I am interested in the interactions of probability and analysis with combinatorics and discrete structures. These interactions arise naturally when trying to find structure and information in large growing networks such as the world wide web and social networks. My particular focus has been on studying the properties of random graphs. My research on component sizes in random directed graphs and some properties in random hypergraphs, and how the analyses of algorithms were important to proving these results, is explained below. I conclude with some continuing projects and other future directions for research as well as potential undergraduate research opportunities.

Random Graphs

Component sizes

First, I describe a classical result in undirected random graphs. Then I detail our result in random directed graphs and explain why the determination of the strong component sizes has been a difficult problem.

Of the many parameters of a graph, we often first want to know the sizes of its components. One of the best known models is $G(n,p)$, which has n vertices and each potential edge is present independently with probability p. A classical result due to Erdős and Rényi [ER60] is the following "double jump" phenomenon: let $\mathcal{L}$ denote the number of vertices in the largest component of $G(n,p)$. As customary, we say that $G(n, p = p_n)$ has some property *with high probability*, denoted *w.h.p.*, if the probability that $G(n,p)$ has this property tends to 1 as n tends to infinity.

- If $p = c/n$, $c < 1$, then w.h.p. $\mathcal{L}$ is on the order of $\ln n$.
- If $c = 1$, then w.h.p. $\mathcal{L}$ is on the order of $n^{2/3}$.
- If $c > 1$, then w.h.p. $\mathcal{L} = \alpha\, n + o(n)$, where α is the unique root in $(0,1)$ of $1 - \alpha = e^{-c\,\alpha}$. This component is called giant because the other components have size $O(\ln n)$.

This model is rather unrealistic for many real-world networks due to the independence of the possible edges, but $G(n,p)$ serves as a benchmark for and gives insight into understanding and gauging the difficulty of problems that one may face in other more constrained random graphs.

Researchers have struggled to to extend this classic result to the random directed graph $D(n,p)$, where each potential edge is present with probability p. It was not until 1990 that Karp [Ka90] and Łuczak [Luc90] established a similar phase transition for $\mathcal{L}$, the size of the largest strong component of $D(n, p = c/n)$. They proved that if $c < 1$, then $\mathcal{L}$ is bounded in probability and if $c > 1$, then w.h.p. $\mathcal{L} = \alpha^2 n + o(n)$.

In my thesis, I, along with my advisor B. Pittel, prove a Central Limit Theorem (CLT) for $\mathcal{L}$ if $c > 1$. That is, there is some $\sigma = \sigma(c) > 0$ such that for all $a < b$

fixed,

$$P\left(a \leq \frac{\mathcal{L} - \alpha^2 n}{\sigma n^{1/2}} \leq b\right) \to \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx.$$

In fact, we show that number of vertices and edges of the giant strong component have a jointly asymptotic Gaussian (multivariate normal) distribution. This result generalizes similar CLTs in undirected random graphs due to Stepanov [Ste70], Pittel [Pit90], and Pittel and Wormald [PW05].

Other than Karp and Łuczak's phase transition and our CLT, there are not many parameters known about the strong giant component, especially in light of the wealth of results about the size and properties of the giant component in random undirected graphs and hypergraphs. This is because the notion of connectivity morphs into two harder-to-handle notions of sink sets and source sets in directed graphs. The strong component containing some vertex v is the set of vertices that are both reachable from v **and** can reach v. In $D(n,p)$, the two search processes that find those vertices in a strong component with v are *interdependent*; this interdependence is very difficult to analyze and is a primary reason that considerably less is known about the strong components of $D(n,p)$.

To prove our CLT, we deal with the difficulty of interdependence of the search processes by introducing a deletion algorithm that considers only "local" information but terminates with a slightly larger subgraph called the $(1,1)$-core. A key lemma in our proof is that w.h.p. the $(1,1)$-core is essentially the largest strong component along with "few" peripheral vertices and edges (whose orders are dwarfed by the fluctuations of the giant component parameters). I discuss this deletion algorithm further in the section describing deletion algorithms.

Random Hypergraphs

I additionally prove three properties about random d-uniform random hypergraphs. In particular, I find where these random hypergraphs are likely to be k-connected and weak Hamiltonian as well as their diameter near the connectivity threshold. A d-uniform hypergraph is a pair (V,E), of a vertex set V and a (hyper)edge set E, where each hyperedge is a set of d vertices. In the random d-uniform model $H_d(n,p)$, there are n vertices and each potential edge of cardinality d is present with probability p. Note that for $d=2$, this model is simply $G(n,p)$. I provide some definitions before stating our results.

k-connectivity

There are multiple measures of the strength of connectivity of a connected graph and I considered k-connectivity. A (hyper)graph is k-(vertex)connected, if whenever $k-1$ vertices are deleted with their adjacent edges, the remaining graph is connected. Necessarily, a k-connected graph must have minimum degree at least k.

