Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Distributed Computing

Approximate Clone Robustness in Games via Graphs

Master's Thesis

Supanat Wangsutthitham

swangsutthit@ethz.ch

Distributed Computing Group
Computer Engineering and Networks Laboratory
ETH Zürich

Supervisors:
Damien Berriaud
Prof. Dr. Roger Wattenhofer

September 17, 2026

Thank you for everything, Nello.

Acknowledgements

I would like to thank my direct advisor, Damien Berriaud, for all his helpful guidance, generous time, and unparalleled flexbility throughout the duration of the thesis. His advice has been instrumental in writing the thesis and in helping me develop as a researcher.

I would also like to thank Professor Wattenhofer for the opportunity to work with the DISCO group and his insights that helped shape the thesis.

I am grateful to my friends, especially to Leo for entertaining my daily quirky conversations, to Bhudhi for constantly helping me out with technical discussions, and to Dan for being there and giving me emotional support whenever I need.

Finally, I would like to express my deepest gratitude to my family for their unconditional love and encouragement. Their unwavering support, both financial and emotional, has been the main driving force from the very beginning through to the end.

As an international student in a foreign country, the beginning of this journey was, frankly speaking, rough—so much so I could never have imagined how much it would come to mean to me. However, having met so many wonderful people along the way, many of whom have become my very good friends, I have come to view it as one of the most cherished chapters of my life. To everyone who contributed, directly or indirectly, to this journey, I will carry these memories with me. I sincerely thank you.

Contents

Acknowledgements ii

1 Introduction 1

2 Clone-robust Graph Weighting Functions 3

2.1 Preliminaries . . . 3

2.2 Vertex Sharing . . . 4

2.3 Edge Sharing . . . 5

2.4 Clone Sharing . . . 7

2.4.1 Constructing Interpretable Clone-robust Graph Weighting Functions . . . 13

3 Clone-robust Games 21

3.1 Preliminaries . . . 21

3.2 Clone-robust Games . . . 22

Chapter 1

Introduction

Designing mechanisms whose global behavior depends on the intrinsic information, rather than the redundancy of its representation, is one of the foundational objectives of theoretical computer science and game theory. When a model represents distinct elements—such as vertices in an information network or strategies in an interactive game—it routinely operates under the implicit assumption that distinct labels correspond to fundamentally distinct alternatives.

In realistic and complex environments, this assumption collapses. Systems frequently exhibit redundancy, where multiple entities are not genuinely independent, but rather functional replicas of one another. Worse, this vulnerability is exacerbated in the adversarial settings, where strategic actors may deliberately inject duplicative entities into the system to manipulate the outcome in their favor.

While the concept of clone invariance has been studied across social choice, network ranking, and classical game theory, these works focus almost exclusively on the idealized setting of exact clones [Tid87; Bal+18]. In these classical formulations, duplicate entities are mathematically indistinguishable, allowing them to be trivially factored out via discrete equivalence classes or symmetry reductions. However, this discrete approach offers no guarantees when redundancy is approximate, where entities deviate by some positive metric error.

In such settings, standard discrete reductions collapse, often leading to severe discontinuities or dynamical instabilities. One notable exception is [PSZ25]: it introduced approximate clone robustness by assigning weights proportionally to the size of their Voronoi regions. This approach however fails upon systems that have dense clusters. The earliest work to develop a formal framework of weights assignment under approximate cloning, to our best knowledge, is by [BW26b]. It highlighted the limitations of the Voronoi-based approach and presented clone-robust weighting functions on Euclidean spaces. [BW26a] built on the work of [BW26b], where it generalized Euclidean spaces to any metric spaces via the discrete graph spaces.

Related Work. The formal study of clone-robustness originated in social choice theory as a desired property for electoral systems. [Tid87] introduced the *Independence of Clones* criterion, positing that the outcome of an election should not change when a candidate is replaced by a set of alternatives that are perceived as equal by voters. In modern computational social choice, this criterion has been extensively explored: [EFS14] investigated the algorithmic complexity of candidate cloning to manipulate election outcomes, while comprehensive syntheses in [16] frame clone-invariance as an essential condition for robust preference aggregation.

Concerns regarding identity duplication also emerged in network analysis and distributed systems. In their seminal work on identity manipulation, [Dou02] formalized the Sybil attack, where an adversary artificially manufactures entities to usurp control over decentralized networks. Within axiomatic graph ranking and search algorithms, [AT08] formalized how centrality and ranking functions (such as *PageRank*) react to cloned vertices, while [CF06] investigated Sybil-proof graph scoring rules.

According to the available literature, the first explicit effort to study clone-robustness beyond per-

fect clones was introduced in the context of clone-robust AI alignment [PSZ25]. This work recognized that when pooling preferences, evaluations, or data from generative models, agents often produce near-duplicate outputs rather than exact replicas. However, this early formulation exhibited notable limitations as previously mentioned: it lacked a rigorous metric-space formalization, was vulnerable to metric boundary instabilities, and did not establish axiomatic guarantees regarding continuous mass conservation.

A major theoretical advance was achieved by [BW26b], which established the first formal axiomatic framework for clone-robust weighting in metric spaces. Designed primarily to combat redundancy bias in benchmark aggregation and multi-criteria evaluation, it introduced formal continuity and mass-preservation axioms for continuous metric point clouds. Nevertheless, the feasibility of its mechanism remained constrained: its closed-form weighting formulas relied on restrictive geometric assumptions—only on Euclidean spaces—and did not generalize to topologically equivalent structures.

To overcome these geometric limitations, [BW26a] introduced a graph-based framework for arbitrary metric spaces. By mapping metric configurations into parameterized radius graphs, they demonstrated that clone-robust weighting functions could be computed across arbitrary metric topologies without geometric embedding constraints.

Contribution. The first half (chapter 2) of the thesis directly continues the work of [BW26a]; it explores various methods of sharing weights on graphs, where the main emphasis is on *clone-sharing*, i.e., how the weights of the vertices change upon adding a clone in a graph. Furthermore, it proposes interpretable properties on clone-sharing and presents a construction that satisfies them.

The second half (chapter 3) of the thesis introduces a groundwork for clone-robust games, where it first extends the properties obtained from the construction of clone-robust weighting function in [BW26a]. It then proposes a construction of update rules for games, which under a certain condition admits a unique fixed point. Finally, the thesis shows the construction admits a hyperrectangle that contains new every fixed point upon addition of clones, through the improved properties.

Chapter 2

Clone-robust Graph Weighting Functions

2.1 Preliminaries

From here, we will refer to graphs as undirected unweighted finite graphs. Now, we begin by defining common terminologies we will be using throughout the section.

Definition 2.1 (graph weighting function)**.** A graph weighting function w associates each finite graph G with a probability distribution over the vertices of G, i.e.

$$w : (V, E) \in \mathcal{G} \to p_V \in \Delta(V)$$

where $\mathcal{G}$ denotes the set of all finite graphs and $\Delta(S) \coloneqq \left\{p_S : S \to [0,1] \middle| \sum_{x \in S} p_S(x) = 1\right\}$ denotes the simplex over the elements of V.

As clones are the main topic of the thesis, we must first define what it means to be a *clone*. Here, we employ the notion of *closed-neighborhood*. Formally, we denote $N_G[v] \coloneqq \{v\} \cup \{u \in V(G) : uv \in E(G)\}$ a closed-neighborhood of a vertex v in graph G, i.e., a union of a vertex and its neighborhood. This naturally gives rise to an equivalence relation $\equiv_G$ between vertices in graph G where two vertices are in the same equivalence class if and only if they have the same closed-neighborhood, i.e., $u \equiv_G v$ if and only if $N_G[u] = N_G[v]$. We also say interchangably that u is a clone of v in this case and denote $[v]_G \coloneqq \{u \in V(G) : u \equiv_G v\}$ the equivalence class containing v in G. Moreover, we extend the notion of closed-neighborhood to *k-closed-neighborhood* $N_G^k[v]$, defined as $N_G^k[v] \coloneqq \{u \in V(G) : d_G(u, v) \leq k\}$.

With the definition of clones, we begin describing graphs obtained by cloning the vertices.

Definition 2.2. Let G be a graph. For $i \geq 1$, denote the graph obtained by adding i clones of x to G as

$$G_{x^i} \coloneqq G \cup \left\{x^1, x^2, \ldots, x^i\right\}$$

where x^j is the j-th clone of x. We let $G_{x^0} = G$ for consistency.

Now, we define properties that we will use to capture the notion of *clone-robustness*:

Definition 2.3 (symmetry)**.** Let w be a graph weighting function. We say that w is symmetric if for any finite graph G, G', vertex $v \in V(G)$, and graph isomorphism $\phi : V(G) \to V(G')$,

$$w(G)(v) = w(G')(\phi(v)).$$

Definition 2.4 (k-locality)**.** Let w be a graph weighting function and $k \geq 0$. We say that w is k-local if for any finite graph G, vertex $u, v \in V(G)$,

$$d_G(u, v) > k \implies w(G)(v) = w(G_{u^1})(v).$$

We say that a graph weighting function w is *k-clone-robust* if it is both symmetric and k-local.

2.2 Vertex Sharing

We begin by defining the following definitions to try to capture the notion of vertex sharing:

Definition 2.5. A graph weighting function w satisfies **vertex-sharing expansion** if for each graph G and vertex $x \in V(G)$, there exists an expansive constant $\eta_{G,x} \geq 1$ such that for any distinct vertex $y \neq x$,

$$w(G\backslash\{x\})(y) \geq \eta_{G,x} \cdot w(G)(y),$$

where we get equality if $y \notin N_G[x]$.

Note that since $\eta_{G,x} \geq 1$, the following theorem shows that no k-clone-robust graph weighting function can simultaneously satisfy vertex-sharing multiplicative rescaling for any $k \geq 1$.

