Research Statement

Steven Kay Butler

My research has been focused primarily on using spectral techniques in graph theory, but has included various other topics such as mathematical juggling, induced universal graphs, algebraic combinatorics, game theory, and mathematical juggling. In this research statement I will provide a summary of my past and current research in these fields, as well as state some open problems that I will pursue in future research.

1 Quasirandom graphs and edge discrepancy

Graph theory has gotten increased attention in the last ten years with the need to examine structures involving large data sets. A classic example is the internet which is composed of pages (vertices) and hyper-links (edges). A number of different properties of graphs have been defined in the literature (e.g., Hamiltonian, connected, bridgeless, etc.) and many papers are of the form "a graph having property A must also have property B". Chung, Graham and Wilson [18] noticed that a large number of nontrivial properties are related in the sense that if a graph satisfies any single property it must needs satisfy them all, they termed these properties quasirandom.

We will make mention of three quasirandom properties here (without being too rigorous in our definition): (i) any small graph shows up as a subgraph approximately the expected number of times; (ii) the nontrivial eigenvalues of the adjacency matrix are small; (iii) for most pairs of vertices u and v the number of vertices which are adjacent to both u and v or are adjacent to neither u and v is about half of the vertices.

A closely related property is discrepancy, which is a measurement of how "randomly" edges are placed between various subsets of graphs. Namely, the discrepancy of a graph is the minimal α so that for all subsets X and Y

$$\left| e(X,Y) - \frac{\operatorname{vol}(X)\operatorname{vol}(Y)}{\operatorname{vol}(G)} \right| \le \alpha\sqrt{\operatorname{vol}(X)\operatorname{vol}(Y)},$$

where $\operatorname{vol}(X)$ denotes the sum of the degrees for a subset X of the vertices and $e(X,Y)$ is the number of edges with the first endpoint in X and the second endpoint in Y.

It has been known for some time that $\alpha \le \lambda_2$, so that if the graph had a tight bound on eigenvalues then it also has a tight control on discrepancy. Recently Bilu and Linial [3] were able to show that for d-regular graphs that λ_2 could be bounded in terms of α in such a way that a small discrepancy implied a small bound on the spectral gap. By taking their approach and mixing it with the approach of Bollobás and Nikiforov [4] I was able to generalize this result to all graphs (more generally all nonnegative matrices) and established the following for directed graphs [6, 7].

Theorem 1. *For a directed graph G let A denote its adjacency matrix D_{in} the diagonal in-degree matrix and D_{out} the diagonal out-degree matrix. Let $e(X\to Y)$ count the number of directed edges starting in X and ending in Y while $\operatorname{vol}_{in}(X)$ and $\operatorname{vol}_{out}(X)$ denote respectively the sum of the in and out degrees of vertices of the set X. Then if $\operatorname{disc}(G)$ is the minimal α so that for all X and Y*

$$\left| e(X\to Y) - \frac{\operatorname{vol}_{out}(X)\operatorname{vol}_{in}(Y)}{\operatorname{vol}(G)} \right| \le \alpha\sqrt{\operatorname{vol}_{out}(X)\operatorname{vol}_{in}(Y)},$$

it follows that

$$\mathrm{disc}(G) \le \sigma_2(D_{out}^{-1/2} A D_{in}^{-1/2}) \le 150\,\mathrm{disc}(G)\big(1 - 8\log(\mathrm{disc}(G))\big),$$

where $\sigma_2(M)$ denotes the second largest singular value of matrix M.

Thus the discrepancy of a directed graph G and the second singular value of the normalized adjacency matrix of a directed graph both lie in the same quasirandom class of properties of directed graphs. One of the most important quasirandom properties for undirected graphs mentioned above is that it contains each small graph the expected number of times. It is easy to show that this property is not in the same quasirandom class and is a stronger property in that it implies the small discrepancy and small second singular value, so we call the graph containment a strong quasirandom property while the other two will be termed weak quasirandom properties.

My future research will look at more clearly defining these two quasirandom classes. In particular, what additional assumptions are needed for weak quasirandomness to imply strong quasirandomness? [Comparing to the undirected case the problem seems to lie in property (iii) mentioned above. It is easy to show that when the graph is undirected that this reduces to a condition on the eigenvalues but for the corresponding statement when the graph is directed it does not.]

Further research in the area of discrepancy is also planned for the problem of discrepancy in hypergraphs. As noted above the discrepancy result follows from a matrix result for nonnegative (not necessarily square) matrices. There are a number of ways to use this result to define discrepancy, it remains to work out which ones are useful in application and can be related in a natural way to other properties of hypergraphs.

