Social Impact of Network Science - SINS 2016

JUNE 8-10 2016
PALAZZO FRANCHETTI, VENICE

Human Decisions and Machine Predictions
Jon Kleinberg (Cornell University)

An increasing number of domains are providing us with detailed trace data on human decisions, often made by experts with deep experience in the subject matter. This provides an opportunity to use machine-learning prediction algorithms to ask several families of questions --- not only about the extent to which algorithms can outperform expert-level human decision-making in specific domains, but also whether we can use algorithms to analyze the nature of the errors made by human experts, to predict which instances will be hardest for these experts, and to explore some of the ways in which prediction algorithms can serve as supplements to human decision-making in different applications.
In this talk, I'll explore this theme by drawing on a line of recent projects; all are joint with Sendhil Mullainathan, and include collaborations with Ashton Anderson, Himabindu Lakkaraju, Jure Leskovec, Annie Liang, and Jens Ludwig.

Discrimination in Human vs. Algorithmic Decision Making
Krishna Gummadi (Max Planck Institute for Software Systems)

Algorithmic (data-driven) decision making is increasingly being used to assist or replace human decision making in a variety of domains ranging from banking (rating user credit) and recruiting (ranking applicants) to judiciary (profiling criminals) and journalism (recommending news-stories). Against this background, in this talk, I will pose and attempt to answer the following high-level questions: (a) Can algorithmic decision making be discriminatory? (b) Can algorithmic discrimination be controlled? i.e., can algorithmic decision making be made more fair? (c) Can algorithmic decisions be used to detect and avoid implicit biases in human decisions?

Is Your Big Data Too Big Or Too Small: Sample Complexity and Generalization Error
Eli Upfal (Brown University)

Conversational markers of social dynamics
Cristian Danescu-Niculescu-Mizil (Cornell University)

More and more of life is now manifested online, and many of the digital traces that are left by human activity are in natural-language format. Exploiting these resources under a computational framework can bring a new understanding of online social dynamics.
In this work I will introduce a computational framework for modeling conversation dynamics and for extracting the social signals they encode. I will show how these signals can be predictive of the future evolution of a dyadic relationship. In particular, I will characterize friendships that are unlikely to last and examine temporal patterns that foretell betrayal in the context of the Diplomacy strategy game.

MapReduce and Streaming Algorithms for Diversity Maximization in Doubling Metric Spaces
Geppino Pucci (University of Padova)

Given a point dataset in a metric space, a diversity maximization algorithm must identify a small subset of points that maximizes some objective diversity measure, such as the minimum distance between any pair of points in the selected set, or the average distance between a pair of points in the set. Diversity maximization problems are computationally hard, hence only approximate solutions can be hoped for. Moreover, since the applications of diversity maximization are mostly in the realm of massive data analysis, we are interested in efficient algorithms for computational settings that can handle very large data sets, in particular, the MapReduce and the Streaming models of computation. In this paper, we focus on metric spaces of bounded doubling dimension, which include Euclidean spaces of constant dimension. Our results show that despite the limitations of the computational models under consideration, for a variety of objective diversity measures we can efficiently achieve an $(\alpha+\epsilon)$-approximation, for any constant $0<\epsilon<1$, where α the is best approximation ratio achieved by a standard linear-space polynomial time sequential algorithm for the same diversity criterion.
Joint work with Matteo Ceccarello, Andrea Pietracaprina, and Eli Upfal

Models for Discrete Choice
Andrew Tomkins (Google)

The online world is rife with scenarios in which a user must select one from a finite set of alternatives: which movie to watch, which song to play, which camera to order, which website to visit. This talk will begin with an overview of these problems and the models used to study them, and will describe some recent work on large-scale learning of the structure of choice.

AI Dangers, Imagined and Real
Devdatt Dubhashi (Chalmers University)

Thompson sampling is a powerful and versatile method for exploiting partial information in decision making under uncertainty. We apply Thompson sampling for the case of graph structured feedback and show that both theoretically and empirically it outperforms all previous methods.