Theorem 2.6. *For every k-clone-robust graph weighting function w, there exists a graph G and distinct vertices $x, y \in V(G)$ such that $w(G \setminus \{x\})(y) < w(G)(y)$.*

Proof. Let w be a k-clone-robust graph weighting function. Consider the following graph $G = P_3 \cup \bar{K_3}$:

Suppose for contradiction that for every distinct vertices $x, y \in V(G)$, $w(G \setminus \{x\})(y) \geq w(G)(y)$. Since w is symmetric, $w(G)(4) = w(G)(5) = w(G)(6)$ and $w(G)(2) = w(G)(3)$. Denote by $x = w(G)(4)$ the weight of the isolated vertices, $y = w(G)(2)$ the weight of the leaf vertices in P_3, and $z = w(G)(1)$ the weight of vertex 1. Since weights must sum up to 1, we know that

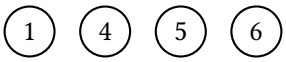
$$3x + 2y + z = 1.$$

Let $G' = K_2 \cup \bar{K_3}$ be the graph obtained by removing a leaf vertex in P_3:

We call the new weights $x' = w(G')(4)$ and $y' = w(G')(1)$ for the vertices in $\bar{K_3}$ and K_2 respectively. Again, as w is a probability distribution over the vertices we have

$$3x' + 2y' = 1.$$

Then, we further delete a vertex in K_2, resulting in the graph $G'' = \bar{K_4}$:



Here, all the vertices have the same weight under w by symmetry, which we know is equal to $\frac{1}{4}$. Furthermore, locality applies to the weights of the isolated vertices as vertex 1 and 2 were clones. This gives us $x' = \frac{1}{4}$, so we also get that $y' = \frac{1}{8}$. By the assumption, we know that

$$y, z \leq \frac{1}{8}.$$

Now, we consider another graph $G^\circ = \bar{K_5}$ obtained by deleting vertex 1 from G:

Similarly to G'', we get that there is only one weight here by symmetry, which is $\frac{1}{5}$. By the assumption, we have that

$$x \leq \frac{1}{5}.$$

Summing these inequalities together gives

$$1 = 3x + 2y + z \leq \frac{3}{5} + \frac{2}{8} + \frac{1}{8} = \frac{39}{40},$$

which is a contradiction. □

2.3 Edge Sharing

Similarly to how we wanted to capture the notion of vertex sharing, we begin by defining the following definitions to try to capture the notion of sharing weight via edges:

Definition 2.7. A graph weighting function w satisfies **edge-sharing endpoint-monotonicity** if for each graph G and edge $e = xy \in E(G)$,

$$w(G - e)(x) \geq w(G)(x).$$

Definition 2.8. A graph weighting function w satisfies **edge-sharing bystander-monotonicity** if for each graph G, edge $e = xy \in E(G)$, and vertex $z \in V(G)$ such that $z \notin e$,

$$w(G - e)(z) \leq w(G)(z).$$

The idea is that, an edge forms when two endpoints are relatively close to each other in the original metric space. This means they should share weights, so upon the removal of the edge, the endpoints are no longer constrainted by this sharing factor. Hence, they should receieve more weight upon the removal.

Now, the goal is to find either a counterexample to show that clone-robustness and the definitions are incompatible or to find a closed formula for a graph weighting function that satisfies all the above. This is proved to be elusive; we tried using linear programming to find the former, which upon running up to graphs of $n = 7$ does not give any infeasibility. Given this information, we strengthened our definitions into the following:

Definition 2.9. A graph weighting function w satisfies **edge-sharing endpoint-expansion** if for every graph G and edge $e = xy \in E(G)$, there exists an expanding constant $\mu_{G,e} \geq 1$ such that

$$w(G - e)(x) = \mu_{G,e} w(G)(x).$$

Definition 2.10. A graph weighting function w satisfies **edge-sharing bystander-contraction** if for every graph G and edge $e = xy \in E(G)$, there exists a contracting constant $\eta_{G,e} \in (0, 1]$ such that for every non-incident vertex $z \notin e$,

$$w(G - e)(z) \leq \eta_{G,e} w(G)(z),$$

with equality holding for all $z \notin N_G[x] \cup N_G[y]$.

Observe that theorem 2.9 and theorem 2.10 imply theorem 2.7 and theorem 2.8 respectively as $\mu \geq 1$ and $\eta \leq 1$. Here, we present the following theorem to show incompatability between the strengthened edge-sharing properties and clone-robustness.

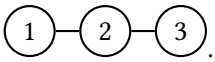
Theorem 2.11. *There exists no* 1*-clone-robust graph weighting function that simultaneously satisfies endpoint expansion and bystander contraction.*

Proof. Suppose for contradiction that w is a 1-clone-robust graph weighting function that satisfies both endpoint expansion and bystander contraction. Observe that for any graph G that is a union of m complete graphs, i.e.,

$$G = \bigcup_{i=1}^{m} K_{n_i}, \tag{2.1}$$

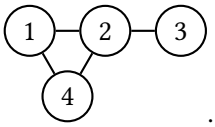
each vertex in K_{n_i} gets $\frac{1}{m \cdot n_i}$ weight.

We will construct a sequence of graphs that will eventually lead to contradiction: consider the graph P_3:



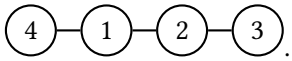
If we remove the edge $(2,3)$, then we have $K_2 \cup K_1$. By the observation eq. (2.1) and endpoint expansion, we have $\frac{1}{4} = \mu w(P_3)(2)$ and $\frac{1}{2} = \mu w(P_3)(3)$. This gives $2w(P_3)(2) = w(P_3)(3)$. By simplexity, $w(P_3)(2) = \frac{1}{5}$ and $w(P_3)(1/3) = \frac{2}{5}$.

We consider the paw graph G_p, which is obtained from cloning vertex 1 in P_3:



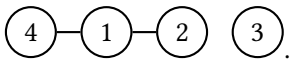
By locality, $w(G_p)(3) = \frac{2}{5}$. Then, consider the graph $K_3 \cup K_1$ obtained from removing the edge $(2,3)$. By endpoint expansion, $\frac{1}{2} = \mu\frac{2}{5}$, so $\mu = \frac{5}{4}$. Similarly, $\frac{1}{6} = \mu w(G_p)(2)$, so $w(G_p)(2) = \frac{2}{15}$. By simplexity and symmetry, $w(G_p)(1/4) = \frac{7}{30}$.

From here, remove the edge $(2,4)$ from G_p to get P_4:



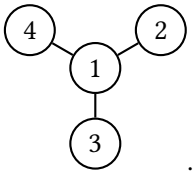
By endpoint expansion, $w(P_4)(4) = \mu\frac{7}{30}$ and $w(P_4)(2) = \mu\frac{2}{15}$. By symmetry and simplexity, we have $1 = 2w(P_4)(4) + 2w(P_4)(2)$. Solving for μ gives $\mu = \frac{15}{11}$. This implies $w(P_4)(3/4) = \frac{7}{22}$ and $w(P_4)(1/2) = \frac{2}{11}$.

Now, we remove the edge $(2,3)$ to get the graph $G_0 \coloneqq P_3 \cup K_1$:



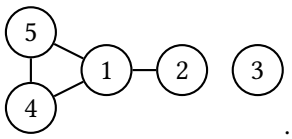
By endpoint expansion, $w(G_0)(3) = \mu\frac{7}{22}$ and $w(G_0)(2) = \mu\frac{2}{11}$. Hence, we can rewrite the weight of the leaf vertex of P_3 as the isolated vertex: $w(G_0)(2) = \frac{4}{7}w(G_0)(3)$. Further remove edge $\mu_2 \coloneqq (1,2)$ to get the graph $K_2 \cup 2K_1$. Again, by the observation in eq. (2.1) and endpoint expansion, we get $\frac{1}{6} = \mu_2 w(G_0)(1)$ and $\frac{1}{3} = \mu_2 w(G_0)(2)$. Rearranging allows us to write $w(G_0)(1) = w(G_0)(2)/2$. By simplexity, $1 = 2w(G_0)(2) + w(G_0)(3) + w(G_0)(1) = \frac{17}{7}w(G_0)(3)$. This means $w(G_0)(3) = \frac{7}{17}$, and consequently $w(G_0)(2/4) = \frac{4}{17}$ and $w(G_0)(1) = \frac{2}{17}$.

Now we consider the star graph $S_3 = K_1 \cup K_3$, which is obtained by drawing the edge $(1,3)$ in G_0:



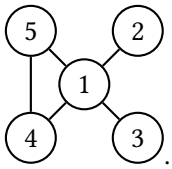
First, we write the weight of the leaf in terms of the hub. By endpoint expansion, $\mu w(S_3)(3) = \frac{7}{17}$ and $\mu w(S_3)(1) = \frac{2}{17}$. Thus, $w(S_3)(3) = \frac{7}{2}w(S_3)(1)$. By simplexity and symmetry, we get $1 = 3w(S_3)(3) + w(S_3)(1)$. Substituting in the leaf weights gives $w(S_3)(1) = \frac{2}{23}$ and consequently $w(S_3)(2/3/4) = \frac{7}{23}$.

We then consider the graph G_1 obtained by cloning vertex 4 in G_0:



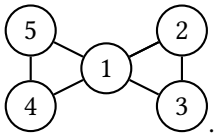
By locality, $w(G_1)(2) = \frac{4}{17}$ and $w(G_1)(3) = \frac{7}{17}$. Upon removing the edge $(1,2)$, the observation eq. (2.1) gives $\frac{1}{9} = \mu w(G_1)(1)$ and $\frac{1}{3} = \mu w(G_1)(2) = \mu\frac{4}{17}$. We solve for μ, giving $\mu = \frac{17}{12}$ and $w(G_1)(1) = \frac{4}{51}$. By simplexity and symmetry, we can conclude that $w(G_1)(4/5) = \frac{7}{51}$.

We clone vertex 4 of the star graph S_3, getting the cricket graph G_c:



By locality and symmetry, we know $w(G_c)(2/3) = \frac{7}{23}$. Also, observe that this graph can be obtained by adding the edge $(1,3)$ in the graph G_1. By endpoint expansion we have $\frac{7}{17} = \mu\frac{7}{23}$ and $\frac{4}{51} = \mu w(G_c)(1)$. Solving for μ gives $\mu = \frac{23}{17}$, so we know $w(G_c)(1) = \frac{4}{69}$. At last, by simplexity and symmetry, we solve for $w(G_c)(4/5)$, giving $\frac{1}{6}$.

Now we consider the final graph of the sequence: the bowtie graph G_b obtained by cloning vertex 3 in graph G_p (we re-index the vertices to make it consistent with G_c):



By locality and symmetry, the leaf vertices all have weight $w(G_b)(2/3/4/5) = \frac{7}{30}$. By simplexity, the hub has weight $w(G_b)(1) = \frac{1}{15}$. On the other hand, observe that we can obtain G_b from adding the edge $(2,3)$ in G_c. By bystander contraction, we get $\frac{1}{6} = \eta\frac{7}{30}$ as 4 and 5 are not in the neighborhood of 2 and 3. Solving for η gives $\eta = \frac{5}{7}$, which by bystander contraction on vertex 1, we have

$$\frac{4}{69} \leq \eta\frac{1}{15} = \frac{5}{7}\cdot\frac{1}{15} = \frac{1}{21},$$

which is a contradiction. □

2.4 Clone Sharing

Since the sharing approaches via vertices and edges did not work out, we look to another approach that is inherenently compatible with locality: addition of clones. We start by defining a quantity that represents clone-sharing:

Definition 2.12. Let w be a graph weighting function, G be a graph, $x, y \in V(G)$ be vertices, and $i \geq 1$ be an integer. Define **clone-sharing of y to x** as

$$\chi_{y\to x}(w)(G_{x^i}) \coloneqq w(G_{x^i})(y) - w(G_{x^{i+1}})(y)$$

to represent how much weight y is giving to the newly added clone x^{i+1}.

Observe that by definition, we have

$$\sum_{y\in V(G_{x^i})} \chi_{y\to x}(w)(G_{x^i}) = w(G_{x^{i+1}})(x^{i+1}) = w(G_{x^{i+1}})(x).$$

With this quantity now defined, we define some properties to give some structure to it:

Definition 2.13. A graph weighting function w satisfies **clone-sharing non-negativity** if for every finite graph G, vertex $x, y \in V(G)$,

$$\chi_{y\to x}(w)(G) \geq 0.$$

We say w satisfies ***k*-clone-sharing positivity** if the inequality is strict whenever $d_G(x, y) \leq k$.

