil + r\Big\} \quad \text{for } 0 \leq r \leq r' \leq n^+_{I_{\tau-1}}(s_{\tau-1}) - 1$$

$$N^+_{I_\tau}(n^+_{I_{\tau-1}}(s_{\tau-1}) - r - 1) = \Big\{j + n^+_{I_{\tau-1}}(s_{\tau-1}), \cdots, j + \lceil k/2\rceil + r\Big\} \quad \text{for } 0 \leq r' < r \leq n^+_{I_{\tau-1}}(s_{\tau-1}) - 1$$

Letting $s_\tau = n^+_{I_{\tau-1}}(s_{\tau-1})-r$ so that, for $r \leq r'$, $r = n^+_{I_{\tau-1}}(s_{\tau-1})-s_\tau$ and $s'_\tau = n^+_{I_{\tau-1}}(s_{\tau-1})-r'$, where $0 \leq r \leq r' \leq n^+_{I_{\tau-1}}(s_{\tau-1}) - 1$ implies that $1 \leq s_\tau \leq s'_\tau \leq n^+_{I_{\tau-1}}(s_{\tau-1})$, then

$$N^+_{I_\tau}(s_\tau) = \Big\{j + n^+_{I_{\tau-1}}(s_{\tau-1}), \cdots, j + n^+_{I_{\tau-1}}(s_{\tau-1}) + \lceil k/2\rceil - s_\tau\Big\} \quad \text{for } 1 \leq s_\tau \leq s'_\tau \leq n^+_{I_{\tau-1}}(s_{\tau-1}) \tag{A.14}$$

If on the other hand $s_\tau = n^+_{I_{\tau-1}}(s_{\tau-1})-r-1$, so that, for $r > r'$, $r = n^+_{I_{\tau-1}}(s_{\tau-1})-s_\tau-1$, where $0 \leq r' < r \leq n^+_{I_{\tau-1}}(s_{\tau-1}) - 1$ implies that $1 \leq s'_\tau < s_\tau \leq n^+_{I_{\tau-1}}(s_{\tau-1})$, then

$$N^+_{I_\tau}(s_\tau) = \Big\{j + n^+_{I_{\tau-1}}(s_{\tau-1}), \cdots, j + n^+_{I_{\tau-1}}(s_{\tau-1}) + \lceil k/2\rceil - s_\tau - 1\Big\} \quad \text{for } 1 \leq s'_\tau < s_\tau \leq n^+_{I_{\tau-1}}(s_{\tau-1}) \tag{A.15}$$

When j has $\lceil k/2 \rceil$ neighbours to its right but all players in $\{j + \lceil k/2\rceil + 1, \cdots, j + \lceil k/2\rceil + n^+_{I_{\tau-1}}(s_{\tau-1}) - 1\}$ have $\lfloor k/2 \rfloor$ neighbours to their left, then $\lceil k/2\rceil - (n^+_{I_{\tau-1}}(s_{\tau-1}) - 1)$ players have $n^+_{I_{\tau-1}}(s_{\tau-1})$ neighbours in $N^+_{I_{\tau-1}}(s_{\tau-1})$, but $\lceil k/2\rceil - (n^+_{I_{\tau-1}}(s_{\tau-1}) - r - 1)$ players have at least $n^+_{I_{\tau-1}}(s_{\tau-1}) - r - 1$ neighbours in $N^+_{I_{\tau-1}}(s_{\tau-1})$. Following the same steps above, we find that

$$N^+_{I_\tau}(s_\tau) = \Big\{j + n^+_{I_{\tau-1}}(s_{\tau-1}), \cdots, j + n^+_{I_{\tau-1}}(s_{\tau-1}) + \lceil k/2\rceil - s_\tau - 1\Big\} \quad \text{for } 1 \leq s_\tau \leq n^+_{I_{\tau-1}}(s_{\tau-1}) \tag{A.16}$$

Next, we consider a scenario where j has $\lfloor k/2 \rfloor$ neighbours to the right. The only feasible arrangement that players in $\Big\{j + \lfloor k/2\rfloor + 1, \cdots, j + \lfloor k/2\rfloor + n^+_{I_{\tau-1}}(s_{\tau-1}) - 1\Big\}$ can take is one where they all have $\lfloor k/2 \rfloor$ neighbours to their left. To see why, let $j \in \mathcal{C}(k + 1)$ and let

$\mathcal{C}(k+1) = \{h, h+1, \cdots, h+k\}$. If j has $\lfloor k/2 \rfloor$ neighbours to the right, then by definition of $\mathcal{C}(k+1)$, $j \in \{h + \lceil k/2 \rceil, \cdots, h+k\}$, which implies that $j + \lfloor k/2 \rfloor \in \{h + \lceil k/2 \rceil, \cdots, h+k\}$ but $j + \lfloor k/2 \rfloor + 1 \notin \mathcal{C}(k+1)$. That is, players in $\left\{j + \lfloor k/2 \rfloor + 1, \cdots, j + \lfloor k/2 \rfloor + n^+_{I_{\tau-1}}(s_{\tau-1}) - 1\right\}$ belong to a different but adjacent $\mathcal{C}'(k+1)$ group than j, and hence, they all have $\lfloor k/2 \rfloor$ neighbours to the left and $\lceil k/2 \rceil$ neighbours to the right.

Since j has $\lfloor k/2 \rfloor$ neighbours to its right, and hence, j is directly connected to $j + \lfloor k/2 \rfloor$, it follows that a total of $j + \lfloor k/2 \rfloor - (j + n^+_{I_{\tau-1}}(s_{\tau-1}) - 1) = \lfloor k/2 \rfloor - (n^+_{I_{\tau-1}}(s_{\tau-1}) - 1)$ players have $n^+_{I_{\tau-1}}(s_{\tau-1})$ neighbours in $N^+_{I_{\tau-1}}(s_{\tau-1})$. Additionally, since player $j + \lfloor k/2 \rfloor + 1$ has $\lfloor k/2 \rfloor$ neighbours to the left, she is directly connected to $j+1$; player $j + \lfloor k/2 \rfloor + 2$ is directly connected to $j+2$; and more generally, for $0 \le r \le n^+_{I_{\tau-1}}(s_{\tau-1}) - 1$, player $j + \lfloor k/2 \rfloor + r$ is directly connected to player $j + r$. This implies that a total of $\lfloor k/2 \rfloor - (n^+_{I_{\tau-1}}(s_{\tau-1}) - 1)$ players have $n^+_{I_{\tau-1}}(s_{\tau-1})$ neighbours in $N^+_{I_{\tau-1}}(s_{\tau-1})$; $\lfloor k/2 \rfloor - (n^+_{I_{\tau-1}}(s_{\tau-1}) - 2)$ players have at least $n^+_{I_{\tau-1}}(s_{\tau-1}) - 1$ neighbours in $N^+_{I_{\tau-1}}(s_{\tau-1})$; $\lfloor k/2 \rfloor - (n^+_{I_{\tau-1}}(s_{\tau-1}) - 3)$ players have at least $n^+_{I_{\tau-1}}(s_{\tau-1}) - 2$ neighbours in $N^+_{I_{\tau-1}}(s_{\tau-1})$; and more generally, for $0 \le r \le n^+_{I_{\tau-1}}(s_{\tau-1}) - 1$, $\lfloor k/2 \rfloor - (n^+_{I_{\tau-1}}(s_{\tau-1}) - r - 1)$ players have at least $n^+_{I_{\tau-1}}(s_{\tau-1}) - r$ neighbours in $N^+_{I_{\tau-1}}(s_{\tau-1})$. Consequently,

$$N^+_{I_\tau}(n^+_{I_{\tau-1}}(s_{\tau-1}) - r) = \left\{j + n^+_{I_{\tau-1}}(s_{\tau-1}), \cdots, j + \lfloor k/2 \rfloor + r\right\} \quad \text{for } 0 \le r \le n^+_{I_{\tau-1}}(s_{\tau-1}) - 1$$

Letting $s_\tau = n^+_{I_{\tau-1}}(s_{\tau-1}) - r$ so that $r = n^+_{I_{\tau-1}}(s_{\tau-1}) - s_\tau$, we have

$$N^+_{I_\tau}(s_\tau) = \left\{j + n^+_{I_{\tau-1}}(s_{\tau-1}), \cdots, j + n^+_{I_{\tau-1}}(s_{\tau-1}) + \lfloor k/2 \rfloor - s_\tau\right\} \quad \text{for } 1 \le s_\tau \le n^+_{I_{\tau-1}}(s_{\tau-1}) \tag{A.17}$$

The derivation for all possible compositions of $N^-_{I_\tau}(s_\tau)$ follows the same steps above. The only difference is that now $j + n^+_{I_{\tau-1}}(s_{\tau-1}) - 1$ takes up the role of j. That is, we consider the two scenarios where $j + n^+_{I_{\tau-1}}(s_{\tau-1}) - 1$ has $\lceil k/2 \rceil$ and $\lfloor k/2 \rfloor$ neighbours to the left. Under these two scenarios, we replicate the three different arrangements of players in $\{j + n^+_{I_{\tau-1}}(s_{\tau-1}) - \lceil k/2 \rceil, \cdots, j - \lceil k/2 \rceil\}$ and respectively $\{j + n^+_{I_{\tau-1}}(s_{\tau-1}) - \lfloor k/2 \rfloor, \cdots, j - \lfloor k/2 \rfloor\}$. We find that the respective compositions of $N^-_{I_{\tau-1}}(s_{\tau-1})$ corresponding to the four scenarios in (A.14), (A.15), (A.17) and (A.17) are respectively:

$$N^-_{I_\tau}(s_\tau) = \{j-1, j-2, \cdots, j - \lceil k/2 \rceil + s_\tau - 1\} \quad \text{for } 1 \le s_\tau \le s'_\tau \le n^+_{I_{\tau-1}}(s_{\tau-1}) \tag{A.18}$$

$$N^-_{I_\tau}(s_\tau) = \{j-1, j-2, \cdots, j - \lceil k/2 \rceil + s_\tau\} \quad \text{for } 1 \le s'_\tau < s_\tau \le n^+_{I_{\tau-1}}(s_{\tau-1}) \tag{A.19}$$

$$N^-_{I_\tau}(s_\tau) = \{j-1, j-2, \cdots, i - \lceil k/2 \rceil + s_\tau - 1\} \quad \text{for } 1 \le s_\tau \le n^+_{I_{\tau-1}}(s_{\tau-1}) \tag{A.20}$$

$$N^-_{I_\tau}(s_\tau) = \{j-1, j-2, \cdots, j - \lfloor k/2 \rfloor + s_\tau - 1\} \quad \text{for } 1 \le s_\tau \le n^+_{I_{\tau-1}}(s_{\tau-1}) \tag{A.21}$$

From (A.14), (A.15), (A.17) and (A.17), we see that $n^+_{I_\tau}(s_\tau)$ takes on any of the values $\lceil k/2\rceil-(s_\tau-1)$, $\lceil k/2\rceil-s_\tau$ and $\lfloor k/2\rfloor-(s_\tau-1)$. Note, however, that, for k odd, $\lfloor k/2\rfloor-(s_\tau-1) = \lfloor k/2\rfloor+1-s_\tau = \lceil k/2\rceil - s_\tau$ (since $\lfloor k/2\rfloor+1 = \lceil k/2\rceil$). Thus, $\lceil k/2\rceil - s_\tau \le n^+_{I_\tau}(s_\tau) \le \lceil k/2\rceil-(s_\tau-1)$. Similarly, from (A.18), (A.19), (A.20) and (A.21), $n^-_{I_\tau}(s_\tau)$ takes on any of the values $\lceil k/2\rceil-(s_\tau-1)$ and $\lceil k/2\rceil-s_\tau$, and hence, $\lceil k/2\rceil-s_\tau \le n^-_{I_\tau}(s_\tau) \le \lceil k/2\rceil-(s_\tau-1)$. Finally, we have $2\lceil k/2\rceil-2s_\tau \le n_{I_\tau}(s_\tau) \le 2\lceil k\rceil - 2(s_\tau-1)$. □

Appendix B. Proof of Lemma 5

The proof of Lemma 5 (i) constitutes of two steps. The first step, which is already proved in Section 4, show that if $p \le \frac{\lfloor k/2\rfloor}{k}$, then starting from a state where any $\lfloor k/2\rfloor+1$ adjacently placed players play strategies in A^τ, strategies in $A^{\tau+1}$ spread contagiously to the whole network, and hence, $(U, N, \mathcal{G}_k, P)$ will converge to an absorbing set of states containing only strategies in $A^{\tau+1}$. The second step, which we prove here, shows that when $p \le \frac{\lfloor k/2\rfloor}{k}$, $(U, N, \mathcal{G}_k, P)$ does not contain absorbing states where strategies in $\bar{A}^\tau$ and $A^{\tau+1}$ coexist within a subgroup of $\lfloor k/2\rfloor+1$ adjacently placed players. These two statements then together imply that when $p \le \frac{\lfloor k/2\rfloor}{k}$, strategies in $\bar{A}^\tau$ cannot co-exists with strategies in $A^{\tau+1}$ in an absorbing state.

We prove the second statement above by contradiction. That is, suppose there exists an absorbing state, $\mathbf{x} \in \mathbf{A}$, where strategies in $\bar{A}^\tau$ and $A^{\tau+1}$ coexist. Pick any $i \in N$ for whom $x^i \in A^{\tau+1}$, where x^i is the ith coordinate of $\mathbf{x}$. Let $N_i(A^{\tau+1};\mathbf{x})$ be the set of i's neighbours playing strategies in $A^{\tau+1}$ in state $\mathbf{x}$, and $n_i(A^{\tau+1};\mathbf{x})$ the respective cardinality. By definition of absorbing states, x^i is a best response to $\mathbf{x}_{-i}$ (state $\mathbf{x}$ with i excluded). We show that, under this set up, there exists at least one I, with $i \in I$, whereby, some $l \in I$, $l \ne i$, with $x^l \in \bar{A}^\tau$, has $n_l(A^{\tau+1};\mathbf{x}) \ge n_i(A^{\tau+1};\mathbf{x})$. This implies that $\frac{n_l(A^{\tau+1};\mathbf{x})}{k} \ge \frac{n_i(A^{\tau+1};\mathbf{x})}{k}$, and hence, some $a \in A^{\tau+1}$ is a best response to l. This in turn implies that $\mathbf{x}$ is transient or a state in an absorbing cycle, a contradiction.

Now, given $i \in N$ with $x^i \in A^{\tau+1}$, consider the following labelling of the closed neighbourhood, B_i, of i (i.e. the direct neighbours of i, N_i, with i included): $B_i = \{i-\lfloor k/2\rfloor, \cdots, i-1, i, i+1, \cdots, i+\lceil k/2\rceil\}$, where for k even, $\lfloor k/2\rfloor = \lceil k/2\rceil = \frac{k}{2}$. Let $B^-_i = \{i-\lfloor k/2\rfloor, \cdots, i-1, i\}$ and $B^+_i = \{i, i+1, \cdots, i+\lceil k/2\rceil\}$. When k is even, the cardinalities of B^-_i and B^+_i are $|B^-_i| = |B^+| = \frac{k}{2}+1$, and when k is odd, $|B^-_i| = \lfloor k/2\rfloor+1$ and $|B^+| = \lceil k/2\rceil+1$. We show that for all possible distributions of the elements of $N_i(A^{\tau+1};\mathbf{x})$ over B_i, there exists at least one $l \in B^-_i$ or $l \in B^+_i$ with $x^l \in \bar{A}^\tau$, for whom $n_l(A^{\tau+1};\mathbf{x}) \ge n_i(A^{\tau+1};\mathbf{x})$. For the remaining steps of this proof, we minimize notational clutter by writing β_i and β_l for $n_i(A^{\tau+1};\mathbf{x})$ and $n_l(A^{\tau+1};\mathbf{x})$ respectively.

First, consider a scenario where $N_i(A^{\tau+1};\mathbf{x}) \subset B_i^-$. For this scenario $\beta_i = n_i(A^{\tau+1};\mathbf{x}) < \lfloor k/2 \rfloor$, otherwise, if $n_i(A^{\tau+1};\mathbf{x}) = \lfloor k/2 \rfloor$, then $|B_i^-| = \lfloor k/2 \rfloor + 1$, and hence, the evolutionary process $(U, N, \mathcal{G}_k, P)$ will converge to an absorbing set containing only strategies in $A^{\tau+1}$. Since $\beta_i < \lfloor k/2 \rfloor$, there exists at least on $l \in B_i^-$, who is directly connected to all players in B_i^-, and hence, has $\beta_l \geq \beta_i$. The same argument applies to a scenario where $N_i(A^{\tau+1};\mathbf{x}) \subset B_i^+$.

Second, consider a scenario where $N_i(A^{\tau+1};)$ is composed of the β_i closest neighbours of i:

$$N_i(A^{\tau+1};\mathbf{x}) = \{i - \lfloor \beta_i/2 \rfloor, \cdots, i-1, i+1, \cdots, i + \lceil \beta_i/2 \rceil\} \tag{B.1}$$

If $N_i(A^{\tau+1};\mathbf{x})$ players are distributed as in (B.1), then each

$$l \in \{i + \lceil \beta_i/2 \rceil - \lfloor k/2 \rfloor - 1, i + \lceil \beta_i/2 \rceil - \lfloor k/2 \rfloor, \cdots, i - \lfloor \beta_i/2 \rfloor - 1, i + \lceil \beta_i/2 \rceil + 1, \cdots, i - \lfloor \beta_i/2 \rfloor + \lfloor k/2 \rfloor\} \tag{B.2}$$

is directly connected to i and $\beta_i - 1$ players in $N_i(A^{\tau+1};\mathbf{x})$, and hence, $\beta_l \geq \beta_i$. Note that each l described in (B.2) belongs to either B_i^- or B_i^+; and that at least one l in set (B.2) must have $x^l \in \bar{A}^\tau$, otherwise, $N_i(A^{\tau+1};\mathbf{x})$ would consist of $\beta_i \geq \lfloor k/2 \rfloor + 1$ players.

