Socioeconomic Clustering and Racial Segregation on Lattices with Heterogeneous Sites

Zhanzhan Zhao\textsuperscript{*1} and Dana Randall\textsuperscript{†1}

\textsuperscript{1}School of Computer Science, Georgia Institute of Technology.

\textbf{Abstract}

The Schelling model of segregation moves colored agents (or particles) on the vertices of a graph (typically a lattice), with unhappy agents trying to move to new positions if the number of their neighbors with a different color exceeds some threshold. A well-studied stochastic variant of this model assumes that agents are geometrically happier with each additional neighbor of the same color, moving with probability proportional to their unhappiness. The stochastic version behaves like a fixed magnetization Ising model, so it is known that we will have segregation at equilibrium if the homophily, or preference for the neighbors of the same color, is large enough.

In this work, we consider race and socioeconomic status simultaneously to understand how carefully architected placement of urban infrastructure that mostly helps low-income individuals might affect segregation. In our model, agents are both two-colored and also labeled as rich or poor. We designate certain vertices on the graph as “urban sites,” providing civic infrastructure that most benefits the poorer population, incentivizing occupation of these vertices by poor agents of either color. Infrastructure that is centralized, like a city center or mall, encourages poor agents to cluster centrally in addition to their homophily preferences, while infrastructure that is well distributed, like a large grid of inner-city bus routes, tends to disperse the low-income agents. We ask what effect these two scenarios have on segregation.

We find that centralized infrastructure simultaneously causes segregation and the “urbanization of poverty” (i.e., occupation of urban sites primarily by poor agents) when the homophily and incentives drawing the poor to urban sites are large enough. However, even when homophily preferences are very small, as long as the incentives drawing the poor on the urban sites is large, under income inequality where one race has a significantly higher proportion of the poor, we get racial segregation on urban sites but integration on non-urban sites. However, we find there is an overall mitigation of segregation when the urban sites are distributed throughout the lattice and the incentive drawing the poor on the urban sites exceeds the homophily preference. We prove that in this case, no matter how strong homophily preferences are, it will be exponentially unlikely that a configuration chosen from stationarity will have large, homogeneous clusters of similarly colored agents, thus promoting integration in the city with high probability.

\textsuperscript{*}zhanzhan@gatech.edu. Supported by Georgia Tech ARC-TRIAD Fellowship and ARO MURI award W911NF-19-1-0233.

\textsuperscript{†}randall@cc.gatech.edu. Supported by ARO MURI award W911NF-19-1-0233.
1 Introduction

Over fifty years ago, Nobel prize winning economist Thomas Schelling studied segregation by modeling residents as colored particles on a chessboard. Each particle is considered happy if its color agrees with more than a fixed fraction of its neighbors and tries to move to a new site if it is not content [16]. Simulations reveal the surprising fact that even mild biases favoring one’s own color, known as homophily, is sufficient to cause segregation on a macroscopic scale [29]. Extensive work has been done by economists and sociologists to expand and explain Schelling’s model using statistical analyses and simulation tools [2, 18, 20, 33]. This work typically focuses on how the dynamics determine the limiting distribution and connecting the model to the real world population dynamics [8–10,19,25,30].

The first rigorous analyses of these dynamics appears in [5,12], where the authors consider the one dimensional Schelling model with two different races. For a set of agents $I$ living in a ring, the utility function is a simple threshold function computed as follows. There is a critical value $\theta \in [0,1]$. For each agent $j \in I$, if the fraction of the neighboring agents with the same race as that of $j$ is more than $\theta$ then $j$ is happy. Otherwise, $j$ is unhappy. The convergence of the dynamics is guaranteed with the stationary configuration being proved. Further rigorous work has been done by extending the neighborhood interactions to a broad class of utility functions. The exponential mixing time and the segregated stationary configurations are proved under certain preference thresholds [1,26,32].

A natural generalization of Schelling’s original model can be described as follows. A set of agents $I$ occupy some of the vertices of a graph $G = (V,E)$. Each agent $j \in I$ belongs to a race $R_j \in R$. Each configuration $\sigma$ of the system is a placement of agents on the vertices of the graph. Also, for $j$ there is the utility function $u_j : R^{N_\sigma(j)} \rightarrow \{\text{“happy”, \ “unhappy”}\}$, where $N_\sigma(j)$ is a small neighborhood around $j$ in $G$, which specifies the happiness of $j$. It is direct that given the utility functions and the configuration we can specify which agents are happy and which are not. In Schelling’s model, each agent is content or not depending on whether the number of like neighbors exceeds a threshold, while in the particle model each agent’s happiness increases incrementally with more neighbors of the same spin [31].

Bhakta et al. [3] considered a reversible variant of the Schelling dynamics using tools from statistical physics. Each configuration $\sigma$ has a Gibbs measure based on the happiness of each agent, parameterized by $\lambda$, the racial bias parameter. Specifically, each agent $j$ contributes the number of same-colored neighbors $N_\sigma(j)$ to the Hamiltonian, or energy of the system. The probability of a configuration is given by $\pi(\sigma) = \prod_{j \in I} \sqrt{X^{N_\sigma(j)}} / Z$, where $Z$ is the normalizing constant. A reversible Markov chain can be constructed that converges to this Gibbs distribution in which every agent has some probability of moving, and those with more neighbors of the other color move with increasing probability. The dynamics are similar to the well-studied Ising model from statistical physics, and the model inherets many of its properties such as a phase transition from a disordered phase to an ordered one when $\lambda$ surpasses a threshold.

The Schelling model and its variants typically assume all agents of each race are homogeneous and have the same homophily, regardless of location. However, it has been observed that the Schelling model can be used to explain other social phenomena where like individuals are inclined to prefer similar individuals over those who differ from them, including an empirical study of both segregation of ethnicity and wealth [11,24,28].

Here we explore the impact of wealth disparity and potential civic interventions to mitigate segregation, taking into account two widely observed phenomena: centralization, or the degree to which a group is spatially located near the the city center, and urbanization of poverty, whereby city centers and other areas dense with civic infrastructure disproportionately attract the poor.
Centralization is widely-used to measuring segregation in metropolitan areas [17, 22]. Urban economists show that urbanization of poverty results from better access to public transportation in central cities and other resources [13].

**Segregation by race and wealth:** To better understand these socioeconomic distinctions and the effects of economic disparity within a city, we introduce a new *Socioeconomic Schelling Model*, where individuals of each race are also designated *rich* or *poor*. We label some vertices to be urban sites if they provide useful infrastructure (or resources) that are most beneficial to poor citizens. For instance, the urban sites can be grouped centrally, representing a metropolitan city center, or distributed in a large spanning grid, representing a network of public transportation (see Fig 1). While all individuals have uniform homophily preferences, like the standard Schelling model, we add additional incentives that benefit poor people residing on urban sites, so more racial intolerance can be justified if more of the poor people are benefitting from proximity to urban sites. We are interested in understanding when urban infrastructure can help mitigate racial biases and lessen segregation for various typologies of urban sites.

Specifically, we represented the city by a finite torus on the triangular lattice, with each site accommodating exactly one person. Each person (or agent) is either blue or red, representing race, and *rich* or *poor*, representing wealth and \( U \subseteq V \) are the vertices that are designated urban sites. Each pair of neighbors has a homophily (or racial) bias \( \lambda \), representing how much they each prefer neighbors of their own color. Setting \( \lambda > 1 \) corresponds to the “ferromanetic setting,” where agents prefer same-colored neighbors. Further, poor agents have an affinity for urban sites with a wealth bias parameter \( \gamma \), in addition to their homophily biases. Setting \( \gamma > 1 \) biases poor agents to preferring residing on urban sites. When \( \gamma = 1 \), we recover the pure homophily model where wealth of individuals is not considered. The stationary probability of any configuration \( \sigma \) is given by

\[
\pi(\sigma) = \lambda^{-h(\sigma)}\gamma^{p(\sigma)}/Z,
\]

where \( h(\sigma) \) is the number of racially heterogeneous edges (whose endpoints do not share the same color), \( p(\sigma) \) is the number of poor agents on urban sites, and \( Z \) is the normalizing constant.

A randomized algorithm \( M \) for sampling from \( \pi \) can be described as follows. At each time step, two random agents are selected, and they swap locations with the appropriate Metropolis probabilities so as to converge to \( \pi \). In particular, they are more likely to swap if they are each in less homogeneous neighborhoods, as previously studied in [3, 6], with an additional bias toward keeping poor agents on urban sites, so happier individuals are less likely to move.

We are interested in how these dynamics might be influenced by the placement of urban sites based on the racial and wealth demographics of the population. We note that when there are no urban sites (or all vertices are urban sites), then the wealth of individuals becomes irrelevant and we recover the racial segregation model studied in [6]. When it comes to the size and topology of the urban sites, the case of no urban-sites at all or full urban sites have been studied in [6], where the dichotomy of the phase change between integration and segregation has been proved. Here we are interested in the effects in heterogeneous cases where both urban and non-urban sites are present. We also require the size of the urban sites to be of a constant fraction of the total sites. For topology, we study the impact of the centralized or distributed placement of the urban sites to segregation.

**Our results:** First, we show that our model yields urbanization of poverty when \( \gamma \) is sufficiently large, with all but an arbitrarily small fraction of urban sites being occupied by poor agents. When the urban sites are centralized (i.e., occupying a dense hexagon in the lattice), the urban sites will be densely occupied with poor agents for any value of \( \lambda \) as long as \( \gamma \) is large enough. Moreover, we
will see a similar urbanization of poverty even for arbitrarily distributed urban sites as long as the wealth bias $\gamma$ exceeds the racial bias $\lambda$ sufficiently. In the centralized case $\lambda$ does not play a role, while in the distributed case we need the incentive for the poor to be on urban sites to outweigh their racial biases. Conversely, we show that for any $\lambda > 1$, if $\gamma > 1$ is small enough, then it is exponentially unlikely that poor agents will be concentrated on urban sites for any constant fraction of poor agents in the city.

Moreover, when the urban sites are centralized and both $\lambda$ and $\gamma$ are large enough, urbanization of poverty and racial segregation will occur simultaneously. However, even when $\lambda$ is small, as long as $\gamma$ is large, under severe income inequality where most of the poor are from one race, we will have racial segregation on urban sites and racial integration on the non-urban sites. This suggests that the urbanization of poverty can enhance segregation when the infrastructure is centralized, such as with a dense city center with civic services and perhaps subsidized housing, providing a primary location that incentivizes occupation by the poor.

However, we find a dramatically different outcome when the urban sites are distributed throughout the city, such as with public transportation evenly partitions the entire city. First, we prove under income inequality, where one race has a higher proportion of poor people, no matter how large $\lambda$ is, as long as $\gamma$ exceeds $\lambda$ sufficiently, both the urban and nonurban sites will be integrated. That is, the probability of large clusters that are predominantly one race forming is exponentially small. This suggests that distributing urban infrastructure equitably throughout the city will have a better effect for mitigating segregation when the incentives are large enough compared to the inherent racial biases.

**Techniques:** Our proofs build on analysis for the integration and separation of heterogeneous particles in the context of programmable matter [6]. Although not formulated as segregation, their proofs show that when there are no urban sites, slight or no local homophily produces integration, whereas stronger homophily produces segregation. The proofs use mappings known as Peierls arguments to show that configurations not in the target sets are exponentially unlikely.

Peierls arguments were developed in physics to prove the existence of phase transitions, such as the sudden change in magnetization in the two-dimensional Ising model [14, 15], and have been applied in computer science to study slow mixing of Markov chains [4, 27]. The essential idea is to map the set not satisfying a target property to configurations that have exponentially larger probability at stationarity so that the inverse maps do not require significant information, thus proving that configurations outside of the target set must have small probability by evaluating “energy/entropy” balancing the probabilities and the number of preimages.

These techniques form the backbone of our proofs when the racial bias is high and we get segregation, but the introduction of urban sites and the wealth bias make the proofs significantly more challenging as we need to maintain the same number of rich and poor members of both races before and after the mapping $\nu = f(\sigma)$. When there are only two types of agents, red and blue, we are able to allow the sizes to deviate and restore the appropriate numbers of each race in a designed way to increase the weight significantly through the map. Here, all four groups may deviate under the maps and it requires careful arguments to be able to restore the cardinalities of all the sets without losing too much information. In particular, we extend Peierls and bridging arguments used in [6, 23] to suit the multi-dimensional information about race and wealth embedded in each agent accounting for more complex tradeoffs between the homophily and wealth biases, and the number of configurations with each weight, requiring much more delicate mappings and counts. We note that in addition to the typical energy-entropy tradeoffs of Peierls arguments, here we additionally have to balance several tradeoffs within the energy term $\frac{\pi(\sigma)}{\pi(\nu)}$, due to the different relationships between wealth and racial biases induced by different placements of the urban sites.
2 The Socioeconomic Schelling Model

In our proposed model, a city is represented by a finite toroidal region of the triangular lattice $G_\Delta = (\mathcal{V}, \mathcal{E})$, shown in Figure 1a. Each vertex in $\mathcal{V}$ represents a potential residence or site. Two adjacent vertices are neighboring sites, and each site has six nearest neighbors on $G_\Delta$. Some vertices $\mathcal{U} \subseteq \mathcal{V}$ are designated urban sites. We denote the set of agents as $\mathcal{A}$ and the poor agents as $\mathcal{P} \subseteq \mathcal{A}$. Figure 1a shows an example of centralized placement of urban sites, whereas Figure 1b shows the distributed placement, with urban sites depicted as yellow hexagons.

