Clustering and Mixing Times for Segregation Models on $\mathbb{Z}^2$

Prateek Bhakta [*] Sarah Miracle [†] Dana Randall [‡]

Abstract

The Schelling Segregation Model was proposed by Thomas Schelling in 1971 as a means of explaining possible causes of racial segregation in cities. He considered residents of two types, say red and blue, where each person prefers the majority of his or her neighbors to have the same color. He showed through simulations that even mild preferences of this type can lead to segregation if residents move whenever they are not happy with their local environments. Economists have considered many natural variants on the model, with similar findings, but there has been very little rigorous work justifying the claims about segregation. Here we introduce a General Influence Model to capture some of these variants in a utillity function, which we study in the context of a reversible Markov chain in which the more unhappy residents are with the demographics of their local environment the more likely they are to move. We use insights from convergence rates of Markov chains to then characterize the stationary distributions to determine whether cities are likely to be integrated or segregated (or clustered).

The General Influence Model considers open cities (where residents can move away) in a saturated or non-saturated setting (so we can allow unoccupied houses), with neighborhoods of any radius, and where moving is based on the product of everyone's happiness. An individual's happiness depends on the demographics in his local neighborhood, and the influence function dictates the probability of moving depending on this function. We show that for any influence function, the dynamics will be rapidly mixing and cities will be integrated (i.e., there will not be clustering) if the racial bias is sufficiently low. We also show complementary results for two broad classes of influence functions. The first is for Increasing Bias Functions (IBF), where an individual's likelihood of moving increases each time someone of the other color moves close or someone of the same color leaves (this does not include Schelling's threshold models). The second is for Threshold Bias Functions (TBF) when the threshold is at least one half, including the model that Schelling originally proposed. Here a resident is happy as long as the majority of his neighbors share his color, and is unhappy otherwise, regardless of the actual percentage. For both classes (IBF and TBF) we show that when the bias is sufficiently high, the dynamics take exponential time to mix and we will have segregation, which in the case of an open city means that a large ghetto will form.

[*]Supported in part by NSF CCF-1219020; pbhakta@gatech.edu.
[†]Supported by a DOE Office of Science Graduate Fellowship and NSF CCF-1219020; sarah.miracle@gatech.edu.
[‡]Supported in part by NSF CCF-1219020; randall@cc.gatech.edu.

1 Introduction

The Schelling Segregation Model was introduced in 1971 to explain how global behavior can arise from small individual preferences [20]. In Schelling's original model, agents are one of two colors and move if there are too many neighbors of the opposite color within their immediate neighborhood. Simulations show that configurations rapidly become segregated with like colored neighbors clustered together. Schelling used this simple model to argue that "micro-motives" can determine "macro-behavior," thereby forming the basis for Agent-Based Computational Economics.

Despite extensive research on the Schelling model and its many variants, almost everything remains non-rigorous. Our goal here is to consider families of Schelling models in an attempt to put them on firmer footing. There are many natural extensions worth considering: How large a neighborhood is relevant to one's happiness, and do all neighbors within this neighborhood influence us equally? Can residents move away, or are they restricted to remain in the city? Are all houses occupied, or are there empty houses (say, foreclosures) that might be even less desirable to have in one's proximity? Is one's happiness determined solely by the color of the majority of one's neighbors, as Schelling originally proposed, or does one get increasingly happy or unhappy as new people of one color or the other move into the neighborhood? Are decisions to move somewhere based on each person's relative happiness, or is one less likely to move to a house where he is not wanted if doing so decreases the happiness of his new neighbors?

Economists and social scientists use statistical and non-rigorous computational tools to study the dynamics and limiting distributions, as well as for connecting the model to real world populations [7, 19, 1, 25, 26]. Even the concept of segregation or clustering typically is not formally defined or characterized. An exception is the rigorous analysis of the Schelling model in the one-dimensional setting by Young [28]. More recently, Schelling's original model was analyzed in one-dimension by Brandt et al. [4], and Lewis-Pye et al. [13]. Additional rigorous work has considered further variations designed to simplify the neighbors' interactions for some specific, basic models [8, 29, 13, 19].

The concept of micro-motives effecting macro-behavior is well-studied and far better understood in the statistical physics community, where it is used to explain fundamental concepts such as phase transitions. The Schelling model itself is reminiscent of many physical models, most notably the Ising model of ferro-magnetism. In the Ising model, vertices of a graph, say a finite region $G = (V, E)$ of $\mathbb{Z}^2$, are assigned + or - spins, and neighboring vertices prefer to have the same spin. Although in the original Schelling model a person's happiness depends only on the color of the majority of his neighbors, in the Ising analogue everyone is incrementally more likely to move as more people of the opposite color move into their neighborhood.

Specifically, in the Ising model we are given a parameter λ that is a function of temperature, and the stationary probability of a configuration $\sigma \in \{\pm 1\}^V$ is

$$\pi(\sigma) = \lambda^{|\{x,y:\ (x,y)\in E,\ \sigma(x)=\sigma(y)\}|}/Z,$$

where $Z = \sum_{\sigma\in\{\pm1\}^V} \lambda^{|\{x,y:\ (x,y)\in E,\ \sigma(x)=\sigma(y)\}|}$ is the normalizing constant known as the partition function. Glauber dynamics is a Markov chain on Ising configurations that changes one spin at a time using Metropolis probabilities to force the chain to converge to π. The Ising model on $\mathbb{Z}^2$ is known to undergo a *phase transition*, i.e., there exists a value λ_c such that when $\lambda < \lambda_c$, the Glauber dynamics for the Ising model mixes in time polynomial in $|V|$ and when $\lambda > \lambda_c$, it mixes in exponential time [12, 17, 14, 24]. Moreover, the phase transition in the mixing time is accompanied by a corresponding transition in the stationary distribution of the Markov chain; at low λ, an average sample from the steady state is "evenly mixed" with regards to the proportions of spins, while at high lambda, an average sample is *clustered*, and has large regions of predominantly one spin type. Indeed, the Ising model has been studied empirically as an alternative to the Schelling model [19, 22, 23]. In open systems at low temperature (high bias) the population will become predominantly one color or the other, and in closed systems (arising as a fixed magnetization Ising model), large clusters of one color (or spin) will form, indicating segregation [24, 27].

While extensions of the Ising model on $\mathbb{Z}^2$ have been examined extensively by physicists and mathematicians, the resulting models are typically less-tractable and give little insight into Schelling variants (such as neighborhoods of size larger than 4, unoccupied houses, or bias functions that do not scale geometrically with the number of differently colored neighbors). A lot is known about the Ising model on graphs with more than nearest-neighbor interactions see, e.g., Chapters 2 and 9 of [16] and general spin systems on $\mathbb{Z}^d$ have been shown to have a phase transition whenever there is a phase transition in the associated mean field model for certain classes of interactions [3, 2, 6]. However, while these results apply only to certain classes of interactions, they fail to give insight into more general utility functions which more closely resemble the original Schelling model.

Results: We consider a generalization of the Schelling model called the General Influence Model (GIM) and give rigorous results demonstrating a dichotomy in mixing times and clustering for two broad classes. The GIM considers *open cities* in a *non-saturated* setting, with neighborhoods of *any radius*, and where moving is based on the *product of everyone's happiness*. *Open cities* allow residents to move away, while *closed cities* require fixed racial demographics. *Unsaturated cities* allow houses to be unoccupied. An individual's happiness is a function depending only on the number of unoccupied, red and blue houses within a certain radius. This function can be a threshold, as suggested by Schelling, a geometric function, similar to the Ising model, or anything else. Moreover, these influence functions are controlled by parameters measuring the strength of these biases, so for any influence function we can study the effects of large or small racial bias.

First, we consider a natural extension of the Schelling dynamics where people move according to the relative global happiness and we analyze the mixing time, or the time to approach equilibrium. The relevance of bounding the mixing time to understanding Schelling dynamics is indirect and will help us discern properties of the stationary distribution. Second, we formalize a concept of clustering in order to predict when typical configurations are likely to be segregated or integrated. We show that for any influence function, the dynamics will be fast mixing and cities will be integrated (i.e, there will not be clustering) if the racial bias is sufficiently low. Next, we show complementary results for two broad classes of influence functions. The first is for Increasing Bias Functions (IBF), where an individual's likelihood of moving increases each time someone of the other color moves close or someone of the same color leaves (this does not include Schelling's threshold model). The second is for Threshold Bias Functions (TBF) when the threshold is at least one half, including the model that Schelling originally proposed. Here a resident is happy as long as the majority of his neighbors share his color, and is unhappy otherwise, regardless of the actual percentage. For both classes (IBF and TBF) we show that when the bias is sufficiently high, the dynamics take exponential time to mix and we will have segregation. Note that because we are considering open cities, segregation means the city will become predominantly one color, a large ghetto, and slow mixing means that it will take exponentially long for the city to transition from a ghetto of one color to one of the other color. (We have initial results showing that these results can be extended to closed cities where our definition of clustering also holds for populations with any fixed racial demographics.)

