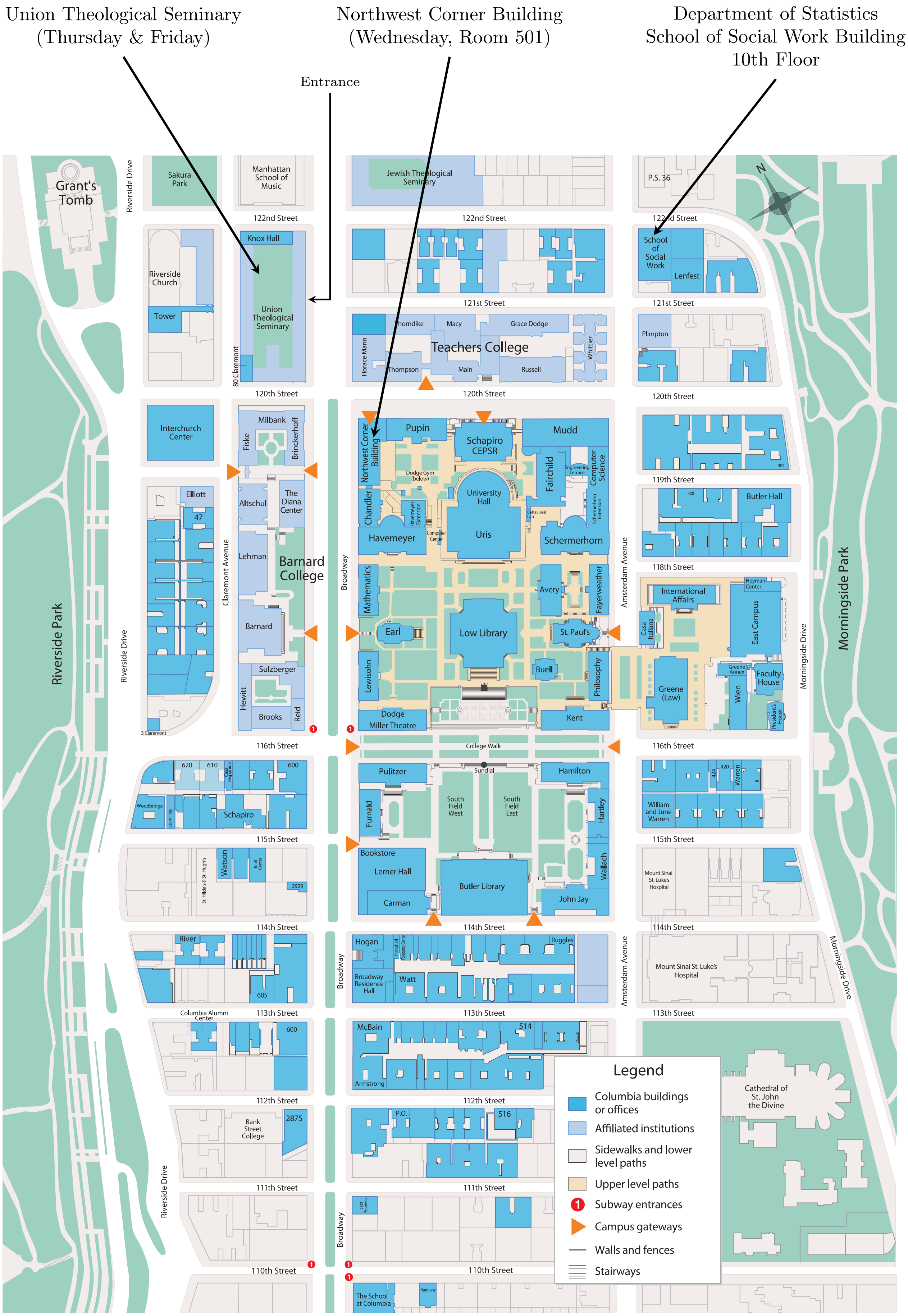
Networks, Random Graphs and Statistics

4–6 May 2016

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK


Union Theological Seminary
(Thursday & Friday)
Northwest Corner Building
(Wednesday, Room 501)
Department of Statistics
School of Social Work Building
10th Floor
Entrance
Grant's Tomb
Sakura Park
Manhattan School of Music
Jewish Theological Seminary
P.S. 36
122nd Street
Knox Hall
Riverside Church
Tower
Union Theological Seminary
80 Claremont
School of Social Work
Lenfest
121st Street
Thorndike
Macy
Grace Dodge
Horace Mann
Teachers College
Whittier
Thompson
Main
Russell
Plimpton
120th Street
Interchurch Center
Milbank
Fiske
Brinckerhoff
Northwest Corner Building
Pupin
Schapiro CEPSR
Mudd
Fairchild
Computer Science
Engineering Terrace
Dodge Gym (below)
University Hall
Uris
Chandler
Havemeyer
Schermerhorn
119th Street
Butler Hall
118th Street
Elliott
47
Altschul
The Diana Center
Lehman
Barnard College
Barnard
Sulzberger
Hewitt
Brooks
Reid
Mathematics
Earl
Low Library
St. Paul's
Avery
Fayerweather
Buell
Philosophy
Lewisohn
Dodge Miller Theatre
Kent
International Affairs
Heyman Center
East Campus
Casa Italiana
Greene (Law)
Greene Annex
Faculty House
Wien
President's House
Claremont Avenue
Riverside Drive
Broadway
Amsterdam Avenue
Morningside Drive
Riverside Park
Morningside Park
116th Street
College Walk
Pulitzer
Sundial
Hamilton
Furnald
South Field West
South Field East
Hartley
Wallach
620
610
600
Woodbridge
Schapiro
William and June Warren
Warren
115th Street
Watson
2929
Bookstore
Lerner Hall
Carman
Butler Library
John Jay
Mount Sinai St. Luke's Hospital
114th Street
River
605
Hogan
Ruggles
Broadway Residence Hall
Watt
Columbia Alumni Center
113th Street
McBain
514
600
Armstrong
112th Street
Bank Street College
2875
P.O.
516
Cathedral of St. John the Divine
111th Street
110th Street
The School at Columbia
Harmony
Legend
Columbia buildings or offices
Affiliated institutions
Sidewalks and lower level paths
Upper level paths
Subway entrances
Campus gateways
Walls and fences
Stairways

	Wednesday	**Thursday**	**Friday**
Venue	Room 501, NWC	Union Seminary	Union Seminary
09:00	*Opening*		
09:15	Peter Bickel	Olav Kallenberg	Mark Handcock
10:00	*Coffee break*	*Coffee break*	*Coffee break*
10:30	Sofia Olhede	Daniel Roy	Harry Crane
11:15	David Choi	Cameron Freer	Andrew Gelman
12:00	*Lunch break*	*Lunch break*	*End of Workshop*
14:00	Hanna Wallach	Patrick Wolfe	
14:45	Chao Gao	Ismaël Castillo	
