



MONTCLAIR STATE
UNIVERSITY

Montclair State University
**Montclair State University Digital
Commons**

Theses, Dissertations and Culminating Projects

5-2018

The Dispersion Process for Particles on Graphs

Adam Cartisano
Montclair State University

Follow this and additional works at: <https://digitalcommons.montclair.edu/etd>



Part of the [Mathematics Commons](#)

Recommended Citation

Cartisano, Adam, "The Dispersion Process for Particles on Graphs" (2018). *Theses, Dissertations and Culminating Projects*. 125.

<https://digitalcommons.montclair.edu/etd/125>

This Thesis is brought to you for free and open access by Montclair State University Digital Commons. It has been accepted for inclusion in Theses, Dissertations and Culminating Projects by an authorized administrator of Montclair State University Digital Commons. For more information, please contact digitalcommons@montclair.edu.

THE DISPERSION PROCESS FOR PARTICLES ON GRAPHS

Abstract

In this thesis, we study a process called Dispersion, in which M particles are dispersed among the vertices of a graph G . All particles initially occupy a single vertex called the origin vertex. At each discrete time step, all particles which share a vertex with at least one other, move to a randomly (though not necessarily uniformly) chosen neighbor of the currently occupied vertex. The process ends when each vertex is occupied by at most one particle.

We will explore various aspects of the Dispersion process. One of these is the expected time to completion, $\mathbf{E}[\mathbf{T}_{Disp}]$ for 3 particles on an n -cycle. Another point of analysis will be the differences in the behavior of particles on even-length cycles vs. odd-length cycles.

MONTCLAIR STATE UNIVERSITY

The Dispersion Process for Particles on Graphs

by

Adam Cartisano

A Master's Thesis Submitted to the Faculty of

Montclair State University

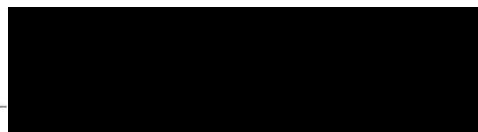
In Partial Fulfillment of the Requirements

For the Degree of

Master of Science

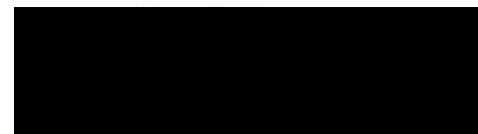
May 2018

College of Science and Mathematics



Dr. Deepak Bal, Thesis Advisor

Department of Mathematics



Thesis Committee:

Dr. Aihua Li, Committee Member



Dr. Jonathan Cutler, Committee Mem-

ber

THE DISPERSION PROCESS FOR PARTICLES ON GRAPHS

A THESIS PROJECT

Submitted in partial fulfillment of the requirements
for the degree of Master of Science

by

Adam Cartisano

Montclair State University

Montclair, NJ

May 2018

Copyright © 2018 by *Adam Cartisano*. All rights reserved.

Contents

1	Introduction	1
1.1	Current Results	1
1.2	New Results	3
2	Definitions and Notation	5
2.1	General Definitions	5
2.2	Definitions on C_n	6
3	Markov Chains	7
4	State Diagrams and Reachable States	9
4.1	State Diagrams	9
4.2	Comparison of Ω' and Λ'	11
5	Closed-form Results	13
5.1	Example of Technique	13
6	Three Particles on C_n	20
7	The Parity of Vertices	28
7.1	Proof for Lemma 1	28
7.2	Proof for Proposition 1	30
7.3	Proof for Theorem 5	31
7.4	Proof for Corollary 1	39
8	Future Endeavors and Open Questions	40

1 Introduction

1.1 Current Results

Imagine you are in a boat with your friends moving down a river. You come around a bend and you see that someone has laid stepping stones all across the river, and your boat crashes into one of them. You and all your friends are now stacked up on one stone, and in order to avoid falling into the river, you need to move. You each decide to flip a coin and move right or left depending on whether you see heads or tails respectively (please ignore the fact that you've all probably fallen in by now). Everyone moves at the same time, and then you repeat the following process: everyone who is still sharing a stone flips a coin and moves left or right according to the outcome. The process ends when everyone has their own stone and you can wait until help arrives.

This is the fundamental idea behind the process known as *Dispersion*. Formally, consider M particles initially placed on a distinguished vertex (called the origin vertex) of a graph G . The process is synchronous and proceeds in discrete steps. Whenever two or more particles occupy the same vertex at some step, they move independently to a random neighbor. If only a single particle occupies a vertex, it stays there until another particle arrives [1]. The process ends when each vertex is occupied by at most one particle.

In general, a *dispersion process* is one in which a collection of $M \in \mathbb{N}$ identical particles are located at an initial vertex (called the origin vertex) of a graph G . The particles move about the graph in a distributed fashion until no more than one particle occupies any given vertex. When this occurs, we say that the particles have dispersed.

Dispersion is symmetric across all particles, in that there is no prioritization, communication, or any other form of asymmetry [1]. This process can be an interesting way to think about how a collection of positively charged physical particles would move if placed close enough together, or how robots could distribute themselves in an area without the knowledge of how the other robots would move. The river example is actually a specific case where G

is the infinite 2-regular tree.

Another example of this process is in terms of non-aggressive swarm behavior. Take for example a swarm of bees trying to pollinate a rose bush. In this case, the rose bush would be the graph G with the flowers as the vertices $V(G)$, and M bees would be the particles. The rule is that if two bees land on the same flower, each one automatically defers to the others, and leaves to find a new flower.

Dispersion was very recently introduced by Cooper, McDowell, Radzik, Rivera and Shraga in [1]. We are interested in various points of analysis for this process. Foremost among them, given a graph G and an initial number of particles, M , along with an origin vertex, we would like to find the number of time steps to achieve dispersion, \mathbf{T}_{Disp} . We are also interested in the farthest distance of a particle from the origin vertex \mathbf{D}_{Disp} after the particles have dispersed, in the case where G is the infinite path. We use $V(G)$ to denote the vertex set of G .

In [1], Cooper et al. proved the following result for complete graphs and stars.

Theorem 1 (Cooper et al. [1]). *For the complete graph K_n , and the star S_n the following hold for any constant $\delta > 0$*

1. *If the number of particles M satisfies $M/n \leq (1/2)(1 - \delta)$, then with probability $1 - O(1/n)$, the dispersion process terminates in $\mathbf{T}_{Disp} = O(\log n)$ steps.*
2. *If the number of particles M satisfies $M/n \geq (1/2)(1 + \delta)$, then there is a constant $c = c(\delta) > 0$ s.t. the probability that $\mathbf{T}_{Disp} \leq e^{cn}$ is less than e^{-cn} .*

This theorem states that if the ratio of the particles M to the number of vertices $n = |V(G)|$ on either K_n or S_n is less than $1/2$, the number of time steps that the process takes to reach dispersion (\mathbf{T}_{Disp}) is logarithmic. But if that ratio is greater than $1/2$, \mathbf{T}_{Disp} is exponential. This shows a drastic change in behavior at the ratio of $1/2$ particles to vertices.

Cooper et al. define a *threshold* for a graph G on n vertices as a value of M for which on G , $M(1 + \epsilon)$ particles will require $\Omega(e^n)$ iterations to reach dispersion, while $M(1 - \epsilon)$

particles will disperse in polynomially many iterations.

The next theorem concerns the case where G is the infinite (or at least a sufficiently long) path, rather than stars or complete graphs. We say a sequence of events A_n holds *with high probability* or *w.h.p.* if $\mathbb{P}[A_n] \rightarrow 1$ as $n \rightarrow \infty$.

Theorem 2 (Cooper et al. [1]). *For a sufficiently long path, and M particles initially placed at the origin vertex, the following holds w.h.p. (as $M \rightarrow \infty$) for any $\epsilon > 0$. When the dispersion process terminates, the maximum distance \mathbf{D}_{Disp} any particle is from the origin is bounded by*

$$\lfloor M/2 \rfloor \leq \mathbf{D}_{Disp} \leq 4(1 + \epsilon)M \log M, \quad (1)$$

and $\mathbf{T}_{Disp} = O(M^3 \log M)$.

Cooper et al. essentially proved an upper bound on the distance of the farthest particle from the origin vertex. The lower bound is trivial for the following reason. Any smaller value would indicate that the process has not finished, since more than one particle would occupy the same vertex by Pigeonhole Principle. In Section 7, we actually show that this lower bound can be improved to $\lceil \frac{M}{2} \rceil$. Very recently, in [2], Frieze and Pegden improved the upper bound on \mathbf{D}_{Disp} from $O(M \log M)$ to $O(M)$. This answered one of the open problems in [1].

1.2 New Results

In this thesis, we will prove the following results. First, we will draw various conclusions regarding the dispersion of exactly 3 particles on the n -cycle, and on the infinite path. Beginning with our first main theorem, we have

Theorem 3. *Let $k = \lfloor \frac{n}{2} - 1 \rfloor$. For the unbiased dispersion of 3 particles on C_n with $n > 3$, $A_n = \frac{1}{3} \left(10 + \frac{4}{3 - \alpha_{k-1}} \right)$, where $\alpha_k = \frac{1}{4 - \alpha_{k-1}}$, and*

1. if n is even, then $\alpha_1 = \frac{1}{2}$, and
2. if n is odd, then $\alpha_1 = \frac{1}{3}$.

Theorem 4. *For the unbiased dispersion of 3 particles on the infinite path, P , the expected time to dispersion is $\frac{1}{3} \left(10 + \frac{4}{1 + \sqrt{3}} \right)$.*

Another topic of study is the difference between odd and even cycles. We proved quite a few results within this topic, since this is the largest portion of our study. First we prove a lemma which is interesting on its own, but also is our main tool for analysis of odd vs even cycle behavior.

Lemma 1. *Let $M > 3 \in \mathbb{N}$ particles be dispersed on the graph C_n . If $n \in \mathbb{N}$ is even, all unhappy particles have the same parity.*

Next, we will show that the lower bound for \mathbf{D}_{Disp} in theorem 2 is actually $\lceil \frac{M}{2} \rceil$.

Proposition 1. *For the dispersion of M particles on the infinite path, \mathbf{D}_{Disp} , the greatest distance from the origin vertex to any particle, is at least $\lceil \frac{M}{2} \rceil$.*

We will then turn our attention to cycles with an odd number of vertices, and prove our next main theorem, which basically states that if we have enough particles, and we are using the dispersion process on an odd-length cycle with more than 3 vertices, then every configuration of particles is reachable.

Theorem 5. *All possible arrangements of M particles on C_n are reachable through dispersion if $n > 3$ is an odd integer and $M > \frac{3}{5}n$. Moreover, this bound cannot be improved.*

This result gives rise to the following corollary when paired with the previous lemma.

