Workshop 2: Introduction to network analysis with Python

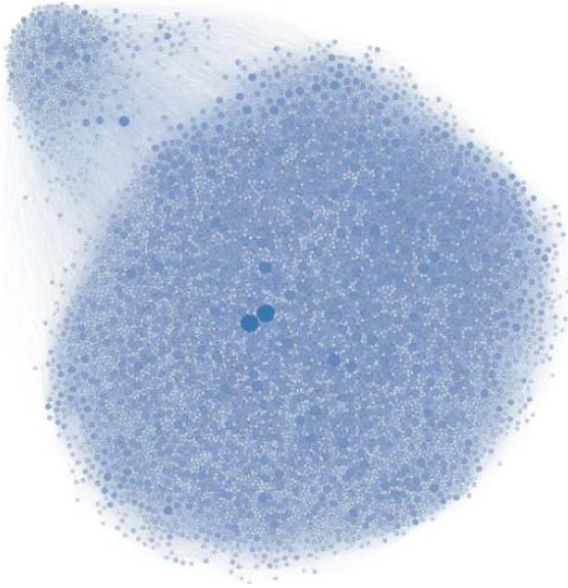Dr. Haiko Lietz[1] & Dr. N. Gizem Bacaksizlar Turbic[1,2]

[1]*GESIS – Leibniz Institute for the Social Sciences*, Cologne, Germany

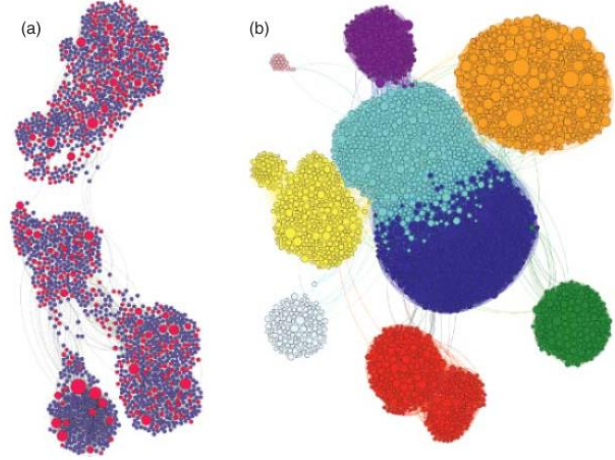[2]*RWTH Aachen University*, Aachen, Germany

Introduction to Computational Social Science methods with Python

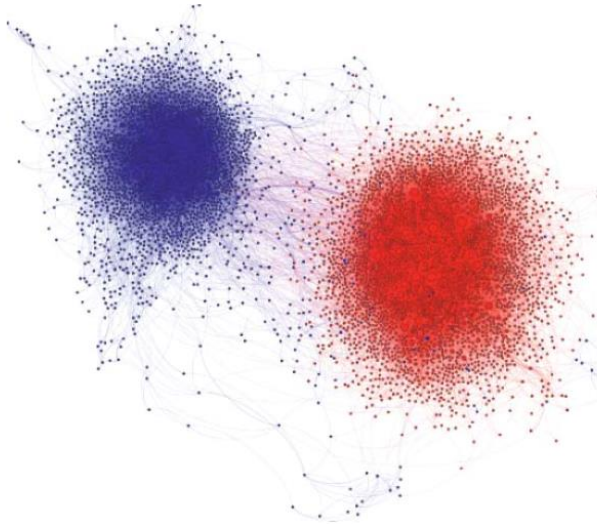Workshop series, Koç University, 11 April 2023

