ASYMPTOTICS OF LARGE-SCALE INTERACTING NETWORKS

Organizers: B. Hajek (University of Illinois at Urbana-Champaign)
Peter Marbach (University of Toronto)
Sujay Sanghavi (University of Texas at Austin)

Workshop dates: February 25 - March 1, 2013

1 Overview of the Field

The focus of the workshop was interacting networks where agents infer and act on local viewpoints, with global consequences. Of particular interest are scenarios where either the number of agents, or the size of the inference problem, is large and the system behavior can be characterized by an *asymptotic analysis.* Interacting networks with these properties arise in several contexts such as biological networks, financial and economic networks, social networks, and energy and communication networks. While significant progress has been made in recent years on characterizing the asymptotic topological properties of large-scale networks, there is still a lack of understanding of large-scale *interactive* networks.

Asymptotic analysis of such networks draws from, and combines, techniques from probability theory, graph theory, statistical physics, game theory, and control and information theory. The researchers attending the workshop spanned the disciplines of mathematics, engineering, applied probability, computer science, economics, and control and information theory.

The workshop covered the following *foundational aspects:*

- local algorithms – strengths and limitations aren't totally understood; progress was made via combining algorithmic complexity theory with random graphs, statistical physics and high-dimensional statistics.
- universality – robustness of graphical methods to underlying statistical assumptions.
- spectral methods - there are gaps remaining between information theoretic bounds and performance.

- game theory and markets – especially its interactions with graph theory, including both random graphs and mean-field.
- understanding the joint dynamics of processes on graphs, and the graph structure itself.
- tools for asymptotics: large deviations, fluid limits, concentration, replica economy, mean field limits.

and the following *new applications:*

- kidney exchange (via cycle-detecting algorithms facilitating multi-party exchanges)
- identifying interactions of drugs and genes (via detecting bi-cliques in bipartite graphs)
- mechanism design (both for markets of exchanges, and those with money)
- peer to peer systems (video streaming and electronic currency)
- cloud services and providers
- contagions (spread of things across the network)
- community detection and clustering in large-scale sparse graphs.
- geopolitics

We observed in several instances a persistent gap between theory and practice. Often iterative (expectation-maximization and belief propagation algorithms) or tuned algorithms work best but performance guarantees are nonexistent. Indeed a theme of several talks and open problems has been the understanding of performance of such algorithms.

2 Presentation Highlights

The presentations fell into a small number of categories, which are reviewed below.

2.1 ALLOCATION

Two of the presentations dealt with the problem of allocating kidney donors to recipients in a kidney exchange. This is an important application with life and death consequences for many people. A complicating aspect is that multiple kidney operations need to be conducted simultaneously in order to lock in the commitments of multiple kidney donors.

Speaker: **Yashodhan Kanoria** (Microsoft Research New England)
Title: *A Dynamic Graph Model of Kidney Exchange*
Abstract: We study average waiting time in a dynamic random graph model. A node arrives at each time step. A directed edge is formed with probability p with each node currently in the system. If a cycle is formed, of length no more than 3, then that cycle of nodes is

removed immediately. We show that the average number of nodes in the system, which is the same as the average waiting time for a node, scales as $\Theta(1/p^{3/2})$ for small p, for this policy. Moreover, we prove that no batching policy can achieve better delay scaling than this policy. This question is motivated by kidney exchange programs. The insight offered by our analysis is that the benefit of waiting for additional incompatible patient-donor pairs to arrive (batching) into kidney exchange clearinghouses is not substantial and is outweighed by the cost of waiting.

Speaker: **David Gamarnik** (Massachusetts Institute of Technology)
Title: *Combinatorics of Kidney Exchanges Problems*
Abstract: Kidney exchanges between incompatible pairs is one of the most challenging medical problems. It concerns donor-recipient pairs (say brother and sister) where the donor is willing to give a kidney, but the recipient cannot accept it due to blood or tissue incompatibility. In this case medical practitioners seek one or several incompatible pairs in similar situation where compatibility between the pairs can be achieved and conduct a simultaneous transplant within a cycle of pairs. Alternatively a chain of exchanges can be generated starting from a so-called altruistic donor - a healthy person willing to donate a kidney.

It has been argued in the past that there is no need to consider exchanges (cycles) longer than three and, moreover, the incremental benefit from altruistic donors is marginal. We built an algorithm for finding maximum cycle cover subject to cycle lengths constraints running on the actual pool of incompatible pairs. The results show that, despite earlier claims, there is a substantial benefit from considering longer cycles and from chains generated from altruistic donors. We substantiate these empirical findings by constructing a random graph model which a is mixture of dense and sparse random graphs. In such a graph we prove that increasing the upper bound on cycle lengths increases the number of covered patients by a linear (in the size of the graph) additive factor. Similarly, we prove that increasing the upper bound on chain lengths again increases the number of covered patients by a linear factor.

