REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
	Technical Report	-

4. TITLE AND SUBTITLE
Eaxct MArkov Propagator for the Voter Models

5a. CONTRACT NUMBER
W911NF-12-1-0546

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
611102

6. AUTHORS
C Lim, William Pickering

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
Rensselaer Polytechnic Institute
110 8th Street

Troy, NY 12180 -3522

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
62449-NS.16

12. DISTRIBUTION AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT
Exact solutions for the standard Voter model in terms of the spectrum and eigenvectors of the Markov propagator for macroststes are first obtained for the complete graph and then extended to the case of uncorrelated networks of inhomogeneous degrees. Using this method, we have recalculated the expected times to consensus of Sood and Redner. More importantly, we have used this method to calculate the variance of consensus times and all higher moments of the Voter models.

15. SUBJECT TERMS
Complete Spectral Resolution of the Markov Propagator

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Chjan Lim
UU	UU	UU	UU		19b. TELEPHONE NUMBER 518-276-6904

Report Title

Eaxct MArkov Propagator for the Voter Models

ABSTRACT

Exact solutions for the standard Voter model in terms of the spectrum and eigenvectors of the Markov propagator for macroststes are first obtained for the complete graph and then extended to the case of uncorrelated networks of inhomogeneous degrees. Using this method, we have recalculated the expected times to consensus of Sood and Redner. More importantly, we have used this method to calculate the variance of consensus times and all higher moments of the Voter models.

Exact Solution of the Markov Propagator for the Voter Model on the Complete Graph

William Pickering and Chjan Lim
Rensselaer Polytechnic Institute
(Dated: July 1, 2014)

The voter model is a simple and well studied model in network science in which nodes in a graph have one of two states. We assume throughout the discussion that every node is connected to all other nodes. By restricting the network topology in such a way, we are able to find a very detailed solution of the model. A single step propagator, as well as the solution to its spectral problem, will both be derived exactly by means of generating function techniques. The eigenvalue problem will then be cast in the thermodynamic limit as the number of nodes, N, becomes infinite, and it is found that the eigenvectors will approach polynomials. Also, using the exact solution to the spectral problem, we can find the time m probability distribution. We then use this to find two quantities. This first is the expected time until the system reaches total agreement and the second is the expected frequency that each macrostate is achieved before consensus is attained. Other quantities of interest will follow in closed form given the solutions provided here.

I. INTRODUCTION

Quantitative insight into social opinion dynamics can be found from a modeling perspective. One of the most well studied of which is the voter model [1, 2]. This model has been solved analytically for several quantities [3–7]. It also has applications in many areas of physical science, such as particle interactions and kinetics of catalytic reactions [6]. Social opinion dynamics is a typical view of this particle interaction model, where there are two opposing opinions in a political discussion. The fundamental social assumption about the model is that individuals are strongly influenced by the beliefs of their neighbors. Another interpretation of the voter model is that species compete for territories, where a change in state corresponds to and invasion [2, 8]. Extensions and generalizations of this model also have been studied in detail [9–14].

A specific and direct application of the voter model is to population genetics. In the context of a binary coded genetic algorithm, the voter model on a complete graph can be interpreted as the crossover operator of bit strings over a single bit. This is the exchange of bits once two parent strings have been selected [15]. The voter model is an instance of the genetic algorithm that examines the case where selection is uniform (constant fitness), only one child string is generated per crossover, and there are no mutations. A biased voter model, which introduces a fitness value for each state, also has been studied [9]. The genetic algorithm, with or without mutations, is a subject of interest from an analytical perspective [15–17]. The work presented here will provide some additional analytical insight into unbiased genetic drift.

Many mathematical models for opinion dynamics can be viewed as discrete time Markov chains. It is known that the eigenvalues and eigenvectors of the transition matrix of the Markov chain have a vital role in the dynamics of the model. In general, the second largest eigenvalue is related to an approximate rate of convergence to equilibrium [18]. For the voter model, this is when all nodes have the same state. All dynamics of the model halt entirely once this network state is attained. On a connected graph, the expected time to attain this unanimous state can be bounded in terms of the spectral gap of the transition matrix [19]. If one had access to the complete spectral decomposition of the transition matrix, then any future probability distribution can be computed in closed form. In this paper, we offer a method for finding the closed form solution of the spectral problem that yields such a decomposition for an instance of the voter model.

The voter model can be cast as an urn problem similar to the Ehrenfest model [20]. In this model, there are two urns with N balls divided amongst them. In a single time step, one ball is chosen at random and is transferred to the other urn. This process is repeated ad infinitum. The voter model on the complete graph has a similar interpretation. Here, two balls are selected, one after the other. After selection, both of the balls are placed in the urn from which the second ball came. When cast as an urn problem, the network topology is identical to a complete graph. A closed form diagonalization of the Markov transition matrix for the Ehrenfest model was found by Mark Kac in 1947 using similar techniques we will utilize here [21]. We will propose a generalization of those techniques which allows us to solve the spectral problem for the voter model. This particular solution of the model will give exact analytical expressions for several quantities of interest. The two specific applications of the solution to the spectral problem that we will provide are the expected time to consensus, and the expected time each macrostate is visited before reaching consensus.