Five Things I learned from a Great Man
Prabhakar Raghavan (Google)

Spatio-temporal Networks: Reachability, Centrality, and Robustness
Mirco Musolesi (University College London)

While recent advances in spatial and temporal networks have enabled researchers to more-accurately describe many real-world systems, existing models do not capture the combined constraint that space and time impose on the relationships and interactions present in a spatio-temporal complex network. This has important consequences, often resulting in an over-simplification of the resilience of a system and obscuring the network's true structure. In this paper, we study the response of spatio-temporal complex networks to random error and systematic attack. Firstly, we propose a model of spatio-temporal paths in time-varying spatially embedded networks. This model captures the property that, in many real-world systems, interaction between nodes is

non-instantaneous and governed by the space in which they are embedded. Secondly, using numerical experiments on four empirical examples of such systems, we study the effect of node failure on a network's topological, temporal, and spatial structure. We find that networks exhibit divergent behaviour with respect to random error and systematic attack. Finally, to identify weaknesses specific to the behaviour of a spatio-temporal system, we introduce centrality measures that evaluate the importance of a node as a structural bridge and its role in supporting temporally efficient flow through the network. We explore the disruption to each system caused by attack strategies based on each of these centrality measures. This exposes the complex nature of fragility in a spatio-temporal system, showing that there is a variety of failure modes when a network is subject to systematic attack.
Paper available at this web-address: http://arxiv.org/abs/1506.00627

Hover and QR-code Based Attacks in Android OS to Steal User Input and Bitcoins
Julinda Stefa (Sapienza University of Rome)

We will present two new types of attack to the security and privacy of Android users. The first type targets devices supporting the floating touch or hover technology, that allows users to interact with the device without actually touching it. This attack aims at stealing all of the input typed by the user. It threatens not only user passwords and sensitive information, but also her personal communication with family members, friends, colleagues, and so on, through apps like WhatsApp, Facebook Messenger, Viber, or similar. The second type of attack is related to a software-enabled technology: the Bitcoin mobile wallet apps and their QR-code based transaction methodology. The goal is to steal bitcoins during a transaction. Both attacks are witnesses of the not-so-seldom improper support that current mobile OSs give to new technology; be it hardware or software enabled.
We prototype and test the two attacks on Android devices of different sizes and brands. Our experimental results show that the hover related attack can detect user passwords with 88 % of accuracy and correctly capture 96.5 % of regular text in private emails or text messages. The bitcoin wallet attack succeeds on stealing bitcoins during QR-code based transactions from the payee's wallet without any of the two parties involved in the transaction noticing any anomaly. Finally, we discuss and analyze the system flaws that enable these attacks, propose possible countermeasures and solutions that would help to mitigate them in the future, and to design more secure systems in view of the above threats.

The Limits of Recommendations
Marco Bressan (Sapienza University of Rome)

Modeling Contagion and Systemic Risk
Monica Billio (Ca'Foscari University of Venice)

Learning from Low Regret Algorithms
Sergei Vassilvitskii (Google)

Suppose you are in a competitive setting, and the entity in charge is using low-regret learning to decide between the options available. In this work we show how to use the actions taken by the learner to infer the valuations of others, and effectively compete against them.

ABRA: Venice, Sampling, and Betweenness Centrality Estimation
Matteo Riondato (Two Sigma Investments)

From the 14th century to the 18th, Venice was the most central node in the graph of trade routes between Europe, Asia, and Africa: most shortest paths went through Venice. Indeed the communications between the near East and Europe suffered a serious hit when Venice fell in 1797: the network was not robust, as could be measured by the Venice's betweenness centrality in the trade graph. Today's networks are much larger than those of the past and include millions of vertices and edges. Betweenness centrality plays an important role of measuring the network robustness. It is extremely expensive to compute the exact betweenness centrality of all vertices in a massive graph. We show how to use Rademacher Averages and progressive sampling to fastly obtain a probabilistically guaranteed high quality approximation of the betweenness of all vertices.

Rank Monotonicity in Centrality Measures
Sebastiano Vigna (University of Milan)

A measure of centrality is rank monotone if after adding an arc from x to y, all nodes with a score smaller than (or equal to) y have still a score smaller than (or equal to) y. If in particular all nodes with a score smaller than or equal to y get a score smaller than y (i.e., all ties with y are broken in favor of y) the measure is called strictly rank monotone. We prove that harmonic centrality is strictly rank monotone, whereas closeness is just rank monotone on strongly connected graphs, and that some other measures, including betweenness, are not rank monotone at all (sometimes not even on strongly connected graphs). Among spectral measures, damped scores such as Katz's index and PageRank are strictly rank monotone on all graphs, whereas the dominant eigenvector is strictly monotone on strongly connected graphs only.

