# Analysis of Large Scale Social Networks

# Exercise Session 2: Graph Clustering

## Software

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

## Data Sets

* Newman: <http://www-personal.umich.edu/~mejn/netdata/>

[Condensed matter collaborations 2003](http://www-personal.umich.edu/%7Emejn/netdata/cond-mat-2003.zip) (available on Toledo)

## Objectives

* Apply different cluster or community detection algorithms to a network
* Compare results

## Pre-processing

*Tasks*

1. *Create an R-script file and log all the commands.*
2. *Read data in gml-format in R – iGraph*
3. *Check if the network is undirected/directed and weighted/unweighted*
4. *Calculate degree and plot degree distribution*
5. *Check if the network is connected*
6. *Identify the largest connected component*
7. *Retain only this largest component*
8. *Calculate degree and plot degree distribution again.*
9. *Save new network in gml and pajek format*
10. *Save the label list (author names)*

## Clustering & Visualization

*Tasks:*

1. *Run three community detection algorithms from the igraph packages on both networks (original data set and the largest connected component)  
   cluster\_leading\_eigen; cluster\_louvain; cluster\_walktrap*
2. *Calculate modularity scores for each cluster solution*
3. *Try to visualize (a sample of) at least one network with nodes in different colour according to their cluster membership*
4. *Compare the obtained cluster solutions*
5. *In addition to the igraph community detection implementations you can try to use Gephi for the Louvain method. Compare the results.*