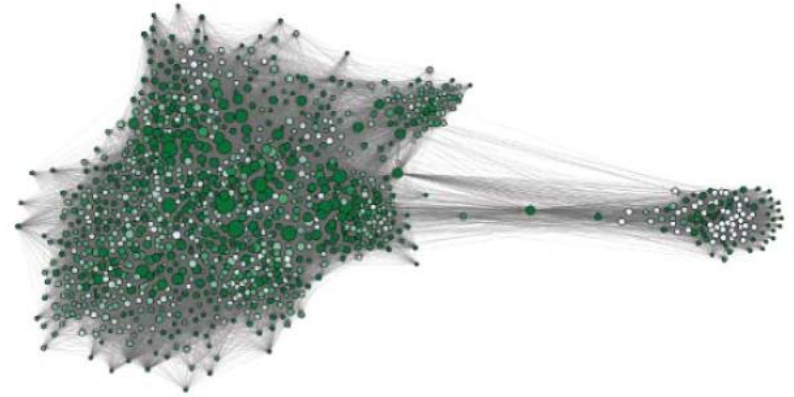# Social networks

Facebook users at Northwestern University

(a) Movie star (b) Movie co-star networks

[1] Menczer, F., Fortunato, S., & Davis, C. A. (2020). *A First Course in Network Science*. Cambridge University Press. 10.1017/9781108653947.
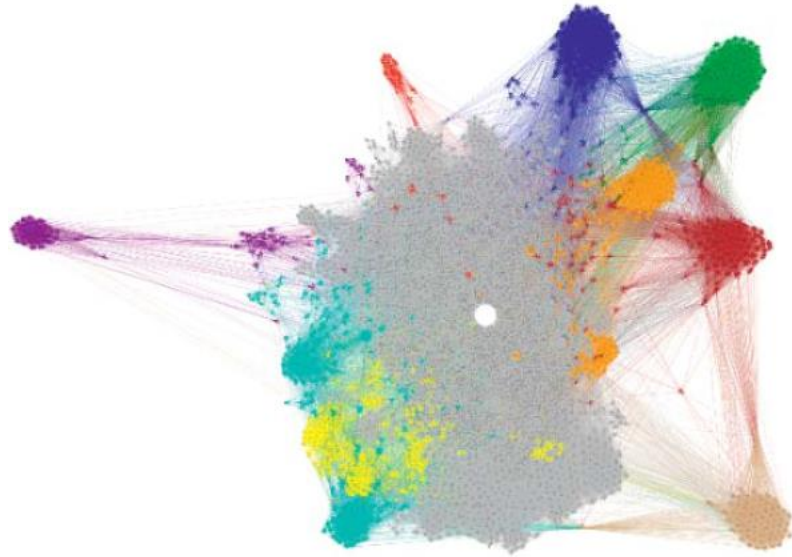
Communication networks

Retweet networks on US politics
(2010 US midterm elections) #tcot and #p2

Emails generated by Enron energy company
during its investigation of its collapse

[1] Menczer, F., Fortunato, S., & Davis, C. A. (2020). *A First Course in Network Science*. Cambridge University Press. 10.1017/9781108653947.
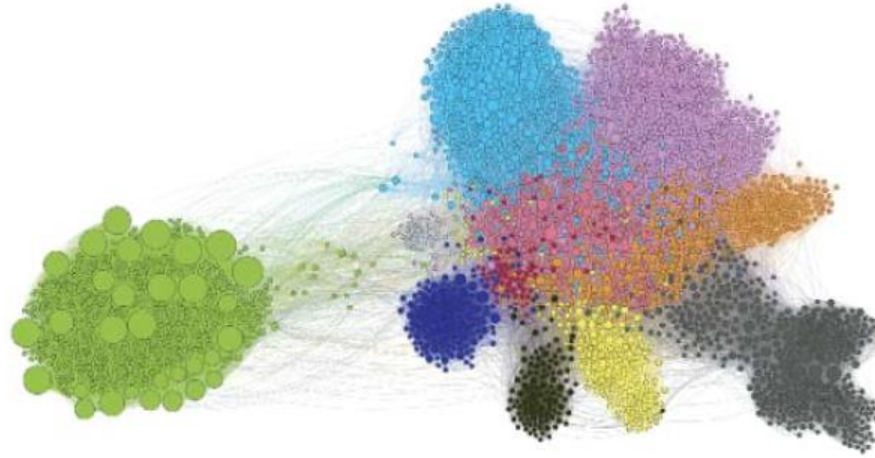
The web and Wikipedia

Wikipedia information network on Math articles
Node size by importance and colors are communities discussed in text

[1] Menczer, F., Fortunato, S., & Davis, C. A. (2020). *A First Course in Network Science*. Cambridge University Press. 10.1017/9781108653947.

4

The internet

Internet router network snapshot generated
by the Center for Applied Internet Data analysis (colors by community detection)

[1] Menczer, F., Fortunato, S., & Davis, C. A. (2020). *A First Course in Network Science*. Cambridge University Press. 10.1017/9781108653947.

Transportation network
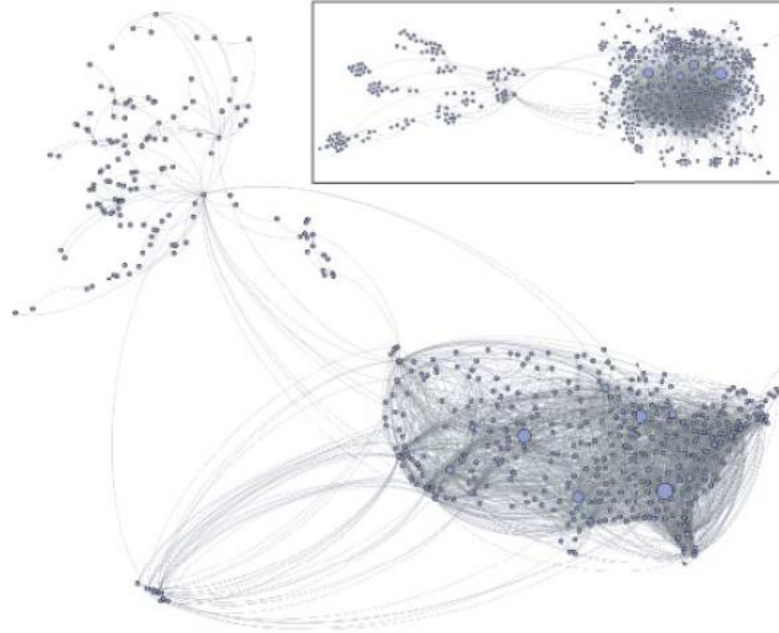
US air transportation network form flight data OpenFlights.org