Behavioral Phenotyping of Digital Health Tracker Data
Luca Foschini (Evidation Health)

Wearable technologies have seen a tremendous development in recent years: step and calorie counters have long made their way to our phones and watches, and new consumer-grade sensors can now measure a breadth of physiological functions that until recently could only be found in the monitoring equipment of intensive care units. However, despite the undisputed short-term benefits due to the user increased awareness, quantifying the potential value of wearable technologies in improving longer-term health outcomes remains an open question. In this talk we will present evidence that activity tracking data contains a wealth of information that is predictive of metrics directly related to health outcomes, ranging from medication adherence to lifestyle. To this end, we will show how machine learning tools need to be adapted to take full advantage of densely sampled, multi-variate time series of tracker data. Finally, we will reflect on how the predictive power of wearable data can be harnessed to inform behavior change interventions.

Algorithms for Network Analysis of Cancer Mutations
Fabio Vandin (University of Padova)

Cancer is a disease that is driven by somatic mutations accumulating in the genome during an individual’s lifetime. Recent advances in DNA sequencing technology have enabled genome-wide measurements of these mutations in large cohorts of cancer patients. A major challenge in analyzing these data is to distinguish functional "driver" mutations responsible for cancer progression from

“passenger”, random mutations not related to the disease. Recent cancer sequencing studies have shown that somatic mutations are distributed over a large number of genes. This mutational heterogeneity is due in part to the fact that somatic mutations target pathways, or groups of interacting genes, and that a mutation in any of dozens possible genes might be sufficient to perturb a pathway. While some of the cancer driver pathways are well characterized, many others are only approximately known.
I will describe algorithms that attempt to identify genes important for cancer by analyzing mutations in the context of interaction networks. These algorithms rigorously identify statistically significant combinations of interacting genes, where the significance is given by the frequency of mutation of the set of genes or its association with survival time.

Markov Clustering: Consolidation and Renewed Bearing
Srinivas Parthasarathy (Ohio State University)

Since its introduction in the late nineties, the idea of Markov Clustering, a graph clustering approach based on the principle of simulating stochastic flows (random walks) has seen wide use -- particularly in the area of bioinformatics. In this talk I will review this basic idea and then describe several enhancements to this approach that in turn improve the quality (via regularization, and the accommodation of overlapped clustering) and speed (via sparsification, and a multi-level mechanism) of such stochastic flow algorithms so that they can be deployed on large scale problems. Results on real world interaction networks demonstrate both the efficacy and efficiency of the approach. Time permitting I will discuss some preliminary efforts on leveraging these ideas in the setting of remote sensing and flood mapping for emergency response.
Joint work with Peter Jacobs, Albert Liang, Venu Satuluri, Duygu Ucar and Yu-Keng Shih.

Planning Problems for Sophisticated Agents with Present Bias
Manish Raghavan (UC Berkeley)

Present bias, the tendency to weigh costs and benefits incurred in the present too heavily, is one of the most widespread human behavioral biases. It has also been the subject of extensive study in the be- havioral economics literature. While the simplest models assume that decision-making agents are naive, reasoning about the future without taking their bias into account, there is considerable evidence that peo- ple often behave in ways that are sophisticated with respect to present bias, making plans based on the belief that they will be present-biased in the future. For example, committing to a course of action to re- duce future opportunities for procrastination or overconsumption are instances of sophisticated behavior in everyday life.
Models of sophisticated behavior have lacked an underlying formalism that allows one to reason over the full space of multi-step tasks that a sophisticated agent might face, and this has made it correspondingly difficult to make comparative or worst-case statements about the performance of sophisticated agents in arbitrary scenarios. In this paper, we incorporate the framework of sophistication into a graph-theoretic model that we used in recent work for modeling naive agents. This new synthesis of two formalisms — sophistication and graph-theoretic planning — uncovers a rich structure that wasn’t apparent in the earlier behavioral economics work on this problem, including a range of findings that shed new light on sophisticated behavior.
In particular, our graph-theoretic model makes two kinds of new results possible. First, we give tight worst-case bounds on the performance of sophisticated agents in arbitrary multi-step tasks relative to the optimal plan, along with worst-case bounds for related questions. Second, the flexibility of our formalism makes it possible to identify new phenomena about sophisticated agents that had not been seen in prior literature: these include a surprising non-monotonic property in the

use of rewards to motivate sophisticated agents; a sharp distinction in the performance of agents who overestimate versus underestimate their level of present bias; and a framework for reasoning about commitment devices that shows how certain classes of commitments can produce large gains for arbitrary tasks.