Techniques: The proofs of fast mixing and integration at low bias use standard coupling and information-theoretic arguments. The proofs of slow mixing and segregation at high bias are subtle and significantly more challenging. In fact, it is not clear whether the latter results extend to the whole class of GIMs, as our proofs only verify that segregation occurs in the IBF and TBF settings.

The strategy used to show slow mixing of Markov chains and clustering effects is a *Peierls ar-*

gument, which originated in physics in order to study Gibbs measures on the infinite lattice. The argument works by showing certain types of configurations are exponentially unlikely by using combinatorial maps and information theory. In the context of Markov chains, Peierls arguments can be used to show that cut sets in the state space are exponentially unlikely, and this is sufficient to show that the Markov chain will require exponential time to converge to equilibrium. Similarly, in the context of clustering, we can use a similar argument to show that configurations that are integrated, or lack large clustered components, also have exponentially small probability at equilibrium.

The proofs of slow mixing build on some techniques established previously, but these pieces had to be put together in novel ways. We use a strategy introduced in [18] to partition the state space according to topological features, namely monocrhomatic crosses (similarly colored neighboring houses that connect all four sides of the housing region) and fault lines, or long paths separating houses of different colors. Configurations with fault lines form the cut in the state space, and our objective is then to show that they have exponentially small probability. For the Ising model on $\mathbb{Z}^2$, for instance, completing the argument is simple because we can reverse the spins (or flip the colors) of all houses on one side of the fault to move to a new configuration with exponentially larger stationary probability. The introduction of unoccupied houses complicates this approach, but we use a technique used in [10] by characterizing the cut as configurations with "fat faults." The greater challenge occurs when the radius of influence is larger than 1 and residents are equally influenced by neighbors up to r houses away, for $r > 1$. In this case faults or fat faults are not sufficient and reversing the colors on one side of a fault can actually decrease the probability of a configuration. To address this we introduce the notion of bridges and build a complex of fat faults connecting components that are within distance r.

The arguments are fine tuned to the specific classes, IBF, where everyone gets increasingly happy as more people of their color move into their neighborhood, and TBF, where residents are unhappy unless some threshold over 50% is reached. Either of these conditions give us the leverage to push through the Peierls argument and show that the cutset has exponentially small probability. The significance of 50% is that if we change the color of a resident who is currently happy then he necessarily becomes unhappy, and this only happens in a threshold model when the threshold is beyond one half.

Many elements of the slow mixing results are then used to establish clustering for the IBF and TBF models when the bias is high. We characterize clustering in configurations by the existence of a region R that has large (quadratic) area, small (linear) perimeter, and whose interior is dense with one of the two colors. A similar notion of clustering was used in [15], but the proofs required the introduction of r-bridges and fat contours to handle unoccupied houses and large radii of influence.

In Section 2 we formalize the GIM, define clustering, and give a brief overview of Markov chains and mixing times. In Section 3 we provide the proofs of fast mixing for all influence functions at low bias and slow mixing for the IBF and TBF classes at high bias. Finally, in Section 4 we give the corresponding proofs for integration at low bias and segregation at high bias.

2 Preliminaries

We first formalize our generalization of the Schelling model, which we call the *General Influence Model* (GIM), and present some fundamental facts about mixing times of Markov chain and clustering. Let Ω be the set of all 3-colorings of the faces of the n x n grid G_n, where the colors represent the types of occupants in a housing grid. We label the possible colors B,R and U where B and R represent two types of residents, red and blue, U represents an unoccupied house and we refer to each of these as B, R, or U-faces respectively (see e.g., Figure 1). An *occupied* face refers to a B or

R-face. We denote the color of face x in configuration σ as $\sigma(x)$. To simplify our notation, we let $\sigma_{x_1=c_1,x_2=c_2,\ldots}$ denote the configuration σ with face x_i colored c_i, for each specified i.

We consider a natural Markov chain $\mathcal{M}$ on Ω whose transitions alter the color of one face at a time. We select a face $x \in G_n$ and a color $c \in \{B, R, U\}$ uniformly at random, then set face x to color c with probability that depends on the total change in "happiness" of the configuration. The happiness of any occupied face is determined by the colors of faces within a radius of r, and the weight of a configuration is the product of the happiness of each occupied face.

Formally, we are given a fixed radius r as a parameter of the model. Each resident (or occupied face) is influenced equally by all $N = 2r^2 + 2r$ *neighbors* which we define as faces within taxicab distance r. We are aslo given a utility function $u : \{(s, o) : s, o \in [0, N], s \geq o\} \to [0, 1]$, that relates the coloring of a resident's neighborhood to its happiness with an arbitrary bias (or utility) function. Let $s(\sigma, x)$ be the number of neighbors of x that have the same color as x in σ and $o(\sigma, x)$ be the number of neighbors of x which are occupied (i.e. R- or B-faces) in σ. The happiness of an occupied face x is defined to be $u(s(\sigma, x), o(\sigma, x))$. For our model, we require that $u(0, 0) = 0$ and $u(N, N) = 1$ for normalization purposes. We also require that u is a non-decreasing function in both parameters, so that one prefers an oppositely colored neighbor to an abandoned house.

The weight π of a configuration is

$$\pi(\sigma) = \prod_{x:\sigma(x)\neq U} \lambda^{u(s(\sigma,x),o(\sigma,x))}/Z,$$

where $Z = \sum_{\sigma\in\Omega} \prod_{x:\sigma(x)\neq U} \lambda^{u(s(\sigma,x),o(\sigma,x))}$ is the normalizing constant.

The Markov chain $\mathcal{M}$.[1] Starting at any σ_0, at step t iterate the following:

- Choose a face x of G_n, and a color $c \in \{B, R, U\}$ uniformly at random.
- If $\sigma_t(x) = U$, with probability 1 let $\sigma_{t+1} = \sigma_{t,x=c}$.
- If $\sigma(x) = R$ and $c = U$, with probability $\pi(\sigma_{t,x=U})/\pi(\sigma_{t,x=R})$ let $\sigma_{t+1} = \sigma_{t,x=c}$.
- If $\sigma(x) = B$ and $c = U$, with probability $\pi(\sigma_{t,x=U})/\pi(\sigma_{t,x=B})$ let $\sigma_{t+1} = \sigma_{t,x=c}$.
- With the remaining probability, let $\sigma_{t+1} = \sigma_t$.

This Markov chain connects the state space since we can always reach the empty configuration.

The General Influence Model (GIM) is a generalization of many well-studied models on the grid. For example, if we let $r = 1 (N = 4)$, and $u(s, o) = s/4$, then (after a suitable change of variables), this model is equivalent to the non-saturated Ising model on the grid [9]. Here, B-faces correspond to + spins and R-faces to $-$ spins. The influence on a site is the number of matching neighbors, and the fact that $u(s, o) = s/4$ means that this influence is linearly proportional to the corresponding exponent of λ in the weight of the configuration. If instead we let $r = 1$ and $u(s, o) = U_0(2s - o)$, where U is a step function, then this model corresponds to a reversible version of the Schelling Model [23, 19]. Here, a site is "happy" if it has at least as many neighbors of the same color as the opposite color. If we let $r = 1$, and $u(s, o) = U_{N/2}(s)$, we have another variant of the Schelling Model where a site is "happy" if at least half of its neighbors are of the same color.

We will state our results in terms of bounds on the discrete partial derivatives of u. In particular, let $u'_\alpha = \min_{a,b}\{u(a + 1, b) - u(a, b)\}$, and similarly let $u'_\beta = \max_{a,b}\{u(a + 1, b) - u(a, b)\}$, $u'_\kappa = \min_{a,b}\{u(a + 1, b + 1) - u(a, b)\}$, and $u'_\gamma = \max_{a,b}\{u(a + 1, b + 1) - u(a, b)\}$.