We assume each agent $i$ is assigned a race $r(i) \in \{\text{blue, red}\}$ and wealth $w(i) \in \{\text{rich, poor}\}$. Each site in $\mathcal{V}$ can accommodate at most one agent. For simplicity of analysis, we assume that $n$ agents fully occupy all the sites on $G_\Delta$, where $|\mathcal{V}| = n$. The size of the urban sites is assumed to be of a constant fraction of all the sites, i.e., $|\mathcal{U}| = c \cdot n$, where $c \in [0, 1]$. As shown in Figure 1c, we represent the race of an agent by color and the wealth of an agent by the shade of each color; poor blue agents are referred to as cyan, poor red agents are pink and blue and red are reserved for the rich members of each color class.

Among the $n$ agents, $\mathcal{P}$ are the subset that are poor and the fraction that are poor is denoted by $p$, so $|\mathcal{P}| = p \cdot n$. Similarly, the fraction of the red $\mathcal{R}$ is $r$, and $|\mathcal{R}| = r \cdot n$. Among the red agents, we further denote the fraction of the poor red $\mathcal{R}_p$ as $r_p$, and the fraction of rich red $\mathcal{R}_r$ as $r_r$, so that $|\mathcal{R}_p| = r_p n$, and $|\mathcal{R}_r| = r_r n$. Similarly, we define the fraction of the blue $\mathcal{B}$ as $b$, and the fraction of poor blue $\mathcal{B}_p$ as $b_p$, and the fraction of rich blue as $\mathcal{B}_r$ as $b_r$.

A configuration (or a state) $\sigma$ is the set of vertices of $G_\Delta$ occupied by an arrangement of $n$ agents with the urban sites assignment information embedded, along with the race and wealth of each agent. The state space (or configuration space) $\Omega$ is the set of all possible configurations.

For a configuration $\sigma$, we denote agent $i$ resides at a site $\ell_\sigma(i) \in \mathcal{V}$. Agents living at neighboring sites are neighbors and each agent can at most have six neighbors. Each agent $i$ is assigned a race $r(i)$, wealth $w(i)$, and occupies a site $\ell_\sigma(i)$, which it can recognize as an urban site or not. We define an indicator function that takes agent $i$ as input and outputs true when $i$ is poor and currently on the urban sites as the following:

$$u_\sigma(i) \equiv \begin{cases} 1, & \text{if } i \in \mathcal{P}, \ell_\sigma(i) \in \mathcal{U} \\ 0, & \text{otherwise} \end{cases}$$

For a configuration $\sigma$, the number of agents that are poor and on the urban sites is defined to be $p(\sigma) \equiv \sum_{i \in \mathcal{A}} u_\sigma(i)$.

Each agent is able to identify each neighbor’s race, and thus can determine the number of neighbors that share its color, $N_\sigma(i)$. An edge in a configuration $\sigma$ with vertices occupied by agent $i$ and $j$ is racially homogeneous if they are the same color (i.e., $r(i) = r(j)$) and racially heterogeneous otherwise. We define the total number of racially homogeneous edges of a configuration $\sigma$ as $h(\sigma)$, and the total number of racially homogeneous edges as $e(\sigma)$.

The Markov chain $\mathcal{M}$ is defined so that it will converge to $\pi(\sigma) = \lambda^{-h(\sigma)} \gamma^{p(\sigma)} / Z$, which generalizes the Schelling probabilities to reflect the additional effect of urban sites. Each agent $i$ is able to swap locations with any agent $j \in \mathcal{A}, j \neq i$ in the city $G_\Delta$, and we denote it a swap move $s_{ij}$. Beginning at any configuration $\sigma_0 \in \Omega$, at each time step, given the current configuration $\sigma$, the algorithm randomly picks two agents $i$ and $j$ at sites $\ell_\sigma(i) \in \mathcal{V}$ and $\ell_\sigma(j) \in \mathcal{V}$ and tries to swap their positions with the appropriate Metropolis probabilities (so agents are more likely to move if they have fewer racially homogeneous neighbors, and with a dampening factor $\frac{1}{\gamma} < 1$ if the agent is poor and currently at the urban site. Mathematically,

$$P(\sigma : i \rightarrow j) = \frac{\lambda^{-N_\sigma(i)} \gamma_u}{\gamma u_\sigma(i)},$$
Figure 1: A city lattice region $G_{\triangle}$ (a) with centralized urban sites, (b) with distributed urban sites, and (c) fully occupied by agents.

where $\lambda > 1$, and $\gamma > 1$. The probability of agents $i$ and agent $j$ swapping positions satisfies

$$P(\sigma : s_{ij}) = \frac{1}{n^2} \lambda^{-N_\sigma(i) - N_\sigma(j)} \gamma^{-\sum_{k \in \{i,j\}} u_\sigma(k)}.$$  \hfill (1)

Algorithm 1 Markov Chain $\mathcal{M}$.

1: Beginning at any configuration $\sigma_0$ with $n$ agents, repeat:
2: Choose two agents $i$ and $j$ uniformly at random in the current configuration $\sigma$.
3: Choose $q \in (0, 1)$ uniformly at random.
4: if $q < \lambda^{-N_\sigma(i) - N_\sigma(j)} \gamma^{-\sum_{k \in \{i,j\}} u_\sigma(k)}$, then then agents $i$ and $j$ swap positions.
5: else agents $i$ and $j$ keep their current locations.

It is easy to see that the Markov chain $\mathcal{M}$ is ergodic on the state space $\Omega$, since swap moves of $\mathcal{M}$ suffice to transform any configuration to any other configuration (irreducible) and there is a non-zero self-loop probability for $\lambda > 1$ and $\gamma > 1$ (aperiodic). Using detailed balance it is easy to confirm that the Markov chain converges to

$$\pi(\sigma) = \lambda^{-h(\sigma)} \gamma^{p(\sigma)} / Z,$$ \hfill (2)

where $h(\sigma)$ is the number of racially heterogeneous edges in $\sigma$, $p(\sigma)$ is the number of the poor on the urban sites in $\sigma$, $Z = \sum_{\sigma \in \Omega} \lambda^{-h(\sigma)} \gamma^{p(\sigma)}$ is the partition function, and $\Omega$ is the configuration space.

3 Urbanization of Poverty

We start by showing the urbanization of poverty, defined as the urban sites being occupied by nearly the maximal number of the poor agents, with an $\epsilon$--fraction of error tolerance, whether the urban sites are centralized or distributed. We prove in Theorem 3 that under centralized placement, for any $\lambda$ and $\epsilon$, as long as $\gamma$ is bigger than a threshold depending only on $\epsilon$, we will observe the urbanization of the poor at stationarity with high probability. See Figure 2a for simulations. We prove in Corollary 4 that when urban sites are distributed throughout the lattice and the contribution of the wealth bias $\gamma$ exceeds the contribution of the racial bias $\lambda$, we will also see the urbanization of poverty (simulations in Figure 2b). Finally, we prove that when urban sites are centralized, and for any $\lambda > 1$, if $\gamma > 1$ is sufficiently small, then we are very unlikely to observe the urbanization of poverty and both urban and nonurban sites will have a nontrivial mix of rich and poor agents.
**Definition 1.** For any $\epsilon \in (0, \frac{1}{2})$, a city is said to have $\epsilon$–urbanization of poverty if the number of the poor on the urban sites is at least $\min\{c, p\}n - \epsilon n$.

Here, $\epsilon$ expresses the tolerance for having rich agents on the urban sites: smaller $\epsilon$ requires stricter urbanization of poverty, whereas larger $\epsilon$ allows for smaller density of the poor on the urban sites. $\min\{c, p\}n$ refers to the maximal possible number of the poor on the urban sites. $\epsilon$–urbanization of poverty can be understood as the urban sites will be accommodated by the maximal possible number of the poor minus a small error tolerance of no more than $\epsilon n$.

**Remark 2.** When the size of the urban sites can roughly accommodate all the poor, where $p - m\epsilon < c < p + (m + 1)\epsilon$, a configuration satisfying the above definition of $\epsilon$–urbanization of poverty can also be viewed as having $(\beta, (m + 1)\epsilon)$–wealth segregation: there are at most $\beta\sqrt{n}$ heterogeneous edges in terms of wealth, where $\beta > 4\sqrt{n}$; and the number of poor in the urban area $\mathcal{U}$ is at least $pn - (m + 1)\epsilon n$.

We prove that with centralized urban sites or distributed urban sites, for large enough $\gamma$, we will most likely observe the urbanization of poverty at stationarity in Theorem 3 and Corollary 4.

**Theorem 3 (Centralized Urbanization of Poverty).** If $\gamma > 16\frac{3(2n+2)}{\epsilon^2} + 1$ and $\lambda > 1$, with the centralized urban sites, when $n$ is sufficiently large, then for $\mathcal{M}$, configurations drawn from distribution $\pi$ have $\epsilon$–urbanization of poverty with probability at least $1 - \xi_1^n$, where $0 < \xi_1 < 1$.

First, we define $\Omega_{\text{urb}}$ to be the set of the configurations that do not have $\epsilon$–urbanization of poverty. To show Theorem 3, it suffices to show $\pi(\Omega_{\text{urb}}) \leq \xi_1^n$. Then we use Peierls argument (see Appendix A.1 for a short tutorial): mapping from non-urbanized configurations to urbanized configurations with the bridging technique shows that the map has an exponential gain in probability weight. Thus the non-urbanized configurations are exponentially unlikely compared with urbanized ones, even though the number of non-urbanized configurations is much more than the urbanized ones, and some of those configurations have large probability weights in terms of $\lambda^{-h(\sigma)}$.

Compared with [6, 23], the state space is enlarged by adding the wealth dimension, which requires modifying the bridge system to encode both race and wealth information of each agent, defined as a $\delta$–color-and-richness bridge system specified in Appendix A.2. Moreover, compared with [6, 23], due to the new term about $p(\sigma)$ in the stationary distribution (2), besides the more complex energy-entropy tradeoffs between $\frac{\pi(\sigma)}{\pi(\nu)}$ and $|f^{-1}(\nu)|$, now there is another tradeoff inside the energy term (energy-energy tradeoff): because a mapped configuration with big probability weight of $\gamma^{p(\sigma) - p(\nu)}$ can correspond to small probability weight of $\lambda^{-h(\sigma) - h(\nu)}$. We design more careful mapping rules to balance and bound $\frac{\pi(\sigma)}{\pi(\nu)}$ and $|f^{-1}(\nu)|$, and see Appendix A.3, A.4 and A.5 for the mapping technique summaries.

**Proof of Theorem 3.** For any $\sigma \in \Omega_{\text{urb}}$, we first construct a $\delta$–color-and-richness bridge system (see Appendix A.2 for definition and Figure 5a for illustration) and define the mapping $f(\sigma) = (f_3 \circ f_4 \circ f_3 \circ f_2 \circ f_1)(\sigma)$, where $\psi = f_1(\sigma)$ is the richness inversion mapping (defined in Appendix A.3 and see Figure 5b for illustration), and $\tau = f_2(\psi)$ is the color inversion mapping (defined in Appendix A.3 and see Figure 5c for illustration). $\tau = (f_2 \circ f_1)(\sigma)$ eliminates the bridged racially heterogeneous edges and the bridged poor agents. For $(f_3 \circ f_4 \circ f_3)(\tau)$ (defined in Appendix A.4), we first assume the urban sites are centralized, under which we recover the same ratios of each color and richness as in $\sigma$ in the centralized way defined in Appendix A.5 (also see Figure 6 for illustrations). Then the upper bounds of $h(\nu) - h(\sigma) \leq 3\alpha\sqrt{n} - z_c$ and $p(\sigma) - p(\nu) \leq -\delta n$ and $|f^{-1}(\nu)| \leq (z_c + 1)^{9\alpha\sqrt{n} 4^{\frac{1+3\epsilon}{3n}}}(z_c + 3n)$ can be obtained from Claim 26 and 27 which are shown in
Appendix B. The color contour length \( z_c \) is defined in the bridge system (Appendix A.2), which is the sum of length of the contours separating the red (or pink) from the blue (or cyan) in \( \pi \) with a no more than \( \delta \)-fraction omission. Finally, substituting (2) and the bounds into the Peierls argument yields

\[
\pi(\Omega_{\text{urb}}) = \sum_{\sigma \in \Omega_{\text{urb}}} \pi(\sigma) \leq \sum_{\nu \in \Omega} \sum_{\sigma \in f^{-1}(\nu)} \pi(\sigma) \leq \sum_{\nu \in \Omega} \sum_{\sigma \in f^{-1}(\nu)} \pi(\nu) \frac{\pi(\sigma)}{\pi(\nu)} \\
\leq \sum_{\nu \in \Omega} \pi(\nu) \sum_{\sigma \in f^{-1}(\nu)} \lambda^h(\nu) - h(\sigma) \gamma^p(\sigma) - p(\nu) \\
\leq \sum_{\nu \in \Omega} \pi(\nu) \sum_{z_c = \sqrt{p \cdot n}}^{3n} n^6(3n + 1)(z_c + 1)9^\alpha \sqrt{n} \left( \frac{4^{344+1}}{\lambda} \right)^\frac{z_c}{\gamma^4} \left( \frac{64^{344+1}}{\gamma^4} \right)^n, \tag{3}
\]

where \( z_c \leq 3n \) is because the color contour length is upper bounded by the sum of all edges of \( G_\Delta \), and \( z_c \geq \sqrt{p \cdot n} \) is due to the triangular lattice geometry, which is proved in Lemma 2.1 in [7].