Message impartiality in social media
Mohammed Bilal Zafar (Max Planck Institute for Software Systems)

Discourse on social media platforms is often plagued by acute polarization, with different camps promoting different perspectives on the issue at hand—compare, for example, the differences in the liberal and conservative discourse on the U.S. immigration debate. A large body of research has studied this phenomenon by focusing on the affiliation of groups and individuals. In this talk, I will discuss studying polarization at a finer-grained perspective: impartiality of individual messages.
While the notion of message impartiality is quite intuitive, the lack of an objective definition and of a way to measure it directly has largely obstructed scientific examination. I will discuss our recent work on defining and operationalizing message impartiality objectively, and introducing a methodology for quantifying it automatically. Unlike a supervised machine learning approach, our method can be used in the context of emerging events where impartiality labels are not immediately available.
Lastly, I will present our study on the effects of (im)partiality on social media discussions at scale, which includes investigation into relative popularity of partial and impartial content and dynamics of message impartiality in highly polarized discussions.

A Network Model Characterized by a Latent Attribute Structure with Competition
Corrado Monti (University of Milan)

Networks often arise from a complex interweaving of some features of the nodes. For example, in a co-authorship network, a link stems more easily between authors with similar interests; similarly, in a genetic regulatory network, links are affected by the different biological functions of the regulators.
We defined and analysed a statistical model for such a network. Features evolve through a modified Indian buffet process: every new node introduces new features, while also copying the popular ones from previous nodes, in a rich-get-richer fashion; moreover, some nodes are supposed to be more effective than others in transmitting their features. Links are then formed as a consequence of the features they bear: we allow not only homophily and heterophily, but also more complex interactions between features.
Through this model, we provide estimators to gain information from real-world networks; finally, we show how the model displays the same global topological properties of a real-world social network.

TRIEST: Counting Local and Global Triangles in Fully-dynamic Streams with Fixed Memory Size
Alessandro Epasto (Google)

Modern graphs like the Web and online social networks are extremely large and inherently dynamic: edges and nodes are continuously inserted and deleted at very high rates. As a result, the network properties in these graphs are constantly evolving and keeping track of an high-quality approximation of these quantities is computationally challenging.

I will present our new method (named TRIEST) that allows to approximate efficiently the number of triangles in a fully-dynamic graph stream with adversarial edge insertions and deletions.
Our algorithms use sampling techniques to exploit a user-specified memory space at all times and provide concentration precise bounds on the quantities estimated. Our experimental results on billion edge streams show that our method outperforms state-of-the-art approaches in accuracy while exhibiting very fast update times.
Time permitting I will also introduce some recent results on preserving dense subgraphs in dynamic graph streams.
This is joint work with L. De Stefani, M. Riondato and E. Upfal. Paper to appear at KDD2016.

On Sampling Nodes in a Network
Silvio Lattanzi (Google)

Random walk is an important tool in many graph mining applications including estimating graph parameters, sampling portions of the graph, and extracting dense communities. In this paper we consider the problem of sampling nodes from a large graph according to a prescribed distribution by using random walk as the basic primitive. Our goal is to obtain algorithms that make a small number of queries to the graph but output a node that is sampled according to the prescribed distribution. Focusing on the uniform distribution case, we study the query complexity of three algorithms and show a near-tight bound expressed in terms of the parameters of the graph such as average degree and the mixing time. Both theoretically and empirically, we show that some algorithms are preferable in practice than the others. We also extend our study to the problem of sampling nodes according to some polynomial function of their degrees; this has implications for designing efficient algorithms for applications such as triangle counting.

Expanders via Local Edge Flips
Aditya Bhaskara (University of Utah)

Can making "local" structural changes to a graph result in global properties, such as expansion? In this talk, I will show that performing a small number of random "local flip" operations in a graph will yield an expander with high probability. This was first proposed by Mahlmann and Schindelhauer in the context of peer-to-peer networks. We give an upper bound of O(n^2) on the time required to reach an expander (w.h.p.), significantly improving the best known bounds.