Corollary 1. *Let $n > 3 \in \mathbb{N}$. The dispersion process for exactly n particles on C_n terminates in a finite number of steps if and only if n is odd.*

This will be our final result, and the rest of the paper will proceed as follows. In Section 2, we introduce the definitions and notation required for the remainder of the thesis. In Section 3 we introduce the concepts from the theory of Markov Chains which will be used along with an example. In Section 4, we describe the state diagrams and corresponding transition matrices of the Markov Chains which appear in dispersion. In Section ??, we apply the techniques described in Section 3 to the chains arising from dispersion. In 6 we prove our first main theorem regarding unbiased dispersion of three particles on an n -cycle. In Section 7 we prove our second main theorem which compares the number of reachable states in dispersion to the total number of states,. Finally, in Section 8 we discuss future work and open problems.

2 Definitions and Notation

2.1 General Definitions

Consider the system modeled by the standard dispersion process as described in [1] on a graph G on n vertices, with M particles. Cooper et al. define a *state* of the process simply as a configuration of M particles on the n vertices of G . The initial state of a dispersion system is the one in which all M particles occupy a single vertex. They also define a particle to be *unhappy* if at time t , another particle occupies the same vertex as itself, and they say that the particle is *happy* otherwise. We will actually redefine these terms in the last section, where we describe an open problem regarding a generalization of Theorem 1.

We define the set of particles which occupy a single vertex the *stack* of particles on that vertex at time t . An unhappy stack then is a set of particles which all occupy the same vertex at a given time step t . We say that a state is *reachable* from another state if there is a way to distribute all unhappy particles in the former state to adjacent vertices to obtain the latter state. We can then say that a state is reachable by the dispersion process if there

exists a sequence of reachable states leading to it from the initial state. Lastly, the dispersion process can be either biased or unbiased, the latter only if the adjacent vertex to which an unhappy particle moves is chosen uniformly at random.

2.2 Definitions on C_n

By convention, since we are mostly going to consider cases where G will be C_n , we will use the following notation to denote a state of the system, and we will specify if other notation is implemented.

Given n and M , let $\Omega = \Omega(M, n)$ be the set of all strings of length n whose entries are over the alphabet $\{0, 1, 2, \dots, M\}$ and sum to M . Next, let $x, y \in \Omega$. Say $x \sim y$ if x is a cyclic permutation of y . Then we define $\Omega' = \Omega'(M, n)$ to be the set of all equivalence classes under \sim .

In other words, Ω represents the set of all configurations of M particles on a *labeled* n -cycle. Ω' on the other hand is the set of all rotationally asymmetric configurations of M particles on C_n . Note that $|\Omega(M, n)|$ can be counted using “Pirates and Gold”, so

$$|\Omega(M, n)| = \binom{M+n-1}{n-1}.$$

It can also be shown that

$$|\Omega'(M, n)| = \frac{1}{n} * \sum_{d|(M+n, n)} \varphi(d) \binom{\frac{M+n}{d} - 1}{\frac{n}{d} - 1},$$

where $(M+n, n)$ is the gcd of $M+n$ and n , and φ represents Euler’s Totient Function.

Lastly, when G is the labeled C_n or the labeled infinite path P , we would like to define the parity of a particle, as well as that of a stack of particles. Consider an element of Ω whose

positions are denoted by v_1, v_2, \dots, v_n . Let x represent a particle or a stack on $v_i, i \in [n]$ at time t . Then the *parity* of x , $Par(x)$ is

$$Par(x) = \begin{cases} 0 & \text{if } i \text{ is even} \\ 1 & \text{if } i \text{ is odd} \end{cases}$$

In other words, at time t , a particle (or stack of particles) has the same parity as the labeled vertex it occupies at that time.

3 Markov Chains

In this section we will introduce the concept of absorbing Markov chains and relevant results which we will make use of later in the thesis. A *Markov chain* is a stochastic process consisting of discrete states, X_1, X_2, \dots, X_A , which takes place over sequential, discrete time steps. At each time step, the system transitions from one state to another with a certain probability. An *absorbing* Markov chain is one in which there exists at least one state whose probability of transitioning to itself exactly 1, and it is possible to reach such a state from any other state in the system. Such a state, which can never be left once entered, is called an *absorbing* state, and any other state is called a *transient* state.

The relationships between the states can be described in a matrix, called the probability transition matrix, or the adjacency matrix. The ij -th entry of the transition matrix describes the probability of transitioning to state j from state i . It is common practice to group all of the absorbing states at the bottom right corner of the matrix. Since each absorbing state transitions to itself with probability 1, this causes the identity matrix of size equal to the number of absorbing states to occupy that corner of the matrix.

$$P = \left(\begin{array}{c|c} Q & R \\ \hline 0 & I \end{array} \right)$$

In this form, Q represents the probability of one transient state transitioning to another, R is the probability of a transient state transitioning to an absorbing state, 0 is a zero matrix, and I is the identity. Note that if P_{ij} represents the probability of transitioning from the i -th state to the j -th state, then P_{ij}^2 is the probability of transitioning from i to j in exactly 2 time steps. More generally, P^k is the probability transition matrix after k time steps.

One of the fundamental properties of absorbing Markov Chains is that the process must eventually end. Consider the sub-matrix Q , the probability transition matrix of transient states transitioning between one another only. It is clear that $\lim_{k \rightarrow \infty} Q^k = \vec{0}$. Linear algebra gives us the following result regarding the geometric series,

$$N = (I - Q)^{-1} = I + Q + Q^2 + Q^3 + \dots$$

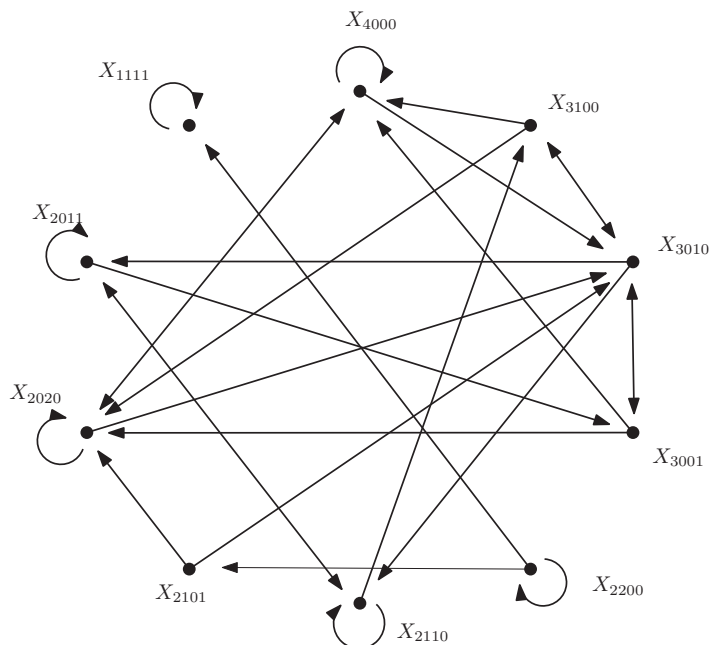
N is often called the Fundamental Matrix of the absorbing Markov Chain. The ij -th entry of N can be interpreted as the expected number of visits to state j given that the initial state of the system was state i . This matrix has a multitude of useful properties, but specifically we want to build on the one regarding the expected number of visits to a state.

If we take $N\vec{1}$ where $\vec{1}$ is a vector of 1's with length equal to the dimensions of N , we obtain a column vector of the row sums of N . The i -th entry of $N\vec{1}$ is exactly the expected number of time steps required for the system to reach an absorbing state, given that the system started in state i . Note that multiplying a matrix by a vector of 1's is equivalent to summing the rows of the matrix. Thus we have all the tools we need to make the connection from Markov Chains back to particle dispersion.

4 State Diagrams and Reachable States

4.1 State Diagrams

It is possible to encompass all of the information regarding the relationships between states of the system using a directed graph (with loops) with weighted edges, called a *state diagram*. We first define $Disp(M, n)$ as the directed graph with vertex set Ω' , and directed edges representing the possibility of the former state changing into the latter state. For example, $Disp(4, 4)$ gives us the following state diagram. We label the states in the diagram based on the state which that vertex represents. For example, $X_{(4000)}$ represents the initial state, where all 4 particles occupy a single vertex. Also, we temporarily omit the transition probabilities while we describe the underlying graphs themselves.

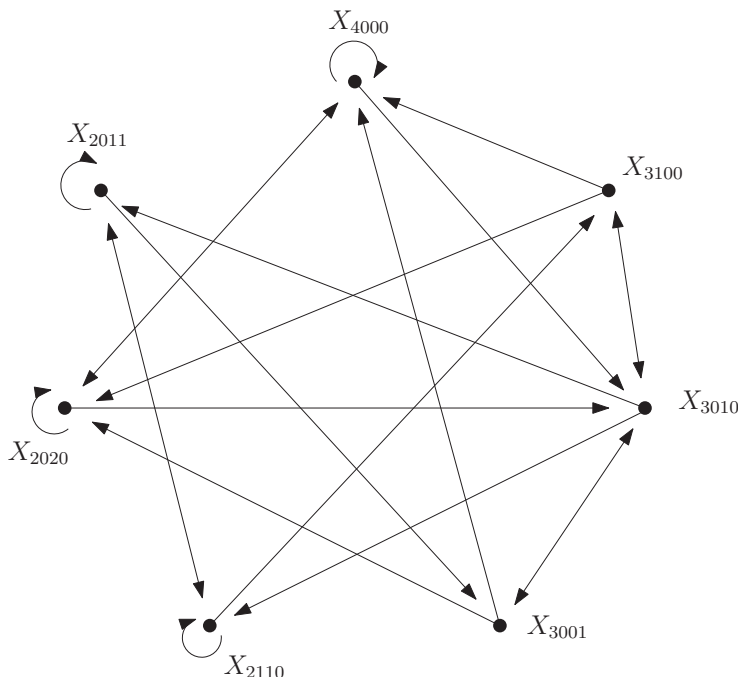


Consider all the edges coming in and out of $X_{(4000)}$. There is a loop at that vertex because there is a chance that all the particles in the stack move left or right as a unit, which would send the system to a rotationally symmetric state, which is equivalent to $X_{(4000)}$ in Ω' . It has a one-directional arrow going to $X_{(3010)}$, because the pile could split into one happy particle and three unhappy particles. The only other way the particles can split is into two pairs,

which gives $X_{(2020)}$. This edge goes both directions, because there is a chance that that state will return to the original, which is impossible for $X_{(3010)}$. Lastly we see two incoming arrows from $X_{(3100)}$ and $X_{(3001)}$, which represent the chance that the stacks of three particles move onto the happy particle, forming a stack of 4 particles.