[1]We present the results in the unsaturated setting where we allow empty houses. For the saturated model the Markov chain allows houses to move between B and R in one move, indicating that a new resident will move in as soon as one vacates a house. All of the proofs carry over in this case and are in fact simpler.

For all $\epsilon > 0$, the *mixing time* $\tau(\epsilon)$ of $\mathcal{M}$ is defined as $\tau(\epsilon) = \min\{t : \max_{x\in\Omega} \frac{1}{2}\sum_{y\in\Omega}|P^t(x,y) - \pi(y)| \leq \epsilon, \forall\ t' \geq t\}$. We say that a Markov chain is *rapidly mixing* if the mixing time is bounded above by a polynomial in n and $\log(\epsilon^{-1})$ and *slowly mixing* if it is bounded below by an exponential function. In Section 3, we bound the mixing time of the Markov chain $\mathcal{M}$ under different conditions.

Next, we formalize a definition for *clustering*, building on the approach given in [15]. We will define a configuration as clustered if it contains a large region densely filled with R or B-faces.

Formally, we will define a *region* $C = (C_F, C_E)$ where C_F is a set of faces in the grid G_n and C_E is a set of edges where C_E is connected and any edge e which is adjacent to a face in C_F and a face in $\bar{C_F} = G_n \backslash C_F$ satisfies $e \in C_E$. The perimeter of a region C is $|C_E|$.

Definition 2.1: Given a configuration $\sigma \in \Omega$, we say that the X-faces are *c-clustered* if σ contains a set of tiles (region) C satisfying:

1. the perimeter of C is at most cn,
2. the density of X-faces in C is at least c and in $\bar{C}$ is at most $1-c$ and
3. all faces contained in C_F and adjacent to an edge in C_E are colored X.

In Section 4, we show that a random sample from our model will be exponentially likely to be clustered in some conditions, and exponentially unlikely to be clustered in other conditions.

3 Mixing time results

We begin by studying the mixing time of $\mathcal{M}$. First, we show that for any IBF and TBF utility function, $\mathcal{M}$ is slowly mixing when λ is sufficiently high. Then we show for all utility functions u, $\mathcal{M}$ is rapidly mixing if λ is sufficiently low.

3.1 Slow Mixing At High λ

We introduce the idea of r-extended contours to show that for sufficiently large values of λ, $\mathcal{M}$ mixes exponentially slow when the utility function is in the IBF or TBF classes. We make use of the well known relationship between the *conductance* and mixing time of a Markov chain to show that three sets Ω_B, Ω_R and Ω_F, which we will define shortly, partition the state space with Ω_F being a cutset with exponentially small weight. This lets us conclude that the *conductance* of the chain is small, and we can conclude the chain mixes exponentially slowly. (See [11, 21] for details.) For an ergodic Markov chain $\mathcal{M}$ with distribution π, the conductance of a subset $S \subseteq \Omega$ is defined as $\Phi(S) = \sum_{s_1\in S, s_2\in\bar{S}} \pi(s_1)P(s_1,s_2)/\pi(S)$. The conductance of the chain $\mathcal{M}$ is then $\min_{S\subset\Omega}\{\Phi(S) : \pi(S) \leq 1/2\}$ and the mixing time $\tau(\epsilon)$ satisfies [11]

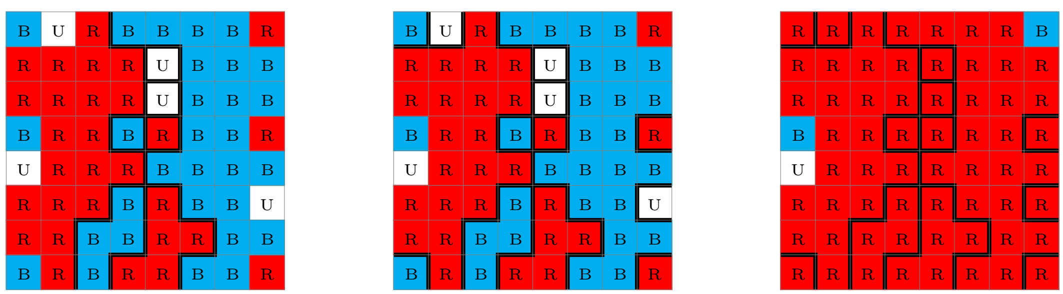
$$\tau(\epsilon) \geq \left(\frac{1-2\Phi}{2\Phi}\right)\ln \epsilon^{-1}.$$

In order to define the three sets that form our cut we start with some terminology. We call a pair of faces within taxicab distance r to be an *influence*, and refer to this as a *bad influence* if the two faces are colored differently or are both U-faces. Influences at distance 1, adjacent faces, we call *edges* since they correspond to edges of the $n \times n$ grid. We define a *contour* to be a connected set of bad edges and a *fat contour* (see [10]) to be a maximally connected set of bad edges.

A fat contour, or set of fat contours, partitions the faces of the grid into regions whose border along any single fat contour is monochromatic. With respect to a single contour, we call these R-regions, B-regions, etc. to denote the color along their border. Note that the entire regions are not necessarily monochromatic, as a B-bordered region may fully enclose a set of R faces that do not border the contour. Also note that U-regions are single squares, since all 4 sides of a U-face are bad edges. Given two fat contours c_1 and c_2, c_1 is within distance r of c_2 if there exists a face adjacent to c_1 that is within taxicab distance r of a face adjacent to c_2, and these faces are in different regions, where the regions are the unique regions defined by c_1 and c_2. We can think of all the disjoint fat contours of a configuration to be connected to each other in an auxiliary graph if they are within distance r of each other. We then define an *r-extended contour* to be the union of all fat contours in a maximally connected component of this auxiliary graph.

We say that a configuration has a *monochromatic cross* if it has a B- or R-region that connects all four sides of the grid. A fat contour that spans from the top to bottom or left to right of the grid is a *fault line*. We use the fact that every configuration falls into one of three disjoint classes: Ω_B (those with a B-cross), Ω_R (those with a R-cross), and Ω_F (those with a fault line). It is known that Ω_B, Ω_R, and Ω_F partition the state space Ω, and moves of the Markov chain $\mathcal{M}$ cannot directly move from Ω_B to Ω_R or vice-versa, and thus must move through Ω_F [10].

Our goal is to show that Ω_F is an exponentially small cut in our state space by exhibiting a mapping $\phi_r : \Omega_F \to \Omega$ such that for any $\sigma \in \Omega_F$, the image $\phi_r(\sigma)$ "fixes" a fault line by reversing the colors in some of the monochromatic regions that border the r-extended contour containing the fault line. This causes many more same-color interactions, yielding a gain $\pi(\phi_r(\sigma))/\pi(\sigma)$ that is exponentially large in n. This gain is exponentially larger than the total weight of all potential preimages $\in \Omega_F$ of any state $\in \Omega$, from which we can conclude that $\pi(\Omega_F)$ is exponentially small.



(a) (b) (c)

Figure 1: (a) A configuration σ with a fault line, (b) the 1-extended fault (c) and $\phi(\sigma)$.

We construct $\phi_r(\sigma)$ for $\sigma \in \Omega_F$ as follows (see Figure 1).

- Take the lexicographically first fault line in σ.
- Find the r-extended contour (and associated regions) which contains this fault line.
- Finally, for the regions defined by the r-extended contour, map all U-regions to R-faces and within any B-region change all R-faces to B-faces and all B-faces to R-faces.

We note that all faces within r of the fat fault line in σ will map to R-faces in $\phi_r(\sigma)$. This map causes all elements within distance r of the fault line to be mapped to R-faces. We also note that no bad influences are created by the map ϕ_r between previously good influences - this can only happen to faces P and Q if they are within r of each other, and also in different fault regions. However, if they are in different fault regions, some fault edge must pass through any shortest path between P and Q, and the r-extended contour would necessarily pick up the borders of the monochromatic regions containing P and Q. Thus, the mapping ϕ_r would cause both P and Q to map to R-faces.

We now bound the number of pre-images of a configuration $\phi_r(\sigma)$. Starting on one of $4n$ points on the border, a r-extended contour can be expressed by a depth first search of m edges, using at most $2m$ steps, and each step travels in up to $2r^2 + 2r$ directions. Each monochromatic region is surrounded by at least four edges, and each edge is on the boundary of two regions. Thus, there are at most $m/2$ distinct regions bordering this contour, each of which can be colored one of 3 ways. Therefore, there are at most $4n3^{m/2}(2r^2+2r)^m$ pre images σ such that $\phi_r(\sigma)$ fixes this contour.