Third, let $N_i(A^{\tau+1};\mathbf{x})$ consist of peripheral neighbours of i: that is,

$$N_i(A^{\tau+1};\mathbf{x}) = \Big\{i - \lfloor k/2 \rfloor, \cdots, i - \lfloor k/2 \rfloor + \lfloor \beta_i/2 \rfloor - 1, i + \lceil \beta_i/2 \rceil, \cdots, i + \lceil k/2 \rceil\Big\} \tag{B.3}$$

When $N_i(A^{\tau+1};\mathbf{x})$ is distributed as in (B.3), then at least one $l \in \{i-1, i+1\}$ is directly connected to i and $\beta_i - 1$ players in $N_i(A^{\tau+1};\mathbf{x})$, and hence, $\beta_l \geq \beta_i$. Note that by definition of $N_i(A^{\tau+1};\mathbf{x})$, $x^l \in \bar{A}^\tau$.

Finally, let $N_i(A^{\tau+1};\mathbf{x})$ assume the following intermediate forms:

$$N_i(A^{\tau+1};\mathbf{x}) = \Big\{i - \lfloor k/2 \rfloor, \cdots, i - \lfloor k/2 \rfloor + \lfloor \beta_i/2 \rfloor - 1, i+1, \cdots, i + \lceil \beta_i/2 \rceil\Big\} \tag{B.4}$$

$$N_i(A^{\tau+1};\mathbf{x}) = \Big\{i - \lfloor \beta_i/2 \rfloor, \cdots, i-1, i + \lceil \beta_i/2 \rceil, \cdots, i + \lceil k/2 \rceil\Big\} \tag{B.5}$$

$$\begin{aligned} N_i(A^{\tau+1};\mathbf{x}) = \Big\{&i - \lfloor k/2 \rfloor + \lfloor \beta_i/2 \rfloor, i - \lfloor k/2 \rfloor + \lfloor \beta_i/2 \rfloor + 1, \cdots, i - \lfloor k/2 \rfloor + 2\lfloor \beta_i/2 \rfloor - 1, \\ &i + \lceil \beta_i/4 \rceil, \cdots, i + 2\lceil \beta_i/4 \rceil + 1\Big\} \end{aligned} \tag{B.6}$$

For all the three distributions of $N_i(A^{\tau+1};\mathbf{x})$ in (B.4), (B.5) and (B.6) above, at least one $l \in \{i-1, i+1\}$ is directly connected to i and $\beta_i - 1$ players in $N_i(A^{\tau+1};\mathbf{x})$, and hence, $\beta_l \geq \beta_i$; and by definition of $N_i(A^{\tau+1};\mathbf{x})$, $x^l \in \bar{A}^\tau$.

Thus, for all possible distributions of $N_i(A^{\tau+1};\mathbf{x})$ over B_i, there exists at least one group of $\lceil k/2 \rceil + 1$ adjacently placed players, I, with $i \in I$, whereby, some $l \in I$, $l \neq i$, with $x^l \in \bar{A}^\tau$, has $n_l(A^{\tau+1};\mathbf{x}) \geq n_i(A^{\tau+1};\mathbf{x})$.

To prove Lemma 5 (ii), it is sufficient to provide an example of an absorbing cycle where strategies in $\bar{A}^{\tau}$ coexist with strategies in $A^{\tau+1}$. Consider some $G_2(n) \in \mathcal{G}_2$ with $N = \{1, 2 \cdots, n\}$. Assume that n is odd and let $\mathbf{x}$ be a strategy configuration where players $1, 3, 5, \cdots, n$ play some $a_j \in \bar{A}^{\tau}$ and players $2, 4, \cdots, n-1$ play some $a_l \in A^{\tau+1}$. Let $\mathbf{y}$ be a configuration where players $1, 3, 5, \cdots, n$ play $a_l \in A^{\tau+1}$ and players $2, 4, \cdots, n-1$ play $a_j \in \bar{A}^{\tau}$. Then, the probability of a direct transition from $\mathbf{x}$ to $\mathbf{y}$, and vice versa, is $P(\mathbf{x}, \mathbf{y}) = P(\mathbf{y}, \mathbf{x}) = 1$, which implies that $\mathbf{x}$ and $\mathbf{y}$ form an absorbing cycle. Thus, we cannot rule out the existence of absorbing cycles where strategies in $\bar{A}^{\tau}$ coexist with strategies in $A^{\tau+1}$, even when $p \leq \frac{\lfloor k/2 \rfloor}{k}$.

Appendix C. Proof of Lemma 6

For $T \geq 2$, we aim to derive the condition (i.e. the minimum value of n) under which at least $\gamma(n(A^T) + 1)$ mutations, for $\gamma \geq 1$, are necessary to trigger an exit from $D(\mathbf{A}^T)$. We do so by examining the evolution of the process $(U, N, \mathcal{G}_k, P)$ out of $D(\mathbf{A}^T)$, after $\mu \geq n(A^T)$ mutations to strategies in $\bar{A}^{\tau}$ by a group $I = \{i, i+1, \cdots, i+\mu-1\}$ of adjacently placed players.

Following the definitions in Section Appendix A.2, $N_{I_\tau}(s_\tau)$, for $0 \leq \tau \leq d_I$, is the set of players in $N \backslash \{I \cup N_{I_1}(s_1) \cup \cdots \cup N_{I_{\tau-1}}(s_{\tau-1})\}$ with at least s_τ neighbours in $N_{I_{\tau-1}}(s_{\tau-1})$, where $N_{I_0}(s_0) = I$, and d_I is the value of τ at which $N_{I_\tau}(s_\tau) \neq \emptyset$ but $N_{I_{\tau+r}}(s_{\tau+r}) = \emptyset$ for all $r \geq 1$.

Now, let process $(U, N, \mathcal{G}_k, P)$ start from some $\mathbf{x} \in \mathbf{A}^T$. At $t = 1$, let μ adjacently placed players in I mutate to strategies in $\bar{A}^{\tau}$. We consider μ with lower and upper bounds of $n^*_{T(\tau+1)}(\sigma^q_{T\tau}) \leq \mu \leq \lceil pk \rceil + 1$.

The lower bound for μ is because, when $\mu < n^*_{T(\tau+1)}(\sigma^q_{T\tau})$, all players revert to strategies in A^T at $t = 2$. Firstly, when $\mu < n^*_{T(\tau+1)}(\sigma^q_{T\tau})$, each $j \in I \cup N_{I_1}(s)$ has less than $n^*_{T(\tau+1)}(\sigma^q_{T\tau})$ neighbours play strategies in $\bar{A}^{\tau}$ at $t = 1$. By definition of $n^*_{T(\tau+1)}(\sigma^q_{T\tau})$, strategies in $A^{\tau+1} \backslash A^T$ are not best responses to all $j \in I \cup N_{I_1}(s)$. Secondly, when $p \leq \frac{\lfloor k/2 \rfloor}{k}$, we have $n^*_{T(\tau+1)}(\sigma^q_{T\tau}) \leq \lceil pk \rceil < \lceil (1-p)k \rceil$. Thus, when $\mu < n^*_{T(\tau+1)}(\sigma^q_{T\tau})$, each $j \in I \cup N_{I_1}(s)$ has at most $\mu < \lceil (1-p)k \rceil$ neighbours in I playing strategies in $\bar{A}^{\tau}$ at $t = 1$ and the rest play strategies $A^T \subseteq A^{\tau+1}$. This in turn implies that strategies in $\bar{A}^{\tau}$ are not best responses to all $j \in I \cup N_{I_1}(s)$ because they are best responses only when played by at least $\lceil (1-p)k \rceil$ neighbours and the rest play strategies $A^{\tau+1}$. Hence, when $\mu < n^*_{T(\tau+1)}(\sigma^q_{T\tau})$, the process $(U, N, \mathcal{G}_k, P)$ reverts to some $\mathbf{x} \in \mathbf{A}^T$.

Under these conditions, the evolutionary process $(U, N, \mathcal{G}_k, P)$ evolves from $t = 2$ onward as follows, where we write $n' \to A'$ to mean that n' players play strategies in A':

$$t = 1 \;\left|\; \begin{array}{l} \mu \to \bar{A}^{\tau}; \\ n - \mu \to A^T. \end{array}\right.$$

$t=2$ $\mu \to A^{\tau+1}$. Specifically, when $n^*_{T(\tau+1)}(\sigma^q_{T\tau}) < \mu \leq \lceil pk \rceil + 1$, we have $\mu \to A^{\tau+1}$. This is because each $j \in I$ has at least $n^*_{T(\tau+1)}(\sigma^q_{T\tau})$ neighbours play strategies in $\bar{A}^\tau$ and the rest play strategies in A^T, which, by definition of $n^*_{T(\tau+1)}(\sigma^q_{T\tau})$, makes strategies in $A^{\tau+1}$ best responses. Strategies in $\bar{A}^\tau$ are not best responses to all $j \in I$ even when $\mu = \lceil pk \rceil + 1$ because each would have $\mu - 1 = \lceil pk \rceil < \lceil (1-p)k \rceil$ neighbours within I play strategies in $\bar{A}^\tau$ at $t=1$; but strategies in $\bar{A}^\tau$ are best responses only when played by at least $\lceil (1-p)k \rceil$ neighbours and the rest play strategies in $A^{\tau+1} \supseteq A^T$. When $\mu = n^*_{T(\tau+1)}(\sigma^q_{T\tau})$, we have $\mu \to A^T$. This is because, at $t=1$, each $j \in I$ would have $n^*_{T(\tau+1)}(\sigma^q_{T\tau}) - 1 < n^*_{T(\tau+1)}(\sigma^q_{T\tau})$ neighbours within I play strategies in $\bar{A}^\tau$ and the rest play strategies in A^T. By definition of $n^*_{T(\tau+1)}(\sigma^q_{T\tau})$, strategies in A^T are best responses to all $j \in I$. But since $A^T \subseteq A^{\tau+1}$, we simply write $\mu \to A^{\tau+1}$.

$n_{I_1}(s_1) \to A^{\tau+1}$, for $s_1 = n^*_{T(\tau+1)}(\sigma^q_{T\tau})$, where, from Lemmas 8 and 9, $s_1 \leq \mu$. This is because, for $s_1 = n^*_{T(\tau+1)}(\sigma^q_{T\tau})$, each player in $N_{I_1}(s_1)$ has at least $n^*_{T(\tau+1)}(\sigma^q_{T\tau})$ neighbours play strategies in $\bar{A}^\tau$ and the rest play strategies in A^T, which, by definition of $n^*_{T(\tau+1)}(\sigma^q_{T\tau})$, makes strategies in $A^{\tau+1}$ best responses. Note that each $j \in N_{I_1}(s_1)$ has at least $\lfloor k/2 \rfloor$ neighbours in $N \backslash I$. This implies that even when $p = \frac{\lfloor k/2 \rfloor}{k}$ so that $\mu = \lfloor pk \rfloor + 1 = \lfloor k/2 \rfloor + 1$, each $j \in N_{I_1}(s_1)$ has at most $\lceil k/2 \rceil$ neighbours in I (this follows by definition of $\mathcal{G}_k$ networks) playing strategies in $\bar{A}^\tau$ while $\lfloor k/2 \rfloor$ neighbours in $N \backslash I$ play strategies in $A^T \subseteq A^{\tau+1}$. Thus, by definition of p-best response sets, strategies in $A^{\tau+1}$ are best responses to any $j \in N_{I_1}(s_1)$.

$n - (\mu + n_{I_1}(s_1)) \to A^T$.

$t=3$ $\mu \to A^{\tau+1}$ because each player in I has all her neighbours play strategies in $A^{\tau+1}$ (considering that $A^T \subseteq A^{\tau+1}$) at $t=2$.

$n_{I_1}(s_1) \to A^{\tau+1}$ because each player in $N_{I_1}(s_1)$ has all her neighbours play strategies in $A^T \subseteq A^{\tau+1}$ at $t=2$. Note that some players in $N_{I_1}(s_1)$ have at least $\lfloor k/2 \rfloor$ neighbours in $N \backslash \{I \cup N_{I_1}(s_1)\}$ play strategies in A^T, and hence, have strategies in $A^{\tau+2}$ as best responses.

$n_{I_2}(s_2) \to A^{\tau+2}$, for $s_2 = n^*_{T(\tau+2)}(\sigma^q_{T(\tau+1)})$, where from Lemmas 8 and 9, $s_2 \leq n^+_{I_1}(s_1)$ This is because, at $t=2$, each $j \in N_{I_2}(s_2)$ has at least $n^*_{T(\tau+2)}(\sigma^q_{T(\tau+1)})$ (and at most $\lceil k/2 \rceil$) neighbours in $I \cup N_{I_1}(s_1)$ play strategies in $\bar{A}^{\tau+1} \subseteq A^{\tau+1}$ and the rest play strategies in A^T. Thus, by definition of $n^*_{T(\tau+2)}(\sigma^q_{T(\tau+1)})$ and p-best response sets, strategies in $A^{\tau+2}$ are best responses to all $j \in N_{I_2}(s_2)$.

$n - (\mu + n_{I_1}(s_1) + n_{I_2}(s_2)) \to A^T$.

The evolution of $(U, N, \mathcal{G}_k, P)$ from $t = 4$ onward depends on the patterns of overlap between the sets $N_{I_\tau}(s_\tau)$ and $N_{I_{\tau+1}}(s_{\tau+1})$ for $1 \leq \tau \leq d_I - 1$. From the definition of $N_{I_\tau}(s_\tau)$, we see that although any pair $N_{I_{\tau-1}}(s_{\tau-1})$ and $N_{I_\tau}(s_\tau)$, for $1 \leq \tau \leq d_I$, are disjoint sets, it is possible that some players in $N_{I_\tau}(s_\tau)$ have neighbours in both $N_{I_{\tau-1}}(s_{\tau-1})$ and $N_{I_{\tau-2}}(s_{\tau-2})$. This type of overlaps determine best responses to players in every $N_{I_\tau}(s_\tau)$ from $t = 4$ onward.

We use the following notations and definitions to account for these overlaps. When I, $s_1, \cdots, s_\tau, s_{\tau+1}$ and $N_{I_1}(s_1), \cdots, N_{I_\tau}(s_\tau), N_{I_{\tau+1}}(s_{\tau+1})$ are clearly defined, we write $N_{I_3}[N_{I_1}]$ for the set of players in $N_{I_3}(s_1)$ with at least one neighbour in $N_{I_1}(s_1)$; $N_{I_4}[N_{I_3}[N_{I_1}]]$ is the set of players in $N_{I_4}(s_4)$ with at least one neighbour in $N_{I_3}[N_{I_3}]$; and more generally, $N_{I_\tau}[N_{I_{\tau-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$ is the set of players in $N_{I_\tau}(s_\tau)$ with at least one neighbour in $N_{I_{\tau-1}}[N_{I_{\tau-2}}[\cdots[N_{I_3}[N_{I_1}]]]]$. The respective cardinality is $n_{I_\tau}[N_{I_{\tau-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$. The analogous definitions to $N^+_{I_\tau}(s_\tau)$ and $N^-_{I_\tau}(s_\tau)$ are $N^+_{I_\tau}[N_{I_{\tau-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$ and $N^-_{I_\tau}[N_{I_{\tau-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$, with the corresponding cardinalities $n^+_{I_\tau}[N_{I_{\tau-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$ and $n^-_{I_\tau}[N_{I_{\tau-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$ respectively.

Given the configuration of strategies of $(U, N, \mathcal{G}_k, P)$ at $t = 3$, the configurations at $t = 4$ and $t = 5$ are as follows:

$t = 4$ | $\mu \to A^{\tau+1}$.

$n_{I_1}(s_1) \to A^{\tau+1}$. All $j \in n_{I_1}(s_1)$ with at least $\lceil pk \rceil$ neighbours in $N_{I_2}(s_2)$ play strategies in $A^{\tau+2}$; we return to this dynamics below, but for now, since $A^{\tau+2} \subseteq A^{\tau+1}$, we simply write $n_{I_1}(s_1) \to A^{\tau+1}$.

$n_{I_2}(s_2) \to A^{\tau+2}$ because, at $t = 3$, each $j \in N_{I_2}(s_2)$ has at least $\lfloor k/2 \rfloor$ neighbours in $N \setminus \{I \cup N_{I_1}(s_1)\}$ play strategies in $A^{\tau+2}$ (considering that $A^T \subseteq A^{\tau+2}$) and at most $\lceil k/2 \rceil$ neighbours in $I \cup N_{I_1}(s_1)$ play strategies in $\bar{A}^{\tau+1} \subset A^{\tau+1}$, which, by definition of p-best response sets, makes strategies in $A^{\tau+2}$ best responses to all $j \in N_{I_2}(s_2)$.

For $s_3 = n^*_{T(\tau+3)}(\sigma^q_{T(\tau+2)})$, where from Lemmas 8 and 9, $s_3 \le n^+_{I_2}(s_2)$, we have:

$n_{I_3}[N_{I_1}] \to A^{\tau+2}$. This is because at $t = 3$, each $j \in N^+_{I_3}[N_{I_1}]$ has at least $\lfloor k/2 \rfloor$ neighbours in $N \setminus \{I \cup N_{I_1}(s_1) \cup N_{I_2}(s_2)\}$ play strategies in A^T, $n^+_{I_2}(s_2)$ neighbours in $N^+_{I_2}(s_2)$ play strategies in $A^{\tau+2}$, and the rest (i.e. neighbours of j that belong to $N^+_{I_1}(s_1) \cup I$) play strategies in $A^{\tau+1}$. That is, at least $\lfloor k/2 \rfloor$ neighbours of j play strategies in $A^{\tau+2}$ (considering that $A^T \subseteq A^{\tau+2}$) and at most $\lceil k/2 \rceil$ play strategies in $\bar{A}^{\tau+1} \subset A^{\tau+1}$; by definition of p-best response sets, strategies in $A^{\tau+2}$ are best responses to all $j \in N^+_{I_3}[N_{I_1}]$. Similarly, each $j \in N^-_{I_3}[N_{I_1}]$ has at least $\lfloor k/2 \rfloor$ neighbours in $N \setminus \{I \cup N_{I_1}(s_1) \cup N_{I_2}(s_2)\}$ play strategies in A^T, $n^-_{I_2}(s_2)$ neighbours in $N^-_{I_2}(s_2)$ play strategies in $A^{\tau+2}$, and the rest (i.e. neighbours of j that belong to $N^-_{I_1}(s_1) \cup I$) play strategies in $A^{\tau+1}$, and hence, strategies in $A^{\tau+2}$ are best responses to all $j \in N^-_{I_3}[N_{I_1}]$. Thus, $n_{I_3}[N_{I_1}] = (n^+_{I_3}[N_{I_1}] + n^-_{I_3}[N_{I_1}]) \to A^{\tau+2}$.

