Final report: SAMSI Program on Complex Networks (2010-11)

Background

This year-long program focussed on the emerging area of *network science*. This highly interdisciplinary field is characterized by novel interactions in the mathematical sciences which are occurring at the interface of applied mathematics, statistics, computer science, and statistical physics, as well as those areas with network-oriented thrusts in biology, computer networks, engineering, and the social sciences.

The program considered several interconnected research foci as a mean to identify and explore the common key mathematical and statistical issues which underlie the empirical, analytical and applied approaches described above.

Overall Program Leaders: Eric Kolaczyk (Boston U.), Alessandro Vespignani (Indiana U.)

Scientific Advisory Committee: Pierre Degond (Institut de Mathématiques de Toulouse), Stephen Fienberg (Carnegie Mellon U.), Martina Morris (U. of Washington)

Local Scientific Coordinators: Alun Lloyd (NCSU), Peter Mucha (UNC)

Directorate Liaison: Rick Durrett (Cornell/Duke) and Pierre Gremaud (NCSU)

National Advisory Committee Liaison: Bin Yu (UC Berkeley)

1 Workshops

1.1 Opening workshop

1.1.1 Summary

The Opening Workshop for the SAMSI program on Complex Networks was held on Sunday-Wednesday, August 29-September 1, 2010, at the Radisson RTP in Research Triangle Park, NC. There were around 150 participants. Tutorial sessions were held on Sunday, August 29 on

- Eric Kolaczyk (Boston University), *statistical analysis of network data*,
- Alessandro Vespignani (Indiana University), *diffusion and epidemic processes in complex techno-social networks*,
- Rick Durrett (Duke University), *some features of the speed of epidemics and opinions on a random graph*,
- Michael Mahoney (Stanford University), *geometric tools for identifying structure in large social and information networks.*

Videos of these presentations are available at the SAMSI website. Invited talks were presented Monday to Wednesday. There was a poster session and reception on Monday, August 30. Immediately following the workshop, on Thursday and Friday, research working groups convened for initial meetings at SAMSI.

1.1.2 Activities

The workshop focused on five complementary themes at the forefront of current research in complex networks, incorporating perspectives ranging from theory to applications, in disciplines spanning applied mathematics, computer science, probability, statistics, and statistical mechanics. The five focus areas were

Network Sampling and Inference: Theory and methods relating to the sampling of network data and the corresponding inference of network characteristics, including applications to tasks like community detection and estimating the size of 'hard-to-count' populations. This session was led by Eric Kolaczyk and featured the following presentations

- Peter Bickel (UC-Berkeley) *statistical inference for unlabelled graphs*,
- Mark Newman (U. of Michigan) *community structure and link prediction in networks*,
- Matt Salganik (Princeton U.) *improvements to the network scale-up method for estimating the sizes of hard-to-count populations*,
- Panel discussion: Stan Wasserman (Indian University), Liza Levina (U. of Michigan) and Bruce Spencer (Northwestern U.)

Dynamic Networks: Modeling and inference of networks in the context of dynamical systems evolving in time, such as time-varying gene regulatory interactions or email social networks. Steve Fienberg (Carnegie Mellon U.) led this session. The following talks were given

- Raissa D'Souza (UC Davis) *what are dynamic networks?*
- Tom Snijders (Oxford U.) *actor-oriented models for network dynamics*,
- Sidney Redner (Boston U.) *dynamics of voter models on heterogeneous networks*,
- Panel discussion: Peter Mucha (U. of North Carolina), Hugh Chipman (Acadia U.), Josh Socolar (Duke U.).

Percolation and Diffusion on Networks: Theory regarding the behavior of phenomena like social contact processes and epidemics on networks. This session was directed by Rick Durrett; the following talks were offered

- Zoltan Toroczkai (U. of Notre Dame) *modeling functional networks for the primate cortex*,
- Erik Volz (U. of Michigan) *simple models for infection disease epidemics in complex networks*,
- Pierre Degond (CNRS) *continuum models for complex systems*,
- Panel discussion: James Moody (Duke U.), Ginestra Bianconi (Northeastern U.), Mason Porter (Oxford U.).

Spectral Analysis and Geometric Algorithms: Topics in network analysis integrating aspects of spectral graph theory and algorithms based on geometric embeddings. Michael Mahoney chaired this session which included the following presentations

- Fan Chung Graham (UC San Diego) *PageRank algorithms with applications to graph sparsification and partitioning*,
- Aaron Clauset (Sante Fe Institute) *the trouble with community detection*,
- Mauro Maggioni (Duke U.) *multi scale methods for analysis of graphs.*

Biological Applications of Networks: Various applications of network-based approaches to biological problems, such as in the context of cell biology or the epidemic spread of disease. This session was chaired by Alex Vespignani and featured the following presentations

- Eric Xing (Carnegie Mellon U.) *time varying networks; reverse engineering and analyzing rewiring social and genetic interactions*,
- Hongzhe Li (U. of Pennsylvania) *statistical methods for network-based analysis of genomic data*,
- Michelle Girvan (U. of Maryland) *effects of network topology in simple models of gene regulation*,
- Desmond Higham (U. of Strathclyde) *algorithms for evolving networks.*

In addition, several shorter talks were given by beginning researchers

- Crystal Linkletter (Brown U.) *explaining network structure: the importance of modeling pair-wise preferences*,
- Joe Bliztstein (Harvard U.) *respondent-driven sampling: degrees of uncertainty with with uncertain degrees*,
- Alexander Gutfraind (Los Alamos NL) *dark networks and vital infrastructure*,
- Natalia Katenka (Boston U.) *the impact of partial information on network inference and characterization*,
- Edoardo Airoldi (Harvard U.) *integer polytope samplers with applications to network analysis.*

1.1.3 Working groups

The opening workshop resulted in the formation of five working groups which we now briefly describe. Report from each individual group are included in Section 3.