[1] Menczer, F., Fortunato, S., & Davis, C. A. (2020). *A First Course in Network Science*. Cambridge University Press. 10.1017/9781108653947.

Variety of networks

- **Nodes** (the fundamental units) can be anything from real persons to Facebook or Twitter users, Wikipedia articles, or cities.
- **Edges** (the links connecting two nodes) can be anything from friendships to co-starring in a movie, retweeting, sending eMail, hyperlinks, or transportation.

Networks as descriptions and functions
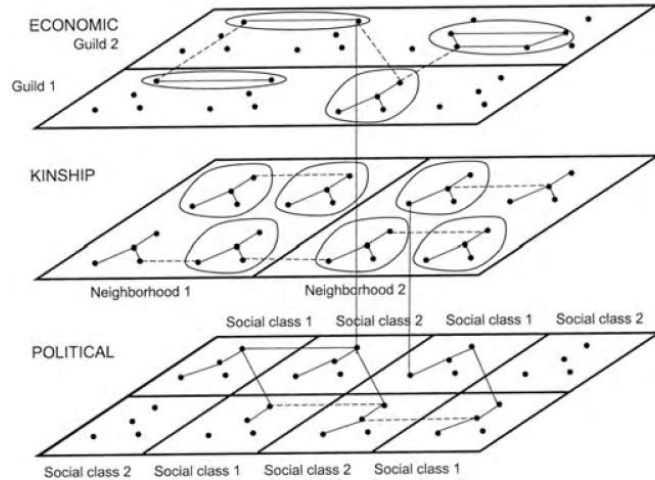
Figure 1.1: Netdom switching is not just for small-scale and informal settings, but is part of business and power life as seen in Padgett and McLean (2006). Concepts on figure are from this source, not from *Identity and Control*. Solid lines are "constitutive ties," dotted lines "relational social exchanges," and oblongs formal organizations. Dots are individuals.

- Social formations cannot only be **described** in terms of networks.
- They actually **function** as networks.

[2] White, H. C. (2008). *Identity and Control: How Social Formations Emerge*. Princeton University Press.

From a richness of data to a richness of networks
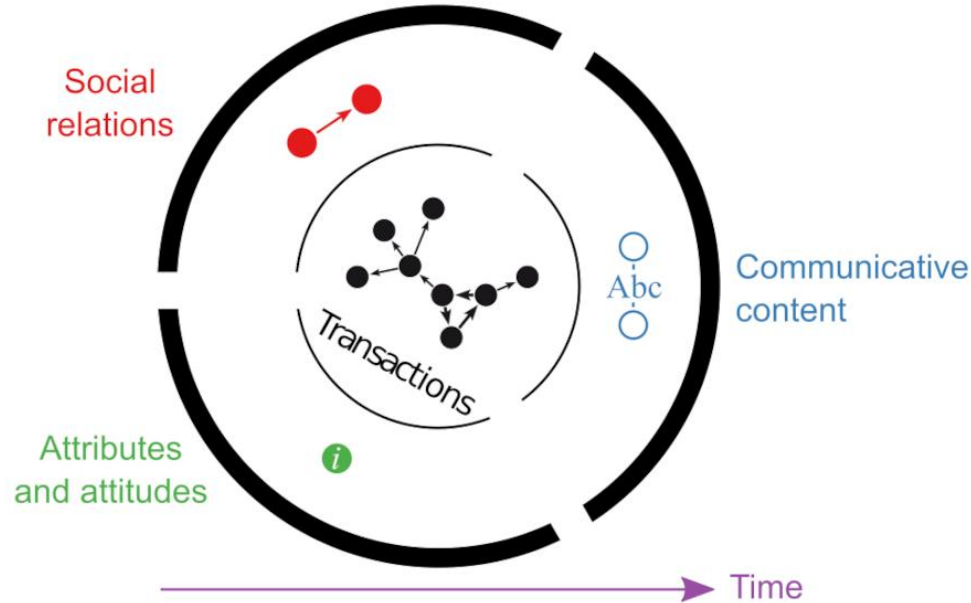
[3] Lietz, H., Schmitz, A., & Schaible, J. (2021). "Social Network Analysis with Digital Behavioral Data". *easy_social_sciences* 66:41–48. 10.15464/easy.2021.005.