$n_{I_3}(s_3) - n_{I_3}[N_{I_1}] \to A^{\tau+3}$. This is because at $t = 3$, each $j \in N_{I_3}(s_3) \setminus N_{I_3}[N_{I_1}]$ has at least $\lfloor k/2 \rfloor$ neighbours in $N \setminus \{I \cup N_{I_1}(s_1) \cup N_{I_2}(s_2)\}$ play strategies in A^T and the rest (i.e. neighbours of j that belong to $N_{I_2}(s_2)$) play strategies in $A^{\tau+2}$. By definition of $n^*_{T(\tau+3)}(\sigma^q_{T(\tau+2)})$, strategies in $A^{\tau+3}$ are best responses to all $j \in N_{I_3}(s_3) \setminus N_{I_3}[N_{I_1}]$.

$n - (\mu + n_{I_1}(s_1) + n_{I_2}(s_2) + n_{I_3}(s_3)) \to A^T$.

$t = 5$ | $\mu \to A^{\tau+1}$.

$n_{I_1}(s_1) \to A^{\tau+1}$.

$n_{I_2}(s_2) \to A^{\tau+2}$.

$n_{I_3}[N_{I_1}] \to A^{\tau+2}$.

$n_{I_3}(s_3) - n_{I_3}[N_{I_1}] \to A^{\tau+3}$.

For $s_4 = n^*_{T(\tau+4)}(\sigma^q_{T(\tau+3)})$, we have:

$n_{I_4}[N_{I_3}[N_{I_1}]] \to A^{\tau+3}$. This is because at $t = 4$, each $j \in N_{I_4}[N_{I_3}[N_{I_1}]]$ has at least $\lfloor k/2 \rfloor$ neighbours in $N\backslash \{I \cup N_{I_1}(s_1) \cup N_{I_2}(s_2) \cup N_{I_3}(s_3)\}$ play strategies in A^T, $n_{I_3}(s_3) - n_{I_3}[N_{I_1}]$ neighbours in $N_{I_3}(s_3)\backslash N_{I_3}[N_{I_1}]$ play strategies in $A^{\tau+3}$, and the rest (i.e. neighbours of j that belong to $N_{I_3}[N_{I_1}] \cup N_{I_2}(s_2)$) play strategies in $A^{\tau+2}$. Thus, j has at least $\lfloor k/2 \rfloor$ neighbours play strategies in $A^{\tau+3}$ and at most $\lceil k/2 \rceil$ play strategies in $\bar{A}^{\tau+2} \subset A^{\tau+2}$, which makes strategies in $A^{\tau+3}$ best responses to j.

The best responses to players in $N_{I_4}(s_4)\backslash N_{I_4}[N_{I_3}[N_{I_1}]]$ depends on whether $n^*_{T(\tau+4)}(\sigma^q_{T(\tau+3)}) \leq n^+_{I_3}(s_3) - n^+_{I_3}[N_{I_1}]$ or $n^*_{T(\tau+4)}(\sigma^q_{T(\tau+3)}) > n^+_{I_3}(s_3) - n^+_{I_3}[N_{I_1}]$. Here, without loss of generality, we follow neighbourhood overlaps for players on the right hand side of I; we could alternatively follow neighbourhood overlaps on the left hand side of I and consider scenarios where $n^*_{T(\tau+4)}(\sigma^q_{T(\tau+3)}) \leq n^-_{I_3}(s_3) - n^-_{I_3}[N_{I_1}]$ or $n^*_{T(\tau+4)}(\sigma^q_{T(\tau+3)}) > n^-_{I_3}(s_3) - n^-_{I_3}[N_{I_1}]$:

If $n^*_{T(\tau+4)}(\sigma^q_{T(\tau+3)}) \leq n^+_{I_3}(s_3) - n^+_{I_3}[N_{I_1}]$, then $n_{I_4}(s_4) - n_{I_4}[N_{I_3}[N_{I_1}]] \to A^{\tau+4}$. This is because, at $t = 4$, each $j \in N_{I_4}(s_4)\backslash N_{I_4}[N_{I_3}[N_{I_1}]]$ would have at least $n^*_{T(\tau+4)}(\sigma^q_{T(\tau+3)})$ neighbours in $N_{I_3}(s_3)\backslash N_{I_3}[N_{I_1}]$ playing strategies in $A^{\tau+3}$ and the rest play strategies in A^T. This, by definition of $n^*_{T(\tau+4)}(\sigma^q_{T(\tau+3)})$, makes strategies in $A^{\tau+4}$ best responses.

If $n^*_{T(\tau+4)}(\sigma^q_{T(\tau+3)}) > n_{I_3}(s_3) - n^+_{I_3}[N_{I_1}]$, then $n_{I_4}(s_4) - n_{I_4}[N_{I_3}[N_{I_1}]] \to A^T$. This is because, at $t = 4$, each $j \in N_{I_4}(s_4)\backslash N_{I_4}[N_{I_3}[N_{I_1}]]$ would have less than $n^*_{T(\tau+4)}(\sigma^q_{T(\tau+3)})$ neighbours in $N_{I_3}(s_3)\backslash N_{I_3}[N_{I_1}]$ playing strategies in $A^{\tau+3}$ and the rest play strategies in A^T, which makes strategies in A^T best responses.

$n - (\mu + n_{I_1}(s_1) + n_{I_2}(s_2) + n_{I_3}(s_3) + n_{I_4}(s_4)) \to A^T$.

The strategy configurations from $t = 6$ onward resemble the strategy configuration at $t = 5$. That is, they depend on whether $n^*_{T(\tau+r+1)}(\sigma^q_{T(\tau+r)}) \leq n^+_{I_r}(s_r) - n^+_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$ or $n^*_{T(\tau+r+1)}(\sigma^q_{T(\tau+r)}) > n^+_{I_r}(s_r) - n^+_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$. We consider two extremes, one where $n^*_{T(\tau+r+1)}(\sigma^q_{T(\tau+r)}) \leq n^+_{I_r}(s_r) - n^+_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$ for all $r = 3, 4, \cdots, T-1$, and the other where $n^*_{T(\tau+r+1)}(\sigma^q_{T(\tau+r)}) > n^+_{I_r}(s_r) - n^+_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]]$ for all $r = 3, 4, \cdots, T-1$. As it becomes clear below, these two extremes determine the minimum value of n above which $\mu + 1$

strategies are not sufficient to trigger an exit from $D(\mathbf{A}^T)$; all other cases lie between these two extremes.

The case of $\boldsymbol{n^*_{T(\tau+r+1)}(\sigma^q_{T(\tau+r)}) \leq n^+_{I_r}(s_r) - n^+_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]}$

For this scenario, the strategy configurations evolve from $t = 6$ onward as follows:

$t = 6$

$\mu \to A^{\tau+1}$.

$n_{I_1}(s_1) \to A^{\tau+1}$.

$n_{I_2}(s_2) \to A^{\tau+2}$.

$n_{I_3}[N_{I_1}] \to A^{\tau+2}$.

$n_{I_3}(s_3) - n_{I_3}[N_{I_1}] \to A^{\tau+3}$.

$n_{I_4}[N_{I_3}[N_{I_1}]] \to A^{\tau+3}$.

$n_{I_4}(s_4) - n_{I_4}[N_{I_3}[N_{I_1}]] \to A^{\tau+4}$.

$n_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]] \to A^{\tau+4}$, for $s_5 = n^*_{T(\tau+5)}(\sigma^q_{T(\tau+4)})$. This is because at $t = 5$, each $j \in N_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]]$ has at least $\lfloor k/2 \rfloor$ neighbours in $N \backslash \{I \cup N_{I_1}(s_1) \cup N_{I_2}(s_2) \cup N_{I_3}(s_3) \cup N_{I_4}(s_4)\}$ playing strategies in A^T, $n_{I_4}(s_4) - n_{I_4}[N_{I_3}[N_{I_1}]] \geq 0$ neighbours in $N_{I_4}(s_4) \backslash N_{I_4}[N_{I_3}[N_{I_1}]]$ play strategies in $A^{\tau+4}$, and the rest (i.e. neighbours of j that belong to $N_{I_4}[N_{I_3}[N_{I_1}]] \cup N_{I_3}(s_3) \backslash N_{I_3}[N_{I_1}]$) playing strategies in $A^{\tau+3}$. Thus, j has more than $\lfloor k/2 \rfloor$ neighbours play strategies in $A^{\tau+4}$ (where $A^T \subseteq A^{\tau+4}$) and at most $\lceil k/2 \rceil$ play strategies in $\bar{A}^{\tau+3} \subset A^{\tau+3}$, which makes strategies in $A^{\tau+4}$ best responses to j.

$n_{I_5}(s_5) - n_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]] \to A^{\tau+5}$. This is because, at $t = 5$, each $j \in N_{I_5}(s_5) \backslash N_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]]$ has at least $n^*_{T(\tau+5)}(\sigma^q_{T(\tau+4)})$ neighbours in $N_{I_4}(s_4) \backslash N_{I_4}[N_{I_3}[N_{I_1}]]$ playing strategies in $A^{\tau+4}$ and the rest play strategies in A^T. This, by definition of $n^*_{T(\tau+5)}(\sigma^q_{T(\tau+4)})$, makes strategies in $A^{\tau+5}$ best responses.

$n - (\mu + n_{I_1}(s_1) + n_{I_2}(s_2) + n_{I_3}(s_3) + n_{I_4}(s_4) + n_{I_5}(s_5)) \to A^T$.

– –   –

$t = T-\tau$	$\mu + n_{I_1}(s_1) \to A^{\tau+1}$.
	$n_{I_2}(s_2) \to A^{\tau+2}$.
	$n_{I_3}[N_{I_1}] \to A^{\tau+2}$.
	$n_{I_3}(s_3) - n_{I_3}[N_{I_1}] \to A^{\tau+3}$
	$n_{I_4}[N_{I_3}[N_{I_1}]] \to A^{\tau+3}$
	$n_{I_4}(s_4) - n_{I_4}[N_{I_3}[N_{I_1}]] \to A^{\tau+4}$.
	$n_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]] \to A^{\tau+4}$.
	$n_{I_5}(s_5) - n_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]] \to A^{\tau+5}$.
	$-\;-\;-\;-\;-\;-\;-\;-\;-\;-\;-\;-\;-\;-$
	$n_{I_{T-\tau-1}}[N_{I_{T-\tau-2}}[\cdots[N_{I_3}[N_{I_1}]]]] \to A^{T-2}$, for $s_{T-\tau-1} = n^*_{T(T-1)}(\sigma^q_{T(T-2)})$.
	$n_{I_{T-\tau-1}}(s_{T-\tau-1}) - n_{I_{T-\tau-1}}[N_{I_{T-\tau-2}}[\cdots[N_{I_3}[N_{I_1}]]]] \to A^{T-1}$.
	$n - \Big(\mu + n_{I_1}(s_1) + n_{I_2}(s_2) + n_{I_3}(s_3) + \cdots + n_{I_{T-\tau-1}}(s_{T-\tau-1})\Big) \to A^T$.
$t = T -$	$\mu + n_{I_1}(s_1) \to A^{\tau+1}$.
$\tau + 1$	$n_{I_2}(s_2) \to A^{\tau+2}$.
	$n_{I_3}[N_{I_1}] \to A^{\tau+2}$.
	$n_{I_3}(s_3) - n_{I_3}[N_{I_1}] \to A^{\tau+3}$
	$n_{I_4}[N_{I_3}[N_{I_1}]] \to A^{\tau+3}$
	$n_{I_4}(s_4) - n_{I_4}[N_{I_3}[N_{I_1}]] \to A^{\tau+4}$.
	$n_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]] \to A^{\tau+4}$.
	$n_{I_5}(s_5) - n_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]] \to A^{\tau+5}$.
	$-\;-\;-\;-\;-\;-\;-\;-\;-\;-\;-\;-\;-\;-$
	$n_{I_{T-\tau-1}}(s_{T-\tau-1}) - n_{I_{T-\tau-1}}[N_{I_{T-\tau-2}}[\cdots[N_{I_3}[N_{I_1}]]]] \to A^{T-1}$.
	$n_{I_{T-\tau}}[N_{I_{T-\tau-1}}[\cdots[N_{I_3}[N_{I_1}]]]] \to A^{T-1}$, for $s_{T-\tau} = n^*_{TT}(\sigma^q_{T(T-1)})$.
	$n_{I_{T-\tau}}(s_{T-\tau}) - n_{I_{T-\tau}}[N_{I_{T-\tau-1}}[\cdots[N_{I_3}[N_{I_1}]]]] \to A^T$
	$n - \Big(\mu + n_{I_1}(s_1) + \cdots + n_{I_{T-\tau-1}}(s_{T-\tau-1}) + n_{I_{T-\tau}}(s_{T-\tau})\Big) \to A^T$.

Thus, after $t = T-\tau$ iterations, the number of players that play strategies in A^T is at least $n - \Big(\mu + \sum_{r=1}^{T-\tau-1} n_{I_r}(s_r)\Big)$. Let z, Z and ϕ be defined as follows.

$$z = n - \left(\mu + \sum_{r=1}^{T-\tau-1} n_{I_r}(s_r) + n_{I_{T-\tau}}[N_{I_{T-\tau-1}}[\cdots[N_{I_3}[N_{I_1}]]]]\right)$$

$$Z = N \setminus \Big\{I \cup N_{I_1}(s_1) \cup \cdots \cup N_{I_{T-\tau-1}}(s_{T-\tau-1}) \cup N_{I_{T-\tau}}[N_{I_{T-\tau-1}}[\cdots[N_{I_3}[N_{I_1}]]]]\Big\}$$

$$\phi = \mu + \sum_{r=1}^{T-\tau-1} n_{I_r}(s_r) + n_{I_{T-\tau}}[N_{I_{T-\tau-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$$

From the strategy configuration at $t = T - \tau$, if $z + n_{I_{T-\tau}}[N_{I_{T-\tau-1}}[\cdots[N_{I_3}[N_{I_1}]]]] \leq \lceil pk \rceil$, so that, for every $j \in Z$ and $i \in N_{I_{T-\tau}}[N_{I_{T-\tau-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$, the number of neighbours playing strategies in A^T is less than $\lceil pk \rceil$, then at $t = T - \tau + 1$, all players in $Z \cup N_{I_{T-\tau}}[N_{I_{T-\tau-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$ can switch to strategies in $\bar{A}^{T-1}$. For this scenario, the evolutionary process can converge to an absorbing set containing either strategies in $A \backslash A^T$ or strategies in A^T together with some strategies in $A \backslash A^T$. When this happens, the evolutionary process exits the basin of attraction of $\mathbf{A}^T$ with μ mutations to strategies in $\bar{A}^{\tau}$.

However, if $z \geq \lceil pk \rceil + 1$, then all players in Z play strategies in A^T from $t = T - \tau + 1$ onward. This is because each $j \in Z$ will have at least $\lceil pk \rceil$ neighbours playing strategies in A^T and the rest play strategies in A^{T-1}.

Now, recall that there is an iterative process that unfolds at the background of the above described evolutionary process. It starts at $t = 4$, whereby, all $j \in N_{I_1}(s_1)$ with at least $\lceil pk \rceil$ neighbours in $N_{I_2}(s_2)$ switch to strategies in $A^{\tau+2}$. Let $N^1_{I_r}[N_{I_{r+1}}]$ be a set of players in $N_{I_r}(s_r)$ with at least $\lceil pk \rceil$ neighbours in $N_{I_{r+1}}(s_{r+1})$; $N^2_{I_r}[N_{I_{r+1}}]$ is a set of players in $N_{I_r}(s_r)$ with at least $\lceil pk \rceil$ neighbours in $N_{I_{r+1}}(s_{r+1}) \cup N^1_{I_r}[N_{I_{r+1}}]$; and more generally, $N^v_{I_r}[N_{I_{r+1}}]$ is a set of players in $N_{I_r}(s_r)$ with at least $\lceil pk \rceil$ neighbours in $N_{I_{r+1}}(s_{r+1}) \cup N^1_{I_r}[N_{I_{r+1}}] \cup \cdots \cup N^{v-1}_{I_r}[N_{I_{r+1}}]$.

Then, at $t = 4$, players in $N^1_{I_1}[N_{I_2}]$ play strategies in $A^{\tau+2}$; at $t = 5$, players in $N^1_{I_1}[N_{I_2}] \cup N^2_{I_1}[N_{I_2}]$ play strategies in $A^{\tau+2}$; and so on, until some $t = t_1$ when all players in $N_{I_1}(s_1)$ play strategies in $A^{\tau+2}$. Between $t = 4$ and $t = t_1$, players in $N_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$, for $3 \leq r \leq T - \tau - 1$, need not switch to strategies in $A^{\tau+r}$ even if some have at least $\lceil pk \rceil$ neighbours in $N_{I_r}(s_r) \backslash N_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$ playing strategies in $A^{\tau+1}$. This is because each $j \in N_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$ has at least one neighbour in $N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]$ playing strategies in $A^{\tau+r-2}$.