Sampling/modeling/inference This group aimed to work towards moving the current state of knowledge on these inter-related tasks – in the specific context of networks – to rest on a more principled and integrated mathematical and statistical foundation. We are pursuing this goal by focusing on a handful of specific prototype problems in the context of certain application areas, ranging from information networks to animal communities to neuroscience.

The group leader was Eric Kolaczyk (Boston U.).

Dynamics OF networks The dynamics of networks working group explored a variety of mathematical and statistical approaches for describing and understanding the changing connection topology of networks over time, the interplay of these network dynamics with other dynamic processes on the network, and the connections between these different mathematical and statistical methodologies.

The group leaders were David Banks (Duke U.), James Moody (Duke U.) and Peter Mucha (U. of North Carolina).

Dynamics ON networks Random graphs are useful models of social and technological networks. To date most of the research in this area has concerned geometric properties of the graphs. This working group will focus on processes taking place ON the network. In particular we are interested in how their behavior on networks differs from that in homogeneously mixing populations or on regular lattices of the type commonly used in ecology and physics.

The group leaders were Rick Durrett (Duke U.) and Alun Lloyd (NCSU).

Geometrical/spectral analysis This working group was concerned with the following topics: detection of communities in networks, multiscale spectral methods for the analysis of the geometry of networks, algorithms that simplify graphs into simpler graphs in order to speed up certain optimization problems, metrics for comparing graphs, and multiscale homogenization of random walks. These topics have applications biology and to spread of "epidemics" in financial networks.

The group leaders were Mauro Maggioni (Duke U.) and Michale Mahoney (Stanford U.).

Modeling flows This working group consisted of two subgroups: (i) modeling traffic flows and (ii) modeling smart grid networks. In the traffic flows subgroup, people were interested in both stochastic and deterministic models to represent microscopic and macroscopic behavior of traffic flows. In modeling smart grid networks, the subgroup worked on mathematically formulating a graph/topology reduction problem of representing the larger network to a reduced number of local area networks, then modeling the corresponding dynamics and parameter estimation for the reduced system.

The group leader was Taufiquar Khan (Clemson U.).

1.2 Complex networks modeling workshop, October 20-22, 2010, SAMSI

1.2.1 General description

The analysis of network data has become a major endeavor across the sciences, and network modeling plays a key role. Frequently, there is an inferential component to the process of network modeling i.e., inference of network model parameters, of network summary measures, or of the network topology itself. For most standard types of data (e.g., independent and identically distributed, time series, spatial, etc.), there is a well-developed mathematical infrastructure guiding sampling, modeling and inference in practice. In the context of network data, however, such an infrastructure is only beginning to be developed.

The goal of this workshop was to bring together researchers working on the sampling, modeling, and inference of networks, for the purpose of helping move the current state of knowledge on these inter-related tasks to rest on a more principled and integrated mathematical and statistical foundation. Topics of focus include recent advances in network sampling (e.g., respondant driven sampling), inference from partially sampled (e.g., ego-centric) network data, and the confluence of traditional models (e.g., stochastic block-models, Gaussian graphical models) with modern tools for high-dimensional data analysis (e.g., l1-penalized optimization, spectral partitioning).

Organizers:

David Banks (Duke U.) and Eric Kolaczyck (Boston U.).

1.2.2 Talks

- Stephen Fienberg (Carnegie Mellon U.) *statistical challenges in network modeling*,
- Edo Airoldi (Harvard U.) *network representation*,
- Tian Zheng (Columbia U.) *statistical methods for studying social networks using aggregated relational data*,
- Purnamitra Sarkar (Carnegie Mellon U.) *theoretical justification of popular link prediction heuristics*,
- Andrew C. Thomas (Carnegi Mellon U.) *exploring the limits of conditionally independent dyadic network models*,

- Lucy Robinson (Johns Hopkins U.) *change point detection in networks*,
- Bruce Spencer (Northwestern U.) *sampling research questions*,
- Krista Gile (U. of Washington) *self-consistent network model-assisted prevalence estimation from respondent-driven sampling data*,
- Bin Yu (UC Berkeley) *spectral clustering and the high-dimensional stochastic Block model*,
- Aarti Singh (Carnegie Mellon U.) *identifying graph-structured network activitations*,
- Stephane Robin (Agro Paris Tech) *uncovering structure in interactions networks*,
- Denise Scholtens (Northwestern U. Medical School) *sequential sampling designs for estimating local connectivity in bait-prey graphs.*

In addition, a poster reception took place on Wednesday, Oct. 20., 2010.

1.3 Workshop on the dynamics of networks, January 10-12, 2011, SAMSI

1.3.1 General description

The changing structure of networks over time impact and are indeed inherent in the study of a broad array of network phenomena. The network of contacts for the spread of an infectious disease varies in time, with that variation playing a potentially important role in the course of the disease. Ad hoc communications networks between roaming elements must continously readjust and renavigate between nodes according to the changing landscape of connections. Political networks of association connections or voting similarities vary from one legislative session to the next.