Joint work with Itai Ashlagi (MIT) and Alvin Roth (Harvard).

2.2 AVERAGING, DIFFUSION, AND VOTING

The workshop highlighted the commonalities and differences among various problems having to do with the spread of information in a network through averaging and diffusion of information.

Speaker: **Angelia Nedich** (University of Illinois)
Title: *On Hegselmann-Krause Dynamics and its Extensions*
Abstract: We consider multi-dimensional Hegselmann-Krause model for opinion dynamics in discrete-time for a set of homogeneous agents. Using dynamic system point of view, we investigate stability properties of the dynamics and show its finite time convergence. The novelty of this work lies in the use of dynamic system approach and the development of Lyapunov-type tools for the analysis of the Hegselmann-Krause model. Furthermore, we discuss some possible extensions of this model.

Speaker: **Giacomo Como** (Lund University)
Title: *Resilience of Distributed Averaging over Large-Scale Networks*
Abstract: Distributed averaging has become ubiquitous both as algorithms for distributed optimization, estimation, and control, as well as a behavioral model, e.g., for social influence and opinion dynamics in social networks. This dynamics can be related by duality to the evolution of a Markov chain with state space coinciding with the node set of the network. When the transition probability matrix P associated to the Markov chain is irreducible, a key quantity is its invariant probability distribution $\pi = P'\pi$. In this work, we study how π is affected by, possibly non-reversible, perturbations of P. In particular, we are interested in perturbations which are localized on a small fraction of nodes but are not necessarily small in any induced norm. While classical perturbation results based on matrix analysis cannot be applied in this context, we present various bounds on the effect on π of changes of P obtained using coupling and other probabilistic techniques. Such results allow one to find sufficient conditions for the l_1-distance between π and its perturbed version to vanish in the large-scale limit, depending on the mixing time and one additional local property of the original chain P.

Speaker: **R. Srikant** (University of Illinois at Urbana-Champaign)
Title: *Random Hamiltonian Cycles and Peer-to-Peer Networks*
Abstract: We consider graphs formed by unions of directed random Hamiltonian cycles. We exploit the fact that such graphs have small diameters to design peer-to-peer networks with high throughput and small delays. In particular, we show the following result: consider the union of two directed random Hamiltonian cycles over N nodes, where each edge in one of the cycles is possibly removed with some fixed probability. We show that the graph formed by the union of these two cycles has a diameter of O(log N) hops, with high probability. Joint work with Joohwan Kim.

Speaker: **Francois Baccelli** (École Normale Supérieure and University of Texas at Austin)
Title: *Infinite Peer to Peer Networks and the Laws of Super-Scalability*
Abstract: Joint work with F. Mathieu and I. Norros.

We analyze a spatial birth and death process in the infinite Euclidean space where the birth rate is constant and the death rate of a given point is the shot noise created at his location by the other points of the current configuration.

This model is motivated by the analysis of distributed content delivery networks and more precisely of network-limited peer to peer systems.

We construct the stationary regimes of this process and give a sufficient condition for the existence and uniqueness of the stationary regime. We then prove that the stationary point process exhibits some form of repulsion between its points.

We establish a hierarchy of balance equations between the stationary moment measures.

We then analyze two asymptotic regimes of this class of spatial birth and death processes: the fluid regime and the hard–core regime. We get closed form expressions for the mean (and in some cases the law) of the lifetime of a typical point as well as for the spatial density of points in the steady state of each asymptotic regime. We also propose an accurate approximation that holds for all regimes.

We conclude by showing that this birth and death process exhibits a property that we call super-scalability: the lifetime of a typical point is a decreasing function of the birth rate.

Speaker: **Zhenming Liu** (Princeton University)
Title: *The Diffusion of Networking Technologies*
Abstract: There has been significant interest in the networking community on the impact of cascade effects on the diffusion of networking technology upgrades in the Internet. Thinking of the global Internet as a graph, where each node represents an economically-motivated Internet Service Provider (ISP), a key problem is to determine the smallest set of nodes that can trigger a cascade that causes every other node in the graph to adopt the protocol. We design the irst approximation algorithm with a provable performance guarantee for this problem, in a model that captures the following key issue: a node's decision to upgrade should be influenced by the decisions of the remote nodes it wishes to communicate with.