An outline of the paper is as follows. Section II will describe more of the details of the model and the associated random walk. In section III, we shall derive a straightforward procedure to exactly solve the eigenvalue problem of the Markov transition matrix for the macrostates of the system. With such a solution, one can compute the probability distribution at any future time in closed form. In section IV, we use this solution to find two quantities: the expected time before reaching consensus and the expected time spent at each macrostate. These expressions are functions of the initial probability distribution, which is kept arbitrary. Although an asymptotic treatment of consensus times is known [9, 22], we will improve upon those results significantly by providing an exact value for the expected time to consensus for any initial probability distribution.

II. THE 2-STATE VOTER MODEL

Although the model that we consider is imposed upon a complete graph of N nodes, more general graphs are studied as well [3, 9, 23, 24]. Each node is assigned one of two states, A or B. In a single time step, a node is chosen randomly and will assume the state of one of its neighbors, also chosen randomly [2]. Note that in this procedure, it is possible that the network state may remain unchanged for several time steps.

Let $n_A(m)$ and $n_B(m)$ represent the total number of agents taking opinion A and B respectively at time m. Since the total number of nodes must be conserved and all nodes must take one of these two opinions, it is necessary that $n_A(m) + n_B(m) = N$. Since N is a constant, this will allow us to simplify the model to a random walk in a single variable, say n_A.

Now let us formalize the problem as a random walk in n_A. We can write it as

$$n_A(m+1) = n_A(m) + \Delta n_A(m). \tag{1}$$

For a given time step m, n_A is considered to be a known constant, and the random behavior is exhibited in Δn_A. Since only a single node is updated per time step, Δn_A only takes values from $\{-1, 0, 1\}$. Let $p_j = \frac{j(N-j)}{N(N-1)}$. Then, from the definition of the model, the probabilities of taking these values are:

$$Pr(\Delta n_A(m) = 1 | n_A(m) = j) = p_j \tag{2}$$

$$Pr(\Delta n_A(m) = -1 | n_A(m) = j) = p_j \tag{3}$$

$$Pr(\Delta n_A(m) = 0 | n_A(m) = j) = 1 - 2p_j \tag{4}$$

Thus, the random walk is determined given an initial condition $n_A(0) = n$.

A. Markov Propagator By Generating Functions

In this section, we will outline a general procedure that provides a recurrence relation for the probability distribution of general random walks. The strategy is to construct a sequence of generating functions for the probability distribution of the random walk. This process can also be applied to other models, incomplete graphs, or to multiple dimensions. An advantage of this approach is that it is highly generalizable in these ways and only requires minor modification to do so. The procedure can be applied to find the probability distribution of either the microstates or the macrostates of the model. Here, the macrostate approach will be utilized since the model is imposed on a complete graph.

To begin, represent the probability distribution in generating function form. Let $a_j^{(m)} = Pr(n_A(m) = j)$. We introduce a sequence of generating functions $R^{(m)}(x) = \sum_j a_j^{(m)} x^j$ and seek to find a relationship between $R^{(m+1)}(x)$ and $R^{(m)}(x)$. From the random walk form of the model, the probability generating function for Δn_A is $D_j(x) = p_j x + (1 - 2p_j) + p_j x^{-1}$. We will make use of the following properties of generating functions in the derivations to follow:

Product rule: If X and Y are integer random variables with probability generating functions $F(x)$ and $G(x)$ respectively, then the generating function of $X + Y$ is $F(x)G(x)$.

Sum rule: If the probability space is partitioned into N events, each with generating function $F_j(x)$, then the generating function for the entire space is $\sum_{j=1}^N F_j(x)$ [25–27].

For time step m, suppose that $n_A(m) = j$. Note that $a_j^{(m)} x^j$ is the corresponding generating function for this event. Now, utilize the product rule in equation (1) to deduce that the probability generating function for time $m+1$ is $a_j^{(m)} x^j D_j(x)$ in the event that $n_A(m) = j$. By the sum rule, we have that

$$R^{(m+1)}(x) = \sum_{j=0}^N a_j^{(m)} x^j D_j(x). \tag{5}$$

This is the generating function form of the Markov propagator of the random walk. This can be easily generalized to other models simply by specifying the appropriate expression for $D_j(x)$ in equation (5). For the given network topology, the sum is simple enough to collect terms and obtain an explicit equation:

$$a_j^{(m+1)} = p_{j-1} a_{j-1}^{(m)} + (1 - 2p_j) a_j^{(m)} + p_{j+1} a_{j+1}^{(m)}. \tag{6}$$

While the procedure to find a relationship between $R^{(m+1)}$ and $R^{(m)}$ can be highly generalizable, we pose a new power series that will be utilized directly to solve the formulation in equation (6). Let $Q^{(m)}(x, y) = \sum_j a_j^{(m)} x^j y^{N-j}$. Some properties of generating functions of this type are listed below:

1. Multiply $Q^{(m)}$ by x/y to shift $a_j^{(m)} \to a_{j-1}^{(m)}$.
2. Multiply $Q^{(m)}$ by y/x to shift $a_j^{(m)} \to a_{j+1}^{(m)}$.
3. The generating function for $p_j a_j^{(m)}$ is $\frac{xy}{N(N-1)} Q_{xy}^{(m)}$.

