Research Statement

Janine E. Janoski

1 Introduction

My research is in discrete mathematics and combinatorics. I have worked on a wide variety of research topics including, enumeration problems, eigenvalue problems, game theory, and the partition function. I enjoy problems that involve a computational component in order to gain insight.

Since I have worked on many computational type problems I have experience in several mathematical languages as well as experience in parallel computing. At Supercomputing 2010, I had the opportunity to help run a workshop on using Sage in mathematics. This was an opportunity to work with math teachers, mathematicians, professionals in industry, and professionals in other fields outside of mathematics to show the power that technology has as our knowledge of math and science move forward. During this conference I also had the opportunity to expand my knowledge about parallel computing as well as share my experiences using Sage in parallel.

For two summers, I had the opportunity to mentor during a Research Experience for Undergraduates (REU). Both summers I had three undergraduate students and one advanced high school student working as a group. A typical REU day would begin with a group meeting to talk about their progress, problem solve, and give them suggestions for future directions. Both groups worked on problems that had a computational component, so I helped them develop and implement efficient algorithms. I have kept in touch with the members of my groups; each year I encourage them to give a presentation at a conference. I think that an important part of research is being able to communicate it to a wide audience, and so, I help my former REU students fine tune their presentation skills.

I will now discuss some research projects that I have worked on and possible future directions of work related to these areas.

2 Partition Function

A sequence, a_n is said to be log-concave if for every n

$$a_n^2 \geq a_{n+1}a_{n-1}. \tag{1}$$

Let $p(n)$ be the number of partitions of n, for example there are five partitions of the number 4, namely $4, 3+1, 2+2, 2+1+1, 1+1+1+1$. Computationally looking at $p(n)$ we see that for $n \geq 26$ the partition function is log-concave [2]. To prove this we did an asymptotic analysis of Rademacher's formula. For $n < 26$, we observed that (1) is satisfied precisely when n is even. To prove this we looked at the expansion of Rademacher's formula. The leading term dominates the asymptotic behavior. When n is small the remaining terms of the expansion are dominated by the second term, which gives a negative contribution when n is odd.

This research was begun during the Summer 2010 REU held at Clemson University. The students were studying the Sperner Property of $\mathcal{P}_n$, the set of all partitions of n ordered by the following poset (partially ordered set). We say that a partition λ is covered by λ' if two summands of λ can be added to form λ'. Our poset is under the operation of transitive closure the covering property.

We say that two partitions are comparable if one covers the other. Using the idea of this covering relation, we can represent $\mathcal{P}_n$ using a Hasse Diagram. We create a vertex for each partition and arrange them in rows, called level sets, according to rank. We then create an edge between any two partitions such that one is covered by the other. $\mathcal{P}_n$ is Sperner if it satisfies unimodality and the maximum matching property for every consecutive pair of levels in the Hasse Diagram. A sequence, a_n is unimodal if there exists m such that $a(n) < a(n+1)$ for all $n < m$ and $a(n) > a(n+1)$ for all $n > m$. $\mathcal{P}_n$ satisfies the maximum matching property if for ever pair of consecutive level sets there is a bipartite matching.

We were able to prove that for $\lceil \frac{n}{2} \rceil \leq k \leq n$ there is a bipartite matching between consecutive levels. We were also able to show that this property held for several other levels, but were unable to generalize this proof.

Conjecture. *For $k < \lceil \frac{n}{2} \rceil$ there is a bipartite matching between consecutive levels.*

In 2003, Canfield showed that $\mathcal{P}_n$ satisfies the maximum matching property for all $1 \leq n \leq 45$ by using the Ford Fulkerson Algorithm [5].

Open Problem. *Are there particular edges which need to be used in the bipartite matching?*

If we can prove that such edges exist then we could begin Ford Fulkerson with a partial matching to help speed up computation time. Other computational approaches such as using the permanent of the adjacency matrix [11] [13] related to the Hasse Diagram.

In 1952, Szekeres [15] proved unimodality holds for sufficiently large n. In 2003, Canfield showed computationally that unimodality holds for $n \leq 2000$ [5]. Computationally we were able to show $\mathcal{P}_n$ is unimodal for $n \leq 25,000$. We would like to find bounds to determine the size of "sufficiently large". One technique to do this would be to prove the stronger property of log-concavity. We have computationally examined this property for the sequence formed by our level sets, namely, $p(n,k)$, the number of partitions with size k. We noticed that $p(n,k)$ appeared to be log-concave for all but the last 26 terms, where we observed an alternating behavior. The second half of the sequence $p(n,k)$ is the reverse of the first $\frac{n}{2}$ terms of $p(n)$. Since we have shown $p(n)$ is log-concave for $n \geq 26$, it remains to show that first half of $p(n,k)$ is log-concave.

