

# ABRAHAM D. FLAXMAN

## Research Statement

Roads, neurons, the Internet, social relationships, mycelia. Everywhere we look, we find complex networks. This has been the case since time immemorial. But only recently has technology allowed us to record the structure of these huge networks and use this information in decision making. Hyperlinking web pages together and ranking them based on network structure has already revolutionized the way we manage information (“Just google it”). This is only the beginning.

To benefit from our new wealth in networked data, we need to really understand networks. We need accurate measurements, relevant models, and efficient algorithms. This is the focus of my research. It sits at the intersection of mathematics and computer science, and incorporates elements of economics as well.

### **Current work: Network sampling, modeling, algorithms, and economics**

I use the term **network sampling** to refer to the process of gathering network data from the natural or artificial environment. The mathematical field of random graph theory was greatly invigorated by the observation of real-world graphs in the late 1990s [25, 6, 3]. In some of these works, such as the analysis of the power grid of the western United States, the network structure is known with little or no error. But in many other cases, we only have access to a noisy sample of the network. For example, the structure of the Autonomous Systems graph of the Internet is not accurately recorded anywhere, and the existence of links must be inferred from indirect measurements. As shown experimentally and theoretically, such network sampling introduces systematic bias in the popular network statistics of the degree distribution [23, 1]. My colleague Juan Vera and I have developed a bias-reduction technique to address this problem. It shows that power-law degree distributions truly appear in the AS graph and are not an artifact of the sampling procedure [22]. I am currently advising the master’s thesis of University of Washington student Elisa Celis on the related topic of sampling bias introduced in another form of network sampling that is widely used in gathering the web graph. This will have immediate implications for ranking web-search results.

**Network modeling** has exploded as an area in the last decade, expanding from a subfield of probabilistic combinatorics to an interdisciplinary area with contributions from fields as diverse as physics, biology, economics, and sociology, as well as math and computer science. My work here has focused on formulating models that capture important features of real-world graphs *and* can be subjected to rigorous mathematical proof. This includes extensive work on variants of the preferential attachment graph [15, 14, 7, 8] as well as a new approach to network modelling that I’ve developed with inspiration from smoothed analysis, which I call *Randomly Perturbed Graphs (RPGs)* [13, 11]. I have successfully applied RPGs in computer systems design in collaboration with Intel research on a system to guard against

Sybil attacks [26].

Randomly Perturbed Graphs grew out of the smoothed analysis paradigm [24]. Instead of choosing a specific generative model, with RPGs I have proven theorems that hold simultaneously for a wide variety of networks so long as they contain a small amount of randomness. For example, the central observation of my first RPG paper [13] is that if  $\epsilon n$  random edges are added to any  $n$ -node connected graph or digraph then the resulting graph has diameter  $\mathcal{O}(\log n)$  with high probability. This provides a theoretical explanation of the “six degrees of separation” phenomenon observed experimentally by Stanley Milgram. It shows that any large-enough network that contains a small amount of uniform randomness will have low diameter. There is great potential for this approach to aid in the solution of problems that are considered intractable on the grounds of worst-case analysis. For example, the Sybil attack is a major security issue in online systems, where an attacker controls an unlimited number of agents. In [26] we designed a system that is proven to work when the underlying network is an RPG. I plan to continue to demonstrate the potential of this approach in the near future, for example by designing next-generation Internet protocols which take advantage of the RPG nature of real-world computer networks.

My time at Microsoft Research has made it abundantly clear to me that harnessing the power of massive networks will require new developments in algorithms, especially in the areas of **online, streaming, and message-passing algorithms**. (I also find input-output and cache-oblivious algorithms interesting for the same reasons.) With Adam Kalai and Brendan McMahan, I have developed an algorithm for online convex optimization [18]. This is a very general problem where we have a convex set  $S$  and an unknown sequence of cost functions  $(c_1, c_2, \dots)$ , and, in each round, we choose a feasible point  $x_t$  in  $S$ , and then learn the cost  $c_t(x_t)$ . Our approach can be viewed as a type of *simultaneous perturbation stochastic approximation*, which approximates the gradient by evaluating  $c_t$  at a single (random) point. In the future, I plan to continue this work on online algorithms and to work in the domain of streaming and message-passing algorithms as well. In particular, I am currently working to understand the power of *survey propagation*, a message-passing algorithm, which grew out of statistical physics and is currently the most successful algorithm for solving large-scale random constraint satisfaction problems.