I prove that this trivial necessary condition that $\min\deg(H_d(n,p)) \geq k$ is in fact a statistically sufficient condition [2]. In particular, if $p = (d-1)!\frac{\ln n+(k-1)\ln\ln n+c}{n^{d-1}}$, then

$$P(H_d(n,p) \text{ is } k\text{-connected}) = P(\min\deg(H_d(n,p)) \geq k) + o(1) \to e^{-e^{-c}/(k-1)!}.$$

This is a generalization to hypergraphs of the k-connectivity result established by Erdős and Rényi [ER61].

In fact, I prove a stronger result which concerns the random hypergraph process, $\tilde{H}_d$. The hypergraph process is a random sequence of hypergraphs, which begins as a hypergraph with no edges and then edges are added one after another at random. I prove that w.h.p., at the precise moment this process loses its last vertex of degree less than k, the process also becomes k-connected. This result generalizes Bollobás and Thomason's [BT85] hitting time result for the graph process (albeit for fixed k).

Hamiltonicity

A number of researchers have studied the problem of determining the threshold value p such that w.h.p. a Hamilton cycle exists in $G(n,p)$. In generalizing to hypergraphs, the first question that arises is how to even define a hypergraph cycle. In one version, a cycle is a sequence $C = (v_0, e_1, v_1, \ldots, e_\ell, v_\ell = v_0)$, where $v_0, v_1, \ldots, v_{\ell-1}$ are distinct vertices, $e_1, \ldots, e_\ell$ are distinct edges, and $v_{i-1}, v_i \in e_i$ for each i. A *weak cycle* is defined similarly, except *edges* are allowed to be repeated.

I found that the main barrier to weak Hamiltonicity is the presence of isolated vertices [4]. In particular, if $p = (d-1)!\frac{\ln n+c}{n^{d-1}}$, then

$$P(H_d(n,p) \text{ is weak Hamiltonian}) = P(\min\deg(H_d(n,p)) \geq 1) + o(1) \to e^{-e^{-c}}.$$

The previous result is extension of similar results for the random graph models due to Korshunov [Kor77], Komlós and Szemerédi [KS83], Bollobás [Bol83].

Loose Connectivity and Diameter

A *loose path* in a hypergraph from v to w is a sequence of edges $e_1, e_2, \ldots, e_k$ such that $v \in e_1$, $w \in e_k$, $|e_i \cap e_{i+1}| = 1$ for each i, and $e_i \cap e_j = \emptyset$ if $j \neq i-1, i, i+1$. A hypergraph is *loosely connected* if between any two vertices there is a loose path between them. A loosely connected hypergraph is necessarily connected. By analyzing a version of a breadth-first search, I prove that w.h.p. at the moment the random hypergraph process loses its last isolated vertex, the process becomes loosely connected [3]. In fact, w.h.p. each pair of vertices is loosely connected by a path of length at most

$$\left[\frac{\ln n}{\ln((d-1)\ln n)} - 1, \frac{\ln n}{\ln((d-1)\ln n} + 9\right].$$

This result generalizes an analogous theorem due to Bollobás [Bol84] about the diameter of the graph process at the moment of connectivity.

Analysis of Algorithms

One key tool in the field of random graphs is the use of algorithms to find some desired structure. Many problems in combinatorics, such as finding the longest path in a graph or whether a graph has a Hamilton cycle, are (provably) very difficult to do for all graphs. However, by constraining the set of graphs and wanting to solve the "average-case" scenario, these difficult questions can be susceptible to attack. A common technique is to introduce randomness into an algorithm, through either random inputs or random decisions in each step, hoping to avoid the "worst-case scenario". I discuss two such algorithms that we introduced and how we used them in the proofs of some of the results discussed earlier.