Using these properties, we rewrite equation (6) as

$$(x - y)^2 Q_{xy}^{(m)} = N(N-1) \Delta_{+m} Q^{(m)}. \tag{7}$$

Here, Δ_{+m} is the forward difference operator in the discrete variable m. The goal of the subsequent section is to solve this equation explicitly for m to find the probability distribution for arbitrary time steps. To do so, we turn to the spectral problem.

III. THE SPECTRAL PROBLEM

It is clear from equation (6) that the future probability distribution can be expressed as a tridiagonal transition matrix multiplied by the probability vector $\mathbf{a}^{(m)} = [a_j^{(m)}]_{j=1}^N$. Solving the spectral problem provides a basis for the initial distribution, which allows future time steps to be computed in closed form. We find the solution to this problem here.

A. Matrix Eigenvalue Problem

For eigenvalue λ with eigenvector $\mathbf{v} = [c_j]_{j=0}^N$, we have that the spectral problem for the propagator can be written as

$$p_{j-1}c_{j-1} + (1-2p_j)c_j + p_{j+1}c_{j+1} = \lambda c_j. \tag{8}$$

The solution to the problem begins by defining a multivariate generating function for c_j as

$$G(x,y) = \sum_j c_j x^j y^{N-j}. \tag{9}$$

We seek a differential equation for the eigenvalue problem. This procedure is similar to Kac's solution of the Ehrenfest model [21]. Here, however, we will obtain a PDE instead of an ODE. Using the formulation in section II gives the following partial differential equation for G:

$$(x-y)^2 G_{xy} = N(N-1)(\lambda-1)G. \tag{10}$$

Note that this equation has no boundary or initial conditions. To find the appropriate solutions to this equation, we assume that it must take the form specified in equation (9), not all coefficients $c_j = 0$, and that $c_j = 0$ for $j < 0$ and $j > N$.

To solve this, use the change of variables $u = x-y$ and $H(u,y) = G(x,y)$ to transform the partial differential equation into

$$u^2(H_{uy} - H_{uu}) = N(N-1)(\lambda-1)H. \tag{11}$$

Since this is a linear change of variables, we still expect solutions to be of the form $H(u,y) = \sum_j b_j u^j y^{N-j}$. Substituting this into the new partial differential equation for H and collecting like terms will give the difference equation for b_j:

$$(j-1)(N-j+1)b_{j-1} = (j(j-1) + N(N-1)(\lambda-1))b_j. \tag{12}$$

Recall that we had required $c_j = 0$ for $j < 0$ and $j > N$. This is also true for b_j since we applied a linear transformation. Therefore, this difference equation would suggest that every $b_j = 0$ unless it has singular behavior for some value of $j \in \{0...N\}$, say when $j = k$. At this singular point, set the coefficient of b_k to 0. This allows the value of b_k to be non-zero, and thus not all $b_j = 0$. Therefore, the set of eigenvalues is determined to be

$$\lambda_k = 1 - \frac{k(k-1)}{N(N-1)},\ k = 0...N. \tag{13}$$

Now we can find the eigenvectors. Since the equation for b_j has singular behavior at b_k, the value at this point is arbitrary. This is expected, since any multiple of an eigenvector remains an eigenvector. Without any loss, let $b_k = 1$. Now we can ascertain the explicit solution for b_j when $k < j$:

$$b_j = \prod_{i=k+1}^{j} \frac{(i-1)(N-i+1)}{N(N-1)(\lambda_k - 1) + i(i-1)}. \tag{14}$$

When $j < k$, we have $b_j = 0$. Thus all values of b_j are determined for a given eigenvalue λ_k. Note that in (14), the requirement that $b_j = 0$ is satisfied when $j > N$.

Now, we use b_j to find the components of the eigenvector, c_j. To do this, express $H(u,y)$ in the original x, y variables:

$$\begin{aligned} G(x,y) &= H(u,y) && (15)\\ &= \sum_i b_i (x-y)^i y^{N-i} && (16)\\ &= \sum_{i=0}^{N}\sum_{j=0}^{i} (-1)^{i-j} b_i \binom{i}{j} x^j y^{N-j} && (17)\\ &= \sum_{j=0}^{N}\sum_{i=j}^{N} (-1)^{i-j} b_i \binom{i}{j} x^j y^{N-j}. && (18) \end{aligned}$$

Therefore $c_j = \sum_{i=j}^N (-1)^{i-j} b_i \binom{i}{j}$, and thus the spectral problem is solved in closed form.

B. Differential Eigenvalue Problem

In this section, we will examine the eigenvectors in more detail. In particular, we wish to consider the thermodynamic limit of the model as $N \to \infty$ and study the behavior of the eigenvectors. To do this, note that the spectral problem can be posed as:

$$\Delta_j^2(p_j c_j) = (\lambda_k - 1)c_j \tag{19}$$

Where Δ_j^2 is the second centered difference operator. Let $x_j = j/N$, $\Delta x = 1/N$ and $u(x_j) = c_j$. Then,

$$\frac{\Delta_j^2(p_j u(x_j))}{\Delta x^2} = N^2(\lambda_k - 1)u(x_j). \tag{20}$$