Increasing Bias Functions. We first present result for utility functions u with bounded u'_α.

Theorem 3.1: For the Markov Chain $\mathcal{M}$, with radius r and utility function u with $u'_\alpha > 0$, there exists a constant $\lambda_1 = \lambda_1(r, u'_\alpha)$ such that $\mathcal{M}$ mixes exponentially slowly when $\lambda > \lambda_1$.

PROOF: We partition Ω_F into sets $\Omega_{F,m}$ where $\sigma \in \Omega_{F,m}$ if m is the number of bad edges fixed by ϕ_r. We observe that for two adjacent faces I and J with a bad edge, every face that influences both I and J will share a bad influence with at least one of them. Thus each of these $2r^2-2$ faces, excluding I, J, gains at least one new neighbor of the same type, which causes an increase of happiness of at least u'_α. Any one influence between any P and Q is counted at most 8 times in this way, once for each potential bad edge bordering P or Q. Also, the happiness of both P and Q improve from is. Thus, we see a gain of at least $u'_\alpha((2r^2-2)/4+1)$ per face bordering the fault line. Thus by fixing an r-extended contour with m edges our gain is at least

$$\pi(\phi_r(\sigma))/\pi(\sigma) \geq (\lambda)^{u'_\alpha \frac{m}{4}(2r^2-1)} \geq (\lambda)^{u'_\alpha \frac{mr^2}{4}}. \tag{3.1}$$

Now, let $\lambda > \lambda_1 = (9(4r^2+4r)^4)^{(r^2 u'_\alpha)^{-1}}$. Then we have:

$$\begin{aligned}
\pi(\Omega_F) &= \sum_{m=n}^{2n^2} \sum_{x\in\Omega_{F,m}} \pi(\phi_r(x)) \frac{\pi(x)}{\pi(\phi_r(x))} \leq \sum_{m=n}^{2n^2} \sum_{x\in\Omega_{F,m}} \pi(\phi_r(x))(\lambda^{u'_\alpha})^{-mr^2/4} \\
&\leq \sum_{m=n}^{2n^2} 2n(2r^2+2r))^m \cdot 3^{m/2}(\lambda^{-u'_\alpha mr^2/4}) \leq \sum_{m=n}^{2n^2} 2n2^{-n/4} \leq 4n^3 2^{-n/4}.
\end{aligned}$$

Combining this bound on $\pi(\Omega_F)$ with the detailed balance condition gives the following.

$$\Phi(\Omega_R) = \sum_{s_1\in\Omega_R, s_2\in\bar{\Omega_R}} \pi(s_1)P(s_1,s_2)/\pi(\Omega_R) \leq \sum_{s_1\in\Omega_R, s_2\in\bar{\Omega_F}} \pi(s_2)P(s_2,s_1)/\pi(\Omega_R) \leq \pi(\Omega_F)/\pi(\Omega_R).$$

By symmetry, we know that $\pi(\Omega_R) = \pi(\Omega_B) = (1-\pi(\Omega_F))/2$. Thus, the conductance of $\mathcal{M}$ is at most $\Phi(\Omega_R) \leq \pi(\Omega_F)/\pi(\Omega_R) = 2\pi(\Omega_F)/(1-\pi(\Omega_F)) \leq 2\pi(\Omega_F) \leq 8n^3 2^{-n/4}$. It follows that $\tau(\epsilon)$, the mixing time of $\mathcal{M}$, satisfies $\tau(\epsilon) \geq (n^{-3}2^{n/4-4}-1)\ln \epsilon^{-1}$. ■

Threshold Bias Functions. We now consider the Schelling variant where a face needs θ matching neighbors to be happy, so $u(s,o) = U_\theta(s)$, where U is a step function with threshold θ. Here $u'_\alpha = 0$ so we cannot apply the bounds in the previous subsection. However, a key observation allows us to apply our technique to a certain class of threshold utility functions as follows.

Theorem 3.2: For the Markov Chain $\mathcal{M}$, with radius r, neighborhood size $N = 2r^2+2r$, threshold $\theta > (\frac{1}{2}+\frac{1}{2r+2})N$ and utility function $u(s,o) = U_\theta(s)$, there exist a constant $\lambda_2 = \lambda_2(r)$ such that $\mathcal{M}$ mixes exponentially slow when $\lambda > \lambda_2$.

PROOF: We again partition Ω_F into sets $\Omega_{F,m}$ where $\sigma \in \Omega_{F,m}$ if m is the number of bad edges fixed by ϕ_r. Again, every two adjacent faces I and J with a bad edge shares a neighborhood of $2r^2-2$ faces, excluding I and J. Thus if $\theta > r^2+2r = (2r^2+2r)(\frac{1}{2}+\frac{1}{2r+2})$, *both I and J cannot be happy.* Thus the mapping ϕ_r will cause at least one of I and J to become happy (from unhappy), leading to a gain of 1 per edge of the fault line. This gain is counted at most 4 times, once for each edge bordering the fixed tile. Thus, we see a gain of at least 1/4 per such face, and a gain of at least $m/4$ by fixing a contour of size $m/4$. Again, we let $\lambda > \lambda_2 = (9(4r^2 + 4r)^4)$. Then we have:

$$\pi(\Omega_F) \le \sum_{m=n}^{2n^2} \sum_{x \in \Omega_{F,m}} \pi(\phi_r(x))(\lambda^{u'_\alpha})^{-m/4} \le \sum_{m=n}^{2n^2} 2n(2r^2 + 2r))^m \cdot 3^{m/2}(\lambda^{-m/4}) \le 4n^3 2^{-n/4}.$$

Again, it follows that $\tau(\epsilon)$, the mixing time of $\mathcal{M}$, satisfies $\tau(\epsilon) \ge (n^{-3}2^{n/4-4} - 1)\ln \epsilon^{-1}$. ∎

3.2 Rapid Mixing at Low λ

In contrast, we show that when λ is sufficiently low, we can guarantee that the chain mixes in polynomial time for all utility functions. Our bound on λ depends on the discrete partial derivative

$$u'_\gamma = \max_{a,b}\{u(a+1, b+1) - u(a, b)\}.$$

The proof relies on the now standard path coupling technique (see, e.g., [5]). We prove the following.

Theorem 3.3: For the Markov Chain $\mathcal{M}$, with radius r and utility function u, there exists a constant $\lambda_3 = \lambda_3(r, u'_\alpha)$ such that $\mathcal{M}$ is fast mixing when $1 \le \lambda < \lambda_3$.

Proof: We use a path coupling argument with the natural coupling. Notice that a move of $\mathcal{M}$ consists of selecting a face f and a color c. The coupling simply uses the same face and color for both configurations. The distance metric we use is the minimal number of steps of $\mathcal{M}$ required to change one configuration into another. Given a pair of configurations, each face occupied in exactly one configuration contributes 1 to the distance while each face colored R in one configuration and B in the other contributes 2. Thus, the maximum distance between any two configurations is $2n^2$.

To apply the path coupling theorem, we consider the set of pairs of configurations at distance 1 and show that for any such pair after one step of $\mathcal{M}$ the distance between the configurations decreases in expectation. In order to bound the expected change in distance, without loss of generality, we consider two configurations $(\sigma_{g=R}, \sigma_{g=U})$. Let f be the face selected by $\mathcal{M}$. The distance only decreases if $f = g$ and can increase if $f \ne g$. In either situation we consider three cases based on the color selected. The rest of the proof uses standard arguments we defer the details to Section A.1. ∎

4 Segregation, Integration and the Stationary Distribution

We now use our insights from Section 3 to show a similar dichotomy in terms of configurations chosen from the stationary distribution at high and low λ. When λ is large ghettos will form, and configurations will be predominantly one color; when λ is small there will be no clustering of one type and cities will remain integrated. Our proofs build on the combinatorial insights developed in Section 3.1 and in [15] to show clustering for colloids.

4.1 Segregation at High λ for the IBF and TBF classes

First, we use the combinatorial techniques developed in Section 3.1 to argue that at high λ configurations will be segregated. In open cities we expect a single ghetto of predominantly R- or B-faces. We prove that at high values of λ, a typical configuration will have no large contours and will have high density of either R- or B-faces. We combine techniques used to show clustering [15] with the slow mixing techniques used in Section 3.1. Let ρ_R be the density of R-faces and ρ_B be the density of B-faces. We prove the following theorem showing ghettos will form.