A universal aspect of large-scale complex networks is that they are the results of decentralized agents. Studying the strategic interaction of decentralized agents has long been the domain of **economics**. Its relevance to computer science has been recognized recently. I have been working in game theory and mechanism design for the last few years to understand how economic reasoning applies to the problems arising in complex networks. This has led me to consider several formulations of the Vickery-Clarke-Groves auction (a generalized second-price auction) [20, 4, 17] and has also motivated my recent work on trust-based recommendation systems, which applies the axiomatic approach of the theory of social choice to the problem of making recommendations in a network [2]. I will continue applying ideas from economics to the problems specifically emerging from complex networks, and perhaps this will eventually lead to ideas that economists find useful as well.

## Older work: Random structures and algorithms

My earlier work on the *average-case analysis of algorithms* has taken a new direction during my post-doctoral fellowship at Microsoft Research. In the past, I have worked extensively on proving rigorous bounds on the possibility of solving combinatorial search, optimization, and counting problems on particular random distributions of instances. During my time at Microsoft I have studied many large real-world datasets, including the social networks from Xbox Live and Halo, the web graph and query-click graphs from Live Search, and a collection of 25,000 pictures of cats and dogs (originally used in reverse Turing tests). The nature of these datasets has convinced me that practical results must not rely on overly restrictive assumptions about the random distribution that the problem instances come from, and should instead use an approach like semi-random instances or my RPG framework, as described above. The tools that I have developed for analyzing classical models of random graphs are useful in the rigorous analysis of any network with some randomness present. Some of my previous work includes average-case analysis of algorithms for graph coloring (the first result to work with many fewer colors than the maximum degree), facility location, subset sum, two-stage stochastic minimum spanning tree, and random 3-SAT (my work here resolved a 10-year-old conjecture of Koutsoupias and Papadimitriou) [5, 16, 19, 12, 10]. I have also written two surveys on these techniques [9, 21].

## References

- [1] ACHLIOPTAS, D., CLAUSET, A., KEMPE, D., AND MOORE, C. On the bias of traceroute sampling: or, power-law degree distributions in regular graphs. In *STOC '05: Proceedings of the thirty-seventh annual ACM symposium on Theory of computing* (New York, NY, USA, 2005), ACM Press, pp. 694–703.
- [2] ANDERSEN, R., BORGS, C., CHAYES, J., FEIGE, U., FLAXMAN, A. D., KALAI, A. T., MIRROKNI, V., AND TENNENHOLTZ, M. Trust-based recommendation systems: an axiomatic approach. Manuscript submitted for publication.
- [3] BARABÁSI, A.-L., AND ALBERT, R. Emergence of scaling in random networks. *Science* 286, 5439 (1999), 509–512.
- [4] CAREY, M., FLAXMAN, A. D., HARTLINE, J., AND KARLIN, A. Auctions for structured procurement. To appear in SODA 2008.
- [5] DYER, M., FLAXMAN, A. D., FRIEZE, A. M., AND VIGODA, E. Randomly coloring sparse random graphs with fewer colors than the maximum degree. *Random Struct. Algorithms* 29, 4 (December 2006), 450–465.
- [6] FALOUTSOS, M., FALOUTSOS, P., AND FALOUTSOS, C. On power-law relationships of the internet topology. In *SIGCOMM '99: Proceedings of the conference on Applications, technologies, architectures, and protocols for computer communication* (New York, NY, USA, 1999), ACM Press, pp. 251–262.