15:30	*Coffee break*	*Coffee break*	
16:00	Patrick Perry	Eric Kolaczyk	
Evening	Poster session 17:00-19:00	Dinner for speakers 18:00	

Talk Abstracts

Peter Bickel (UC Berkeley)

Fitting block models with covariates: "Likelihood" methods, sparsity and selection of the number of blocks

We introduce block models with edge and block covariates, along the lines of Hoff, Handcock, Raftery (2002), specializing to covariate forms of the types proposed by Zhao, Levina, Zhu (2014) and generalizing that of Newman, Clauset(2015). We study mean field variational fitting in this situation along the lines of Celisse, Daudin, Pierre (2011) and B., Choi, Chang, Zhang (2013). Using theory, examples, and simulations we argue that these methods, possibly regularized lead to models with a sparse number of blocks. (Joint work with Purna Sarkar (U. of Texas, Austin))

Sofia Olhede (University College London)

Defining networks of tree species

I will discuss defining networks from observations of tree species. This talk will discuss how to quantify co-associations between multiple and inhomogeneous point-process patterns, and how to identify communities, or groups, in such observations. The work is motivated by the distribution of tree and shrub species from a 50 ha forest plot on Barro Colorado Island. We show that our method can be used to construct biologically meaningful subcommunities that are linked to the spatial structure of the plant community.

This is joint work with David Murrell and Anton Flugge

David Choi (CMU)

Co-clustering of non-smooth graphons

Theoretical results are becoming known for community detection and clustering of networks; however, these results assume an idealized generative model that is unlikely to hold in many settings. Here we consider exploratory co-clustering of a bipartite network, where the rows and columns of the adjacency matrix are assumed to be samples from an arbitrary population. This is equivalent to assuming that the data is generated from a nonparametric model known as a graphon. We show that co-clusters found by any method can be extended to the row and column populations, or equivalently that the estimated blockmodel approximates a blocked version of the generative graphon, with generalization error bounded by $n^{-1/2}$. Analogous results are also shown for degree-corrected co-blockmodels and random dot product bipartite graphs, with error rates depending on the dimensionality of the latent variable space.

