An Information Theoretic Approach to Analyzing the Shelling Model

Joshua Parker
Department of Mathematics, UC Davis
jdparker@math.ucdavis.edu

Abstract:
Information theoretic quantities, such as the entropy density and excess entropy, were measured for the 2-D lattice of the Schelling Model. These measurements were then used to describe and analyze the qualitative results previously found in another paper. Methods for finding these quantities in any 2-D model is discussed.

Introduction

In the late 1960's, economist Thomas Schelling developed a model for segregation in a community. Of course, it is no surprise that if a community can be separated into two different categories, such as black people and white people, where each person has a strong preference to live near those who are of the same race, then segregation will occur. However, the original Schelling Model looked at a community in which each person prefers to only have half or more of their neighbors be of their own type and found that such a community will still completely segregate. This was considered a surprising result, given that the people only have a mild preference for like-neighbors. This result doesn't just apply to people of the same race, but can also be applied to people wanting to be near others with similar ideas, habits, preferences, etc. In fact, the model doesn't have to be restricted to just people. It was shown by [1] that the Schelling Model has a physical analogue to any system in which there are two types of particles, and the particles have a tendency to cluster with like-particles. In [1] several qualitative results and conclusions are made from running simulations of the Schelling Model. The goal of this paper is to make these results and conclusions more concrete using some tools of information theory. In class, we primarily discussed the use of information theory in one dimension, but the dynamics of the Schelling model takes place on a two dimensional lattice, where each lattice cell can represent a person in a community or a particle in a physical system or whatever the system being studied requires. The procedure for finding information theoretic quantities, like the entropy density or excess entropy, of a lattice with two spatial dimensions is not as straightforward as in one dimension, and can vary depending on the system being analyzed. The methods used here to study the Schelling Model will closely follow the ones used by Feldman and Crutchfield in [2] and are briefly discussed below.

Background

Here I will review the definitions of entropy density and excess entropy in one dimension, and then introduce their corresponding definitions in two dimensions. First, recall that the block entropy density for a random variable S_L conditioned on a sequence of L random variables is given by:

$$h_\mu(L) = H[S_L | S_{L-1} S_{L-2} \cdots S_1]$$

(1)

and so the entropy density is just this quantity passed to the limit:

(2) $$h_\mu = \lim_{L \to \infty} h_\mu(L).$$

The excess entropy is defined as:

(3) $$E = \sum_{L=1}^{\infty} [h_\mu(L) - h_\mu]$$

Now, to extend these quantities to a lattice in two dimensions we look at a template, at the site X_0 as seen in Figure 1 below:

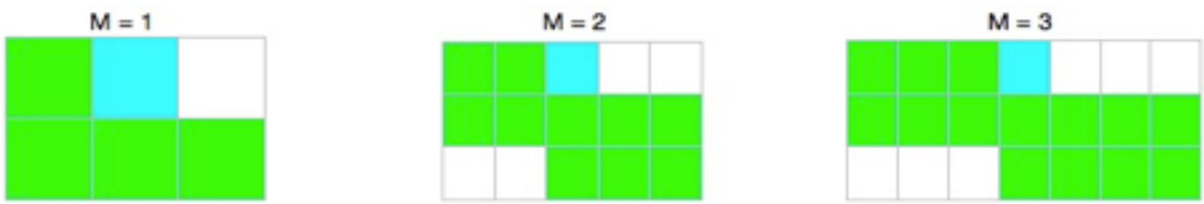

Figure 1: Templates for M = 1, M = 2, M = 3. The blue cell represents and the green cells are the cells being conditioned on.

The templates given here are designed specifically for next-nearest-neighbor interactions, where there are always 2 rows (except when M = 0 or M = 1) and 2*M + 1 columns. Also, we define the template to be the single cell, X_0.