Now take a look at the states $X_{(2200)}$, $X_{(2101)}$, and $X_{(1111)}$. Notice that it is impossible to enter any of these three states by particle dispersion, since dispersion requires that every particle starts on the origin vertex, in the state $X_{(4000)}$. There is no directed path from $X_{(4000)}$ to any of the other three, so the dispersion process will never see those states. Because of this, it is at times completely unnecessary to include them in $Disp(M, n)$. If we “throw away” these states (omit them from the graph), we can obtain a much more insightful description of the relationships between states specifically included in the dispersion process.

Thus the following definition. Given M and n , let $\Lambda' = \Lambda'(M, n)$ be the set of all states which can be reached by a directed path from the initial state in Ω' . We refer to Λ' as the set of *reachable states*. Recall that $Disp(M, n)$ is the graph of all possible arrangements of M particles on C_n . We now define a new directed subgraph (still allowing loops) called $Disp(M, n)'$ as the state diagram whose vertex set is Λ' . For example $Disp(4, 4)'$ looks like



4.2 Comparison of Ω' and Λ'

We developed algorithms which could accurately generate the elements of Ω' and Λ' . While we will not discuss them here, we will pause and observe some data collected via these algorithms. The data we've collected specifically is the number of configurations of M particles on C_n vs the number of reachable states for the same particles and cycles.

For the reader's sake, we show both tables on the same page, to simplify comparison. The first table shows the number of all possible states (equivalent to necklaces of length n with characters summing to M , also the sizes of $\Omega'(M, n)$). The second table shows the number of states in Ω reachable by dispersion (equivalent to the sizes of $\Lambda'(M, n)$).

Possible Arrangements, $\Omega'(M, n)$:

		Cycle Size n									
		1	2	3	4	5	6	7	8	9	10
Number of Particles M	1	1	1	1	1	1	1	1	1	1	1
	2	1	2	2	3	3	4	4	5	5	6
	3	1	2	4	5	7	10	12	15	19	22
	4	1	3	5	10	14	22	30	43	55	73
	5	1	3	7	14	26	42	66	99	143	201
	6	1	4	10	22	42	80	132	217	335	504
	7	1	4	12	30	66	132	246	429	715	1144
	8	1	5	15	43	99	217	429	810	1430	2438
	9	1	5	19	55	143	335	715	1430	2704	4862
	10	1	6	22	73	201	504	1144	2438	4862	9252

Reachable States, $\Lambda'(M, n)$:

		Cycle Size n									
		1	2	3	4	5	6	7	8	9	10
Number of Particles M	1	1	1	1	1	1	1	1	1	1	1
	2	1	1	2	2	2	2	2	2	2	2
	3	1	1	3	5	6	9	10	12	14	16
	4	1	1	5	7	14	18	30	37	52	62
	5	1	1	7	8	26	29	66	76	142	163
	6	1	1	10	10	42	40	132	131	335	337
	7	1	1	12	11	66	51	246	194	715	598
	8	1	1	15	13	99	64	429	276	1430	956
	9	1	1	19	14	143	79	715	368	2704	1433
	10	1	1	22	16	201	94	1144	487	4862	2057

One very interesting observation here is that it seems as though for every odd cycle with

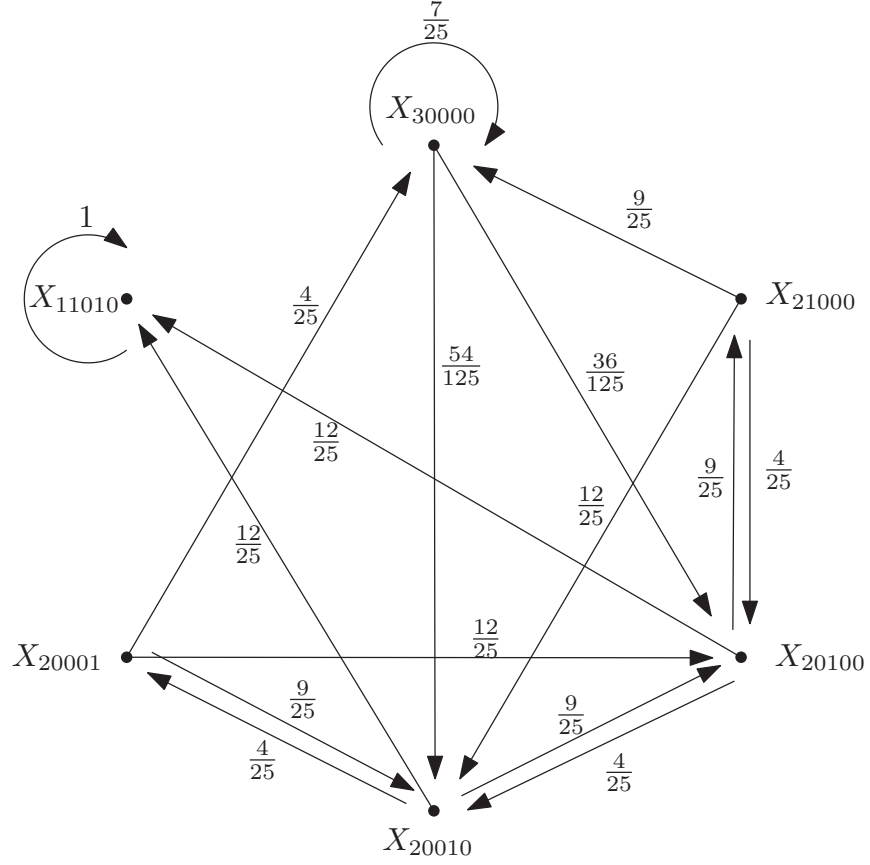
$M > \frac{3}{5}n$ and $n > 3$, every single arrangement of the M particles is reachable by the dispersion process, but for even cycles, the number of reachable states is significantly smaller than the total number of asymmetric particle configurations. This indicates that many of the arrangements of particles are unreachable in the even-cycle cases. Indeed, we confirm these observations in Section 7. This also gives rise to the other interesting conjectures which we will discuss. We will eventually use this discrepancy to improve the lower bound on \mathbf{D}_{Disp} in Theorem 2 from $\lfloor \frac{M}{2} \rfloor$ to $\lceil \frac{M}{2} \rceil$, and show that given exactly n particles on C_n , the dispersion process will terminate if and only if n is odd.

5 Closed-form Results

We now tie together the theory of Markov Chains and the transition matrices we've built. We saw at the end of Section 4.2 that we have the ability to generate the probability matrix for any M, n , provided that these values aren't too large. We also introduce a third variable p , which represents the probability of a particle moving clockwise around the cycle (or moving right on the path P). In both [1] and [2], the authors only consider unbiased dispersion of particles. This generalization does not break the symmetry of the particles, since each particle still moves independently of the others.

5.1 Example of Technique

We will give a demonstration of the entire process for the dispersion of exactly 3 particles on C_5 , where the particles have a probability bias of .6 for moving in the clockwise direction. First, we create the state diagram (now including the probabilities with the directed edges).



From this state diagram, we can extract the following transition matrix,

$$Disp(3, 5) = P = \begin{pmatrix} \frac{7}{25} & 0 & \frac{36}{125} & \frac{54}{125} & 0 & 0 \\ \frac{9}{25} & 0 & \frac{4}{25} & \frac{12}{25} & 0 & 0 \\ 0 & \frac{9}{25} & 0 & \frac{4}{25} & 0 & \frac{12}{25} \\ 0 & 0 & \frac{9}{25} & 0 & \frac{4}{25} & \frac{12}{25} \\ \frac{4}{25} & 0 & \frac{12}{25} & \frac{9}{25} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Continuing in this case, we extract/derive Q, R, N and $N\vec{1}$ as follows.

$$Q = \begin{pmatrix} \frac{7}{25} & 0 & \frac{36}{125} & \frac{54}{125} & 0 \\ \frac{9}{25} & 0 & \frac{4}{25} & \frac{12}{25} & 0 \\ 0 & \frac{9}{25} & 0 & \frac{4}{25} & 0 \\ 0 & 0 & \frac{9}{25} & 0 & \frac{4}{25} \\ \frac{4}{25} & 0 & \frac{12}{25} & \frac{9}{25} & 0 \end{pmatrix} \quad R = \begin{pmatrix} 0 \\ 0 \\ \frac{12}{25} \\ \frac{12}{25} \\ 0 \end{pmatrix}$$

$$N = (I - Q)^{-1} = \begin{pmatrix} -\frac{18}{25} & 0 & -\frac{36}{125} & -\frac{54}{125} & 0 \\ -\frac{9}{25} & 1 & -\frac{4}{25} & -\frac{12}{25} & 0 \\ 0 & -\frac{9}{25} & 1 & -\frac{4}{25} & 0 \\ 0 & 0 & -\frac{9}{25} & 1 & -\frac{4}{25} \\ -\frac{4}{25} & 0 & -\frac{12}{25} & -\frac{9}{25} & 1 \end{pmatrix}^{-1} \approx \begin{pmatrix} 1.6067 & 0.3583 & 0.9953 & 1.0880 & 0.1741 \\ 0.7075 & 1.3325 & 0.9235 & 1.1598 & 0.1856 \\ 0.2826 & 0.5198 & 1.4439 & 0.6395 & 0.1023 \\ 0.1746 & 0.2507 & 0.6963 & 1.3871 & 0.2219 \\ 0.4556 & 0.3971 & 1.1030 & 0.9804 & 1.1569 \end{pmatrix}$$

$$N\vec{1} = \begin{pmatrix} 1.6067 & 0.3583 & 0.9953 & 1.0880 & 0.1741 \\ 0.7075 & 1.3325 & 0.9235 & 1.1598 & 0.1856 \\ 0.2826 & 0.5198 & 1.4439 & 0.6395 & 0.1023 \\ 0.1746 & 0.2507 & 0.6963 & 1.3871 & 0.2219 \\ 0.4556 & 0.3971 & 1.1030 & 0.9804 & 1.1569 \end{pmatrix} * \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 4.2225 \\ 4.3089 \\ 2.9881 \\ 2.7306 \\ 4.0929 \end{pmatrix}$$

Notice that this technique actually gives us more information than we were originally seeking. The expectation of \mathbf{T}_{Disp} is given by the first entry of $N\vec{1}$, but the other entries

of $N\vec{1}$ hold the expected time to completion for all the other states in $Disp(3, 5)$. We can conclude from this information that the dispersion process for 3 particles on C_n will take an expected number of steps approximately equal to 4.2225.