From $t = t_1 + 1$ onward, players in I with at least $\lceil pk \rceil$ neighbours in $N_{I_1}(s_1)$ play strategies in $A^{\tau+2}$. At the same time, all players in $N_{I_3}[N_{I_1}]$ with at least $\lceil pk \rceil$ neighbours in $N_{I_3}(s_3) \backslash N_{I_3}[N_{I_1}] \cup N_{I_4}[N_{I_3}[N_{I_1}]]$ switch to strategies in $A^{\tau+3}$. This process continues until some $t = t_1 + t_2$ when all players in I play strategies in $A^{\tau+2}$ and all players in $N_{I_3}[N_{I_1}]$ play strategies in $A^{\tau+3}$.

From $t = t_1 + t_2 + 1$ onward, all $j \in N_{I_2}(s_2)$ with at least $\lceil pk \rceil$ neighbours in $N_{I_3}(s_3)$ play strategies in $A^{\tau+3}$. At the same time, all players in $N_{I_4}[N_{I_3}[N_{I_1}]]$ with at least $\lceil pk \rceil$ neighbours in $N_{I_4}(s_4) \backslash N_{I_4}[N_{I_3}[N_{I_1}]] \cup N_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]]$ switch to strategies in $A^{\tau+4}$. This step-by-step evolution continues until some $t = t_1 + t_2 + t_3$ when all players in $N_{I_r}(s_r)$, for $1 \leq r \leq T - \tau - 2$, play strategies in $A^{\tau+r+1}$, and players in $Z \cup N_{I_{T-\tau}}[N_{I_{T-\tau-1}}[\cdots[N_{I_3}[N_{I_1}]]]] \cup N_{I_{T-\tau-1}}(s_{T-\tau-1})$ all play strategies in A^T. Eventually, at some $t^* > T - \tau + 1 > t_1 + t_2 + t_3$, it will converge to an absorbing set of states containing only strategies in A^T.

Thus, when $z \geq \lceil pk \rceil + 1$, more than μ (i.e. at least $\mu + 1$) mutations to strategies in $\bar{A}^\tau$ are needed to trigger an exit from $D(\mathbf{A}^T)$. Now, consider two sets, I_1 and I_2, of adjacently placed players, each of size μ. Let I_1 and I_2 be located at opposite regions of the network. That is, given $N = \{1, 2, \cdots, n\}$, with n even, if $I_1 = \{1, 2, \cdots, \mu\}$, then $I_2 = \{\frac{n}{2} + 1, \frac{n}{2} + 2, \cdots, \frac{n}{2} + \mu\}$. Let $(U, N, \mathcal{G}_k, P)$ start from some $\mathbf{x} \in \mathbf{A}^T$ and let players in I_1 and I_2 all simultaneously mutate to strategies in $\bar{A}^\tau$ at $t = 1$. Then, following the same steps above, after $t = T - \tau + 1$ iterations, at least $n - 2\phi$ players play strategies in A^T. And if $n - 2\phi \geq 2(\lceil pk \rceil + 1)$, then $(U, N, \mathcal{G}_k, P)$ will converge to an absorbing set of states containing only strategies in A^T. For this scenario, at least $2(\mu + 1)$ mutations to strategies in $\bar{A}^\tau$ are necessary to trigger an exit from $D(\mathbf{A}^T)$. More generally, if $n - \gamma\phi \geq \gamma(\lceil pk \rceil + 1)$, at least $\gamma(\mu + 1)$ mutations to strategies in $\bar{A}^\tau$ are necessary to trigger an exit from $D(\mathbf{A}^T)$.

Note that, given μ, $n - \gamma\phi$ is minimized when $\tau = 0$, $n_{I_r}(s_r) = 2\lceil k/2 \rceil - 2(s_r - 1)$, at the maximum value of $n_{I_{T-\tau}}[N_{I_{T-\tau-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$ and at the minimum value of s_r. Firstly, the following lemma establishes the upper bound for $n_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$ (see the proof at the end of this section).

Lemma 10. *The upper bound for $n_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$ is*

$$n_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]] \leq \sum_{v=2}^{r-1} \Big(2\lceil k/2 \rceil - n_{I_v}(s_v)\Big) \tag{C.1}$$

Secondly, recall from the above iterative process that $s_r = n^*_{T(\tau+r)}(\sigma^q_{T(\tau+r-1)})$. Taking into account all $r \in [1, T - \tau]$ and $\tau \in [0, T - 1]$, the minimum value of s_r is given by

$$s^* = \min_{\tau \in [0, T-1]} \min_{r \in [1, T-\tau]} s_r = n(A^T)$$

Substituting into the expression for ϕ, it follows that when

$$\begin{aligned}
&n - \gamma \left(\mu + \sum_{r=1}^{T-1} n_{I_r}(s_r) + \sum_{r=2}^{T-1} \Big(2\lceil k/2 \rceil - n_{I_r}(s_r)\Big) \right) \geq \gamma\left(\lceil pk \rceil + 1\right) \\
&\Rightarrow n - \gamma\Big(\mu + 2\lceil k/2 \rceil (T - 2) + n_{I_1}(s_1)\Big) \geq \gamma\left(\lceil pk \rceil + 1\right) \\
&\Rightarrow n \geq \gamma\Big(\mu + \lceil pk \rceil + 3 - 2n(A^T)\Big) + 2\gamma\lceil k/2 \rceil (T - 1) \quad \text{Substituting for } n_{I_1}(s_1) \text{ and } s^*
\end{aligned}$$

at least $\gamma(\mu + 1)$ mutations are necessary to trigger an exit from $D(\mathbf{A}^T)$. Substituting for the lower bound of μ, it follows that at least $\gamma(n(A^T) + 1)$ mutations are needed to trigger an exit from $D(\mathbf{A}^T)$ whenever

$$n \geq \gamma\Big(\lceil pk \rceil + 3 - n(A^T)\Big) + 2\gamma\lceil k/2 \rceil (T - 1) \tag{C.2}$$

Since $n(A^T) \geq n^*(A)$, it follows that at least $\gamma(n^*(A)+1)$ mutations are needed to trigger an exit from $D(\mathbf{A}^T)$ whenever

$$n \geq \gamma\Big(\lceil pk \rceil + 3 - n^*(A)\Big) + 2\gamma\lceil k/2 \rceil (T-1) \tag{C.3}$$

Thus, when $n^*_{T(\tau+r+1)}(\sigma^q_{T(\tau+r)}) \leq n_{I_r}(s_r) - n_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$ for all $r = 3, 4 \cdots, T-1$, $R(\mathbf{A}^T) \geq \gamma(n^*(A)+1)$ whenever (C.3) holds.

The case of $T \geq 4$ and $n^*_{T(\tau+r+1)}(\sigma^q_{T(\tau+r)}) > n^+_{I_r}(s_r) - n^+_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$

We use the following lemma in the analysis that follows (see the proof at the end of this section).

Lemma 11. *For $r = 3, 5, 7, \cdots, T-\tau-2$ (assuming $T-\tau-2$ is an odd number):*

$$\lfloor k/2 \rfloor \leq n^+_{I_r}(s_r) - n^+_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]] + n^+_{I_{r+1}}[N_{I_r}[\cdots[N_{I_3}[N_{I_1}]]]] \leq \lceil k/2 \rceil \tag{C.4}$$

$$\lfloor k/2 \rfloor \leq n^-_{I_r}(s_r) - n^-_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]] + n^-_{I_{r+1}}[N_{I_r}[\cdots[N_{I_3}[N_{I_1}]]]] \leq \lceil k/2 \rceil \tag{C.5}$$

$$2\lfloor k/2 \rfloor \leq n_{I_r}(s_r) - n_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]] + n_{I_{r+1}}[N_{I_r}[\cdots[N_{I_3}[N_{I_1}]]]] \leq 2\lceil k/2 \rceil \tag{C.6}$$

In the steps below, with a slight abuse of notation, we write $N_{I_{r+1}}(s_{r+1})$ to mean both the number of players in $N \backslash \{I \cup N_{I_1}(s_1) \cup \cdots \cup N_{I_r}(s_r)\}$ with at least s_{r+1} neighbours in $N_{I_r}(s_r)$, as defined above, and as the number of players in

$$N \backslash \Big\{ I \cup N_{I_1}(s_1) \cup \cdots \cup N_{I_r}(s_r) \cup N_{I_{r+1}}[N_{I_r}[\cdots[N_{I_3}[N_{I_1}]]]] \Big\}$$

with at least s_{r+1} neighbours in $N_{I_{r+1}}[N_{I_r}[\cdots[N_{I_3}[N_{I_1}]]]] \cup N_{I_r}(s_r) \backslash N_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$. Given these definitions and notations, the strategy configurations evolves from $t = 6$ onward as follows:

$t = 6$

$\mu \to A^{\tau+1}$.

$n_{I_1}(s_1) \to A^{\tau+1}$.

$n_{I_2}(s_2) \to A^{\tau+2}$.

$n_{I_3}[N_{I_1}] \to A^{\tau+2}$.

$n_{I_3}(s_3) - n_{I_3}[N_{I_1}] \to A^{\tau+3}$.

$n_{I_4}[N_{I_3}[N_{I_1}]] \to A^{\tau+3}$.

$n_{I_4}(s_4) \to A^{\tau+4}$, where $n_{I_4}(s_4)$ is the number of players in $N \backslash \{I \cup N_{I_1}(s_1) \cup N_{I_2}(s_2) \cup N_{I_3}(s_3) \cup N_{I_4}[N_{I_3}[N_{I_1}]]\}$ with at least $s_4 = n^*_{T(\tau+4)}(\sigma^q_{T(\tau+3)})$ neighbours in $N_{I_4}[N_{I_3}[N_{I_1}]] \cup N_{I_3}(s_3) \backslash N_{I_3}[N_{I_1}]$. Each $j \in N_{I_4}(s_4)$ plays a strategies in $A^{\tau+4}$ because, at $t = 5$, all j's neighbours in $N_{I_3}(s_3) \backslash N_{I_3}[N_{I_1}] \cup N_{I_4}[N_{I_3}[N_{I_1}]]$ (at least s_4 of them) play strategies in $A^{\tau+3}$ and the rest (at least $\lfloor k/2 \rfloor$) belong to $N \backslash \{I \cup N_{I_1}(s_1) \cup N_{I_2}(s_2) \cup N_{I_3}(s_3) \cup N_{I_4}[N_{I_3}[N_{I_1}]]\}$ and play strategies in A^T. It then follows by definition of $n^*_{T(\tau+4)}(\sigma^q_{T(\tau+3)})$ that strategies in $A^{\tau+4}$ are best responses to each $j \in N_{I_4}(s_4)$. Note that since the cardinality of $N_{I_3}(s_3) \backslash N_{I_3}[N_{I_1}] \cup N_{I_4}[N_{I_3}[N_{I_1}]]$ is either $\lfloor k/2 \rfloor$ or $\lceil k/2 \rceil$, no $j \in N_{I_4}(s_4)$ has any neighbours in $N_{I_3}[N_{I_1}]$, and hence, no $j \in N_{I_4}(s_4)$ has any neighbour playing strategies in $A^{\tau+2}$ at $t = 5$.

$n - (\mu + n_{I_1}(s_1) + n_{I_2}(s_2) + n_{I_3}(s_3) + n_{I_4}[N_{I_3}[N_{I_1}]] + n_{I_4}(s_4)) \to A^T$.

$t = 7$

$\mu \to A^{\tau+1}$.

$n_{I_1}(s_1) \to A^{\tau+1}$.

$n_{I_2}(s_2) \to A^{\tau+2}$.

$n_{I_3}[N_{I_1}] \to A^{\tau+2}$.

$n_{I_3}(s_3) - n_{I_3}[N_{I_1}] \to A^{\tau+3}$.

$n_{I_4}[N_{I_3}[N_{I_1}]] \to A^{\tau+3}$.

$n_{I_4}(s_4) \to A^{\tau+4}$.

$n_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]] \to A^{\tau+4}$, for $s_5 = n^*_{T(\tau+5)}(\sigma^q_{T(\tau+4)})$. This is because at $t = 6$, each $j \in N_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]]$ has at least $\lfloor k/2 \rfloor$ neighbours in $N \backslash \{I \cup N_{I_1}(s_1) \cup N_{I_2}(s_2) \cup N_{I_3}(s_3) \cup N_{I_4}[N_{I_3}[N_{I_1}]] \cup N_{I_4}(s_4)\}$ play strategies in A^T, $n_{I_4}(s_4)$ neighbours in $N_{I_4}(s_4)$ play strategies in $A^{\tau+4}$, and the rest (i.e. neighbours of j in $N_{I_4}[N_{I_3}[N_{I_1}]]$) play strategies in $A^{\tau+3}$. This makes strategies in $A^{\tau+4}$ best responses.

$n_{I_5}(s_5) - n_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]] \to A^{\tau+5}$ because for every $j \in N_{I_5}(s_5) \backslash N_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]]$, at least s_5 (and at most $n_{I_4}(s_4)$) neighbours in $N_{I_4}(s_4)$ play strategies in $A^{\tau+4}$ and the rest (at least $\lfloor k/2 \rfloor$ of them) belong to $N \backslash \{I \cup N_{I_1}(s_1) \cup N_{I_2}(s_2) \cup N_{I_3}(s_3) \cup N_{I_4}[N_{I_3}[N_{I_1}]] \cup N_{I_4}(s_4)\}$ and play strategies in A^T. It then follows by definition of $n^*_{T(\tau+5)}(\sigma^q_{T(\tau+4)})$ that strategies in $A^{\tau+5}$ are best responses to j.

$n - (\mu + n_{I_1}(s_1) + n_{I_2}(s_2) + n_{I_3}(s_3) + n_{I_4}[N_{I_3}[N_{I_1}]] + n_{I_4}(s_4) + n_{I_5}(s_5)) \to A^T$.

$t = 8$

$\mu \to A^{\tau+1}$.

$n_{I_1}(s_1) \to A^{\tau+1}$.

$n_{I_2}(s_2) \to A^{\tau+2}$.

$n_{I_3}[N_{I_1}] \to A^{\tau+2}$.

$n_{I_3}(s_3) - n_{I_3}[N_{I_1}] \to A^{\tau+3}$.

$n_{I_4}[N_{I_3}[N_{I_1}]] \to A^{\tau+3}$.

$n_{I_4}(s_4) \to A^{\tau+4}$.

$n_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]] \to A^{\tau+4}$.

$n_{I_5}(s_5) - n_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]] \to A^{\tau+5}$.

$n_{I_6}[N_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]]] \to A^{\tau+5}$, for $s_6 = n^*_{T(\tau+6)}(\sigma^q_{T(\tau+5)})$.

$n - (\mu + n_{I_1}(s_1) + \cdots + n_{I_4}(s_4) + n_{I_5}(s_5) + n_{I_6}[N_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]]]) \to A^T$. Note that players in $N_{I_6}(s_6) \backslash N_{I_6}[N_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]]]$ stick to strategies in A^T because we are considering a situation where $n^*_{T(\tau+6)}(\sigma^q_{T(\tau+5)}) > n^+_{I_5}(s_5) - n^+_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]], n^-_{I_5}(s_5) - n^-_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]]$.

$t = 9$	$\mu \to A^{\tau+1}$.
	$n_{I_1}(s_1) \to A^{\tau+1}$.
	$n_{I_2}(s_2) \to A^{\tau+2}$.
	$n_{I_3}[N_{I_1}] \to A^{\tau+2}$.
	$n_{I_3}(s_3) - n_{I_3}[N_{I_1}] \to A^{\tau+3}$.
	$n_{I_4}[N_{I_3}[N_{I_1}]] \to A^{\tau+3}$.
	$n_{I_4}(s_4) \to A^{\tau+4}$.
	$n_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]] \to A^{\tau+4}$.
	$n_{I_5}(s_5) - n_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]] \to A^{\tau+5}$
	$n_{I_6}[n_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]]] \to A^{\tau+5}$.
	$n_{I_6}(s_6) \to A^{\tau+6}$, where $n_{I_6}(s_6)$ is the number of players in $N\backslash \{I \cup N_{I_1}(s_1) \cup \cdots \cup N_{I_5}(s_5) \cup N_{I_6}[N_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]]]\}$ with at least s_6 neighbours in $N_{I_5}(s_5)\backslash N_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]] \cup N_{I_6}[N_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]]]$. Every $j \in N_{I_6}(s_6)$ plays a strategy in $A^{\tau+6}$ because, at $t = 8$, j has at least s_6 neighbours in $N_{I_5}(s_5)\backslash N_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]] \cup N_{I_6}[N_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]]]$ playing strategies in $A^{\tau+5}$, and the rest belong to $N\backslash \{I \cup N_{I_1}(s_1) \cup \cdots \cup N_{I_5}(s_5) \cup N_{I_6}[N_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]]]\}$ and play strategies in A^T. Thus, strategies in $A^{\tau+6}$ are best responses to all $j \in N_{I_6}(s_6)$.
	$n - (\mu + n_{I_1}(s_1) + \cdots + n_{I_5}(s_5) + n_{I_6}[N_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]]] + n_{I_6}(s_6)) \to A^T$.
$--$	$-----------------------------$

$t = \frac{3(T-\tau-1)}{2}$	$\mu \to A^{\tau+1}$. $n_{I_1}(s_1) \to A^{\tau+1}$. $n_{I_2}(s_2) \to A^{\tau+2}$. $n_{I_3}[N_{I_1}] \to A^{\tau+2}$. $n_{I_3}(s_3) - n_{I_3}[N_{I_1}] \to A^{\tau+3}$. $n_{I_4}[N_{I_3}[N_{I_1}]] \to A^{\tau+3}$. $n_{I_4}(s_4) \to A^{\tau+4}$. $n_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]] \to A^{\tau+4}$. $n_{I_5}(s_5) - n_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]] \to A^{\tau+5}$ $n_{I_6}[n_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]]] \to A^{\tau+5}$. $n_{I_6}(s_6) \to A^{\tau+6}$ $- - - - - - - - - - - - - -$ Assuming without loss of generality that $T - \tau - 1$ is even, then: $n_{I_{T-\tau-2}}[N_{I_{T-\tau-3}}[\cdots[N_{I_3}[N_{I_1}]]]]] \to A^{T-3}$. $n_{I_{T-\tau-2}}(s_{T-\tau-2}) - n_{I_{T-\tau-2}}[N_{I_{T-\tau-3}}[\cdots[N_{I_3}[N_{I_1}]]]]] \to A^{T-2}$. $n_{I_{T-\tau-1}}[N_{I_{T-\tau-2}}[\cdots[N_{I_3}[N_{I_1}]]]]] \to A^{T-2}$ $n_{I_{T-\tau-1}}(s_{T-\tau-1}) \to A^{T-1}$. $n - \Big(\mu + n_{I_1}(s_1) + \cdots + n_{I_{T-\tau-2}}(s_{T-\tau-2}) + n_{I_{T-\tau-1}}[N_{I_{T-\tau-2}}[\cdots[N_{I_3}[N_{I_1}]]]]] + n_{I_{T-\tau-1}}(s_{T-\tau-1})\Big) \to A^T$.