Conjecture. *$p(n,k)$ is log-concave for $k \leq n - 26$.*

3 Graph Nim

Graph Nim is an impartial two player game, where the players take turns removing any number edges incident to a vertex they choose during a given turn. The object of Graph Nim is to be the person to remove the last set of edges from the graph. For a given graph we can associate a non-negative number, called the Sprague-Grundy number, based on the possible subgraphs we can reach in one move. On a given turn, if the value of the current graph is non-zero then current player has a winning move. This happens because with a positive Sprague-Grundy number the player can always make a move which will force the resulting graph to have a Sprague-Grundy number of zero.

We began by analyzing this game on paths, that is trees with two vertices of degree one and all other vertices of degree two. On a path that has at least one edge, the first player has a winning move. If the graph has an odd number of edges, the player simply takes the edge in the middle and if the graph has an even number of edges, the winning move is to take the two center edges. For the rest of the game the first player will mirror the moves of the second player until the first player has won. Considering the Sprague-Grundy numbers as the length of the paths increase, we see a periodic behavior emerge. In particular, for paths whose length is longer than 72 we have a period of 12. It turns out this game is equivalent to the game of Kayles, where expert bowlers remove one or two adjacent bowling pins from a line of pins [6].

We then begin looking at paths with an extra edge attached to the second vertex. We again were able to prove as the length of the path increased there was a period of 12 after length 156. We examined other paths with one extra edge attached to a given vertex and noticed all appeared to have a period of 12 or 60 that eventually appeared [3]. We define a G-Path to be any graph G with a path appended to a fixed vertex on G.

Conjecture. *The Sprague-Grundy numbers of a G-Path are periodic as the length of the path increases.*

We hope to be able to prove this by doing case analysis of what Sprague-Grundy numbers can appear as the length of the path increases. In particular, we can examine the order in which the Sprague-Grundy numbers can appear as the length of the path increases. We hope to show that there is no arrangement in which all Sprague-Grundy numbers can appear modulo 16, and thus the Sprague-Grundy numbers would have to bounded.

Another interesting phenomenon occurs when we consider graph nim on the complete graph, K_n. We notice that K_1, K_3, K_5 are all losing graphs for the first player and K_2, K_4, K_6 are winning graphs for the first player. We note that if K_n is a losing graph then K_{n+1} must be a winning graph, so there are infinitely many winning complete graphs.

Conjecture. *K_{2k+1} is always a losing graph*

We have computationally shown that this holds for up to K_{11}.

Many other interesting directions can be taken from this problem. In the simple case of paths we were able to determine not only who would win the game but what the winning strategy would be. Are there other graphs where we can determine the winning strategy?

Open Problem. *If we restrict the players' moves so that they must choose a vertex based on some rule, are the Sprague-Grundy numbers periodic?*

Open Problem. *If we only allow the players to choose an even number of edges to remove, are the Sprague-Grundy numbers periodic?*

For questions like these we might need to reconsider what the terminal graph might be. For example, if we are removing an even number of edges, the "winning" graph might be defined to be a graph where there are no more moves instead of a graph with no more edges.

4 Matrices with Fixed Row and Column Sums

How many $n \times m$ matrices with entries in $\{0, 1\}$ and some fixed row and column sum are there? This question has been studied by various mathematicians with several different specifications. Reed [12], Stanley [14], McKay [7], [10], Anand [1], Gao [9] [8] have all studied the enumeration of these types of matrices.

We studied this problem for an $n \times n$ matrix with row and column sum 2 and the restriction that the rows have a lexicographical ordering [4]; we denote this set by $\mathcal{M}(n, 2)$ and $|\mathcal{M}(n, 2)| = M(n, 2)$. We found that for $M(n, 2)$ we could write down a simple recurrence relation. We found a constructive proof for this recurrence relation, which inductively constructs all matrices of size $n \times n$. If we place an extra restriction of the matrices being symmetric then this recurrence becomes the fibonacci sequence.

Let $\mathcal{S}(n)$ be the set of symmetric matrices over $\{0, 1, 2\}$ with trace zero and sum 2 and $|\mathcal{S}(n)| = S(n)$. We were also able to show that $M(n, 2) = S(n)$. To prove this we showed a bijection between $\mathcal{M}(n, 2)$ and $\mathcal{S}(n)$. In particular, the moves to construct a matrix in $\mathcal{M}(n, 2)$ correspond to the moves to construct a matrix in $\mathcal{S}(n)$.