The detailed local social and/or technological processes underlying each of these example applications obviously differ, but many of the basic mathematical and statistical questions regarding such networks and the generalized information they carry are similar. Though the importance of dynamics in networks has of course been long recognized, renewed interest has emerged in part due to the increasing accessibility of dynamic network data, ranging from longitudinal data waves to complete time histories of network evolution. Additionally, most of the theoretical modeling work that has been done on the dynamics of networks has been focused on the statistical equilibria of those models (e.g., growing networks by preferential attachment) or on one-time disruption events (e.g., the effect of knocking out hubs). At the same time, statistical and computational tools for analyzing time-varying networks remain relatively few in number, especially as compared to the wealth of advances in methods for modeling and analyzing static networks.

There thus remains an ongoing need and opportunity for more thorough mathematical and statistical analysis and modeling of dynamic networks. This workshop aimed to bring together researchers interested in pushing forward this extremely fertile area of research.

Organizers:

Raissa D'Souza (UC Davis), Stephen Fienberg (Carnegie Mellon U.), Eric Kolaczyk (Boston U.), Jim Moody (Duke U.), Peter Mucha (U. of North Carolina), Mason Porter (Oxford U.)

1.3.2 Talks

- Peter Mucha (U.of North Carolina) *the "OF" working group*,
- Carter Butts (UC Irvine) *modeling complex social interaction within and across settings via relational events*,

- Reka Albert (Penn State) *modeling the dynamics of biological signaling networks*,
- Carsten Wiuf (Aarhus U., Denmark) *likelihood and likelihood-free inference for certain growing network models*,
- J.P. Onnela (Harvard U.)
- Eric Kolaczyk (Boston U.) *analysis of time-indexed networks in epilepsy*,
- Skye Bender-deMoll (Self-employed) *escaping from the matrix: storing, exploring and explaining dynamic networks*,
- Peter Hoff (U. of Washington) *multiway array models for dynamic networks and relational data*,
- Pavel Krivitsky (Carnegie Melon U.) *a separable model for dynamic networks*,
- Ginestra Bianconi (Northeastern U.) *dynamics of social interactions at short timescales*,
- Dani Bassett (UC Santa Barbara) *dynamic community structure in adaptive systems*,
- Katy Borner (Indiana U.) *visualizing the structure and evolution of science and technology*.

In addition, time was reserved for the discussion of working group specific issues and for a poster reception.

1.4 Workshop on pedestrian traffic flow, February 14-16, 2011, SAMSI

1.4.1 General description

Human crowds and pedestrian groups exhibit complex and coordinated spatio-temporal patterns such as the spontaneous organization of pedestrian ows into lines and the oscillations of fluxes at gates or intersections. Understanding these phenomena requires a deeper knowledge of the laws governing the interactions of individuals with both themselves and their environment. Although a number of experiments can be found in the literature, the available data does not lend itself well to systematic study. The lack of completeness and varying quality of available experimental data make it nearly impossible to identify the fundamental principles underlying pedestrian behavior. Motion capturing techniques have become available in the recent years and permit more complex experiments, yielding comprehensive recordings of individual positions. Such experiments generate a formidable amount of data. Mathematical modeling and numerical simulation can facilitate the extraction of information from such data, helping to identify relevant observables by which a set of biological hypotheses can be validated. Such a dialectic use of experiment and simulation results in the ability to extract from traffic data an ensemble of behavioral rules, which can then be inserted into individual (agent)-based simulation models. For crowds involving several thousand (or more) individuals, continuum fluid-mechanical-like models are attractive, from both the computational and theoretical viewpoints. Deriving these models from the individual behavior by means of statistical physics guarantees their biological relevance and opens the way to qualitative analysis and understanding of many observed emerging phenomena in crowds.

The workshop aimed at bringing together experts from different elds with the main objective to systematically develop and analyze a hierarchy of pedestrian traffic models. The studied hierarchy of models consists of three levels: (i) Microscopic description, (ii) Mesoscopic description and (iii) Macroscopic PDE description. The microscopic description is similar to spin-flip models, an interacting system of cellular automata. All modeling assumptions are only made at the microscopic level. Mesoscopic and macroscopic models are rigorously derived through appropriate averaging and limiting procedures; no ad-hoc assumptions on the density and velocity of the flow are used in the derivation. One of the main goals of the workshop was to naturally link three levels of description, allowing for a detailed analysis of how various terms in the macroscopic PDE model are affected by the microscopic assumptions.

Organizers:

Alina Chertock (NCSU), Pierre Degond (CNRS, France), Alexander Kurganov (Tulane U.)

1.4.2 Talks

- Jian-Guo Liu (Duke U.) *dynamics of orientational alignment and phase transition*,
- Vladislav Panferov (California State U., Northridge) *continuum description for systems of self-propelled particles used in modeling fish migration*,
- Anthony Polizzi (Tulane U.) *modern freeway traffic flow models*,
- Ilya Timofeyev (U. of Houston) *group report on traffic flow modeling*,
- Pierre Degond (CNRS) *current challenges in pedestrian dynamics and crowd modeling*,
- Alexander Kurganov (Tulane U.) *numerical methods for traffic and pedestrian flow models*,
- Sébastien Motsch (U. of Maryland) *a traffic model for pedestrian and its comparison with experimental data*.

In addition, time was reserved for the discussion of working group specific issues.