Given that the MATLAB language allows for symbolic computation, we are actually able to extract even more information for small cases. One way we were able to do this was to calculate the fundamental matrix for some systems in terms of the general probability p . This gives us the opportunity to generalize $Disp(M, n) \forall p \in (0, 1)$. It was by this method that we obtained the matrices for $Disp(4, 4)$ and $Disp(3, 5)$, as opposed to by hand.

Consider $Disp(3, 6)$ for general $p \in (0, 1)$. We calculate the symbolic matrix to appear as follows, using the symbol “p” for our variable bias.

$$Disp(3, 6) = P = \left(\begin{array}{cccccc|ccc} p^3 + q^3 & 0 & 3pq^2 & 0 & 3p^2q & 0 & 0 & 0 & 0 \\ p^2 & 0 & q^2 & 0 & 2pq & 0 & 0 & 0 & 0 \\ 0 & p^2 & 0 & q^2 & 0 & 0 & 0 & 2pq & 0 \\ 0 & 0 & p^2 & 0 & q^2 & 0 & 0 & 0 & 2pq \\ 0 & 0 & 0 & p^2 & 0 & q^2 & 2pq & 0 & 0 \\ q^2 & 0 & 2pq & 0 & p^2 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right)$$

Following the above process, we can see that

$$Q = \begin{pmatrix} p^3 + q^3 & 0 & 3pq^2 & 0 & 3p^2q & 0 \\ p^2 & 0 & q^2 & 0 & 2pq & 0 \\ 0 & p^2 & 0 & q^2 & 0 & 0 \\ 0 & 0 & p^2 & 0 & q^2 & 0 \\ 0 & 0 & 0 & p^2 & 0 & q^2 \\ q^2 & 0 & 2pq & 0 & p^2 & 0 \end{pmatrix}$$

At this point we need to take $N = (I - Q)^{-1}$. Unfortunately, every entry is close to a rational polynomial, with the degree of both the numerator and denominator around 8 and 10 respectively. This would be a waste to include here, so instead we choose to list only the polynomials in the first row, since these are the elements we sum to obtain \mathbf{T}_{Disp} . For the sake of notation, assume we are starting in the first state. Let $\mathbf{E}[X_i]$ be the expected number of visits to state i . Then, in order from left to right in the first row of $N = (I - Q)^{-1}$, we have:

$$\mathbf{E}[X_{(300000)}] = \frac{1 - 2p + 10p^2 - 34p^3 + 65p^4 - 66p^5 + 36p^6 - 12p^7 + 3p^8}{6p^2(1-p)^2(3 - 8p + 14p^2 - 15p^3 + 15p^4 - 9p^5 + 3p^6)}$$

$$\mathbf{E}[X_{(210000)}] = \frac{p - p^2 + p^6}{2(1-p)(14p^2 - 8p - 15p^3 + 15p^4 - 9p^5 + 3p^6 + 3)}$$

$$\mathbf{E}[X_{(201000)}] = \frac{p^5 - p + 1}{2(p - p^2)(14p^2 - 8p - 15p^3 + 15p^4 - 9p^5 + 3p^6 + 3)}$$

$$\mathbf{E}[X_{(200100)}] = \frac{4p^2 - 3p - 5p^3 + 10p^4 - 9p^5 + 3p^6 + 1}{2(p - p^2)(14p^2 - 8p - 15p^3 + 15p^4 - 9p^5 + 3p^6 + 3)}$$

$$\mathbf{E}[X_{(200010)}] = \frac{-4p + 10p^2 - 10p^3 + 5p^4 - p^5 + 1}{2(p - p^2)(14p^2 - 8p - 15p^3 + 15p^4 - 9p^5 + 3p^6 + 3)}$$

$$\mathbf{E}[X_{(200001)}] = \frac{14p^2 - 5p - 20p^3 + 15p^4 - 6p^5 + p^6 + 1}{2p(14p^2 - 8p - 15p^3 + 15p^4 - 9p^5 + 3p^6 + 3)}$$

A far more significant result comes when we sum these polynomials. While the resulting polynomial is no less atrocious to look at, it is a generalization of \mathbf{T}_{Disp} for 3 particles on C_6 .

$$\begin{aligned}\mathbf{T}_{Disp} &= \frac{-27p^8 + 108p^7 - 204p^6 + 234p^5 - 187p^4 + 110p^3 - 44p^2 + 10p + 1}{6p^2(1-p)^2(3p^6 - 9p^5 + 15p^4 - 15p^3 + 14p^2 - 8p + 3)} \\ &= \frac{1}{18(1-p)^2} + \frac{1}{18p^2} + \frac{22}{27(1-p)} + \frac{22}{27p} - \frac{129p^4 - 258p^3 + 288p^2 - 159p + 20}{54(3p^6 - 9p^5 + 15p^4 - 15p^3 + 14p^2 - 8p + 3)}\end{aligned}$$

Again, since we are limited by the matrix size, we were only able to obtain a small collection of these closed forms for \mathbf{T}_{Disp} . This includes $Disp(3, n, p)$ for $4 \leq n \leq 15$, $Disp(4, 5, p)$, and $Disp(4, 6, p)$. Beyond that would require more intensive computing resources or significantly improved algorithms. We will report these formulas in partial fractions form because they are more legible and some require more than one line of text.

We will use the natural notation $\mathbf{T}_{Disp}(M, n, p)$ to represent the expected time for M particles to disperse on C_n given the probability bias p of a particle moving clockwise. The following results are only for the dispersion of 3 particles.

$$\mathbf{T}_{Disp}(3, 4, p) = \frac{1}{6(1-p)^2} + \frac{1}{1-p} + \frac{1}{p} + \frac{1}{6p^2} - \frac{2}{3}$$

$$\mathbf{T}_{Disp}(3, 5, p) = \frac{1}{12(1-p)^2} + \frac{7}{8(1-p)} + \frac{1}{p^2 - p + 1} - \frac{21}{8(p^2 - p + 2)} + \frac{7}{8p} + \frac{1}{12p^2}$$

$$\mathbf{T}_{Disp}(3, 6, p) = \frac{1}{18(1-p)^2} + \frac{1}{18p^2} - \frac{129p^4 - 258p^3 + 288p^2 - 159p + 20}{54(3p^6 - 9p^5 + 15p^4 - 15p^3 + 14p^2 - 8p + 3)} + \frac{22}{27(1-p)} + \frac{22}{27p}$$

$$\mathbf{T}_{Disp}(3, 7, p) = \frac{1}{24(1-p)^2} + \frac{25}{32(1-p)} + \frac{1}{3(p^2 - p + 1)} + \frac{13}{72(2p^2 - 2p + 1)} + \frac{25}{32p} + \frac{1}{24p^2} - \frac{682p^2 - 682p + 589}{288(2p^4 - 4p^3 + 5p^2 - 3p + 4)}$$

$$\mathbf{T}_{Disp}(3, 8, p) =$$

$$\frac{-584p^8 + 2336p^7 - 5790p^6 + 9194p^5 - 9633p^4 + 6668p^3 - 2670p^2 + 479p + 76}{150(4p^{10} - 20p^9 + 61p^8 - 124p^7 + 187p^6 - 211p^5 + 188p^4 - 129p^3 + 68p^2 - 24p + 5)} + \frac{19}{25(1-p)} + \frac{1}{30(1-p)^2} + \frac{19}{25p} + \frac{1}{30p^2}$$

For unbiased cases, the results are much cleaner. Note that in the cases where $M = 3$, this is equivalent to substituting $p = \frac{1}{2}$ for p in the above formulas.

$$\mathbf{T}_{Disp} \left(M, n, \frac{1}{2} \right):$$

		Number of Particles M			
		3	4	5	6
Cycle Size n	4	$\frac{14}{3}$	∞		
	5	4	$\frac{1324}{147}$	$\frac{165244403}{1316285}$	
	6	$\frac{58}{15}$	$\frac{16150}{2343}$	$\frac{66096938}{4264995}$	∞
	7	$\frac{23}{6}$	$\frac{12182338}{1898815}$	10.5911	23.8918
	8	$\frac{218}{57}$	$\frac{41006344}{6520533}$	9.5472	
	9	$\frac{172}{45}$	6.2515		
	10	$\frac{814}{213}$	6.2409		
	11	$\frac{107}{28}$			
	12	$\frac{3038}{795}$			
	13	$\frac{2396}{627}$			
14	$\frac{11338}{2967}$				

6 Three Particles on C_n

Let A_n be the expected time to dispersion for 3 particles on C_n , with unbiased dispersion. We now have enough information about the $M = 3$ case to generalize for any n -cycle. Essentially, we have a sequence of rational numbers that converges to some number. The sequence starts as the first column in the above table, where $n = 4$. Let $A = \{A_4, A_5, A_6, \dots\}$ be the sequence of rationals. In the list of the first 10 terms of A below, we have undone some of the fraction simplification in order to see the trend.

$$A = \left\{ \frac{14}{3}, \frac{24}{6}, \frac{58}{15}, \frac{92}{24}, \frac{218}{57}, \frac{344}{90}, \frac{814}{213}, \frac{1284}{336}, \frac{3038}{795}, \frac{4792}{1254}, \dots \right\} \quad (2)$$

Our first main theorem gives an exact formula for this sequence in terms of a recursively defined constant.