There are many other potential open problems still to be explored on this research.

Open Problem. *Is there a recurrence for the number of $n \times n$ binary matrices with ordered rows and row and column sum 3?*

Open Problem. *Is there a recurrence for the number of $n \times n$ binary matrices with ordered rows and a row and column sum of 3 if we add more restrictions such as require the matrices to be symmetric, have a certain trace, or can't have repeated row?*

Open Problem. *What is the probability that an $n \times n$ binary matrix with specified properties is invertible?*

5 Evolution of Strings

Define a string of length l where each letter is chosen from an alphabet of size k. If a letter is incorrect we select a new letter. If a letter is correct with some probability, p we reselect a letter for that position. We continue to flip letters until we converge to the correct word. How quickly will it take to get to the correct string? To do this we create the transition matrix based on our string length and alphabet size. We then are interested in the size of the eigenvalues to determine the rate of convergence. We know that we have a trivial eigenvalue of 1 and were able to observe that the second largest eigenvalue appeared to be converging to 1 as the size of l and k increased.

In 2010, Wilf and Ewens [16] examined a similar problem; they assumed the probability of reselecting for a position with a correct letter was 0. This research was conducted as a model to see if there was enough time for evolution to have taken place. For their model they used $l = 20,000$, which is the number of gene loci where the replacement process takes place in the genome. In a population with millions of births, about 250 mutations per gene per year occur and only one mutation is 10,000 is favored; thus they let $k = 40$. Using these parameters, how does our modified model behave?

Other parameters we might run in our model include gene mutations and amino acids. A protein is a chain of amino acids. We can use our model to study non-standard amino acids found in proteins. These amino acids are formed by post-transitional modification, which occurs after translation.

Open Problem. *Can our model be used to explore the behavior of non-standard amino acids found in proteins?*

Open Problem. *If we allow each position in the string to have a different probability that there would be a mutation, how does the model behave?*

References

[1] H. Anand, V.C. Dumir, and H. Gupta. A combinatorial distribution problem. *Duke Math J.*, 33:757–770, 1966.

[2] Brian Bowers, Neil Calkin, Kerry Gannon, Janine E. Janoski, Katie Joes, and Anna Kirkpatrick, 2011. In Preparation.

[3] Neil Calkin, Kevin James, Janine E. Janoski, Sarah Leggett, Bryce Richards, and Stephanie Thomas. Computing strategies for graphical nim. *Congressus Numerantium*, (202):171–186, 2010.

[4] Neil J. Calkin and Janine E. Janoski. Matrices of row and column sum 2. *Congressus Numerantium*, (192):19–32, 2008.

[5] E. Rodney Canfield. Integer partitions and the sperner property. *Theoretical Computer Science*, 307(3):515–529, 2003.

[6] Thomas S. Ferguson. Game theory. Lecture Notes, Math 167, School, 2000.

[7] Catherine Greenhill and Brendan D. McKay. Asymptotic enumeration of dense 0-1 matrices with specified line sums and forbidden positions. arXiv:0701600v1, January 2007.

[8] Ken Matheis, Gao Shanzhen, and Tan Zhonghua. Some formulas of (0,1)-matrices. *Congressus Numerantium*, 182:53–63, 2006.

[9] Ken Matheis, Gao Shanzhen, and Tan Zhonghua. Enumeration of nonnegative integer matrices. *Congressus Numerantium*, 184:193–207, 2007.

[10] Brendan D. McKay and Xiaoji Wang. Asymptotic enumeration of 0-1 matrices with equal row sums and equal column sums. *Linear Algebra and its Applications*, 373:272–287, 2003.

[11] R. C. Mittal and Ahmad Al-Kurdi. Efficient computation of the permanent of a sparse matrix. *Intern. J. Computer Math.*, 77:189–199, 2001.

[12] R. C. Reed. *Some Enumeration Problems in Graph Theory.* PhD thesis, University of London, 1958.

[13] H.J. Ryser. *Combinatorial Mathematics.* Mathematical Association of America, 1963.

[14] Richard P. Stanley. *Enumerative Combinatorics*, volume 2. Cambridge, 1999.

[15] G. Szekeres. Some asymptotic formulae in the theory of partitions(ii). *Quarterly Journal of Mathematics*, 2(4):96–111, 1953.

[16] Herbert S. Wilf and Warren J. Ewens. There's plenty of time for evolution. *Proceedings of the National Academy of Sciences*, 2009.