If \( \lambda \geq 4^{344+1} \), as long as \( \gamma^4 \lambda^3 > 4^{344+1} \), the sum will be exponentially small for sufficiently large \( n \). Or if \( 1 \leq \lambda < 4^{344+1} \), the sum further yields \( \pi(\Omega_{\text{urb}}) \leq n^6 \cdot (3n + 1) \cdot 9^\alpha \sqrt{n} \cdot 3n \cdot \left( \frac{16^{344+1}}{\lambda^2 \gamma^4} \right)^3n \). As long as \( \gamma^4 \lambda^3 > 16^{344+1} \) \( \geq 16^{344+1} / \lambda \), the sum will still be exponentially small for sufficiently large \( n \). Combining the two cases, we can see that as long as \( \gamma^4 \lambda^3 > 16^{344+1} \) and \( \lambda > 1 \), \( \pi(\Omega_{\text{urb}}) \leq \xi^n \), for \( \xi_1 \in (0, 1) \). Substituting \( \delta = \epsilon/2 \) into \( \gamma^4 \lambda^3 > 16^{344+1} \) yields Theorem 3.

In the above theorem, under the centralized urban sites placement, to realize the urbanization of poverty, we find that for \( \lambda \) in certain threshold, it suffices to have \( \lambda \cdot \gamma^4 \lambda^3 > 16^{344+1} \), where the wealth bias and the racial bias both help the urbanization of poverty. Instead, in the following corollary where urban sites are distributed, we find the competing relationship between \( \gamma \) and \( \lambda \), where when \( \gamma \) is larger than \( \lambda \) to a certain extent, the urbanization of poverty is achieved.

**Corollary 4 (Distributed Urbanization of Poverty).** If \( \gamma > 4^{344+1} \cdot \frac{1}{\pi^4} \cdot \max \{ \lambda^6 / \epsilon, 4^{344+1} \} \) and \( \lambda > 1 \), with the distributed urban sites, when \( n \) is sufficiently large, then for \( \mathcal{M} \), configurations drawn from distribution \( \pi \) have \( \epsilon \)-urbanization of poverty with probability at least \( 1 - \xi_1^n \), where \( 0 < \xi_1 < 1 \).

**Proof of Corollary 4.** The bridge system and mapping are the same as the proof of Theorem 3 except that: for \( (f_3 \circ f_1 \circ f_3)(\tau) \), we recover the same ratios of each color and richness as in \( \pi \) in the distributed way defined in section A.5. Claim 27 still holds true. For Claim 26, it now follows from Lemma 24 and 19 that \( h(\nu) - h(\sigma) \leq 2cn - z_c \leq 3n - z_c \), and \( p(\sigma) - p(\nu) \leq -\delta n \) holds true, where \( \delta = \epsilon/2 \). Substituting the bounds into the Peierls argument yields

\[
\pi(\Omega_{\text{urb}}) \leq \sum_{\nu \in \Omega} \pi(\nu) \sum_{z_c = \sqrt{p \cdot n}}^{3n} n^6(3n + 1)(z_c + 1)9^\alpha \sqrt{n} \left( \frac{4^{344+1}}{\lambda} \right)^\frac{z_c}{\gamma^4} \left( \frac{64^{344+1}}{\gamma^4} \right)^n. \tag{4}
\]

Similarly, when \( \lambda \geq 4^{344+1} \), if \( \gamma^4 \lambda^3 > \lambda \cdot 4^{344+1} \), the sum will be exponentially small for sufficiently large \( n \). When \( 1 \leq \lambda < 4^{344+1} \), the sum further yields \( \pi(\Omega_{\text{urb}}) \leq n^6 \cdot (3n + 1) \cdot 9^\alpha \sqrt{n} \cdot 3n \cdot \left( \frac{16^{344+1}}{\lambda^2 \gamma^4} \right)^3n \). As long as \( \gamma^4 \lambda^3 > 16^{344+1} \), the sum will still be exponentially small for sufficiently large \( n \). Combining the two cases, we can see that as long as \( \gamma^4 \lambda^3 > 4^{344+1} \cdot \max \{ \lambda, 4^{344+1} \} \) and \( \lambda > 1 \), \( \pi(\Omega_{\text{urb}}) \leq \xi^n \), for \( \xi_1 \in (0, 1) \). Substituting \( \delta = \xi / 2 \) the lower bound of \( \gamma \) yields Corollary 4. \( \square \)
To complement Theorem 3, we further prove that for any $\lambda > 1$, if $\gamma > 1$ but is smaller than a threshold, it is exponentially unlikely we will observe urbanization of poverty at stationary under certain demographic parameter choices. See the proof details in Appendix 5.

Theorem 5 (Dispersion of Poverty). Given the centralized urban sites, $p < c < p + \epsilon$, $r_p < r_l - \delta$ and $b_p < b_l - \delta$, for any $\lambda > 1$, if $1 < \gamma < \left(\frac{r_p}{b_p}\right)^{\frac{1}{\delta}} \left(\frac{b_p}{r_p}\right)^{\frac{\delta}{2}}$, when $n$ is sufficiently large, then for $M$, configurations drawn from distribution $\pi$ have $\epsilon-$urbanization of poverty with probability at most $\xi_2^3$ for some constant $0 < \xi_2 < 1$ and $\delta = \frac{\epsilon}{2}$.

4 Urbanized Racial Segregation

In this section, we prove urbanized racial segregation, where major racial clusters will be formed, and the race that is predominantly poor will predominantly occupy the urban area. We begin by formally define $(\beta, \delta)-$racial segregation as follows

Definition 6. For $\beta > 4\sqrt{\tau}$ and $\delta \in (0, \frac{1}{2})$, a city configuration $\sigma$ is said to be $(\beta, \delta)-$racial segregated if there is a subset of agents $R$ such that:

- there are at most $\beta\sqrt{n}$ racially heterogeneous edges of $\sigma$ with exactly one endpoint in $R$;
- the number of red agents in $R$ is at least $\tau n - \delta n$.

The parameter $\delta$ is the tolerance for having agents of the wrong color within the red region $R$, with smaller $\delta$ corresponding to a stricter segregation class. If one color class has fewer than $\delta n$ agents, then the entire configuration space will be $(\beta, \delta)-$segregated, with $R = \emptyset$, or $\bar{R} = \emptyset$. We require that each color class has more than $\delta n$ agents and, accordingly, we need $\delta < 1/2$. The parameter $\beta$ in the definition controls how small the boundary is between the red region $R$ and the rest of the configuration, and the minimal value $4\sqrt{\tau}$ corresponds to the case where the red forms a homogeneous hexagon cluster.

Then urbanized segregation is defined as the union of Definition 6 and Definition 1 as follows.

Definition 7. For centralized urban sites, $\beta > 4\sqrt{\tau}$, and $\epsilon \in (0, \frac{1}{2})$, a city has $(\beta, \epsilon)-$urbanized racial segregation if it is $(\beta, \epsilon)-$racially segregated and have $\epsilon-$urbanization of poverty.

Like Definition 6, $\beta$ here controls how small the boundary is between the red region $R$ and the rest of the city. $\epsilon$ expresses the tolerance for having agents of the wrong color within the monochromatic color regions and having rich agents on the urban sites. In the following theorem, we show that for large enough $\lambda$ and $\gamma$, then with high probability, $M$ leads to urbanized segregation. See Figure 2c for simulated visualizations.

Theorem 8 (Urbanized Racial Segregation). With centralized urban sites, $\lambda > 3^{3/4} 4^{3/4+1/45}$, $\gamma^{3/\beta} > 4^{3/4+1/45}$, and $n$ sufficiently large, configurations from $M$ drawn from distribution $\pi$ have $(\beta, \epsilon)-$urbanized segregation with probability at least $1 - \xi^{\sqrt{n}}$ for some constant $0 < \xi < 1$, and $\delta = \frac{\epsilon}{2}$.

Proof of Theorem 8. First, we define $U_{\beta, \epsilon} \subset \Omega$ to be the configurations that have $(\beta, \epsilon)-$urbanized segregation. To prove Theorem 8, it suffices to prove $\pi(\Omega \setminus U_{\beta, \epsilon}) \leq \xi^{\sqrt{n}}$, where $\xi \in (0, 1)$. We can further divide $\Omega \setminus U_{\beta, \epsilon}$ into two parts: $\Omega_{\text{urb}}$ that do not have $\epsilon-$urbanization of poverty, and $\Omega_{\text{urb} \land \text{seg}}$ that have $\epsilon-$urbanization of poverty and do not have $(\beta, \epsilon)-$segregation. Thus it suffices to prove $\pi(\Omega_{\text{urb}}) \leq \xi_0^n$ and $\pi(\Omega_{\text{urb} \land \text{seg}}) \leq \xi^{\sqrt{n}}_0$, for $0 < \xi_1, \xi_0 < 1$. It follows from the proof of Theorem 3 that if $\lambda \geq 4^{3/4+1/45}$, as long as $\gamma^{3/\beta} > 4^{3/4+1/45}$, $\pi(\Omega_{\text{urb}}) \leq \xi^n_0$. It is proved in Claim 31
in Appendix D that if \( \lambda > 3\pi \frac{3\pi + 4}{4\pi} \), \( \pi(\Omega_{\text{urb} \land \neg \text{seg}}) \leq \xi_0^{\sqrt{\pi}} \) for some \( \xi_0 \in (0, 1) \). Combining the two parts, to have \( \pi(\Omega \setminus S_{\beta, \epsilon}) \) to be exponentially small for large \( n \), it suffices to have \( \lambda > 3\pi \frac{3\pi + 4}{4\pi} \) and \( \gamma^{\delta/3} > 4 \frac{3\pi + 4}{4\pi} \).

To complement Theorem 8, in the following theorem, we prove that for large enough \( \gamma \) but \( \lambda > 1 \) smaller than a threshold, we will most likely observe urbanization of poverty but racial integration outside the urban area under certain demographic parameter choices. The proof technique is very similar to the proof of Theorem 5. See Appendix E for proof details. A special case of Theorem 9 is shown in Remark 10, where segregation of the poor red is observed in the urban area with racial integration outside. See Figure 2d for simulated visualizations.

**Theorem 9 (Coexistence of Urbanization and Racial Integration).** With the centralized urban sites, for the demographics choices such that \((\frac{p-\delta}{b_p})(\frac{1-p-\delta}{b_1})^r > \beta \), if \( 1 < \lambda^3 < (\frac{p-\delta}{b_p})(\frac{1-p-\delta}{b_1})^r \), and \( \gamma^{\delta/3} > 16 \frac{3\pi + 4}{4\pi} \), when \( n \) is sufficiently large, then for \( \mathcal{M} \), configurations drawn from distribution \( \pi \) have \( \epsilon \)-urbanization of poverty, and \( (\beta, \delta) \)-racial integration outside the urban area, with probability at least \( 1 - \xi_3^n \) for some constant \( 0 < \xi_3 < 1 \) and \( \delta = \frac{\gamma}{2} \).

**Remark 10.** As a special case when the size of the urban sites can roughly accommodate all the poor, where \( p < c < p + \epsilon \), if the demographics satisfies Theorem 9 with \( b_p \leq m \epsilon \), then under the same bias parameter choices as Theorem 9, the stationary configuration will have high probability of having urbanized segregation of the poor red, where the density of the poor red on the urban sites is at least \( 1 - (m + 1)\epsilon \), and racial integration among the rich outside the urban area.