1.5 Workshop on dynamics on networks, March 21-23, 2011, SAMSI

1.5.1 General description

It has now been clearly established that many social and technological systems are complex networks. However after the structures have been estimated and the geometric properties of the graphs such as their "small world" nature have been studied, there remains the question: How does the structure of the network affect the behavior of processes taking place on the network? The SAMSI working group Dynamics ON Networks studied this question for evolutionary games and various models of the spread of opinions and epidemics, both for tree like random networks and for clustered networks. The main purpose of this workshop was to facilitate progress on group projects by bringing long distance collaborators to SAMSI and by bringing researchers whose work had been important to the group's investigations.

Organizers:

Rick Durrett (Duke U.), Alun Lloyd (NCSU), Peter Mucha (U. of North Carolina) and Alex Vespignani (Indiana U.)

1.5.2 Talks

- Vittoria Colizza (ISI and INSERM) *human mobility in an emerging epidemic: a key aspect for response planning*,
- Elizabeth Leicht (Oxford U., UK) *interacting networks: formalism and significance*,
- Joel Miller (Harvard U.) *epidemic spread in networks with one equation*,
- Duygu Balcan (Indiana U.) *phase transition in contagion processes mediated by recurrent mobility patterns*,

- Charlie Brummitt (UC Davis) *sandplie cascades on interacting tree-lie networks*,
- James Gleeson (U. of Limerick, Ireland) *analytical results for cascades on networks*,
- Mason Porter (Oxford U., UK) *cascades on networks*,
- Sid Redner (Boston U.) *the role of reinforcement in social dynamics*,
- Nicolas Lanchier (Arizona State U.) *The Axelrod model for the dissemination of culture revisited*,
- Sara Solia (Northwestern U.) *fast and slow dynamics in neural networks with small-world connectivity*,
- Deena Schmidt (Ohio State U.) *network structure and dynamics of sleep-wake regulation*,
- John McSweeney (Concordia U. and SAMSI) *single seed cascades on networks with triangles*,
- Rachel Kranton (Duke U.) *strategic interactions and networks*,
- David Sivakoff (SAMSI) *contact process on modular networks*,
- Shirhsendu Chatterjee (Cornell U.) *latent voter model on random regular graphs*,
- Shankar Bhamidi (U. of North Carolina) *flows, first passage percolation and random disorder in networks.*

In addition, time was reserved for the discussion of working group specific issues and for a poster reception.

1.6 Transition workshop, June 6-7, 2011, SAMSI

1.6.1 General description

This workshop provided a time to look back over the achievements of the program and to highlight directions for future research. This two-day workshop featured half-day sessions on the topics of the five working groups: sampling, modeling and inference; geometric and spectral properties; dynamics of networks; dynamics on networks; and flows and networks.

Organizers:

Eric Kolaczyk (Boston U.), Michael Mahoney (Stanford U.), Mason Porter (Oxford U.) and Alex Vespignani (Indiana U.)

1.6.2 Talks

- Kash Balachandran (Duke U.) *comparison of local spectral clustering algorithms*,
- Alan Lenarcic (University of North Carolina) *multiple latent trait model for interaction in expander networks*,
- Blair Dowling Sullivan (Oak Ridge NL)
- David Banks (Duke U.) *grooming networks in a baboon troop*,
- Bruce Rogers (SAMSI) *fitting monkey models*,
- Amanda Traud (NCSU) *ant modeling*,

- Tyler McCormick (Columbia U.) *surveying hidden populations through sampled respondents in a social network: a comparison of two strategies*,
- Ali Shojaie (U. of Michigan) *reconstructing directed regulatory networks from multiple steady-state and perturbed gene expression profiles*,
- Eric Kolaczyk (Boston U.) *some progress on asymptotics for ERGMs*
- Tipan Verella (U. of Virginia) *the dual of the random intersection graph*,
- Bill Shi (U. of North Carolina) *robust scaling behavior in dynamics voter models*,
- David Sivakoff (SAMSI) *evolving voter results.*

In addition, time was reserved for the discussion of working group specific issues and for a poster reception.

2 Courses and workshop for students

A one semester course for graduate students and a workshop for undergraduate students were offered as part of this program.

2.1 Graduate course: Complex Networks: Theory and Applications

This course was offered during the Fall semester of 2010. Ten graduate students from the University of North Carolina at Chapel Hill , Duke University and North Carolina State University took the course for credit while another ten people attended the class on a regular basis. Lectures were given at SAMSI on Tuesdays, 4:30-7:00 p.m.

This course focused on the mathematical and statistical analysis and modeling of networked systems, such as arise in biological, social, and technological contexts. Both static and dynamic perspectives were studied. Specific topics included network graph construction and relevant sampling issues, characterization of networks, community detection, and network modeling and inference. Various applications were considered, including in social networking, biology, and epidemiology.

Texts included

- Kolaczyk, E.D. (2009). Statistical Analysis of Network Data: Methods and Models. Springer, New York.
- Durrett, R. (2006). Random Graph Dynamics. Cambridge University Press.
- Durrett, R., (2010), "Some features of the spread of epidemics and information on a random graph", Proc. Nat. Acad. USA, (107) 2010, pp. 4491-4498

In addition, various additional publications, handouts, etc, were provided. The course outline and the corresponding instructors are listed below

1. September 7: Introduction and motivation; network mapping. (Eric Kolaczyk)
2. September 14: Network characterization. (Eric Kolaczyk)
3. September 21: Network sampling and inference. (Eric Kolaczyk)
4. September 28: Community detection (Peter Mucha)