Theorem 6. *Let $k = \lfloor \frac{n}{2} - 1 \rfloor$. For the unbiased dispersion of 3 particles on C_n with $n > 3$, the expected time to dispersion A_n , satisfies $A_n = \frac{1}{3} \left(10 + \frac{4}{3 - \alpha_{k-1}} \right)$, where $\alpha_k = \frac{1}{4 - \alpha_{k-1}}$, and*

1. *if n is even, then $\alpha_1 = \frac{1}{2}$, and*
2. *if n is odd, then $\alpha_1 = \frac{1}{3}$.*

Proof. We begin by splitting the expected time to completion for odd and even cycles into two distinct subsequences, A_{odd} , and A_{even} respectively. The motivation behind this is discussed heavily in the next section. Superficially, we do this because the subsequences are significantly easier to quantify separately. Moreover, we will show that both individual sequences converge to the same number. Bear in mind that the expected number of steps to completion for C_4 occurs at $n = 4$, so also the sequences begin incrementing from $n = 4$.

$$A_{even} = \left\{ \frac{14}{3}, \frac{58}{15}, \frac{218}{57}, \frac{814}{213}, \frac{3038}{795}, \dots \right\}$$

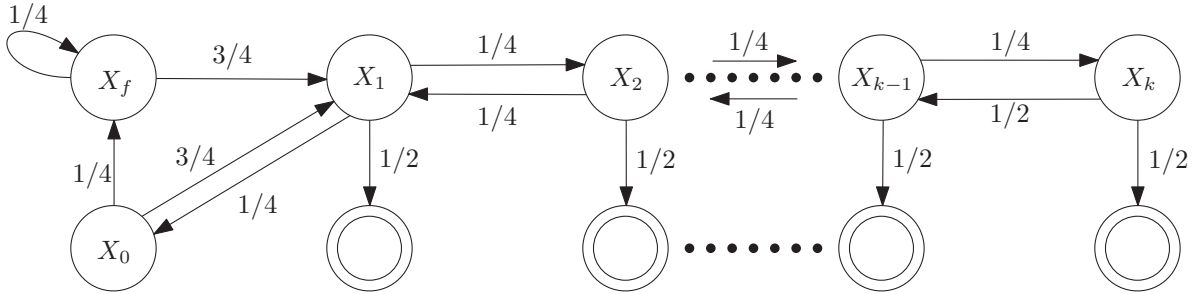
$$A_{odd} = \left\{ \frac{24}{6}, \frac{92}{24}, \frac{344}{90}, \frac{1284}{336}, \frac{4792}{1254}, \dots \right\}$$

Let X_f be the “first” state, and X_i for $i \in 0, 1, \dots, k$ be the transient state where the smallest number of empty vertices between the unhappy stack of 2 and the happy particle is exactly i , and $k = \lfloor \frac{n}{2} - 1 \rfloor$. For example, $X_0 = X_{(210\dots 0)}$, $X_1 = X_{(2010\dots 0)}$, etc. In unbiased

dispersion, it is convenient to consider $X_{(21000)} = X_{(20001)}$ (reflections of states are part of the same equivalence classes), since in terms of probability, clockwise and counterclockwise motion of particles is equal. In a slight abuse of notation, in what follows, we use X_f and X_i to represent the corresponding value in the first row of the fundamental matrix N corresponding to this Markov chain. Recall that the expected time to dispersion appears as the sum of the entries in the first row of N . Thus we have

$$A_n = X_f + \sum_{i=0}^k X_i. \quad (3)$$

Case 1: Assume n is even. By analyzing the possible transient states of three particles on an n -cycle, we arrive at the following state diagram with transition probabilities included.



Here we make note that from this state diagram, we can write down the general form of the transition matrix and relevant submatrix Q . Again, the expected value of \mathbf{T}_{Disp} is the sum of the first row of the matrix $N = (I - Q)^{-1}$. Since we are only concerned with the first row of N , we obtain a system of $k + 2$ equations (shown below) and $k + 2$ unknowns. We have multiplied each equation through by 4 (to clear denominators):

$$\begin{aligned}
3X_f &= X_0 + 4 \\
4X_0 &= X_1 \\
4X_1 &= 3X_f + 3X_0 + X_2 \\
4X_2 &= X_3 + X_1 \\
&\vdots \\
4X_{k-2} &= X_{k-1} + X_{k-3} \\
4X_{k-1} &= 2X_k + X_{k-2} \\
4X_k &= X_{k-1}
\end{aligned}$$

Summing up the left and right hand sides of this equations and making use of equation (3), we have

$$\begin{aligned}
4A_n - X_f &= 4 + 3X_f + 4 \sum_{i=0}^k X_i \\
&= 4 + 3X_f + 4X_0 + 2X_1 + \dots + 2X_{k-1} + 2X_k \\
&= 4 + X_f + 2X_0 + 2A_n
\end{aligned}$$

which implies that

$$A_n = 2 + X_f + X_0. \tag{4}$$

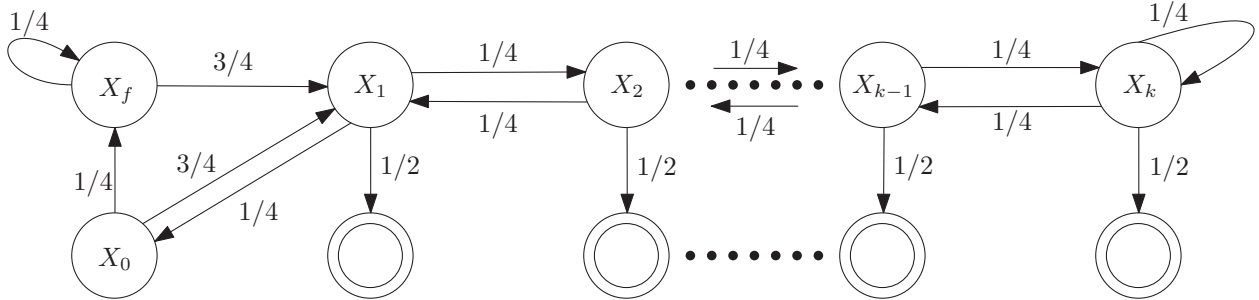
Next, we need to obtain the system's dependence on k , since the above is true for all n . We would like to specifically represent X_0 and X_f in terms of k . Observe that $\forall k > 1$, X_k is

dependent only on X_{k-1} . This fact does change in the odd n case, but only slightly. Then

$$\begin{aligned}
 X_k &= \frac{1}{4}X_{k-1} \\
 \implies 4X_{k-1} - \frac{1}{2}X_{k-1} &= X_{k-2}, \text{ where } \alpha_1 = \frac{1}{2}, \\
 \implies X_{k-1} &= \frac{1}{4 - \frac{1}{2}}X_{k-2} = \alpha_2 X_{k-2}, \text{ where } \alpha_2 = \frac{1}{4 - \alpha_1} \\
 \implies X_{k-2} &= \frac{1}{4 - \frac{1}{4 - \frac{1}{2}}}X_{k-3} = \alpha_3 X_{k-3}, \text{ where } \alpha_3 = \frac{1}{4 - \alpha_2} \\
 &\vdots \\
 \implies X_2 &= \alpha_{k-1}X_1
 \end{aligned}$$

Case 2: Assume n is odd.

The proof for this case is almost identical to that of the previous case. The only difference is that at X_k , there is a $\frac{1}{4}$ chance the state returns to itself, since the system is unbiased, and X_k is defined as k being the *minimum* number of empty vertices separating the happy particle from the unhappy stack. The diagram for this is as follows.



We derive almost the same equations, except we further solve X_k to be in terms of only X_{k-1} .

$$\begin{aligned}
3X_f &= X_0 + 4 \\
4X_0 &= X_1 \\
4X_1 &= 3X_f + 3X_0 + X_2 \\
&\vdots \\
4X_{k-1} &= X_k + X_{k-2} \\
3X_k &= X_{k-1}
\end{aligned}$$

Again combining these relationships with Equation (4), we obtain exactly Equation (5) from the previous case,

$$A_n = 2 + X_f + X_0. \quad (5)$$

In order to apply this equation, we must derive values for α_k in a similar fashion to that of the previous case.

$$\begin{aligned}
X_k &= \frac{1}{3}X_{k-1} \\
\implies 4X_{k-1} - \frac{1}{3}X_{k-1} &= X_{k-2}, \text{ where } \alpha_1 = \frac{1}{3}, \\
\implies X_{k-1} &= \frac{1}{4 - \frac{1}{3}}X_{k-2} = \alpha_2 X_{k-2}, \text{ where } \alpha_2 = \frac{1}{4 - \alpha_1} \\
\implies X_{k-2} &= \frac{1}{4 - \frac{1}{4 - \frac{1}{3}}}X_{k-3} = \alpha_3 X_{k-3}, \text{ where } \alpha_3 = \frac{1}{4 - \alpha_2} \\
&\vdots \\
\implies X_2 &= \alpha_{k-1} X_1
\end{aligned}$$

At this point, we can treat the cases as one, and finish the argument. Because of the dependence on k only in the coefficient α_k , we can reduce the system of equations from $k+2$

unknowns down to 2, namely X_0 and X_f if we substitute $4X_0 = X_1$. This gives us

$$\begin{aligned} 3X_f - X_0 &= 4 \\ 3X_f + (4\alpha_{k-1} - 13)X_0 &= 0 \end{aligned}$$

or equivalently,

$$\left(\begin{array}{cc|c} 3 & -1 & 4 \\ 3 & 4\alpha_{k-1} - 13 & 0 \end{array} \right)$$

By solving the system of two equations, we obtain

$$\begin{aligned} X_0 &= \frac{1}{3 - \alpha_{k-1}} \\ X_f &= \frac{1}{3} \left(4 + \frac{1}{3 - \alpha_{k-1}} \right) \end{aligned}$$

Finally, substituting into equation (3), we have

$$A_n = 2 + \frac{1}{3} \left(4 + \frac{1}{3 - \alpha_{k-1}} \right) + \frac{1}{3 - \alpha_{k-1}} = \frac{1}{3} \left(10 + \frac{4}{3 - \alpha_{k-1}} \right). \quad (6)$$

□

Theorem 7. *For the unbiased dispersion of 3 particles on the infinite path, P , the expected time to dispersion is $\frac{1}{3} \left(10 + \frac{4}{1 + \sqrt{3}} \right)$.*

Proof. We know that dispersion on P behaves almost exactly like that on C_n with sufficiently large n , so $\mathbf{T}_{Disp} = \lim_{n \rightarrow \infty} A_n$. We will show that $\alpha = \lim_{k \rightarrow \infty} \alpha_k = 2 - \sqrt{3}$ when n is both odd and even. First note that α satisfies

$$\alpha = \frac{1}{4 - \alpha} \implies \alpha^2 - 4\alpha + 1 = 0 \implies (\alpha - r_1)(\alpha - r_2) = 0$$

where $r_1 = 2 - \sqrt{3}$ and $r_2 = 2 + \sqrt{3}$. We will show the limit is r_1 by induction on k , and give

different base cases for the different parities of n . Remember that $k \rightarrow k+1 \implies n \rightarrow n+2$.

Induction Hypothesis: Assume for some k that $\alpha_k \leq \alpha_{k-1}$. It follows that

$$\frac{1}{4 - \alpha_{k-1}} \leq \alpha_{k-1} \implies \alpha_{k-1}^2 - 4\alpha_{k-1} + 1 \leq 0.$$

Moreover, we can say that $(\alpha_{k-1} - r_1)(\alpha_{k-1} - r_2) \leq 0$. Since the polynomial is upwards facing, we can claim that $\alpha_{k-1} \in (r_1, r_2)$.