Hanna Wallach (Microsoft Research)
Bayesian Poisson Tensor Decomposition for International Relations

Like their inhabitants, countries interact with one another: they consult, negotiate, trade, threaten, and fight. These interactions are seldom uncoordinated. Rather, they are connected by a fabric of overlapping communities, such as security coalitions, treaties, trade cartels, and military alliances. A single country can belong to multiple communities, reflecting its many overlapping identities, and can engage in both within- and between-community interactions, depending on the capacity in which it is acting. In this talk, I will introduce two tensor decomposition models for modeling interaction events of the form "country i took action a toward country j at time t." The first model (Bayesian Poisson CP decomposition) discovers coherent threads of events, characterized by sender countries, receiver countries, action types, and time steps; the second model (Bayesian Poisson Tucker decomposition) discovers latent country–community memberships, including the number of latent communities, as well as directed community–community interaction networks that are specific to "topics" of similar action types. I will demonstrate that these models infer interpretable latent structures that conform to and inform our knowledge of international relations. Many existing models for discrete data (such as networks and text) are special cases of these models, including infinite relational models, stochastic block models, and latent Dirichlet allocation. As a result, Bayesian Poisson tensor decomposition is a general framework for analyzing and understanding discrete data sets in the social sciences.

Chao Gao (Yale)
Adaptive Graphon Estimation

Graphon is a central object in network analysis. In this talk, I will establish an intuitive connection between nonparametric graphon estimation and nonparametric regression. Then I will show that the minimax rate of graphon estimation consists of two parts: the nonparametric part and the clustering part. An interesting implication is that the smoothness of the graphon does not affect the rate once it is greater than 1. When the smoothness is not given, an adaptive Bayes procedure is proposed to achieve optimal posterior contraction. The Bayes procedure is a special case of a general Bayes framework for structured linear models. Under this general framework, the problems of graphon estimation, sparse linear regression, dictionary learning and many other statistical problems can be understood and solved from a unified perspective.

Patrick Perry (New York University)
Block models for clustering directed networks and nonsymmetric data matrices

Much of the work on network clustering (âĂIJcommunity detectionâĂİ) focuses on symmetric, binary networks. In many real-world applications, including recommender systems and communications, it is more natural to treat the underlying network as directed and weighted. In this talk, I will show how to extend the block models and the profile-likelihood-based clustering method popularized by Bickel and Chen (2009) to arbitrary nonsymmetric data matrices. This leads to a biclustering procedure with similar consistency properties to the analogous method for symmetric binary networks. I will demonstrate the method with applications to congressional voting data and microarray analysis. This talk is based on work done jointly with Cheryl Flynn: http://arxiv.org/abs/1206.6927 .

Olav Kallenberg (Auburn University)
An introduction to multivariate probabilistic symmetries

My plan is to give a general survey of exchangeable and related random arrays, some of which arise naturally in the study of random networks. For three of the basic probabilistic symmetries, we can derive general representations that give a fairly complete understanding of independence and other structural properties. Though some of those results are extremely tricky to prove, I shall try to keep my talk on a rather elementary level that avoids the technical subtleties of the subject.

Daniel Roy (University of Toronto)
Sparse Random Graphs arising from Exchangeable Random Measures

We introduce a class of random graphs on $\mathbb{R}$ defined by the exchangeability of their vertices. A straightforward adaptation of a result by Kallenberg yields a representation theorem: every such random graph is characterized by three (potentially random) components: a nonnegative real $I \in \mathbb{R}_+$, an integrable function $S : \mathbb{R}_+ \to \mathbb{R}_+$, and a symmetric measurable function $W : \mathbb{R}_+^2 \to [0, 1]$ that satisfies several weak integrability conditions. We call the triple (I, S, W) a graphex, in analogy to graphons, which characterize the (dense) exchangeable graphs on $\mathbb{N}$. I will present some joint work with Victor Veitch on the structure and consistent estimation of these random graphs.

Cameron Freer (MIT)
Exchangeable constructions of countable structures

The Aldous-Hoover-Kallenberg theorem and the theory of graph limits connect three kinds of objects: sequences of finite graphs, random countably infinite graphs, and certain continuum-sized measurable "limit" objects (graphons). Graphons induce exchangeable countably infinite graphs via sampling, and all exchangeable graphs arise from a mixture of such sampling procedures – a two-dimensional generalization of de Finetti's theorem.

This naturally leads to the question of which countably infinite graphs (or other structures) can arise via an exchangeable construction. More formally, consider a random structure with a fixed countably infinite underlying set. The random structure is *exchangeable* when its joint distribution is invariant under permutations of the underlying set. For example, the countably infinite Erdős-Rényi graph is exchangeable; moreover, it is almost surely isomorphic to a particular graph, known as the Rado graph, and so we say that the Rado graph admits an exchangeable construction. On the other hand, one can show that no infinite tree admits an exchangeable construction.