5. October 5: Community detection continued. (Peter Mucha)
6. October 19: Network role/positional analysis (James Moody)
7. October 26: Network change: How/why in social settings (James Moody)
8. November 2: Characterization of dynamic networks (James Moody)
9. November 9: Epidemic processes on networks (Alun Lloyd)
10. November 16: Epidmic processes on networks continued. (Alun Lloyd)
11. November 30: Dynamics on networks (Rick Durrett)
12. December 7: Dynamics on networks continued (Rick Durrett)

2.2 Undergraduate student workshop

A two-Day Undergraduate Workshop was held at SAMSI on October 29-30, 2010, on the general theme of Complex Networks. This workshop was part of SAMSIs Education and Outreach Program for 2010-2011. The following lectures and events took place

- Pierre Gremaud (SAMSI) *welcome and introduction*,
- Rick Durrett (Duke U.) *what are graphs and why are they important?*
- Yi Sun (SAMSI) *MATLAB demonstration*,
- David Sivakoff (SAMSI) *random graphs*,
- Mauro Maggioni (Duke U.) *graphs, high dimensional data and visualization*,
- Hongziao Zhu (SAMSI) *R demonstration*,
- Joshua Mendelsohn (Duke U.), *using R for scientific research; R lab*,
- Pierre Gremaud (SAMSI) *career options*,
- David Banks (Duke U.) *a history of network modeling*,
- Bruce Rogers (SAMSI) and Mandi Traud (NCSU) *graph clustering with random walks; MATLAB lab*.

3 Working group reports

3.1 Sampling/modeling/inference working group

Leader: Eric Kolaczyk (Boston U.)

3.1.1 Participants

Edo Airoldi (Harvard U.), Shankar Bhamidi (UNC-CH), David Banks (Duke U.), Shirshendu Chatterjee (Cornell U.), Eric Kolaczyk (Boston U.), Pavel Krivitsky (Carnegie Melon), Yingbo Li (Duke U.), Crystal Linkletter (Brown U.), Tyler McCormick (Columbia U.), Bruce Rogers (SAMSI), Cosma Shalizi (Carnegie Melon U.), Ali Shojie (U. of Michigan), Sean Simpson (NCSU), Bruce Spencer (Northwestern U.), Tian Zheng (Columbia U.).

3.1.2 Research activities and results

The working group focused on topics that arguably involve some or all of sampling, modeling, and inference of networks. Much of the focus was restricted to static networks, but some projects had nontrivial overlap with interests in the Dynamics OF working group as well. Members of the working group are active on six group projects.

Project 1: Inference on Agent-Based Models for Networks Agent based models (ABM) specify local interaction rules between individuals and implement these rules through computer simulation. Agents are placed together in a simulated environment, and then one can observe repeated realizations of the evolution of social networks among the agents. But such applications need methods for statistical inference that relate the parameters (agent rules) to the network behavior. Important open problems include how to estimate parameters given a time series of the (simulated) observables or how assess the goodness-of-fit of the model from the given data. These challenges are difficult because each agent can have its own set of parameters, so the dimension of the parameter space grows geometrically in the number of agents. To perform dimension reduction in parameter space and make inference possible, David Banks, Bruce Rogers and Cosma Shalizi built statistical emulators for agent based models using a Bayesian framework for the calibration of computer models. As a proof of concept, they used the "Sugarscape" agent-based model from Axtel and Epstein's book Growing Artificial Societies.

Project 2: Fit Assessment for Attachment Models Much of the recent literature on real-world networks concerns mathematical models which relate their characteristics, including degree distributions, small worldedness, and clustering coefficients to parameters in various preferential attachment models. At the simplest level new nodes enter the system and decide to link themselves to nodes in the pre-existing network with probability proportional to some functional of the vertices in the network, for example their present degree. Models based on such schemes arise in a number of areas, ranging from economic systems wherein nodes try to optimize two competing functionals to evolutionary biology (Yule processes). Often one looks at the degree distribution of the real network at hand and matches it to an appropriate attachment model and then uses various properties of the model to infer properties about the network (e.g., epidemic durations). However, there is no mathematical theory adequate to judge such fits. Shankar Bhamidi, Shirshendu Chatterjee and Eric Kolaczyk explored this area in order to provide the necessary inferential theory.

Project 3: Baboon Grooming Data A consortium of primatologists has collected data on social interactions in twelve baboon troops in Kenya. A subgroup consisting of David Banks, Yingbo Li, Bruce Rogers and Ali Shojie is attempting to fit a dynamic latent space model to grooming relationships over time, using covariate information on kinship, relatedness, age, health, gender, reproductive status, dominance hierarchy, and rainfall. The key aim is to product the observed fission in the baboon troop, which entails a block model for the the proximity matrix in the latent space.

Project 4: Effective Sample Size in Network Modeling Stochastic models for network structure and processes have applications spanning the social, informational, and biological sciences. However, complex networks have complex structure, and when sophisticated models reflect this in their own dependence structure, ordinary sampling and asymptotic results used to quantify uncertainty often do not apply. Indeed, the only unambiguously independent sample for many network processes is repeated independent observations of realizations of that network over the same set of actors — something that is, with a few exceptions, not possible. And yet, observing more actors in a network clearly provides more information about the network's structure than observing fewer. Pavel Krivitsky, Sean Simpson, Crystal Linkletter, and Eric Kolaczyk worked toward establishing notions of effective sample size for network models, taking into account the dependence structure and parametrization of these models and the process by which the network is observed.