We want to show that $\alpha_{k+1} \leq \alpha_k$, or equivalently that $\alpha_k^2 - 4\alpha_k + 1 \leq 0$. Applying our definition of α_k , we have

$$\begin{aligned} & \left(\frac{1}{4 - \alpha_{k-1}} \right)^2 - 4 \left(\frac{1}{4 - \alpha_{k-1}} \right) + 1 \\ &= \frac{1}{(4 - \alpha_{k-1})^2} (1 - 16 + 4\alpha_{k-1} + 16 - 8\alpha_{k-1} + \alpha_{k-1}^2) \\ &= \frac{\alpha_{k-1}^2 - 4\alpha_{k-1} + 1}{(4 - \alpha_{k-1})^2} = \frac{(\alpha_{k-1} - r_1)(\alpha_{k-1} - r_2)}{(4 - \alpha_{k-1})^2} \end{aligned}$$

Since the denominator is always positive, and by the inductive hypothesis, we assume that the numerator is not positive. For base cases, we have $\alpha_1 = \frac{1}{2}, \alpha_2 = \frac{2}{7}$ for even cycles, and $\alpha_1 = \frac{1}{3}, \alpha_2 = \frac{3}{11}$ for odd cycles, all of which hold for the induction. Therefore, $\alpha_k^2 - 4\alpha_k + 1 \leq 0$, so $\alpha_{k+1} \leq \alpha_k$. Thus, the sequence of α_k is decreasing and bounded below by r_1 , and $\alpha = \lim_{k \rightarrow \infty} \alpha_k = 2 - \sqrt{3}$. Moreover,

$$\mathbf{T}_{Disp} = \lim_{n \rightarrow \infty} A_n = \frac{1}{3} \left(10 + \frac{4}{3 - \lim_{k \rightarrow \infty} \alpha_k} \right) = \frac{1}{3} \left(10 + \frac{4}{1 + \sqrt{3}} \right)$$

□

7 The Parity of Vertices

In the table for $\mathbf{T}_{Disp}(M, n, \frac{1}{2})$, notice that $\mathbf{T}_{Disp}(4, 4, \frac{1}{2})$ and $\mathbf{T}_{Disp}(6, 6, \frac{1}{2})$ are both ∞ , but $\mathbf{T}_{Disp}(5, 5, \frac{1}{2})$ has a finite, reasonable value. As it turns out, $\mathbf{T}_{Disp}(7, 7, \frac{1}{2})$ also has a finite value, and in general, it seems that odd values for n particles on the n cycle all eventually terminate. Conversely, none of the even values of n particles on C_n terminate. We mentioned this phenomenon briefly earlier, and now we revisit it, in addition to some other observations. In this section, we will prove the following results.

Lemma 1: Let $M > 3 \in \mathbb{N}$ particles be dispersed on the graph C_n . If $n \in \mathbb{N}$ is even, all unhappy particles have the same parity.

Proposition 1: For the dispersion of M particles on the infinite path, \mathbf{D}_{Disp} , the greatest distance from the origin vertex to any particle, is at least $\lceil \frac{M}{2} \rceil$.

Theorem 5: All possible arrangements of M particles on C_n are reachable through dispersion if $n > 3$ is an odd integer and $M > \frac{3}{5}n$. Moreover, this bound cannot be improved.

Corollary 1: Let $n > 3 \in \mathbb{N}$. The dispersion process for exactly n particles on C_n terminates in a finite number of steps if and only if n is odd.

7.1 Proof for Lemma 1

First, we will prove a lemma about even cycles. This lemma also has ties into many other ideas regarding dispersion, each with discussions to follow.

Assume n is an even number.

Recall that the definition of a particle's parity is rooted in the labeling of the vertices of

C_n . Specifically, let $V(G) = \{v_1, v_2, \dots, v_n\}$, where v_1 is the origin vertex, and the vertices are labeled sequentially clockwise around C_n . For the sake of notation, let $|v_i|$ denote the number of particles at v_i (e.g. $|v_i| \geq 2 \implies$ all particles at v_i are unhappy). Also note that $v_0 = v_n$, and $v_{n+1} = v_1$.

Assume now, for the purpose of contradiction, that for some time step t , $\exists a, b \in [n]$ s.t. $|v_a| \geq 2$, $|v_b| \geq 2$, and $|a - b|$ is an odd integer. In other words, we are assuming there exist 2 stacks of unhappy particles with opposite parities at time t . since the following holds for both v_a and v_b , WLOG we will only consider what happens at v_a .

Note that when the state jumped from time $t - 1$ to t , there are two cases for the particles at v_a .

Case 1: At $t - 1$, $|v_a| \leq 1$. Because at time t the particles at v_a are unhappy, we know that at least one unhappy particle came from either v_{a-1} or v_{a+1} .

Case 2: At $t - 1$, $|v_a| \geq 2$. In this case, the particles at v_a are all unhappy. This means that when the state changes, all the particles must leave v_a . But again because at t the particles at v_a are unhappy, we know that at least *two* particles must have come from some combination of either v_{a-1} or v_{a+1} .

In either case, we know this: that at time $t - 1$, either v_{a-1} or v_{a+1} contained a stack of unhappy particles. Because n is even, both v_{a-1} and v_{a+1} must have the opposite parity as v_a . Since these hold for v_b in addition to v_a , we can claim the following.

Let v_A be the vertex with $\max\{|v_{a-1}|, |v_{a+1}|\}$ at $t-1$, and v_B the vertex with $\max\{|v_{b-1}|, |v_{b+1}|\}$, also at $t - 1$. Then $|v_A| \geq 2$, $|v_B| \geq 2$, and $|A - B|$ is an odd number.

Therefore, for any time step t for which our assumption is true, it must also have been true at $t - 1$. By reverse induction on t , we arrive at our contradiction, since the initial state for particle dispersion does not meet this condition. Thus, through the dispersion process of M particles on C_n with n even, it is impossible for two stacks of unhappy particles to have the opposite parity. In other words, all unhappy particles share the same parity. \square

7.2 Proof for Proposition 1

In and of itself, the above result is rather satisfying, but we will use it to show the following. This item has to do with what happens in particle dispersion on the infinite path. Theorem 2 talks about the bounds on \mathbf{D}_{Disp} , the distance of the farthest particle from the origin vertex. Specifically, the theorem claims that

$$\lfloor \frac{M}{2} \rfloor \leq \mathbf{D}_{Disp} \leq 4(1 + \epsilon)M \log M.$$

Freize and Pegden improved the upper bound in [2] to $O(n)$, but we would like to mention (and slightly improve) the lower bound to $\lceil \frac{M}{2} \rceil$

Let P be the infinite path, and $M \in \mathbb{N}$ the number of particles.

The number $\lfloor \frac{M}{2} \rfloor$ is almost trivial to prove; if the greatest distance of a particle from the origin is less than this, then by Pigeonhole Principle, there must exist a vertex occupied by more than 1 particle.

The additional 1 comes from this problem of parity. The infinite path behaves like an even cycle in terms of the parity of particles. Label the vertices of P as exactly $V(P) = \{\dots v_{-1}, v_0, v_1, v_2, \dots\}$, where v_0 is the origin vertex, v_i with $i > 0$ is the i -th vertex to the right of the origin, and v_i with $i < 0$ the i -th vertex to the left of the origin. Then at every distinct time step, all unhappy particles change parities, and the result from Lemma 1 holds that all unhappy particles share the same parity.

The ending condition still holds that all unhappy stacks in the state before the ending state must be of exactly size 2. Coupling these two facts, we see that any unhappy stack in this state splits into two happy particles, and must leave behind an empty vertex between the origin and the particle at \mathbf{D}_{Disp} . This fact guarantees that there will be an empty vertex somewhere between the farthest particle to the left and that to the right.

Thus, the distance of the farthest particle from the origin must be at least $\lceil \frac{M}{2} \rceil$. □

7.3 Proof for Theorem 5

Next, we will revisit the phenomenon observed in Table 1 and Table 2, where the number of states reachable through dispersion is equal to the number of arrangements of particles when n is odd.

Let $X(t)$ be an arbitrary state at time t called the target state, and as always, label the vertices $V(C_n) = \{v_1, \dots, v_n\}$. In what follows, we present a scheme in which the particles can reach $X(t)$ by moving according to the rules of dispersion, provided that various conditions around $X(t)$ are met. Then we discuss cases where some of these conditions fail, but the state is still reachable. Finally, we will show that if all arrangements are reachable, then $n > 3$, n is odd, and $M > \frac{3}{5}n$. Note that the probability of the particles moving according to our scheme is inconsequential, so long as it is *possible* for them to do so.

We begin by stating the following lemmas, the first of which lets us guarantee that any two of the stacks in $X(t)$ can have either the same or opposite parities.

Lemma 2. *Let n be an odd integer with $n \geq 3$. Given any two stacks in state $X(t)$, the parity of those stacks can be either opposite or equal, depending on the labeling of $V(C_n)$.*

Next, we must define a *downbeat* as a value of t which has the same parity as the time step in which $X(t)$ is reached. The motivation behind this definition is to subdivide the sequence of discrete time steps into two alternating subsequences (one of downbeats, the other as *upbeats*), so that at each downbeat, we would like each stack to have the same parity as its target vertex. An *upbeat* is a time step which is not a downbeat.

The following lemma gives us the power to claim that unhappy stacks can move independently throughout the cycle as long as they avoid crossing from v_n to v_1 or vice-versa. The proof is rather trivial, given our definition of a downbeat. Nevertheless we will give the proof at the end of this section, along with that for Lemma 2.

Lemma 3. *At each time step t for dispersion of M particles on an odd-cycle, no two unhappy stacks of opposite parities will have any particles move to the same vertex unless that*

vertex is the origin vertex.

We will now discuss the algorithm which can be used to reach most states, and then we will address cases which require a more creative approach at the end.

First, we list the assumptions necessary for this algorithm:

1. There are at least 2 unhappy stacks in $X(t)$.
2. Every unhappy stack in $X(t)$ is either adjacent to another unhappy stack, or an unoccupied vertex.

We will demonstrate in what follows that $X(t)$ is reachable by an alternate strategy even if the assumptions are relaxed, although doing so requires some work. Next, we list the rules governing particle motion throughout the process:

1. Any time an unhappy stack lands on a vertex with a happy particle, that unhappy stack returns the happy particle to its vertex two time steps later by splitting it off as the stack continues to move.
2. If an unhappy stack arrives at its target vertex in $X(t)$ before time step t , it must *wait* there until step t . “Waiting” is possible if and only if there is an empty vertex or unhappy stack adjacent to the stack which is waiting. The process of waiting involves the whole stack moving as a unit between its target vertex (on the downbeats) and the viable neighboring vertex (on the upbeats).
3. After the stack of unhappy particles (in $X(t)$) splits into the odd stack and even stack (defined below), these stacks never land on the origin vertex again. This ensures that Lemma 3 holds, and the two parities maintain their independence of one another.