Deletion Algorithms

As mentioned above, we proved that the $(1, 1)$-core of $D(n, p)$, which is the maximal subgraph with minimum in-degree and out-degree at least 1, is not much larger than its largest strong component. This is important because the $(1, 1)$-core can be found by a greedy deletion algorithm where semi-isolated vertices, those with either in-degree zero or out-degree zero, are successively deleted. Deterministically, no matter the order in which you delete these semi-isolated vertices, this deletion process ends with the $(1, 1)$-core. Although given an initial directed graph, the terminal directed graph produced from the algorithm is determined, we introduce randomness into our deletion algorithm; then, we prove our result using techniques from probability and differential equations.

Here is how we introduce and use randomness in the algorithm: first, at each step, we randomly chose the semi-isolated vertex to be deleted. By introducing this specific kind of randomness, we can show that the 4-tuple of the number of vertices, number of in-degree zero vertices, number of out-degree zero vertices, and number of edges after t steps form a Markov chain. In order to show that the terminal number of vertices and edges is indeed jointly asymptotically Gaussian, we "approximate" the likely realization of the deletion algorithm with a deterministic trajectory. Using this Fourier-based technique, this assumed approximation forces the mean and covariance parameters to satisfy a certain system of PDEs, which we convert to some systems of ODEs by the method of characteristics. Once these ODEs are "solved", by an exponential supermartingale argument, we prove our deterministic trajectory is in fact a close approximation to the actual deletion process, which nearly finishes the proof.

Edge-finding Algorithms

In the proof of establishing the weak Hamiltonicity threshold, we encountered the probability in $H_d(n, p)$ that a set of vertices, say B, are all adjacent to at least one vertex from another disjoint set of vertices, say A; let's denote this probability by $P(A, B)$. For the random graph $G(n, p)$, $P(A, B)$ is easily computable, but for hypergraphs, there is interdependence with the events $\{v \text{ is adjacent to } A\}$ for different

$v \in B$. Needing to take on this interdependence, we introduced an edge-finding algorithm that necessarily discovers at least $|B|/(d-1)$ edges on the event corresponding to $P(A,B)$. Analyzing this deterministic algorithm (with random inputs), we find that

$$P(A,B) \leq \left(1-(1-p)^{\binom{|A|+|B|}{d-1}-\binom{|B|}{d-1}}\right)^{\lceil \frac{|B|}{d-1} \rceil}.$$

Future Directions

The following are some continuing projects and open problems in which I am interested.

Evolution of random directed graphs

Many parameters of the strong giant component are still unknown. For example, the asymptotic joint distribution of the size of the giant strong component along with its ancestor and descendant set is unknown. Further, there are many properties known about the components of $G(n,p)$ for p through the critical window $(1+\epsilon)/n$, $\epsilon = O(n^{-1/3})$ (such as sizes of the components, the number of components with more edges than vertices, planarity, diameter, etc.). There should be analogues to these statements for directed graphs (this critical range is also near $p = 1/n$).

Hamiltonicity in random hypergraphs

The (sharp) threshold's location for where Berge Hamilton cycles begin to exist is still an open question. Further, there are other notions of Hamiltonicity in hypergraphs. The thresholds corresponding to Hamiltonian ℓ-overlapping cycles have been established, but there remains the sharp threshold determination (for most cases, see Dudek and Frieze [DF13]). Also, I would like to know if there is some sort of local obstruction, such as vertices of degree less than 2 for Hamiltonicity in $G(n,p)$, that prevents Hamilton cycles from forming in these random hypergraphs.

Cores

In $G(n,p)$, Pittel, Spencer and Wormald [PSW96] established the sharp thresholds for the appearance of the k-core, the maximal subgraph of minimum degree k, if it exists. As for its directed analogue, even the threshold of the existence of the (k_i, k_o)-core, the maximal subgraph with minimum in-degree at least k_i and minimum out-degree at least k_0, is yet to be determined. Boris Pittel and I have made recent gains on this problem and believe we are near proving the sharp threshold for the appearance of the (k_i, k_o)-core.

Recently, Krivelevich, Lubetzky and Sudakov [KLS14] proved that w.h.p. the k-core, for $k \geq 15$, is Hamiltonian at its birth. It is conjectured that this holds for $k \geq 3$ as well (trivially true for $k = 2$). Is the (k,k)-core Hamiltonian at its birth as well?

Kahn-Kalai Conjecture

As long as $p > e/n$, the *expected* number of Hamilton cycles in $G(n,p)$ tends to infinity, but not until $p > (\ln n)/n$ can we say that w.h.p. at least one Hamilton cycle exists. This gap of order $\ln n$ is conjectured by Kahn and Kalai [KK07] to be the worst possible order between the "expectation threshold" and the threshold for a relatively general class of properties.