2 Eigenvalues interlacing and graph structure

Eigenvalues of graphs can be used to give information about the structure of a graph, and vice-versa. Therefore graphs which are structurally "close" should have eigenvalues that are also "close" in some sense.

One idea of being close is that two graphs differ by either the removal or addition of a small number of edges. Recently Chen et al. [17] were able to give an interlacing result for the normalized Laplacians for simple graphs where the eigenvalues can spread by at most the number of edges removed. (The normalized Laplacian of a graph $\mathcal{L}(G)$ is $D^{-1/2}(D-A)D^{-1/2}$ where, as before, D is the diagonal degree matrix and A the adjacency matrix of the graph, more information about the normalized Laplacian is found in Chung [19].) I was able to push their result to work for arbitrary weighted graphs and show that the spread is at most the number of *vertices* in the subgraph being removed (an improvement when removing dense graphs).

Theorem 2 ([8])**.** *Let G be a graph and H a subgraph of G with t nonisolated vertices. If $\lambda_0 \le \lambda_1 \le \cdots \le \lambda_{n-1}$ and $\theta_0 \le \theta_1 \le \cdots \le \theta_{n-1}$ are the eigenvalues of $\mathcal{L}(G)$ and $\mathcal{L}(G-H)$ respectively then for $k = 0, 1, \ldots, n-1$ we have*

$$\lambda_{k-t+1} \le \theta_k \le \begin{cases} \lambda_{k+t-1} & H \text{ is bipartite,} \\ \lambda_{k+t} & \text{otherwise,} \end{cases}$$

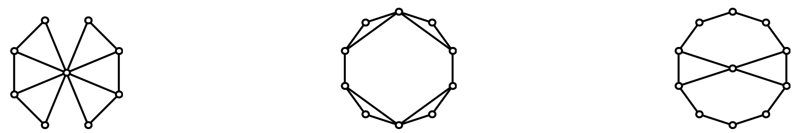
where $\lambda_{-t+1} = \cdots = \lambda_{-1} = 0$ and $\lambda_n = \cdots = \lambda_{n+t-1} = 2$.

In other words by inserting/removing a local structure the eigenvalues will not shift very far. The original motivation for examining this was considering an interlacing problem for the Laplacian of directed graphs defined by Chung [20], while some progress towards finding such an interlacing result were given in [8] it still is not completely known what can be said for this situation. More research into the directed Laplacian is planned.

3 Eigenvalues and coverings

Some graphs can share local similar structure, this happens for instance when one graph "covers" another and again we would expect the eigenvalues of the graphs to be related. A classical example of a covering coming from topology is when S^1 double covers itself by the map $e^{i\theta} \mapsto e^{2i\theta}$.

One special type of covering I have examined [9] is termed a 2-edge-covering. As the name indicates this is a structure preserving homomorphism where each edge in the larger graph covers two edges in the smaller graph. My interest from this was examining graphs that shared many eigenvalues but did not cover the same graph. For instance, consider the three graphs shown below.



By a calculation it can be seen that for both the adjacency matrix and the normalized Laplacian these graphs share four common eigenvalues. In this case it turns out that these eigenvalues do not correspond to a common cover but rather can be found by use of an "anti-cover" graph (used to find the eigenvalues not found in the covering). The examination of these anti-covers leads to considering the normalized Laplacian for graphs with negative edge weights and vertex weights. Future research will be looking at the normalized Laplacians for such graphs and looking for applications.

Besides 2-edge-coverings one could consider k-edge-coverings and other types of coverings that preserve structure. These coverings can help to explain how spectras of graphs relate and why some graphs share many eigenvalues in common.

4 Work in several other areas

In addition to problems related to eigenvalues of graphs mentioned in the preceding sections, I have also looked at a variety of other combinatorial problems. In this section I will mention some of my favorite problems, results I have obtained and further research that I will be pursuing along these lines.

4.1 Induced universal graphs

Given a family of graphs $\mathcal{F}$ a graph U is universal if it contains each graph of $\mathcal{F}$ as a subgraph. The idea being that we are packing the family of $\mathcal{F}$ into one universal structure. For example, if $\mathcal{F}$ is the family of all graphs on n vertices then U can be taken to be the complete graph on n vertices since it contains every such graph as a subgraph.