4. At every downbeat, each stack which has reached the vertex it will eventually end on returns to that vertex.

Finally, here are the steps for the algorithm itself:

1. By Lemma 2, we can relabel the vertices $V(C_n)$ so that at least 2 unhappy stacks in the final arrangement of particles are on the opposite parity. We can choose to move the initial stack from the origin vertex to the new vertex labeled v_1 , but this will not affect the execution.
2. The first $n + 1$ time steps simply let the stack of particles run clockwise around the cycle, dropping off happy particles as it goes on vertices which, in $X(t)$, contain happy particles. At the end of this loop, all particles which belong to an unhappy stack in $X(t)$ should occupy the origin vertex, along with the happy particle which ends at that vertex, if such is the case in $X(t)$.
3. Let N_1 be the number of particles in unhappy stacks with parity 1 in $X(t)$ (the sum of the odd stacks), and N_0 be the number of particles in unhappy stacks with parity 0 in $X(t)$ (the sum of the even stacks). At the next time step, the current stack splits into two unhappy stacks, one of size N_0 called the *even stack*, one of size N_1 , the *odd stack*.
4. The odd stack and the even stack begin by moving clockwise and counter-clockwise, respectively. They march all the way around the cycle splitting off unhappy stacks on the downbeats. Those unhappy stacks *wait* until all the others have reached their target vertices.

If either of the two assumptions we've made are not met, then this algorithm fails, so we will deal with the cases where $X(t)$ is a state which does not meet these conditions. Note that if we show that a state which can itself reach the ending state $X(t)$, and that that state

(call it $X(t - 1)$) is reachable, then this is equivalent to showing the target state $X(t)$ is reachable.

First, assume that condition 1 fails, that there are not at least 2 unhappy stacks in $X(t)$ ($\implies \exists$ at most 1 unhappy stack). Then we have the following cases:

Case 1: Assume there is exactly one stack in $X(t)$.

Since condition 2 still holds, the process ends immediately after the unhappy stack drops off all its happy particles

Case 2: Assume there are no unhappy stacks in $X(t)$.

Since $\frac{3}{5}n < M$, by averaging, we know that there exists a sequence of 5 adjacent vertices in $X(t)$ with at least 4 of them having happy particles. If these particles are all adjacent, then at $X(t - 1)$, there were 2 adjacent stacks of size 2 which both split. Otherwise, in $X(t - 1)$ there was a stack of size 2 at the empty vertex in $X(t)$ which split into two happy particles.

Now we assume condition 2 fails, so that there exists in $X(t)$ an unhappy stack with both adjacent vertices occupied by happy particles, say at v_j, v_{j-1}, v_{j+1} respectively for $j \in [n]$. This formation is only reachable by $X(t - 1)$ having a stack of exactly 2 at v_j , and the unhappy stack occupying either v_{j-1} or v_{j+1} (it doesn't matter which). All of the other unhappy stacks would be on their respective adjacent upbeat vertices.

But this both assumptions are met for $X(t - 1)$ which implies that $X(t - 1)$ is reachable through the algorithm, so $X(t)$ is reachable by dispersion.

Note that condition 2 states that every unhappy stack needs a viable adjacent vertex at which it can wait during the upbeats. For this condition to fail for a given unhappy stack, that stack must be straddled by two unhappy particles in the ending state. But there could be a chain of unhappy stacks all on the same parity with happy particles in between them, causing condition 2 to fail for each stack. There could also be multiple independent cases around the cycle of this condition.

The solution generalizes from the previous example in that at $X(t - 1)$, each of those happy particles must have come from somewhere. For each independent chain of particle stacks following the pattern $1, k_1, 1, k_2, 1, k_3, \dots, 1, k_l, 1$, where $\forall i \in [l], k_i \geq 2$, that section of $X(t - 1)$ looks like $k_1, 2, k_2, 0, k_3, 2, k_4, 0, \dots$. In this state, the happy particles have been paired up into stacks of 2 which will split when the time step updates from $t - 1$ to t , and the unhappy stacks all sifted (arbitrarily) clockwise. All other unhappy stacks did not conform to this pattern so they must have been able to wait, and thus at $t - 1$ they each occupied a viable adjacent vertex. If l was odd, then one happy particle was not paired up, and occupied the same location at both $t - 1$ and t . Now, $X(t - 1)$ satisfies all five conditions and thus $X(t)$ is reachable by dispersion.

Now, we will show that the bound $M > \frac{3}{5}n$ cannot be improved.

First, we will show the necessity that n is odd and $n > 3$. We will prove both of these conditions by counterexample. Specifically, assume each condition is not true, and show a state which cannot be reached by dispersion.

First, assume that $n = 3$, and $M = 3$. Then the ending state $X_{(111)}$ is unreachable, since the initial state can either go to itself or the pile can split into stacks of size 1 and 2. Because when the 2 splits to “finish”, it actually wraps around and one of the 1’s combines with the currently happy particle, the game will never end. For $n \leq 2$, we are not interested in this result.

Next, assume n is an even number. By Lemma 1, any arrangement of particles in which there can be found 2 unhappy stacks of opposite parities is unreachable.

Lastly, assume $M \leq \frac{3}{5}n$. Realistically, this bound is only a signpost result. Indeed, there are 5 different cases to consider where this assumption breaks down. In what follows, we discuss all 5 and give the necessary bound for each.

Remark: Heuristically, the idea is that what causes an arrangement of particles to be

unreachable on account of insufficient quantity is when there is too much space between small strings of happy particles.

Under our assumption, the source of the contradiction arises from the fact that no unhappy stack can deposit particles in the arrangement of $X(t)$. This gives rise to a new set of conditions which must apply in order to break the algorithm. If any of these are not true, the state would satisfy the necessary conditions above.

1. The target arrangement $X(t)$ cannot contain any unhappy stacks.
2. any unoccupied vertex between happy particles must be adjacent to another unoccupied vertex (creating a gap between happy particles of length at least 2).
3. A string of happy particles in the ending state can be at most 3 particles long, since a string of 4 happy particles could have come from a state with two adjacent unhappy pairs.

Combining these rules, we see that the particle arrangements $X(t)$ which cause the algorithm to fail in this way all take the form of a set of strings of adjacent happy particles with length at most 3, separated by strings of at least 2 adjacent empty vertices. Therefore, the greatest number of particles we can use in order to find a state which cannot be reached by dispersion can be found by packing the happy particles in to the final state as efficiently as possible, obtaining the highest ratio of happy particles to empty vertices as possible.

Since every string of happy particles is adjacent to at least 2 empty vertices, we try to minimize the empty vertices and maximize the happy particles. It is from this minimization process that we obtain the following cases. Note that 3 happy particles per 2 empty vertices is the best ratio possible, so we use as many sequences of “1, 1, 1, 0, 0” as possible. In what follows, let $k \in \mathbb{N}$ be odd, and $n = 5k + b$ for b even in \mathbb{N}

Case 1: Assume $b = 0 \implies n = 5k$.

This case gives us our signpost result, because there is no more efficient way to arrange particles on C_n than by forming a sequence of $\frac{n}{5}$ strings, each of 3 happy particles followed by 2 empty vertices. Thus, if $n = 5k$, $M > \frac{3}{5}n$ for all states to be reachable by dispersion.

Case 1: Assume $b = 2 \implies n = 5k + 2$.

Notice that in our above example, we have exactly one state which cannot be reached by dispersion. When we add 2 vertices to the cycle, it is actually impossible to add more particles to the system, because no matter where we place those happy particles, there will be at most 1 empty vertex between the inserted particle and a string of 3 happy particles.

While the number of particles does not increase, the number of states which cause dispersion to fail does increase, since there are more ways to arrange the particles in such a way so as to satisfy the three conditions above. Thus, if $n = 5k + 2$, $M > \frac{3}{5}(n - 2) = \frac{3n-6}{5}$ for all states to be reachable by dispersion.

Case 3: Assume $b = 4 \implies n = 5k + 4$.

When we add 4 vertices to the first case, we can add exactly 2 particles before we hit capacity. The state looks like that of the case 1, but one of the strings of happy particles is only length 2. This creates the bound for when $n = 5k + 4$, $M > \frac{3}{5}(n - 4) + 2 = \frac{3n-2}{5}$.

Case 4: Assume $b = 6 \implies n = 5k + 6$.

Adding 2 vertices from the previous case only allows us to add one more particle, which completes the string of 3 unhappy particles. The other vertex must remain unoccupied, otherwise it will cause the state to become reachable. In this case that also means that there is a string of at least 3 unoccupied vertices. Thus we obtain the bound $n = 5k + 6, \implies M > \frac{3}{5}(n - 6) + 3 = \frac{3n-3}{5}$.

Case 5: Assume $b = 8 \implies n = 5k + 8$.

From the last case, if we add two more vertices, then we can add one more particle, which will allow for an isolated happy particle. There are other arrangements of this many particles which fail to be reached by the dispersion process, but none with more particles. Thus, for $n = 5k + 8$, we have that $M > \frac{3}{5}(n - 8) + 4 = \frac{3n-4}{5}$.

The greatest proportion of these is $M = \frac{3n}{5}$, so we use this as our lower bound for the theorem. Thus, if all possible arrangements of M particles on C_n are reachable by dispersion, then $M > \frac{3n}{5}$, n is odd, and $n > 3$. \square

Proof of Lemma 2:

This Lemma is called the spin-lemma, because by relabeling the vertices of C_n , or “spinning” the labels on the vertices, we can control whether two stacks have the same or opposite parities.

Let $X(t)$ be any arrangement of M particles on C_n with n an odd integer > 3 . Fix two vertices v_a and v_b in $V(C_n)$ with $a, b \in [n]$. Now label the vertices $V(C_n) = \{v_1, v_2, \dots, v_n\}$, with $a = 1$. Then $b - a = b - 1$ is either odd or even. Next, relabel the vertices in the same way, except $b = 1$. Then we have $b - a = 1 - a$ is also either odd or even. We want to show that $b - a$ in the first labeling is odd $\implies b - a$ in the second labeling is even (and also the case where $b - a$ even $\implies b - a$ odd).

Note that because v_a and v_b are fixed, we have that in the first labeling, $b - a = b - 1$ is the number of vertices between v_a and v_b , including the latter but not the former. The significance behind the labeling is that we are *counting clockwise*, starting from v_a and going to v_b . Likewise, in the second labeling, $b - a = 1 - a$ is the number of vertices between v_a and v_b counting clockwise starting from (but not including v_b).

Then the sum of these two numbers is exactly n . Moreover, since n is odd, one of them must be odd and the other must be even. Therefore, if the labeling of $V(C_n)$ starts at $v_a = v_1$ and b and a have the same parity, then labeling $V(C_n)$ with $v_b = v_1$ gives that b and a have the opposite parity. \square

Proof of Lemma 3:

Let v_a and v_b be vertices in $V(C_n)$, with $a, b \in [n]$, such that $a - b$ is an odd number, and $|v_a| \geq 2$, and $|v_b| \geq 2$. At time step t , the unhappy particles on v_a and v_b all move to

an adjacent vertex, v_{a+1} , v_{a-1} , v_{b+1} , and v_{b-1} . since v_a and v_b had the opposite parity, v_{a+1} and v_{a-1} share the same parity which is opposite to that of v_{b+1} and v_{b-1} .