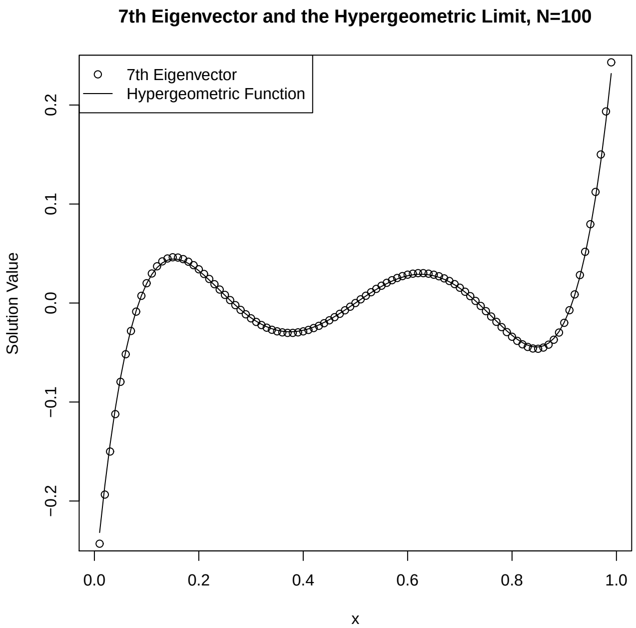

FIG. 1. 7th Eigenvector of the discrete problem plotted with the exact solution for the limit as $N \to \infty$. The hypergeometric function in the figure is a fifth degree polynomial.

Thus, take $N \to \infty$ to obtain the differential equation for the eigenfunctions of the continuous propagator:

$$\frac{d^2}{dx^2}[x(1-x)u(x)] = -k(k-1)u(x). \tag{21}$$

Expanding all derivatives yields

$$x(1-x)\frac{d^2u}{dx^2} + (2-4x)\frac{du}{dx} + (k(k-1)-2)u = 0, \tag{22}$$

which is valid on $0 < x < 1$. This is a form of the hypergeometric differential equation, and as such, the basis for the solutions are

$$u_k(x) = {}_2F_1(k+1, 2-k, 2, x) \tag{23}$$

for $k = 0, 1, 2...$ This is a special case of the hypergeometric function in which the series expression terminates for each k [28]. This implies that the kth eigenvector of the voter model approaches a polynomial of degree $k-2$ as $N \to \infty$. When $k = 0$ and $k = 1$, the solution is $u(x) = 0$. This differential equation does not describe the behavior of the eigenvectors at the boundary, which explains why the first two eigenfunctions are trivial. In figure 1, the seventh eigenvector for the discrete model when $N = 100$ has very close agreement with the continuous solution posed in (23).

IV. APPLICATIONS OF THE SPECTRAL SOLUTION

The solution of the spectral problem yields exact expressions for the probability distribution at any future time step. With such a strong result, quantities that depend on macrostate probabilities naturally follow from it. In this section, we will apply the solution of the spectral problem to find two quantities. The first is the expected time for the system to reach consensus and the second is the expected time spent at each macrostate before consensus. Both quantities are considered functions of the initial distribution of macrostates.

A. Consensus Time

A topic of particular importance with social models such as this is the expected time until all agents in the network have the same state. Once such a state is achieved, the dynamics halt entirely. In this section, we will provide an exact expression for the consensus time. This calculation depends on the initial distribution of the model, however an estimate that is independent of initial data can also be determined. We know that if q_m is the probability of reaching consensus at step m, then the consensus time is $\sum_{m=1}^{\infty} q_m m$. We can determine q_m from the boundary conditions of the transition equation. The probability of reaching consensus at time step $m+1$ is the probability that the model is in consensus at time $m+1$ minus the probability it was already in consensus at time m. Symbolically, this is $q_{m+1} = a_0^{(m+1)} + a_N^{(m+1)} - a_0^{(m)} - a_N^{(m)}$. Now from the boundary conditions, we have that

$$a_{0,N}^{(m+1)} - a_{0,N}^{(m)} = \frac{1}{N} a_{1,N-1}^{(m)}. \tag{24}$$

Thus $q_m = (a_1^{(m-1)} + a_{N-1}^{(m-1)})/N$. If $\mathbf{a^{(m)}}$ takes components $a_j^{(m)}$, and letting $\mathbf{v_k}$ be the kth eigenvector, then we can express any such distribution as $\mathbf{a^{(m)}} = \sum_{k=0}^{N} d_k \lambda_k^m \mathbf{v_k}$. Here, d_k is the initial probability distribution expressed in the eigenbasis. We only need components 2 and $N-1$ from to find the consensus time, so we let $s_k = d_k[\mathbf{v_k}]_2 + [\mathbf{v_k}]_{N-1}$ and write

$$a_1^{(m)} + a_{N-1}^{(m)} = \sum_{k=2}^{N} s_k \lambda_k^{m-1}. \tag{25}$$

We exclude the $k = 0$ and $k = 1$ terms since $a_1^{(m)}$ and $a_{N-1}^{(m)}$ are independent of those eigenvectors. Letting $\mathbf{V}$ be the matrix of eigenvectors, we will write the expected time to consensus as

$$E[T|\mathbf{a^{(0)}}=\mathbf{Vd}] = \sum_{m=1}^{\infty}\frac{1}{N}\sum_{k=2}^{N} s_k\lambda_k^{m-1}m \tag{26}$$