- [7] FLAXMAN, A., FRIEZE, A., AND FENNER, T. High degree vertices and eigenvalues in the preferential attachment graph. *Internet Math.* 2, 1 (2005), 1–19.
- [8] FLAXMAN, A., FRIEZE, A. M., AND VERA, J. A geometric preferential attachment model of networks. In *Algorithms and Models for the Web-Graph: Third International Workshop, WAW 2004, Rome, Italy, October 16, 2004, Proceedings* (2004), vol. 3243 of *Lecture Notes in Computer Science*, Springer, pp. 44–55.
- [9] FLAXMAN, A. D. Algorithms for random 3-SAT.
- [10] FLAXMAN, A. D. A spectral technique for random satisfiable 3CNF formulas. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms (Baltimore, MD, 2003)* (New York, 2003), ACM, pp. 357–363.
- [11] FLAXMAN, A. D. Expansion and lack thereof in randomly perturbed graphs. In *Web Algorithms Workshop 2007* (2007). To appear in *Internet Mathematics*.
- [12] FLAXMAN, A. D., FRIEZE, A., AND KRIVELEVICH, M. On the random 2-stage minimum spanning tree. *Random Structures Algorithms* 28, 1 (2006), 24–36.
- [13] FLAXMAN, A. D., AND FRIEZE, A. M. The diameter of randomly perturbed digraphs and some applications. In *Proc. of the 8th Int. Workshop on Randomization and Computation (RANDOM)* (2004), K. Jansen, S. Khanna, J. D. P. Rolim, and D. Ron, Eds., vol. 3122 of *Lecture Notes in Computer Science*, Springer, pp. 345–356.
- [14] FLAXMAN, A. D., FRIEZE, A. M., AND VERA, J. Adversarial deletions in a scale-free random graph process. *Combinatorics, Probability, and Computing* 16 (2007), 261–270.
- [15] FLAXMAN, A. D., FRIEZE, A. M., AND VERA, J. A geometric preferential attachment model of networks II. To appear in *Web Algorithms Workshop 2007*.
- [16] FLAXMAN, A. D., FRIEZE, A. M., AND VERA, J. C. On the average case performance of some greedy approximation algorithms for the uncapacitated facility location problem. In *STOC'05: Proceedings of the 37th Annual ACM Symposium on Theory of Computing* (New York, 2005), ACM, pp. 441–449.
- [17] FLAXMAN, A. D., GAMARNIK, D., AND SORKIN, G. B. First-passage percolation on a width-2 strip and the path cost in a vcg auction. In *Internet and Network Economics, Second International Workshop, WINE 2006, Patras, Greece, December 15-17, 2006, Proceedings* (2006), vol. 4286 of *Lecture Notes in Computer Science*, Springer, pp. 99–111.
- [18] FLAXMAN, A. D., KALAI, A. T., AND MCMAHAN, B. H. Online convex optimization in the bandit setting: gradient descent without a gradient. In *SODA '05: Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms* (Philadelphia, PA, USA, 2005), Society for Industrial and Applied Mathematics, pp. 385–394.

- [19] FLAXMAN, A. D., AND PRZYDATEK, B. Solving medium-density subset sum problems in expected polynomial time. In *STACS 2005*, vol. 3404 of *Lecture Notes in Comput. Sci.* Springer, Berlin, 2005, pp. 305–314.
- [20] FLAXMAN, A. D., AND REED, B. Fluctuating frugality: Vcg path auction in the giant component. Manuscript in preparation.
- [21] FLAXMAN, A. D., AND SORKIN, G. B. Tools for average-case analysis of optimization heuristics. In *Handbooks of Operations Research and Management Science: Approximation and Heuristics*. In preparation.
- [22] FLAXMAN, A. D., AND VERA, J. Bias reduction in traceroute sampling: towards a more accurate map of the Internet. To appear in Web Algorithms Workshop 2007.
- [23] LAKHINA, A., BYERS, J., CROVELLA, M., AND XIE, P. Sampling biases in ip topology measurements. In *22nd Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2003)* (April 2003), vol. 1, IEEE, pp. 332–341.
- [24] SPIELMAN, D., AND TENG, S.-H. Smoothed analysis of algorithms: why the simplex algorithm usually takes polynomial time. In *Proceedings of the Thirty-Third Annual ACM Symposium on Theory of Computing* (New York, 2001), ACM, pp. 296–305 (electronic).
- [25] WATTS, D. J., AND STROGATZ, S. H. Collective dynamics of “small-world” networks. *Nature* 292 (1998), 440–442.
- [26] YU, H., KAMINSKY, M., GIBBONS, P. B., AND FLAXMAN, A. D. Sybilguard: defending against sybil attacks via social networks. In *SIGCOMM '06: Proceedings of the 2006 conference on Applications, technologies, architectures, and protocols for computer communications* (New York, NY, USA, 2006), ACM Press, pp. 267–278.