Computer Science Seminar

Relational Learning Methods and Complex Networks with Applications to Health Informatics

Pablo Robles-Granda, PhD
University of Notre Dame

Abstract: Statistical relational models are widely used in network science to reason about the properties of complex systems where the nodes represent entities (e.g., individuals, disease vectors, patients etc.) and the links represent relationships (e.g., friendship, contagion paths, bio- and social-markers, etc.). A type of relational models is generative network models that aim to capture the observed characteristics of real world networks to reproduce them via sampling. To acquire a better understanding of the underlying properties of the system (e.g., a social network), many challenges must be overcome. An example of these challenges is that a good descriptive model of networks not always facilitate sampling. In order to overcome this challenge, it is important to develop relational models that facilitate sampling methods for networks with and without vertex-attributes. However, this task remains a challenging problem because most current methods work with relatively simple or specific models. Another challenge is to understand the dynamics that appear in a complex system and how these dynamics evolve and what are the dependencies across components of the system that drive these dynamics. In my talk, I will discuss some developments for these two types of problems starting with an application to health informatics. First, nowadays, the popularity of machine learning has produced a significant body of research on social and medical sciences. In my work, I have developed mathematical descriptions of health and wellness of the individuals. I investigated various techniques of network science to extract the graph-theoretic bio-markers. Also, I applied machine learning to predict stress and wellness variables. With my colleagues, we showed how to use these additional biomarkers to improve the accuracy of prediction of wellness of individuals. I will present a few other examples where I apply some methodological insights acquired from relational learning. Second, I will present observations about a number of recent models that share a common structure. This structure allows to increase both the variance of the models and the space of characteristics possible to be modeled. Building on these results I will discuss general strategies to combine models and density functions of the node attributes to sample attributed-networks. Since in this structure the probability mass is allocated to certain regions of the network space it is hard to identify candidate networks to sample attributed networks. I will present constrained sampling to bias the search of edge-candidates using the topological space with higher likelihood. Finally, identifying the laws that govern a complex system based on its network representation is an open problem. In my work, I identified strategies to represent changes in a dynamical system using its footprint data. The persistent challenge is that most approaches for incident detection in complex non-linear systems use nominal systems that rely on domain expertise. With my colleagues, we have studied the problem of incident detection in non-linear systems using the trajectory of the causal relationships described only by the time-series data.

Friday, January 24, 2020, 10:30 am
Mathematics and Science Center: MSC W201

Computer Science
Emory University