Definition 2.14. A graph weighting function w satisfies **clone-sharing dilution** if for every finite graph G, vertex $x, y \in V(G)$, integers $0 \leq i < j$,

$$\chi_{y\to x}(w)(G_{x^i}) \geq \chi_{y\to x}(w)(G_{x^j}).$$

We say w satisfies ***k*-strong clone-sharing dilution** if the inequality is strict whenver $d_G(x, y) \leq k$.

We observe that there are direct connections between *clone-sharing dilution* and *clone-sharing non-negativity* (and the strong variants):

Lemma 2.15. *Clone-sharing dilution (resp. k-strong dilution) implies clone-sharing non-negativity (resp. k-positivity). However, the converse does not hold, i.e., clone-sharing positivity does not imply clone-sharing dilution.*

Proof. Let w be a graph weighting function. Assume that w satisfies *clone-sharing dilution*. Let G be a finite graph, $x \in V(G)$ and $y \in N[x]$. Suppose for contradiction that

$$\chi_{y\to x}(w)(G) = -c$$

for some $c > 0$. By dilution, for any $i \geq 0$, we get

$$\chi_{y\to x}(w)(G_{x^i}) \leq -c.$$

Denote $w_i \coloneqq w(G_{x^i})(y)$. By unpacking the definition of clone sharing and rearranging the inequalities above, we have

$$w_{i+1} - w_i \geq c \quad \forall i \geq 0. \tag{2.2}$$

Then, consider the following partial sum of clone sharing of y to x:

$$\sum_{i=0}^{N-1} \chi_{y\to x}(w)(G_{x^n}) = (w_0 - w_1) + (w_1 - w_2) + \ldots (w_{N-1} - w_N) = w_0 - w_N.$$

We evaluate w_N, the weight of y after adding N clones of x, by rearranging the above sum:

$$w_N = w_0 - \sum_{i=0}^{N-1} \chi_{y\to x}(w)(G_{x^i}) = w_0 + \sum_{i=0}^{N-1} (w_{i+1} - w_i).$$

Substituting the inequality (2.2) into the inside of the summation gives

$$w_N \geq w_0 + \sum_{i=0}^{N-1} c = w_0 + N \cdot c. \tag{2.3}$$

As N gets sufficiently large, $w_N > 1$, which is a contradiction as each weight must be between 0 and 1 (inclusive). Thus, we are done.

The proof for the strong variant is almost exactly the same except we instead suppose that

$$\chi_{y\to x}(w)(G) \leq 0.$$

In particular, we do not get a negative number from the first iteration. Thus, we have to apply strong dilution once, giving some $c > 0$ such that

$$\chi_{y\to x}(w)(G_{x^1}) = -c.$$

As a consequence, the inequalty (2.2) holds for $i > 0$ instead and the inequality (2.3) becomes

$$w_N \geq w_0 + 0 + \sum_{i=1}^{N-1} c = w_0 + (N-1)c,$$

i.e., we take the first term in the sum out. Still, with sufficiently large N, we get the same contradiction.

To show that the converse does not hold, define unnormalized weight for any finite graph G as

$$m_G(v) \coloneqq 1 + \frac{1}{|[v]_G|^2}.$$

Define the total normalizing mass of the graph as

$$M(G) \coloneqq \sum_{v\in V(G)} m(v).$$

Now, define the weighting function w as

$$w(G)(v) \coloneqq \frac{m(v)}{M(G)} = \frac{1 + 1/|[v]_G|^2}{M(G)}.$$

We verify that w is a graph weighting function, i.e., the weights sum up to 1 and that each weight is between 0 and 1 (inclusive):

$$\sum_{v\in V(G)} w(G)(v) = \sum_{v\in V(G)} \frac{m_G(v)}{M(G)} = \frac{1}{M(G)} \sum_{v\in V(G)} m_G(v) = \frac{M(G)}{M(G)} = 1$$

$$w(G)(v) = \frac{m_G(v)}{M(G)} = \frac{m_G(v)}{m_G(v) + \sum_{u\in V(G);u\neq v} m_G(u)} \in (0,1] \subset [0,1],$$

where the second line holds because the numerator is strictly positive.
Now, observe that one can rewrite $M(G)$ in terms of equivalence classes as follows:

$$M(G) = \sum_{v\in V(G)} m(v) = \sum_{C\in G/\equiv} C + \frac{1}{C}.$$

In particular, the function $1 + \frac{1}{x}$ is a strictly increasing function for $x \geq 1$, so adding a clone of any class in G would increase the value of M, i.e. $M(G) < M(G \cup \{x'\})$ for any clone $x' \notin G$.

Let us consider a graph G and vertices $x, v \in V(G)$. If we clone x, the denominator of $w(G)(v)$ strictly increases, whereas the numerator decreases if v belongs in the same class as x and otherwise remains the same. Thus, $\chi_{v\to x}(w)(G) > 0$, implying w satisfies k-clone-sharing positivity for any $k \in \mathbb{N}$.

Now, consider the graph $H \coloneqq P_3$, which we denote the vertices as $0{-}1{-}2$. Let $x = 1$ and $v = 0$. Now, we add a clone of x twice and calculate the weight of v in each graph: $w(H)(v) = \frac{1}{3}$; $w(H_{x^1})(v) = \frac{4}{13}$; and $w(H_{x^2})(v) = \frac{3}{11}$. Consequently,

$$\chi_{v\to x}(w)(H_{x^1}) - \chi_{v\to x}(w)(H_{x^2}) = \left(\frac{1}{3} - \frac{4}{13}\right) - \left(\frac{4}{13} - \frac{3}{11}\right) = \frac{1}{39} - \frac{5}{143} < 0,$$

so w does not satisfy clone-sharing dilution. □

We can apply the lemma to show a graph weighting function does not satisfy clone-sharing dilution by showing it does not satisfy clone-sharing non-negativity instead since the latter is generally easier to show (and respectively for the strong version).

Now, we introduce other one last property that we considered.

Definition 2.16. A clone-robust graph weighting function w satisfies ***k*-clone-sharing domination** if for every graph G, vertex $x, y, z \in V(G)$ such that $d(x,y), d(x,z) \leq k$, $N_G^k[x] \cap N_G^k[y] \supseteq N_G^k[x] \cap N_G^k[z]$ implies that

$$\chi_{y\to x}(w)(G) \geq \chi_{z\to x}(w)(G).$$

We say w satisfies ***k*-strong clone-sharing domination** when we have equality if and only if the two intersections are equal.

Strong clone-sharing domination, as the name suggests, is rather restrictive. Namely, it is in fact not compatible with locality:

Lemma 2.17. *k-strong clone-sharing domination is not compatible with $(2k-1)$-locality.*

Proof. Let w be a graph weighting function. Suppose for contradiction that w satisfies k-strong clone-sharing domination and $(2k-1)$-locality. Consider the graph $G \coloneqq P_{2k+2}$ whose vertices are labelled $0, 1, \ldots, 2k+1$. Let x be the vertex 0, u be the vertex $2k$, and v be the vertex $2k+1$. Observe that $d_G(x,v) = 2k$ and $d_G(x,u) = 2k+1$, so by $(2k-1)$-locality we have

$$\chi_{u\to x}(w)(G) = 0 = \chi_{v\to x}(w)(G).$$

Now, observe their intersections: $N_G^k[x] \cap N_G^k[u] = \{k\}$ and $N_G^k[x] \cap N_G^k[v] = \varnothing$. By k-strong clone-sharing domination, we get

$$0 = \chi_{u\to x}(w)(G) > \chi_{v\to x}(w)(G) = 0,$$

which is a contradiction. □

Remark 2.18. Since k-locality implies $k+1$-locality, by the contraposition k-strong clone-sharing domination is not compatible with j-locality for all $j < 2k$.

Finally, we show that clone-sharing dilution and domination are independent from each other in the following lemma.

Lemma 2.19. *Clone-sharing dilution and k-clone-sharing domination do not imply each other.*

Proof. We first prove the forward direction. Let $\mathcal{F} \coloneqq \{K_a \cup K_b : a \geq 1, b \geq 1\}$, i.e., set of union of two disconnected complete graphs. Observe that $\mathcal{F}$ is closed under cloning. Define the positive sequence $H : \mathbb{Z}^+ \to \mathbb{R}_{>0}$ by $H(c) \coloneqq c+1$ for all $c \geq 1$. Now, define graph weighting function w as the following:

- If $G \in \mathcal{F}$ (with vertex sets V_a and V_b):

$$w(G)(v) \coloneqq \begin{cases} \frac{H(a)}{a(H(a)+H(b))} = \frac{1+1/a}{a+b+2}, & v \in V_a \\ \frac{H(b)}{b(H(a)+H(b))} = \frac{1+1/b}{a+b+2}, & v \in V_b \end{cases}.$$

  Note that the weight assignment depends only on which set v is in.

- If $G \notin \mathcal{F}$:

$$w(G)(v) \coloneqq \frac{1}{n} \quad \forall v \in V(G),$$

  i.e., w is a uniform graph weighting function.

We first show that w is a graph weighting function on $G \in \mathcal{F}$. By construction, w is positive, and

$$\sum_{v \in V(G)} w(G)(v) = a \cdot \frac{H(a)}{a(H(a)+H(b))} + b \cdot \frac{H(b)}{b(H(a)+H(b))} = 1.$$

Thus, w is a graph weighting function. Now, we show that w satisfies dilution. Let G be a graph, $x, v \in V(G)$ be vertices of G, and $i \geq 0$ be some integers. By transitivity of $\geq$, it suffices to show $\chi_{v \to x}(G_{x^i}) \geq \chi_{v \to x}(G_{x^{i+1}})$. This is also equivalent to $w(G_{x^i})(v) - 2w(G_{x^{i+1}})(v) + w(G_{x^{i+2}})(v) \geq 0$, i.e., the sequence of weights $(w(G_{x^i})(v))_{i \geq 0}$ is convex in i.

We do case analysis on G: for $G \notin \mathcal{F}$,

$$\chi_{v \to x}(w)(G_{x^i}) = \frac{1}{n+i} - \frac{1}{n+i+1} = \frac{1}{(n+i)(n+i+1)}$$

Since $(n+i)(n+i+1)$ is strictly increasing in i, $\chi_{v \to x}(w)(G_{x^i})$ is strictly decreasing in i.