Project 5: Estimation of the Degree Distributions The degree distribution of a sampled network can differ from the degree distribution of the underlying population network, as pointed out, for example, in Subnets of Scale- Free Networks are Not Scale-Free: Sampling Properties of Networks Stumpf, Wiuf, May

(PNAS, 2005). Work by Frank has shown how to obtain unbiased estimates of the number of nodes of any degree given a simple random sample of nodes, and normalizing that distribution yields an estimate of the degree distribution. Bruce Spencer and Eric Kolaczyk worked to extend those results in two directions. First, the mean and variance of the degree distribution can be calculated from the triad census, and hence they can be estimated from an induced graph sample. Work is underway to improve Frank's estimators by adjusting them to estimates of the mean and variance based on the sample triads. Second, work is underway to extend the results from samples where nodes are included with equal probabilities to samples based on unequal probability sampling.

Project 6: Comparison of Aggregated Relation Data and Ego-Centric Sampling Tyler McCormick, Tian Zheng, and Eric Kolaczyk studied a framework for inference in multi-relational data measured through standard surveys. Our goals are (i) formalizing the information contained in two sampling methods (aggregated relational data and ego-centric fixed-size nominations) (ii) exploring the relationship between these two methods (iii) learning about the relationship between two nested networks. Mini-Project: Minimum Description Length Inference Edo Airoldi and David Banks have written an NSF proposal that includes a study of minimum description length inference for a class of latent space models. If funded, this project will surely move forward; if not, its pursuit will depend upon available energy and time.

3.1.3 Working papers and publications

1. Krivitsky, P.N. and Kolaczyk, E.D. (2012). On the question of effective sample size in network modeling. Submitted.
2. McCormick, T.H., He, R., Kolaczyk, E.D., and Zheng, T. (2012). Surveying hard-to-reach groups through sampled respondents in a social network: a comparison of two survey strategies. *Statistics in Biosciences* Special Issue on Networks, 4:1, 177-195

3.2 Dynamics OF networks working group

Leaders: David Banks (Duke U.), James Moody (Duke U.) and Peter Mucha (UNC-CH)

3.2.1 Participants

See Dynamics ON networks list.

3.2.2 Research activities and results

This working group formed as a natural outgrowth of common interests at the Opening Workshop. The group included significant overlap of interests with the other CN working groups, including but not limited to analysis of longitudinal data (overlapping with the Geometrical/Spectral group), the modeling of networks via agent based models (overlapping with Sampling/Inference), and the exploration of abstract voter models (as just one specific project overlapping with the Dynamics ON Networks group). Indeed, because of the significant overlap with the ON group, the weekly meeting of the two groups were merged for the Spring semester.

Starting at the OW, there has been intense early interest in the working group on bridging the different network perspectives commonly attributed to communities in statistics, probability, and statistical physics. Much of this interest in the group has focused on the phenomenon of explosive percolation (described at the OW in the presentation of Raissa DSouza), namely the observation in some constructive models where the giant component emerges in a seemingly sharp transition, cf. in classic models such as the Erdos-Renyi random graph model. Quantifying this notion mathematically, understanding the scaling window and understanding how small components merge into larger components eventually leading to the formation of the giant component brings in very interesting connections to the famous Smoluchowski equations in colloidal chemistry, and Markov processes such as the multiplicative coalescent. One of the hopes of the working group is to develop the techniques that will enable us to prove the first mathematically rigorous results in this context. Our early efforts in this direction after the opening workshop included conversations in the group led by Steve Fienberg and Raissa DSouza, among others. Our ongoing efforts aimed towards rigorous results are currently led by Shirsendu Chatterjee, Charlie Brummitt, Shankar Bhamidi, Raissa DSouza, and Rick Durrett.

A second model problem of focus in the working group is the analysis of Dynamic Voter Models [Holme & Newman, 2006], hybrid models that incorporate both a rewiring dynamic, so the network changes in time, and a voter interaction process which modifies the node/actor opinion to drive agreement between connected nodes in the network. When the selected rewiring dynamic is to drop connections between disagreeing actors, the two processes cooperate in pushing the network towards a final state of complete consensus within each component, but the distribution of properties of 228 those components depends on the relative frequency of activity of the rewiring and interaction processes (codified by a parameter). Our primary goals include the demonstration of the existence of a critical value for this rewiring-frequency parameter, at which the model changes behavior between the two extremes of a purely voter model dynamic and a purely rewiring dynamic, the development of model moment-closure equations for approximating this process, and the rigorous analysis of the process (especially at and near criticality). Numerical simulations are being used to explore the behavior of this system, including the different behaviors observed at different rewiring frequencies, edge densities, and opinion initial conditions. We are using a pair- approximation approach to gain a heuristic understanding of the process and the location of the critical value. We are also using a rigorous large deviations approach to demonstrate the existence of the critical value. This work is led by David Sivakoff, Bill Shi, Rick Durrett, Alun Lloyd, Peter Mucha, and Mason Porter.

A third subgroup project is investigating the role of concurrent partners (overlapping in time) in the spread of sexually transmitted diseases. STD transmission through the dynamic contact network must obviously obey an arrow of time, requiring that the start time and duration of each edge must be respected, and the resulting questions about the influence of concurrency in enhancing rapid spread are important for modeling such systems and for properly targeting behavioral and biological interventions. Multiple researchers have developed competing measures for the level of concurrency in a networks, and a wide variety of approaches have been used to study disease dynamics as a function of concurrency, including but not limited to extensive use of simulations. One part of this working groups activity in this area is in addressing the utility of different network representations thereof. For instance, the temporal information can be used to create a directed network of possible transmissibility (either exactly or nearly obeying all temporal rules). An alternative approach is to explore the allowed permutations in a fully ordered edge list (those which do not change the available transmission paths). These representations will be used to develop formal models of dynamic sexual networks, targeting bounds on the size of the largest or expected disease outbreak as a function of an appropriate global measure of concurrency. This work is led primarily by Bruce Rogers, Alun Lloyd, Jim Moody, and Peter Mucha.