## 5 Integration for Distributed Urban Sites

With the centralized urban sites, the wealth and homophily biases align to cause segregation, as shown in Theorem 8. However, when the placement of urban sites is distributed, the racial and wealth biases work against each other, and we get integration if the wealth bias is strong enough to overcome the homophily bias. (See Remark 38 in Appendix F for additional explanation.)

**Theorem 11.** We assume we have a bank of urban sites with size \( |\mathcal{U}| = \epsilon n \) evenly partitioning the city like we find with bus routes. If the number of the poor blue is small enough, \( b_p < b_p \), no matter how large \( \lambda \) is, as long as \( \gamma^{h(b_p)} > \lambda^\epsilon \frac{64}{4\pi} \frac{3\pi + 4}{4\pi} \), and \( n \) is sufficiently large, then configurations drawn from distribution \( \pi \) will be \((\beta, \delta)\)-segregated with exponentially small probability \( \xi_4^n \), for some constant \( 0 < \xi_4 < 1 \).

Integration occurs because no matter how large the homophily bias weight \( \lambda^{-(h(b_p)-h(\nu))} \) is, as long as the energy term arising from the wealth bias, \( \gamma^{\nu(\sigma)-\nu(\nu)} \), is larger, then the stationary distribution will be very unlikely to be segregated. See Appendix F for proof details.

## 6 Simulations

We supplement the theorems with simulations of \( \mathcal{M} \), shown in Figure 2, for a city with income inequality starting from random initial locations of agents. Figure 2 compares configurations after running \( \mathcal{M} \) for the same number of iterations, varying only the values of \( \lambda \), \( \gamma \), and the placement of urban sites. Note that the parameter values of \( \lambda \) and \( \gamma \) in the simulations are better than in our theorems, confirming that our proven bounds are likely not tight.
Figure 2: Simulations of $\mathcal{M}$ after five million iterations with 40% poor red (pink), 10% rich red (red), 40% rich blue (blue), and 10% poor blue (cyan) and a 46% fraction of the urban sites.

Figure 2a verifies Theorem 3 and 9, showing the coexistence of urbanization of poverty and racial integration outside the urban area under strong wealth bias but slight racial bias. Specially, since the chosen urban area can accommodate all the poor in a city and the city has severe income inequality, Figure 2a can also be viewed as a verification of Remark 10, showing segregation of the poor red in the urban area and integration outside. Individuals in Figure 2a have small racial bias, so the wealth biases can also drive racial segregation under centralized placement of urban sites. Figure 2b verifies Corollary 4, showing the urbanization of poverty with distributed urban sites. Compared with Figure 2a, the pink cluster gets dispersed via the distributed urban sites. Figure 2c verifies Theorem 8, showing the urbanized segregation. Due to income inequality where most of the poor are red, we can see the pink predominantly occupies the urban area. To contrast, in Figure 3, where the demographic is without income equality, we can see urbanized segregation and roughly the same amount of the poor red and poor blue occupying the urban sites. Figure 2d verifies Theorem 11, showing the mitigation of segregation via distributing the urban sites in the existence of agents’ strong racial bias, which should lead to Figure 2c if not distributing the urban sites. Compared with Figure 2b, whose segregation level is even smaller, the difference is that agents in Figure 2b have little racial bias, whereas in Figure 2d each agent has strong racial bias. Figure 2e and 2f provide baselines of the main work. Figure 2e shows segregation under strong racial bias without wealth bias, which is proved in [6, 21]. Figure 2f shows integration under little racial and wealth bias, which is proved in Theorem 5.
Figure 3: Urbanized segregation without income inequality under strong wealth and racial biases ($\gamma = 200, \lambda = 1.01$). The racial and wealth distribution is 25% poor red, 25% rich red, 25% rich blue, and 25% poor blue.

References

[1] George Barmpalias, Richard Elwes, and Andy Lewis-Pye. Digital morphogenesis via schelling segregation. In 2014 IEEE 55th Annual Symposium on Foundations of Computer Science, pages 156–165. IEEE, 2014.

[2] Patrick Bayer and Robert McMillan. Tiebout sorting and neighborhood stratification. Journal of Public Economics, 96(11-12):1129–1143, 2012.

[3] Prateek Bhakta, Sarah Miracle, and Dana Randall. Clustering and mixing times for segregation models on $\mathbb{Z}^2$. In Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms, pages 327–340. SIAM, 2014.

[4] Christian Borgs, Jennifer T Chayes, Alan Frieze, Jeong Han Kim, Prasad Tetali, Eric Vigoda, et al. Torpid mixing of some monte carlo markov chain algorithms in statistical physics. In 40th Annual Symposium on Foundations of Computer Science (Cat. No. 99CB37039), pages 218–229. IEEE, 1999.

[5] Christina Brandt, Nicole Immorlica, Gautam Kamath, and Robert Kleinberg. An analysis of one-dimensional schelling segregation. In Proceedings of the forty-fourth annual ACM symposium on Theory of computing, pages 789–804, 2012.

[6] Sarah Cannon, Joshua J. Daymude, Cem Gokmen, Dana Randall, and AndrÉa W. Richa. A local stochastic algorithm for separation in heterogeneous self-organizing particle systems. Proceedings of RANDOM, 2019.

[7] Sarah Cannon, Joshua J. Daymude, Dana Randall, and AndrÉa W. Richa. A markov chain algorithm for compression in self-organizing particle systems. In Proceedings of the 2016 ACM Symposium on Principles of Distributed Computing, pages 279–288, 2016.

[8] William A.V. Clark. Residential preferences and neighborhood racial segregation: A test of the schelling segregation model. Demography, 28(1):1–19, 1991.

[9] William A.V. Clark and Mark Fossett. Understanding the social context of the schelling segregation model. Proceedings of the National Academy of Sciences, 105(11):4109–4114, 2008.
[10] Marie Cottrell, Madalina Olteanu, Julien Randon-Furling, and Aurelien Hazan. Multidimensional urban segregation: an exploratory case study. In 2017 12th International Workshop on Self-Organizing Maps and Learning Vector Quantization, Clustering and Data Visualization (WSOM), pages 1–7. IEEE, 2017.

[11] Mark Fossett. Generative models of segregation: Investigating model-generated patterns of residential segregation by ethnicity and socioeconomic status. The Journal of mathematical sociology, 35(1-3):114–145, 2011.

[12] Stefan Gerhold, Lev Glebsky, Carsten Schneider, Howard Weiss, and Burkhard Zimmermann. Computing the complexity for schelling segregation models. Communications in Nonlinear Science and Numerical Simulation, 13(10):2236–2245, 2008.

[13] Edward L. Glaeser, Matthew E. Kahn, and Jordan Rappaport. Why do the poor live in cities? the role of public transportation. Journal of urban Economics, 63(1):1–24, 2008.

[14] Robert B. Griffiths. Peierls proof of spontaneous magnetization in a two-dimensional ising ferromagnet. Physical Review, 136(2A):A437, 1964.

[15] Robert B. Griffiths. The peierls argument for the existence of phase transitions. In Mathematical Aspects of Statistical Mechanics, volume 5, pages 13–26. Amer. Math. Soc Providence, 1972.

[16] Erez Hatna and Itzhak Benenson. Combining segregation and integration: Schelling model dynamics for heterogeneous population. arXiv preprint arXiv:1406.5215, 2014.

[17] John Iceland, Daniel H. Weinberg, and Erika Steinmetz. Racial and ethnic residential segregation in the United States 1980-2000, volume 8. Bureau of Census, 2002.

[18] Katherine Kortum, Rajesh Paleti, Chandra R. Bhat, and Ram M. Pendyala. Joint model of residential relocation choice and underlying causal factors. Transportation research record, 2303(1):28–37, 2012.

[19] Alexander J. Laurie and Narendra K. Jaggi. Role of ‘vision’in neighbourhood racial segregation: a variant of the schelling segregation model. Urban Studies, 40(13):2687–2704, 2003.

[20] Sonia Lehman-Frisch. Segregation, spatial (in) justice, and the city. Berkeley Planning Journal, 24(1), 2011.

[21] Shengkai Li, Bahnisikha Dutta, Sarah Cannon, Joshua J. Daymude, Ram Avinery, Enes Aydin, Andréa W. Richa, Daniel I. Goldman, and Dana Randall. Programming active cohesive granular matter with mechanically induced phase changes. Science Advances, 7(17):eabe8494, 2021.

[22] Douglas S. Massey and Nancy A. Denton. The dimensions of residential segregation. Social forces, 67(2):281–315, 1988.

[23] Sarah Miracle, Dana Randall, and Amanda Pascoe Streib. Clustering in interfering binary mixtures. In Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, pages 652–663. Springer, 2011.

[24] Rocco Paolillo and Jan Lorenz. How different homophily preferences mitigate and spur ethnic and value segregation: Schelling’s model extended. Advances in Complex Systems, 21(06n07):1850026, 2018.
[25] Liliana Perez, Suzana Dragicevic, and Jonathan Gaudreau. A geospatial agent-based model of the spatial urban dynamics of immigrant population: A study of the island of montreal, canada. Plos one, 14(7):e0219188, 2019.

[26] Mark Pollicott and Howard Weiss. The dynamics of schelling-type segregation models and a nonlinear graph laplacian variational problem. Advances in Applied Mathematics, 27(1):17–40, 2001.

[27] Dana Randall. Slow mixing of glauber dynamics via topological obstructions. In Symposium on Discrete Algorithms: Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm, volume 22, pages 870–879, 2006.

[28] Anand Sahasranaman and Henrik Jeldtoft Jensen. Ethnicity and wealth: The dynamics of dual segregation. PloS one, 13(10):e0204307, 2018.

[29] Thomas C. Schelling. Dynamic models of segregation. Journal of mathematical sociology, 1(2):143–186, 1971.

[30] Abhinav Singh, Dmitri Vainchtein, and Howard Weiss. Schelling’s segregation model: Parameters, scaling, and aggregation. Demographic Research, 21:341–366, 2009.

[31] Dietrich Stauffer. Social applications of two-dimensional ising models. American Journal of Physics, 76(4):470–473, 2008.

[32] Dietrich Stauffer and Sorin Solomon. Ising, schelling and self-organising segregation. The European Physical Journal B, 57(4):473–479, 2007.

[33] Tiit Tammaru, Szymon Marcińczak, Raivo Aunap, Maarten van Ham, and Heleen Janssen. Relationship between income inequality and residential segregation of socioeconomic groups. Regional Studies, 54(4):450–461, 2020.
Appendix A: Technical Summary

A.1 Peierls arguments

The goal of a Peierls argument is to show that the probability that a sample drawn from the stationary distribution $\pi$ of a Markov chain falls into a target set is exponentially small in $n$, which yields $\pi(\Omega_t) \leq \xi^n$ for some constant $\xi \in (0,1)$. It maps from configurations in the target set $\Omega_t$ to configurations in the configuration space $\Omega$ and shows that the map has an exponential gain in probability. Thus the targeted configurations are exponentially unlikely in $\Omega$. Mathematically, the mapping $\nu = f(\sigma)$ is defined from $\sigma \in \Omega_t$ to $\nu \in \Omega$, which yields

$$\pi(\Omega_t) = \sum_{\sigma \in \Omega_t} \pi(\sigma) \leq \sum_{\nu \in \Omega} \sum_{\sigma \in f^{-1}(\nu)} \pi(\sigma) = \sum_{\nu \in \Omega} \pi(\nu) \frac{\sum_{\sigma \in f^{-1}(\nu)} \pi(\sigma)}{\pi(\nu)}. \quad (5)$$

In order to show $\pi(\Omega_t) \leq \xi^n$, the mapping needs to be carefully defined to get the upper bounds of the probability ratio $\frac{\pi(\sigma)}{\pi(\nu)}$ and the number of the preimages $|f^{-1}(\nu)|$ for any given $\nu$. Usually a mapped configuration $\nu$ with large probability ratio $\frac{\pi(\sigma)}{\pi(\nu)}$ will also have many possible preimages, which leads to the tradeoffs between $\frac{\pi(\sigma)}{\pi(\nu)}$ and $|f^{-1}(\nu)|$, known as an energy/entropy tradeoff.

To facilitate the mapping operations $f(\cdot)$ and the counting of $|f^{-1}(\nu)|$, certain bridge systems have been defined in [6, 23] to systematically encode the agents’ colors to facilitate inverting the map (to bound the number of preimages). Compared to [6, 23], here the configuration space is enlarged by adding the wealth dimension, which requires extending the bridge system to encode the multi-dimensional information simultaneously. Moreover, because of the additional wealth term is reflected in the stationary distribution, more careful mapping rules are required to account for both the tradeoffs between $\frac{\pi(\sigma)}{\pi(\nu)}$ and $|f^{-1}(\nu)|$, and the tradeoffs between the wealth bias and the homophily bias in the probability measure.

A.2 Bridge Systems

Throughout the proofs, we use red and blue to represent the races, and richness of the color to represent the wealth of each agent (rich red is red; poor red is pink; rich blue is blue; poor blue is cyan). We first need to extend the bridging technique to expand the encoded information dimension from color only to both color and richness over the methods in [6, 23], within our context. The following shows our adapted bridging technique.