Here, let $G \in \mathcal{F}$, i.e., $G = K_a \cup K_b$ for some $a, b \geq 1$. Without loss of generality, assume $x \in V_a$. Then, $G_{x^i} = K_{a+i} \cup K_b$. Observe that for any $u \in V_b$,

$$w(G_{x^i})(u) = \frac{H(b)}{b(H(a+i)+H(b))} = \frac{1+1/b}{a+b+i+2}.$$

Consequently, we have

$$\chi_{u \to x}(w)(G_{x^i}) = \frac{1+1/b}{a+b+i+2} - \frac{1+1/b}{a+b+i+3}.$$

Again, this is strictly decreasing in i.
Now, we consider $u \in V_a$. Let $t \coloneqq a+i$ and $B \coloneqq H(b)+1 = b+2$. Then,

$$w(G_{x^i})(u) = \frac{H(t)}{t(H(t)+H(b))} = \frac{t+1}{t(t+B)}.$$

Decomposing the above into partial fractions gives

$$\frac{t+1}{t(t+B)} = \frac{1}{B} \cdot \frac{1}{t} + \frac{B-1}{B} \cdot \frac{1}{t+B}.$$

Observe that both $\frac{1}{t}$ and $\frac{1}{t+B}$ are strictly decreasing and strictly convex functions on t for all $t > 0$. Furthermore, the coefficients are also positive as $B > 1$. Since any positive linear combination of strictly convex functions is strictly convex, $(w(G_{x^i})(u))$ is strictly convex in i. Thus, we have showed that w satisfies clone-sharing dilution.

Consider $G \coloneqq K_3 \cup K_1 \in \mathcal{F}$ where $a = 3$ and $b = 1$. Let $x, v \in V_a$ and $u \in V_b$. Indeed, we have $N_G^k[v] \cap N_G^k[x] = V_a \supset \varnothing = N_G^k[u] \cap N_G^k[x]$. Now, we compute the clone sharing of v:

$$\begin{aligned}
\chi_{v \to x}(w)(G) &= w(G)(v) - w(G_{x^1})(v) \\
&= \frac{1+1/3}{3+1+2} - \frac{1+1/4}{4+1+2} \\
&= \frac{11}{252}.
\end{aligned}$$

We compute the clone sharing of u as well:

$$\begin{aligned}
\chi_{u \to x}(w)(G) &= w(G)(u) - w(G_{x^1})(u) \\
&= \frac{1+1/1}{3+1+2} - \frac{1+1/1}{4+1+2} \\
&= \frac{12}{252}.
\end{aligned}$$

Therefore, w does not satisfy dilution as desired.

For the backward direction, first define a positive sequence $H : \mathbb{Z}^+ \to \mathbb{R}_{>0}$ by

$$H(1) \coloneqq 1 \qquad H(2) = \frac{3}{2} \qquad H(3) = \frac{3}{17} \qquad H(c) = \frac{1}{c^2} \quad \forall c \geq 4.$$

Now, we define w as the following:

- If $G \in \mathcal{F}$ (with vertex sets V_a and V_b):

$$w(G)(v) \coloneqq \begin{cases} \frac{H(a)}{a(H(a)+H(b))}, & v \in V_a \\ \frac{H(b)}{b(H(a)+H(b))}, & v \in V_b \end{cases}.$$

  Again, note that the weight assignment depends only on which set v is in.

- If $G \notin \mathcal{F}$:

$$w(G)(v) \coloneqq \frac{1}{n} \quad \forall v \in V(G),$$

  i.e., w is a uniform graph weighting function.

First, we show that w is a graph weighting function on $G \in \mathcal{F}$. Here, w is positive as H is positive, and the weights sum up to 1 as before. Thus, w is a graph weighting function. Now, we show that w satisfies k-clone-sharing domination. Let G be a graph, $x, v, u \in V(G)$ be vertices such that $N_G^k[v] \cap N_G^k[x] \supseteq N_G^k[u] \cap N_G^k[x]$. Now, we again do case analysis on G:

Case 1: $G \notin \mathcal{F}$. For any $z \in V(G)$, $w(G)(z) = \frac{1}{n}$ and $w(G_{x'})(z) = \frac{1}{n+1}$. Hence, $\chi_{z\to x}(w)(G) = \frac{1}{n(n+1)}$. Since this holds for every vertex in G, $\chi_{v\to x}(w)(G) = \chi_{u\to x}(w)(G)$, so k-domination holds.

Case 2: $G \in \mathcal{F}$. Without loss of generality, assume $x \in V_a$. Indeed, if u and v belong to the same vertex set, then their k-neighborhoods intersect with the k-neighborhood of x in the same way, which is V_a or $\emptyset$, depending on if u and v are in V_a or V_b respectively. This implies they are assigned the same weight by w before and adding the clone, so $\chi_{v\to x}(w)(G) = \chi_{u\to x}(w)(G)$. It remains to check when they are in different vertex sets. In particular, we only need to check when $v \in V_a$ and $u \in V_b$ as otherwise the assumption does not hold. Thus, we get

$$\begin{aligned} \chi_{v\to x}(w)(G) - \chi_{u\to x}(w)(G) &= \left(\frac{H(a)}{a(H(a)+H(b))} - \frac{H(a+1)}{(a+1)(H(a+1)+H(b))}\right) \\ &\quad - \left(\frac{H(b)}{b(H(a)+H(b))} - \frac{H(b)}{b(H(a+1)+H(b))}\right) \\ &= \frac{g(a,b)}{(H(a)+H(b))(H(a+1)+H(b))}, \end{aligned}$$

where

$$g(a,b) = H(b)\left(\left(\frac{H(a)}{a} - \frac{H(a+1)}{a+1}\right) - \frac{H(a+1)-H(a)}{b}\right) + \frac{H(a)}{a}\frac{H(a+1)}{a+1}.$$

Here, it suffices to verify that $g(a,b) \geq 0$ for any $a, b \geq 1$. Observe that $H(c) > H(c+1)$ and consequently $\frac{H(c)}{c} > \frac{H(c+1)}{c+1}$ for all $c \geq 3$. Now, we do case analysis on a:

(i) $a = 1$:

$$g(1,b) = H(b)\left(0.25 - \frac{0.5}{b}\right) + 0.75.$$

- For $b = 1$, $g(1,1) = 1(0.25 - 0.5) + 0.75 = 0.5 > 0$.
- For $b = 2$, $g(1,2) = 1.5(0.25 - 0.25) + 0.75 = 0.75 > 0$.

- For $b \geq 3$, $g(1,b) > 0$ as $0.25 - \frac{0.5}{b} > 0$.

(ii) $a = 2$:

$$g(2,b) = H(b)\left(\frac{47}{68} + \frac{45}{34b}\right) + \frac{3}{68} > 0 \quad \forall b \geq 1.$$

(iii) $a \geq 3$: since $\frac{H(a)}{a} - \frac{H(a+1)}{a+1} > 0$ and $H(a+1) - H(a) < 0$, we can conclude that $g(a,b) > 0$.

Therefore, w satisfies k-clone-sharing domination.

Now, we consider the graph $G \coloneqq K_1 \cup K_1 \in \mathcal{F}$ (with $a = 1$ and $b = 1$). Let $x \in V(G)$ be one of the isolated vertices. Now we evaluate the weight assigned to x by w on graphs with clones added:

- $w(G)(x) = \frac{1}{1(1+1)} = \frac{1}{2} = 0.5$.
- $w(G_{x^1})(x) = \frac{1.5}{2(1.5+1)} = \frac{1.5}{5} = 0.3$.
- $w(G_{x^2})(x) = \frac{3/17}{3(3/17+1)} = \frac{1}{20} = 0.05$.

Thus, we finally have

$$\begin{aligned}\chi_{x\to x}(w)(G) - \chi_{x\to x}(w)(G_{x^1}) &= (w(G)(x) - w(G_{x^1})(x)) - (w(G_{x^1})(x) - w(G_{x^2})(x)) \\ &= (0.5 - 0.3) - (0.3 - 0.05) \\ &= -0.05 \\ &< 0,\end{aligned}$$

implying that w does not satisfy clone-sharing dilution. □

2.4.1 Constructing Interpretable Clone-robust Graph Weighting Functions

Indeed, the remaining question is whether the above clone-sharing properties can be satisfied. In this section, we present a framework to construct such graph weighting functions. To do so, we introduce two operators. First, we denote Q the *quotient lift operator* defined as the following:

$$Qw(G)(v) \coloneqq \frac{w(G/\equiv)([v]_G)}{|[v]_G|}$$

where $G/\equiv$ denotes the quotient graph of G and $[\cdot]_G : V(G) \to V(G/\equiv)$ the canonical quotient map. Second, we denote D the *diffusion operator*, which is defined as

$$D_k w(G)(v) \coloneqq \sum_{u \in V(G)} w(G)(u) P_G(u \to v) = \sum_{u \in N_G[v]} \frac{w(G)(u)}{|N_G[u]|}$$

where the lazy random walk kernel $P_G(u \to v)$ is defined as

$$P_G(u \to v) \coloneqq \begin{cases} \frac{1}{|N_G[u]|}, & u \in N_G[v] \\ 0, & \text{otherwise} \end{cases}.$$

We consider two natural extensions of the operator D to any integer $k \geq 2$. First, we apply the operator iteratively, which we denote as D^k. Second, we extend the diffusion uniformly over the k-neighborhood instead, which we denote as D_k. We will use the latter extension as the former is not compatible with our desired properties.

In the following theorem, we will show that applying Q and D_k consequently to a symmetric graph weighting function creates a k-clone-robust graph weighting function that satisfies the above clone-sharing properties.

Theorem 2.20. *Let w be a symmetric graph weighting function. Then, Qw is a 1-clone-robust graph weighting function satisfying clone-sharing dilution, but it fails clone-sharing positivity and 1-clone-sharing domination. Moreover, $D_k \circ Qw$ is a $(k+1)$-clone-robust graph weighting function satisfying both clone-sharing dilution and k-clone-sharing domination for $k \geq 1$, and the k-strong variants are satisfied as well if w is positive, i.e., $w(G)(v) > 0$ for all $G \in \mathcal{G}$ and $v \in V(G)$.*

Proof. We will prove the theorem in the following steps.

Probability Distribution. Here, it suffices to show that the functions satisfy non-negativity and the weights sum up to 1. By assumption, $w(G/\equiv)$ is a probability distribution over $G/\equiv$, so Qw is non-negative by construction. Summing over the vertices of G gives

$$\begin{aligned}\sum_{v \in V(G)} Qw(G)(v) &= \sum_{v \in V(G)} \frac{w(G/\equiv)([v]_G)}{|[v]_G|} \\ &= \sum_{[v]_G \in G/\equiv} \sum_{u \in [v]_G} \frac{w(G/\equiv)([v]_G)}{|[v]_G|} \\ &= \sum_{[v]_G \in G/\equiv} w(G/\equiv)([v]_G) \\ &= 1.\end{aligned}$$

Observe that D_k preserves non-negativity and that Qw is non-negative as we have showed earlier, so we get that $D_k \circ Qw$ is also non-negative. Thus, it suffices to show that the weights sum up to 1:

$$\begin{aligned}\sum_{v \in V(G)} D_k \circ Qw(G)(v) &= \sum_{v \in V(G)} \sum_{u \in N_G^k[v]} \frac{w(G/\equiv)([u]_G)}{|[u]_G| \cdot |N_G^k[u]|} \\ &= \sum_{u \in V(G)} \sum_{v \in N_G^k[u]} \frac{w(G/\equiv)([u]_G)}{|[u]_G| \cdot |N_G^k[u]|} \\ &= \sum_{u \in V(G)} \frac{w(G/\equiv)([u]_G)}{|[u]_G|} \\ &= 1,\end{aligned}$$

where the second line comes from the fact that k-neighborhood relation is symmetric.