Data used in this workshop: Copenhagen Networks Study
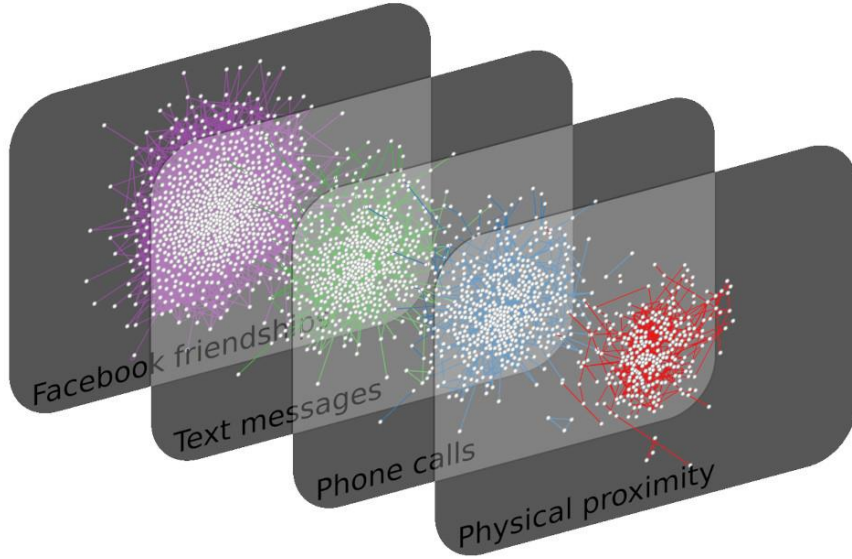
- Nodes: more than 700 freshmen at the Technical University of Denmark
- Edges:
  - Facebook friendship
  - Text messages
  - Phone calls
  - Physical proximity
- Time: one month in 2013

[4] Sapiezynski, P., Stopczynski, A., Lassen, D. D., & Lehmann, S. (2019). "Interaction data from the Copenhagen Networks Study". *Scientific Data* 6:315 (2019). 10.1038/s41597-019-0325-x.

Graph theory to handle the richness of data

Simple graphs (directed or undirected)	Attributed graphs (e.g., node categories, edge weights)
Multilayer graphs	Multimodal graphs

Graph theory and keeping your data tidy

$$G = (V, E)$$

Set of nodes V stored in **nodelist**

	user	sex
840	845	0
841	846	0
842	847	1
843	848	2
844	850	2

Set of edges E stored in **edgelist**

	timestamp	sender	recipient
0	18	370	512
1	37	512	370
2	126	370	512
3	134	0	512
4	312	137	136

Network analysis

Micro level	Meso level	Macro level
Centrality analysis	Community detection	Connectivity analysis
Social capital analysis	Blockmodeling	Inequality analysis
		Network modeling

Python packages for network preprocessing and analysis

	Advantages	Disadvantages
NetworkX	Huge set of methods implemented Very large user base Easy to understand (written in pure Python)	Slow for large networks
igraph	Fast for large networks Many methods implemented	
graph-tool	Fast for large networks Stochastic Blockmodeling Excellent network drawing	Hard to understand and master Not many methods implemented

Recommended textbooks and NetworkX resources

- McLevey, J. (2022). Doing Computational Social Science: A Practical Introduction. SAGE. https://us.sagepub.com/en-us/nam/doing-computational-social-science/book266031

- Menczer, F., Fortunato, S., & Davis, C. A. (2020). A First Course in Network Science. Cambridge University Press. https://doi.org/10.1017/9781108653947

- Platt, E. L. (2019). Network Science with Python and NetworkX Quick Start Guide. Packt. https://www.packtpub.com/product/network-science-with-python-and-networkx-quick-start-guide/9781789955316

- Artime, O., Benigni, B., Bertagnolli, G., d'Andrea, V., Gallotti, R., Ghavasieh, A., Raimondo, S., & De Domenico, M. (2022). Multilayer Network Science. Cambridge University Press. https://doi.org/10.1017/9781009085809

Workshop schedule

10 min | Introduction

80 min | Part I: Network construction

10 min | Break

80 min | Part II: Network analysis

Feel invited to contact us!

Haiko Lietz <haiko.lietz@gesis.org>

N. Gizem Bacaksizlar Turbic <gizem.bacaksizlarturbic@gesis.org>