Binary to Multivariate Voting

Comparing Voting Model Structure for Consensus Formation through Simulation

Daniel Lane
Dept. of Mechanical Engineering, Embry-Riddle Aeronautical University

Abstract— **The voter model is a popular tool for studying group opinion formation, with many variations studied. This project uses simulation to understand the effects of scale and interconnectivity on the stability of a network representing the shifting opinions of a population. The study follows a paper comparing a binary-choice voter model (BVM) to Robert Axelrod's multivariate voter model (MVM). Both models consider the distribution of a feature within a network based on the similarity to its neighbors' features. By investigating how initial conditions and model parameters affect consensus formation, this study could contribute to a better understanding of how beliefs evolve in diverse communities.**

Keywords— complex systems, system modeling, voter model, stochastic modeling,

I. Introduction

The voter model has been a definitive tool for studying group interaction since its evolution and distinction from as a model for describing territory won in conflict in 1973 [1]. Over the years, many variations and extensions of the BMV have been developed and analyzed in-depth for their efficacy. Despite this, the model remains simple and robust at its core. This paper serves as an introduction to the voter model for those with a background in science who are interested in learning about application of system modeling, graph theory, and complex dynamics. It provides an overview of the BVM, some of its applications and insights into opinion formation as a complex system. By the end of this paper, readers will have a solid understanding of the voter model and its potential for modeling and analyzing complex systems.

Cellular automata (CA) and stochastic models have become popular tools for studying complex systems across various domains such as math, physics, chemistry, biology, and sociology[1]–[5] They are discrete dynamical systems governed by simple rules that can simulate the behavior of cells in 2D lattice structures. One well-known example is Conway's Game of Life which is capable of modeling simulations with itself [6]. Additionally, CA have been used to model the spread of forest fires [7], traffic models [8], and magnetic fields theory[3]. While these types of models are useful to illustrate and understand when presented in 2D, they often can be applied to any number of network topologies to simulate a wide range of phenomena. This paper provides an introduction complex system modeling by demonstrating the flexibility of the BVM to adapt to new environments and how small changes in these environments can produce significantly different results.

This paper will extend the BVM onto a more complex network topology allowing for a wider range of applications in real-world scenarios. It presents elements from multiple authors, but the structure closely follows research performed by San Miguel, et. al. in 2005 [9] and attempts to simultaneously validate their results as well as reframe the analysis within the broader contextual framework of complex systems analysis. Three different scenarios will be evaluated. First, we use CA to measure the rate of convergence for the BVM over a range of network sizes to establish a baseline for model performance to be used for later comparison. The second and third phases repeat the sane data collection methods and differ from the baseline only by altering the network structure and then only the algorithm, respectively. In the third scenario, the MVM has additional hyperparameters. Therefore, we begin this task by maintaining as many of the BVM's hyperparameters as possible while performing additional simulations over a range of potential values of the new parameters. By evaluating each scenario with iterative Monty Carlo simulations times on each set of initial conditions, we ensure that the data from these stochastic models are reliable; and by isolating and changing only one variable at a time, we can more easily distinguish how the changed variable contributes to the model dynamics.

II. METHODS

A. *BVM*

The BVM is a stochastic method for detailing the interactions of agents within any interconnected network or system. For this model network a 2D lattice of size N x N is initialized with a binary choice for the opinion of votes $q = \{0,1\}$ with the desired distribution. The vote of a particular node q_i is updated with the probability equal to the percentage of opposite valued votes existing amongs the nearest neighbors q_k plus added gaussian noise σ, and only active edges have the possibility of being updated. An edge is said to be active when its vertices possess opposing votes. By measuring the ratio of active to inactive edges at every time-step (ρ), we can measure the level of disorder present in the system. The initial value of ρ_0 has an expected value equal to an expected measurement from the chosen distribution used to assign initial votes in the model. The value of ρ will decrease to zero over time as the model converges to a vote of 1 or 0 for all discretely sized networks.

The BVM model progresses along the following simple algorithm:

Algorithm 1: Bivariate voter model

Input: Vote of input node: n_i, votes of k nearest-neighbors to current node $\boldsymbol{q}_j^k = \{ q_j^1, q_j^2, \ldots q_j^k \}$

Output: Updated vote of input node: q_i

1. Select a node at random: n_i
2. Select a nearest neighbor of n_i at random: n_j
3. While $q_i \neq q_j$:

   update $q_i = q_j$ with probability: $P(\Sigma_k\, q_k/k) + \sigma$

This process continues for any number of time steps as desired, or until consensus is achieved. A Monty Carlo or similar method of iteration may be utilized to compensate for the stochasticity of the model and improved analysis.