With this, we can now define the entropy density and excess entropy density in 2-D:

(4) $$h_\mu = \lim_{M \to \infty} h_\mu(M).$$

(5) $$E = \sum_{M=0}^{\infty} [h_\mu(M) - h_\mu]$$

Where we have,

$$h_\mu(M) = H[X_0 | X_{-1} X_{-2} \cdots] = H[BlueCell | GreenCells]$$

(6)

Model Description

The model is fairly simple: We start out with a square lattice of arbitrary size, where each cell is one of three values: red, blue, or black. The red and blue cells are 'agents', while the black cells are 'empty'. The system operates in discrete time where after each iteration we get a new grid (same size), but with the same number of empty cells and same number of red and blue agents. In other words, the number of red agents, blue agents, and empty cells are preserved throughout the process. Therefore, it is more intuitive to visualize the cells moving around the grid on each iteration. The dynamics of the system is governed by a set of rules, which are best described as follows. First, the red and blue agents are assigned a threshold value of 'satisfaction'. If the percentage of cells in their Moore neighborhood that are of the same color is above or equal to this threshold then the agent is satisfied and stays the same color for that iteration. However, if the percentage of like-neighbors is less than this threshold, then this agent 'moves' to an empty cell that will make it satisfied. So, on each iteration, the model finds all the unsatisfied agents, and moves them to new cells. This, of course, will make other agents unsatisfied, and so the process continues until it reaches an invariant configuration. An invariant configuration is one where either all the agents are satisfied or there are no empty cells that will make any of the unsatisfied agents satisfied.

There are several ways the dynamics of this model can be changed. For one, we can make the boundaries periodic. However, for all the results in this paper, the boundaries were kept non-periodic. Also, the rules of the model can be changed. In the original model, the agents can move anywhere on the lattice, but one could make it so that agents move to the closest empty cell that will make them satisfied, which would correspond to a slower diffusion rate in the physical analogue. Additionally, one could make it so that satisfied agents are allowed to move as long as they don't become unsatisfied, that way the system is always in flux and never reaches a fixed point. This rule change, serves as a model for the mixture of liquids in the physical analogue, where our original model serves as model for a mixture forming into a solid.

To conclude our description of the Schelling model, we list the most important parameters that are involved with this system:

1. X: Threshold value, percentage of like-neighbors needed to make agents satisfied
2. EP: Percentage of empty cells.
3. R: Percentage of red cells
4. B: Percentage of blue cells

In [2], there is a qualitative analysis of how the Schelling Model depends on the values X and EP . The steady states are presented for threshold values of X = 37.5, 50.0, 62.5 and for empty cells percentages EP = 2, 10, 30 with 100 x 100 latices (see Figure 2). It is easy to observe that in general, the steady states become more clustered as X gets larger or EP gets smaller, except for the case where the combination of high threshold and low number of empty cells makes it so there are unsatisfied agents that have no empty cell to go to, and so the system isn't able to evolve properly and becomes fixed in an unsatisfied state. However, as more empty cells are added with this threshold kept fixed, we see a phase transition-like phenomenon where the steady state suddenly becomes highly clustered like in Fig. 2 with X = 62.5, EP = 10. Notice that in this case, the empty cells act like a boundary layer between the clusters.