Given an internetwork G(V,E) and threshold function θ, we assume that node u activates (upgrades to the new technology) when it is adjacent to a connected component of active nodes in G of size exceeding node u's threshold $\theta(u)$. Our objective is to choose the smallest set of nodes that can cause the rest of the graph to activate. Our main contribution is an approximation algorithm based on linear programming, which we complement with computational hardness results and a near-optimum integrality gap. Our algorithm, which does not rely on submodular optimization techniques, also highlights the substantial algorithmic difference between our problem and similar questions studied in the context of social networks.

The talk will be based on joint work with Sharon Goldberg.

Speaker: **Shreyas Sundaram** (University of Waterloo)
Title: *Robustness of Complex Networks with Implications for Consensus and Contagion*
Abstract: We consider averaging dynamics on networks that contain stubborn or malicious nodes. In particular, we study a class of dynamics where each normal node removes extreme values from its neighborhood before averaging the remaining values. We show that traditional graph metrics such as connectivity are no longer sufficient to characterize the behavior of such dynamics; instead, we show that a property termed r-robustness is required. This property is much stronger than other graph properties such as connectivity, in that one can construct graphs with high connectivity but low robustness. We investigate the robustness of common random graph models for complex networks (Erdos-Renyi, geometric random, and preferential attachment graphs), and show that the notions of connectivity and robustness coincide on these graphs: the properties share the same threshold function in Erdos-Renyi graphs, cannot be very different in 1-d geometric random graphs, and are equivalent in preferential attachment graphs. This indicates that a variety of purely local diffusion dynamics (such as resilient consensus and contagion) will be effective at spreading information in such networks. Joint work with Haotian Zhang (U. Waterloo).

Speaker: **Mohammed Abdullah** (Imperial College, London)
Title: *Majority Consensus on Sparse Graphs by Local Majority Polling*
Abstract: We study the local majority protocol on simple graphs of a given degree sequence, for a certain class of degree sequences. We show that for almost all such graphs,

subject to a sufficiently large bias, within time $A \log_d \log_d n$ the local majority protocol achieves consensus on the initial global majority with probability $1 - n^{-\Omega((\log n)^{\epsilon})}$, where $\epsilon > 0$ is a constant. A is bounded by a universal constant and d is a parameter of the graph; the smallest integer which is the degree of $\Theta(n)$ vertices in the graph. We further show that under the assumption that a vertex v does not change its colour if it and all of its neighbours are the same colour, *any* local protocol $\mathcal{P}$ takes time at least $(1 - o(1)) \log_d \log_d n$, with probability $1 - e^{-\Omega(n^{1-o(1)})}$ on such graphs. We further show that for almost all d-regular simple graphs with d constant, we can get a stronger probability to convergence of initial majority at the expense of time. Specifically, with probability $1 - O(c^{-n^{\epsilon}})$, the local majority protocol achieves consensus on the initial majority by time $O(\log n)$. Finally, we show how the technique for the above sparse graphs can be applied in a straightforward manner to get bounds for the Erdős–Renyi random graphs in the connectected regime.

Speaker: **Ayalvadi Ganesh** (University of Bristol)
Title: *Tutorial: Rumors and Consensus on Networks*
Abstract: We provide a brief overview of mathematical models of the spread of information over a network, as well as models of competing information or opinions. We review techniques for studying the time required for information to spread or consensus to be reached, and for obtaining bounds on this quantity in arbitrary networks. These bounds relate to topological properties of the network. The talk will conclude with a discussion of some open problems.

2.3 COMPLEXITY AND LOCAL ALGORITHMS

A theme that recurred several times is the apparent inability of local algorithms, such as belief propagation, to solve combinatorially difficult problems in random graphs in dense enough regimes.

Speaker: **David Gamarnik** (Massachusetts Institute of Technology)
Title: *Tutorial: Power and Limitations of Local Algorithms for Network Optimization Problems*
Abstract: Algorithms which can be run in parallel on large networks based on local information (local algorithms) gained a lot of prominence recently in a variety of applications, primarily as a way of addressing the scaling issues of computations on large instances. Some local algorithms such as, for example, the Belief Propagation algorithm emerged as strong contenders for solving a variety of optimization and inference problems on large scale network models, including random instances of hard constraint satisfaction problems. In this talk we will survey the known positive results about the power of local algorithms, but we will also discuss limitations of local algorithms for solving constraint satisfaction problems. In particular, we will discuss a (negative) resolution of a recent conjecture by Hatami, Lovasz and Szegedy regarding the power of local algorithms for solving random instances of constraint satisfaction problems.