Problems on universal graphs deal with trying to find the minimum size of a universal graph for a family subject to some constraints. An example of a constraint is that the graphs show up as *induced* subgraphs of U which then requires that the packing be a little "loose" (so the previous example with the complete graph now fails miserably). Recently Alon and Capalbo [1] made progress in constructing small universal graphs for the family of bounded-degree graphs. I considered the same family of graphs with the additional constraint that the graphs were induced subgraphs and was able to establish the following result [10] (it should be noted that the Alon-Capalbo result was heavily dependent on the probabilistic method while the methods that I used were completely different and relied on elementary methods to produce an explicit construction and embedding of all graphs).

Theorem 3. *Let $\mathcal{F}$ be the family of all graphs with maximum degree at most r on n vertices. Then there exists a graph U so that every graph in $\mathcal{F}$ is an induced subgraph of U and*

$$|V(U)| \le Cn^{\lfloor (r+1)/2 \rfloor}, \qquad \textit{and} \qquad |E(U)| \le Dn^{2\lfloor (r+1)/2 \rfloor -1},$$

where C and D are constants which depend only on r. Further for r even the number of vertices is within a constant factor of the minimal number of vertices.

The same construction used for this family also worked to give "induced" universal graphs for multigraphs and directed graphs with bounded degree. For the case when $r = 3$ a slightly better construction has been found [2] using probabilistic tools, and the general odd case is still open. An interesting problem which I am investigating is looking for similar results for hypergraphs. As a step in this direction it will be useful to search for a hypergraph decomposition result similar to Petersen's Theorem.

4.2 Graphs associated with permutations

Answering a question of Woo and Yong [22] I was able to prove a conjecture which shows that permutations which avoid the patterns 1324 and 2(14)3 characterize Schubert varieties which are locally factorial. These permutations also have an interesting characterization in terms of graphs. Given a permutation π we can construct a graph G_π on the vertices $\{1, 2, \ldots, n\}$ by joining i and j if and only if (i) $i < j$ and $\pi(i) < \pi(j)$ and (ii) there is no k such that $i < k < j$ and $\pi(i) < \pi(k) < \pi(j)$. This second property is closely related to Bruhat ordering (more information about a generalization of this construction can be found in [21]).

Theorem 4. *A permutation π avoids the pattern* 1324 *and* 2(14)3 *if and only if the graph G_π is a forest.*

In joint work with Mireille Bousquet-Mélou [5] we were able to give generating functions and asymptotics for the number of such permutations. There are of course many interesting questions that can be asked about this construction, perhaps the most interesting is what graphs can be realized by a G_π for some permutation π. From the definition it easily follows that the graph must be triangle free but this is not sufficient since an exhaustive search on permutations on 8 elements show that there is no G_π which is the 3-cube (Q_3). In future research I will return to this question of realizability.

4.3 Hat guessing games

Hat guessing games give a simple way to present a challenging mathematical problem. Consider the following game, there are n players who will soon have hats placed on their heads, each hat one of k different colors. Before the hats are placed the players meet together to form a deterministic strategy (i.e., each persons guess is determined by the hats they see); overhearing this strategy though is the person who is putting the hats on their head and so they need to form a strategy which will maximize the number of correct guesses given that the strategy is publicly known. What should the players do and what is the maximum number of correct guesses they can achieve? It turns out that the maximum number of correct guesses that can be guaranteed is $\lfloor n/k \rfloor$ and it is easy to form a strategy achieving this bound.

In joint work with Mohammad Hajiaghayi, Robert Kleinberg and Tom Leighton we considered variations of this game. One version of this game is that players may not have perfect information, i.e., they may only get to see a subset of the other players. Let G be the graph with a directed arc between player i and j if player i will be able to see player j's hat. In the case when information is symmetric and there are 2 colors it is easy to show that the maximum number of correct guesses corresponds to the size of the largest matching. For the directed case the problem is more interesting and only bounds are known, neither of which is sharp.

Theorem 5. *The maximum number of correct guesses is at least as large as the maximum number of disjoint cycles in the graph and is no more than the minimum number of vertices whose removal leaves the graph acyclic.*

An interesting problem which I will be looking at in future research is to see if there is some simple (to compute) graph parameter which will give the maximum number of correct guesses. Another variation of the game is to restrict the hat supply of the adversary (now giving an edge to the players). For instance if there are three players and two blue hats and two red hats it can be shown that if the players form a directed three cycle and guess the opposite color of what they see in front of them then they will always have 2 correct guesses (an improvement over their previous 1).