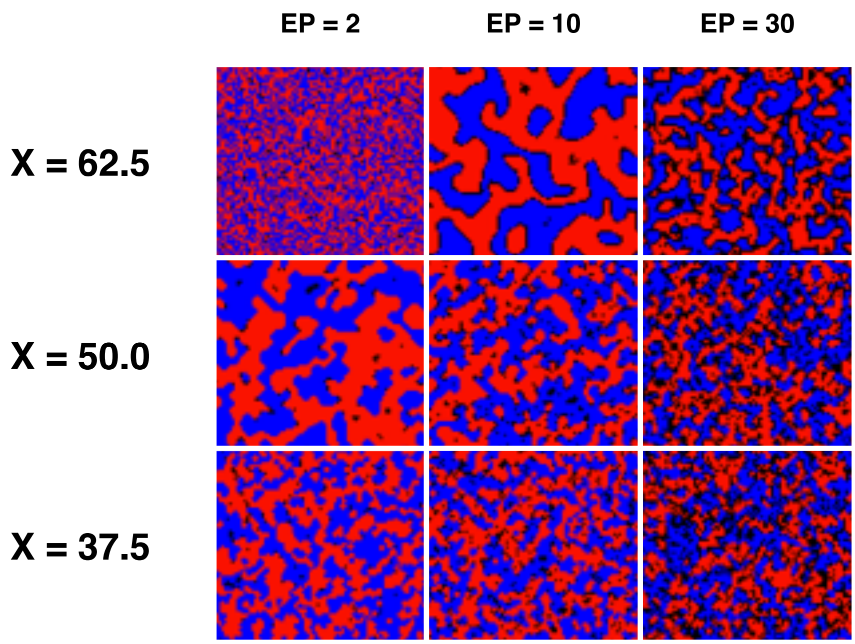

Figure 2:
Replication of figure in physical analogue paper [2] which shows qualitatively how clustering structure varies with threshold value and percentage of empty cells in the system.

Again, these results are purely qualitative. To see if these results could be described more quantitatively, I found the entropy densities and excess entropies for these steady states with the hypothesis being that a lower entropy density and higher excess entropy will correspond to more clustering.

Methods

To find the block entropy density one has to find the probability distribution for all the possible configurations of the template. So, for example, if M = 3, then the template consists of 15 cells and so there are 3^8 possible configurations, since there are three possible values[1] for each cell: red, blue, and empty. This results in a very large event space making it difficult to get enough data samples to cover the entire space, which results in a biased estimation of the entropy density. To limit the bias, I used the following entropy estimator[3]:

(7) $$\hat{H}_2 = \sum_{i=1}^{M} \frac{k_i}{N}\left(\psi(N) - \psi(k_i) + \log 2 + \sum_{j=1}^{k_i-1} \frac{(-1)^j}{j}\right)$$

Where k_i is the number of times the ith configuration occurred for N data samples, and ψ is the digamma function.

This equation has an upper bound bias given by:

(8) $$B \leq \frac{M+1}{N}$$

So, if we rewrite Eq. 6 as,

(9) $$H[BlueCell|GreenCells] = H[BlueCell] - H[GreenCells]$$

Then we can directly apply Eq. 7 to find the block entropy density for a given M after using a computer program to count the frequency of each template configuration for N data samples. To find the excess entropy, all one has to do is find h_u (M) for $M < M_c$, where M_c is the value for which the entropy density converges, and then apply Eq. 5.

Results

Unfortunately, the execution time for my simulation of the Schelling Model made it difficult to get good statistics, so I started off by estimating the entropy density with a *nearest-neighbor* interaction template at M = 3, where the template only has a thickness of one row instead of two. This resulted in an overestimate of the true entropy density, but the value did converge with respect to M. I calculated this entropy density for various values of X and EP. Figures 3a and 3c show how entropy density changes as more empty cells are added to the system. As expected, when empty cells are added, we get a larger value for h_u which corresponds to less clustering,

---

[1] Pedantic note: The choice to label each cell 'red', 'blue', or 'empty' is an arbitrary one. A more natural labeling might use the alphabet {-1,0,1} for example.