Speaker: **Jinwoo Shin** (IBM Research and KAIST)
Title: *Belief Propagation for Maximum Weighted Matching using Odd Cycles*

Abstract: The maximum weighted matching (MWM) is one of classical combinatorial optimization problems which have been studied since the seminar work by Edmonds in 1961. In this work, we study the MWM problem via the max-product Belief Propagation (BP) algorithm, which is a popular heuristic solving inference problems arising in graphical models. We show that the BP algorithm converges to a correct answer if the MWM linear programming relaxation using non-intersecting odd cycle inequalities is tight. Our main idea is constructing a novel graphical model using non-intersecting odd cycles so that the resulting BP algorithm works. Furthermore, our result naturally guides a cutting-plane-like method using the BP algorithm for the MWM problem, where our experimental results show that it solves 95% of random MWM instances.

This is a joint work with Misha Chertkov (LANL) and Andrew Gelfand (UCI)

2.4 GAMES

Games in a network setting offer a rich variety of behaviors and applications. One application covered by several of the talks is to determine the level of investments interconnected firms make for their security

Speaker: **Venkat Anantharam** (University of California, Berkeley)
Title: *Nash Equilibrium Structure of a Class of Blocking Games with Applications to Network Security*
Abstract: We study a game-theoretic model for security/availability in a networking context. To perform some desired task, a defender needs to choose a subset from a set of resources. To perturb the task, an attacker picks a resource to attack. We model this scenario as a 2-player game and are interested in describing its set of Nash equilibria. The games we study have a particular structure, for which we can use the theory of blocking pairs of polyhedra, pioneered by Fulkerson, to arrive a reasonably satisfactory understanding of the Nash equilibria. The subsets of resources that support Nash equilibrium strategies of the attacker, called ”vulnerability sets”, are of particular interest, and we identify them in several specific games of this type. An example of a game of this sort is when the set of resources is the set of edges of a connected graph, the defender chooses as its subset the edges of a spanning tree, and the attacker chooses an edge to attack with the aim of breaking the spanning tree.

(joint work with Assane Gueye, Aron Laszka, and Jean Walrand)

Speaker: **Augustin Chaintreau** (Columbia University)
Title: *Convergence of Coloring Games with Collusions and Why It's Difficult to Keep Your Friends When You Have Enemies*
Abstract: Various social networks, and some computational problems, require forming groups in the presence of de facto antagonistic relationships denoting local incompatibility. This can be posed as a vertex graph coloring problem, and the configurations attainable under these constraints through coordinated algorithms (e.g., aiming at using the least amount of colors) are well characterized. In contrast, when groups are formed in a distributed manner by self-interested nodes, we know surprisingly little of the configurations that emerge from the dynamics of such a coloring game. Collusions multiple players joining a new

group together so that they all improve their scores are likely to occur in such group formation settings; they greatly affect the stability and efficiency of the configurations.

In this presentation, we characterize the convergence of coloring games, revealing their intricate complexity properties and proving that previous bounds are arbitrarily loose. First, in a uniform game, we show that sequences of moves relate to physics models of condensed matter and integer partition. More- over, we prove that combinatorial complexity of coloring games suddenly increases: no polynomial bound exists when- ever this analysis fails. This solves an open problem and explains the failure of previous proposed methods using potential functions. In non-uniform games where nodes preferences are given in a set of weights, we prove generally the maximum collusion size that guarantees convergence. As we show, deciding the convergence beyond this point is never computationally feasible. Our results extend to variants where nodes join multiple groups, have asymmetrical weights, and utility goes beyond pairwise relationship. Finally, we show these games create a tension between stability and efficiency, as collusions and using multiple colors both have a beneficial effect on the price of anarchy.

(based on joint work with Guillaume Ducoffe and Dorian Mazauric).

Speaker: **Mallik Rao** (Indian Institute of Technology, Bombay)
Title: *Topic: Multiple Mutations in Evolutionary Games*
Abstract: In this talk, we re-look at evolutionary stability in bimatrix games and introduce evolutionary stability against multiple mutations. We provide some characterizations and properties. (Joint work with Ghatak and Shaiju.)

Speaker: **Vijay Subramanian** (Northwestern University)
Title: *On the Stability and Efficiency of Network Bargaining with Search Friction*
Abstract: In this paper, we consider the dynamics of a market trading process modeled as a non-cooperative networked bargaining game. Our goal is to study how search friction inuences the markets stability and efficiency. We show that with search friction, stationary equilibrium can be sustained, but the equilibrium can be inefficient. In particular, the strategy in which agents trade immediately whenever they meet a potential trading partner cannot be sustained at equilibrium; trade sometimes is delayed. This is joint work with Thanh Nguyen and Randy Berry at Northwestern.

