RESEARCH STATEMENT
FALL 2008
STEPHEN J. YOUNG

In many ways my current research efforts were shaped by my experiences at my undergraduate institution, the Rose-Hulman Institute of Technology. Rose-Hulman is a small, primarily undergraduate, engineering school in rural Indiana. As a result, the mathematics department offerings tended towards the eclectic with courses in the traditional core of research mathematics rarely offered. These courses were instead replaced by a motley collection of theoretical and applied courses which would appeal to the engineering student body. As a result of Rose's engineering focus, in addition to the rather unusual collection of courses which I took as an undergrad, there was also a philosophical viewpoint towards problem solving I absorbed in my time at Rose. Because of this focus, there was an emphasis on group/collaborative work throughout the curriculum. Even on assignments that were ostensibly individual, the expectation was that although all the work would be your own, you would have discussed/consulted with your fellow students.

The range of mathematics courses at Rose led me naturally to the program in Algorithms, Combinatorics, and Optimization (ACO) at the Georgia Institute of Technology. The ACO program is an interdisciplinary degree-granting program offered jointly by the School of Mathematics, the School of Computer Scinece, and the H. Milton Stewart School of Industrial and Systems Engineering. However, the interdisciplinary nature of the program did pose some problems for me, as I had too many options for potential advisors and thus had a hard time settling on one. Ultimately, it was a chance encounter with Prof. Ed Scheinerman of Johns Hopkins University that led me to my thesis topic and finalized my choice of advisor. Prof. Scheinerman described to me his recent work with Kraetzl and Nickel on the random dot product graph model [19]. In this model, they considered a random graph formed by assigning to every vertex a single value, X, distributed on $[0, 1]$ such that $\mathbb{P}(X \leq t) = t^{1/\alpha}$ for some $\alpha > 0$. Each edge was then present independently with probability equal to the product of the endpoints. They were able to show that this model exhibits positive clustering, has a power-law-like degree distribution for large degrees, and the giant component has diameter at most 5. In describing this model, Prof. Scheinerman brought to my attention several extensions of the model whose behavior he wished to resolve. Specifically, can X be replaced by a vector and multiplication be replaced by the dot product; can the model be modified so as to be directed; and can the model be generalized to a sparse(r) model?

In response to Prof. Scheinerman's questions, independently and in collaboration with my advisor Prof. Milena Mihail and Prof. Scheinerman, I have developed answers to these questions. Specifically, we generalized the model to the following: Let μ be a distribution on $\mathbb{R}^d$ satisfying mild technical conditions and let $g(n)$ be a monotonically increasing function. Then to each vertex v assign a vector X_v distributed according to μ. Each edge $\{u, v\}$ is then present independently with probability $\langle X_u, X_v \rangle / g(n)$. With this generalized model, we were able to show that unless μ is concentrated on a single point, the resulting graph exhibits positive clustering. By proving a new concentration result for binomial random variables with random parameters, we were also able to describe the relationship between μ and the log-histogram of the degree distribution. We also characterized the diameter of the model by reducing the problem to the asymptotic behavior of a class of Erdős-Rényi models. To our knowledge, this is the first diameter result that has proceeded via a reduction rather than through a direct calculation. Combining these results with the natural semantic interpretation of the vectors assigned to vertices, we feel that this model could be a significant advance in the study of complex networks and their models.