3.2.3 Working papers and publications

1. Rombach, M.P., Porter, M.A., Fowler, J.H., Mucha, P.J. Core-periphery structure in networks. Submitted.

3.3 Dynamics ON networks working group

Leader: Rick Durrett (Duke U.) and Alun Lloyd (NCSU)

3.3.1 Participants

Shankar Bhamidi (UNC), Charlie Brummitt (UC Davis), Shirshendu Chatterjee (Cornell visiting Duke), Ariel Cintron-Arias (East Tennessee State U), Luis Gordillo (U Puerto Rico), Christian Gromoll (Virginia), Peter Kramer (Rensselaer), Jim Lynch (South Carolina), John McSweeney (Concorida), Peter Mucha (UNC), Mason Porter (Oxford), Michael Roberts (NC State), Puck Rombach (Oxford), Bill Shi (UNC), David Sivakoff (SAMSI), Mandy Traud (NC State), Tipan Verella (Virginia)

3.3.2 Research activities and results

A number of projects have emerged for investigation.

SIS models on networks with communities. During the Stochastic Dynamics program, John McSweeney and Bruce Rogers considered the contact process (SIS epidemic) on a strongly clustered graph, described as a stochastic block model. In the simplest case there are two communities with the probability of an edge connecting two individuals within the same community p1 and the probability of an edge connecting two individuals between communities p2 much smaller than p1. In a range of parameters, a plot of the number of infected sites versus time shows a stair step: the epidemic reaches equilibrium in one community before spreading to the other. Work on this problem is continuing in the Dynamics ON Networks working group with David Sivakoff and Rick Durrett. A paper has been published.

Nonlinear voter models. The ordinary voter model with a linear response function is now fairly well understood on complex networks due to its duality with coalescing random walk. Redner et al have studied various nonlinear versions. Chatterjee and Durrett will try to use ideas of Cox, Durrett, and Perkins to show that the mean field predictions for the nonlinear voter model are accurate when the nonlinear response function is almost linear. They think that this method can be used to cover the latent voter model where voters who have just changed their opinions enter an inert state for an exponential amount of time with mean λ if the rate λ is large. Although this is a small perturbation of the voter model, the behavior changes discontinuously, the density converges to 1/2 starting from any positive level.

In **Axelrods model**, voters have a vector of F opinions, each of which can take one of q values. Each voter x at rate 1 picks a neighbor y, and one of his opinions i. If the two agree on issue i then x picks a j and imitates ys opinion on that issue. When $q = F = 2$ this reduces to the model with leftists (00), centrists (01, 10) and rightists (11). Lanchier has proved some results in one dimension about the clustering in the case $q = F = 2$ and freezing when q is larger than F . Durrett and J.C. Li will work on generalizing the second result to two dimensions, starting with the case when both q and F are large.

Pair approximation. Mandy Traud and Michael Roberts are coding and testing the full pairwise approximation model for SIS epidemics using the equations exhibited in Keeling and Eames - Modeling Dynamic and Network Heterogeneities in the Spread of Sexually Transmitted Diseases. They plan to test actual network data with these equations, code few different closure approximations, and test different selections of parameters.

Suppressing cascades of load in interdependent networks Charlie Brummitt and his collaborators studied cascades. Understanding how interdependence among systems affects cascading behaviors is increasingly important across many fields of science and engineering. Inspired by cascades of load shedding in coupled electric grids and other infrastructure, they studied the BakTangWiesenfeld sandpile model on modular random graphs and on graphs based on actual, interdependent power grids. Starting from two isolated networks, adding some connectivity between them is beneficial, for it suppresses the largest cascades in each system. Using a multitype branching process and simulations they showed these effects and estimated the optimal level of interconnectivity that balances their trade-offs.

The Dual of a Random Intersection Graph. Christian Gromoll and Tipan Verella are interested in studying the emergence of a giant component in the dual of a generalized random intersection graph.

Conceptually, given a set of nodes and a set of attributes, a random intersection graph is a graph on nodes induced by constructing a random bipartite graph between the nodes and the attributes. Interest in random intersection graphs, has increased in recent years. However, very little attention has been devoted to studying the dual graph naturally constructed by the same process that produces the random intersection graph.

3.3.3 Working papers and publications

1. Brummitt, C.D., D'Souza, R.M., Leicht, E.A. Suppressing cascades of load in interdependent networks. *Proc. Nat'l. Acad. Sci.* 109 (2012), E680-E689.

2. Durrett, R., Gleeson, J., Lloyd, A., Mucha, P., Shi, F., Sivakoff, D., Varghese, C. Graph fission in an evolving voter model. *Proc. Nat'l. Acad. Sci.* 109 (2012), 3682-3687.

3. Durrett, R. and Sivakoff, D. Contract process on modular networks. In progress.

4. Verella, J.T. A randomized approximation algorithm for community detection in complex networks. In preparation.

3.4 Geometrical/spectral analysis working group

Leaders: Mauro Maggioni (Duke U.) and Michael Mahoney (Stanford U.)