Symmetry. Since w is symmetric, for any graph G, automorphism ϕ, and vertex v we have

$$\begin{aligned}Qw(G)(\phi(v)) &= \frac{w(G/\equiv)([\phi(v)]_G)}{|[\phi(v)]_G|} \\ &= \frac{w(G/\equiv)([v]_G)}{|[v]_G|} \\ &= Qw(G)(v)\end{aligned}$$

where the second line follows from w being symmetric and ϕ preserving equivalence class sizes.

Now, we show that $D_k \circ Qw$ is symmetric. Note that the k-neighborhood is preserved under

automorphisms, i.e., $N_G^k[\phi(v)] = \phi(N_G^k[v])$ for any $v \in V(G)$. Thus,

$$\begin{aligned} D_k \circ Qw(G)(\phi(v)) &= \sum_{u \in N_G^k[\phi(v)]} \frac{Qw(G)(u)}{|N_G^k[u]|} \\ &= \sum_{u \in N_G^k[v]} \frac{Qw(G)(\phi(u))}{|N_G^k[\phi(u)]|} \\ &= D_k \circ Qw(G)(v). \end{aligned}$$

Locality. Let G be a graph, $x, v \in V(G)$ be vertices such that $d_G(x, v) > 1$. This means that x and v do not belong in the same equivalence class, so adding a clone of either one to the graph does not change the size of the equivalence class of the other. Then,

$$\begin{aligned} Qw(G_{x^1})(v) &= \frac{w(G_{x^1}/\equiv)([v]_{G_{x^1}})}{\left|[v]_{G_{x^1}}\right|} \\ &= \frac{w(G_x/\equiv)([v]_G)}{|[v]_G|} \\ &= Qw(G)(v) \end{aligned}$$

where the second line comes from the fact that graphs obtained by adding clones share the same quotient graph (and representatives).

Now, we consider $(k+1)$-locality for $D_k \circ Qw$ for some $k \geq 1$, where we instead assume $d_G(x, v) > k + 1$ and get the following:

$$\begin{aligned} D_k \circ Qw(G_{x^1})(v) &= \sum_{u \in N_{G_{x^1}}^k[v]} \frac{Qw(G_{x^1}(u))}{|N_{G_{x^1}}^k[u]|} \\ &= \sum_{u \in N_G^k[v]} \frac{Qw(G(u))}{|N_G^k[u]|} \\ &= D_k \circ Qw(G)(v) \end{aligned}$$

where the second line follows from the triangle inequality: $d_G(u, x) \geq d_G(x, v) + d_G(u, v) > 1$.

Clone-sharing dilution and domination. For this part, we will adopt the following notations:

- $w_v \coloneqq w(G/\equiv)([v]_G)$;
- $c_v^i \coloneqq \left|[v]_{G_{x^i}}\right|$;
- $n_v^{k,i} \coloneqq \left|N_{G_{x^i}}^k[v]\right|$ (omit k when $k = 1$).

Now, we show the claims for each function separately.

(a) Qw: unpacking the definition of $\chi_{v \to x}(Qw)(G_{x^i})$ gives the following:

$$\begin{aligned} \chi_{v \to x}(Qw)(G_{x^i}) &= Qw(G_{x^i})(v) - Qw(G_{x^{i+1}})(v) \\ &= \frac{w_v}{c_v^i} - \frac{w_v}{c_v^{i+1}}. \end{aligned}$$

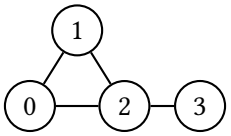
If v is a clone of x itself, then $c_v^{i+1} = c_v^i + 1$. Otherwise, they are equal, so $\chi_{v \to x}(Qw)(G_{x^i}) = 0$. Thus, Qw cannot satisfy clone-sharing positivity: if v is not a clone of x, then clearly v does not share any of its weight to the newly added clone of x.

Now, we show that Qw satisfies clone-sharing dilution: for any $0 \le i < j$,

$$\begin{aligned}\chi_{v\to x}(Qw)(G_{x^i}) - \chi_{v\to x}(Qw)(G_{x^j}) &= \left(\frac{w_v}{c_v^i} - \frac{w_v}{c_v^{i+1}}\right) - \left(\frac{w_v}{c_v^j} - \frac{w_v}{c_v^{j+1}}\right)\\ &\ge \left(\frac{w_v}{c_v^i} - \frac{w_v}{c_v^i + 1}\right) - \left(\frac{w_v}{c_v^j} - \frac{w_v}{c_v^j + 1}\right)\\ &\ge 0\end{aligned}$$

where the second line follows from the above observation and the last line follows from $i < j$.

At last, we suppose for contradiction that Qw satisfies 1-clone-sharing domination. Consider the following graph G':



Observe that vertex 0 and vertex 1 are clones in this graph and that the intersection of the closed-neighborhoods of vertex 0 and vertex 1 are equal to that of vertex 0 and vertex 2, i.e., $N_G[0] \cap N_G[1] = \{0, 1, 2\} = N_G[0] \cap N_G[2]$. Since vertex 2 is not a clone of vertex 0, its weight never changes upon adding clones of vertex 0, while vertex 1 strictly does as per the earlier section of the proof. Thus, by clone-sharing domination of Qw, we have that

$$0 < \chi_{1\to 0}(Qw)(G') \le \chi_{2\to 0}(Qw)(G) = 0,$$

which is a contradiction.

(b) $D_k \circ Qw$: denote $C_v^i := \{[v]_{G_{x^i}} : u \in N^k_{G_{x^i}}[v]\}$, the set of equivalence classes contained in k-neighborhood of v. Then, we group the vertices in the k-neighborhood in the definition of $D_k \circ Qw(G_{x^i})(v)$ into equivalence classes, giving

$$\begin{aligned}D_k \circ Qw(G_{x^i})(v) &= \sum_{u \in N^k_{G_{x^i}}[v]} \frac{w_u}{c_u^i n_u^{k,i}}\\ &= \sum_{C \in C_v^i} \sum_{u \in C} \frac{w_u}{c_u^i n_u^{k,i}}\\ &= \sum_{C \in C_v} \frac{w_C}{n_C^{k,i}},\end{aligned}$$

where w_C and n_C are defined by taking any vertex $u \in C$. Observe that the number of the classes remain the same regardless of how many clones added, so we can drop the i from C as we only need one representative from each class anyway.

To continue, we parition the equivalence classes into two groups: (1) C_1 containing classes that are contained in the k-neighborhood of x and (2) C_2 containing the remaining classes. For classes in C_1, the sizes of their k-neighborhoods increase by the number of clones added. For each class in the second group, however, the size of its k-neighborhood never changes upon adding clones of x.

We evaluate $\chi_{v\to x}(D_k \circ Qw)(G_{x^i})$:

$$\begin{aligned}\chi_{v\to x}(D_k \circ Qw)(G_{x^i}) &= \sum_{C \in C_v} w_C \left(\frac{1}{n_C^{k,i}} - \frac{1}{n_C^{k,i+1}}\right)\\ &= \sum_{C \in C_1} w_C \left(\frac{1}{n_C^k + i} - \frac{1}{n_C^k + i + 1}\right)\\ &= \sum_{C \in C_1} \frac{w_C}{(n_C^k + i)(n_C^k + i + 1)},\end{aligned}$$

where the third line follows from the above observation. Then, for $i < j$, we immediately have that

$$\chi_{v\to x}(D_k \circ Qw)(G_{x^i}) \geq \chi_{v\to x}(D_k \circ Qw)(G_{x^j}). \tag{2.4}$$

Now, we prove that $D_k \circ Qw$ satisfies k-clone-sharing domination. Let $u, v \in V(G)$ such that $N_G^k[u] \cap N_G^k[x] \subseteq N_G^k[v] \cap N_G^k[x]$. Observe that the former intersection is in fact $C_{1,u}$ and the latter $C_{1,v}$. By the derivation above, we have

$$\chi_{v\to x}(D_k \circ Qw)(G) - \chi_{u\to x}(D_k \circ Qw)(G) = \sum_{C\in C_{1,v}\setminus C_{1,u}} \frac{w_C}{n_C^k(n_C^k+1)} \geq 0 \tag{2.5}$$

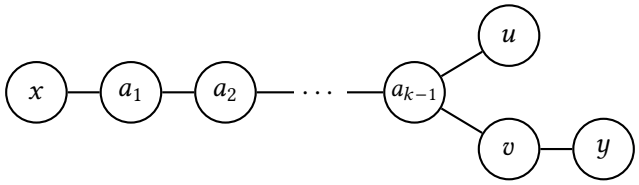
as other classes canceled out.

(c) $D \circ Qw$ ($w > 0$): we follow the same clone-sharing dilution and domination proof as in (b). The main differences are in eq. (2.4) and eq. (2.5). In particular, the inequalities would be strict instead with w being positive. Thus, it remains to show that the k-clone-sharing domination is strong, i.e., we get equality in *eq.* (2.5) if and only if $C_{1,v} = C_{1,u}$. Indeed, if the two sets are equal, the equality is immediate. Similarly, if there exists a class $C \in C_{1,v}$ such that $C \notin C_{1,u}$, then we must necessarily get a strict inequality, so we are done. □

Although we have proven that D_k generalizes smoothly with the other clone-sharing properties, we will show in the following lemma why applying the diffusion operator iteratively, D^k, does not work out as cleanly.

Lemma 2.21. *There exists a symmetric graph weighting function w such that for any $k \geq 2$, $D^k \circ Qw$ does not satisfy clone-sharing dilution and k-clone-sharing domination.*

Proof. Let w be a uniform graph weighting function, i.e., $w(G)(v) = \frac{1}{|V(G)|}$ for all $v \in V(G)$. Denote $\hat{w} \coloneqq D^k \circ Qw$. We first show $\hat{w}$ does not satisfy k-clone-sharing domination. Consider the graph G defined as the following:



where C_x is the equivalence class containing x.

Denote $G' \coloneqq G_{x^1}$. We begin by simplifying $\Delta\chi \coloneqq \chi_{v\to x}(\hat{w})(G) - \chi_{u\to x}(\hat{w})(G)$:

$$\begin{aligned}
\Delta\chi &= (\hat{w}(G)(v) - \hat{w}(G')(v)) - (\hat{w}(G)(u) - \hat{w}(G')(u)) \\
&= \left(\sum_{z\in V(G)} \left(Qw(G)(z)P_G^k(z\to v) - Qw(G')(z)P_{G'}^k(z\to v)\right) - Qw(G')(x)P_{G'}^k(x\to v)\right) \\
&\quad - \left(\sum_{z\in V(G)} \left(Qw(G)(z)P_G^k(z\to u) - Qw(G')(z)P_{G'}^k(z\to u)\right) - Qw(G')(x)P_{G'}^k(x\to u)\right)
\end{aligned}$$

From here, it suffices to consider vertex z such that the value of $Qw(G')(z)$ or $P_{G'}^k(z \to v/u)$ changes. Only $z = x$ changes the former, while changes in the latter occur when there exists a k-step (potentially lazy) walk such that (1) it starts at z, (2) it ends at v/u, and (3) it includes a vertex whose degree has changed—x or a_1. For conditions (1)-(3) to hold, z can only be either x, a_1, or a_2. Moreover, if $z = x$ or $z = a_2$, the walk must be $x \to a_1 \to \cdots \to a_{k-1} \to v/u$ or $a_2 \to a_1 \to a_2 \to \cdots \to a_{k-1} \to v/u$ respectively, which are identical for v and u except at the last step. These last steps share the same

probability as both vertices are neighbors of a_{k-1}, implying $P_H^k(z \to v) = P_H^k(z \to u)$ for $H \in \{G, G'\}$. Thus, we can further simplify $\Delta\chi$ to the following:

$$\Delta\chi = Qw(G)(a_1)[\underbrace{(P_G^k(a_1 \to v) - P_G^k(a_1 \to u))}_{\Delta P_G^k} - \underbrace{(P_{G'}^k(a_1 \to v) - P_{G'}^k(a_1 \to u))}_{\Delta P_{G'}^k}]$$

as all other terms canceled out or evaluated to 0.