$$
\begin{array}{l|l}
t \;=\; \frac{3(T-\tau-1)}{2}+1 &
\begin{array}{l}
\mu \to A^{\tau+1}. \\
n_{I_1}(s_1) \to A^{\tau+1}. \\
n_{I_2}(s_2) \to A^{\tau+2}. \\
n_{I_3}[N_{I_1}] \to A^{\tau+2}. \\
n_{I_3}(s_3) - n_{I_3}[N_{I_1}] \to A^{\tau+3}. \\
n_{I_4}[N_{I_3}[N_{I_1}]] \to A^{\tau+3}. \\
n_{I_4}(s_4) \to A^{\tau+4}. \\
n_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]] \to A^{\tau+4}. \\
n_{I_5}(s_5) - n_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]] \to A^{\tau+5} \\
n_{I_6}[n_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]]] \to A^{\tau+5}. \\
n_{I_6}(s_6) \to A^{\tau+6} \\
- - - - - - - - - - - - - - \\
n_{I_{T-\tau-1}}[N_{I_{T-\tau-2}}[\cdots[N_{I_3}[N_{I_1}]]]] \to A^{T-2} \\
n_{I_{T-\tau-1}}(s_{T-\tau-1}) \to A^{T-1}. \\
n_{I_{T-\tau}}[N_{I_{T-\tau-1}}[\cdots[N_{I_3}[N_{I_1}]]]] \to A^{T-1} \\
n_{I_{T-\tau}}(s_{T-\tau}) - n_{I_{T-\tau}}[N_{I_{T-\tau-1}}[\cdots[N_{I_3}[N_{I_1}]]]] \to A^{T}. \\
n - \Big(\mu + n_{I_1}(s_1) + \cdots + n_{I_{T-\tau-2}}(s_{T-\tau-2}) + n_{I_{T-\tau-1}}[N_{I_{T-\tau-2}}[\cdots[N_{I_3}[N_{I_1}]]]] + \\
n_{I_{T-\tau-1}}(s_{T-\tau-1}) + n_{I_{T-\tau}}(s_{T-\tau})\Big) \to A^{T}.
\end{array}
\end{array}
$$

Let z, Z and ϕ be defined as follows.

$$\phi = \mu + \sum_{r=1}^{T-\tau-1} n_{I_r}(s_r) + \sum_{r=2}^{\frac{T-\tau-1}{2}} n_{I_{2r}}[N_{I_{2r-1}}[\cdots[N_{I_3}[N_{I_1}]]]] + n_{I_{T-\tau}}[N_{I_{T-\tau-1}}[\cdots[N_{I_3}[N_{I_1}]]]]]$$

$$Z = N \backslash \Big\{ I \cup N_{I_1}(s_1) \cup \cdots \cup N_{I_{T-\tau-1}}(s_{T-\tau-1}) \cup N_{I_{T-\tau}}[N_{I_{T-\tau-1}}[\cdots[N_{I_3}[N_{I_1}]]]] \Big\}$$

$$z = n - \phi$$

In the above iterative process, at least $z + n_{I_{T-\tau}}[N_{I_{T-\tau-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$ players play strategies in A^T after $t = \frac{3(T-\tau-1)}{2}$ iterations. If $z + n_{I_{T-\tau}}[N_{I_{T-\tau-1}}[\cdots[N_{I_3}[N_{I_1}]]]] \leq \lceil pk \rceil$, so that each $j \in Z$ and $i \in N_{I_{T-\tau}}[N_{I_{T-\tau-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$ has less than $\lceil pk \rceil$ play strategies in A^T, then at $t = \frac{3(T-\tau-1)}{2} + 1$, all players in $Z \cup N_{I_{T-\tau}}[N_{I_{T-\tau-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$ may switch to strategies in $\bar{A}^{T-1}$. For this scenario, $(U, N, \mathcal{G}_k, P)$ can converge to an absorbing set containing either strategies in $A \backslash A^T$ or both strategies in A^T and some strategies in $A \backslash A^T$. When this happens, $(U, N, \mathcal{G}_k, P)$ exits $D(\mathbf{A}^T)$ with μ mutations to strategies in $\bar{A}^{\tau}$.

However, if $z \geq \lceil pk \rceil + 1$, then all $j \in Z$ play strategies in A^T from $t = \frac{3(T-\tau-1)}{2} + 1$ onward because each has at least $\lceil pk \rceil$ neighbours playing strategies in A^T and the rest play strategies in A^{T-1}.

At the background of the above evolutionary process is an iterative process whereby, at $t=4$, all $j \in N^1_{I_1}[N_{I_2}]$ play strategies in $A^{\tau+2}$. At $t=5$, all $j \in N^1_{I_1}[N_{I_2}] \cup N^2_{I_1}[N_{I_2}]$ play strategies in $A^{\tau+2}$; followed by all $j \in N^1_{I_1}[N_{I_2}] \cup N^2_{I_1}[N_{I_2}] \cup N^3_{I_1}[N_{I_2}]$ at $t=6$; and so on, until some $t=t_1$ when all players in $N_{I_1}(s_1)$ play strategies in $A^{\tau+2}$.

At $t=t_1+1$, players in I with at least $\lceil pk \rceil$ neighbours in $N_{I_1}(s_1)$ play strategies in $A^{\tau+2}$. At the same time, all players in $N_{I_3}[N_{I_1}]$ with at least $\lceil pk \rceil$ neighbours in $N_{I_3}(s_3) \backslash N_{I_3}[N_{I_1}] \cup N_{I_4}[N_{I_3}[N_{I_1}]]$ switch to strategies in $A^{\tau+3}$. This process continues until some $t=t_1+t_2$ when all players in I play strategies in $A^{\tau+2}$ and all players in $N_{I_3}[N_{I_1}]$ play strategies in $A^{\tau+3}$.

At $t=t_1+t_2+1$, all $j \in N_{I_2}(s_2)$ with at least $\lceil pk \rceil$ neighbours in $N_{I_3}(s_3)$ play strategies in $A^{\tau+3}$. At the same time, all players in $N_{I_4}[N_{I_3}[N_{I_1}]]$ with at least $\lceil pk \rceil$ neighbours in $N_{I_4}(s_4) \backslash N_{I_4}[N_{I_3}[N_{I_1}]] \cup N_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]]$ switch to strategies in $A^{\tau+4}$. This continues until some $t=t_1+t_2+t_3$ when all players in $N_{I_r}(s_r)$, for $1 \leq r \leq T-\tau-2$, play strategies in $A^{\tau+r+1}$, and players in $Z \cup N_{I_{T-\tau}}[N_{I_{T-\tau-1}}[\cdots[N_{I_3}[N_{I_1}]]]] \cup N_{I_{T-\tau-1}}(s_{T-\tau-1})$ all play strategies in A^T. Eventually, at some $t^* > \frac{3(T-\tau-1)}{2}+1$ the process will converge to an absorbing set of states containing only strategies in A^T.

Thus, when $z \geq \lceil pk \rceil + 1$, at least $\mu+1$ mutations to strategies in $\bar{A}^\tau$ are needed to trigger an exit from $D(\mathbf{A}^T)$. And more generally, if $n-\gamma\phi \geq \gamma(\lceil pk \rceil+1)$, at least $\gamma(\mu+1)$ mutations to strategies in $\bar{A}^\tau$ are needed to trigger an exit from $D(\mathbf{A}^T)$ (see the discussion above).

Given μ, $n-\gamma\phi$ is minimized when $\tau=0$, $n_{I_r}(s_r)=2\lceil k/2 \rceil - 2(s_r-1)$, $s_r=s^*=n(A^T)$, and at the maximum values of $n_{I_{2r}}[N_{I_{2r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$ and $n_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$. The following lemma establishes the upper bound for the latter (see the proof of the following lemma at the end of this section).

Lemma 12. *For $r=2,3,\cdots,\frac{T-\tau-1}{2}$, where $T-\tau-1$ is an even number, we have*

$$n_{I_{2r}}[N_{I_{2r-1}}[\cdots[N_{I_3}[N_{I_1}]]]] \leq 4\lceil k/2 \rceil - \Big(n_{I_{2r-1}}(s_{2r-1}) + n_{I_{2r-2}}(s_{2r-2})\Big) \tag{C.7}$$

$$n_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]] \leq 2\lceil k/2 \rceil - n_{I_{r-1}}(s_{r-1}) \tag{C.8}$$

Substituting $\tau=0$, $n_{I_r}(s_r)=2\lceil k/2 \rceil - 2(s_r-1)$ and the expressions in (C.7) and (C.8) into $n-\gamma\phi$ then implies that at least $\gamma(\mu+1)$ mutations are needed to trigger an exit from $D(\mathbf{A}^T)$

whenever

$$n-\gamma\left(\mu+\sum_{r=1}^{T-1} n_{I_r}(s_r)+\sum_{r=2}^{\frac{T-1}{2}}\Big(4\lceil k/2\rceil-\Big(n_{I_{2r-1}}(s_{2r-1})+n_{I_{2r-2}}(s_{2r-2})\Big)\Big)+\Big(2\lceil k/2\rceil-n_{I_{T-1}}(s_{T-1})\Big)\right)$$

$$\geq\gamma\Big(\lceil pk\rceil+1\Big)$$

$$\Rightarrow n-\gamma\left(\mu+4\lceil k/2\rceil\left(\frac{T-1}{2}-1\right)+2\lceil k/2\rceil+\sum_{r=1}^{T-1} n_{I_r}(s_r)-\sum_{r=2}^{T-2} n_{I_r}(s_r)-n_{I_{T-1}}(s_{T-1})\right)\geq\gamma\Big(\lceil pk\rceil+1\Big)$$

$$\Rightarrow n-\gamma\Big(\mu+2\lceil k/2\rceil(T-2)+n_{I_1}(s_1)\Big)\geq\gamma\Big(\lceil pk\rceil+1\Big)$$

$$\Rightarrow n-\gamma\Big(\mu+2\lceil k/2\rceil(T-1)+2-2n(A^T)\Big)\geq\gamma\Big(\lceil pk\rceil+1\Big)\quad \text{Substituting for } n_{I_1}(s_1) \text{ and } s^*$$

$$\Rightarrow n\geq\gamma\Big(\mu+\lceil pk\rceil+3-2n(A^T)\Big)+2\gamma\lceil k/2\rceil(T-1)$$

Thus, when $n^*_{T(\tau+r+1)}(\sigma^q_{T(\tau+r)}) > n_{I_r}(s_r) - n_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$ for all $r = 3, 4, \cdots, T-1$, at least $\gamma(n(A^T)+1)$ mutations are needed to trigger an exit from $D(\mathbf{A}^T)$ whenever

$$n\geq\gamma\Big(\lceil pk\rceil+3-n(A^T)\Big)+2\gamma\lceil k/2\rceil(T-1) \tag{C.9}$$

Since $n(A^T)\geq n^*(A)$, it follows that at least $\gamma(n^*(A)+1)$ mutations are needed to trigger an exit from $D(\mathbf{A}^T)$ whenever

$$n\geq\gamma\Big(\lceil pk\rceil+3-n^*(A)\Big)+2\gamma\lceil k/2\rceil(T-1) \tag{C.10}$$

Finally, the right hand sides of (C.10) and (C.3) are equal, which implies that, for all $T\geq 2$, $R(\mathbf{A}^T)\geq\gamma(n^*(A)+1)$ whenever any of (C.3) and (C.10) holds.

Appendix C.1. Proof of Lemma 10

First note that for $r=3$, there are $n_{I_2}(s_2)$ players between $N_{I_3}[N_{I_1}]$ and $N_{I_1}(s_1)$; for $r = 4,\cdots,T-\tau$, there are $n_{I_{r-1}}(s_{r-1})-n_{I_{r-1}}[N_{I_{r-2}}[\cdots[N_{I_3}[N_{I_1}]]]]$ players between $N_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$ and $N_{I_{r-1}}[N_{I_{r-2}}[\cdots[N_{I_3}[N_{I_1}]]]]$. Let $n^+_{I_{r-1}}(s_{r-1})-n^+_{I_{r-1}}[N_{I_{r-2}}[\cdots[N_{I_3}[N_{I_1}]]]]=\alpha$ so that

$$N^+_{I_{r-1}}(s_{r-1})\backslash N^+_{I_{r-1}}[N_{I_{r-2}}[\cdots[N_{I_3}[N_{I_1}]]]]=\{j,j+1,\cdots,j+\alpha-1\}$$

Using this labeling, it follows by definition that player $j-1\in N^+_{I_{r-1}}[N_{I_{r-2}}[\cdots[N_{I_3}[N_{I_1}]]]]$. The upper bound for $n^+_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$ obtains when player $j-1$ has $\lceil k/2\rceil$ neighbours to the right. For this scenario, it follows by definition that $j-1+\lceil k/2\rceil\in N^+_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$, and that:

$$N^+_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]=\{j+\alpha,j+\alpha+1,\cdots,j-1+\lceil k/2\rceil\}$$

The cardinality of $N^+_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$ is then:

$$\begin{aligned} n^+_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]] &\leq j-1+\lceil k/2\rceil-(j+\alpha-1)=\lceil k/2\rceil-\alpha \\ &=\lceil k/2\rceil-\Big(n^+_{I_{r-1}}(s_{r-1})-n^+_{I_{r-1}}[N_{I_{r-2}}[\cdots[N_{I_3}[N_{I_1}]]]]\Big) \end{aligned} \tag{C.11}$$

The same argument follows for $n^-_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$, so that

$$\begin{aligned} n_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]] &= n^+_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]] + n^-_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]] \\ &\leq \lceil k/2 \rceil - \Big(n^+_{I_{r-1}}(s_{r-1}) - n^+_{I_{r-1}}[N_{I_{r-2}}[\cdots[N_{I_3}[N_{I_1}]]]]\Big) \\ &+ \lceil k/2 \rceil - \Big(n^-_{I_{r-1}}(s_{r-1}) - n^-_{I_{r-1}}[N_{I_{r-2}}[\cdots[N_{I_3}[N_{I_1}]]]]\Big) \\ &= 2\lceil k/2 \rceil - n_{I_{r-1}}(s_{r-1}) + n_{I_{r-1}}[N_{I_{r-2}}[\cdots[N_{I_3}[N_{I_1}]]]] \end{aligned} \tag{C.12}$$

Following the same steps, we find that

$$n_{I_{r-1}}[N_{I_{r-2}}[\cdots[N_{I_3}[N_{I_1}]]]] \leq 2\lceil k/2 \rceil - n_{I_{r-2}}(s_{r-2}) + n_{I_{r-2}}[N_{I_{r-3}}[\cdots[N_{I_3}[N_{I_1}]]]] \tag{C.13}$$

Substituting (C.13) into (C.12) yields

$$n_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]] \leq 4\lceil k/2 \rceil - n_{I_{r-1}}(s_{r-1}) - n_{I_{r-2}}(s_{r-2}) + n_{I_{r-2}}[N_{I_{r-3}}[\cdots[N_{I_3}[N_{I_1}]]]]$$

By iteratively substituting for the upper bound of $n_{I_v}[N_{I_{v-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$ for all $v = r-2, r-3, \cdots, 3$, we find that

$$n_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]] \leq \sum_{v=2}^{r-1} \Big(2\lceil k/2 \rceil - n_{I_v}(s_v)\Big) \tag{C.14}$$

Appendix C.2. Proof of Lemma 11

To prove Lemma 11, let $n^+_{I_r}(s_r) - n^+_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]] = \beta$ and consider the following labeling of the elements of set $N^+_{I_r}(s_r)\backslash N^+_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$

$$N^+_{I_r}(s_r)\backslash N^+_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]] = \{j, j+1, \cdots, j+\beta-1\}$$

Using this labeling, it follows by definition that player $j-1 \in N^+_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$. If player $j-1$ has $\lceil k/2 \rceil$ neighbours to the right, then $j-1+\lceil k/2 \rceil \in N^+_{I_{r+1}}[N_{I_r}[\cdots[N_{I_3}[N_{I_1}]]]]$, and that:

$$N^+_{I_{r+1}}[N_{I_r}[\cdots[N_{I_3}[N_{I_1}]]]] = \{j+\beta, j+\beta+1, \cdots, j-1+\lceil k/2 \rceil\}$$

$$\begin{aligned} &N^+_{I_r}(s_r)\backslash N^+_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]] \cup N^+_{I_{r+1}}[N_{I_r}[\cdots[N_{I_3}[N_{I_1}]]]] \\ &\qquad = \{j, j+1, \cdots, j+\beta-1, j+\beta, \cdots, j-1+\lceil k/2 \rceil\} \end{aligned}$$

Similarly, if player $j-1$ has $\lfloor k/2 \rfloor$ neighbours to the right, then

$$\begin{aligned} &N^+_{I_r}(s_r)\backslash N^+_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]] \cup N^+_{I_{r+1}}[N_{I_r}[\cdots[N_{I_3}[N_{I_1}]]]] \\ &\qquad = \{j, j+1, \cdots, j+\beta-1, j+\beta, \cdots, j-1+\lfloor k/2 \rfloor\} \end{aligned}$$

Thus, the cardinality of set $N^+_{I_r}(s_r)\backslash N^+_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]] \cup N^+_{I_{r+1}}[N_{I_r}[\cdots[N_{I_3}[N_{I_1}]]]]$ takes on the value $\lfloor k/2 \rfloor$ or $\lceil k/2 \rceil$, which proves the bounds in (C.4), and respectively the bounds for $N_{I_r}(s_r)\backslash N_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]] \cup N_{I_{r+1}}[N_{I_r}[\cdots[N_{I_3}[N_{I_1}]]]]$ in (C.6).