$$= \sum_{k=2}^{N}\frac{1}{N}s_k\frac{1}{(1-\lambda_k)^2} \tag{27}$$

$$= N(N-1)^2\sum_{k=2}^{N}\frac{s_k}{(k(k-1))^2}. \tag{28}$$

This is an exact formula given the initial distribution. Now let us estimate this quantity to ascertain asymptotic information of consensus times. To do this, we need to find a restriction on s_k. Consider the limit of the system as $m\to\infty$. Since the system will always reach consensus in finite time, we have that $a_0^{(\infty)}+a_N^{(\infty)}=1$. Let us now estimate the right hand side of (25) by using the largest available eigenvalue, λ_2, to obtain that for some s,

$$a_1^{(m)}+a_{N-1}^{(m)}\leq s\lambda_2^{m-1}. \tag{29}$$

If we sum both boundaries in (24) over all m, the left hand side is telescopic. If we assume that the initial probability of consensus is not $O(1)$, then adding the two boundaries together gives that

$$\binom{N-1}{2}s = O(1) \tag{30}$$

$$s = O(1/N) \tag{31}$$

With this, we obtain the asymptotic behavior of the expected time to consensus as

$$E[T|\mathbf{a^{(0)}}=\mathbf{Vd}] \leq O\left(\frac{1}{N^2}\right)\sum_{m=1}^{\infty}\lambda_2^{m-1}m \tag{32}$$

$$= O\left(\frac{1}{N^2}\right)\frac{1}{(1-\lambda_2)^2} \tag{33}$$

$$= O(N^2) \tag{34}$$

Thus, as a uniform estimate for general initial probability distributions of the model, the asymptotic behavior of the expected time to consensus is quadratic, which is consistent with previous results [9, 22, 23]. Using continuous time methods, the expected time to reach consensus given a opinion density ρ is $E[T]=N^2\left[(1-\rho)\ln\frac{1}{1-\rho}+\rho\ln\frac{1}{\rho}\right]$ [9]. The exact solution and the bound we found here are improvements upon this result since they are valid for small values of N and generalize the theory to general initial probability distributions.

B. Macrostate Times

We define *macrostate times* as the expected frequency of each state of the random walk prior to consensus. Unlike the consensus time, macrostate times will be organized as a vector whose components correspond to each state. The consensus states are not included because they have an infinite macrostate time. For non-consensus states, we expect macrostate times to be finite since the consensus time is also finite. Macrostate times are stronger quantities than the consensus time since the sum of all macrostate times is identical to the consensus time.

1. Discrete Time Solution

Since the solution of the spectral problem is known, computing macrostate times becomes straightforward. Let $M_j(m)$ be the total number of visitations of macrostate j by time m and let $\Delta M_j(m)=M_j(m)-M_j(m-1)$. Note that $M_j(m)$ depends on the outcome of the random walk for n_A. Since $a_j^{(m)}$ is the defined as the probability that $n_A(m)=j$, we have that $\Delta M_j(m)$ takes value 1 with probability $a_j^{(m)}$. Otherwise, it takes value 0. Thus, for $j=1...N-1$, we can write the macrostate time as

$$E[M_j]=\sum_{m=0}^{\infty}E[\Delta M_j(m)]=\sum_{m=0}^{\infty}a_j^{(m)}. \tag{35}$$

We can use the solution to the spectral problem to compute this infinite series exactly. Let $\mathbf{M}$ be a vector whose components are $E[M_j]$. Also, letting d_k be the initial distribution of the macrostates expressed in the eigenbasis, we have that the time m probability distribution is $\mathbf{a^{(m)}}=\sum_{k=0}^{N}d_k\lambda_k^m\mathbf{v}_k$. The $k=0$ and $k=1$ terms in the sum are the contributions of the consensus states to the probability distribution. Since consensus is a frozen state, the probability distribution for the relevant macrostates are independent of the first two terms in this sum. The macrostate times can be found once these terms are discarded:

$$E[\mathbf{M}]=\sum_{m=0}^{\infty}\sum_{k=2}^{N}d_k\mathbf{v}_k\lambda_k^m \tag{36}$$

$$= N(N-1)\sum_{k=2}^{N}\frac{d_k}{k(k-1)}\mathbf{v}_k \tag{37}$$

The first and last components of this vector are irrelevant since it is understood that $M_0=M_N=\infty$. The remaining components of $\mathbf{M}$ are the exact values for the macrostate times.

We perform Monte Carlo simulation to reinforce this result. Three cases of the initial distribution are considered: $a_j^{(0)}=\delta_{j,N/2}$, $a_j^{(0)}=\delta_{j,N/4}$, and $a_j^{(0)}=\frac{1}{N+1}\ \forall j=0...N$. Figure 2 shows that there is good agreement between the exact solutions and the results from the simulations. A particularly interesting case is when the initial distribution is uniform. This distribution happens to be the eigenvector corresponding to the second largest eigenvalue of the transition matrix. So, in equation (37), we