In the course of developing these results, Prof. Mihail and I came across several additional algorithmic and structural questions regarding the random dot product graph model. First, what is the spectral gap and conductance of the model and can it be finely controlled? Many algorithmic properties of complex networks such as the Internet can be explained by the underlying network exhibiting "good" spectral gap/conductance. For instance, the ability to effectively calculate PageRank$^{\text{TM}}$ on the World Wide Web is a consequence of the spectral properties of the network. Additionally, the ability of the physical layer of the Internet to transmit massive amounts of data with minimal congestion can be explained, in part, by the conductance properties of the network [10, 24]. While attempting to determine the spectral gap/conductance properties of the random dot product graph, I discovered an extension of an earlier result of Chung, Lu, and Vu. In [6] they use Wigner's high moment method and the Füredi and Komlós bound on the number of a specific type of closed walks [9] to show that the expected degree sequence model, with mild restrictions on the degree sequence, has a large spectral gap. I extend these results to random graphs where each edge is present independently with arbitrary probability and show that the asymptotic behavior of the spectral gap can be calculated explicitly from the spectrum of a deterministic matrix. We are able to use this result to show that the recently-introduced Stochastic Kronecker Graph model of Leskovec, Chakrabarti, Kleinberg and Faloutsos [20, 21] has large spectral gap. Unfortunately, we do not currently believe that this result will extend to resolve the conductance/spectral gap properties of the random dot product graph model. However, from a modeling point of view, this is preferable, as recent studies by Leskovec, Lang, Anirban, and Mahoney [22, 23] of the World Wide Web have shown that the Web does not have universally good conductance/spectral gap properties. Rather, there is a core section of the Web that has good spectral gap/conductance properties and then there are small(er) communities that are loosely connected to the core of the network and, within the community, have good spectral gap/conductance properties. Currently the only way to effectively model these sparse cuts is to artificially "plant" them within the network rather than having them arise organically from the parameters of the model. In contrast, we strongly believe that such sparse cuts can arise organically from the random dot product graph or some generalization thereof.

In addition to the issue of the conductance of the random dot product graph model, there are several other algorithmic and practical application questions we are working towards answering. For instance, we were able to show that the assortativity ($\sum_{\{u,v\}\in E(G)} \deg(u)\deg(v)$) of the random dot product graph model is larger than that of the Erdős-Rényi model. However, there are some real world networks, such as the physical level of the Internet, which exhibit assortativity less than the Erdős-Rényi model with same edge density. Is there some way of modifying/generalizing the random dot product graph model to capture this behavior? From a practical application point of view, the most pressing question is how should the distribution μ be derived from real world data? One potential direction we are exploring is to "fit," in some manner, vectors to the vertices of a real world network and then use a noisy sampling method such as that described by Hörmann and Leydold [12]. In fact, we will soon begin a collaboration with Prof. Joel Sokol to explore the applicability of the random dot product graph model to modeling the spread of disease using data collected by the Centers for Disease Control.

One of the seminal works in the study of complex networks was Stanley Milgram's 1967 experiment showing that there exist short paths in the social network of the time [25], that is, there are "six degrees of separation." Recently, Kleinberg has shed new light on this work by showing that there exist geometric networks where there are short paths but they cannot be found efficiently in a decentralized manner [17, 18]. However, we know from Milgram's experiments that these short paths can be found in a decentralized manner within a social network. Thus, a natural question is whether the geometry of the random dot product graph model can be exploited in such a manner that the network can be efficiently navigated.

Independently, Bollobás, Janson, and Riordan developed a random graph model that has a very similar flavor and behavior with respect to the diameter, degree distribution, and clustering [3, 4] as the random dot product graph model. In their model, there is a significant relaxation in terms of the conditions placed on μ and the ability to replace the inner product with a general symmetric positive valued function. However, to achieve this additional generality, they must sacrifice some generality in terms of the function $g(n)$. In fact, a key factor in their analysis is that $g(n)$ is precisely n. Thus, in some sense, the random dot product graph model and the inhomogeneous graph model of Bollobás, Janson, and Riordan are orthogonal models in the same space of random graph models. We naturally wonder whether similar results hold for the "convex combination" of these two models. Although we believe these results should hold in the common generalization of both models, we believe that neither of the proof techniques will suffice in the general case.