This algorithm only requires nodes to be connected to at least two other nodes, and therefore can be applied to any network with this property.

B. *MVM*

The MVM has a similar structure the BVM made to be multivariate simply by modifying one rule, adding one rule to the algorithm and 2 new constants. The first added constant F notes the number of votes each node in the network possesses. We also allow for the each opinion q in F to be of any integer valued $q = \{2\!:\mathbb{Z}\}$, such that all F_n have the same number of opinions for all nodes. This change takes us from a binary choice voter algorithm to one with q^F choices. Active edges are now defined as ones which connect nodes with at least one differing opinion on any vote: $q_j^k \neq q_j^k$. The second new constant l_{ij} is the ratio of the shared to F for a given edge. Active edges are those with $l_{ij} = 0 < l_{ij} < 1$.

Algorithm 2: Multivariate voter model

Input: Vote of input node: q_i, votes of k nearest-neighbors to current node $\boldsymbol{q}_j^k = \{ q_j^1, q_j^2, \ldots q_j^k \}$

Output: Updated vote of input node: q_i

1. Select a node at random: n_i
2. Select a nearest neighbor of n_i at random: n_j
3. Calculate the ratio of shared to total votes between n_i and n_j: l_{ij}.
4. FF $0 < l_{ij} < F$, select one unshared vote between n_i and n_j at random, which will change from $q_i \neq q_j$ to $q_i = q_j$ with a probability $P\left(\frac{l_{ij}}{F}\right) + \sigma$

Similarly, this process continues for any number of time steps as desired, or until consensus is achieved.

C. *Simulation*

For each modeling scenario described below, specific simulations may be performed as required to understand each model's unique structure, however networks with metrics $\langle k \rangle \approx 8$ and $N = 1024$ were evaluated at least 10 times to facilitate comparisons between them.

Scenario 1 - The BVM is evaluated with a 2-D lattice network. As this is a well-studied system, it provides a baseline for comparison to any future model alterations. Multiple network sizes ranging from $N = 2^4, 2^3, \ldots, 2^7$ were examined, and initial votes were assigned with a probability $p = 0.5$. The time required to achieve consensus with this model configuration is given by $\langle p \rangle \sim \ln(t)^{-1}$, such that $\langle p \rangle$ is an ensemble average.

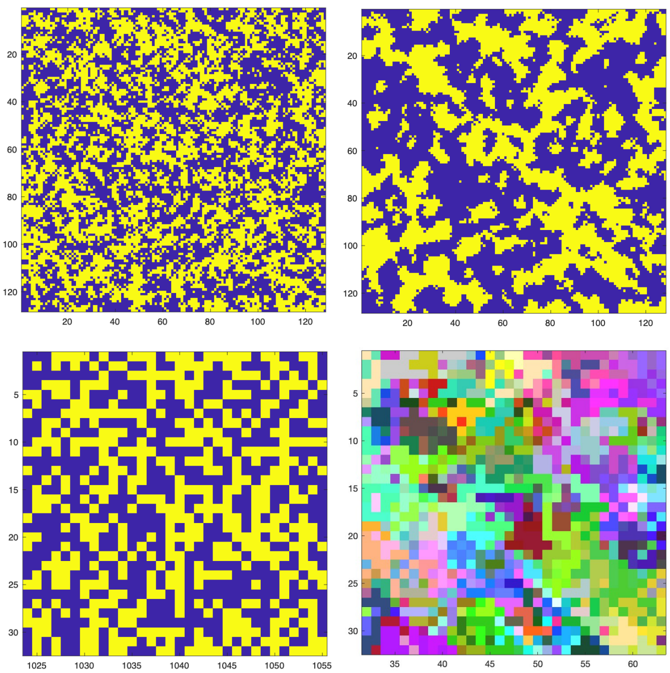

Figure 1: Visualizations from simulation scenarios: (top-left) BVM with 2D lattice network at $t = 30{,}000$. (Top-right) BVM with 2D lattice network at $t = 180{,}000$. (Bottom-left) BVM with scale-free network at $t = 50{,}000$. (Bottom-right) MVM with 2D lattice network at $t = 50{,}000$.

Scenario 2 - The BVM is evaluated with a sale-free network. A network is said to be scale-free when the probability for the number of edges of any given vertex is $P(k) \sim k^{-\gamma}$. The scale-free model consisted of N = 1024 nodes and was constructed via method of preferential attachment [10], with $\langle k \rangle = 7.96 \approx 8$ to maintain similarity to the 2-D lattice described in scenario 1. The average interface density ρ is again measured at each of 50,000 time-steps. The algorithm described in IIA (Algorithm 1) was followed according to the adjacency map created for the network. However, for ease of comparability the node values were assigned a 'pixel' in the same visualization matrix used for the 2-D lattice simulations. This allows for an easier estimation for the number of votes present in the model at any given time step across model construction.