Undergraduate Research Opportunities

Combinatorics and Random Graphs greatly lend themselves to undergraduate research due to the relatively low barrier of entry. Some of the most useful techniques in Random Graphs, such as Markov's and Chebyshev's inequalities, are usually taught in undergraduate probability classes. Although my previously mentioned research projects are beyond the scope of undergraduate research, there are plenty of smaller questions that are appropriate. These include:

- For what p values, do "small" Berge (weak Berge, ℓ-overlapping) cycles begin to emerge in $H_d(n,p)$? How many cycles are there and what are their lengths?
- When is the $(1,1)$-core of $D(n,p)$ w.h.p. Hamiltonian? More difficult: when does the (k,k)-core w.h.p. have k edge-disjoint Hamilton cycle?
- Is the (k,k)-core of $D(n, p = c/n)$, for c large enough, w.h.p. k-strongly connected?

References (own work)

[1] Pittel, B. and D. Poole (2015). Asymptotic distribution of the numbers of vertices and arcs of the giant strong component in sparse random digraphs. To appear in *Random Structures & Algorithms.*

[2] Poole, D. (2015). On the Strength of Connectedness of a Random Hypergraph. *Electronic Journal of Combinatorics*, 22(1), P1-69.

[3] Poole, D. (2014). Loose connectivity of a random hypergraph. *In preparation.*

[4] Poole, D. (2014). On Weak Hamiltonicity of a Random Hypergraph. *Submitted. Preprint available at arXiv 1410.7446*

References (others)

[Bol83] Bollobás, B. (1983). Almost all regular graphs are Hamiltonian, *Europ. J. Combinatorics* **4**, 97-106.

[Bol84] Bollobás, B. (1984). The evolution of sparse graphs. In *Graph Theory and Combinatorics*. Proc. Cambridge Conf. in honour of Paul Erdős (Bollobás B., Ed.) Academic Press, 35-57.

[BT85] Bollobás, B. and A.G. Thomason (1985). Random graphs of small order, In *Random Graphs*, Annals of Discr. Math., 47-97.

[DF13] Dudek, A. and A. Frieze (2013). Tight Hamilton cycles in random uniform hypergraphs, *Random Structures & Algorithms*, **42**, 374-385.

[ER60] Erdős, P. and A. Rényi (1960). On the evolution of random graphs, *Magyar Tud. Akad. Math. Kutató Int. Közl*, **5**, 17-61.

[ER61] Erdős, P. and A. Rényi (1961). On the strength of connected of a random graph, *Acta Mathematica Hungarica*, **12(1)**, 261-267.

[GT11] Graham, R. and E. Tressler (2011). Open problems in Euclidean Ramsey theory, In *Ramsey Theory*, 115-120. Birkhäuser Boston.

[KK07] Kahn, J. and G. Kalai (2007). Thresholds and Expectation Thresholds, *Combinatorics, Probability and Computing* **16**, 495-502.

[Ka90] Karp, R.M. (1990). The transitive closure of a random digraph, *Random Structures & Algorithms*, **1** 125-135.

[KS83] Komlós J. and Szemerédi, E. (1983). Limit distributions for the existence of Hamilton cycles in a random graph, *Discrete Mathematics*, **43**, 55-63.

[Kor77] Korshunov, A.D. (1977). A solution of a problem of P. Erdős and A. Rényi about Hamilton cycles in non-oriented graphs, *Soviet Mat. Dokl.* **17**, 760-764.

[KLS14] Krivelevich, M., Lubetzky, E. and B. Sudakov (2014). Cores of random graphs are born Hamiltonian, *Proceedings of the London Mathematical Society*, pdu003.

[Luc90] Łuczak, T. (1990). The phase transition in the evolution of random digraphs, *Journal of graph theory*, **14**, 131-138.

[Pit90] Pittel, B. (1990). On tree census and the giant component in sparse random graphs. *Random Structures & Algorithms*, **1**, 311-342.

[PSW96] Pittel, B., Spencer, J. and N. Wormald (1996). Sudden emergence of a giant k-core in a random graph, *Journal of Combinatorial Theory, Series B.*, **67**, 111-151.

[PW05] Counting connected graphs inside-out, *Journal of Combinatorial Theory, Series B*, **93**, 127-172.

[Ste70] Stepanov, V.E. (1970). Some features of connectedness of a random graph $G_m(t)$, *Theory Probab. Applics*, **15**, 55-67.