Theorem 4.1: Assume a valid utility function u with radius r such that $u'_\alpha > 0$ or u is a threshold utility function with $\theta > (\frac{1}{2} + \frac{1}{2r+2})N$, where $N = 2r^2 + 2r$. Given a constant density $d_1 > 1/2$, there exist constants $\gamma_1 = \gamma_1(d_1) < 1$ and $\lambda_1 = \lambda_1(u'_\alpha, r, d_1)$ such that for all $\lambda \geq \lambda_1$ a random sample from Ω will have no contours with more than $d_1 n$ edges and either the density $\rho_R > d_1$ or $\rho_B > d_1$ with probability at least $(1 - \gamma_1^n)$.

PROOF: Using an extension of the techniques from 3.1 we show that it is exponentially unlikely for a configuration to have any contour with size greater than $d_1 n$ and also that it is exponentially unlikely for $\rho_R, \rho_B < d_1$. Using the union bound we then combine these two results. To show there is no large contour, we construct a map ϕ_{d_1} from configurations with contours of size greater than $d_1 n$ to configurations which have at least one less contour of size greater than $d_1 n$. As in Section 3.1, ϕ_{d_1} takes the lexicographically first contour of size greater than $d_1 n$, finds the r-extended contour which contains this contour, changes all U-faces bordering the r-extended contour to R-faces and flips all B-bordered regions adjacent to the contour. Unlike Section 3.1, our contour is not necessarily anchored to the border but the extension is straightforward and we leave the details to Section A.2.

To show that it's exponentially unlikely for $\rho_R, \rho_B < d_1$ we construct a map ϕ_S which locates a sufficiently large set of r-extended contours and removes them. Given a set S of r-extended contours, the *size* of the set which we denote as $|S|$ is the sum of the sizes of the distinct r-extended contours contained in S. We show there exists a row P in the grid and a set S of r-extended contours with $|S| \geq (\frac{1-d_1}{2})n$ such that each r-extended contour in S contains at least one vertical edge along P.

Next, we bound the number of pre-images of a configuration under ϕ_S, using an argument similar to Section 3.1. There are n possible rows P and for each choice of P there are 2^n different sets of starting points for our depth first search. Given the set of starting points, a depth first search of m edges takes at most $2m$ steps and each move travels in up to $2r^2+2r$ directions. Thus there are now $n2^n3^{m/2}(2r^2+2r)^m$ configurations with a row P and set S with $|S| = m$ and each r-extended contour in S intersecting P. Unlike Section 3.1, we only guarantee that we are "fixing” at least $\frac{1-d_1}{2}n$ bad edges instead of n edges since $|S| \geq (\frac{1-d_1}{2})n$. If our utility function u satisfies $u'_\alpha > 0$, then we have a gain of at least $\lambda^{u'_\alpha r^2/4}$ per bad edge. In this case, assume $\lambda \geq \lambda_2 = (2^{2/(1-d_1)}3(2r^2+2r))^{4/u'_\alpha r^2}$, and let $\gamma_1 = 3^{-3(1-d_1)/8}$. Let $\Omega_{F,m}$ be defined as in Section 3.1. Combining this gives the following.

$$\pi(\Omega_S) \leq \sum_{m=\frac{(1-d_1)n}{2}}^{2n^2} \sum_{x \in \Omega_{F,m}} \pi(\phi_S(x))\lambda^{-mu'_\alpha r^2/4} \leq \sum_m n2^n 3^{\frac{1-d_1}{4}n}(2r^2+2r)^{\frac{1-d_1}{2}n}\lambda^{\frac{-(1-d_1)}{2}\frac{nu'_\alpha r^2}{4}} \leq \gamma_1^n.$$

Otherwise, if u is a threshold utility function with $\theta > (\frac{1}{2} + \frac{1}{2r+2})N$, then we have a gain of at least $\lambda^{1/4}$ per bad edge. In this case, assume $\lambda \geq \lambda_2 = (2^{2/(1-d_1)}3(2r^2+2r))^4$, and let $\gamma_1 = 3^{-3(1-d_1)/8}$.

$$\pi(\Omega_S) \leq \sum_{m=\frac{(1-d_1)n}{2}}^{2n^2} \sum_{x \in \Omega_{F,m}} \pi(\phi_S(x))\lambda^{-m/4} \leq \sum_m n2^n 3^{\frac{1-d_1}{4}n}(2r^2+2r)^{\frac{1-d_1}{2}n}\lambda^{\frac{-(1-d_1)}{2}\frac{n}{4}} \leq \gamma_1^n.$$

It remains to show that there exists a row P and a set S of r-extended contours with $|S| \geq (\frac{1-d_1}{2})n$ such that each r-extended contour in S contains at least one vertical edge along P. First consider the case where the density of B- and R-faces along any row P is low specifically, $\rho_R + \rho B <$

$\frac{1+d_1}{2}$. This implies that along this row there are at least $(1 - \frac{1+d_1}{2})n = (\frac{1-d_1}{2})n$ U-faces this implies that the maximum set S of r-extended contours which intersect P satisfies $|S| \geq (\frac{1-d_1}{2})n$ (for each U-faces either the edge above or the edge below must be included in S). Next, we can assume the density of B- and R-faces along each row is at least $\frac{1+d_1}{2}$. Let γ_R be the number of R-faces along the left and right boundaries of the grid and similarly let γ_B be the number of B-faces. Since $\gamma_R + \gamma_B \leq 2n$, either $\gamma_R < n$ or $\gamma_B < n$. We assume $\gamma_R < n$. Next, assume there is a row P with at least $(\frac{1-d_1}{2})n$ R-faces. Consider the maximum set S of r-extended contours which intersect P. This set S divides the grid into regions. Now for each R-face t along P, this face is contained within some region which implies that there is an edge of S in the same column as t or the region containing t spans the entire column. If there are no such regions that span entire columns then the size of S is at least as large as the number of R-faces along P implying, $|S| \geq (\frac{1-d_1}{2})n$ as desired. Otherwise we have a region with boundary ψ that is bordered by R-faces and spans an entire column. Since ψ spans an entire column, each row of the grid contains 2 edges of ψ. Since there are at most n R-faces along the boundary, there are at most n boundary edges contained in ψ implying ψ contains at least n non-boundary edges which implies $|S| \geq n \geq (\frac{1-d_1}{2})n$, as desired.

Finally, if there is no row P with at least $(\frac{1-d_1}{2})n$ R-faces then, since every row has at least $(\frac{1+d_1}{2})n$ B- and R-faces, there must be at least $d_1 n^2 B$-faces implying $\rho_B \geq d_1$, a contradiction. ∎

4.2 Integration at Low λ

At last we provide complementary results showing that at low λ an average sample from the steady state is integrated there will not be a high density of R or B-faces and we won't see clustering. We prove the following theorem which shows that at low bias, ghettos are unlikely to form.

Theorem 4.2: Given a valid utility function u with radius r and constant $c_2 > 10/11$, there exist constants $\gamma_2 = \gamma_2(c_2) < 1$ and $\lambda_2 = \lambda_2(u'_\beta, r, c_2)$ such that for $\lambda \leq \lambda_2$ a random sample from Ω will be c_2-clustered or the density ρ_R or $\rho_B > c_2$ with probability at most γ_2^n.