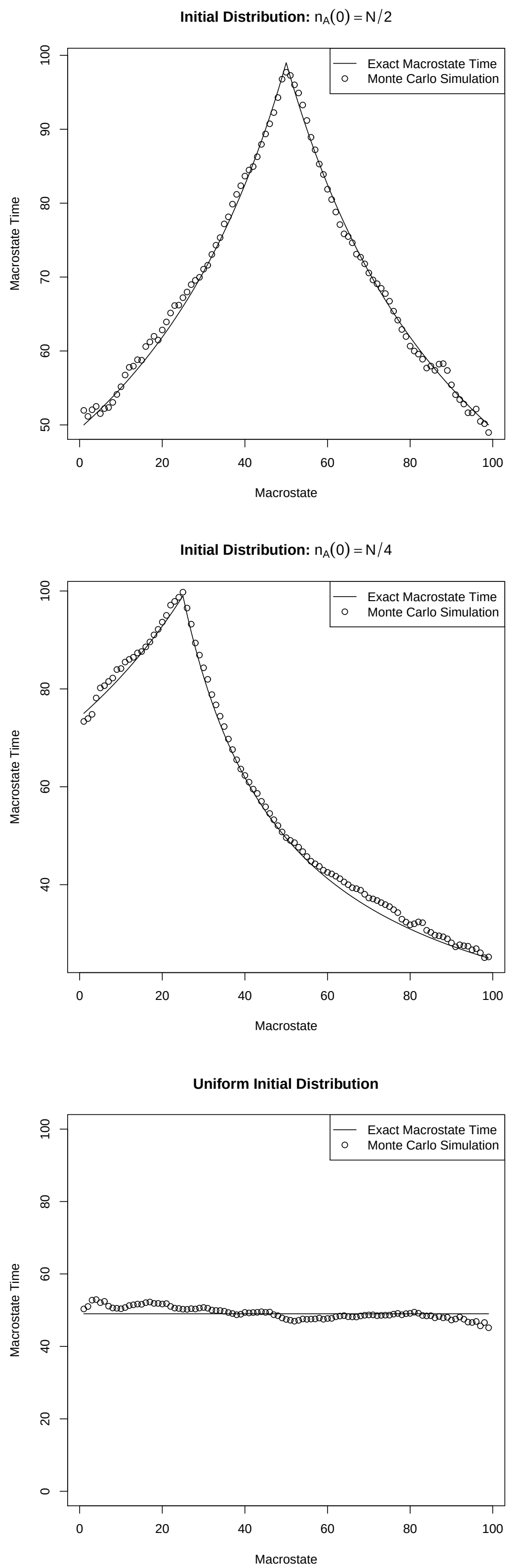

FIG. 2. Three cases of the initial distribution are examined. The exact expression for the macrostate time in equation (37) is compared with averaged results from a Monte Carlo simulation over 3000 runs of the voter model when $N = 100$.

know that $d_2 = \frac{1}{N+1}$ and $d_k = 0$ otherwise. Therefore, the time spent at each macrostate is also uniform with value $\frac{N(N-1)}{2(N+1)} \sim \frac{N}{2}$.

2. Continuous Time Solution

Here, we will find the expected time spent per macrostate as $N \to \infty$ using continuous time methods. While the discrete solution is valid for every N, the continuous time solution will provide some additional insight into its behavior for large N. Let $\rho_j = j/N$, $t_m = m/N$, and $u(\rho_j, t_m) = N a_j^{(m)}$. Using this, the Fokker-Plank equation for the model is

$$\frac{\partial u}{\partial t} = \frac{1}{N} \frac{\partial^2}{\partial \rho^2} [\rho(1-\rho) u(\rho, t)]. \tag{38}$$

Furthermore, let $M(\rho)$ be the expected time spent at density ρ prior to consensus. As $N \to \infty$, we have that

$$M(\rho) = \sum_{m=0}^{\infty} a_j^{(m)} \to \int_0^{\infty} u(\rho, t) dt. \tag{39}$$

Now, integrate equation (38) from $t = 0$ to $t = \infty$ to obtain

$$u(\rho, \infty) - u(\rho, 0) = \frac{1}{N} \frac{\partial^2}{\partial \rho^2} [\rho(1-\rho) M(\rho)]. \tag{40}$$

Since the system reaches consensus in finite time, we have that $u(\rho, \infty) = 0$. Let $f(\rho) = u(\rho, 0)$ and $T(\rho) = \rho(1-\rho)M(\rho)/N$. Then, the problem becomes

$$-f(\rho) = \frac{\partial^2 T}{\partial \rho^2} \tag{41}$$

with boundary conditions $T(0) = T(1) = 0$. The solution of this problem is found by determining the Green's function of the differential operator [29]. For this problem, the Green's function is given by

$$\tilde{g}(\rho, \xi) = \begin{cases} \rho(1-\xi) & \rho < \xi \\ \xi(1-\rho) & \rho > \xi \end{cases}. \tag{42}$$

With this solution, we have that

$$T(\rho) = \int_0^1 f(\xi) \tilde{g}(\rho, \xi) d\xi. \tag{43}$$

Therefore, the expression for macrostate time for large N is

$$M(\rho) \sim N \int_0^1 f(\xi) g(\rho, \xi) d\xi. \tag{44}$$

where

$$g(\rho,\xi) = \begin{cases} \frac{1-\xi}{1-\rho} & \rho < \xi \\ \frac{\xi}{\rho} & \rho > \xi \end{cases} . \tag{45}$$

The Green's function, $g(\rho,\xi)$, is the solution to macrostate times when the initial density is specified as a given value ξ. Furthermore, when the initial probability distribution is uniform, the solution is also uniform with value $M(\rho) = N/2$ as we have observed with the discrete formulation.