**Lattice Duality.** The hexagonal dual to the triangular lattice $G_\Delta$ is obtained by creating a vertex at the centroid of each unit triangle in $G_\Delta$ and connecting two of these vertices if their corresponding unit triangles have a common edge (as shown in Figure 4a, the obtained hexagonal lattice is denoted by $G_{\text{hex}}$). Each edge $e' \in G_{\text{hex}}$ crosses a unique edge $e \in G_\Delta$ and separates two adjacent agents living at the of $e$. There is a bijection between edges of $G_\Delta$ and edges of $G_{\text{hex}}$, associating an edge of $G_\Delta$ with the unique edge of $G_{\text{hex}}$, and vice versa.

**Color Contours and Color Bridges.** If an edge $e' \in G_{\text{hex}}$ separates two agents heterogeneous in race, we call it a color edge. We define a color contour to be made up of color edges and is a self-avoiding polygon in $G_{\text{hex}}$ that never visits the same vertex twice except to start and end at the same place. The color contour is denoted in green as shown in Figure 4b. The color bridges are shown in dashed green. They are self-avoiding walks on $G_{\text{hex}}$ that connect color contours to the boundary.
Figure 4: (a) The hexagon lattice $G_\triangle$ dual to the city lattice $G_\triangledown$. (b) Demonstration of the color contours and color bridges. (c) Demonstration of the richness contours and richness bridges.

**Richness Contours and Richness Bridges.** If an edge $e' \in G_{\text{hex}}$ separates two agents heterogeneous in wealth, we call it a **richness edge**. We define a **richness contour** to be made up of richness edges and is a self-avoiding polygon, which is denoted in orange as shown in Figure 4c. The **richness bridges** are shown in dashed orange, which connects richness contours to the boundary.

**δ−Color-and-Richness Bridge System.** We first define the $\delta$−color bridge system similarly as [6,23] as the following. For any color component $F$, let $C_c$ be a collection of color contours of $F$. The color bridges collection $B_c$ connects each color contour in $C_c$ to the boundary of $F$.

An agent $P$ is bridged in terms of color in $F$ if there exists a path through agents of the same race as $P$ to the boundary of $F$ or a bridged color contour in $C_c$. An agent is unbridged in terms of color if such a path does not exist. Then we define that $(B_c, C_c)$ is a $\delta$−**color bridge system** for $F$ if

1. $|B_c| \leq |C_c|(1 - \delta)/2\delta$, where $|B_c|$ is the total number of edges in $B_c$ and $|C_c|$ is the total number of edges in $C_c$;

2. the number of unbridged agents in terms of color in $F$ is $\leq \delta|F|$, where $|F|$ is the number of agents in $F$.

Note $\delta \in [0, 1]$ controls how much color information is omitted by the $\delta$−color bridge system, and see proof for Lemma 7.2 in [6] for the construction way of a $\delta$−color bridge system for any $F$.

**Lemma 12 ([6]).** For any color component $F$, there exists a $\delta$−**color bridge system** for $F$.

Similarly, we can also define the $\delta$−richness bridge system, and the richness component. The following lemma holds similarly.

**Lemma 13 ([6]).** For any richness component $F$, there exists a $\delta$−**richness bridge system** for $F$.

We call the joint $\delta$−color and $\delta$−richness bridge system a $\delta$−**color-and-richness bridge system**. For any configuration $\sigma \in \Omega_t$, we can construct a $\delta$−color-and-richness bridge system. See Figure 5a for illustrations, where at most $\delta n$ agents are not bridged in terms of color, and at most $\delta n$ agents are not bridged in terms of richness. Combining Lemma 12 and Lemma 13, the following lemma holds.

**Lemma 14.** For any finite region $F$, there exists a $\delta$−**color-and-richness bridge system** for $F$.

---

A color component is a maximal simply connected subset $F$ of agents where all agents in $F$ adjacent to a location not in $F$ have the same race, which we call the color of $F$. 

Figure 5: (a) A configuration $\sigma$ with a $\delta$–color-and-richness bridge system. (b) Richness inversion in $\psi = f_1(\sigma)$. (c) Color inversion in $\tau = f_2(\psi)$.

Crossing Contours and Non-Crossing Contours. The contour that touches the boundary of the defined domain is called a crossing contour. The contour that does not touch the boundary is called a non-crossing contour. For example, Figure 4b has one crossing color contour and three non-crossing color contours.

The sum of the number of edges of a contour is called the length of the contour. We denote the length of the bridged non-crossing color contours as $y_c$, and length of the crossing contours as $x_c$. The length of the bridged non-crossing richness contours is denoted by $y_r$. The length of the crossing richness contours is $x_r$. We can also define non-crossing richness contours $y_r$ and crossing richness contours $x_r$ in similar way. It follows that $y = y_c + y_r$ and $x = x_c + x_r$. We call $z_c \triangleq x_c + y_c$ the bridged color contour length, and $z_r \triangleq x_r + y_r$ the bridged richness contour length.

For a given bridged color or richness contour length, the following lemmas bound the number of possible bridge systems, which can be counted in a depth-first way (see proof details in [6]).

Lemma 15 (Lemma 7.6 in [6]). For a given $z_r = y_r + x_r$, there are at most $(z_r + 1)^3 4^{3\delta+1} \alpha \sqrt{n}$ ways of constructing a $\delta$–richness bridge system.

Lemma 16 (Lemma 7.6 in [6]). For a given $z_c = y_c + x_c$, there are at most $(z_c + 1)^3 4^{3\delta+1} \alpha \sqrt{n}$ ways of constructing a $\delta$–color bridge system.

Combining Lemma 15 and 16 yields the following lemma.

Lemma 17. For a given $z_c = y_c + x_c$, there are at most $(z_r + 1)(z_c + 1)^3 4^{3\delta+1} \alpha \sqrt{n}(z_c + \max\{z_r\})$ ways of constructing a $\delta$–color-and-richness bridge system.

A.3 The Inversion Mappings

For any configuration $\sigma \in \Omega_t$, after bridging, in order to get the upper bounds of $h(\nu) - h(\sigma)$ and $p(\sigma) - p(\nu)$, we define the mappings to first eliminate the bridged racially heterogeneous edges $h(\sigma)$ and bridged poor agents, and then define mappings to recover the racially heterogeneous edges and the poor agents on the urban sites in various designed ways. Different proof targets will lead to different recovery mappings, shown in the theorems in the following sections. For this part, we define the richness inversion function $f_1(\cdot)$ and color inversion function $f_2(\cdot)$ which eliminate most of the racially heterogeneous edges and poor agents respectively.

Richness Inversion. First, we represent each agent with two bits, with poor red denoted 00, poor blue denoted 01, rich red denoted 10 and rich blue denoted 11. We define the richness inversion function $f_1(\cdot)$ as follows: for any agent $ij$, where $i$ is the richness bit, and $j$ is the color bit, the
richness bit is flipped to \((i + b)(\text{mod}2)\) for agent \(ij\) that is surrounded by \(b\) bridged richness contours or unbridged crossing richness contours (see the left corner’s contour as an example of richness contours in Figure 5b). The color bit remains unchanged. See Figure 5b for illustrations.

**Lemma 18** (Lemma 7.5 in [6]). For any configuration \(\psi = f_1(\sigma)\), there are at most \(\delta n\) poor agents, and they are unbridged; no additional color edges are introduced; and for any mapped configuration \(\psi\), there is only one preimage for a given \(\delta\)-richness bridge system.

**Color Inversion.** We define the function *color inversion* \(f_2(\cdot)\) as: for any agent \(ij\), where \(i\) is the richness bit, and \(j\) is the color bit, the color bit is flipped to \((j + b)(\text{mod}2)\) for the agent \(ij\) that is surrounded by \(b\) bridged color contours or unbridged crossing color contours (like the red agent on the right boundary in Figure 5b). The richness bit remains unchanged. See Figure 5c for illustrations.

**Lemma 19** (Lemma 7.8 in [6]). For any \(\tau = f_2(\psi), (x_c + y_c)\) of the original racially heterogeneous edges in \(\psi\) are eliminated: \(h(\tau) - h(\psi) \leq -(x_c + y_c)\); no additional poor agents are introduced; and for any mapped configuration \(\tau\) with a given bridge system, there is only one preimage that can be mapped to it.

**A.4 The Color and Richness Recovery Mappings**

After eliminating the bridged poor agents and the bridged racially heterogeneous edges in \(f_1(\cdot)\) and \(f_2(\cdot)\), we need to recover the same ratio of each color and richness as in \(\sigma\), which is defined in \(f_3(\cdot)\), \(f_4(\cdot)\) and \(f_5(\cdot)\) as the following.

**Pink Recovery.** For any \(\tau = (f_2 \circ f_1)(\sigma)\), we define the *pink recovery* function \(f_3(\tau)\) as to flip the agents’ colors to pink starting from a fixed place in a given order except when encountering the following unbridged agents: we flip the unbridged pink to cyan, cyan to red, and red to blue. The flipping process stops once reaching the correct number of the pink agents as in \(\sigma\).

**Lemma 20.** For any mapped configuration \(\zeta\), if the starting location of \(f_3\) and the flipping order are specified, there are at most \(n\) preimages that can be mapped to it: \(|f_3^{-1}(\zeta)| \leq n\).

*Proof.* Given \(\zeta\), it suffices to recover its preimage if we are given the stopping place and there are at most \(n\) possible stopping places.

**Cyan Recovery.** For any \(\zeta = (f_3 \circ f_2 \circ f_1)(\sigma)\), we define the *cyan recovery* function \(f_4(\zeta)\) as to flip the agents starting from the stopping place of \(f_3(\cdot)\) in a given order to cyan except when encountering the following unbridged agents: we remain the unbridged pink to pink, red to red, and flip cyan to blue. The flipping will be stopped after reaching the right number of the cyan agents as in \(\sigma\). The proof of Lemma 21 is similar to the proof of Lemma 20. Given any mapped configuration \(\phi\), it suffices to recover its preimage \(f_4^{-1}(\phi)\) if we are given the starting location and the stopping place of \(f_4(\cdot)\), which is bounded by \(n^2\).

**Lemma 21.** For any mapped configuration \(\phi\), if the flipping order is specified, there are at most \(n^2\) preimages that can be mapped to it: \(|f_4^{-1}(\phi)| \leq n^2\).

**Red Recovery.** For any \(\phi = (f_1 \circ f_3 \circ f_2 \circ f_1)(\sigma)\), we define the *red recovery* function \(f_5(\cdot)\) as to flip agents starting from the stopping place of \(f_4(\cdot)\) in a given order to red except when encountering the following unbridged agents: we flip the unbridged red to blue, cyan to pink, and pink to cyan. To guarantee the right number of the cyan and pink in the mapped configuration \(\nu\), whenever we
Figure 6: (a) Demonstration of $\zeta = f_3(\tau)$. (b) Demonstration of $\phi = f_4(\zeta)$. (c) Demonstration of $\nu = f_5(\phi)$.

flip an unbridged cyan to pink during this phase, we flip one pink back to cyan starting from the stop location of $f_3(\cdot)$. If we encounter non-pink agents, we first recover its colors before $f_3(\cdot)$ and use the rule of flipping the unbridged agents for $f_4(\cdot)$. Whenever we flip an unbridged pink to cyan, we flip one cyan to pink starting from the starting location of $f_4(\cdot)$, and if we encounter agents that are not cyan, we first recover its colors before $f_4(\cdot)$ and use the rule of flipping the unbridged agents for $f_3(\cdot)$. We stop such operations after reaching the right number of the red for $\nu$.

The proof of Lemma 22 is similar as Lemma 20. To show the upper bound of the number of preimages $|f_5^{-1}(\nu)|$ for a given $\nu$: we first need to find the stop location of $f_4(\cdot)$, which has at most $n$ possibilities. Then we complement the colors of the first $k$ elements of $\nu$, where $k \in \{0, 1, ..., n - 1\}$, and possibly we also need to find the stop location of $f_3$ (same location as the starting location of $f_4$), which has at most $n$ possibilities. Altogether there are at most $n^3$ different preimages.

Lemma 22. For any mapped configuration $\nu$, if the flipping order is specified, there are at most $n^3$ preimages that can be mapped to it: $|f_5^{-1}(\nu)| \leq n^3$.

A.5 The Centralized Recovery and Distributed Recovery

Centralized Recovery. If the urban sites are centralized, like shown in Figure 1a, for the pink, cyan, and red recoveries defined in $f_3$, $f_4$ and $f_5$, the starting location of $f_3$ and the flipping order of each function can be specified in the centralized way: the starting location of $f_3$ is specified to be the center of the urban area, and the flipping order for $f_3$, $f_4$, $f_5$ are specified as in clockwise direction and loop to the immediate outer layer when completing flipping one clockwise cycle like shown in Figure 6a, 6b, and 6c. In such a way, the following upper bound can be obtained.