PROOF: First we show that a configuration is exponentially unlikely to be c_2-clustered. We use a similar technique to show that $\rho_R, \rho_B < c_2$ and so defer this part of the proof to Appendix A.3. Let $\Omega_C \subset \Omega$ be the set of configurations that are c_2-clustered. We will show that under the conditions stated in the theorem, $\pi(\Omega_C)$ is exponentially small. Throughout this proof we will assume that the R-faces are c_2-clustered. To show Ω_C is exponentially small, we construct a map $\phi_C : \Omega_C \to \Omega$, which maps a configuration $\sigma \in \Omega_C$ to the set of all configurations which corresponding to removing a c_2-cluster region C and then selecting $(1-c_2)n^2$ B-faces or U-faces and changing them to R-faces. Given $\sigma \in \Omega_C$, define $N(\sigma)$ to be the set of all configurations obtained from σ by removing a c_2-cluster region C and changing exactly $(1-c_2)n^2$ B-faces or U-faces to R-faces. To remove C, we change (or flip) all R-faces to B-faces within C. Once we flip the R-faces and B-faces in C there are at most $(1-c_2)n^2$ R-faces remaining so $|N(\sigma)| \geq \binom{c_2 n^2}{(1-c_2)n^2}$. For each configuration $\tau \in \Omega$ we bound the number of configurations σ such that $\tau \in N(\sigma)$. If there exists σ such that $\tau \in N(\sigma)$, then the number of R-faces in τ is at most $2(1-c_2)n^2$ Since C is a c_2-cluster region with perimeter at most $c_2 n$, there are at most $2n^2 3^{c_2 n} 3^{2(1-c_2)n^2}$ possible pre-images of any configuration τ. The factor of 2 is because the configuration could have been R or B-clustered. Next, given configurations σ, τ such that $\tau \in N(\sigma)$ we derive an upper bound on the ratio $\pi(\sigma)/\pi(\tau)$. Recall the map ϕ_C first removes a c_2-cluster region C by flipping the $R-$ and $B-$ faces within C. This procedure only changes the "happiness" of faces within distance r of the border. Since there are at most $(2r^2 + 2r + 1)c_2 n$ of these, removing C decreases the weight by at most a factor of $\lambda^{c_2 n(2r^2+2r+1)}$. Changing the color of a single face can decrease the weight of a configuration by at most a factor of $\lambda^{2u'_\beta(2r^2+2r)}$. Thus, changing $(1-c_2)n^2$ B-faces or U-faces to R-faces decreases the weight by at most a factor of $\lambda^{2u'_\beta(1-c_2)n^2(2r^2+2r)}$. Combining these shows that $\pi(\sigma)/\pi(\tau) \leq \lambda^\Delta$, where $\Delta = c_2 n(2r^2+2r+1) + 2u'_\beta(1-c_2)n^2(2r^2+2r)$.

We define a weighted bipartite graph $G(\Omega_D, \Omega, E)$ with an edge weight $\pi(\sigma)$ between $\sigma \in \Omega_D$ and $\tau \in \Omega$ if $\tau \in N(\sigma)$. The total weight of edges is

$$\sum_{\sigma\in\Omega_D} \pi(\sigma)|N(\sigma)| \geq \sum_{\sigma\in\Omega_D} \pi(\sigma)\binom{c_2 n^2}{(1-c_2)n^2} \geq \pi(\Omega_D)\left(\frac{c_2}{(1-c_2)}\right)^{(1-c_2)n^2}.$$

Also, the weight of edges is at most

$$\sum_{\tau\in\Omega} \pi(\tau) 2n^2 3^{c_2 n} 3^{2(1-c_2)n^2} \lambda^\Delta \leq 2n^2 3^{c_2 n} 3^{2(1-c_2)n^2} \lambda^{2(1-\mu)\Delta}.$$

Combining these equations, assuming $\lambda \leq \lambda_2 = \left(\frac{c_2}{10(1-c_2)}\right)^{(4u'_\beta(r^2+r))^{-1}}$, and letting $\gamma_2 = (10/11)^{1-c_2}$ gives

$$\pi(\Omega_D) \leq \left(\frac{1-c_2}{c_2}\right)^{(1-c_2)n^2} 2n^2 3^{c_2 n} 3^{2(1-c_2)n^2} \lambda^{2(1-\mu)\Delta} \leq \gamma_2^n.$$

References

[1] Sylvain Barde. Back to the future: A simple solution to schelling segregation. 2011.

[2] M. Biskup and L. Chayes. Rigorous analysis of discontinuous phase transitions via mean-field bounds. *Communications in Mathematical Physics*, 2003.

[3] M. Biskup, L. Chayes, and N. Crawford. Mean-field driven first-order phase transitions in systems with long-range interactions. *Journal of Statistical Physics*, 2004.

[4] C. Brandt, N. Immorlica, G. Kamath, and R. Kleinberg. An analysis of one-dimensional schelling segregation. In *Proceedings of the 44th symposium on Theory of Computing*, STOC '12, 2012.

[5] R. Bubley and M. Dyer. Faster random generation of linear extensions. In *Proceedings of the ninth annual ACM-SIAM symposium on Discrete algorithms*, SODA '98, 1998.

[6] L. Chayes. Mean field analysis of low-dimensional systems. *Communications in Mathematical Physics*, 2009.

[7] S. Grauwin. Effect of local coordination on a schelling-type segregation model. *CoRR*, 2008.

[8] S. Grauwin, F. Goffette-Nagot, and P. Jensen. Dynamic models of residential segregation: Brief review, analytical resolution and study of the introduction of coordination. *CoRR*, 2008.

[9] S. Greenberg, A. Pascoe, and D. Randall. Sampling biased lattice configurations using exponential metrics. In *Proceedings of the twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '09, 2009.

[10] S. Greenberg and D. Randall. Convergence rates of markov chains for some self-assembly and non-saturated ising models. *Theoretical Computer Science*, 2009.

[11] M. Jerrum and A. Sinclair. Approximate counting, uniform generation and rapidly mixing markov chains. In *Information and Computation*, 1989.

[12] M. Jerrum and A. Sinclair. Poly-time approximation algorithms for the ising model. In *Society for Industrial and Applied Mathematics*, 1993.

[13] E. Lewis-Pye, G. Barmpalias, and R. Elwes. Dynamic models of residential segregation: Brief review, analytical resolution and study of the introduction of coordination. *CoRR*, 2008.

[14] S. Lubetzy and A. Sly. Critical ising on the square lattice mixes in polynomial time. *CoRR*, 2012.

[15] S. Miracle, D. Randall, and A. Streib. Clustering in interfering binary mixtures. In *Proceedings of the 14th international workshop and 15th international conference on Approximation, randomization, and combinatorial optimization: algorithms and techniques*, APPROX'11/RANDOM'11, pages 652–663, 2011.

[16] E. Presutti. *Scaling Limits in Statistical Mechanics and Microstructures in Continuum Mechanics.* Theoretical and mathematical physics. Springer Berlin Heidelberg.

[17] D. Randall and D. Wilson. Sampling spin configurations of an ising system. In *Proceedings of the tenth annual ACM-SIAM symposium on Discrete algorithms*, SODA '99, 1999.

[18] Dana Randall. Slow mixing of glauber dynamics via topological obstructions. In *Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm*, SODA '06, 2006.

[19] T. Rogers and A. McKane. A unified framework for schelling's model of segregation. *CoRR*, 43, 2011.

[20] T. Schelling. Dynamic models of segregation. *Journal of Mathematical Sociology*, 1971.

[21] A. Sinclair. *Algorithms for Random Generation & Counting: A Markov Chain Approach.* Birkhäuser, 1993.

[22] D. Stauffer. Social applications of two-dimensional ising models. *CoRR*, 2007.

[23] D. Stauffer and S. Solomon. Ising, schelling, and self-organising segregation. *CoRR*, 2007.

[24] R. Swendsen and J. Wang. Nonuniversal critical dynamics in monte carlo simulations. *Phys. Rev. Lett.*, 58:86–88, 1987.

[25] D. Vinkovik. Segregation: Society as a liquid or a solid. *Proceeedings of the National Academy of Sciences of the United States of America*, 2006.

[26] D. Vinkovik and A. Kirkman. A physical analogue of the schelling model. *Proceeedings of the National Academy of Sciences of the United States of America*, 2006.

[27] U. Wolff. Collective monte carlo updating for spin systems. *Phys. Rev. Lett.*, 62:361–364, 1989.

[28] H.P. Young. *Individual strategy and social structure: an evolutionary theory of institutions.* Princeton paperbacks. 2001.

[29] J. Zhang. A dynamic model of residential segregation. *Journal of Mathematical Sociology*, 2004.

A Appendix: Additional Details of Proofs

In this appendix we provide additional details missing from the proofs in Section 3 and Section 4.

A.1 Rapid Mixing at Low λ: Proof of Theorem 3.3

Here we provide the missing details from Theorem 3.3 which says that for sufficiently low λ we will have rapid mixing in our model. We restate Theorem 3.3 and the outline of the proof for clarity.

We use a path coupling argument with the natural coupling. A move of $\mathcal{M}$ consists of selecting a face f and a color c. The natural coupling simply uses the same face and color for both configurations. The distance metric we use is the minimal number of steps of $\mathcal{M}$ required to change one configuration into another. Given a pair of configurations, each face occupied in exactly one configuration contributes 1 to the distance while each face colored R in one configuration and B in the other contributes 2. Thus, the maximum distance between any two configurations is $2n^2$.

