Tutorial: Large Scale Network Analytics with SNAP

http://snap.stanford.edu/proj/snap-www

Rok Sosič, Jure Leskovec
Stanford University

WWW-15, Florence, Italy

May, 2015
What is SNAP?

- **Stanford Network Analysis Platform (SNAP)** is a general purpose, high-performance system for analysis and manipulation of large networks
  - [http://snap.stanford.edu](http://snap.stanford.edu)
  - Scales to massive networks with hundreds of millions of nodes and billions of edges

- **SNAP software**
  - Snap.py for Python, SNAP C++

- **SNAP datasets**
  - Over 70 network datasets
SNAP Tutorial: Content

- Motivation
- Introduction to SNAP
- Snap.py for Python
- Network analytics
- SNAP network datasets
- SNAP for C++
- Hands-on exercise

Contact information:
Rok Sosič, rok@cs.stanford.edu

Slides available at:
http://snap.stanford.edu/proj/snap-www
Why Networks?

Networks are a general language for describing complex systems.
Friends & Family
Media & Information
World economy
Roads
Human cell
Networks!
SNAP Resources

Rok Sosič, Jure Leskovec
Stanford University

WWW-15, Florence, Italy  May, 2015
Snap.py Resources

- **Prebuilt packages** available for Mac OS X, Windows, Linux
  http://snap.stanford.edu/snappy/index.html

- **Snap.py documentation:**

- **SNAP user mailing list**
  http://groups.google.com/group/snap-discuss

- **Developer resources**
  - Software available as open source under BSD license
  - GitHub repository
  https://github.com/snap-stanford/snap-python
SNAP C++ Resources

- **Source code** available for Mac OS X, Windows, Linux
  http://snap.stanford.edu/snap/download.html

- **SNAP documentation**
  http://snap.stanford.edu/snap/doc.html
  - Quick Introduction, User Reference Manual
  - Source code, see tutorials

- **SNAP user mailing list**
  http://groups.google.com/group/snap-discuss

- **Developer resources**
  - Software available as open source under BSD license
  - GitHub repository
    https://github.com/snap-stanford/snap
  - SNAP C++ Programming Guide
Collection of over 70 web and social network datasets:
http://snap.stanford.edu/data

Mailing list: http://groups.google.com/group/snap-datasets

- **Social networks**: online social networks, edges represent interactions between people
- **Twitter and Memetracker**: Memetracker phrases, links and 467 million Tweets
- **Citation networks**: nodes represent papers, edges represent citations
- **Collaboration networks**: nodes represent scientists, edges represent collaborations (co-authoring a paper)
- **Amazon networks**: nodes represent products and edges link commonly co-purchased products