wh sity is a m
sys

have simi
thr ngth of 1/
reg Moore nei
onl

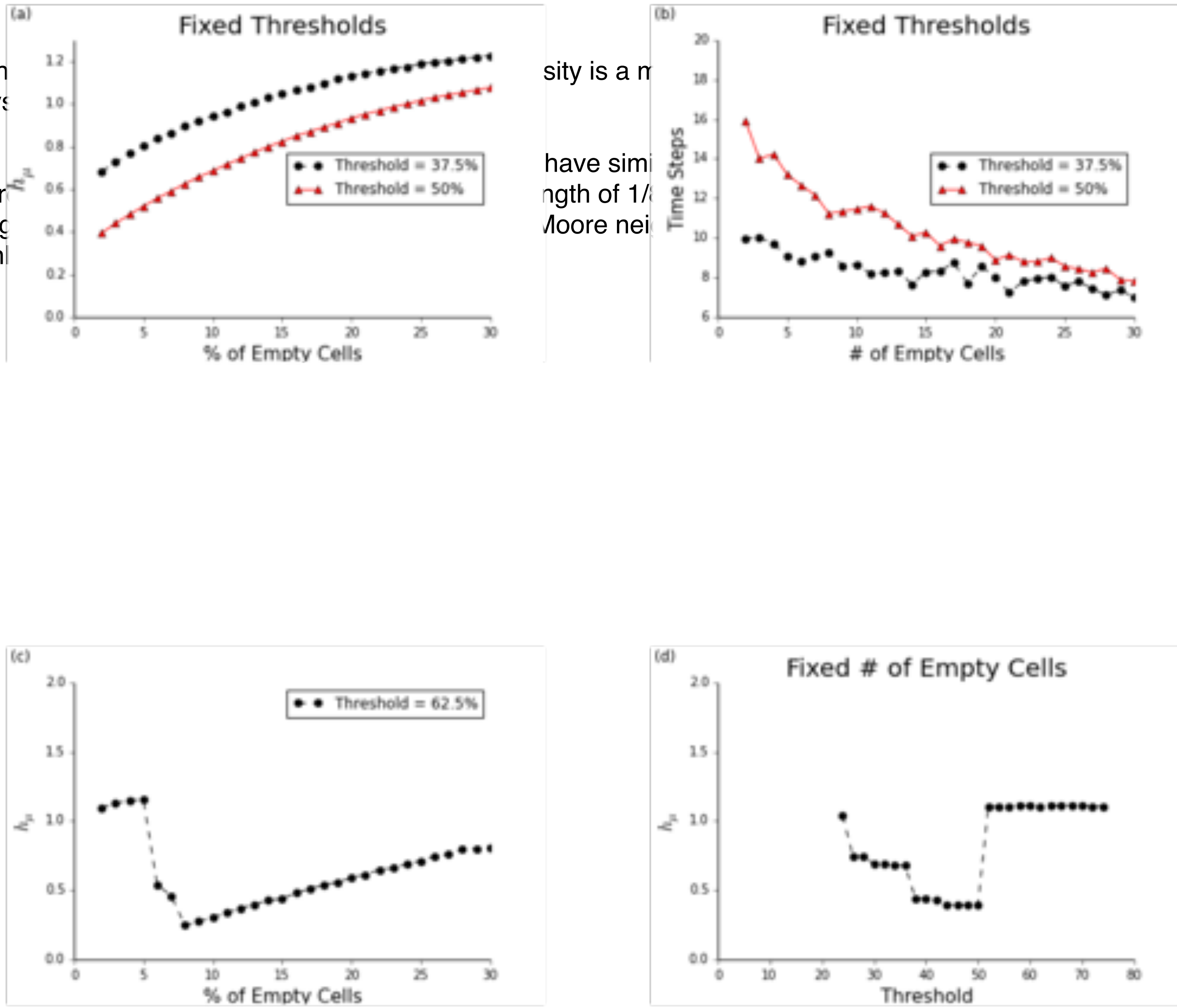

Figure3:
(a) and (c) plot values of h_u for different percentages of empty cells while keeping the threshold value fixed. (b) plots the number of time steps needed for equilibrium to be reached as a function of percentage of empty cells. (d) Plots h_u as a function of threshold value while keeping percentage of empty cells fixed at 2. The values of R and B are always such that R=B.

