Online Proceedings

2A

Graphs and Geometry

Session Moderator: Werner Stuetzle, Statistics
085 MGH
3:45 PM to 5:15 PM

* Note: Titles in order of presentation.

Classifying Freight Carriers in Supply Chains

Andrea Jeanne (Andrea) Gagliano, Senior, Business Administration (Finance), Mathematics
Mary Gates Scholar
Mentor: Anne Goodchild, Civil & Environmental Engineering

Computer-generated transportation models are used by State Departments of Transportation to simulate flows of goods and people on roadways. Infrastructure investments and policy decisions can be supported by results of these models. Passenger travel is well developed and uses activity-based models using characteristics of the household to more accurately model travel behavior. Freight, on the other hand, is not as developed, and uses simple algorithms to predict goods movement. This research focuses on characterizing freight travel in the context of supply chains to more appropriately model frieght behavior. A survey was conducted of motor carriers in Oregon and Washington to capture fleet statistics, carrier services, travel distances, time of day travel patterns, and company characteristics to find which factors differentiate motor carriers. The results revealed a key distinction between "Supply Chain Node Carriers" and "Only Transportation Carriers," with "Supply Chain Node Carriers" being split further into those companies linked to raw materials, manufacturing facilities, storage facilities, distribution centers, and retail outlets. Suggestions are made on how to implement these findings into state transportation model developments to enhance them. These more precise models allow for evidence-backed infrastructure investments and policy decisions.

A Political Redistricting Tool for the Rest of Us

Evan Kleiner, Junior, Mathematics, Whitman College
Mentor: Albert Schueller, Mathematics, Whitman College

For our research project, we asked the question: 'Is there some way to automate the redrawing of district lines in each state mandated by the constitution after every decennial census?' This presentation will consist of a brief overview of the factors considered in redistricting (population, compactness, alignment with city and county lines, etc.) followed by a demonstration and discussion of the software application that we developed to address this problem. The presentation will conclude with a mention of potential political and social drawbacks of automated redistricting.

Stable Marriage when You are Not Sure Who Would Make a Good Long-Term Partner

Eric Christopher Lei, Senior, Computer Science, Economics, Mathematics
Mary Gates Scholar
Mentor: Anna Karlin, Computer Science and Engineering

Can a group of men and women be paired together to guarantee no divorces or scandals? In 1962, two economists posed this question mathematically and offered a solution. A 2012 Nobel Prize went to two of the key researchers that worked on this problem. Their solution took the form of an algorithm, a systematic procedure that could be performed by a computer. The solution to the "stable marriage problem" has both interesting theoretical properties and real-world applications, such as the assignment of medical students to hospitals. In the original version of the problem, each man and woman know their preferences over all their potential partners. We are considering a variant in which they only know some of their preferences. The algorithm not only needs to figure out whom to match with whom, but also what information to gather on preferences. We present our preliminary results on efficient algorithms for this problem. This is joint work with Anna Karlin and Jamie Morgenstern.

Structured Learning of Gaussian Graphical Models

Palma Alise Den Nijs (Palma) London, Senior, Mathematics, Electrical Engineering
Mary Gates Scholar
Mentor: Maryam Fazel, Electrical Engineering
Mentor: Daniela Witten, Department of Biostatistics
Mentor: Su-In Lee, Computer Science & Engineering
Mentor: Karthik Mohan

We estimate a Gaussian graphical model corresponding to a covariance matrix using machine learning techniques. A motivating application is to identify genes associated with cancer in the human genome. Suppose we have genetic data from two people, one of whom is healthy, while the other has cancer. Our goal is to identify gene expressions that differ between datasets and contribute to cancer. We model each dataset as a Gaussian graphical model, which is a network of edges and nodes. Each node represents a random variable, or gene. Each edge represents the dependence between the two variables it connects. An edge with zero value indicates the two variables are independent when conditioned on the rest of the variables. Such networks are analogously represented as inverse covariance matrices. Our goal is to use empirical genetic information from cancer patients to recover the true inverse covariance matrices. We suppose that the inverse covariance matrices have a certain structure and we set up a problem to encourage that structure. Structure refers to the relative magnitudes of matrix elements. A structure of interest is a *hub node*, which is highly connected to other nodes. These nodes play a regulatory role in gene expression and may be related to cancer. Our goal is to identify hub nodes in the difference between two inverse covariance matrices. We formulate this question as a convex optimization problem. We encourage the solution to have a certain structure by adding regularizer terms to the objective function. We recover the overall structure of the inverse matrices, while simultaneously encouraging the solution to have specific structures of interest. An alternating directions method of multipliers (ADMM) algorithm is used to solve the optimization problem. We run the algorithm on synthetic genetic data, and on real gene expression data to identify cancer related genes.

Generalized Ramsey Numbers for Near-Diagonal Pan Graphs and Relatively Prime Tadpole Graphs

Timothy Perisho, Fifth Year, Mathematics, Philosophy, Seattle Pacific University
Mentor: Steve Johnson, Mathematics, Seattle Pacific University

Ramsey Theory is an intricate field of pure mathematics, but it also has applications in computer science, where graphs are commonlu used as data structures. Every pair of ”forbidden” graphs has a Ramsey number, which indicates when one of those graphs must occur. A Ramsey number for a pair of forbidden graphs is the smallest number of points R for which every edge-coloring of the complete graph on R points includes one of the forbidden graphs in its respective color. In this paper we use two original methods to find and prove new Ramsey number formulas for certain tadpole graphs based upon known Ramsey number formulas for arbitrary cycle graphs. A tadpole graph Qn,t is formed by connecting one end of a path graph on t points (called the “tail”) to a cycle of size n (called the “head”). A pan graph Qn is defined as a tadpole graph with t=1. We first use an intuitive counting method to prove exact Ramsey number formulas for all “near-diagonal” pan graphs, i.e. those pairs of pan graphs where the maximum cycle size of the pair is no more than 5/2 times the minimum cycle size. Then, we demonstrate a number-theoretic method to find exact Ramsey numbers for arbitrarily far-from-diagonal tadpole pairs with arbitrarily long tails. However, the second method requires that the tadpoles meet number-theoretic constraints (rather than size constraints). For example, in the diagonal cases, n-1 must be relatively prime to n+t, where either n-1 has an odd order (mod n+t) or the additive inverse of n-1 has an even index relative to n-1 (mod n+t). These “scattered” on- and off-diagonal results provide upper bounds on the Ramsey numbers for other tadpole pairs that fail to meet the number-theoretic constraints.