Lemma 23. If the recoveries are specified in the centralized way, for any $\nu = (f_5 \circ f_4 \circ f_3)(\tau)$, there are at most $3\alpha \sqrt{n}$ racially heterogeneous edges introduced by the recovery operations: $h(\nu) - h(\tau) \leq 3\alpha \sqrt{n}$.

Proof. In the defined way of centralized recovery, racially heterogeneous edges are created along the hexagon boundaries between the pink and cyan regions, the cyan and red regions, and the red and the rest region. Each boundary can be upper bounded by the perimeter of the fundamental domain, which is $\alpha \sqrt{n}$, and in total $3\alpha \sqrt{n}$.

Inside each pink, cyan, and red region, the unbridged agents will not create additional racially heterogeneous edges: for any $\tau = (f_2 \circ f_1)(\sigma)$, It follows from the definition of $f_2$ and Lemma 19 that the bridged agents in $\tau$ are either cyan or blue. The flipping rules for the unbridged agents in $f_3$, $f_4$ and $f_5$ thus can be verified to not introduce additional racially heterogeneous edges. Hence we get $h(\nu) - h(\tau) \leq 3\alpha \sqrt{n}$. \qed
Distributed Recovery. If the urban sites even partition the city (diamond-shaped city) like shown in Figure 1b, for the pink, cyan, and red recoveries defined in \( f_3, f_4 \) and \( f_5 \), the starting location of \( f_3 \) and the flipping order of each function can be specified in the distributed way: the starting location of \( f_3 \) is specified to be the most top left corner of the urban site, and the flipping order for \( f_3, f_4, f_5 \) are specified as to only flip the agents on the urban sites row after row, and then flip agents on the non-urban sites row after row after finishing flipping all the agents on the urban sites. In such a way, the following upper bound can be obtained.

**Lemma 24.** If the recoveries are specified in the distributed way, for any \( \nu = (f_5 \circ f_4 \circ f_3)(\tau) \), there are at most \( 2cn \) racially heterogeneous edges introduced by the recovery operations: \( h(\nu) - h(\tau) \leq 2cn \).

**Proof.** In the defined way of distributed recovery, compared with \( \tau \), racially heterogeneous edges are possibly created along the diamond boundaries between the urban and non-urban sites (see Figure 1b for demonstration), which can be upper bounded by the sum of perimeters of all the small diamonds.

The perimeter of a diamond can be upper bounded by \( 4 \cdot k \): since the urban sites, with total size \( c \cdot n \), evenly partitions the finite lattice, the number of row urban sites is \( \frac{c}{2} \). Each row of urban sites has \( \sqrt{n} \) sites, so there are \( \frac{\sqrt{n}}{2} \) rows of urban sites. Since each column of urban sites also has \( \sqrt{n} \) sites, we can get the side length of a diamond to be \( k = \frac{\sqrt{n}}{\sqrt{2} - 1} \geq \frac{\sqrt{n}}{2} \). Since the side length of one diamond is \( k \), the total number of the diamonds can be upper bounded by \( k^2 \). Hence the sum of perimeters of all the diamonds can be upper bounded by \( \frac{4n}{k} \leq 2cn \), which yields \( h(\nu) - h(\tau) \leq 2cn \) \( \square \).

**Lemma 25.** For any \( \nu = (f_5 \circ f_4 \circ f_3)(\sigma) \), if the recovery is specified in the centralized way or distributed way, the number of the poor agents on the urban sites is lower bounded by \( p(\nu) \geq (\min\{c, p\} - \delta)n \).

**Proof.** It follows from Lemma 18 that the number of poor agents is at most \( \delta n \) in the mapped \( \psi = f_1(\sigma) \) and they are unbridged. It follows from Lemma 19 that no additional poor agents will be introduced for any \( \tau = f_2(\psi) \). In the worst case scenario, these \( \delta n \) poor agents will not be recovered on the urban sites in \( \nu = (f_5 \circ f_4 \circ f_3)(\tau) \). For the rest of the poor agents, because of the recovery way in the order of pink followed by cyan then red, it yields that \( p(\nu) \geq (\min\{c, p\} - \delta)n \). \( \square \)

**Appendix B: Proof of Theorem 3 (Urbanization of Poverty)**

The claims in the proof of Theorem 3 are shown and proved as follows.

**Claim 26.** For any \( \sigma \in \Omega_{urb} \) with a given bridged color contour length \( z_c \), for the defined mapping \( \nu = f(\sigma) \), \( h(\nu) - h(\sigma) \leq 3\alpha \sqrt{n} - z_c \) and \( p(\sigma) - p(\nu) \leq -\delta n \), where \( \delta = \epsilon/2 \).

**Proof of Claim 26.** It follows from Lemma 19 that for any \( \sigma \in \Omega_{urb} \), \( h(\tau) - h(\sigma) \leq -z_c \). It follows from Lemma 23 that \( h(\nu) - h(\tau) \leq 3\alpha \sqrt{n} \). Combining the two inequalities, we get \( h(\nu) - h(\sigma) \leq 3\alpha \sqrt{n} - z_c \).

For any \( \sigma \in \Omega_{urb} \), \( p(\sigma) \leq (\min\{c, p\} - \epsilon)n \) is satisfied. It follows from Lemma 25 that \( p(\nu) \geq (\min\{c, p\} - \delta)n \). Combining the inequalities, setting \( \epsilon = 2\delta \), it follows that \( p(\sigma) - p(\nu) \leq -\epsilon n + \delta n = -\delta n \). \( \square \)

**Claim 27.** For a given color contour length \( z_c \), for any \( \nu = f(\sigma) \), the number of configurations in \( \Omega_{urb} \) that can map to \( \nu \) is upper bounded by: \( |f^{-1}(\nu)| \leq (z_c + 1)9^{9\sqrt{n}}(\frac{14}{48})(z_c+3n) \).
Proof of Claim 27. We denote \( \nu = f_5(\phi) \), \( \phi = f_3(\zeta) \), \( \zeta = f_3(\tau) \), \( \tau = f_2(\psi) \), and \( \psi = f_1(\sigma) \). Since \( f(\sigma) = (f_5 \circ f_4 \circ f_3 \circ f_2 \circ f_1)(\sigma) \), for any \( \nu \), it follows that \(|f^{-1}(\nu)| \leq |f_5^{-1}(\nu)| \cdot |f_4^{-1}(\phi)| \cdot |f_3^{-1}(\zeta)| \cdot |f_2^{-1}(\tau)| \cdot |f_1^{-1}(\psi)| \). It follows from Lemma 22 that \(|f_5^{-1}(\nu)| \leq n^3 \). It follows from Lemma 21 that \(|f_4^{-1}(\phi)| \leq n^2 \). It follows from Lemma 20 that \(|f_3^{-1}(\zeta)| \leq n \).

For a given color contour length \( z_c \), the number of all possible bridge systems \( b \) is upper bounded by \( b \leq (\max\{z_i\} + 1)(z_c + 1)9^\alpha \sqrt{n}4^{\frac{3\alpha + 1}{4\epsilon}}(z_c + \max\{z_i\}) \) in Lemma 17. For the triangular lattice with \( n \) vertices, the sum of all edges is \( 3n \). Because of the lattice duality, we conclude \( z_t \leq 3n \). Thus \( b \leq (3n + 1)(z_c + 1)9^\alpha \sqrt{n}4^{\frac{3\alpha + 1}{4\epsilon}}(z_c + 3n) \). It follows from Lemma 18 and 19 that for a given bridge system and a given image, the number of the corresponding preimages of both \( f_2 \) and \( f_1 \) is one.

Hence, we conclude that for a given \( z_c \), \(|f^{-1}(\nu)| \leq |f_5^{-1}(\nu)| \cdot |f_4^{-1}(\phi)| \cdot |f_3^{-1}(\zeta)| \cdot b \leq n^6(3n + 1)(z_c + 1)9^\alpha \sqrt{n}4^{\frac{3\alpha + 1}{4\epsilon}}(z_c + 3n) \).

\[ \square \]

Appendix C: Proof of Theorem 5 (Dispersion of Poverty)

Proof of Theorem 5. Let \( \Omega_{urb} \subseteq \Omega \) be the set of configurations that have \( \epsilon \)-urbanization of poverty. To prove Theorem 5, it suffices to prove \( \pi(\Omega_{urb}) \leq \xi_2^\alpha \) for some \( \xi_2 \in (0, 1) \) and large enough \( n \).

For each \( \sigma \in \Omega_{urb} \), we construct a \( \delta \)-richness bridge system and define the mapping \( \nu = d(\sigma) = (d_2 \circ f_1)(\sigma) \) as the following: we first do the richness inversion and obtain \( \tau = f_1(\sigma) \); next for \( \tau \), we randomly flip the blue to cyan until the right number of the cyan, and we randomly flip the red to pink until the right number of the pink and obtain \( \nu = d_2(\tau) \).

Claim 28. For any \( \sigma \in \Omega_{urb} \), for the defined mapping \( \nu = d(\sigma), h(\nu) - h(\sigma) \leq 0 \) and \( p(\sigma) - p(\nu) \leq cn \).

Proof of Claim 28. Because the color information remains unchanged for \( f_1 \) and \( d_2 \), \( h(\nu) - h(\sigma) \leq 0 \). For any \( \sigma, p(\sigma) \leq \min\{c, p\} n \). The scenario under discussion follows that \( p < c \), thus \( p(\sigma) \leq pn \).

Since for any \( \nu, p(\nu) \geq 0 \), hence \( p(\sigma) - p(\nu) \leq pn \).

\[ \square \]

Claim 29. For a given richness contour length \( z_t \), for any \( \nu = d(\sigma) \), the number of configurations in \( \Omega_{urb} \) that can map to \( \nu \) is upper bounded by: \(|d^{-1}(\nu)| \leq (\beta \sqrt{n} + 1)3^{\alpha \sqrt{n}4^{\frac{3\alpha + 1}{4\epsilon}}(z_t + 3n)} \).

Proof of Claim 29. We denote \( \tau = f_1(\sigma) \) and \( \nu = d_2(\tau) \). Since \( d(\sigma) = (d_2 \circ f_1)(\sigma) \), for any \( \nu \), it follows that \(|d^{-1}(\nu)| \leq |f_1^{-1}(\tau)| \cdot |d_2^{-1}(\nu)| \). For any \( \nu \), the number of preimages \( |d_2^{-1}(\nu)| \) can be upper bounded by \( 2^{pn} \) by recording whether each poor agent is flipped or not.

Because every configuration \( \sigma \in \Omega_{urb} \) satisfies \( \epsilon \)-urbanization, assuming we also have \( p < c < p + \epsilon \), then \( \sigma \) also satisfies \((\beta, \epsilon)\)-wealth segregation (see definition in Remark 2). Hence \( z_t \) can be upper bounded by \( z_t \leq \beta \sqrt{n} \) (See Lemma 7.4 in [6] for details). Thus it follows from Lemma 15 that for any given \( z_t \), the number of \( \delta \)-richness bridge systems can be upper bounded by \( (\beta \sqrt{n} + 1)3^{\alpha \sqrt{n}4^{\frac{3\alpha + 1}{4\epsilon}}(z_t + 3n)} \). It follows from Lemma 18 that for any \( \tau \), for a given \( \delta \)-richness bridge system, the number of preimages is one. Hence we conclude for any \( \tau \) with a given \( z_t \), the number of preimages is upper bounded by: \(|f_1^{-1}(\tau)| \leq (\beta \sqrt{n} + 1)3^{\alpha \sqrt{n}4^{\frac{3\alpha + 1}{4\epsilon}}(z_t + 3n)} \). Combining the two inequalities, we conclude \(|d^{-1}(\nu)| \leq (\beta \sqrt{n} + 1)3^{\alpha \sqrt{n}4^{\frac{3\alpha + 1}{4\epsilon}}(z_t + 3n)} \).

\[ \square \]

Claim 30. For any given \( \sigma \in \Omega_{urb} \), we denote \( D(\sigma) \) to be the set of all possible images \( \nu = d(\sigma) \) mapped from \( \sigma \). It follows that \(|D(\sigma)| \geq (\frac{b - \delta}{b_0})^{n}(\frac{b_0 - \delta}{b_0})^{n} \).