Speaker: **Adam Wierman** (California Institute of Technology)
Title: *Congestion and Competition in the Cloud*
Abstract: The cloud marketplace has evolved into a large, highly complex economic system made up of a variety of services. As a result of this complicated marketplace, the performance delivered to users by cloud services depends on the the resource allocation design of the service itself and the strategic incentives resulting from the large-scale multi-tiered economic interactions among cloud providers. In this talk I will present work developing and analyzing some new models capturing the interaction this multi-tiered interaction in a manner that exposes the interplay of congestion, pricing, capacity provisioning, and performance. This talk will present joint work with Jonatha Anselmi, Danilo Ardagna, Urtzi Ayesta, Yunjian Xu, and Zichao Yang.

Speaker: **Ayalvadi Ganesh** (University of Bristol)

Title: *On the secure connectivity of wireless sensor networks*

Abstract: We consider a model of a wireless sensor network in which nodes are distributed uniformly at random over a square, and can communicate other nodes within a fixed radius. This is the well-known random geometric graph model. In addition, each node is assigned a subset of cryptographic keys chosen uniformly at random from a key pool. Two nodes can communicate only if they are within range of each other, and possess at least one common key. We establish a connectivity threshold for these random graph models. This resolves a conjecture of Osman Yagan.

Joint work with Bikash Dey, Santhana Krishnan and D. Manjunath.

Speaker: **Ashish Goel** (Stanford University)

Title: *Trust and Mistrust in Social Networks*

Abstract: Automated reputation and trust systems play an ever increasing role in the emerging networked society. We will first describe a model of networked trust that functions by exchange of IOUs among participants. Informally, every node acts as a bank and prints its own currency, which is then used to purchase services within the network. Such ”trust networks” are robust to infiltration, since any node only accepts currency printed by other nodes that it directly trusts. We will analyze the liquidity of this model, i.e., the number of transactions that such a network can support. We will show that in many interesting cases, the liquidity of these trust networks is comparable to a system where currency is issued by a single centralized bank. We will also present results on the formation of credit networks given self-interested agents.

While social networks and the Internet have made it much easier to communicate, there is considerable anecdotal and research evidence that we are getting more polarized as a society. We will present an analysis of the dynamics of opinion polarization in a social network, and discuss its implications for the design of collaborative systems as well as recommendation algorithms.

Speaker: **Asu Ozdaglar** (Massachusetts Institute of Technology)

Title: *Network Security and Contagion*

Abstract: This paper develops a theoretical model of investments in security in a network of interconnected agents. The network connections introduce the possibility of cascading failures depending on exogenous or endogenous attacks and the profile of security investments by the agents. The general presumption in the literature, based on intuitive arguments or analysis of symmetric networks, is that because security investments create positive externalities on other agents, there will be underinvestment in security. We show that this reasoning is incomplete because of a first-order economic force: security investments are also strategic substitutes. In a general (non-symmetric) network, this implies that underinvestment by some agents will encourage overinvestment by others. We demonstrate by means of examples that not only there will be overinvestment by some agents but also aggregate probabilities of infection can be lower in equilibrium than in the social optimum. We then provide sufficient conditions for underinvestment. This requires both sufficiently convex cost functions (just convexity is not enough) and networks that are either symmetric or locally tree-like (i.e., either trees or in the case of stochastic networks, without local cycles with high probability). We also characterize the impact of network structure

on equilibrium and optimal investments. Finally, we show that when the attack location is endogenized (by assuming that the attacker chooses a probability distribution over the location of the attack in order to maximize damage), there is another reason for overinvestment: greater investment by an agent shifts the attack to other parts of the network.

This is joint work with Daron Acemoglu and Azarakhsh Malekian.

Speaker: **Bert Zwaart** (VU University Amsterdam and Centrum Wiskunde & Informatica (CWI))
Title: *Dimensioning Reliable Power Grids*
Abstract: Reliability and robustness are essential features in the design, control and maintenance of power grids with wind and solar energy production. To this end, stringent demands are imposed on system reliability (e.g. max 2.25 minutes of power interruption per year). However, wind and solar generation are dependent on meteorological conditions and therefore inherently uncertain, leaving the (small) possibility of large fluctuations that lead to failure of power grids. This leads to an exciting class of stochastic network problems.

2.5 INFERENCE

A major topic of the workshop was inference in graphs. Problems included the detection of communities of nodes based on the links connecting individuals. An interesting variety of new methods have been applied recently, from spectral graph methods to low rank and sparse recovery methods.