For $z = a_1$, the walk must be a path from a_1 to v/u with one lazy step on any vertices in the path. This is where the difference occurs: the walk can perform a lazy step at v and u, whose degrees are different. In particular, we have the following:

$$P_H^k(a_1 \to v/u) = \prod_{i=1}^{k-1} \frac{1}{|N_H[a_i]|} \left(\sum_{j=1}^{k-1} \frac{1}{|N_H[a_j]|} + \frac{1}{|N_H[v/u]|} \right) \text{ for } H \in \{G, G'\},$$

where the sum terms come from the probability of taking a lazy step on a vertex in the path.
We now evaluate ΔP_H^k:

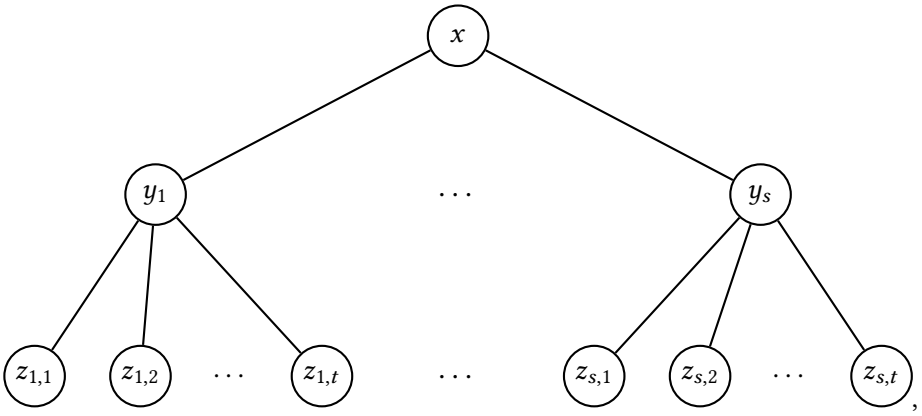
$$\Delta P_H^k = \prod_{i=1}^{k-1} \frac{1}{|N_H[a_i]|} \left(\frac{1}{N_H[v]} - \frac{1}{N_H[u]} \right) = \frac{1}{6} \prod_{i=1}^{k-1} \frac{1}{|N_H[a_i]|}.$$

With this, we are now ready to evaluate $\Delta\chi$:

$$\begin{aligned}
\Delta\chi &= Qw(G)(a_1)(\Delta P_G^k - \Delta P_{G'}^k) \\
&= \frac{1}{6} Qw(G)(a_1) \left(\frac{1}{|N_G[a_1]|} - \frac{1}{|N_{G'}[a_1]|} \right) \prod_{i=2}^{k-1} \frac{1}{|N_G[a_i]|} \\
&= \frac{1}{72} Qw(G)(a_1) \prod_{i=2}^{k-1} \frac{1}{|N_G[a_i]|} \\
&> 0.
\end{aligned}$$

Observe that $N_G^k[v] \cap N_G^k[x] = N_G^k[u] \cap N_G^k[x]$, which implies that $\hat{w}$ does not satisfy clone-sharing domination, as otherwise we must get $\Delta\chi = 0$.

Now we show $\hat{w}$ does not satisfy clone-sharing dilution. Consider the following graph G:



that is, $V(G) = \{x\} \cup Y \cup Z$ where $Y \coloneqq \{y_1, \ldots, y_s\}$ and $Z \coloneqq \{z_{i,j} : i \in [s], j \in [t]\}$ for some $s \geq 2$ and large t. We connect x to y_i for all $i \in [s]$ and each y_i to $z_{i,j}$ for all $j \in [t]$. Observe that G is a quotient graph and that we have $1 + s + st$ equivalence classes in total.

Let $G_i \coloneqq G_{x^i}$. We define by $T(t)$ the 3×3 transition matrix between the set $X \coloneqq [x]_{G_i}$, Y, and Z:

$$T(t) \coloneqq \begin{pmatrix} \dfrac{i+1}{i+s+1} & \dfrac{s}{i+s+1} & 0 \\ \dfrac{i+1}{i+t+2} & \dfrac{1}{i+t+2} & \dfrac{t}{i+t+2} \\ 0 & 1/2 & 1/2 \end{pmatrix},$$

where $T(t)_{A,B}$ is the probability that a vertex from A steps to a vertex in B. Then, given that w is symmetric, we can write the weight $\hat{w}(G_i)(x)$ as the following:

$$\hat{w}(G_i)(x) = \frac{1}{(1+s+st)(i+1)}\Big(1 \cdot (T(t)^k)_{X,X} + s \cdot (T(t)^k)_{Y,X} + st \cdot (T(t)^k)_{Z,X}\Big),$$

where the $i+1$ factor in the denominator comes from the fact that the walk must end at x and each of the sum term comes from the contribution of vertices in X, Y, and Z respectively.

From here, we define

$$L_k(i) \coloneqq \lim_{t\to\infty}(st \cdot \hat{w}(G_i)(x)),$$

which we will later use to compute $\chi_{x\to x}(\hat{w})(G_i)$. Let $x = 1/t$. We parameterize the transition matrix by x:

$$T(x) = \begin{pmatrix} a & 1-a & 0 \\ \dfrac{x(i+1)}{x(i+2)+1} & \dfrac{x}{x(i+2)+1} & \dfrac{1}{x(i+2)+1} \\ 0 & 1/2 & 1/2 \end{pmatrix},$$

where $a \coloneqq \frac{i+1}{i+s+1}$. Observe that we can rewrite the limit of the contribution from vertices in Z as $t \to \infty$ as the definition of derivative at $x = 0$:

$$\lim_{t\to\infty} st \cdot (T(t)^k)_{Z,X} = s \cdot \lim_{x\to 0} \frac{(T(x)^k)_{Z,X} - (T(0)^k)_{Z,X}}{x} = s \cdot \frac{d}{dx}\Big[T(x)^k\Big]_{Z,X}\Big|_{x=0}.$$

By the matrix product rule, the derivative of $T(x)^k$ at $x = 0$ is

$$\frac{d}{dx}\Big[T(x)^k\Big]\Big|_{x=0} = \sum_{m=0}^{k-1} T(0)^m T'(0) T(0)^{k-1-m}.$$

We evaluate $T(x)$ and its derivative $T'(x)$ at $x = 0$:

$$T(0) = \begin{pmatrix} a & 1-a & 0 \\ 0 & 0 & 1 \\ 0 & 1/2 & 1/2 \end{pmatrix}, \qquad T'(0) = \begin{pmatrix} 0 & 0 & 0 \\ i+1 & 1 & -(i+2) \\ 0 & 0 & 0 \end{pmatrix}.$$

Let $w_m \coloneqq (T(0)^m)_{Z,Y}$ be the probability of Z transitioning to Y after m steps in the asymptotic limit as $t \to \infty$. Observe that the product $T'(0)T(0)^{k-1-m}$ contains only one non-zero entry in column X, which is $(i+1)a^{k-1-m}$ at row Y. Thus, we have

$$\frac{d}{dx}\Big[T(x)^k\Big]_{Z,X}\Big|_{x=0} = (i+1)\sum_{m=0}^{k-1} a^{k-1-m} w_m.$$

Here, we compute the limit for other terms, given by the following:

$$\lim_{t\to\infty} \frac{st}{(1+s+st)(i+1)} = \frac{1}{i+1}, \qquad \lim_{t\to\infty}(T(t)^k)_{X,X} = a^k, \qquad \lim_{t\to\infty} s(T(t)^k)_{Y,X} = 0.$$

Hence, the limit $L_k(i)$ resolves to

$$L_k(i) = \frac{1}{i+1}\left(a^k + s(i+1)W_k(a)\right) = f_k(i) + sW_k(a),$$

where $f_k(i) \coloneqq \frac{(i+1)^{k-1}}{(i+s+1)^k}$ and $W_k(a) \coloneqq \sum_{m=0}^{k-1} a^{k-1-m} w_m$.

Now, we evaluate $\chi_{x\to x}(\hat{w})(G_i)$ at $i = 0$, which is asymptotically governed by $L_k(0) - L_k(1)$. Here, we consider the two components of $L_k(i)$:

1. f_k: we compare the ratio of $f_k(1)$ and $f_k(2)$ as follows:

$$\frac{f_k(1)}{f_k(0)} = \frac{1}{2}\left(\frac{2(s+1)}{s+2}\right)^k > 1 \quad \forall s \geq 2, k \geq 2,$$

   implying $f_k(0) < f_k(1)$.

2. $sW_k(a)$: since w_m is a probability, it is non-negative for any $m \geq 0$. Hence, $W_k(a)$ is a polynomial in a with non-negative coefficients. Observe that $a_0 = \frac{1}{s+1}$ at $i = 0$ and $a_1 = \frac{2}{s+2}$ at $i = 1$. Since $s \geq 2$, we have $0 < a_0 < a_1$, so $W_k(a_0) \leq W_k(a_1) \iff sW_k(a_0) \leq sW_k(a_1)$.

Together, they imply $L_k(0) < L_k(1)$, so we have

$$\lim_{t\to\infty} st \cdot \chi_{x\to x}(\hat{w})(G_0) = L_k(0) - L_k(1) < 0.$$

Therefore, there exists a sufficiently large $t \in \mathbb{Z}^+$ such that $st \cdot \chi_{x\to x}(\hat{w})(G) < 0$. We divide both sides by $st > 0$ without flipping the inequality, giving $\chi_{x\to x}(\hat{w})(G) < 0$. By theorem 2.15, we are done. □

Chapter 3

Clone-robust Games

3.1 Preliminaries

First, we introduce a way of converting metric spaces into graphs.

Definition 3.1. Let (M, d) be a metric space. For every finite subset $S \subseteq M$, we construct $G_r(S)$, the r-radius graph over S, by connecting every pair of points $x, y \in S$ if and only if $r > d(x, y)$.