Appendix C.3. Proof of Lemma 12

The upper bound for $N_{I_{2r}}[N_{I_{2r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$ can be derived from (C.6) since for $r = 3, 5, 7, \cdots, T-\tau-2$, $r+1$ is an even number. From (C.6), we have

$$n_{I_{r+1}}[N_{I_r}[\cdots[N_{I_3}[N_{I_1}]]]] \leq 2\lceil k/2\rceil - \Big(n_{I_r}(s_r) - n_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]\Big) \tag{C.15}$$

We now derive the bounds for $n_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$. For any for $r = 3, 5, 7, \cdots, T-\tau-2$, there are $n_{I_{r-1}}(s_{r-1})$ players between set $N_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$ and set $N_{I_{r-1}}[N_{I_{r-2}}[\cdots[N_{I_3}[N_{I_1}]]]]$. Consider the subsets of each of these three sets that consist of players to the right of I, that is, $N^+_{I_{r-1}}[N_{I_{r-2}}[\cdots[N_{I_3}[N_{I_1}]]]]$, $N^+_{I_{r-1}}(s_{r-1})$ and $N^+_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$. Let $n^+_{I_{r-1}}(s_{r-1}) = \alpha$ and consider the following labeling of the elements of set $N^+_{I_{r-1}}(s_{r-1})$

$$N^+_{I_{r-1}}(s_{r-1}) = \{j, j+1, \cdots, j+\alpha-1\}$$

Using this labeling, it follows by definition that player $j-1 \in N^+_{I_{r-1}}[N_{I_{r-2}}[\cdots[N_{I_3}[N_{I_1}]]]]$. First consider a scenario where player $j-1$ has $\lceil k/2\rceil$ neighbours to the right. For this scenario, it follows by definition that $j-1+\lceil k/2\rceil \in N^+_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$, and that

$$N^+_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]] = \{j+\alpha, j+\alpha+1, \cdots, j-1+\lceil k/2\rceil\}$$

Thus, the cardinality of $N^+_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$ is:

$$n^+_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]] = j-1+\lceil k/2\rceil - (j+\alpha-1) = \lceil k/2\rceil - \alpha = \lceil k/2\rceil - n^+_{I_{r-1}}(s_{r-1})$$

Similarly, if player $j-1$ has $\lfloor k/2\rfloor$ neighbours to the right, then

$$N^+_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]] = \{j+\alpha, j+\alpha+1, \cdots, j-1+\lfloor k/2\rfloor\}$$

The respective cardinality of $N^+_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$ in this scenario is $n^+_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]] = \lfloor k/2\rfloor - n^+_{I_{r-1}}(s_{r-1})$. The same argument follows for $n^-_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$, so that

$$\begin{aligned} n_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]] &= n^+_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]] + n^-_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]] \\ &\leq 2\lceil k/2\rceil - n_{I_{r-1}}(s_{r-1}) \end{aligned} \tag{C.16}$$

This completes for the upper bound of $n_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$.

To derive the upper bound of $n_{I_{2r}}[N_{I_{2r-1}}[\cdots[N_{I_3}[N_{I_1}]]]] \equiv n_{I_{r+1}}[N_{I_r}[\cdots[N_{I_3}[N_{I_1}]]]]$, first substitute (C.16) into (C.15) to obtain

$$n_{I_{r+1}}[N_{I_r}[\cdots[N_{I_3}[N_{I_1}]]]] \leq 4\lceil k/2\rceil - \Big(n_{I_r}(s_r) + n_{I_{r-1}}(s_{r-1})\Big) \tag{C.17}$$

This can equivalently be written as

$$n_{I_{2r}}[N_{I_{2r-}}[\cdots[N_{I_3}[N_{I_1}]]]] \leq 4\lceil k/2\rceil - \Big(n_{I_{2r-1}}(s_{2r-1}) + n_{I_{2r-2}}(s_{2r-2})\Big) \tag{C.18}$$

Appendix D. Proof of Lemma 7

The proof of Lemma 7 follows in two steps. Recall the structure of absorbing sets of $(U, N, \mathcal{G}_k, P)$ as:

$$\mathbf{A} \equiv \bigcup_{\tau=0}^{T} \left(M(\bar{\mathbf{A}}^\tau) \cup Q(\bar{\mathbf{A}}^\tau)\right) \bigcup L(\mathbf{A}) \equiv \bigcup_{\tau=0}^{T} \left(M(\bar{\mathbf{A}}^\tau) \cup Q(\bar{\mathbf{A}}^\tau) \cup L(\mathbf{A}^\tau)\right) \tag{D.1}$$

where $M(\bar{\mathbf{A}}^\tau)$ and $Q(\bar{\mathbf{A}}^\tau)$ are the sets of monomorphic and polymorphic absorbing states containing only strategies in $\bar{A}^\tau$; $L(\mathbf{A})$ is the set of all absorbing cycles; $L(\mathbf{A}^\tau)$ is the set of all absorbing cycles containing either only strategies in $\bar{A}^\tau$ or strategies in $\bar{A}^\tau$ together with some strategies in $A^{\tau+1}$; $\bar{\mathbf{A}}^\tau = M(\bar{\mathbf{A}}^\tau) \cup Q(\bar{\mathbf{A}}^\tau) \cup L(\mathbf{A}^\tau)$; and $\mathbf{A}^\tau$ is the set of all absorbing sets containing only strategies in A^τ, that is,

$$\mathbf{A}^\tau \equiv \bigcup_{v=\tau}^{T} \left(M(\bar{\mathbf{A}}^v) \cup Q(\bar{\mathbf{A}}^v) \cup L(\mathbf{A}^v)\right) \tag{D.2}$$

We first show that, for any $W \subseteq \bar{\mathbf{A}}^\tau$, there exists at least one $W_1 \subseteq \mathbf{A}^\tau$ with $C_D(W, W_1) \leq n(p,k)$. We then show that, if inequalities (C.3) and (C.10) hold, then $C_D(W, W'') \geq \gamma(n(A^T)+1)$ for any $W \subseteq \bar{\mathbf{A}}^\tau$ and $W'' \subseteq \bigcup_{v=0}^{\tau-1} \left(M(\bar{\mathbf{A}}^v) \cup Q(\bar{\mathbf{A}}^v) \cup L(\mathbf{A}^v)\right)$.

The preceding two statements imply that, for any $W \subseteq \bar{\mathbf{A}}^\tau$, if (C.3) and (C.10) hold, and $\gamma(n(A^T)+1) > n(p,k)$, then $R(W) = C_D(W, W')$, for some $W' \subseteq \mathbf{A}^\tau$. Finally, we show that since $R(W) = C_D(W, W')$ for every $W \subseteq \bar{\mathbf{A}}^\tau$ and some $W' \subseteq \mathbf{A}^\tau$, the modified coradius of $\mathbf{A}^T$ is given by:

$$CR^*(\mathbf{A}^T) = \max_{\tau \in [0, T-1]} \max_{W \subseteq \bar{\mathbf{A}}^\tau} \min_{W_1 \subseteq \mathbf{A}^\tau} C_D(W, W_1) \leq n(p,k)$$

We now derive the upper bound of $C_D(W, W_1)$, for any $W \subseteq \bar{\mathbf{A}}^\tau$ with corresponding $W_1 \subseteq \mathbf{A}^\tau$. Let the process $(U, N, \mathcal{G}_k, P)$ start from any $\mathbf{x} \in W \subseteq \bar{\mathbf{A}}^\tau$ containing either only strategies in $\bar{A}^\tau$ or both strategies in $\bar{A}^\tau$ and $A^{\tau+1}$. Let all players in $I = \{j, j+1, \cdots, j+s-1\}$, where $s = \lceil pk \rceil$, mutate to strategies in $A^{\tau+1}$ at $t = 1$. Then at $t = 2$, some players in I revert to strategies in $\bar{A}^\tau$ but all players in $N_{I_1}(s)$ switch to strategies in $A^{\tau+1}$. There are three scenarios that determine the dynamics from $t = 3$ onward.

First, if $n_{I_1}(s)^+ \geq s$ and/or $n_{I_1}^-(s) \geq s$, then, at $t = 3$, all players in I, $N_{I_2}(s)$ and some players within $N_{I_1}(s)$ all switch to strategies in $A^{\tau+1}$. From $t = 4$ onward, strategies in $A^{\tau+1}$ spread to $N_{I_3}(s)$, then to $N_{I_4}(s)$, and so on. The process $(U, N, \mathcal{G}_k, P)$ will eventually converge to either an absorbing set of states containing only strategies in $A^{\tau+1}$ or an absorbing cycle containing strategies in both $\bar{A}^\tau$ and $A^{\tau+1}$ (recall that there are no absorbing states where strategies in $\bar{A}^\tau$ coexist with strategies in $A^{\tau+1}$). That is, $(U, N, \mathcal{G}_k, P)$ cannot revert to $\mathbf{x}$ because if $n_{I_1}(s)^+ \geq s$ and/or $n_{I_1}^-(s) \geq s$, so that $n_{I_1}(s) \geq s$, then either all players in $I \cup N_{I_1}(s)$ all switch to strategies in $A^{\tau+1}$ at some $t^* \geq 2$, triggering a network-wide cascade of strategies in $A^{\tau+1}$, or players in I and $N_{I_1}(s)$ alternate between strategies in $\bar{A}^{\tau}$ and $A^{\tau+1}$. The latter follows because all players in I have at least $n_{I_1}(s) \geq s$ neighbours in $N_{I_1}(s)$, and hence, will play strategies in $A^{\tau+1}$ at $t+1$ if players in $N_{I_1}(s)$ play strategies in $A^{\tau+1}$ at t, and players in $N_{I_1}(s)$ will play strategies in $A^{\tau+1}$ at $t+1$ if players in I play strategies in $A^{\tau+1}$ at t.

Second, if both $n_{I_1}(s)^+ < s$ and $n^-_{\bar{I}_1}(s) < s$ but $n_{I_1}(s) \geq s$ then, $(U, N, \mathcal{G}_k, P)$ will converge to an absorbing cycle where players in I and $N_{I_1}(s)$ alternate between strategies in $\bar{A}^{\tau}$ and $A^{\tau+1}$. This is because, when $n_{I_1}(s) \geq s$, each $i \in I$ has at least $s = \lceil pk \rceil$ neighbours in $n_{I_1}(s)$. Thus, when $n_{I_1}(s) \geq s$, players in I play strategies in $A^{\tau+1}$ at $t+1$ if players in $N_{I_1}(s)$ play strategies in $A^{\tau+1}$ at t, and vice versa.

Third, when $n_{I_1}(s) < s$, which by definition implies both $n_{I_1}(s)^+ < s$ and $n^-_{\bar{I}_1}(s) < s$, then $(U, N, \mathcal{G}_k, P)$ can revert to $\mathbf{x}$ at $t = 3$ because all players in $I \cup N_{I_1}(s)$ have less than $\lceil pk \rceil$ neighbours play strategies in $A^{\tau+1}$, and hence, these strategies need not be best responses. For this scenario $\lceil pk \rceil + 1$ mutations to strategies in $A^{\tau+1}$ are sufficient to trigger an exit from $D(W)$. That is, if $\lceil pk \rceil + 1$ players in $I' = \{j, j+1, \cdots, j+s\}$ mutate to strategies in $A^{\tau+1}$ at $t = 2$, starting from $\mathbf{x} \in W \subseteq \bar{\mathbf{A}}^{\tau}$, then from $t = 2$ onward, all players in I' play strategies in $A^{\tau+1}$ since they are played by at least $s\lceil pk \rceil$ neighbours; similarly, all players in $N_{I'_1}$ play strategies in $A^{\tau+1}$ from $t = 2$ onward; from $t = 3$ onward, all players in $I' \cup N_{I'_1} \cup N_{I'_2}$ play strategies in $A^{\tau+1}$; and so on, until the entire network eventually switches to strategies in $A^{\tau+1}$.

Thus, when $n_{I_1}(s) \geq s$, at most $\lceil pk \rceil$ mutations trigger an exit from the basin of attraction of $W \subseteq \bar{\mathbf{A}}^{\tau}$ to some $W_1 \subseteq \mathbf{A}^{\tau}$. And when $n_{I_1}(s) < s$, at most $\lceil pk \rceil + 1$ mutations trigger an exit from W. Recall that that $n_{I_1}(s) \leq 2\lceil k/2 \rceil - 2(s-1)$, so that $n_{I_1}(s) \geq s$ implies that $3\lceil pk \rceil \leq 2(\lceil k/2 \rceil + 1)$. Thus, when (C.3) and (C.10) hold, we have

$$C_D(W, W_1) \leq \begin{cases} \lceil pk \rceil & \text{if } 3\lceil pk \rceil \leq 2(\lceil k/2 \rceil + 1) \\ \lceil pk \rceil + 1 & \text{if } 3\lceil pk \rceil > 2(\lceil k/2 \rceil + 1) \end{cases} \tag{D.3}$$

Next, we show that when (C.3) and (C.10) hold, $C_D(W, W'') \geq \gamma(n(A^T)+1)$ for any $W \subseteq \bar{\mathbf{A}}^{\tau}$ and $W'' \subseteq \bigcup_{v=0}^{\tau-1} \left(M(\bar{\mathbf{A}}^v) \cup Q(\bar{\mathbf{A}}^v) \cup L(\mathbf{A}^v) \right)$. That is, given that $(U, N, \mathcal{G}_k, P)$ starts from some $\mathbf{x} \in W \subseteq \bar{\mathbf{A}}^{\tau}$, if less than $\gamma(n^*(A)+1)$ players mutate to strategies in $\bar{A}^v$, for all $0 \leq v \leq \tau - 1$, then $(U, N, \mathcal{G}_k, P)$ will either revert to some state in W or converge to an absorbing set containing only strategies in A^{τ} whenever (C.3) and (C.10) hold.

Consider an analogous process to the dynamic process leading up to condition (C.3). That is, let $(U, N, \mathcal{G}_k, P)$ start from $\mathbf{x} \in W \subseteq \bar{\mathbf{A}}^{\tau}$, and at $t = 1$, let μ adjacently placed players in $I = \{j, j+1, \cdots, j+\mu-1\}$, for $n^*_{\tau(v+1)}(\sigma^q_{\tau v}) \leq \mu \leq \lceil pk \rceil + 1$, mutate to strategies in $\bar{A}^v$.

We first consider a scenario where $n^*_{\tau(v+r+1)}(\sigma^q_{\tau(v+r)}) \leq n^+_{I_r}(s_r) - n^+_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$ for all $r = 3, 4, \cdots, \tau - 1$. Then, from $t = 2$ onward, the strategy configurations of $(U, N, \mathcal{G}_k, P)$ evolve as follows.

$t = 1$ | $\mu \to \bar{A}^v$;

$n - \mu \to A^\tau$.

$t = 2$ | $\mu \to A^{v+1}$. That is, when $n^*_{\tau(v+1)}(\sigma^q_{\tau v}) < \mu \leq \lceil pk \rceil + 1$, each player in I has at least $n^*_{\tau(v+1)}(\sigma^q_{\tau v})$ neighbours play strategies in $\bar{A}^v$ and the rest play strategies in A^τ. This, by definition of $n^*_{\tau(v+1)}(\sigma^q_{\tau v})$, implies that strategies in A^{v+1} are best responses. When $\mu = n^*_{T(\tau+1)}(\sigma^q_{T\tau})$, we have $\mu \to A^\tau$. This is because, at $t = 1$, each $j \in I$ would have $n^*_{T(\tau+1)}(\sigma^q_{T\tau}) - 1 < n^*_{T(\tau+1)}(\sigma^q_{T\tau})$ neighbours within I play strategies in $\bar{A}^v$ and the rest play strategies in A^τ. By definition of $n^*_{T(\tau+1)}(\sigma^q_{T\tau})$, strategies in A^τ are best responses to all $j \in I$. Since $A^\tau \subseteq A^{v+1}$ (assuming $v + 1 \leq \tau$), we simply write $\mu \to A^{v+1}$.

For the same reason, $n_{I_1}(s_1) \to A^{v+1}$, for $s_1 = n^*_{\tau(v+1)}(\sigma^q_{\tau v})$, where, from Lemmas 8 and 9, $s_1 \leq \mu$. That is, at $t = 1$, each $j \in N_{I_1}(s_1)$ has at least s_1 neighbours in I play strategies in $\bar{A}^v$ and the rest play strategies in A^τ; by definition of $n^*_{\tau(v+1)}(\sigma^q_{\tau v})$ strategies in A^{v+1} are best response to all $i \in N_{I_1}(s_1)$.

$n - (\mu + n_{I_1}(s_1)) \to A^\tau$.

$t = 3$ | $\mu \to A^{v+1}$, because each $i \in I$ has all her neighbours play strategies in A^{v+1} at $t = 2$.

For the same reason, $n_{I_1}(s_1) \to A^{v+1}$ (considering that $A^\tau \subseteq A^{v+1}$). Note that some players in $N_{I_1}(s_1)$ with a sufficiently large number of neighbours in $N \backslash \{I \cup N_{I_1}(s_1)\}$ will switch to strategies in A^{v+2}; but since $A^{v+2} \subseteq A^{v+1}$, we simply write $n_{I_1}(s_1) \to A^{v+1}$.

If $v + 2 \leq \tau$ so that $A^\tau \subseteq A^{v+2}$, then $n_{I_2}(s_2) \to A^{v+2}$, for $s_2 = n^*_{\tau(v+2)}(\sigma^q_{\tau(v+1)})$. This is because, at $t = 2$, each $i \in N_{I_2}(s_2)$ has at least $\lfloor k/2 \rfloor$ neighbours in $N \backslash \{I \cup N_{I_1}(s_1)\}$ play strategies in A^τ and the rest play strategies in $A^{v+1} \backslash \bar{A}^\tau$. Thus, by definition of $n^*_{\tau(v+2)}(\sigma^q_{\tau(v+1)})$, strategies in A^{v+2} are best responses to all $i \in N_{I_2}(s_2)$.