Proof of Claim 30. It follows from the definition of \( f_1 \) that for any given \( \sigma \), there is only one configuration can be obtained from \( f_1(\sigma) \). For any given \( \tau \), we define \( D_2(\tau) \) to be the set of all possible configurations obtained from \( d_2(\tau) \) by flipping the pink and cyan back; then \(|D_2(\tau)| =
\[(\frac{rn - a_\tau}{r_p n - a_\tau})(\frac{bn - a_\tau}{b_p n - a_\tau})\], where \(a_\tau\) is the number of unbridged agents in \(\tau\), \(a_\tau'\) is the number of unbridged pink, and \(a_\tau''\) is the number of unbridged cyan. Thus \(|D_2(\tau)|\) can be further lower bounded by

\[|D_2(\tau)| \geq \left(\frac{r - a_\tau}{r_p - a_\tau}\right)(r_p a_\tau')(b - a_\tau')(b_p a_\tau'')(b - \delta)(b_p - \delta)n \geq \left(\frac{r - \delta}{r_p}\right)(r_p - \delta)n \left(\frac{b - \delta}{b_p}\right)(b_p - \delta)n.\] (6)

Hence \(|D(\sigma)|\) can be lower bounded by \(|D(\sigma)| \geq 1 \cdot |D_2(\tau)| \geq \left(\frac{r - \delta}{r_p}\right)(r_p - \delta)n \left(\frac{b - \delta}{b_p}\right)(b_p - \delta)n.\]

Finally, we define a weighted bipartite graph \(G(\Omega_{urb}, \Omega, E)\) with an edge of weight \(\pi(\sigma)\) between \(\sigma \in \Omega_{urb}\) and \(\nu \in \Omega\). The total weight of edges is

\[\sum_{\sigma \in \Omega_{urb}} \pi(\sigma) \cdot |D(\sigma)| \geq \pi(\Omega_{urb})\left(\frac{r - \delta}{r_p}\right)(r_p - \delta)n \left(\frac{b - \delta}{b_p}\right)(b_p - \delta)n.\] (7)

On the other hand, the weight of the edges is at most

\[\sum_{\nu \in \Omega} \sum_{\sigma \in \Omega_{urb}} \max_{\pi(\sigma)} \pi(\sigma) = \sum_{\nu \in \Omega} \pi(\nu) \sum_{\sigma \in \Omega_{urb}} \frac{\max_{\pi(\sigma)}}{\pi(\nu)} |d^{-1}(\nu)| \leq \sum_{\nu \in \Omega} \pi(\nu) \sum_{\nu' = \nu \pi} \lambda^{\max(h(\nu) - h(\sigma))} \max(p(\sigma) - p(\nu)) (\beta \sqrt{n} + 1)(3^{n-1} \beta \sqrt{n})^{3^{n-1} \beta \sqrt{n} 2 \rho n} \leq \gamma^{3n} \beta \sqrt{n}(\beta \sqrt{n} + 1)(3^{n-1} \beta \sqrt{n})^{3^{n-1} \beta \sqrt{n} 2 \rho n},\] (8)

where the inequalities in Claim 28 and 29 has been substituted in the above derivation. Combining (7) and (8), we have

\[\pi(\Omega_{urb})\left(\frac{r - \delta}{r_p}\right)(r_p - \delta)n \left(\frac{b - \delta}{b_p}\right)(b_p - \delta)n \leq \beta \sqrt{n}(\beta \sqrt{n} + 1)(3^{n-1} \beta \sqrt{n})^{3^{n-1} \beta \sqrt{n} 2 \rho n} \gamma^{3n} 2 \rho n.\]

For large enough \(n\), to have \(\pi(\Omega_{urb}) \leq \xi_3^n\) for some \(\xi_3 \in (0, 1)\), it suffices to have

\[\gamma^{3n} 2 \rho n < \left(\frac{r - \delta}{r_p}\right)(r_p - \delta)n \left(\frac{b - \delta}{b_p}\right)(b_p - \delta)n < \left(\frac{r - \delta}{r_p}\right)r_p n \left(\frac{b - \delta}{b_p}\right)b_p n,\]

which can be rewritten as

\[\gamma < \left(\frac{r - \delta}{r_p}\right) \frac{r_p n}{b_p n} \left(\frac{b - \delta}{b_p}\right) \frac{b_p}{b_p} / 2.\]

Since \(\gamma > 1\), to make the right hand side of the above inequality greater than one, it suffices to have \(r_p > 2\) and \(b_p > 2\), which can be rewritten as \(r_p < r_t - \delta\) and \(b_p < b_t - \delta.\]

**Appendix D: Proof of Theorem 8 (Urbanized Racial Segregation)**

Claim 31 in the proof of Theorem 8 are shown and proved as follows.

**Claim 31.** If \(\lambda > 3^{n-1} \beta \sqrt{n} 2 \rho n\), \(\pi(\Omega_{urb \neg seg}) \leq \xi_0 \sqrt{n}\) for some \(\xi_0 \in (0, 1).\)
Proof of Claim 31. To prove \( \pi(\Omega_{\text{urb} \wedge \text{-seg}}) \leq \xi_0 \sqrt{n} \), we define the bridge system and the mapping \( g(\cdot) = (g_3 \circ g_2 \circ f_2)(\cdot) \) from the set \( \Omega_{\text{urb} \wedge \text{-seg}} \) to \( \Omega \) as the following: for any \( \sigma \in \Omega_{\text{urb} \wedge \text{-seg}} \), we construct a \( \delta \)-color bridge system for it. Then we do color inversion like defined in \( f_2(\cdot) \), and get \( \psi = f_2(\sigma) \).

\( \tau = g_2(\psi) \) is defined as the following: starting from the center of the urban area, we flip the color bit \((0 \text{ to } 1 \text{ and } 1 \text{ to } 0)\) of each agent layer by layer in a given order until the right number of pink and cyan is reached.

\( \nu = g_3(\tau) \) is then defined as the following: starting from outside the urban area boundary, we flip the color bit \((0 \text{ to } 1 \text{ and } 1 \text{ to } 0)\) of each agent layer by layer in a given order until the right number of red and blue is reached. During this phase, whenever we flip the color bit for an unbridged pink, we go back to the stopping location of \( g_2(\cdot) \) to flip one more cyan to pink, and vice versa.

Claim 32. For any \( \sigma \in \Omega_{\text{urb} \wedge \text{-seg}} \) with a given bridged color contour length \( z_c \), for the defined mapping \( \nu = g(\sigma) \), \( h(\nu) - h(\sigma) \leq 2\alpha \sqrt{n} - z_c \) and \( p(\sigma) - p(\nu) \leq 0 \).

Proof of Claim 32. It follows from Lemma 19 that \( h(\psi) - h(\sigma) \leq -z_c \) and the richness information remains unchanged. For the color recoveries \((g_3 \circ g_2)(\psi)\), similar as the proving strategy of Lemma 23, we at most introduce \( 2\alpha \sqrt{n} \) racially heterogeneous edges compared with \( h(\psi) \) due to the hexagon boundaries created between pink and cyan, and cyan and red. Each boundary can be upper bounded by the perimeter of the fundamental domain \( \alpha \sqrt{n} \). The richness information remains unchanged. Thus \( h(\nu) - h(\sigma) \leq h(\nu) - h(\psi) + h(\psi) - h(\sigma) \leq 2\alpha \sqrt{n} - z_c \) and \( p(\sigma) - p(\nu) \leq 0 \).

Claim 33. For a given color contour length \( z_c \), for any \( \nu = g(\sigma) \), the number of configurations in \( \Omega_{\text{urb} \wedge \text{-seg}} \) that can map to \( \nu \) is upper bounded by: \( |g^{-1}(\nu)| \leq n^3(z_c + 1)3^{\alpha \sqrt{n}} \frac{\delta^3 + 1}{4 \beta^4} z_c \).

Proof of Claim 33. For any given mapped \( \tau \), the number of preimages \( |g_2^{-1}(\tau)| \) is at most \( n \), since we only need to know the stopping locations of the flipping operations, and there are at most \( n \) possibilities of the stopping location. For any given mapped \( \nu \), the number of preimages \( |g_3^{-1}(\tau)| \) is at most \( n^2 \), since we need the stopping location information of \( g_2(\cdot) \), which is upper bounded by \( n \) and the stopping location of \( g_3(\cdot) \) is also upper bounded by \( n \).

To bound \( |g^{-1}(\nu)| \) for a given \( \nu \) and a given \( z_c = x_c + y_c \), we can first bound the number of possible bridge systems for a given \( z_c \), which yields \( (z_c + 1)3^{\alpha \sqrt{n}} \frac{\delta^3 + 1}{4 \beta^4} z_c \). See proof details of this bound from Lemma 7.6 in [6]. For any configuration \( \psi \in f_2(\sigma) \) with a given bridge system, there is only one configuration \( \sigma \) that can be mapped to \( \psi \) (Lemma 19). Thus combining with \( |g_2^{-1}(\tau)| \) and \( |g_3^{-1}(\nu)| \), it yields \( |g^{-1}(\nu)| \leq n^3(z_c + 1)3^{\alpha \sqrt{n}} \frac{\delta^3 + 1}{4 \beta^4} z_c \).

Finally, substituting the bounds into Peierls Argument (5) yields

\[
\pi(\Omega_{\text{urb} \wedge \text{-seg}}) \leq \sum_{\nu \in \Omega} \pi(\nu) \sum_{z_c = \beta \sqrt{n}} n^3(z_c + 1)3^{\alpha \sqrt{n}} \left( \frac{\delta^3 + 1}{4 \beta^4} \right) z_c \\
\leq \sum_{\nu \in \Omega} \pi(\nu) \sum_{z_c = \beta \sqrt{n}} n^3(z_c + 1) \left( \frac{3^\alpha \delta^3 + 1}{4 \beta^4} \right) z_c,
\]

where \( z_c \geq \beta \sqrt{n} \) is due to \( \sigma \in \Omega_{\text{urb} \wedge \text{-seg}} \) does not satisfy \((\beta, \epsilon)\)-segregation (see Lemma 7.4 in [6] for details), and \( z_c \leq \beta \sqrt{n} \), which is the sum of all edges of \( G_\Delta \). If \( \lambda > 3^\alpha \delta^3 + 1 \), the sum will be exponentially small for sufficiently large \( n \), which means \( \pi(\Omega_{\text{urb} \wedge \text{-seg}}) \leq \xi_0 \sqrt{n} \) for some \( \xi_0 \in (0, 1) \).
Appendix E: Proof of Theorem 9 (Urbanization and Integration)

Proof of Theorem 9. First, we define $\Omega^\text{urb\&seg} \subset \Omega$ to be the configurations that have $\epsilon-$urbanization of the poor and $(\beta, \delta)-$integration outside the urban area. To prove Theorem 9, it suffices to prove with all but exponentially small probability, a sample drawn from (2) is not in $\Omega^\text{urb\&seg}$: 

$$\pi(\Omega \setminus \Omega^\text{urb\&seg}) \leq \xi_3^{\sqrt{n}},$$

where $\xi_3 \in (0,1)$, and $n$ is sufficiently large.

We can further divide the configuration space $\Omega \setminus \Omega^\text{urb\&seg}$ into two parts: the set of configurations $\Omega^\text{urb}$ that do not have $\epsilon-$urbanization of poverty, and the set of configurations $\Omega^\text{urb\&seg}$ that have $\epsilon-$urbanization of poverty and $(\beta, \delta)-$segregation. Since $\Omega \setminus \Omega^\text{urb\&seg} = \Omega^\text{urb} + \Omega^\text{urb\&seg}$, to prove $\pi(\Omega \setminus \Omega^\text{urb\&seg}) \leq \xi_3^{\sqrt{n}}$, it suffices to prove $\pi(\Omega^\text{urb}) \leq \xi_1^n$ and $\pi(\Omega^\text{urb\&seg}) \leq \xi_0^{\sqrt{n}}$, for some constant $0 < \xi_1, \xi_0 < 1$, and sufficiently large $n$.

It follows from Theorem 3 that for $\gamma^{3/3} > 10^{344 + 4}$ and $\lambda > 1$, $\pi(\Omega^\text{urb}) \leq \xi_1^n$ for some $\xi_1 \in (0,1)$. To prove the second part $\pi(\Omega^\text{urb\&seg}) \leq \xi_0^{\sqrt{n}}$, for each $\sigma \in \Omega^\text{urb\&seg}$, we construct a $\delta-$color bridge system. Then we define the mapping $s = (s_2 \circ f_2)(\cdot)$: we do the color inversion and obtain $\tau = f_2(\sigma)$; next for $\tau$, we randomly flip the cyan to pink until the right number of the pink, and we randomly flip the blue to red until the right number of the red and obtain $\nu = s_2(\tau)$.

Claim 34. For any $\sigma \in \Omega^\text{urb\&seg}$ with bridged color contour length $z_c$, for the defined mapping $\nu = s(\sigma)$, $h(\nu) - h(\sigma) \leq -z_c + 3n$ and $p(\sigma) - p(\nu) \leq 0$.