Here, we define what it means to be an approximate clone on metric spaces: let S be a subset of a metric space (M, d). Then, x' is an *ϵ-clone* of an element $x \in S$ if $d(x, x') \leq \epsilon$. With this, we present a construction of a weighting function that does not depend on the number of clones we add to the original set S but instead on how similar they are to the original set in the following theorem:

Theorem 3.2 (ϵ-Lipschitz-continuity)**.** *Let w be a symmetric graph weighting function, $\alpha > 0$ be some positive radius, and ν be a probability density function with support $[0, \alpha]$ and value is bounded by $\bar{\nu}$. For every finite subset S of the (psuedo) metric space (M, d) and every $x \in S$, we define f as the following:*

$$f_{\nu,w}(S)(x) \coloneqq \int_0^\alpha \nu(r) w(G_r(S))(x) dr.$$

*Then, f is a weighting function of (M, d) satisfying **ϵ-Lipschitz-continuity**: for all $\epsilon \geq 0$, set $C \subseteq M$ containing ϵ-clones of elements in S, and bijection $\pi : S \cup C \to M$,*

$$\left| f_{\nu,w}(S \cup C)(x) - f_{\nu,w}(\pi(S \cup C))(\pi(x)) \right| \leq 2\bar{\nu}|S|^2 (\max_{x \in S} d(z, \pi(z)) + \epsilon).$$

Proof. Denote $S' \coloneqq S \cup C$. We first show that f is a weighting function. Given that the density function ν is a bounded measurable function, the map $r \mapsto \nu(r) w(G_r(S'))(x)$ is integrable and $f_{\nu,w}(S')(x)$ is well-defined. Now, we sum over all elements in S':

$$\begin{aligned}
\sum_{x \in S'} f_{\nu,w}(S')(x) &= \sum_{x \in S'} \int_0^\alpha \nu(r) w(G_r(S'))(x) dr \\
&= \int_0^\alpha \nu(r) \sum_{x \in S'} w(G_r(S'))(x) dr \\
&= \int_0^\alpha \nu(r) dr \\
&= 1
\end{aligned}$$

where the second line comes from the fact that w is a graph weighting function.

Let $\epsilon \geq 0$. Denote $\delta \coloneqq \max_{z \in S} d(z, \pi(z))$ as the maximum distance created by the bijection π of any original elements in S. Define $\mathcal{D}_S \coloneqq \{d(x, y) : x, y \in S\}$ and $B \coloneqq \bigcup_{d' \in \mathcal{D}_S} [d' - 2(\epsilon + \delta), d' + 2(\epsilon + \delta)]$ as the set of pair-wise distances and the extended set of *bad radii* respetively. Note that the cardinality of B is the same as $\mathcal{D}_S$, which is bounded above by $|S|^2$.

From here, we show that for every radius $r \in (0, \alpha] \setminus B$, the graph $G_r(S')$ is isomorphic to the graph $G_r(\pi(S'))$. Let $x', y' \in C$ be the newly inserted clones and $x, y \in S$ be their original elements respectively. It suffices to show that x' is connected to y' in graph $G_r(S')$ if and only if $\pi(x')$ is connected to $\pi(y')$ in graph $G_r(\pi(S'))$. Now, we apply the triangle inequality together with the ϵ-clone assumption to get

$$\begin{aligned} |d(x', y') - d(x, y)| &\leq 2\epsilon \\ |d(\pi(x'), \pi(y')) - d(x, y)| &\leq 2(\epsilon + \delta). \end{aligned} \tag{3.1}$$

We do the following case analysis on r:
Case $r > d(x, y) + 2(\epsilon + \delta)$. By eq. (3.1), we have

$$\begin{aligned} r &> d(x, y) + 2\epsilon \geq d(x', y') \\ r &> d(x, y) + 2(\epsilon + \delta) \geq d(\pi(x'), \pi(y')). \end{aligned}$$

Case $r < d(x, y) - 2(\epsilon + \delta)$. Similarly, by eq. (3.1) we get

$$\begin{aligned} r &< d(x, y) < 2\epsilon \leq d(x', y') \\ r &< d(x, y) < 2(\epsilon + \delta) \leq d(\pi(x'), \pi(y')). \end{aligned}$$

Hence, we showed the graphs are isomorphic outside of the bad radii. This gives the following:

$$\begin{aligned} \left| f_{\nu,w}(S')(x') - f_{\nu,w}(\pi(S'))(\pi(x')) \right| &\leq \int_0^\alpha \nu(r) |w(G_r(S'))(x') - w(G_r(\pi(S')))(\pi(x'))| dr \\ &\leq \int_B \nu(r) dr \\ &\leq \bar{\nu} \cdot \mu(B) \\ &\leq 2\bar{\nu} |S|^2 (\epsilon + \delta) \end{aligned}$$

where the second line follows from the fact that w is symmetric, implying outside of the bad radii the weights canceled out. □

Remark 3.3. Note that for sets S' without clones, we simply set $\epsilon = 0$, obtaining true Lipschitz continuity.

Finally, we list terminologies that we will use throughout the section. A set of *pure strategies* X_i of player i is a finite set containing abstract representations of actions player i can perform. Denote by X the space of all pure strategies and by $X_{-i} \coloneqq \prod_{j \neq i} X_j$ the space of pure strategies of opponents of i. A *mixed strategy* p_i of player i is a probability distribution over the set of pure strategies X_i. In particular, p_i is a vector with $|X_i|$ entries such that for each pure strategy $x \in X_i$, $p_i(x)$ is non-negative and upon summing the entries we get 1. An entry $A_i(x)$ of the *payoff matrix* $A_i : X \to \mathbb{R}$ of player i indicates the payoff player i receives when the combination of strategies x is played. We denote by A_i^* the *maximum payoff* player i could get of any combination of strategies, i.e., $A_i^* \coloneqq \max_{x \in X} A_i(x)$. An *embedding* $\pi_{p_{-i}} : X_i \to \mathbb{R}^{|X_{-i}|}$ of pure strategies of player i is defined by

$$\pi_{p_{-i}}(x_i) \coloneqq [A_i(x_i, x_{-i}) \cdot p_{-i}(x_{-i})]_{x_{-i} \in X_{-i}},$$

where p_{-i} is a combination of mixed strategies of other players. This naturally induces the set of embeddings $S_i(p_{-i}) \coloneqq \left\{ \pi_{p_{-i}}(x_i) : x_i \in X_i \right\}$ of player i.

3.2 Clone-robust Games

We start by defining an update rule for the game.

Definition 3.4 (update rule)**.** Let X_i be the set of pure strategies of player i with size n_i; $A_i \in \mathbb{R}^{n_1 \times n_2}$ be the payoff matrix for player i; Δ_i be the set of mixed strategies of player i; $\pi_{p_{-i}} : X_i \to \mathbb{R}^{\prod_{j \neq i} n_j}$ be the embedding of pure strategies of player i; $S_i(p_{-i})$ be the set of embedded pure strategies of player i; f_i be the function constructed through theorem 3.2 on $(\mathbb{R}^{n_{-i}}, d_\infty)$ with some symmetric graph weighting function w_i, positive radius α_i, probability density function ν_i with support $[0, \alpha]$, and $\epsilon = 0$. Then, the **n-player update rule** of the game $F : \Delta \to \Delta$ is defined as the following:

$$F(p) \coloneqq \left(f_i(S_i(p_{-i})) \circ \pi_{p_{-i}} \right)_{i=1}^{n}$$

where $\Delta \coloneqq \prod_i \Delta_i$.

One core idea in game theory is to show that a fixed point exists in a game. In the following lemma, we will show that there exists a fixed point in our game setting:

Lemma 3.5. *Let F be an update rule. Then, F admits a fixed point in Δ.*

Proof. Via Brouwer's fixed-point theorem, it suffices to show that Δ is both convex and compact and that F is continuous. For the former, since the Cartesian product preserves both convexity and compactness, it is sufficient to look at Δ_i by symmetry. Convexity follows from definition. Compactness follows from Heine-Borel: boundedness of Δ_i is straightforward and closedness of Δ_i comes from the fact that it is an intersection between two closed sets (a hyperplane and a closed-orthant). Thus, Δ is both convex and compact.

Similarly, one only needs to show that each component of F is continuous. By symmetry, it is enough to show that $f_i(S_i(p_{-i})) \circ \pi_{p_{-i}}$ is continuous for any $p_{-i} \in \Delta_{-i}$. Indeed, as composition preserves continuity, the two functions can be analyzed separately. Since the former is (Lipschitz) continuous by theorem 3.2, we only need to focus on the latter, which is also continuous as it is a linear transformation. □

Though we have established that our game has a fixed point, it could be that there are multiple fixed points. Indeed, if F is a contraction map, then Banach fixed-point theroem gives us a unique fixed point. Thus, we will show that F is a contraction map under certain condition in the following lemma.

Lemma 3.6. *Let F be a n-player update rule on the pure strategies $X_1, \ldots, X_n$. If for each $i \in [n]$,*

$$\bar{\nu}_i < \frac{1}{2|X_i|^2 A_i^* (n-1)},$$

then F is a contraction map on (Δ, d_∞).

Proof. We want to show that for every $p, q \in \Delta$, $\|F(p) - F(q)\|_\infty \leq L\|p - q\|_\infty$ for some $L < 1$. We analyze each component F_i separately:

$$\begin{aligned}
\|F_i(p_{-i}) - F_i(q_{-i})\|_\infty &= \max_{x \in X_i} \left| f_i(S_i(p_{-i}))(\pi_{p_{-i}}(x)) - f_i(S_i(q_{-i}))(\pi_{q_{-i}}(x)) \right| \\
&\leq 2\bar{\nu}_i |X_i|^2 \max_{x \in X_i} \|\pi_{p_{-i}}(x) - \pi_{q_{-i}}(x)\|_\infty \\
&= 2\bar{\nu}_i |X_i|^2 \max_{x \in X_i} \max_{x' \in X_{-i}} A_{x,x'} |p_{-i}(x') - q_{-i}(x')| \\
&\leq 2\bar{\nu}_i |X_i|^2 A_i^* \max_{x' \in X_{-i}} \left| \prod_{j \neq i} p_j(x'_j) - \prod_{j \neq i} q_j(x'_j) \right|
\end{aligned}$$

where the second line is by Lipschitz continuity of f_i. We re-index the opponents of i to $j_1, \ldots, j_{n-1}$. From here, we use the telescopic identity on the difference of products:

$$\begin{aligned}
\left|\prod_{j\neq i} p_j(x'_j) - \prod_{j\neq i} q_j(x'_j)\right| &= \left|\sum_{k=1}^{n-1} p_{j_k}(x'_{j_k}) - q_{j_k}(x'_{j_k}) \cdot \prod_{l=1}^{k-1} q_{j_l}(x'_{j_l}) \cdot \prod_{l=k+1}^{n-1} p_{j_l}(x'_{j_l})\right| \\
&\leq \sum_{k=1}^{n-1} \left|p_{j_k}(x'_{j_k}) - q_{j_k}(x'_{j_k})\right| \cdot \left|\prod_{l=1}^{k-1} q_{j_l}(x'_{j_l})\right| \cdot \left|\prod_{l=k+1}^{n-1} p_{j_l}(x'_{j_l})\right| \\
&\leq \sum_{k=1}^{n-1} \left|p_{j_k}(x'_{j_k}) - q_{j_k}(x'_{j_k})\right| \\
&= \sum_{j\neq i} \left|p_j(x'_j) - q_j(x'_j)\right| \\
&\leq \sum_{j\neq i} \|p - q\|_\infty \\
&= (n-1)\|p - q\|_\infty,
\end{aligned}$$

where the second line os obtained by distributing the absolute inside the sum, the third line follows from the fact that q and p are probabilities, and the fifth line follows from the definition of taking sup-norm. Thus, we have

$$\|F_i(p_{-i}) - F_i(q_{-i})\|_\infty \leq 2\bar{v}_i |X_i|^2 A_i^* (n-1)\|p - q\|_\infty.$$

