Analysis of Large-Scale Networks  
ICPSR Network Analysis Workshop 2012

Synopsis: This course is an introduction to computational network analysis. It is based on learning how to write scripts, or short snippets of code, using Python and NetworkX. The approach is extremely versatile, enabling one to analyze structural properties of networks, generate networks from microscopic rules, simulate dynamical processes on networks, and much more. The approach is also scalable, meaning that the methods can be applied to systems of various sizes, from small networks with a few actors to massive networks consisting of millions of nodes.

Instructor: Jukka-Pekka “JP” Onnela, PhD, Assistant Professor of Biostatistics, Harvard University  
Teaching Assistant: Ravi Goyal, Department of Biostatistics, Harvard University  
Location and Time: Ann Arbor, Michigan, June 18-22, 2012

Approach: Participants will learn the necessary scripting skills that will enable them to build their own tools for carrying out network analysis. We will first cover the basics of the Python programming language and then proceed to NetworkX, the Python network analysis module. While this is not a programming course, Python & NetworkX is a powerful combination that enables the participants to perform customized and scalable network analysis and modeling. One cannot overstate how much more powerful and flexible this approach is compared to dealing with network analysis software. The skills taught in the course will be immediately applicable to a diverse set of network problems.

Prerequisites:
- Basic familiarity with network concepts (we will provide a brief review if necessary)
- Some prior programming experience in any language (ideally)

Syllabus (core): We will start by learning the basics of Python and NetworkX. Python is a fun and addictive language, which enables one to go very quickly from ideas to execution. While it does not have the speed of some of the compiled languages, particularly Fortran, C, or C++, it is still sufficiently powerful for most problems, and it is significantly easier to learn and maintain than those languages. NetworkX is a Python package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks. It makes handling networks, even relatively large networks, very easy.

**Python topics**: Numbers; Strings; Object Types; Lists; Dictionaries; Tuples; Sets; Expression Operators; Dynamic Typing; Statements; Files; Functions; Scope Rules; Modules; Extras

**NetworkX topics**: Nodes and Edges; Node Degree and Neighbors; Basic Network Properties; File Operations on NetworkX; Generating Random Graphs; Graph and Node Attributes; Edge Attributes; Visualization using NetworkX; Exercises

We will then proceed to “case studies.” For example, we will look at network sampling, and examine diffusion processes and/or epidemic spreading models on networks.

Syllabus (optional): Depending on the participants, we can additionally explore the impact of structural perturbations on network properties (e.g., node and/or tie removal), learn about network community detection, or generate networks from microscopic mechanisms. Alternatively, the participants can spend the last day of the workshop working on their own research problems in the area of network analysis, and apply their newly acquired skills while benefitting from the instructors’ help.

What to expect: After the workshop, and a little practice, the participants will be able analyze structural properties of networks, and simulate some simple dynamical processes on networks.