gesis Leibniz Institute for the Social Sciences

39. Methodenseminar: Big Data Module II

Introduction to Social Network Science with Python

Introduction and Micro-Scale Analysis

Fariba Karimi, Haiko Lietz, & Marcos Oliveira

July 15, 2019

Leibniz Association

Introduction to Social Network Science

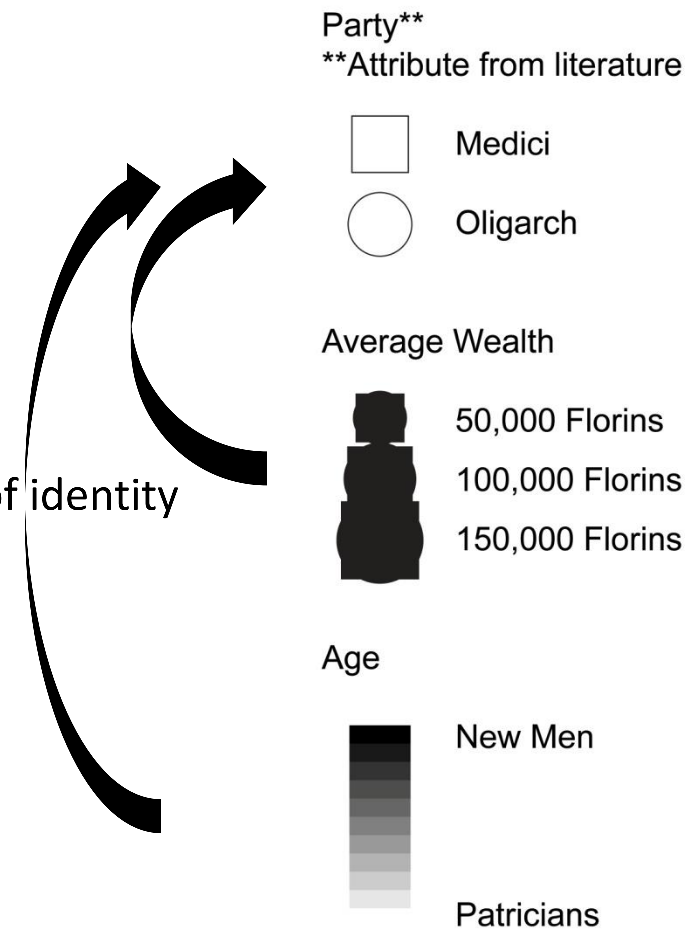
gesis Leibniz Institute for the Social Sciences

The Explanatory Power of Relations

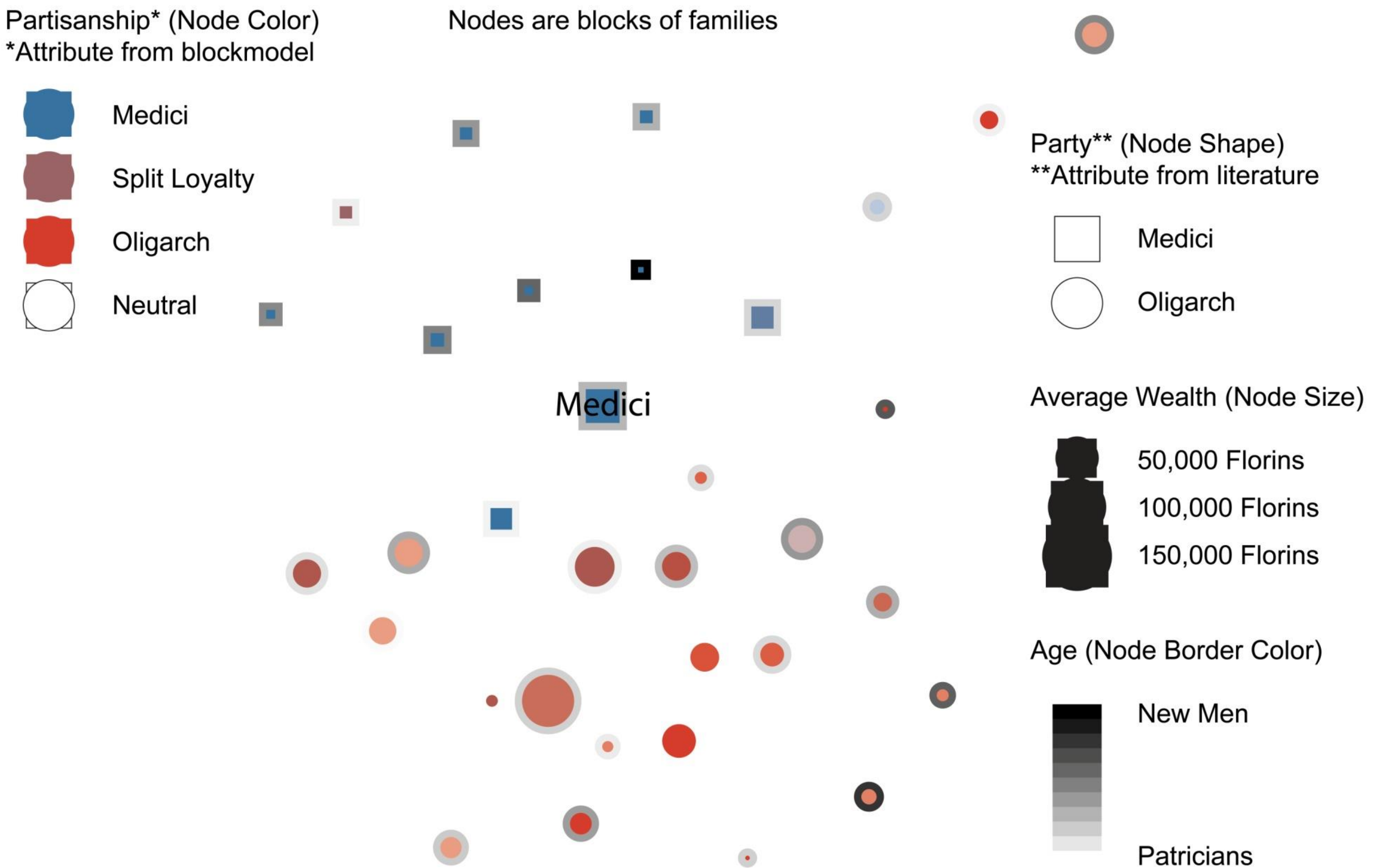
Cosimo de Medici (1389-1464)