Since this is symmetric for any i, we get

$$\begin{aligned}
\|F(p) - F(q)\|_\infty &= \max_{i\in[n]} \|F_i(p_{-i}) - F_i(q_{-i})\|_\infty \\
&\leq \max_{i\in[n]} 2\bar{v}_i |X_i|^2 A_i^* (n-1)\|p - q\|_\infty \\
&< \|p - q\|_\infty,
\end{aligned}$$

where the last line comes from the assumption. □

Now, we show that our game is indeed clone-robust. To do so, we again define notion of clone, this time in the context of a game: we say a pure strategy x'_i is an *ϵ_i-clone strategy* of a pure strategy x_i if $\max_{x_{-i}\in X_{-i}} \left|A_i(x_i, x_{-i}) - A(x'_i, x_{-i})\right| \leq \epsilon_i$. Note that this implies for any $p_{-i} \in \Delta_{-i}$,

$$\|\pi_{p_{-i}}(x_i), \pi_{p_{-i}}(x'_i)\|_\infty = \max_{x_{-i}\in X_{-i}} p_{-i}(x_{-i}) \left|A_i(x_i, x_{-i}) - A_i(x'_i, x_{-i})\right| \leq \epsilon_i.$$

Upon adding multiple clones, however, the size of the set of pure strategies increases. This means the condition on $\bar{v}_i$ of the original game may no longer hold. Thus, we need to allow for some errors, which is captured in the following theorem.

Theorem 3.7 (clone-robust game)**.** *Let F be an n-player update rule for strategies X. If for each player i,*

$$\bar{v}_i < \frac{1}{2|X_i|^2 A_i^* (n-1)},$$

then F admits a unique fixed point $p^* \in \Delta$. Furthermore, let $F' : \Delta' \to \Delta'$ be an extended update rule on strategy space $X' := X \cup C$, where C_i is a set of ϵ_i-clone strategies of X_i. Then, the followings hold:

- *For any $p', q' \in \Delta'$, $\|F'(p') - F'(q')\|_\infty < \|p' - q'\|_\infty + \max_{j\in[n]} E_j$ where $E_j := 2\bar{v}_j |X_j|^2 \epsilon_j$.*

- *There exists a hyperrectangle $\mathcal{R} \subseteq \mathbb{R}^{|X'|}$ such that*
  (a) the length of the sides corresponding to player i is bounded by

$$\ell_i \coloneqq \frac{E_i}{1+K_i} + \frac{K_i}{1+K_i}\left(\frac{\sum_{j=1}^{n}\frac{E_j}{1+K_j}}{1-\sum_{j=1}^{n}\frac{K_j}{1+K_j}}\right),$$

  where $K_i \coloneqq 2\bar{v}_i|X_i|^2 A_i'^*$
  (b) the set of new fixed points P' is contained in $\mathcal{R}$.

Proof. By theorem 3.6, F is a contraction map, so Banach fixed-point theorem tells us that it admits a unique fixed point p^*. Let $p', q' \in \Delta'$. We analyze with respect to player i as before, giving

$$\begin{aligned}
\|F_i'(p'_{-i}) - F_i'(q'_{-i})\|_\infty &= \max_{x_i \in X_i'} \left| f_i(S_i(p'_{-i}))(\pi_{p'_{-i}}(x_i)) - f_i(S_i(q'_{-i}))(\pi_{q'_{-i}}(x_i)) \right| \\
&\leq 2\bar{v}_i|X_i|^2 (\max_{x_i \in X_i'} \|\pi_{p'_{-i}}(x_i) - \pi_{q'_{-i}}(x_i)\|_\infty + \epsilon_i) \\
&\leq 2\bar{v}_i|X_i|^2 A_i'^* \sum_{j \neq i} \|p'_j - q'_j\|_\infty + 2\bar{v}_i|X_i|^2\epsilon_i \\
&\leq 2\bar{v}_i|X_i|^2 A_i'^* (n-1)\|p' - q'\|_\infty + E_i \\
&< \|p' - q'\|_\infty + E_i,
\end{aligned} \tag{3.2}$$

where the second line is from theorem 3.2 with $\epsilon = \epsilon_i$ and the third line follows from the proof in theorem 3.6. Hence, we obtain the final bound for the whole function:

$$\|F'(p') - F'(q')\|_\infty = \max_{i \in [n]} \|F_i'(p'_{-i}) - F_i'(q'_{-i})\|_\infty < \|p' - q'\|_\infty + \max_{j \in [n]} E_j.$$

Denote by $D_i \coloneqq \sup_{p'^*, q'^* \in P'} \|p_i'^* - q_i'^*\|_\infty$ the maximum distance between any two fixed points along strategy space of Player i. We will show that the quantity is bounded by ℓ_i. First, we take the supremum over all pairs in the third line of eq. (3.2), giving

$$D_i \leq K_i \sum_{j \neq i} D_j + E_i,$$

where the left side of the inequality comes from p' and q' being fixed points. Here, we continue by rewriting the sum as $\sum_{j \neq i} D_j = \sum_{j=1}^{n} D_j - D_i$:

$$D_i \leq K_i \sum_{j=1}^{n} D_j - K_i D_i + E_i$$

Adding $K_i D_i$ to both sides, factoring D_i out on the left hand side of the inequality, and dividing by $1 + K_i$ gives

$$D_i \leq \frac{K_i}{1+K_i} \sum_{j=1}^{n} D_j + \frac{E_i}{1+K_i}. \tag{3.3}$$

Now, we sum over all players:

$$\sum_{i=1}^{n} D_i \leq \underbrace{\left(\sum_{i=1}^{n} \frac{K_i}{1+K_i}\right)}_{\alpha} \sum_{j=1}^{n} D_j + \underbrace{\sum_{i=1}^{n} \frac{E_i}{1+K_i}}_{\beta}.$$

Subtracting both sides with $\alpha \sum D_j$ gives

$$\sum_{j=1}^{n} D_j(1-\alpha) \le \beta.$$

By the assumption, $K_i < \frac{1}{n-1}$, so we have

$$\alpha = \sum_{i=1}^{n} \frac{K_i}{1+K_i} < \sum_{i=1}^{n} \frac{1/(n-1)}{1+1/(n-1)} = \frac{1/(n-1)}{n+n/(n-1)} < 1.$$

Thus, we can divide both sides by $1-\alpha$:

$$\sum_{j=1}^{n} D_j \le \frac{\beta}{1-\alpha}.$$

Substituting this back in eq. (3.3) gives what we wanted

$$D_i \le \frac{K_i}{1+K_i}\left(\frac{\beta}{1-\alpha}\right) + \frac{E_i}{1+K_i} = \ell_i.$$

For each player i and strategy $x \in X_i'$, define the maximum and minimum probabilities of any fixed points as follow:

$$a_i(x) = \inf_{p'^* \in P'} p_i'^*(x) \qquad \text{and} \qquad b_i(x) = \sup_{p'^* \in P'} p_i'^*(x).$$

Then, observe that we have for any $x \in X_i'$,

$$b_i(x) - a_i(x) \le \sup_{p'^*, q'^* \in P} \left|p_i'^*(x) - q_i'^*(x)\right| \le \sup_{p'^*, q'^* \in P} \|p_i'^* - q_i'^*\|_\infty = D_i \le \ell_i,$$

where the first inequality follows from the definition of supremum. Define a hyperrectangle $\mathcal{R}$ as the following:

$$\mathcal{R} := \prod_{i=1}^{n} \prod_{x \in X_i'} [a_i(x), b_i(x)].$$

Indeed, by construction, the length of $\mathcal{R}$ along coordinates belonging to player i is bounded by ℓ_i. Thus, it suffices to show that every fixed point $p'^* \in P$ is in the hyperrectangle. This is trivially true as for each player i and strategy $x \in X_i'$, $a_i(x) \le p'^*(x) \le b_i(x)$ by definition, so we are done.

□

Bibliography

[16] *Handbook of Computational Social Choice*. Cambridge University Press, 2016.

[AT08] A. Altman and M. Tennenholtz. "Axiomatic Foundations for Ranking Systems". In: *Journal of Artificial Intelligence Research* 31 (Mar. 2008), pp. 473–495. ISSN: 1076-9757. DOI: 10.1613/jair.2306. URL: http://dx.doi.org/10.1613/jair.2306.

[Bal+18] David Balduzzi et al. *Re-evaluating Evaluation*. 2018. arXiv: 1806.02643 [cs.LG]. URL: https://arxiv.org/abs/1806.02643.

[BW26a] Damien Berriaud and Roger Wattenhofer. *Breaking the Illusion of Artificial Consensus: Clone-Robust Weighting for Arbitrary Metric Spaces*. 2026. arXiv: 2602.24024 [cs.GT].

[BW26b] Damien Berriaud and Roger Wattenhofer. "Clone-Robust Weights in Metric Spaces: Handling Redundancy Bias in Benchmark Aggregation". In: *Proceedings of the 25th International Conference on Autonomous Agents and Multiagent Systems*. AAMAS '26. Paphos, Cyprus: International Foundation for Autonomous Agents and Multiagent Systems, 2026, pp. 2270–2278. ISBN: 9798400723179. DOI: 10.65109/MJWJ6521. URL: https://doi.org/10.65109/MJWJ6521.

[CF06] Alice Cheng and Eric Friedman. "Manipulability of PageRank under Sybil Strategies". In: NetEcon, 2006.

[Dou02] John R. Douceur. "The Sybil Attack". In: *Peer-to-Peer Systems*. Ed. by Peter Druschel, Frans Kaashoek, and Antony Rowstron. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002, pp. 251–260. ISBN: 978-3-540-45748-0.

[EFS14] Edith Elkind, Piotr Faliszewski, and Arkadii Slinko. "Cloning in elections: Finding the possible winners". In: (Jan. 2014). arXiv: 1401.4602 [cs.GT].

[PSZ25] Ariel D. Procaccia, Benjamin Schiffer, and Shirley Zhang. *Clone-Robust AI Alignment*. 2025. arXiv: 2501.09254 [cs.LG]. URL: https://arxiv.org/abs/2501.09254.

[Tid87] T. N. Tideman. "Independence of clones as a criterion for voting rules". en. In: *Social Choice and Welfare* 4.3 (Sept. 1987), pp. 185–206. ISSN: 1432-217X. DOI: 10.1007/BF00433944. URL: https://doi.org/10.1007/BF00433944 (visited on 05/06/2025).