Speaker: **Andrea Montanari** (Stanford University)
Title: *Hypothesis Testing in High-Dimension*
Abstract: A large amount of work has been devoted, over the last ten years, to high-dimensional statistical estimation problems. Examples range from signal processing to collaborative filtering. The methods developed are mostly based on convex optimization techniques, and typically do not provide tools to reason about uncertainty or confidence. Hypothesis testing provides a classical framework for addressing this problem. I will discuss the challenges related to hypothesis testing in this context. [Joint work with Adel Javanmard]

Speaker: **Aarti Singh** (Carnegie Mellon University)
Title: *Computationally Efficient Detectors for Weak Graph-Structured Activations*
Abstract: The ability to detect weak patterns of anomalous activity is critical to network operations. However, the ever-increasing size of modern networks renders traditional detectors computationally intractable. I will describe two computationally efficient detectors that we have developed by leveraging tools from graph theory, machine learning and signal processing - one is a spectral relaxation of the combinatorial graph scan statistic and the other relies on a graph wavelet construction using the uniform spanning tree. We demonstrate the precise performance gains in terms of signal-to-noise ratio required by these detectors for different graph structures as well as their computational complexity. In particular, we show that the uniform spanning tree wavelet detector achieves information-theoretic limits for a broad class of graphs.

Speaker: **Sujay Sanghavi** (University of Texas at Austin)
Title: *Network Inverse Problems via Convex Optimization*
Abstract: Recent years have seen a revolution in high-dimensional statistics, and in particular the statistical and computational efficiency of algorithms therein that are based on convex optimization. In this talk, we present two results that illustrate the power of this line of work in solving inverse problems in networks: (a) Clustering sparse graphs: for the classic and popular planted partition/stochastic block model, we propose a simple convex program, and show that it outperforms a long line of work in terms of the level of sparsity and size of clusters it can resolve. (b) Finding the graph of cascades: given only node states in an epidemic, we consider the problem of finding the graph on which the epidemic spread. We show, for the classic independent cascade model, this can be posed as a convex program; this yields both an efficient parallel algorithm, and a way to determine statistical guarantees on when the graph is recoverable. We emphasize the common analytical threads in both problems: how familiar concentration results on graphs can be used to establish the success of convex methods for these (otherwise combinatorial) problems.

Speaker: **Siddhartha Banerjee** (The University of Texas at Austin)
Title: *Large-Scale Online Recommendation on Graphs*
Abstract: We consider the problem of large-scale online recommendation, i.e., settings wherein an algorithm has to display relevant items from a large available content set to dynamically arriving users. Such settings are characterized by four critical features: (i) the use of user-feedback for inferring unknown content value (resulting in an exploration vs. exploitation tradeoff), (ii) the presence of an access graph, which constrains which items can be shown to which users, (iii) the inherent non-stochastic nature of content value and (iv) the surfeit of content, which often far exceeds the number of users. We consider a model that captures these features, and study how they affect the design of recommendation algorithms. More precisely, we consider a setting where a graph constrains what items each user can be shown; users earn rewards from these displayed items according to an underlying, unknown reward-function, which can be learnt from user-feedback. Our task is to design an algorithm to select a set of relevant items to display to each user upon arrival. We develop algorithms for both static and dynamic settings, and provide guarantees on their competitive ratio, under very general conditions on the reward-functions. Conversely, we also present upper bounds on the competitive ratio which show that our algorithms are within a constant factor of the optimal.

Speaker: **Alex Dimakis** (University of Texas at Austin)
Title: *Sparse Principal Component Analysis*
Abstract: We present a new algorithm for sparse PCA for low-rank or near low-rank matrices. Under spectral assumptions, we obtain provable performance guarantees.
(Joint work with Dimitris Papailiopoulos)

Speaker: **Anima Anandkumar** (University of California, Irvine)
Title: *A Tensor Spectral Approach to Learning Mixed Membership Community Models*
Modeling community formation and detecting hidden communities in networks is a well studied problem. However, theoretical analysis of community detection has been mostly

limited to models with non-overlapping communities such as the stochastic block model. In this paper, we remove this restriction, and consider a family of probabilistic network models with overlapping communities, termed as the mixed membership Dirichlet model, first introduced in Aioroldi et. al. 2008. This model allows for nodes to have fractional memberships in multiple communities and assumes that the community memberships are drawn from a Dirichlet distribution. We propose a unified approach to learning these models via a tensor spectral decomposition method. Our estimator is based on low-order moment tensor of the observed network, consisting of 3-star counts. Our learning method is fast and is based on simple linear algebra operations, e.g. singular value decomposition and tensor power iterations. We provide guaranteed recovery of community memberships and model parameters and present a careful finite sample analysis of our learning method. Additionally, our results match the best known scaling requirements in the special case of the stochastic block model. This is joint work with Rong Ge, Daniel Hsu and Sham Kakade.