In joint work with Ackerman and Patel, we provide a necessary and sufficient condition for a countably infinite structure to admit an exchangeable construction. We also address related questions, such as what structures admit a unique exchangeable construction, and give examples involving graphs, directed graphs, and partial orders.

Joint work with Nathanael Ackerman, Diana Cai, Alex Kruckman, Aleksandra Kwiatkowska, Jaroslav Nešetřil, Rehana Patel, and Jan Reimann.

Patrick Wolfe (University College London)
TBA

Ismaël Castillo (University of Paris VI)
On uniform estimation of some random graph parameters

Suppose one fits a graphon model to an observed graph. If the model is specifically a stochastic block model, and if the precise number of groups in the block model is known, some of the parameters can be estimated at a "good" rate; this has been clarified in recent contributions by other authors. We show that good rates are not achievable uniformly over the whole parameter space, which has implications in particular if the number of classes is not known: close to the boundary between k- and (k-1)-class models, the rate drops. We provide minimax bounds that make this phenomenon precise; the bounds are non-asymptotic. The results and their proofs also have implications for more general graphon models beyond the block model case.

This is joint work with Peter Orbanz (Columbia)

Eric Kolaczyk (Boston University)
Estimating Network Degree Distributions from Sampled Networks: An Inverse Problem

Networks are a popular tool for representing elements in a system and their interconnectedness. Many observed networks can be viewed as only samples of some true underlying network. We study the problem of how to estimate the degree distribution of a true underlying network from its sampled network, under various common network sampling designs. We show that it can be formulated as an inverse problem that is, in many cases, ill-posed. Accordingly, we offer a penalized least-squares approach to solving this problem, with the option of additional constraints. The resulting estimator is a linear combination of singular vectors of a matrix, relating the expectation of our sampled degree distribution to the true underlying degree distribution, which is defined entirely in terms of the sampling plan. Choice of the penalization parameter is made through a Monte Carlo version of Stein's unbiased risk estimation. We present the results of a simulation study, characterizing the performance of our proposed method, and we illustrate its use in the context of monitoring large-scale social media networks.

Mark Handcock (UCLA)
Some new models for social networks

Random graphs, where the connections between nodes are considered random variables, have wide applicability in the social sciences. Exponential-family Random Graph Models (ERGM) have shown themselves to be a useful class of models for representing complex social phenomena.

In this talk we will consider some new classes of models that generalize ERGM in different ways. First, we model the attributes of the social actors as random variates, thus creating a random model of both the relational and individual data, which we call Exponential-family Random Network Models (ERNM). This provides a framework for expanded analysis of network processes, including a new formulation for network regression where the outcomes, covariates and relations are socially endogenous. We illustrate this with a new class of latent cluster models and network regression. Next we introduce a class of models we call Tapered Exponential-family Random Network Models (TERNM). These models remove the degeneracy properties that hamper ERGM and ERNM while retaining there advantages. We show how these models can provide good fits to large networks. Finally we introduce spatial temporal exponential-family of point processes (STEPP) models to jointly represent the co-evolution of social relations and individual behavior in discrete time.

This is joint work with Ian E. Fellows and Joshua D. Embree.

Harry Crane (Rutgers)
Edge exchangeability: a new foundation for modeling network data

Exchangeable models for vertex labeled graphs cannot replicate the large sample behaviors of sparsity and power law degree distributions observed in many network datasets. To address this issue, we introduce the principle of edge exchangeability, which is more natural for most applications and admits models for networks with sparse and/or power law structure. The vertices in an edge exchangeable network arrive in size-biased order according to their degree, further explaining why vertex exchangeability is an untenable assumption for many applications. Our discussion settles a longstanding question of statistical network modeling and presents a new framework within which to develop theory and methods.

Joint work with Walter Dempsey.

Andrew Gelman (Columbia University)
Learning about networks using sampling

In survey research we are often interested in hidden, hard to reach, and marginalized populations. Even when potential respondents are sitting out in plain sight, it can be difficult to gather a representative sample, to persuade sampled people to respond to a survey, and to get accurate responses. In addition, we can be interested not just in hard-to-reach groups but how they relate to the larger society, in examples ranging from political persuasion to the spread of disease. In this talk we consider two related challenges: (a) learning about networks using sampling, and (b) using network sampling to learn about a population. We also discuss a current project on the study of political penumbras, which is joint research with Yotam Margalit.