During one of the times when we were struggling to understand the behavior of the random dot product graph model, Prof. Mihail suggested that it my be profitable to develop my combinatorial intuition by working with Prof. Tom Trotter. After discussing the possibility with Prof. Trotter, I began sitting in on his research group's weekly meetings. At first I was more of a silent observer as I acclimated myself to the study of the combinatorics of partially ordered sets (posets). Although this collaboration has not yielded a unified research direction as with my work on models for complex networks, it has provided me with several diverging directions for future work.

My first real contribution to the research of the group sprung out the recent work of one of Prof. Trotter's graduate students, Dave Howard. He had been working on the linear discrepancy of posets and had some results partially characterizing the forbidden posets for linear discrepancy at most two [13]. The linear discrepancy of a poset can be most succinctly characterized as the bandwidth of the co-comparability graph, where the bandwidth of a graph is the minimum over all bijections from the vertices to $\{1, 2, \ldots, |V(G)|\}$ of the maximum distance between endpoints of an edge [7]. Together with Mitch Keller, another of Prof. Trotter's students, we were able to extend the width two characterization to completely characterize the forbidden posets for linear discrepancy at most two [14]. Mitch and I were also able to show that there is a Brooks-type theorem for the bandwidth of interval graphs [16]. In the course of our research, we discovered that the behavior of the critical pairs of a poset is fundamentally related to the linear discrepancy of a poset. This is surprising as critical pairs are more commonly seen in the context of the dimension of the poset where their behavior captures the essence of dimension. This lends credence to the conjecture of Trotter that if P is a poset of dimension at least five then the linear discrepancy is at least the dimension [28]. We also hope this insight will be helpful in resolving the conjecture of Tanenbaum, Trenk, and Fishburn that the linear discrepancy of a poset is at most $\lfloor {}^{3\Delta-1}/_{2} \rfloor$ where Δ is the maximum degree in the co-comparability graph [27]. Currently we are working with Prof. Ann Trenk of Wellesley College on generalizations of linear (and weak) discrepancy which we hope will lead to a better bound on the linear discrepancy of a poset.

The second direction of research stemming from my involvement with Prof. Trotter's research group originates outside the group with the work of Herzog, Vladoiu, and Zheng [11]. They show that the Stanley depth of a monomial ideal within a commutative ring $K[x_1, x_2, \ldots, x_n]$ (K a field) can be computed by considering a certain class of partitions of a finite subposet of the n-dimensional lattice. In the case of the maximal ideal, this subposet has nice structural properties which they conjecture can be exploited to show that the Stanley depth is precisely $\lceil {}^{n}/_{2} \rceil$. Biró, Howard, Keller, Trotter, and I were able to resolve this conjecture in the affirmative by constructing the desired partition in two different manners, one proceeds inductively with n and the other proceeds by giving an explicit construction for each n [2]. The natural next step would be to consider non-maximal ideals and the resulting combinatorial structures.

The final line of research I wish to mention that has resulted from my collaboration with Prof. Trotter and his graduate students is the study of the height sequence of posets. Given a poset P and an element x the height sequence of x is $h_0, h_1, h_2, \cdots$, where h_i is the number of linear extensions of P where x has height i. In 1981 Stanley was able to prove that for any poset P and any point x in the poset the height sequence is log-concave, that is, $h_i h_{i+2} \leq h_{i+1}^2$ [26]. However, in order to prove the log-concavity result, Stanley applies the Alexsandrov-Fenchel inequalities for mixed volumes, a powerful theorem from algebraic geometry. Aesthetically, there is no reason for the Alexsandrov-Fenchel inequalities to be essential to the proof of a purely combinatorial results. Thus it is hoped that there is either some, as yet undiscovered, combinatorial interpretation of the mixed volumes inequalities or that there is some proof of the log-concavity of the height sequence that circumvents the application of the Alexsandrov-Fenchel inequalities. Fortunately, from an aesthetic point of view, there is some indication that the latter may be true. Specifically, recently Trotter and Biró have developed a combinatorial proof of the log-concavity of the height sequence when the points incomparable to x are either a chain or a two element antichain. The fundamental idea behind their proof is the construction of a distributive lattice that is a variation on the Shepp lattice where each element of the lattice expresses a linear extension of the original poset in a many-to-one fashion. To this lattice they then apply the inequality of Fourtuin, Kasteleyn, and Ginibre [8] (the FKG inequality) to yield the log-concavity result and an additional understanding of the tightness of the inequality. Now inspired by this, it may be possible to construct a similar Shepp-style distributive lattice so that by the application of the FKG inequality, or its generalization, the Ahlswede-Daykin Four Functions Theorem [1], the general log-concavity result, will follow complete with an understanding of the conditions on the tightness of the inequality.