Proof of Claim 34. Because the richness information remains unchanged for $f_1$ and $s_2$, $p(\sigma) - h(\nu) \leq 0$. It follows from Lemma 19 that for a given $z_c$ and any $\tau = f_2(\sigma)$, $h(\tau) - h(\sigma) \leq -z_c$. The maximal number of racially heterogeneous edges created by $\nu = s_2(\tau)$ can be bounded by $3n$, which is sum of all the edges in $G_\Delta$. Hence $h(\nu) - h(\sigma) = h(\nu) - h(\tau) + h(\tau) - h(\sigma) \leq 3n - z_c$. □

Claim 35. For a given color contour length $z_c$, for any $\nu = s(\sigma)$, the number of configurations in $\Omega^\text{urb\&seg}$ that can map to $\nu$ is upper bounded by: $|s^{-1}(\nu)| \leq (z_c + 1)3^{\sqrt{n}4^{344 + 4}z_c}2^{2n}$.

Proof of Claim 35. We denote $\tau = f_2(\sigma)$ and $\nu = s_2(\tau)$. Since $s(\sigma) = (s_2 \circ f_2)(\sigma)$, for any $\nu$, it follows that $|s^{-1}(\nu)| \leq |f_2^{-1}(\tau)| \cdot |s_2^{-1}(\nu)|$. For any $\nu \in \Omega$, the number of preimages $|s_2^{-1}(\nu)|$ can be upper bounded by $2^{2n}$ by recording whether each red agent is flipped or not.

It follows from Lemma 16 that for any given $z_c$, the number of $\delta-$color bridge systems can be upper bounded by $(z_c + 1)3^{\sqrt{n}4^{344 + 4}z_c}$. It follows from Lemma 19 that for any $\tau$, for a given $\delta-$bridge system, the number of preimages is one. Hence we conclude for any $\tau$ with a given $z_c$, the number of preimages is upper bounded by: $|f_2^{-1}(\tau)| \leq (z_c + 1)3^{\sqrt{n}4^{344 + 4}z_c}$. Combining the two inequalities, we conclude $|s^{-1}(\nu)| \leq (z_c + 1)3^{\sqrt{n}4^{344 + 4}z_c}2^{2n}$. □

Claim 36. For any given $\sigma \in \Omega^\text{urb\&seg}$, we denote $S(\sigma)$ to be the set of all possible images $\nu = s(\sigma)$ mapped from $\sigma$. It follows that $|S(\sigma)| \geq (\frac{2^d}{\rho})^{(n-p-d)a} \cdot (\frac{1-(p-d)}{r})^{(n-p-d)a}$.

Proof of Claim 36. It follows from the definition of $f_2$ that for any given $\sigma$, there is only one configuration can be obtained from $f_2(\sigma)$. For any given $\tau$, we define $S_2(\tau)$ to be the set of all possible configurations obtained from $s_2(\tau)$ by flipping the pink and red back; then $|S_2(\tau)| = \binom{pm - a_\tau}{r_\rho n - a'''_\tau} \cdot \binom{n - pm - a_\tau}{r_\tau n - a''_\tau}$, where $a_\tau$ is the number of unbridged agents in $\tau$, $a'''_\tau$ is the number of unbridged pink, and $a''_\tau$ is the number of unbridged red. Thus $|S_2(\tau)|$ can be further lower bounded
The following claim can be proved after the mapping.

\[ |S(\tau)| \geq \left( \frac{p - \delta}{r_p - a_r} \right)^{(r_p - a'_\nu)} \left( \frac{1 - p - \delta}{r_r - a_r} \right)^{(r_r - a'_\nu)} \geq \left( \frac{p - \delta}{r_p} \right)^{(r_p - \delta)} \left( \frac{1 - p - \delta}{r_r} \right)^{(r_r - \delta)}. \]  

(10)

Hence \( |S(\sigma)| \) can be lower bounded by \( |S(\sigma)| \geq 1 \cdot |S_2(\tau)| \geq \left( \frac{p - \delta}{r_p} \right)^{(r_p - \delta)} \left( \frac{1 - p - \delta}{r_r} \right)^{(r_r - \delta)}. \) \( \square \)

Finally, we define a weighted bipartite graph \( G(\Omega_{\text{urb} \wedge \text{seg}}, \Omega, E) \) with an edge of weight \( \pi(\sigma) \) between \( \sigma \in \Omega_{\text{urb} \wedge \text{seg}} \) and \( \nu \in \Omega \). The total weight of edges is

\[ \sum_{\sigma \in \Omega_{\text{urb} \wedge \text{seg}}} \pi(\sigma) \cdot |S(\sigma)| \geq \pi(\Omega_{\text{urb} \wedge \text{seg}}) \left( \frac{p - \delta}{r_p} \right)^{(r_p - \delta)} \left( \frac{1 - p - \delta}{r_r} \right)^{(r_r - \delta)}. \]  

(11)

On the other hand, the weight of the edges is at most

\[ \sum_{\nu \in \Omega} \sum_{\sigma \in \Omega_{\text{urb} \wedge \text{seg}}} \max_{\sigma \in \Omega_{\text{urb} \wedge \text{seg}}} \pi(\sigma) \cdot \sum_{\nu \in \Omega} \pi(\nu) \cdot \sum_{\sigma \in \Omega_{\text{urb} \wedge \text{seg}}} \max_{\sigma \in \Omega_{\text{urb} \wedge \text{seg}}} \pi(\sigma) \cdot |S^{-1}(\nu)| \]

\[ \leq \sum_{\nu \in \Omega} \pi(\nu) \sum_{z_c = \beta_{\min} \sqrt{n}} \lambda^{\max(h(\nu) - h(\sigma))} \gamma^{\max(p(\sigma) - p(\nu))} (z_c + 1) 3^{\alpha \sqrt{n}} 4^{\frac{3}{45}} 3z_c 2^{rn} \]

\[ \leq \sum_{z_c = \beta_{\min} \sqrt{n}} (z_c + 1) 3^{\alpha \sqrt{n}} 4^{\frac{3}{45}} 3z_c 2^{rn}. \]  

(12)

where the inequalities in Claim 35 and 36 has been substituted in the above derivation. Combining (11) and (12), we have

\[ \pi(\Omega_{\text{urb} \wedge \text{seg}}) \left( \frac{p - \delta}{r_p} \right)^{(r_p - \delta)} \left( \frac{1 - p - \delta}{r_r} \right)^{(r_r - \delta)} \leq \sum_{z_c = \beta_{\min} \sqrt{n}} (z_c + 1) 3^{\alpha \sqrt{n}} 4^{\frac{3}{45}} 3z_c 2^{rn}. \]  

(13)

For large enough \( n \), to have \( \pi(\Omega_{\text{urb} \wedge \text{seg}}) \leq \xi_3^3 \) for some \( \xi_3 \in (0, 1) \), it suffices to have

\[ \lambda^{3n} 2^{rn} \leq \left( \frac{p - \delta}{r_p} \right)^{(r_p - \delta)} \left( \frac{1 - p - \delta}{r_r} \right)^{(r_r - \delta)} \leq \left( \frac{p - \delta}{r_p} \right)^{r_p} \left( \frac{1 - p - \delta}{r_r} \right)^{r_r} \]

which can be rewritten as

\[ \lambda^3 < \left( \frac{p - \delta}{r_p} \right)^{r_p} \left( \frac{1 - p - \delta}{r_r} \right)^{r_r} / 2^r. \]

Since \( \lambda > 1 \), to make the right hand side of the above inequality greater than one, it suffices to have \( (\frac{p - \delta}{r_p})^{r_p} (\frac{1 - p - \delta}{r_r})^{r_r} > 2^r \). Combing the above parameter choices with Theorem 3, which requires \( \gamma^{\delta/3} > 16^{\frac{3}{45}} \) and \( \lambda > 1 \), we conclude Theorem 9. \( \square \)

**Appendix F: Proof of Theorem 11 (Integration for Distributed \( \mathcal{U} \))**

**Proof of Theorem 11.** First we define the configuration space \( S_{\beta, \delta} \) to be the set of configurations that are \((\beta, \delta)-\)segregated. To prove Theorem 11, it suffices to prove \( \pi(S_{\beta, \delta}) \leq \xi_4^4 \), where \( \xi_4 \in (0, 1) \). The bridging and the mapping \( \nu = f(\sigma) = (f_{5\circ f_4 \circ f_3 \circ f_2 \circ f_1})(\sigma) \) are defined as the following: we first construct a \( \delta \)-color-and-richness bridge system for \( \sigma \in S_{\beta, \delta} \) (see Appendix A.2 for details). Then we do richness inversion and color inversion like defined in \( f_1(\cdot) \) and \( f_2(\cdot) \). Then we do the color and richness recovery \( \nu = (f_5 \circ f_4 \circ f_3 \circ \cdot) \) in the distributed way as specified in section A.5. The following claim can be proved after the mapping.
Claim 37. For any \( \sigma \in S_{\beta, \delta} \) with a given bridged color contour length \( z_c \), for the defined mapping \( \nu = f(\sigma) \), \( h(\nu) - h(\sigma) \leq 2cn - z_c \) and \( p(\sigma) - p(\nu) \leq b_p n - \hat{b}_p n \), where \( \hat{b}_p \equiv \min\{c, p\} - (r + \delta)c - 2\delta \).

Proof of Claim 37. For any configuration \( \sigma \in S_{\beta, \delta} \), the poor agents on the urban sites are either in the red region \( R \) or outside \( R \). It follows from the definition of \((\beta, \delta)\)-segregation that the size of \( R \) is at most \((r + \delta)n\). Since the urban sites are evenly distributed with the total size \( c \cdot n \), the maximal number of the urban sites in \( R \) is \((r + \delta)n\). Hence the maximal number of the poor agents on the urban sites in \( R \) is \((r + \delta)n\). Outside \( R \), the poor agents on the urban sites could be unbridged agents whose number is upper bounded by \( \delta \cdot n \), or the cyan agents whose number is \( b_p n \). Hence the total number of the poor on the urban sites for any \( \sigma \) follows that \( p(\sigma) \leq (r + \delta)n + \delta n + b_p n \).

After the richness and color inversions, it follows from Lemma 25 that \( h(\sigma) - h(\nu) \leq -z_c \) and the number of the poor in \( \tau \) is less than \( 64n \) and they are unbridged. After the color and richness recovery, it follows from Lemma 27 that \( p(\nu) \geq \min\{c, p\}n - \delta n \). Hence \( p(\sigma) - p(\nu) \leq b_p n - \hat{b}_p n \), where \( \hat{b}_p = \min\{c, p\} - (r + \delta)c - 2\delta \).

It also follows from Lemma 24 that for any \( \nu = (f_5 \circ f_4 \circ f_3 \circ) (\tau) \), \( h(\nu) - h(\tau) \leq 2cn \). Combining with \( h(\tau) - h(\sigma) \leq -z_c \), we get \( h(\nu) - h(\sigma) \leq 2cn - z_c \).

Remark 38. If the number of the poor blue satisfies \( b_p < \hat{b}_p \equiv \min\{c, p\} - (r + \delta)c - 2\delta \), we can conclude the ratio between the poor blue and the poor red is smaller than the ratio between the blue and the red: \( \frac{b_p}{r_p} < \frac{b}{r} \), which is understood as income inequality.

Proof. If \( c \leq p \): it follows that \( b_p < c - (r + \delta)c - 2\delta = (b - \delta)c - 2\delta < (b - \delta)p < b \cdot p \), which can be written as \( \frac{b_p}{r_p} < \frac{b_p}{r_p} = b + \frac{b_c}{r_p} \). Hence we can get \( \frac{b_p}{r_p} < \frac{b}{r} \). If \( p \leq c \), it follows that \( b_p < p - (r + \delta)c - 2\delta < p - (r + \delta)p - 2\delta < p(b - \delta) < b \cdot p \). Hence the same conclusion \( \frac{b_p}{r_p} < \frac{b}{r} \) follows.

For a given color contour length \( z_c \), for any \( \nu = f(\sigma) \), the number of preimages follows from Claim 27. Similarly, we use Peierls Argument (5), substituting the related bounds into which yields

\[
\pi(S_{\beta, \delta}) \leq \sum_{\nu \in \Omega} \sum_{z_c} \pi(\nu) n^6(3n + 1)(z_c + 1)9^{\alpha \sqrt{n}}(\frac{4^{3n+1}}{\lambda})z_c(\frac{\lambda^{\frac{2e}{2^{3+1}}}}{\gamma^{b_p-b_c}})^n, \tag{14}
\]

where \( z_c \geq \sqrt{r \cdot n} \) is due to the triangular lattice geometry, which is proved in Lemma 2.1 in [7], and \( z_c \leq \beta \sqrt{r} \) is due to \( \sigma \in S_{\beta, \delta} \) and the definition of \((\beta, \delta)\)-segregation. If \( b_p < \hat{b}_p \) and \( \gamma^{b_p-b_c} > \lambda^{\frac{2e}{2^{3+1}}} \), the sum will be exponentially small given large enough \( n \), which means \( \pi(S_{\beta, \delta}) \leq \xi_4^{z_c} \) for some \( \xi_4 \in (0, 1) \).