Note that: We present the results in the unsaturated setting where we allow empty houses. For the saturated model the Markov chain allows houses to move between B and R in one move, indicating that a new resident will move in as soon as one vacates a house. All of the proofs carry over in this case and are in fact simpler.

Theorem 3.3: For the Markov Chain $\mathcal{M}$, with radius r and utility function u, there exists a constant $\lambda_3 = \lambda_3(r, u'_\alpha)$ such that $\mathcal{M}$ is fast mixing when $1 \leq \lambda < \lambda_3$.

Proof: In order to apply the path coupling theorem, we consider the set of pairs of configurations at distance 1 and show that for any such pair, the distance between the configurations decreases in expectation after one step of $\mathcal{M}$. In order to bound the expected change in distance, we consider two configurations that differ at one site, without loss of generality $(\sigma_{g=R}, \sigma_{g=U})$. For ease of notation, in this section we will sometimes refer to $\sigma_{g=U}$ as σ, and write $s(x) = s(\sigma, x)$.

Let f be the face selected by $\mathcal{M}$. The distance only decreases if $f = g$; here we consider three cases.

- If $f = g$ and $c = R$, then we accept both moves with probability 1, thus decreasing the distance by 1.
- If $f = g$ and $c = B$, then configuration $\sigma_{g=U}$ will accept the transition with probability 1, while the move is disallowed for $\sigma_{g=R}$; thus we will increase the distance by 1.
- If $f = g$ and $c = U$, then we decrease the distance with the probability that $\sigma_{g=R}$ transitions,

$$\frac{\pi(\sigma_{g=U})}{\pi(\sigma_{g=R})} = \frac{1}{\lambda^{u(s(g),o(g))}} \prod_{\substack{y:\sigma(y)=R,\\ d(g,y)\leq r}} \frac{\lambda^{u(s(y),o(y))}}{\lambda^{u(s(y)+1,o(y)+1)}} \prod_{\substack{y:\sigma(y)=B,\\ d(g,y)\leq r}} \frac{\lambda^{u(s(y),o(y))}}{\lambda^{u(s(y),o(y)+1)}}$$
$$\geq \frac{1}{\lambda^{u(s(g),o(g))}} \frac{1}{\lambda^{u'_\gamma s(g) + u'_\beta (o(g)-s(g))}}$$

The distance can increase whenever $f \neq g$. We again consider three cases:

- If $f = U$, both transitions are accepted with probability 1 and the distance does not change.

- If $f = R$, the probability that we increase the distance by 1 is the difference in the chance that $\sigma_{g=U}$ becomes U at f but $\sigma_{g=R}$ does not. This quantity is

$$
\begin{aligned}
& |\frac{\sigma_{f=0,g=0}}{\sigma_{f=R,g=0}} - \frac{\sigma_{f=0,g=R}}{\sigma_{f=R,g=R}}| \\
& = |\frac{1}{\lambda^{u(s(f),g(f))}} \prod_{\substack{y:\sigma(y)=A \\ d(y,f)\le r}} \frac{\lambda^{u(s(y)-1,o(y)-1)}}{\lambda^{u(s(y),o(y))}} \prod_{\substack{y:\sigma(y)=B, \\ d(f,y)\le r}} \frac{\lambda^{u(s(y),o(y)-1)}}{\lambda^{u(s(y),o(y))}} - \\
& \frac{1}{\lambda^{u(s(f)+1,g(f)+1)}} \prod_{\substack{y:\sigma(y)=A \\ d(y,f)\le r}} \frac{\lambda^{u(s(y),o(y))}}{\lambda^{u(s(y)+1,o(y)+1)}} \prod_{\substack{y:\sigma(y)=A \\ d(y,f)\le r, d(y,g)>r}} \frac{\lambda^{u(s(y)-1,o(y)-1)}}{\lambda^{u(s(y),o(y))}} \\
& \cdot \prod_{\substack{y:\sigma(y)=B, \\ d(f,y)\le r, d(g,y)\le r}} \frac{\lambda^{u(s(y),o(y)-1)}}{\lambda^{u(s(y),o(y))}} \prod_{\substack{y:\sigma(y)=B, \\ d(f,y)\le r, d(g,y)>r}} \frac{\lambda^{u(s(y),o(y))}}{\lambda^{u(s(y),o(y)+1)}}| \\
& \le \frac{1}{\lambda^{u(s(f),g(f))}} \prod_{\substack{y:\sigma(y)=A \\ d(y,f)\le r, d(y,g)>r}} \frac{\lambda^{u(s(y)-1,o(y)-1)}}{\lambda^{u(s(y),o(y))}} \prod_{\substack{y:\sigma(y)=B, \\ d(f,y)\le r, d(g,y)>r}} \frac{\lambda^{u(s(y),o(y))}}{\lambda^{u(s(y),o(y)+1)}} \\
& \cdot |\frac{1}{\lambda^{u'_\kappa s(g)} \lambda^{u'_\alpha (o(g)-s(g))}} - \frac{1}{\lambda^{u'_\gamma s(g)} \lambda^{u'_\beta (o(g)-s(g))}}| \\
& \le |\frac{1}{\lambda^{u'_\kappa s(g)} \lambda^{u'_\alpha (o(g)-s(g))}} - \frac{1}{\lambda^{u'_\gamma}} \frac{1}{\lambda^{u'_\gamma s(g)} \lambda^{u'_\beta (o(g)-s(g))}}| \\
& \le 1 - \frac{1}{\lambda^{u'_\gamma}} \frac{1}{\lambda^{(u'_\gamma - u'_\kappa) s(g)} \lambda^{(u'_\beta - u'_\alpha)(o(g)-s(g))}}
\end{aligned}
$$

- Similarly, if $f = B$, this probability is

$$\le 1 - \frac{1}{\lambda^{u'_\beta}} \frac{1}{\lambda^{(u'_\beta - u'_\alpha) s(g)} \lambda^{(u'_\gamma - u'_\kappa)(o(g)-s(g))}}$$

Let $\eta = \max u'_\gamma - u'_\kappa, u'_\beta - u'_\alpha$. The expected change in distance is then

$$
\begin{aligned}
\mathrm{E}[\Delta(\sigma_1,\sigma_2)] \le & \frac{1}{3n^2}\Big(\frac{1}{\lambda^{u(s(g),o(g))}} \frac{1}{\lambda^{u'_\gamma s(g) + u'_\beta (o(g)-s(g))}} + (s(g))(1 - \frac{1}{\lambda^{u'_\gamma}} \frac{1}{\lambda^{(u'_\gamma - u'_\kappa)s(g)} \lambda^{(u'_\beta - u'_\alpha)(o(g)-s(g))}}) \\
& + \quad (o(g) - s(g))(1 - \frac{1}{\lambda^{u'_\beta}} \frac{1}{\lambda^{(u'_\beta - u'_\alpha)s(g)} \lambda^{(u'_\gamma - u'_\kappa)(o(g)-s(g))}})\Big) \\
\le & \frac{1}{3n^2}\Big(\frac{1}{\lambda^{2u'_\gamma s(g) + 2u'_\beta (o(g)-s(g))}} + (s(g))(1 - \frac{1}{\lambda^{u'_\gamma}} \frac{1}{\lambda^{\eta N}}) + (o(g) - s(g))(1 - \frac{1}{\lambda^{u'_\beta}} \frac{1}{\lambda^{\eta N}})\Big) \\
\le & \frac{-1}{3n^2}\left(\frac{1}{\lambda^{2u'_\gamma s(g) + 2u'_\beta (o(g)-s(g))}} - \log(\lambda^{(\eta N (u'_\gamma + u'_\beta))(u'_\gamma s(g) + u'_\beta (o(g)-s(g)))})\right).
\end{aligned}
$$

The final step uses the fact that $\lim_{n\to\infty} n(1 - x^{1/n}) \to -\log x$ from below. Thus we see our expected change is negative whenever the value $v = \lambda^{\eta N (u'_\gamma + u'_\beta)}$ satisfies $1/v^2 > \log v$. Thus, we simply need $\eta/\lambda^{2r^1 - 1} > (2r^2 - 1)\log\lambda$. This occurs for $1 \le \lambda \le (1.8)^{\eta/(2r^1-2)} = 1 + O(1/r^2) + O(\eta)$. Setting $\lambda = (1.5)^{\eta/(2r^1-2)}$, the expected change in distance is $\le -.2612/3n^2$ per step. Applying the path coupling theorem [5] gives the result. ∎

A.2 Segregation at High λ: Proof of Theorem 4.1

Here we provide the missing details from Theorem 4.1 which says that at high λ the stationary distribution will be segregated. We restate Theorem 4.1 for clarity.