Although such a result would be appealing from an aesthetic point of view, it would have broader implications as a potential means of attacking the $1/3$–$2/3$ conjecture and other correlation inequalities. More specifically, we consider the proof of Brightwell, Felsner, and Trotter [5] that in any poset that is not a chain, there exist two incomparable points x and y such that the fraction of linear extensions where x is below y is between $5-\sqrt{5}/10$ and $5+\sqrt{5}/10$. This is a generalization of an earlier proof by Kahn and Saks [15] where the bounds were $3/11$ and $8/11$. One of the key steps in both of these results is showing that a certain sequence that is similar to the height sequence is log-concave. In both cases this is achieved by applying the Alexsandrov-Fenchel inequalities for mixed volumes. Thus, it is hoped that by developing a combinatorial proof of the log-concavity of the height sequence we will create a series of techniques that will allow us to avoid the application of the mixed volume inequalities in the proofs working towards the $1/3$–$2/3$ conjecture and make progress towards resolving the conjecture.

In summary, over the course of the next few years I see myself pursuing two major avenues of research. The first being the continuing generalization and refinement of the random dot product graph model, with an emphasis on generalizing the nature of the distribution μ, especially allowing variations in μ with respect to n, and replacing the dot product by a more general similarity function perhaps also varying with n in some reasonable manner. The second would be the construction of distributive lattices which can be analyzed via the Ahlswede-Daykin Four Functions Theorem to yield log-concavity results for the height sequence or other height related sequences. In addition to these major avenues of research, I will continue to pursue the less long-range avenues of research such as the combinatorial interpretation of Stanley depth, linear/weak discrepancy and its generalizations, and the navigation and conductance of the random dot product graphs. However, I will continue to be open to new avenues of research that I come across. In fact, based on my experience in the past few years of graduate school, I will spend a significant amount of time working to solve other's problems in addition to continuing to pursue my own.

References

[1] R. Ahlswede and D. E. Daykin, *An inequality for the weights of two families of sets, their unions and intersections*, Z. Wahrsch. Verw. Gebiete, 43 (1978), pp. 183–185.

[2] C. Biró, D. M. Howard, M. T. Keller, W. T. Trotter, and S. J. Young, *Interval partitions and stanley depth*. submitted.

[3] B. Bollobás, S. Janson, and O. Riordan, *The phase transition in inhomogeneous random graphs*, Random Structures Algorithms, 31 (2007), pp. 3–122.

[4] ———, *Sparse random graphs with clustering*. preprint, August 2008.

[5] G. R. Brightwell and W. T. Trotter, *A combinatorial approach to correlation inequalities*, Discrete Math., 257 (2002), pp. 311–327. Kleitman and combinatorics: a celebration (Cambridge, MA, 1999).

[6] F. Chung, L. Lu, and V. Vu, *The spectra of random graphs with given expected degrees*, Internet Math., 1 (2004), pp. 257–275.

[7] P. C. Fishburn, P. J. Tanenbaum, and A. N. Trenk, *Linear discrepancy and bandwidth*, Order, 18 (2001), pp. 237–245.

[8] C. M. Fortuin, P. W. Kasteleyn, and J. Ginibre, *Correlation inequalities on some partially ordered sets*, Comm. Math. Phys., 22 (1971), pp. 89–103.

