

# Analysis of Large Scale Social Networks

## Exercise Session 1: Basic Network Analysis & Visualization

### Objectives

The objective of the first exercise session is to familiarize students with graph processing software. After the session they should be able to use a graph processing package for these tasks:

- Constructing a network with nodes and edges
- Read network data from file into the software
- Calculate centrality measures and other basic network properties
- Extract subgraphs
- Obtain the shortest path between two nodes
- Visualize the network

### Software

Network analysis:

- Gephi: <https://gephi.org/>
- iGraph in Python: <http://igraph.org/python/>
- Pajek: <http://mrvar.fdv.uni-lj.si/pajek/>
- SNAP: <http://snap.stanford.edu/index.html>

Python:

- Anaconda: <https://www.anaconda.com/distribution/>

On Mac, make sure you have the latest version of these packages:

```
homebrew (/bin/bash -c "$(curl -fsSL
https://raw.githubusercontent.com/Homebrew/install/master/install.sh)")
pkg-config (brew install pkg-config)
cairo (brew install cairo)
```

### Data Sets

- Toledo: USAir97.net (Pajek format)
- Newman: <http://www-personal.umich.edu/~mejn/netdata/>
- SNAP: <http://snap.stanford.edu/data/index.html>

### Exercise 1: Guided Exercise

Reading data files

*Task: Import data file into software package.*

Calculate, plot and analyse centrality measures and network properties

*Task: Calculate the following centrality measures: Degree, Betweenness, average path length, Closeness, PageRank and clustering coefficient. Calculate and plot the degree and betweenness distribution. Calculate the correlation between degree, betweenness and page rank.*

Graph Processing

*Task: Create a second network with only those nodes with a degree at least the average degree in the network*

## Visualize network

*Task: Create a network visualization with the Fruchterman-Reingold layout. Export the plot to a common format like JPEG or PNG.*

## Shortest Path

*Task: calculate and visualize the shortest path between 'Sheppard Afb/Wichita Falls Mun' and 'West Tinian'.*

## Export

*Task: Export the networks and visualizations.*

## Exercise 2: *Real Data Set*

- *Choose a data set with a corresponding research paper (eg. one of the sets provided by Newman such as [Coauthorships in network science](#)):.*
- *Try to define a few research questions that you could solve with this dataset without neglecting validity and reliability. You can make a small notes in your notebook describing your data set and your findings*
- *Post your notebook with code and findings [here](#).*

Your analysis should include these steps:

- *Import dataset in package of your choice*
- *Calculate number of nodes, number of edges, density*
- *Calculate degree and plot degree distribution*
- *Describe degree distribution properties*
- *Calculate betweenness if appropriate*
- *Identify the nodes with the highest degree and betweenness*
- *Plot the network and identify the nodes with the highest centrality*
- *Remove the nodes with the highest centrality and rerun the plot. What has happened? How many nodes do you have to remove before the network structure starts to collapse?*
-