Finally, the entropy rate was measured using the next-nearest-neighbor templates and are shown in Table 1. We see that the previous measurements were indeed overestimates, but we see the same trends we were seeing before. In Table 2 are the corresponding excess entropies. The results here are not as easy to interpret. It should first be noted that the values most susceptible to bias are those with X = 37.5 or EP = 30, especially the latter because the event space of template configurations is much greater for this case. Nevertheless, the results for the most part are promising. As expected, we tend to see the excess entropy go up as the threshold value increases. This is because the clusters are getting bigger. Another way to look at it is like this: a larger amount of excess entropy corresponds to more information stored in the system. If the threshold value is higher then this places more restrictions on the system, so we get a larger predictability gain than by changing the number of empty cells. The excess entropy clearly detects structure that the entropy rate doesn't. For example look at the entropy densities for X = 62.5, EP = 2 and X = 37.5, EP = 30; they are nearly the same. However, looking at Fig. 2 it is clear that the former is much more random than the latter. The reason the entropy rates are the same is because even though one is more clustered, it is also the one with 30% empty cells so the three states of the system are distributed more evenly, while the other system only has 2% empty cells; it is practically a 2-state system and so it is more predictable from that standpoint. Looking at the excess entropy, however, correctly reveals that the one with more empty cells is actually more clustered. I should probably mention that this is the stationary state most susceptible to bias, and would therefore have an overestimated excess entropy, which can be a little worrisome, however the difference between the two values is so large that for this specific case we don't need to worry about it.

	EP = 2	EP = 10	EP = 30
X = 62.5	1.07 bits	0.22 bits	0.69 bits
X = 50.0	0.31 bits	0.58 bits	0.96 bits
X = 37.5	0.60 bits	0.83 bits	1.03 bits

Table1: Entropy densities as threshold and number of empty cells are varied. We see that entropy density still increases with percentage of empty cells and decrease with increases in threshold. My estimation is that all these values have a bias less than 0.01.

	EP = 2	EP = 10	EP = 30
X = 62.5	0.11 bits	1.25 bits	0.99 bits
X = 50.0	0.93 bits	0.93 bits	*0.76 bits
X = 37.5	0.63 bits	*0.70 bits	*0.77 bits

Table2:
Excess entropy as threshold and number of empty cells are varied. Asterisked values are one I believe are more susceptible to bias. Those that aren't asterisked I have estimated to have a bias less than 0.02.

Conclusion

The main goal of this project was to use information theory to quantify some of the results of that were qualitatively described in [2]. However, there is much more that could be done. It would be interesting to perform the same analysis for the different versions of the Schelling Model as described earlier. While it seems unlikely that changing the rules would change the structure of the stationary state, since the rules seem to only affect the dynamics that get the system to the stationary states, it is still worthwhile, in my opinion, to investigate whether or not the entropy density or excess entropy would change after making these changes to the model. Even more interesting, might be how these values change with periodic boundary conditions. However, it is clear that before such investigations can be made, the current procedure must be improved. Firstly, the simulations need to be made faster so that more data samples can be made, and secondly, a more rigorous approach analyzing the errors involved should be developed to better understand how reliable our results are. While I had a good sense of how small the bias was for many of the values measured, it would be nice to have a way of estimating the bias accurately. It would also be good to look at the variance, but this issue is a result of not having enough data. Nevertheless, even with this lack of certainty in our results, it is good to see that they strongly align with what was expected. The values measured for the entropy density clearly exhibit the trends we would expect even for smaller block entropy densities, and the excess entropy proved to be a valuable tool in analyzing the structures of the steady states by detecting clusters or a lack thereof in systems with ambiguous entropy rate densities.

Sources

[1] Feldman, David P., and James P. Crutcheld. "Structural information in two-dimensional patterns: Entropy convergence and excess entropy." Physical Review E 67.5 (2003): 051104.

[2] D. Vinkovi´c and A. Kirman. “A Physical Analogue of the Schelling Model.” Proc. Nat. Amer. Soc. 103(51), 19261 (2006).

[3] Schurmann, Thomas. “A Note on Entropy Estimation.” arXiv:1503.05911v2 [physics.data-an] 5 Jun 2015

[4]Dall’Asta, Luca, Claudio Castellano, and Matteo Marsili. “Statistical Physics of the Schelling Model of Segregation.” arXiv:0707.1681v1 [physics.soc-ph] 11 Jul 2007