Theorem 4.1: Assume a valid utility function u with radius r such that $u'_\alpha > 0$ or u is a threshold utility function with $\theta > (\frac{1}{2} + \frac{1}{2r+2})N$, where $N = 2r^2 + 2r$. Given a constant density $d_1 > 1/2$, there exist constants $\gamma_1 = \gamma_1(d_1) < 1$ and $\lambda_1 = \lambda_1(u'_\alpha, r, d_1)$ such that for all $\lambda \geq \lambda_1$ a random sample from Ω will have no contours with more than $d_1 n$ edges and either the density $\rho_R > d_1$ or $\rho_B > d_1$ with probability at least $(1 - \gamma_1^n)$.

PROOF: In Section 4.1 we proved that at high λ it's exponentially unlikely to have the density of both R-faces and B-faces less than d_1. Here we prove that it's exponentially unlikely for a configuration to have any contour with size greater than $d_1 n$. Using the union bound, it is straightforward to combine these two results.

Let Ω_{d_1} be the set of configuration in Ω which contain a contour more than $d_1 n$ edges. As in Section 3.1, we construct a map $\phi_{d_1} : \Omega_{d_1} \to \Omega$ which maps configurations which have a contour of size greater than $d_1 n$ to configurations which have at least one less contour of size greater than $d_1 n$. Specifically, ϕ_{d_1} takes the lexicographically first contour of size greater than $d_1 n$, finds the r-extended contour which contains this component, changes all U-faces boarding the r-extended contour to R-faces and finally in all B-bordered regions adjacent to the contour exchanges R-faces with B-faces (i.e replaces all R-faces with B-faces and all B-faces with R-faces).

Next, we bound the number of pre-images of a configuration under ϕ_{d_1}, using a combinatorial argument similar to Section 3.1. In Section 3.1 the number of configurations with an r-extended contour with m edges which intersect the border is at most $4n3^{m/2}(2r^2 + 2r)^m$. However, with our new function ϕ_{d_1}, the contour might not be connected to the border so the number of configurations with an r-extended contour with m edges is now $2n^2 3^{m/2}(2r^2 + 2r)^m$, since the number of possible starting points is increased from $4n$ to $2n^2$. Additionally, we only guarantee that the r-extended contour has at least $d_1 n$ edges instead of n edges. Let $\Omega_{F,m}$ be defined as before. The remainder of the proof is the same as in Theorem 3.1. If our utility function u satisfies $u'_\alpha > 0$, then we have a gain of at least $\lambda^{u'_\alpha r^2/4}$ per edge of the r-extended contour. Assume $\lambda \geq \lambda_1 = (3(2r^2 + 2r))^{4/u'_\alpha r^2}$, and let $\gamma_1 = 3^{-3d_1 1/4}$. We have the following.

$$
\begin{aligned}
\pi(\Omega_{d_1}) &\leq \sum_{m=d_1 n}^{2n^2} \pi(\Omega_{F,m}) \\
&\leq \sum_m \sum_{x \in \Omega_{F,m}} \pi(\phi_{d_1}(x)) \lambda^{-m u'_\alpha r^2/4} \\
&\leq \sum_m 2n^2 3^{d_1 n/2} (2r^2 + 2r)^{d_1 n} (\lambda)^{-d_1 n u'_\alpha r^2/4} \\
&\leq 4n^4 3^{-d_1 n/2} \leq \gamma_1^n.
\end{aligned}
$$

Otherwise, if u is a threshold utility function with $\theta > (\frac{1}{2} + \frac{1}{2r+12})N$, then we have a gain of at least $\lambda^{1/4}$ per edge of the r-extended contour. Assume $\lambda \geq \lambda_1 = (3(2r^2 + 2r))^4$, and let

$\gamma_1 = 3^{-3d_1 1/4}$. We have the following.

$$\begin{aligned}
\pi(\Omega_{d_1}) &\leq \sum_{m=d_1 n}^{2n^2} \pi(\Omega_{F,m}) \\
&\leq \sum_m \sum_{x \in \Omega_{F,m}} \pi(\phi_{d_1}(x))\lambda^{-m/4} \\
&\leq \sum_m 2n^2 3^{d_1 n/2}(2r^2+2r)^{d_1 n}(\lambda)^{-d_1 n/4} \\
&\leq 4n^4 3^{-d_1 n/2} \leq \gamma_1^n.
\end{aligned}$$

■

A.3 Integration at Low λ: Proof of Theorem 4.2

Here we provide the missing details from Theorem 4.2 which says that at low bias the stationary distribution will be well-integrated. We restate Theorem 4.2 for clarity.

Theorem 4.2: Given a valid utility function u with radius r and constant $c_2 > 10/11$, there exist constants $\gamma_2 = \gamma_2(c_2) < 1$ and $\lambda_2 = \lambda_2(u'_\beta, r, c_2)$ such that for $\lambda \leq \lambda_2$ a random sample from Ω will be c_2-clustered or the density ρ_R or $\rho_B > c_2$ with probability at most γ_2^n.

PROOF: In Section 4.2 we showed that at low λ, Ω will not be c_2-clustered. Here we show that at low λ we will have $\rho_R, \rho_B < d_2$. It is straightforward to combine the two results using a union bound. Let Ω_D be the set of configuration in Ω for which $\rho_R \geq d_2$ or $\rho_B \geq d_2$. We will show that under the conditions stated in the theorem, $\pi(\Omega_D)$ is exponentially small. Throughout this proof we will assume that $\rho_R \geq d_2$. To show this we will construct a map $\phi_D : \Omega_D \to \Omega$, which maps a configuration σ to the set of all configurations which correspond to selecting $(1-d_2)n^2$ R-faces and changing them to B-faces. Define $N(\sigma)$ to be the set of all configurations obtained from σ by changing exactly $(1-d_2)n^2$ R-faces to B-faces. Since there are least $d_2 n^2$ R-faces, $|N(\sigma)| \geq \binom{d_2 n^2}{(1-d_2)n^2}$. For each configuration $\tau \in \Omega$ we need to bound the number of configuration σ such that $\tau \in N(\sigma)$. If there exists a σ such that $\tau \in N(\sigma)$ then this implies that the number of B-faces in σ is at most $2(1-d_2)n^2$ and since our map only changes R-faces to B-faces, there are at most $2^{2(1-d_2)n^2+1}$ possible pre images for σ. The additional factor of 2 is due to the fact that originally either $\rho_R \geq d_2$ or $\rho_B \geq d_2$. We define a weighted bipartite graph $G(\Omega_D, \Omega, E)$ with an edge weight $\pi(\sigma)$ between $\sigma \in \Omega_D$ and $\tau \in \Omega$ if $\tau \in N(\sigma)$. The total weight of edges is

$$\sum_{\sigma \in \Omega_D} \pi(\sigma)|N(\sigma)| \geq \sum_{\sigma \in \Omega_D} \pi(\sigma)\binom{d_2 n^2}{(1-d_2)n^2} \geq \pi(\Omega_D)\left(\frac{d_2}{1-d_2}\right)^{(1-d_2)n^2}.$$

However the weight of the edges is at most

$$\sum_{\tau \in \Omega} \pi(\tau) 2^{2(1-d_2)n^2+1}(\lambda^{2(1-\mu)(2r^2-1)})^{(1-d_2)n^2} \leq 2^{2(1-d_2)n^2+1}(\lambda^{2(1-\mu)(2r^2-1)})^{(1-d_2)n^2}.$$

Combining these equations, assuming $\lambda^{1-\mu} \leq \lambda_2 = \left(\frac{d_2}{5(1-d_2)}\right)^{1/(4r^2-2)}$, and letting $\gamma_2 = (5/6)^{1-d_2}$ gives the following

$$\pi(\Omega_D) \leq \left(\frac{1-d_2}{d_2}\right)^{(1-d_2)n^2} 2^{2(1-d_2)n^2+1}(\lambda^{2(1-\mu)(2r^2-1)})^{(1-d_2)n^2} \leq \gamma_2^n.$$