3.4.1 Participants

Prakash Balachandran (Duke), Pietro Poggi Corradini (Kansas State), Mihail Cucuringu (Princeton), Raymond Falk, Mauro Maggioni (Duke), Michael Mahoney (Stanford), Peter Mucha (UNC Chapel Hill), Mason Porter (Oxford), Ali Shojaie (U. Wisconsin Madison), Blair Sullivan (Oak Ridge Lab).

3.4.2 Research activities and results

The group has been studying a number of topics in spectral and multiscale techniques for graphs, and their geometric and algorithmic implications. It focused on several problems.

1. Multiscale conductance and time-series Networks. Inspired by recent work on multislice networks, and on multiscale conductance, the group studied the behavior of multiscale methods (e.g. conductance, community detection, diffusions) on multislice networks.

2. Soft clustering. Most existing community-detection algorithms find hard partitions in the network, while in some situations it seems more natural to construct soft- partitions. Few methods do that, but it is a new topic and little is well-understood. It would also be interesting to apply new ideas to the case of multislice networks.

3. Tree decompositions. These decompositions simplify graphs, leading to faster optimization/solution of problems which are computationally hard. They are related to important problems in statistics (especially in graphical models) but such ties are not well- understood. Moreover, these decomposition algorithms are right now very combinatorial in nature, and one may wonder if they would become easier/more stable by weaking some of the requirements, while still being useful.

4. Community Detection. Spectral techniques based on intermediate eigenvectors of the Laplacian and/or modularity may be used to detect communities.

5. Scan Statistics/sparse regression. Scan statistics on graphs is not well understood and only few technique exist, albeit many applications seem to be in need of such tools, such as epidemics (early detection), genomics (identification of pathways relevant in a disease), etc...

6. Epidemic/Reactive Networks. Corradini collected and is processing data from questionnaires that allow mapping social interactions and behavior in a community, both under normal conditions and in the case of the emergence of an epidemics. Many questions arise: would the network change enough to prevent the epidemics? How would the network change? What would be efficacious ways of modifying the network to prevent the spread?

3.4.3 Working papers and publications

3.5 Modeling flows

Leader: Taufiquar Khan (Clemson U.)

3.5.1 Participants

Aranya Chakrabortty (FREEDM Center, NCSU), Alina Chertock (NCSU), Alex Kurganov (Tulane U.), George Michailidis (U. of Michigan), Anthony Polizzi (Tulane U.), Sean Simpson (NCSU), Yi Sun (SAMSI), Ilya Timofeyev (U. of Houston), Zifang Wang (UC Davis).

3.5.2 Research activities and results

The group looked at two specific “flow applications” on networks: traffic flows and smart grids.

Traffic flows

Most of the activities of that subgroup were centered around the derivation of new continuum models based microscopic approaches. In particular, new stochastic optimal velocity models were proposed and analyzed. Several numerical issues linked to the resolution of nonlocal interactions were also studied. The group also considered a new multi-class traffic flow model with Arrhenius look-ahead dynamics. They first derived a cellular automatum model and study its numerical solution obtained by the Monte Carlo simulations. They then passed to the PDE limit, which is a hyperbolic system of conservation laws with global fluxes. The solution of the PDE system was numerically studied using the central-upwind scheme. Finally, the group also considered pedestrian traffic flows. One-dimensional models for the behavior of pedestrians in a narrow street or corridor were proposed and investigated. The corresponding model was formulated at the microscopic level as a cellular automata model with explicit rules for pedestrians moving in two opposite directions. Coarse-grained analogs was derived leading to the coupled system of partial differential equations for the density of the pedestrian traffic. The group compared and contrasted the behavior of the microscopic stochastic model and the resulting coarse-grained equations for various parameter settings and initial conditions. Diffusive higher-order corrections were also discussed. A related workshop was held on February 14-16, 2011, see Section 1.4.

Smart grid

There is tremendous growth in scientific research efforts in “smart grid related field. This new trend in growth is particularly clear from the number of organized groups, centers, consortia developed just in the past few years (in addition to the existing ones). The electrical network, a good example of a complex network, has many different aspects and fits well into several areas of applied mathematics and statistics: control and dynamical systems, network optimization and power quality, network communication/information theory.

These aspects of the power network and the new development in the area of power grid brings new challenges that have significant overlap with existing applied mathematical and statistical techniques as well as there is a need to develop new techniques to tackle the relevant pressing problems to develop an efficient and sustainable power network. In particular, the group investigated the following issues.

1. Graph and Model Reduction for Large Scale Power Systems Network. In this thrust, the groups aimed at reducing the complex system of generators, buses with loads etc to a reduced graph problem and then formulate the model reduction problem for the dynamics of the major components.

2. Optimal Load Management in the Network. An optimization problem for the power flow equation was formulated in order to extend to dynamic settings the usual steady state setting (which is more or less already done).

3. Identification of Critical Paths in Power Network. Here is the goal was to identify which vertices/paths of the graph have critical power connections. The group then investigated how the system configuration changes if one or more of these critical paths are removed from the network for example does it have significant impact to cause blackouts etc.

SAMSI held a related Smart grid workshop in the Fall of 2011 as part of the Uncertainty Quantification program. Several papers are in preparation.

3.5.3 Working papers and publications

1. Sun, Y. and Timofeyev, I. New stochastic models for traffic flows. In preparation.

2. Chertock, A., Kurganov, A., Polizzi, A. Multi-Class Traffic Flow Model with Look-Ahead Dynamics. Submitted.

3. Chertock, A., Kurganov, A., Polizzi, A., Timofeyev, I. Pedestrian Flow Models with Slowdown Interactions. Submitted.