



# Projects on (Social) Network Analysis

F.W. Takes

Leiden University, The Netherlands

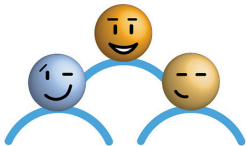
Bachelorklas — December 9, 2015

# Projects

- Topics: **social network analysis**, **network science**, graph mining, data mining, data science, algorithms
- Data: Twitter, Wikipedia, Hyves, ...
- Quantitative studies (so, with concrete algorithms, large real-world datasets and experiments)
- Project work in for example C++ or Python
- Thesis in  $\text{\LaTeX}$
- Relevant courses: algorithms, data structures, data mining
- Interested? Contact me at `f.w.takes@liacs.leidenuniv.nl` or walk by Snellius room 157b

# Social network analysis

- Example: HYVES online social network
  - $n = 8,000,000$  users (vertices)
  - $m = 1,000,000,000$  friendships (edges)
- Assume 4 bytes per int (integer), **storage** is challenging:
  - Adjacency Matrix:  $n^2 = 8,000,000^2 = 64 \cdot 10^{12}$  bits =  $\sim 8$ TB
  - Adjacency List:  $n + m = 1,008,000,000$  ints =  $\sim 4$ GB
- **Computation**: one Breadth First Search takes 6 seconds, what if we need to do one for every node?
- **Data mining**: how can we do clustering on 8 million items?



# Project 1: Comparing Graph Frameworks

- Many social network analysis toolkits
- Not all equally suitable for analyzing “big” network data
- Compare toolkits: SNAP, NetworkX, Graph-Tool, ...
- Data: up to 200 social network datasets of different sizes
- Literature review of suitable comparison measures
- Comparison based on time and memory usage for certain critical operations (distance computation, component analysis, etc.), ...
- Interested? `f.w.takes@liacs.leidenuniv.nl` or find me in 157b







Figure : Board interlock network (200 000 nodes and over 1 000 000 edges)

## Project 2: Interactive web platform

- Modern responsive web platform
  - **Visualization** of (social) networks using for example `sigma.js`
  - Upload network and do online **computation**
- Service-oriented architecture with specialized hardware (16-core, 1.5TB RAM, 12TB SSD)
- Challenges in usability, scalability and performance
- 2 positions available
- Interested? `f.w.takes@liacs.leidenuniv.nl` or find me in 157b



## Titles of previous projects

- Combining Customer Attributes and Social Network Mining for Prepaid Mobile Churn Prediction
- Mining Population Movement Patterns from Cellphone Data During Natural Disasters
- Exploratory Recommendations using Wikipedia's Linking Structure
- Identifying Prominent Actors in Social Networks
- Extracting Information from an Energy Expenditure Dataset
- Topical Influence on Twitter: A Feature Construction Approach
- An Evaluation Method for Nodes in Multiple Dynamic Networks
- Corporate Network Analysis Based on Centrality Methods

## Other projects

- Interested? Questions? Ideas of your own?  
`f.w.takes@liacs.leidenuniv.nl` or find me in 157b