Modern Statistical Machine Learning Methodologies for Understanding Social Media and Social Networks – A Compact Course

Eric Xing
Carnegie Mellon University

Across the sciences, a fundamental setting for representing and interpreting information about entities, the structure and organization of these entities, and their changes over time, is a stochastic network that is topologically rewiring and semantically evolving over time and space. While there is a rich literature in modeling and mining various macroscopic properties of networks, such centrality, connectivity, diameter, conductance, motif spectrum, clustering coefficient, etc., which have led to many important insights of network characteristics such as degree-freeness and small-world, and practical algorithms for network computation such as partition, flow tracking, and path planning, relatively little effort is seen toward enabling microscopic and mesoscopic analysis and operation on network that allows one to gain actionable insights and make prediction on intention, interest, risk, and influence for every individual node or edge in a network. Worse still, even less progress are made on furthering the above problems toward more challenging but realistic scenarios where the networks are rewiring over time, and/or the networks of interest are themselves not even observable (e.g., imagine a covert and evolving drug trafficking network).

In this tutorial, I will present our recent work in addressing the above problems, which help to bridge latest developments in statistical machine learning theory and algorithms for network and multimodal analysis, to the practical needs from industry and other consumers for tools supporting tasks ranging from online nodal/edge level analytics and operations, to overall hidden dynamic network discovery. Specifically, I will introduce new sparse-coding algorithms for estimating the topology of latent evolving networks underlying nonstationary time-series of nodal activities, along with theoretical results on the asymptotic sparsistency of the proposed methods; I will then present a family of new Bayesian model for estimating and visualizing the trajectories of latent multi-functionality of every node, and for learning community structures, in the evolving networks; and finally I will discuss how to combine textual/image information with social networks for user and community modeling in a way that goes beyond simply graph-driven, and how to address computational challenges in massive networks of societal scale.