Theorem 6. *If there are a_i hats of color i for $i = 1, 2, \ldots, k$ then the maximum number of correct guesses is bounded above by*

$$\left\lfloor \sum_{\substack{b_i \le a_i,\, 1\le i\le k \\ b_1+\cdots+b_k=n-1}} n\binom{n-1}{b_1,\ldots,b_k} \Big/ \sum_{\substack{b_i \le a_i,\, 1\le i\le k \\ b_1+\cdots+b_k=n}} \binom{n}{b_1,\ldots,b_k} \right\rfloor.$$

I conjecture that this bound is actually tight (and in several special cases it is known to be tight). This problem reduces to looking at marking edges in a hyper-hypercube (a hypergraph version of the hypercube) and essentially the conjecture states that the hyper-hypercube has no obstructions in the marking of special sets to keep the markings balanced simultaneously at each vertex. Given the fundamental nature of the hypercube in combinatorics this problem on the hyper-hypercube will be an important avenue for my future research.

4.4 Additional research

In addition to the areas mentioned above I have also looked at mathematical juggling in joint work with Ron Graham [12], liar games in joint work with Jia Mao and Ron Graham [13], Hamiltonicity of graphs in joint work with Fan Chung [14] as well as several other mathematical problems [15, 16]. All of my papers mentioned above are available online at `http://www.math.ucsd.edu/~sbutler/`.

References

[1] N. Alon and M. Capalbo, Sparse universal graphs for bounded degree-graphs, *Random Structures & Algorithms* **31** (2007), 123–133.

[2] N. Alon and M. Capalbo, private communication.

[3] Y. Bilu and N. Linail, Constructing exapnder graphs by 2-lifts and discrepancy vs. spectral gap, *FOCS 2004 Proceedings*, 404–412.

[4] B. Bollobás and V. Nikiforov, Hermetian matrices and graphs: singular values and discrepancy, *Discrete Mathematics* **285** (2004), 17–32.

[5] M. Bousquet-Mélou and S. Butler, Forest-like permutations, to appear in *Annals of Combinatorics*.

[6] S. Butler, Using discrepancy to control singular values for nonnegative matrices, *Linear Algebra and its Applications* **419** (2006), 486–493.

[7] S. Butler, Relating singular values and discrepancy of weighted directed graphs, *Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm (Miami, FL, 2006)*, 1112–1116.

[8] S. Butler, Interlacing for weighted graphs using the normalized Laplacian, *Electronic Journal of Linear Algebra* **16** (2007), 90–98.

[9] S. Butler, Eigenvalues of 2-edge-coverings, preprint.

[10] S. Butler, Induced-universal graphs for graphs with bounded maximum degree, preprint.

[11] S. Butler, M. T. Hajiaghayi, R. D. Kleinberg and T. leighton, Hat guessing games, to appear in *SIAM Journal on Discrete Mathematics*.

[12] S. Butler and R. Graham, Enumerating (multiplex) juggling sequences, to appear in *Annals of Combinatorics*.

[13] S. Butler, R. Graham and J. Mao, How to play the majority game with liars, *AAIM 2007*, Lecture Notes in Computer Science **4508**, Springer-Verlag, 221–230.

[14] S. Butler and F. Chung, Small spectral gap in the combinatorial Laplacian implies Hamiltonian, to appear in *Annals of Combinatorics*.

[15] S. Butler, Estimating the number of graphs containing very long induced paths, to appear in *Ars Combinatoria*.

[16] S. Butler, Tangent line transformations, *The College Mathematics Journal* **34** (2003), 105–106.

[17] G. Chen, G. Davis, F. Hall, Z. Li, K. Patel and M. Stewart, An interlacing result on normalized Laplacians, *SIAM Journal on Discrete Mathematics* **18** (2004), 353–361.

[18] F. Chung, R. Graham, and R. Wilson, Quasi-random graphs, *Combinatorica* **9** (1989), 345–362.

[19] F. Chung, Spectral Graph Theory, CBMS Regional Conference Series in Mathematics, AMS, Providence, RI, 1997.

[20] F. Chung, Laplacians and the Cheeger inequality for directed graphs, *Annals of Combinatorics* **9** (2005), 1–19.

[21] P. Keevash, P.-S. Loh and B. Sudakov, Bounding the number of edges in permutations graphs, *Electronic Journal of Combinatorics* **13** (2006), #R44.

[22] A. Woo and A. Yong, When is a Schubert variety Gorenstein?, *Advances in Mathematics* **207** (2006), 205–220,