Speaker: **Jiaming Xu** (University of Illinois at Urbana-Champaign)
Title: *Reconstruction in the Sparse Labeled Stochastic Block Model*
Abstract: The labeled stochastic block model, in its simplest form, consists of two communities of approximately equal size, and the edges are drawn and labeled with probability depending on whether their two endpoints belong to same community. In the sparse graph case with bounded degrees, a reconstruction threshold has been conjectured for finding a partition positive correlated to the true hidden community partition based on the observed edges and labels.

We establish the converse part of this conjecture, i.e., the reconstruction is impossible when below the threshold. For the positive part of the reconstruction, we show that: (1) the minimum bisection algorithm works on an optimally weighted graph when above the threshold by a factor of $64 \ln 2$; (2) the spectral method works when above the threshold by an unknown constant.

Furthermore, it is shown that the labeled stochastic block model is contiguous to a labeled ER random graph when below the reconstruction threshold and orthogonal to it when above the threshold, which further implies that it is impossible to consistently estimate the model parameters when below the reconstruction threshold.

Finally, we consider the general case with multiple symmetric communities. A new reconstruction threshold is conjectured and shown to coincide with a phase transition threshold in a related robust inference problem on the labeled tree.

This is joint work with Marc Lelarge and Laurent Massoulie.

2.6 WIRELESS NETWORKS

The design and performance analysis of networks formed by wireless communication among mobile nodes is an important application area for the theory of distributed algorithms for large scale networks, as illustrated by three of the presentations.

Speaker: **Florian Simatos** (Eindhoven University of Technology (TU/e))
Title: *Lingering Issues in Distributed Algorithms*
Abstract: Recent advances have resulted in queue-based algorithms for medium access

control which operate in a distributed fashion, and yet achieve the optimal throughput performance of centralized scheduling algorithms. However, fundamental performance bounds reveal that the“cautious” activation rules involved in establishing throughput optimality tend to produce extremely large delays, typically growing exponentially in 1/(1-r), with r the load of the system, in contrast to the usual linear growth.

Motivated by that issue, I will discuss in this talk the extent to which more “aggressive” schemes can improve the delay performance. In the simplest possible scenario, I will show how aggressive activation rules induce a lingering effect, where individual nodes retain possession of a shared resource for excessive lengths of time even while a majority of other nodes idle. Using central limit theorem type arguments, it can be proved that the idleness induced by the lingering effect may cause the delays to grow with 1/(1-r) at a quadratic rate.

Speaker: **Ravi Mazumdar** (University of Waterloo)
Title: *On the Typical Number of Links in a SINR Random Graph that can Support a Given Minimum Rate*
Abstract: The talk will discuss the problem of how many simultaneously transmitting links that can deliver bits at a given minimum rate can be sustained in a random wireless network with N transmitter-receiver pairs when each transmitter can interfere with a given receiver resulting in a SINR random graph.

Two scenarios will be considered. The first is when pathless issues are not very significant and all transmitter receiver pairs are within a bounded cell size while channel gains are i.i.d. Rayleigh. Here we show that the number of links (out of N) is $O(\log^2 N)$ asymptotically in probability. The second scenario is when path loss and the random placements are taken into account. Under the assumption that the distances have a light-tailed distribution and the channel gains are i.i.d. Rayleigh we show that the number of links is $O(N^{1/3})$ asymptotically.

I will provide some details of the proof technique based on order statistics in the talk.

Joint work with H. Keshavarz (USB, Iran), R. Roy (ISI, Delhi) and F. Zoghalchi (UToronto).

Speaker: **Rajesh Sundaresan** (Indian Institute of Science, Bangalore)
Title: *Asymptotics of the Invariant Distribution in a Mean Field Model with Jumps*
Abstract: I will discuss the asymptotics of the invariant measure for the process of the empirical spatial distribution of N coupled Markov chains in the limit of a large number of chains. Each chain reflects the stochastic evolution of one particle. The chains are coupled through the dependence of the transition rates on this spatial distribution of particles in the various states. The model is a caricature for medium access interactions in wireless local area networks. It is also applicable to the study of spread of epidemics in a network. The limiting process satisfies a deterministic ordinary differential equation called the McKean-Vlasov equation. When this differential equation has a unique globally asymptotically stable equilibrium, the spatial distribution asymptotically concentrates on this equilibrium. More generally, its limit points are supported on a subset of the ω-limit sets of the McKean-Vlasov equation. I will provide a very brief summary on large deviations of the invariant measure from the limit.

The talk will be based on joint work with Vivek S. Borkar (arXiv:1107.4142v2).

3 Outcome of the Meeting

A group of bright researchers from around the world interacted and will influence directions of research. Researchers from top institutions in Canada, US, France, UK, Netherlands, and India participated.

Videos of the BIRS talks will be a valuable resource for students and other researchers including tutorials that surveyed the state of the art. This is evidenced by the fact that several researchers have linked to these videos from their websites, etc.