Scenario 3 – The MVM was evaluated on a network of size $N = 1024$ and $\langle k \rangle \approx 8$. Constants F and q were initially investigated at values of 3 and 10 respectively to gauge MVM performance with the literature presented [9]. With this model configuration we assign one of the three votes in F to a RGB (red, green, blue) color value to represent specific vote configuration as a unique color value.

Additional performance and metric evaluations were planned for values of $F = 10$ and $q = 10 \times \{1, 2, \ldots 10\}$ to capture the effective relationship between these two parameters. There is a phase change that occurs at approximately $q/F = 5$ where the MVM stops converging to a single vote and converges to a monocultural network with several specific opinions being represented in sections of locally ordered groups.

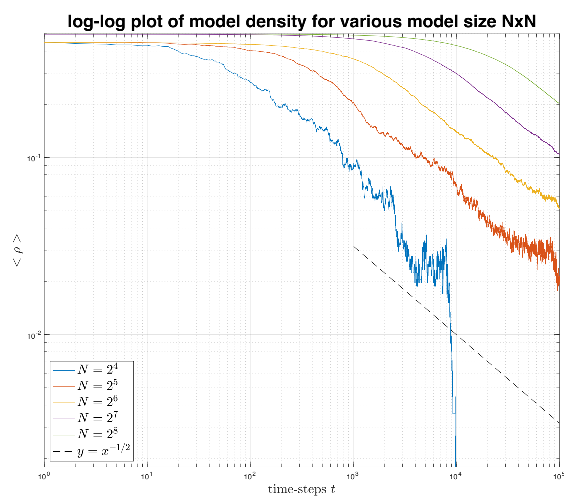

Figure 2: Log plot of average interference density for the BVM on a 2-D lattice for models of size $N = 2^4, 2^3, \ldots, 2^7$ (solid), and an exponential curve of $y = x^{-0.5}$ (dashed).

III. RESULTS

A. *Scenario 1*

When the average network disorder $\langle p \rangle$ for each network size is plotted in time, we find that the rate of decay is independent of network size. Figure 2. shows these data are consistent with a power-law of $\langle p \rangle \sim t^{-0.5}$ Given this rate of decay, it is estimated that the largest model presented of $N = 1024$ would require an order of $t = 10^{10}$ time-steps to converge.

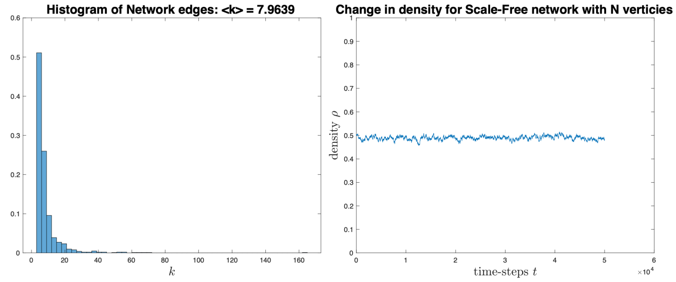

Figure 3: Histogram for scale-free network of N $= 1024$, $\langle k \rangle \approx 8$ *used in scenario 2 (left). Plot of* $\langle$p$\rangle$*(t) used in scenario 2 showing a meta-stabile region in time of* $\langle$p$\rangle \approx 0.49$ *(right)*

B. *Scenario 2*

When used with the scale-free network, the BVM rapidly entered a region of metastability around $\langle$p$\rangle \approx 0.49$ and remained in this configuration for the duration of $50{,}000$ time-steps. Figure 2. (left) shows the non-uniform network structure that is present with scale-free networks. The average number of edges present for nodes in the network was measured at $\langle k \rangle = 7.96$ which is quite similar to that of the 2-D lattice, with a heavy bias towards vertices possessing fewer edges. This skew however was enough to shift the model's ability to converge within the time scale provided.