V. THE COMPLETE BIPARTITE GRAPH

The methods we have developed for the complete graph can be extended to more complex networks. In particular, we will consider the complete bipartite graph. In this case, nodes in the network are divided into two groups. Every node in a group is connected to every node in the other group. The complete bipartite graph can also be defined as the complement of two complete graphs. Let N_1 be the number of nodes in the first group and N_2 be the total number of nodes in the second group. Also, let $n_A^{(1)}(m)$ and $n_A^{(2)}(m)$ be the number of nodes with opinion A in groups 1 and 2 respectively.

Letting $a_{ij}^{(m)} = Pr(n_A^{(1)}(m) = i,\ n_A^{(2)}(m) = j)$, and defining a power series $Q^{(m)}(x,y,u,v) = \sum_{i,j} a_{ij}^{(m)} x^i, y^{N_1-i} u^j v^{N_2-j}$, we can use the procedure laid out in section II to find the single step propagator for the model:

$$\left[\frac{u(x-y)}{NN_2} - \frac{y(u-v)}{NN_1}\right] Q_{yu}^{(m)} + \left[\frac{-v(x-y)}{NN_2} + \frac{x(u-v)}{NN_1}\right] Q_{xv}^{(m)} = \Delta_{+m} Q^{(m)}. \tag{46}$$

Unlike for the complete graph, the exact solution to the spectral problem for this equation will not be found with these methods. However, we can apply some assumptions to the system that can reduce this to obtain an approximation of the size of the spectrum. In particular, we wish to find the approximate size of the spectral gap, since this governs the expected time to consensus. More detailed solutions such as the expected time spent per macrostate will not be found.

If we take the single step propagator to continuous time, it is known that the system approaches equilibrium when $n_A^{(1)}/N_1 \sim n_A^{(2)}/N_2$ [9]. Furthermore, the time to reach this equilibrium state is negligible compared to the time to reach consensus. Along this line, diffusion governs the motion of the macrostate of the system instead of drift. The study of the behavior of the probability distribution is most valuable when drift can be neglected. As such, we will make this assumption in equation (46) to obtain

$$yuQ_{yu}^{(m)} \approx xvQ_{xv}^{(m)}. \tag{47}$$

With this simplification, there are three cases need to be accounted for:

1. $yu \approx xv \longrightarrow Q_{yu}^{(m)} \approx Q_{xv}^{(m)}$
2. $yu \ll xv \longrightarrow Q_{xv}^{(m)} \ll Q_{yu}^{(m)}$
3. $xv \ll yu \longrightarrow Q_{yu}^{(m)} \ll Q_{xv}^{(m)}$

Consider the first case. Since $Q_{yu}^{(m)} \approx Q_{xv}^{(m)}$, we can combine the two terms together to obtain

$$\frac{(u-v)(x-y)}{N_1N_2} Q_{yu}^{(m)} \approx \Delta_{+m} Q^{(m)}. \tag{48}$$

Letting $G(x,y,u,v) = \sum_{ij} c_{ij} x^i y^{N_1-i} u^j v^{N_2-j}$, the spectral problem is given by

$$(u-v)(x-y)G_{yu} \approx N_1N_2(\lambda-1)G. \tag{49}$$

Let $s = u - v$, $r = x - y$, and $H(r,y,s,v) = G(x,y,u,v)$. As with the complete graph, this is a linear transformation, so we expect solutions of the same form: $H(r,y,s,v) = \sum_{ij} b_{ij} r^i y^{N_1-i} s^j v^{N_2-j}$. The differential equation becomes

$$rs(H_{sy} - H_{rs}) \approx N_1N_2(\lambda-1)H. \tag{50}$$

The corresponding difference equation for the coefficients, b_{ij}, is

$$j(N_1-i+1)b_{i-1,j} - ijb_{ij} \approx N_1N_2(\lambda-1)b_{ij}. \tag{51}$$

With similar arguments for the complete graph, we require a singularity in this difference equation so that the solution is not trivial. Therefore, we obtain an approximate spectrum for the diffusive scale of the two dimensional random walk:

$$\lambda_{ij} \approx 1 - \frac{ij}{N_1N_2}. \tag{52}$$

Note that in procedure, the same spectrum would be recovered if one were to use G_{xv} instead of G_{yu} in equation (49). This does not produce the spectrum for the complete bipartite graph, but rather an approximate form when the walk is dominated by diffusion.

Now let us consider cases 2 and 3 together. Without loss of generality, results from case 3 can be recovered by case 2 through interchanging $u \leftrightarrow x$, $y \leftrightarrow v$, and $N_1 \leftrightarrow N_2$. Physically, this corresponds to interchanging the labels on the two groups of nodes. Taking case 2 as the archetype, drop small terms to simplify the propagator to

$$\frac{ux}{NN_2}Q_{yu}^{(m)}+\frac{yv}{NN_1}Q_{yu}^{(m)}-\frac{2yu}{N_1N_2}Q_{yu}^{(m)}\approx\Delta_{+m}Q^{(m)}. \quad (53)$$

There is no need to change variables to solve the corresponding spectral problem since the corresponding finite difference equation explicitly determines c_{ij}. Therefore, the approximate spectrum for this case is found to be

$$\lambda_{ij}\approx 1-\frac{2ij}{N_1N_2}. \quad (54)$$

The two forms of the approximate spectrum found in equations (52) and (54) are very similar. The discrepancy between the two is due to the errors invoked by making the necessary approximations. To find the approximate eigenvalues, we required that the sizes of both sides of the spectral problem should be comparable. Since we used scaling arguments to find the approximate spectrum, there will be no asymptotic discrepancy in the spectral gap through this procedure. These results indicate that the spectral gap is $1-\lambda_2=O\left(\frac{1}{N_1N_2}\right)$.