[9] Z. Füredi and J. Komlós, *The eigenvalues of random symmetric matrices*, Combinatorica, 1 (1981), pp. 233–241.

[10] C. Gkantsidis, M. Mihail, and A. Saberi, *Conductance and congestion in power law graphs*, SIGMETRICS Perform. Eval. Rev., 31 (2003), pp. 148–159.

[11] J. Herzog, M. Vladoiu, and X. Zheng, *How to compute the stanley depth of a monomial ideal*. to appear, 2008.

[12] W. Hörmann and J. Leydold, *Automatic random variate generation for simulation input*, Simulation Conference Proceedings, 2000. Winter, 1 (2000), pp. 675–682 vol.1.

[13] D. M. Howard, G.-B. Chae, M. Cheong, and S.-M. Kim, *Irreducible width 2 posets of linear discrepancy 3*, Order, 25 (2008), pp. 105–119.

[14] D. M. Howard, M. T. Keller, and S. J. Young, *A characterization of partially ordered sets with linear discrepancy equal to 2*, Order, 24 (2007), pp. 139–153.

[15] J. Kahn and M. Saks, *Balancing poset extensions*, Order, 1 (1984), pp. 113–126.

[16] M. T. Keller and S. J. Young, *A brooks-type theorem for the bandwidth of interval graphs*. submitted.

[17] J. Kleinberg, *The small-world phenomenon: an algorithm perspective*, in STOC '00: Proceedings of the thirty-second annual ACM symposium on Theory of computing, New York, NY, USA, 2000, ACM, pp. 163–170.

[18] J. Kleinberg, *Complex networks and decentralized search algorithms*, in International Congress of Mathematicians. Vol. III, Eur. Math. Soc., Zürich, 2006, pp. 1019–1044.

[19] M. Kraetzl, C. Nickel, and E. R. Scheinerman, *Random dot product graphs: A model for social networks*. 2005.

[20] J. Leskovec, D. Chakrabarti, J. Kleinberg, and C. Faloutsos, *Realistic, mathematically tractable graph generation and evolution, using kronecker multiplication*, in European Conference on Principles and Practice of Knowledge Discovery in Database, 2005.

[21] J. Leskovec and C. Faloutsos, *Scalable modeling of real graphs using kronecker multiplication*, in ICML '07: Proceedings of the 24th international conference on Machine learning, New York, NY, USA, 2007, ACM, pp. 497–504.

[22] J. Leskovec, K. J. Lang, A. Dasgupta, and M. W. Mahoney, *Community structure in large networks: Natural cluster sizes and the absence of large well-define clusters*, tech. rep., 2008.

[23] J. Leskovec, K. J. Lang, A. Dasgupta, and M. W. Mahoney, *Statistical properties of community structure in large social and information networks*, in WWW '08: Proceeding of the 17th international conference on World Wide Web, New York, NY, USA, 2008, ACM, pp. 695–704.

[24] M. Mihail, C. Papadimitriou, and A. Saberi, *On certain connectivity properties of the internet topology*, J. Comput. System Sci., 72 (2006), pp. 239–251.

[25] S. Milgram, *The small world problem*, Psychology Today, 1 (1967), pp. 60–67.

[26] R. P. Stanley, *Two combinatorial applications of the Aleksandrov-Fenchel inequalities*, J. Combin. Theory Ser. A, 31 (1981), pp. 56–65.

[27] P. J. Tanenbaum, A. N. Trenk, and P. C. Fishburn, *Linear discrepancy and weak discrepancy of partially ordered sets*, Order, 18 (2001), pp. 201–225.

[28] W. T. Trotter, *Problem 383 in research problems [Discrete mathematics]*, Discrete Math., 257 (2002), pp. 599–624. Kleitman and combinatorics: a celebration (Cambridge, MA, 1999).