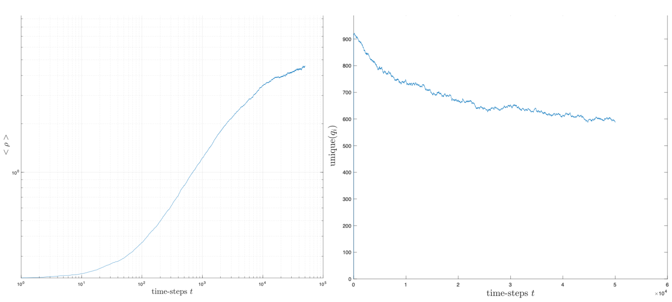

Figure 4: Log plot of average interference density (left) and plot average number of unique voters with time (right) for the Axelrod MVM on a 2-D lattice with model parameters: $N = 1024, F = 3, q = 10$ *in scenario 3.*

C. *Scenario 3*

Figure 2. shows dramatic inconsistency with expected behavior. When the average network disorder $\langle p \rangle$ for each network size is plotted in time, we find that the rate of decay increases well beyond the accepted bounds for the parameter (up to 4.5). Simultaneously the number of unique votes does tend to increase. These data trends are inconsistent with each other and reveal a dramatic error in that was experienced when coding the model.

IV. Discussion

While the results from this research show promise, there are several inconsistencies which need to be addressed. The discrepancy found from the simulation results do not accurately what the literature has detailed for these models.

In scenario 1, the only converging model presented is the smallest with only 16 nodes. While the other models trended towards convergence, the time scale appeared to be extremely exaugurated. At first glance, the decay rate follows a power series of $\sim t^{-0.5}$. The author believed this to be correct until other inconsistencies appeared. Upon further literature review, this same power law decay was confirmed for BVMs, however only in 1-D models. The proper decay rate for 2-D systems is $\sim \ln(t)^{-1}$ [9].

For scenario 2, we did not again did not experience the same convergence. For scale-free networks, we should see this metastability end after $t \sim N^{0.88}$ which for a network size of N=1024 should be two orders of magnitude less than the 50,000 time-steps that have been recorded here. Having performed this experiment ten times, it can reasonably be assumed that this is not some stochastic unlikelihood, but rather an inconsistency in the model or in the interpretation of the literature.

Finally, in scenario 3 wo do not see any data that is consistent with intuition or that is expected from intuition. In fact, there was a significant issue related to computer memory that contributed to the premature termination of data collection.

The data collection for this research ended earlier than necessary to fully analyze the model and the contributions made by network application and multivariate enhancement. Future research may focus on resuming data on the MVM and further expanding the findings presented here to examine the phase change that occurs when examining the ratio of q/F.

V. References

[1] P. CLIFFORD and A. SUDBURY, "A model for spatial conflict," *Biometrika*, vol. 60, no. 3, pp. 581–588, Dec. 1973, doi: 10.1093/biomet/60.3.581.

[2] S. Wolfram, “Cellular automata as models of complexity,” *Nature*, vol. 311, no. 5985, Art. no. 5985, Oct. 1984, doi: 10.1038/311419a0.

[3] E. Ising, “Beitrag zur Theorie des Ferromagnetismus,” *Z. Für Phys.*, vol. 31, no. 1, pp. 253–258, Feb. 1925, doi: 10.1007/BF02980577.

[4] R. Axelrod, “The Dissemination of Culture: A Model with Local Convergence and Global Polarization,” *J. Confl. Resolut.*, vol. 41, no. 2, pp. 203–226, 1997.

[5] G. B. Ermentrout and L. Edelstein-Keshet, “Cellular Automata Approaches to Biological Modeling,” *J. Theor. Biol.*, vol. 160, no. 1, pp. 97–133, Jan. 1993, doi: 10.1006/jtbi.1993.1007.

[6] P. Rendell, “Game of Life Universal Turing Machine,” in *Turing Machine Universality of the Game of Life*, P. Rendell, Ed., in Emergence, Complexity and Computation. Cham: Springer International Publishing, 2016, pp. 71–89. doi: 10.1007/978-3-319-19842-2_5.

[7] I. Karafyllidis and A. Thanailakis, “A model for predicting forest fire spreading using cellular automata,” *Ecol. Model.*, vol. 99, no. 1, pp. 87–97, Jun. 1997, doi: 10.1016/S0304-3800(96)01942-4.

[8] K. Nagel and M. Schreckenberg, “A cellular automaton model for freeway traffic,” *J. Phys. I*, vol. 2, no. 12, pp. 2221–2229, Dec. 1992, doi: 10.1051/jp1:1992277.

[9] M. S. Miguel, V. M. Eguiluz, R. Toral, and K. Klemm, “Binary and multivariate stochastic models of consensus formation,” *Comput. Sci. Eng.*, vol. 7, no. 6, pp. 67–73, Nov. 2005, doi: 10.1109/MCSE.2005.114.

[10] M. George, “B-A Scale-Free Network Generation and Visualization,” *MATLAB Central File Exchange*, Apr. 30, 2023. https://www.mathworks.com/matlabcentral/fileexchange/11947-b-a-scale-free-network-generation-and-visualization (accessed Apr. 25, 2023).