We can repeat the calculation in section IV to find a bound on consensus times for the complete bipartite graph. Following this procedure, we find that the expected consensus time is

$$E[T]=O(N_1N_2). \quad (55)$$

This is consistent with continuous time analysis, which shows that the expected time to consensus is $E[T]=4N_1N_2\left[(1-\omega)\ln\frac{1}{1-\omega}+\omega\ln\frac{1}{\omega}\right]$, where ω is the degree weighted mean of microstates [9]. As before, this bound is valid for all initial probability distributions. However, without detailed information about the eigenvectors, we cannot extract more detailed information about the propagator than the bound on consensus.

VI. CONCLUSIONS

We have successfully derived exact solutions to the voter model on the complete graph. In particular, the solution to the spectral problem can be found exactly, which allows us to find the time m probability distribution of the model. Also, knowing the eigenvalues for the discrete matrix problem, we found the solution to the corresponding differential eigenvalue problem. The solutions of which are found to be hypergeometric functions that have terminating series expressions. Exact formulae for the expected time to consensus and the expected frequency of each macrostate prior to consensus are also given, though there are other quantities of interest that can be found using the theory given above.

The means by which these exact solutions are found also provide mathematical insight. Since the procedures are easily generalizable, there is great potential for applying these techniques to other problems. This includes extensions of the voter model, imposing it upon different graphs, and other models entirely. There are other urn models of interest in the context of statistical physics and network science that may be studied using these techniques. The procedure for determining the single step propagator can be generalized to finding probabilities of microstates instead of macrostates, which may be insightful for a network with specific topological features. With such a formulation, a general network can be studied in detail. Rigorous treatment of these claims are potential areas of future work.

[1] T. Liggett, *Stochastic Interacting Systems: Contact, Voter, and Exclusion Processes* (Springer-Verlag, New York, 1999).
[2] P. Clifford and A. Sudbury, Biometrika **60 (3)**, 581C588 (1973).
[3] F. Vazquez and V. M. Eguíluz, New Journal of Physics **10** (2008).
[4] E. Ben-Naim, L. Frachebourg, and P. Krapivsky, Phys. Rev. E. **53**, 3078 (1996).
[5] D. Vilone and C. Castellano, Phys. Rev. E. **69**, 016109 (2004).
[6] L. Frachebourg and P. Krapivsky, Phys. Rev. E. **53**, R3009(R) (1996).
[7] J. Cox, The Annals of Probability **17**, 1333 (1989).
[8] T. M. Liggett, *Interacting Particle Systems* (Springer, 2005).
[9] V. Sood and S. Redner, Phys. Rev. Lett. **94**, 178701 (2005).
[10] N. Masuda, N. Gilbert, and S. Redner, Phys. Rev. E **82**, 010103(R) (2010).
[11] N. Masuda, Phys. Rev. E **88**, 052803 (2013).
[12] F. Caccioli, L. Dall'Asta, T. Galla, and T. Rogers, Phys. Rev. E **87**, 052114 (2013).
[13] T. Rogers and T. Gross, Phys. Rev. E **88**, 030102(R) (2013).
[14] A. Baronchelli, C. Castellano, and R. Pastor-Satorras, Phys. Rev. E **83**, 066117 (2011).
[15] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning* (Addison-Wesley Professional, Boston, 1989).
[16] J. H. Holland, *Adaptation in Natural and Artificial Systems* (MIT Press, 1992).
[17] M. D. Vose, *The Simple Genetic Algorithm, foundations and Theory* (MIT Press, 1999).
[18] S. Boyd and P. Diaconis, SIAM Review **46 (4)**, 667 (2004).
[19] C. Cooper, R. Elsässer, and T. Radzik, SIAM J. Discrete Math **27(4)**, 1748 (2013).
[20] P. Ehrenfest and T. Ehrenfest, Physik Z. **8**, 311 (1907).
[21] M. Kac, The American Mathematical Monthly , 375

(1947).
[22] W. Zhang, C. Lim, S. Sreenivasan, J. Xie, B. Szymanski, and G. Korniss, Chaos **21**, 025115 (2011).
[23] M. E. Yildiz, R. Pagliari, A. Ozdaglar, and A. Scaglione, Information Theory and Applications Workshop (2010).
[24] K. Suchecki, V. M. Eguíluz, and M. S. Miguel, Phys. Rev. E **72**, 036132 (2005).
[25] M. Newman, S. Strogatz, and D. Watts, Phys. Rev. E. **64**, 026118 (2001).
[26] M. Newman, Phys. Rev. E **66**, 016128 (2002).
[27] E. A. Bender and S. G. Williamson, *Foundations of Combinatorics with Applications* (Dover, New York, 2006).
[28] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (Dover, New York, 1964).
[29] I. H. Herron and M. R. Foster, *Partial Differential Equations in Fluid Dynamics* (Cambridge University Press, New York, 2008).