MANAGEMENT SCIENCES SEMINAR SERIES

The ground truth about metadata and community detection in networks

Prof. Aaron Clauset
Department of Computer Science, University of Colorado, Boulder

November 4, 2016
10:30-11:30 am

S401 Pappajohn Business Building

Abstract

In this talk, I'll describe two exciting new results for community detection in networks. Community detection is the common step of decomposing a network into its underlying structural modules or groups, and is similar to the task of seeking clusters in vector data. Over the past 15 years, thousands of papers have been published on community detection, describing hundreds of different methods. The first result I will present is a general theoretical statement about the feasibility of recovering "ground truth" communities from a network, in which we prove a No Free Lunch theorem for community detection. These results imply that there can be no universal algorithm that always recovers the ground truth, and no algorithm can perform better than all others on all problems. The second result is a new community detection algorithm that exploits node "metadata", such as the age or gender of individuals in a social network, geographic location of nodes in the Internet, or cellular function of nodes in a gene regulatory network, to find more scientifically useful communities. Crucially, this method does not assume that the metadata are correlated with the communities we are trying to find and instead learns whether a correlation exists and then uses or ignores the metadata depending on whether they contain useful information. The learned correlations are also of interest in their own right, allowing us to make predictions about the community membership of nodes whose network connections are unknown. After sketching the method, I'll demonstrate the model on synthetic networks with known structure, where the method performs better than any algorithm can without metadata, and on real-world networks, large and small, drawn from social, biological, and technological domains. This is joint work with Leto Peel, Daniel B. Larremore, and Mark Newman.

Prof. Clauset's Bio

Aaron Clauset is an Assistant Professor in the Department of Computer Science and the BioFrontiers Institute at the University of Colorado Boulder, and is External Faculty at the Santa Fe Institute. He received a PhD in Computer Science, with distinction, from the University of New Mexico, a BS in Physics, with honors, from Haverford College, and was an Omidyar Fellow at the prestigious Santa Fe Institute. Clauset is an internationally recognized expert on network science, data science, and machine learning for complex systems. His work has appeared in many prestigious scientific venues, including Nature, Science, PNAS, JACM, WWW, ICWSM, STOC, SIAM Review, and Physical Review Letters, and has been covered in the popular press by the Wall Street Journal, The Economist, Discover Magazine, New Scientist, Wired, Miller-McCune, the Boston Globe and The Guardian.