The only time this is not true is when either $a = n$ or $b = n$. in this case, $v_{n+1} = v_1$ which is the origin vertex, but more importantly, it shares the same parity as n since n is odd. Then, if the other unhappy stack was at v_2 , both stacks can meet at the origin vertex. Otherwise, The unhappy stacks will not interact. \square

Another way of thinking about this Lemma is to say, “As long as no unhappy stacks touch the origin vertex, their parities will change consistently on every time step.” This allows us to control and monitor the independence of stacks which occupy vertices of opposite parities.

7.4 Proof for Corollary 1

We now have the tools for a very clean proof of the corollary regarding the termination of the dispersion process for exactly n particles on C_n . Consider one of the ending conditions for the general dispersion process, that at the time step before the system reaches its end state, the number of particles in a stack is at most 2. Specifically for n particles being dispersed on C_n , we have the following additional ending condition for the second to last state - that two stacks of two particles must lie on adjacent vertices, with no particles on a vertex adjacent to the pair. Equivalently, one of the permutations of the state label must contain “0220” (e.g. $X_{(220110)}$).

To verify this, simply observe that when the last step occurs, for the process to end, a stack of exactly 2 must split onto two adjacent empty vertices. Since that would leave a vertex unoccupied, another particle must simultaneously move to the vertex from which the pair split. that particle must also come from a stack of 2 which was adjacent to the first pair, and must have an empty vertex on the other side.

By Lemma 1, this is automatically impossible for all even n , since the parity of adjacent stacks is always opposite on an even cycle. If n is odd, we just proved that every possible

arrangement of the particles is reachable by the dispersion process on C_n . Since $X_{(11\dots 1)}$ is a possible arrangement, it can be reached. Since the process will not end until this state has been reached, the process must eventually reach it and thus come to an end. \square

8 Future Endeavors and Open Questions

One of the major obstacles we encountered was trying to calculate the matrix inverse in order to obtain the fundamental matrix for the absorbing Markov Chain. One pattern we noticed specifically for the $M = 3$ case was that the main diagonal of the Q matrix looked almost completely 2 - banded. For example, here is the Q section of the transition matrix for $Disp(3, 10)$, with clockwise bias $\frac{1}{3}$.

$$\begin{pmatrix} 1/3 & 0 & 4/9 & 0 & 0 & 0 & 0 & 0 & 2/9 & 0 \\ 1/9 & 0 & 4/9 & 0 & 0 & 0 & 0 & 0 & 4/9 & 0 \\ 0 & 1/9 & 0 & 4/9 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/9 & 0 & 4/9 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/9 & 0 & 4/9 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/9 & 0 & 4/9 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/9 & 0 & 4/9 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/9 & 0 & 4/9 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/9 & 0 & 4/9 \\ 4/9 & 0 & 4/9 & 0 & 0 & 0 & 0 & 0 & 1/9 & 0 \end{pmatrix}$$

While there are a few strange non-zero entries in this matrix, it is both sparse and organized, which indicates that it may be possible to decrease the time it takes to compute the inverse of $I - Q$.

In [6], they show that it is possible to recursively calculate the inverse of general r -banded matrices. We were able to show that the information contained in our matrix can be rearranged so that all the transition probabilities are preserved, but we can reduce the

problem to a 3-banded matrix. The reason we cannot further reduce the problem is because there exists a state, namely $X_{(20100\dots0)}$ which can be reached from three different states, none of which are itself. This requires three different vertical entries in the matrix at the column corresponding to that state, again none of which can lie on the main diagonal.

Still, this is very useful and even the third band is almost entirely empty. Using the results in [6], it should be possible to find an easier (more computationally efficient) way to invert this matrix. One open problem that arises from this is, can we algorithmically rearrange the order of the rows/columns of Q so as to minimize the number of bands required for the matrix?

One interesting point of future discussion is the proof for Proposition 5. We presented an algorithm which *can* reach all states in an odd cycle, given the proper initial conditions. The probability of this occurring is incredibly low, but still possible.

It would be interesting to look at the ties this idea has to machine learning. In practice, it seemed possible to generate the target state for any (reasonable) n -cycle by simply letting the computer conduct a random simulation, and breaking the loop if the target state has been reached. But machine learning could let us “teach” the computer that less iterations to get to the solution is better. If we did this, it would learn how to take the fastest route from the initial state to any target state, and we could then observe its movements, and formalize them into a much more efficient and probable algorithm.

Perhaps the most difficult question is so simply ask, what is the closed form for the expected time to reach dispersion for M particles on C_n with probability bias p ? While we know that the answer does exist, this question is incredibly vast, and all of the results in this paper, as well as those in [1] and [2] only point us in the direction of answering that question.

We looked for example at the *unbiased* dispersion of exactly 3 particles on any n cycle,

and we were able to draw various conclusions about the behavior of the expected time. But if we ask the same questions for 4 particles on the n cycle, we immediately hit a wall, because the particles are able to interact in a vast number of ways.

All that said, there is a large trade-off to having access to that closed form solution, because literally every single result would fall out of it. We would be able to see intuitively why things behave so differently, taking limits would explain the behavior of dispersion on the infinite path, and we could explore what happens as we sharply increase the number of particles.

The last future project we have is the idea of capacity on the vertices of G . Recall Theorem 1, which states that for the complete graph K_n , and the star S_n the following hold for any constant $\delta > 0$:

1. If the number of particles M satisfies $M/n \leq (1/2)(1 - \delta)$, then with probability $1 - O(1/n)$, the dispersion process terminates in $\mathbf{T}_{Disp} = O(\log n)$ steps.
2. If the number of particles M satisfies $M/n \leq (1/2)(1 + \delta)$, then there is a constant $c = c(\delta) > 0$ s.t. the probability that $\mathbf{T}_{Disp} \leq e^{cn}$ is less than e^{-cn} .

We would like to generalize this theorem, as well as offer a more general way to view the dispersion process in terms of a new variable. We define the *capacity* $k(v)$ of a vertex in $v \in V(G)$ as the maximum number of particles allowed on that vertex before all the particles on that vertex at time t become unhappy.

In every single case we've seen so far, the capacity of all the vertices in G has been exactly

1. We conjecture that the following result holds for $k = 2$.

For the complete graph K_n , and the star S_n with $k(v) = 2 \forall v \in V(K_n)$ or $V(S_n)$, the following hold for any constant $\delta > 0$

1. If the number of particles M satisfies $M/n \leq (1 - \delta)$, then with probability $1 - O(1/n)$, the dispersion process terminates in $\mathbf{T}_{Disp} = O(\log n)$ steps.

2. If the number of particles M satisfies $M/n \leq (1 + \delta)$, then there is a constant $c = c(\delta) > 0$ s.t. the probability that $\mathbf{T}_{Disp} \leq e^{cn}$ is less than e^{-cn} .

Although we failed to give a formal proof for this conjecture, structure of our analysis is similar to that of the proof given for 1 in [1].

We begin by letting H be the number of happy particles at time t , U the number of unhappy particles at t , and we assume that happy particles do not move at each time step t , but unhappy particles do move. Also let H' be the number of happy particles at time $t + 1$, which we will bound.

A particle is happy if either it is the only particle on its vertex, or if it shares that vertex with exactly one other particle. We will distinguish between these two cases by letting H_1 be the number of vertices containing particles which are “one-happy”, in that they are the only particles on their vertex. Similarly, we let H_2 the number of vertices containing “two-happy” particles, which are particles that share a vertex with exactly one other particle.

Note that $H_1 + 2H_2 = H$, and that in the case where $k(v) = 1 \forall v \in V(G)$, $H_1 = H$. The process will end when $H = M$. Define the following variables.

1. X_1 is the number of previously one-happy particles which have become unhappy from at least 2 particles landing on them.
2. X_2 is the number of previously two-happy particles which have become unhappy from at least 1 particle landing on them.
3. Y_1 is the number of previously unhappy particles which have become one-happy by being the only unhappy particle to land on an empty vertex.
4. Y_2 is the number of previously unhappy particles which have become two-happy, either by being 1 of 2 unhappy particles to land on an empty vertex or by being the only unhappy particle to land on a previously one-happy particle (thus making both two-happy).

At any time step t , given values for U, H_1, H_2 , randomly allocating U balls into n boxes, we can obtain the following expected values or X_1, X_2, Y_1 , and Y_2 .

$$\mathbf{E}X_1 = H_1 \left(1 - \left(1 - \frac{1}{n} \right)^U - \frac{U}{n} \left(1 - \frac{1}{n} \right)^{U-1} \right) \quad (7)$$

$$\mathbf{E}X_2 = 2H_2 \left(1 - \left(1 - \frac{1}{n} \right)^U \right) \quad (8)$$

$$\mathbf{E}Y_1 = U \left(\frac{n - (H_1 + H_2)}{n} \right) \left(1 - \frac{1}{n} \right)^{U-1} \quad (9)$$

$$\mathbf{E}Y_2 = U \left(\frac{U-1}{n} \left(\frac{n - (H_1 + H_2)}{n} \right) \left(1 - \frac{1}{n} \right)^{U-2} + \frac{H_1}{n} \left(1 - \frac{1}{n} \right)^{U-1} \right) \quad (10)$$

We cannot say much beyond this without talking about different concentrations, and perhaps the relative expected values of H_1 and H_2 . If we could prove this generalization however, it would be interesting to explore whether the notion of capacity generalizes for an arbitrary capacity κ .

References

- [1] C. Cooper, A. McDowell, T. Radzik, N. Rivera and T. Shiraga, *Dispersion Processes*, 2017, manuscript <https://arxiv.org/abs/1712.03389>
- [2] A. Frieze and W. Pegden, *A note on dispersing particles on a line*, 2017, manuscript <https://arxiv.org/pdf/1711.02192>
- [3] A. Frieze and M. Karoski *Introduction to random graphs* Cambridge University Press, Cambridge, 2016
- [4] W. Feller *An Introduction to Probability Theory and its Applications* John Wiley & Sons Inc, Canada, January, 1950
- [5] V. Shevelev *A Problem of Enumeration of Two-Color Bracelets with Several Variations*, May, 2011 <https://arxiv.org/pdf/0710.1370.pdf>
- [6] E. Kilic and P. Stanica *The Inverse of Banded Matrices* Journal of Computational and Applied Mathematics, September, 2011