Participants learned many problems new to them, due to the fact that the topic of asymptotics in large networks is interdisciplinary. In particular, the work presented will eventually be published in diverse venues:

- theoretical computer science
- machine learning
- applied probability
- operations research
- economics
- commerce
- communication networks
- information theory
- applied mathematics
- optimization

One realization of the workshop was the increasing opportunities presented by data that is generated by large-scale networks; a potential idea for a future workshop would be to take a more data-heavy angle to the problems investigated.

The workshop brought together leading researchers in this area to discuss recent results and open problems and to explore new mathematical techniques and models to study these problems. In addition, the workshop gave some outstanding junior researchers an opportunity to present their own research and become engaged in this field.

Besides tutorials and research presentations, the workshop provided time for participants to present open problems, and discuss problems in small groups. The workshop acted as a catalyst for new research directions and approaches in the emerging research area of asymptotic analysis of interacting networks.

4 Open Problems

Several presentations on open problems were made during the workshop. A partial list follows.

Speaker: **Sujay Sanghavi**
Title: *Open problem: Matrix Completion via Alternating Minimization*
Abstract: Matrix completion refers to the problem of inferring a low-rank matrix from only a few entries; it is central to applications in recommendation systems, manifold learning etc. Theoretically, completion has been shown to be possible under certain (now standard) assumptions. In practice, the dominant and widely used method is to represent the low-rank matrix via a bi-linear form (consisting of much smaller matrices), and doing alternating minimization. There are no theoretical guarantees for this procedure, under any assumptions. A recent result of ours (to appear in STOC 2013) proves that a modified alt-min algorithm works with the standard assumptions; the modification requires fresh resampling in every iteration. Empirically, this is not seen to be required. A proof of the success of the original (and implemented) algorithm would, I believe, be of some interest. However, this involves resolving a statistical dependency issue I will outline.

Speaker: **Aarti Singh** (Carnegie Mellon University)
Title: *Open problem: On the statistical and computational gap in discovering weak and sparse bi-cliques*
Abstract: I will describe the problem of identifying a small bi-clique of interacting nodes, defined by edges with elevated mean weights, based on noisy edge weight measurements. This problem comes up in many application scenarios, for example, discovering groups of interacting drugs and genes in a microarray experiment. If the signal is strong or the size of the bi-clique is large, then the problem is both statistically feasible and computationally tractable. However, in the weak and sparse setting there is a gap between statistical and computational efficiency. Specifically, we establish information-theoretic lower bounds on the signal strength of edges in the bi-clique as a function of the bi-clique size and graph size needed to guarantee identifiability. We demonstrate that a combinatorial procedure achieves these limits, however several natural computationally tractable procedures are unable to identify the bi-clique at the minimax signal strength. Closing this gap remains an open problem.

Speaker: **Bruce Hajek** (University of Illinois at Urbana-Champaign)
Title: *Open problem: A Conjecture on the Distribution of Balanced Load in a Large Network*
Abstract: We review a conjecture posed in B. Hajek, "Performance of global load balancing by local adjustment," *IEEE Trans. Information Theory,* Vol. 36, Nov. 1990. pp. 1398 - 1414, about the distribution of load in a large network after a load balancing algorithm is applied. (Note added May 2013: It appears Rajesh Sundaresan is coming up with a proof of this conjecture, based on the theory of unimodular distributions over the space of random graphs with a distinguished node.)

Speaker: **Vijay Subramanian**

Title: *Open problem: On the stability of a back-pressure scheduling algorithm*
Absract: I had proposed an alternate scheduling policy to back-pressure, but was only able to prove stability for some simple cases. For an even bigger class Tara Javidi and her students proved stability using piece-wise quadratic Lyapunov functions. The general case is still unproved. The policy is described in the following CDC paper
`http://users.eecs.northwestern.edu/~vjsubram/vsubramanian/cdc07_final`
article by Javidi and students is - M. Naghshvar, H. Zhuang, and T. Javidi "A General Class of Throughput Optimal Routing Policies in Multi-hop Wireless Networks," IEEE Transactions on Information Theory, Volume 58, Issue 4, pp 2175 - 2193, 2012 .

Speaker: **Rajesh Sundaresan** (Indian Institute of Science, Bangalore)
Title: *Open problem: Unbalanced fixed points for CSMA*
Abstract: Fixed point analysis has been a very useful tool in analysing CSMA systems. I will describe the notion of a balanced fixed point and an unbalanced fixed point, and will provide some observations on the effect these unbalanced fixed points seem to have on long-term and short-term system behaviour. The open question is to formulate precise statements of these effects and to establish their validity.