$n - (\mu + n_{I_1}(s_1) + n_{I_2}(s_2)) \to A^\tau$.

$t = 4$

$\mu \to A^{v+1}$.

$n_{I_1}(s_1) \to A^{v+1}$; all $j \in n_{I_1}(s_1)$ with at least $\lceil pk \rceil$ neighbours in $N_{I_2}(s_2)$ play strategies in A^{v+2}, but since $A^{v+2} \subseteq A^{v+1}$, we simply write $n_{I_1}(s_1) \to A^{v+1}$; we return to this part of the dynamics below since it determines whether the entire network ultimately switches to strategies in A^{τ}.

$n_{I_2}(s_2) \to A^{v+2}$ because, at $t = 3$, each $j \in N_{I_2}(s_2)$ has at least $\lfloor k/2 \rfloor$ neighbours in $N \setminus \{I \cup N_{I_1}(s_1)\}$ play strategies in A^{v+2} (considering the fact that $A^{\tau} \subseteq A^{v+2}$) and at most $\lceil k/2 \rceil$ neighbours in $I \cup N_{I_1}(s_1)$ play strategies in $\bar{A}^{v+1} \subseteq A^{v+1}$, which, by definition of p-best response sets, makes strategies in A^{v+2} best responses to all $j \in N_{I_2}(s_2)$.

For $s_3 = n^*_{\tau(v+3)}(\sigma^q_{\tau(v+2)})$, and assuming $v + 3 \leq \tau$ so that $A^{\tau} \subseteq A^{v+3}$, we have:

$n_{I_3}[N_{I_1}] \to A^{v+2}$. This is because at $t = 3$, each $j \in N_{I_3}[N_{I_1}]$ has at least $\lfloor k/2 \rfloor$ neighbours in $N \setminus \{I \cup N_{I_1}(s_1) \cup N_{I_2}(s_2)\}$ play strategies in A^{τ}, either $n^+_{I_2}(s_2)$ or $n^-_{I_2}(s_2)$ neighbours in $N_{I_2}(s_2)$ play strategies in A^{v+2}, and the rest (i.e. neighbours of j that belong to $N_{I_1}(s_1) \cup I$) play strategies in A^{v+1}. This implies that more than $\lfloor k/2 \rfloor$ neighbours of j play strategies in A^{v+2} (where $A^{\tau} \subset A^{v+2}$) and at most $\lceil k/2 \rceil$ play strategies in $\bar{A}^{v+1} \subset A^{v+1}$, which, by definition of p-best response sets, implies that strategies in A^{v+2} are best responses to all $j \in N_{I_3}[N_{I_1}]$.

$n_{I_3}(s_3) - n_{I_3}[N_{I_1}] \to A^{v+3}$. This is because at $t = 3$, each $j \in N_{I_3}(s_3) \setminus N_{I_3}[N_{I_1}]$ has at least $\lfloor k/2 \rfloor$ neighbours in $N \setminus \{I \cup N_{I_1}(s_1) \cup N_{I_2}(s_2)\}$ play strategies in A^{τ} and the rest (i.e. neighbours of j that belong to $N_{I_2}(s_2)$) play strategies in A^{v+2}. Thus, by definition of $n^*_{\tau(v+3)}(\sigma^q_{\tau(v+2)})$, strategies in A^{v+3} are best responses to all $j \in N_{I_3}(s_3) \setminus N_{I_3}[N_{I_1}]$.

$n - (\mu + n_{I_1}(s_1) + n_{I_2}(s_2) + n_{I_3}(s_3)) \to A^{\tau}$.

$t = 5$ | $\mu \to A^{v+1}$.

$n_{I_1}(s_1) \to A^{v+1}$.

$n_{I_2}(s_2) \to A^{v+2}$.

$n_{I_3}[N_{I_1}] \to A^{v+2}$.

$n_{I_3}(s_3) - n_{I_3}[N_{I_1}] \to A^{v+3}$.

For $s_4 = n^*_{\tau(v+4)}(\sigma^q_{\tau(v+3)})$, and assuming $v + 4 \leq \tau$ so that $A^\tau \subseteq A^{v+4}$, we have:

$n_{I_4}[N_{I_3}[N_{I_1}]] \to A^{v+3}$. This is because at $t = 4$, each $j \in N_{I_4}[N_{I_3}[N_{I_1}]]$ has at least $\lfloor k/2 \rfloor$ neighbours in $N \backslash \{I \cup N_{I_1}(s_1) \cup N_{I_2}(s_2) \cup N_{I_3}(s_3)\}$ play strategies in A^τ, $n_{I_3}(s_3) - n_{I_3}[N_{I_1}]$ neighbours in $N_{I_3}(s_3) \backslash N_{I_3}[N_{I_1}]$ play strategies in A^{v+3}, and the rest (i.e. neighbours of j that belong to $N_{I_3}[N_{I_1}] \cup N_{I_2}(s_2)$) play strategies in A^{v+2}. Thus, j has at least $\lfloor k/2 \rfloor$ neighbours play strategies in A^{v+3} and at most $\lceil k/2 \rceil$ play strategies in $\bar{A}^{v+2} \subset A^{v+2}$, which makes strategies in A^{v+3} best responses to j.

$n_{I_4}(s_4) - n_{I_4}[N_{I_3}[N_{I_1}]] \to A^{v+4}$, following the assumption that $n^*_{\tau(v+4)}(\sigma^q_{\tau(v+3)}) \leq n^+_{I_3}(s_3) - n^+_{I_3}[N_{I_1}]$. This is because, at $t = 4$, each $j \in N_{I_4}(s_4) \backslash N_{I_4}[N_{I_3}[N_{I_1}]]$ has at least $n^*_{\tau(v+4)}(\sigma^q_{\tau(v+3)})$ neighbours in $N_{I_3}(s_3) \backslash N_{I_3}[N_{I_1}]$ playing strategies in A^{v+3} and the rest play strategies in A^τ. This, by definition of $n^*_{\tau(v+4)}(\sigma^q_{\tau(v+3)})$, makes strategies in A^{v+4} best responses.

$n - (\mu + n_{I_1}(s_1) + n_{I_2}(s_2) + n_{I_3}(s_3) + n_{I_4}(s_4)) \to A^\tau$.

$t=6$	$\mu \to A^{v+1}$. $n_{I_1}(s_1) \to A^{v+1}$. $n_{I_2}(s_2) \to A^{v+2}$. $n_{I_3}[N_{I_1}] \to A^{v+2}$. $n_{I_3}(s_3) - n_{I_3}[N_{I_1}] \to A^{v+3}$. $n_{I_4}[N_{I_3}[N_{I_1}]] \to A^{v+3}$. $n_{I_4}(s_4) - n_{I_4}[N_{I_3}[N_{I_1}]] \to A^{v+4}$. $n_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]] \to A^{v+4}$, for $s_5 = n^*_{\tau(v+5)}(\sigma^q_{\tau(v+4)})$, and assuming $v+5 \leq \tau$ so that $A^\tau \subseteq A^{v+5}$. This is because at $t=5$, each $j \in N_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]]$ has at least $\lfloor k/2 \rfloor$ neighbours in $N \backslash \{I \cup N_{I_1}(s_1) \cup N_{I_2}(s_2) \cup N_{I_3}(s_3) \cup N_{I_4}(s_4)\}$ playing strategies in A^τ, $n_{I_4}(s_4) - n_{I_4}[N_{I_3}[N_{I_1}]]$ neighbours in $N_{I_4}(s_4) \backslash N_{I_4}[N_{I_3}[N_{I_1}]]$ play strategies in A^{v+4}, and the rest (i.e. neighbours of j that belong to $N_{I_4}[N_{I_3}[N_{I_1}]] \cup N_{I_3}(s_3) \backslash N_{I_3}[N_{I_1}]$) playing strategies in A^{v+3}. Thus, j has more than $\lfloor k/2 \rfloor$ neighbours play strategies in A^{v+4} (where $A^\tau \subseteq A^{v+4}$) and at most $\lceil k/2 \rceil$ play strategies in $\bar{A}^{v+3} \subset A^{v+3}$, which makes strategies in A^{v+4} best responses to j. $n_{I_5}(s_5) - n_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]] \to A^{v+5}$. This is because, at $t=5$, each $j \in N_{I_5}(s_5) \backslash N_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]]$ has at least $n^*_{\tau(v+5)}(\sigma^q_{\tau(v+4)})$ neighbours in $N_{I_4}(s_4) \backslash N_{I_4}[N_{I_3}[N_{I_1}]]$ playing strategies in A^{v+4} and the rest play strategies in A^τ. This, by definition of $n^*_{\tau(v+5)}(\sigma^q_{\tau(v+4)})$, makes strategies in A^{v+5} best responses. $n - (\mu + n_{I_1}(s_1) + n_{I_2}(s_2) + n_{I_3}(s_3) + n_{I_4}(s_4) + n_{I_5}(s_5)) \to A^\tau$.
$--$	$-----------------------------$
$t = \tau - v$	$\mu + n_{I_1}(s_1) \to A^{v+1}$. $n_{I_2}(s_2) \to A^{v+2}$. $n_{I_3}[N_{I_1}] \to A^{v+2}$. $n_{I_3}(s_3) - n_{I_3}[N_{I_1}] \to A^{v+3}$ $n_{I_4}[N_{I_3}[N_{I_1}]] \to A^{v+3}$ $n_{I_4}(s_4) - n_{I_4}[N_{I_3}[N_{I_1}]] \to A^{v+4}$. $n_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]] \to A^{v+4}$. $n_{I_5}(s_5) - n_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]] \to A^{v+5}$. $-------------$ $n_{I_{\tau-v-1}}[N_{I_{\tau-v-2}}[\cdots[N_{I_3}[N_{I_1}]]]] \to A^{\tau-2}$, for $s_{\tau-v-1} = n^*_{\tau(\tau-1)}(\sigma^q_{\tau(\tau-2)})$. $n_{I_{\tau-v-1}}(s_{\tau-v-1}) - n_{I_{\tau-v-1}}[N_{I_{\tau-v-2}}[\cdots[N_{I_3}[N_{I_1}]]]] \to A^{\tau-1}$. $n - \left(\mu + n_{I_1}(s_1) + n_{I_2}(s_2) + n_{I_3}(s_3) + \cdots + n_{I_{\tau-v-1}}(s_{\tau-v-1})\right) \to A^\tau$.

$$\begin{array}{c|l}
t \quad = & \mu + n_{I_1}(s_1) \to A^{v+1}. \\
\tau \quad - & n_{I_2}(s_2) \to A^{v+2}. \\
v+1 & n_{I_3}[N_{I_1}] \to A^{v+2}. \\
 & n_{I_3}(s_3) - n_{I_3}[N_{I_1}] \to A^{v+3} \\
 & n_{I_4}[N_{I_3}[N_{I_1}]] \to A^{v+3} \\
 & n_{I_4}(s_4) - n_{I_4}[N_{I_3}[N_{I_1}]] \to A^{v+4}. \\
 & n_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]] \to A^{v+4}. \\
 & n_{I_5}(s_5) - n_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]] \to A^{v+5}. \\
 & - - - - - - - - - - - - - - \\
 & n_{I_{\tau-v-1}}(s_{\tau-v-1}) - n_{I_{\tau-v-1}}[N_{I_{\tau-v-2}}[\cdots[N_{I_3}[N_{I_1}]]]] \to A^{\tau-1}. \\
 & n_{I_{\tau-v}}[N_{I_{\tau-v-1}}[\cdots[N_{I_3}[N_{I_1}]]]] \to A^{\tau-1}, \text{ for } s_{\tau-v} = n^*_{\tau(\tau)}(\sigma^q_{\tau(\tau-1)}). \\
 & n_{I_{\tau-v}}(s_{\tau-v}) - n_{I_{\tau-v}}[N_{I_{\tau-v-1}}[\cdots[N_{I_3}[N_{I_1}]]]] \to A^{\tau}. \\
 & n - \Big(\mu + n_{I_1}(s_1) + \cdots + n_{I_{\tau-v-1}}(s_{\tau-v-1}) + n_{I_{\tau-v}}(s_{\tau-v})\Big) \to A^{\tau}.
\end{array}$$

Let Z, ϕ and z be defined as follows.

$$Z = N \backslash \left\{ I \cup N_{I_1}(s_1) \cup \cdots \cup N_{I_{\tau-v-1}}(s_{\tau-v-1}) \cup N_{I_{\tau-v}}[N_{I_{\tau-v-1}}[\cdots[N_{I_3}[N_{I_1}]]]] \right\}$$

$$\phi = \left(\mu + \sum_{v=1}^{\tau-v-1} n_{I_v}(s_v) + n_{I_{\tau-v}}[N_{I_{\tau-v-1}}[\cdots[N_{I_3}[N_{I_1}]]]] \right)$$

$$z = n - \phi$$

We see from the above iterative process that when $z \geq \lceil pk \rceil + 1$, all players in Z play strategies in A^{τ} from $t = \tau - v + 1$ onward. This is because each has at least $\lceil pk \rceil$ neighbours play strategies in A^{τ} and the rest play strategies in $A^{\tau-1}$.

The following iterative process simultaneously unfolds at the background of the above evolutionary process from $t = 4$ onward. At $t = 4$, all $j \in N^1_{I_1}[N_{I_2}]$ play strategies in A^{v+2}; at $t = 5$, all $j \in N^1_{I_1}[N_{I_2}] \cup N^1_{I_1}[N_{I_2}]$ play strategies in A^{v+2}; and so on, until some $t = t_1$ when all players in $N_{I_1}(s_1)$ play strategies in A^{v+2}. Between $t = 4$ and $t = t_1$, players in $N_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$, for $3 \leq r \leq \tau - v - 1$, need not switch to strategies in A^{v+r} even if some have at least $\lceil pk \rceil$ neighbours in $N_{I_r}(s_r) \backslash N_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$ playing strategies in A^{v+r}. This is because each $j \in N_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$ has at least one neighbour in $N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]$ playing strategies in A^{v+r-2}.

From $t = t_1 + 1$ onward, players in I with at least $\lceil pk \rceil$ neighbours in $N_{I_1}(s_1)$ play strategies in A^{v+2}. At the same time, all players in $N_{I_3}[N_{I_1}]$ with at least $\lceil pk \rceil$ neighbours in $N_{I_3}(s_3) \backslash N_{I_3}[N_{I_1}] \cup N_{I_4}[N_{I_3}[N_{I_1}]]$ switch to strategies in A^{v+3}. This process continues until some $t = t_1 + t_2$ when all players in I play strategies in A^{v+2} and all players in $N_{I_3}[N_{I_1}]$ play strategies in A^{v+3}.

From $t = t_1 + t_2 + 1$ onward, all $j \in N_{I_2}(s_2)$ with at least $\lceil pk \rceil$ neighbours in $N_{I_3}(s_3)$ play strategies in A^{v+3}. At the same time, all players in $N_{I_4}[N_{I_3}[N_{I_1}]]$ with at least $\lceil pk \rceil$ neighbours in $N_{I_4}(s_4) \backslash N_{I_4}[N_{I_3}[N_{I_1}]] \cup N_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]]$ switch to strategies in A^{v+4}. This step-by-step evolution continues until some $t = t_1 + t_2 + t_3$ when all players in $N_{I_r}(s_r)$, for $1 \leq r \leq \tau - v - 2$, play strategies in $A^{\tau + r + 1}$, and players in $Z \cup N_{I_{\tau-v}}[N_{I_{\tau-v-1}}[\cdots[N_{I_3}[N_{I_1}]]]] \cup N_{I_{\tau-v-1}}(s_{\tau-v-1})$ all play strategies in A^τ. Eventually, at some $t^* > \tau - v + 1$, it will converge to an absorbing set of states containing only strategies in A^τ.

Thus, when $z \geq \lceil pk \rceil + 1$, at least $\mu + 1$ mutations to strategies in $\bar{A}^v$, for $0 \leq v \leq \tau - 1$, are needed to trigger an exit from $D(W)$ to any $W'' \subseteq \bigcup_{v=0}^{\tau-1} \Big(M(\bar{\mathbf{A}}^v) \cup Q(\bar{\mathbf{A}}^v) \cup L(\mathbf{A}^v) \Big)$. And following the same steps in Appendix Appendix C, when $n - \gamma\phi \geq \gamma(\lceil pk \rceil + 1)$, at least $\gamma(\mu + 1)$ mutations to strategies in $\bar{A}^v$ are necessary.

