# Analysis of Large Scale Social Networks

# Exercise Session 1: Graph Processing & Visualization

## Software

* Gephi: <https://gephi.org/>
* iGraph in R: <http://igraph.org/r/>

## Data Sets

* Toledo: USAir97.net (Pajek format)
* Newman: <http://www-personal.umich.edu/~mejn/netdata/>

## Reading data files

*Task: .Import data file into software package.*

## Calculate and plot centrality measures and network properties

*Task: Calculate the following centrality measures for the US-Air network in one of the packages:*

* *Degree,*
* *Betweenness,*
* *average path length,*
* *Closeness,*
* *PageRank*
* *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*

* iGraph  
  Combine the functions *induced\_subgraph* and *degree*
* Gephi  
  Use the Filters Pane

## Shortest Path

*Task: Select to random pair of nodes from your networks and calculate and   
visualize the shortest path*

* Gephi  
  Select the airplane button in the Overview Pane. Then select the source and the target node.
* iGraph  
  *get.shortest.path(graph, source\_node, target\_node)*For visualization see: <https://rpubs.com/kateto/netviz>

## Visualize network

*Task: Create a Fruchterman-Reingold of the US-Air network. Export the plot to a common format like jpg, png, bmp.*

### Fruchterman-Reingold

* Gephi: Choose Fruchterman-Reingold in Layout
* iGraph:  
   *plot(graph, layout=layout\_with\_fr)*

### Other layouts

* Gephi: try *Force-Atlas 2*
* iGraph: try DrL  
   *plot(graph, layout=layout\_with\_drl)*

### Export

* Gephi  
  Export as png/pdf or jpg from the Preview Pane after   
  refreshing the layout
* iGraph  
  In R-Studio: use the export button in the Plot window.

## Real dataset

*Task: Choose one of the datasets provided by Newman.*

*Try to define a few research questions that you could solve with this dataset without neglecting validity and reliability. You can make a small text document describing your data set and your findings.*

*Compare your results with the results presented in the original paper. Discuss these results with your neighbour.*

* *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 node with the highest degree*
* *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?*