Given μ, $n - \gamma\phi$ is minimized when $v = 0$, $n_{I_r}(s_r) = 2\lceil k/2 \rceil - 2(s_r - 1)$, $n_{I_{\tau-v}}[N_{I_{\tau-v-1}}[\cdots[N_{I_3}[N_{I_1}]]]] = \sum_{r=2}^{\tau-v-1} \Big(2\lceil k/2 \rceil - n_{I_r}(s_r) \Big)$ (this follows from (10)), and when $s_r = \min_{v \in [0,\tau-1]} \min_{r \in [1,\tau-v]} s_r = n(A^\tau)$. This implies that at least $\gamma(\mu + 1)$ mutations are generally required to trigger an exit from $D(W)$ to any $W'' \subseteq \bigcup_{v=0}^{\tau-1} \Big(M(\bar{\mathbf{A}}^v) \cup Q(\bar{\mathbf{A}}^v) \cup L(\mathbf{A}^v) \Big)$ whenever

$$
\begin{aligned}
& n - \gamma \left(\mu + \sum_{r=1}^{\tau-1} n_{I_r}(s_r) + \sum_{r=2}^{\tau-1} \Big(2\lceil k/2 \rceil - n_{I_r}(s_r) \Big) \right) \geq \gamma(\lceil pk \rceil + 1) \\
& \Rightarrow n - \gamma\Big(\mu + 2\lceil k/2 \rceil (\tau - 2) + n_{I_1}(s_1) \Big) \geq \gamma(\lceil pk \rceil + 1) \\
& \Rightarrow n \geq \gamma\Big(\mu + \lceil pk \rceil + 3 - 2n(A^\tau) \Big) + 2\gamma\lceil k/2 \rceil (\tau - 1) \quad \text{Substituting for } n_{I_1}(s_1) \text{ and } s^*
\end{aligned}
$$

Thus, when $n^*_{\tau(v+r+1)}(\sigma^q_{\tau(v+r)}) \leq n^+_{I_r}(s_r) - n^+_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$ for all $r = 3, \cdots, \tau - 1$, and at least $\gamma(n(A^\tau) + 1)$ mutations are needed to trigger an exit from $D(W)$ to any $W'' \subseteq \bigcup_{v=0}^{\tau-1} \Big(M(\bar{\mathbf{A}}^v) \cup Q(\bar{\mathbf{A}}^v) \cup L(\mathbf{A}^v) \Big)$ whenever

$$
n \geq \gamma\Big(\lceil pk \rceil + 3 - n(A^\tau) \Big) + 2\gamma\lceil k/2 \rceil (\tau - 1) \tag{D.4}
$$

Since $n(A^\tau) \geq n^*(A)$, at least $\gamma(n^*(A) + 1)$ mutations are needed to trigger an exit from $D(W)$ to any $W'' \subseteq \bigcup_{v=0}^{\tau-1} \Big(M(\bar{\mathbf{A}}^v) \cup Q(\bar{\mathbf{A}}^v) \cup L(\mathbf{A}^v) \Big)$ whenever

$$
n \geq \gamma\Big(\lceil pk \rceil + 3 - n^*(A) \Big) + 2\gamma\lceil k/2 \rceil (\tau - 1) \tag{D.5}
$$

This implies that $C_D(W, W'') \geq \gamma(n^*(A) + 1)$ whenever (D.5) holds. Since $T \geq \tau$, it follows that when (C.3) holds, (D.5) also hold.

As discussed in Section Appendix C, when $4 \leq \tau \leq T$, we need to take into account scenarios where $n^*_{\tau(v+r+1)}(\sigma^q_{\tau(v+r)}) > n^+_{I_r}(s_r) - n^+_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$ for all $r = 3, 4, \cdots, \tau$. Following the same steps in the dynamic process leading up to condition (C.10), we find that after $t = \frac{3(\tau - v - 1)}{2}$ and $t = \frac{3(\tau - v - 1)}{2} + 1$ iterations, the strategy configurations are:

$t = \frac{3(\tau-v-1)}{2}$

$\mu \to A^{v+1}$.

$n_{I_1}(s_1) \to A^{v+1}$.

$n_{I_2}(s_2) \to A^{v+2}$.

$n_{I_3}[N_{I_1}] \to A^{v+2}$.

$n_{I_3}(s_3) - n_{I_3}[N_{I_1}] \to A^{v+3}$.

$n_{I_4}[N_{I_3}[N_{I_1}]] \to A^{v+3}$.

$n_{I_4}(s_4) \to A^{v+4}$.

$n_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]] \to A^{v+4}$.

$n_{I_5}(s_5) - n_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]] \to A^{v+5}$

$n_{I_6}[n_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]]] \to A^{v+5}$.

$n_{I_6}(s_6) \to A^{v+6}$

$- - - - - - - - - - - - - - -$

Assuming without loss of generality that $\tau - v - 1$ is even, then:

$n_{I_{\tau-v-2}}[N_{I_{\tau-v-3}}[\cdots[N_{I_3}[N_{I_1}]]]]] \to A^{\tau-3}$.

$n_{I_{\tau-v-2}}(s_{\tau-v-2}) - n_{I_{\tau-v-2}}[N_{I_{\tau-v-3}}[\cdots[N_{I_3}[N_{I_1}]]]]] \to A^{\tau-2}$.

$n_{I_{\tau-v-1}}[N_{I_{\tau-v-2}}[\cdots[N_{I_3}[N_{I_1}]]]]] \to A^{\tau-2}$

$n_{I_{\tau-v-1}}(s_{\tau-v-1}) \to A^{\tau-1}$.

$n - \Big(\mu + n_{I_1}(s_1) + \cdots + n_{I_{\tau-v-2}}(s_{\tau-v-2}) + n_{I_{\tau-v-1}}[N_{I_{\tau-v-2}}[\cdots[N_{I_3}[N_{I_1}]]]]] + n_{I_{\tau-v-1}}(s_{\tau-v-1})\Big) \to A^{\tau}$.

$$
\begin{array}{l|l}
t \quad = & \mu \to A^{v+1}. \\
\frac{3(\tau-v-1)}{2} & n_{I_1}(s_1) \to A^{v+1}. \\
+1 & n_{I_2}(s_2) \to A^{v+2}. \\
& n_{I_3}[N_{I_1}] \to A^{v+2}. \\
& n_{I_3}(s_3) - n_{I_3}[N_{I_1}] \to A^{v+3}. \\
& n_{I_4}[N_{I_3}[N_{I_1}]] \to A^{v+3}. \\
& n_{I_4}(s_4) \to A^{v+4}. \\
& n_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]] \to A^{v+4}. \\
& n_{I_5}(s_5) - n_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]] \to A^{v+5} \\
& n_{I_6}[n_{I_5}[N_{I_4}[N_{I_3}[N_{I_1}]]]] \to A^{v+5}. \\
& n_{I_6}(s_6) \to A^{v+6} \\
& -------------- \\
& n_{I_{\tau-v-1}}[N_{I_{\tau-v-2}}[\cdots[N_{I_3}[N_{I_1}]]]]] \to A^{\tau-2} \\
& n_{I_{\tau-v-1}}(s_{\tau-v-1}) \to A^{\tau-1}. \\
& n_{I_{\tau-v}}[N_{I_{\tau-v-1}}[\cdots[N_{I_3}[N_{I_1}]]]]] \to A^{\tau-1}. \\
& n_{I_{\tau-v}}(s_{\tau-v}) - n_{I_{\tau-v}}[N_{I_{\tau-v-1}}[\cdots[N_{I_3}[N_{I_1}]]]]] \to A^{\tau}. \\
& n - \Big(\mu + n_{I_1}(s_1) + \cdots + n_{I_{\tau-v-2}}(s_{\tau-v-2}) + n_{I_{\tau-v-1}}[N_{I_{\tau-v-2}}[\cdots[N_{I_3}[N_{I_1}]]]]] + \\
& n_{I_{\tau-v-1}}(s_{\tau-v-1}) + n_{I_{\tau-v}}(s_{\tau-v})\Big) \to A^{\tau}.
\end{array}
$$

Let ϕ, Z and z be defined as follows:

$$Z = N \backslash \Big\{ I \cup N_{I_1}(s_1) \cup \cdots \cup N_{I_{\tau-v-1}}(s_{\tau-v-1}) \cup N_{I_{\tau-v}}[N_{I_{\tau-v-1}}[\cdots[N_{I_3}[N_{I_1}]]]]] \Big\}$$

$$\phi = \mu + \sum_{r=1}^{\tau-v-1} n_{I_r}(s_r) + \sum_{r=2}^{\frac{\tau-v-1}{2}} n_{I_{2r}}[N_{I_{2r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]] + n_{I_{\tau-v}}[N_{I_{\tau-v-1}}[\cdots[N_{I_3}[N_{I_1}]]]]]$$

$$z = n - \phi$$

We see in the above iterative process that when $z \geq \lceil pk \rceil + 1$, all players in Z play strategies in A^{τ} from $t = \frac{3(\tau-v-1)}{2} + 1$ onward. This is because each has at least $\lceil pk \rceil$ neighbours play strategies in A^{τ} and the rest play strategies in $\bar{A}^{\tau-1}$. Following the same steps above and the steps leading up to condition (C.10), this iterative process eventually converges to an absorbing set of states containing only strategies in A^{τ}.

Since $n - \gamma\phi$ is minimized when $v = 0$; $n_{I_r}(s_r) = 2\lceil k/2 \rceil - 2(s_r - 1)$, $n_{I_{2r}}[N_{I_{2r-1}}[\cdots[N_{I_3}[N_{I_1}]]]] = 4\lceil k/2 \rceil - (n_{I_{2r-1}}(s_{2r-1}) + n_{I_{2r-2}}(s_{2r-2}))$, $n_{I_{\tau-v}}[N_{I_{\tau-v-1}}[\cdots[N_{I_3}[N_{I_1}]]]]] = 2\lceil k/2 \rceil - n_{I_{\tau-v-1}}(s_{\tau-v-1})$, and when $s_r = n(A^{\tau})$, it follows that when $n^{*}_{\tau(v+r+1)}(\sigma^{q}_{\tau(v+r)}) > n^{+}_{I_r}(s_r) - n^{+}_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]$ for all $r = 3, 4, \cdots, \tau$, at least $\gamma(\mu + 1)$ mutations are needed to trigger an exit from $D(W)$ to any $W'' \subseteq \bigcup_{v=0}^{\tau-1}\Big(M(\bar{\mathbf{A}}^v)\cup Q(\bar{\mathbf{A}}^v)\cup L(\mathbf{A}^v)\Big)$ whenever

$$n-\gamma\left(\mu+\sum_{r=1}^{\tau-1}n_{I_r}(s_r)+\sum_{r=2}^{\frac{\tau-1}{2}}\Big(4\lceil k/2\rceil-\Big(n_{I_{2r-1}}(s_{2r-1})+n_{I_{2r-2}}(s_{2r-2})\Big)\Big)+\Big(2\lceil k/2\rceil-n_{I_{\tau-1}}(s_{\tau-1})\Big)\right)$$

$$\geq\gamma\Big(\lceil pk\rceil+1\Big)$$

$$\Rightarrow n-\gamma\left(\mu+4\lceil k/2\rceil\left(\frac{\tau-1}{2}-1\right)+2\lceil k/2\rceil+\sum_{r=1}^{\tau-1}n_{I_r}(s_r)-\sum_{r=2}^{\tau-2}n_{I_r}(s_r)-n_{I_{\tau-1}}(s_{\tau-1})\right)\geq\gamma\Big(\lceil pk\rceil+1\Big)$$

$$\Rightarrow n-\gamma\Big(\mu+2\lceil k/2\rceil(\tau-2)+n_{I_1}(s_1)\Big)\geq\gamma\Big(\lceil pk\rceil+1\Big)$$

$$\Rightarrow n-\gamma\Big(\mu+2\lceil k/2\rceil(\tau-1)+2-2n(A^\tau)\Big)\geq\gamma\Big(\lceil pk\rceil+1\Big)\quad\text{Substituting for } n_{I_1}(s_1)\text{ and } s_r=n(A^\tau)$$

$$\Rightarrow n\geq\gamma\Big(\mu+\lceil pk\rceil+3-2n(A^\tau)\Big)+2\gamma\lceil k/2\rceil(\tau-1)$$

Thus, when $n^*_{T(\tau+r+1)}(\sigma^q_{T(\tau+r)}) > n_{I_r}(s_r)-n_{I_r}[N_{I_{r-1}}[\cdots[N_{I_3}[N_{I_1}]]]]]$ for all $r=3,\cdots,\tau-1$, at least $\gamma(n(A^\tau)+1)$ mutations are needed to trigger an exit from $D(W)$ to any $W''\subseteq \bigcup_{v=0}^{\tau-1}\Big(M(\bar{\mathbf{A}}^v)\cup Q(\bar{\mathbf{A}}^v)\cup L(\mathbf{A}^v)\Big)$ whenever

$$n\geq\gamma\Big(\lceil pk\rceil+3-n(A^\tau)\Big)+2\gamma\lceil k/2\rceil(\tau-1)\tag{D.6}$$

Since $n(A^\tau)\geq n^*(A)$, at least $\gamma(n^*(A)+1)$ mutations are needed to trigger an exit from $D(W)$ to any $W''\subseteq \bigcup_{v=0}^{\tau-1}\Big(M(\bar{\mathbf{A}}^v)\cup Q(\bar{\mathbf{A}}^v)\cup L(\mathbf{A}^v)\Big)$ whenever

$$n\geq\gamma\Big(\lceil pk\rceil+3-n^*(A)\Big)+2\gamma\lceil k/2\rceil(\tau-1)\tag{D.7}$$

This implies that $C_D(W,W'')\geq\gamma(n^*(A)+1)$ whenever (D.7) holds. Since $T\geq\tau$, it follows that when (C.10) holds, (D.7) also hold.

Now, let either (C.3) or (C.10) hold so that $C_D(W,W'')\geq\gamma(n^*(A)+1)$ for any $W\subseteq\bar{\mathbf{A}}^\tau$ and $W''\subseteq \bigcup_{v=0}^{\tau-1}\Big(M(\bar{\mathbf{A}}^v)\cup Q(\bar{\mathbf{A}}^v)\cup L(\mathbf{A}^v)\Big)$. If $\gamma(n^*(A)+1)>n(p,k)$, then condition (C.3) ensures that, for any $W\subseteq\bar{\mathbf{A}}^\tau$, there exists some $W'\subseteq\mathbf{A}^\tau$ for which $R(W)=C_D(W,W')\leq n(p,k)$. Thus, starting from any $\mathbf{x}\in D(W)$ with $W\subseteq\bar{\mathbf{A}}^\tau$, there exists a sequence of absorbing sets $W_1,W_2,\cdots,W_H$, with $W_1\subseteq\mathbf{A}^\tau$ and $W_H\subseteq\mathbf{A}^T$, along which $R(W)=C_D(W,W_1)$ and $R(W_h)=C_D(W_h,W_{h+1})$, for all $h=1,2,\cdots,H-1$. The respective cost of the minimum path in $\Gamma(\mathbf{x},\mathbf{A}^T)$ and coradius of $\mathbf{A}^T$ when (C.3) holds are then:

$$\begin{aligned}C^*(\mathbf{x},\mathbf{A}^T)&=\min_{(\mathbf{x};W,W_1,W_2,\cdots,W_H)\in\Gamma(\mathbf{x},\mathbf{A}^T)}C_D(W,W_1)+\sum_{h=1}^{H-1}\Big(C_D(W_h,W_{h+1})-R(W_h)\Big)\\&=\min_{W_1\subseteq\mathbf{A}^\tau}C_D(W,W_1)\end{aligned}$$

$$CR^*(\mathbf{A}^T)=\max_{\mathbf{x}\in\mathbf{X}\setminus D(\mathbf{A}^T)}C^*(\mathbf{x},\mathbf{A}^T)=\max_{\tau\in[0,T-1]}\max_{W\subseteq\bar{\mathbf{A}}^\tau}\min_{W_1\subseteq\mathbf{A}^\tau}C_D(W,W_1)\tag{D.8}$$

Since $C_D(W,W')\leq n(p,k)$ for all $W\subseteq\bar{\mathbf{A}}^\tau$ and $W_1\subseteq\mathbf{A}^\tau$, and for all $\tau\in[0,T-1]$, it follows that when (C.3) holds, $CR^*(\mathbf{A}^T)\leq n(p,k)$.

Acknowledgements

I would like to thank the seminar participants at the School of Economics and the African Institute of Financial Markets and Risk Management, University of Cape Town, and at the Department of Economics, Stellenbosch University.

References

Alós-Ferrer, C., Weidenholzer, S., 2007. Partial bandwagon effects and local interactions. Games and Economic Behavior 61, 179–197.

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

Azomahou, T.T., Opolot, D., 2018. Strategic diffusion in networks through contagion. Available at SSRN 3323076 .

Bilancini, E., Boncinelli, L., 2020. The evolution of conventions under condition-dependent mistakes. Economic Theory 69, 497–521.

Blume, L.E., 1995. The statistical mechanics of best-response strategy revision. Games and Economic Behavior 11, 111–145.

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

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

Foster, D., Young, P., 1990. Stochastic evolutionary game dynamics. Theoret. Population Biol .

Goyal, S., Vega-Redondo, F., 2005. Network formation and social coordination. Games and Economic Behavior 50, 178–207.

Jackson, M.O., Watts, A., 2002. On the formation of interaction networks in social coordination games. Games and Economic Behavior 41, 265–291.

Jiang, G., Weidenholzer, S., 2017. Local interactions under switching costs. Economic Theory 64, 571–588.

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

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

Lee, I.H., Valentinyi, A., 2000. Noisy contagion without mutation. Review of Economic Studies 67, 47–56.

Man, P.T., 2012. Efficiency and stochastic stability in normal form games. Games and Economic Behavior 76, 272–284.

Maruta, T., 1997. On the relationship between risk-dominance and stochastic stability. Games and Economic Behavior 19, 221 – 234.

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

Newton, J., 2012. Coalitional stochastic stability. Games and Economic Behavior 75, 842–854.

Opolot, D., 2018. On the relationship between p-dominance and stochastic stability in network games. Http://dx.doi.org/10.2139/ssrn.3234959.

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

Oyama, D., Takahashi, S., 2015. Contagion and uninvadability in local interaction games: The bilingual game and general supermodular games. Journal of Economic Theory 157, 100–127.

Peski, M., et al., 2010. Generalized risk-dominance and asymmetric dynamics. Journal of Economic Theory 145, 216.

Robson, A.J., Vega-Redondo, F., 1996. Efficient Equilibrium Selection in Evolutionary Games with Random Matching. Journal of Economic Theory 70, 65–92.

Staudigl, M., Weidenholzer, S., 2014. Constrained interactions and social coordination. Journal of Economic Theory 152, 41–63.

Tercieux, O., 2006. p-best response set. Journal of Economic Theory 131, 45–70.

Weidenholzer, S., 2012. Long-run equilibria, dominated strategies, and local interactions. Games and Economic Behavior 75, 1014–1024.

Young, H.P., 1993. The evolution of conventions. Econometrica 61, 57–84.

Young, H.P., 1998. Individual Strategy and Social Structure: An Evolutionary Theory of Institutions. Princeton University Press.