Rinaldo degli Albizzi (1370-1442)

No explanation of identity



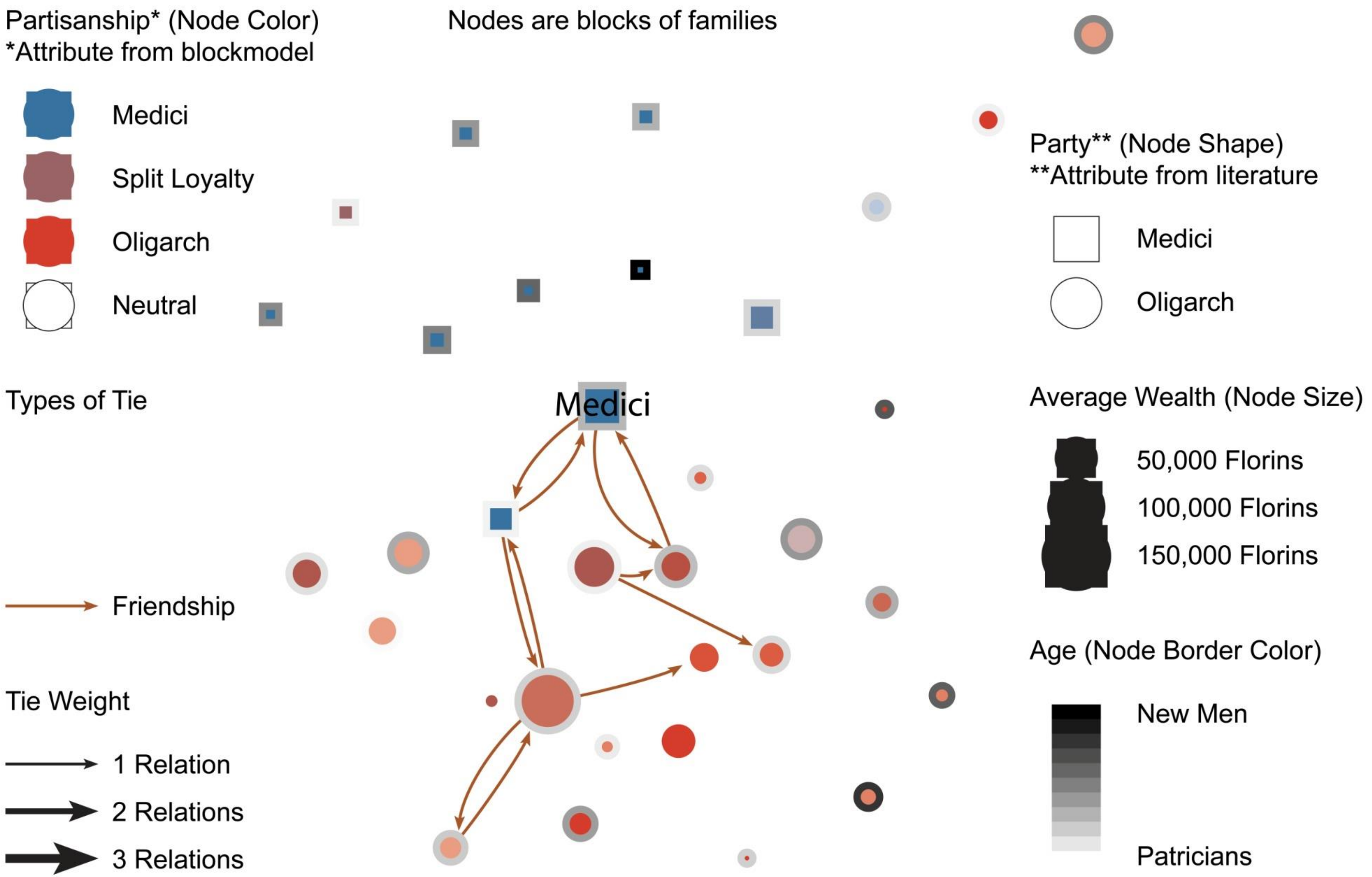
The Explanatory Power of Relations



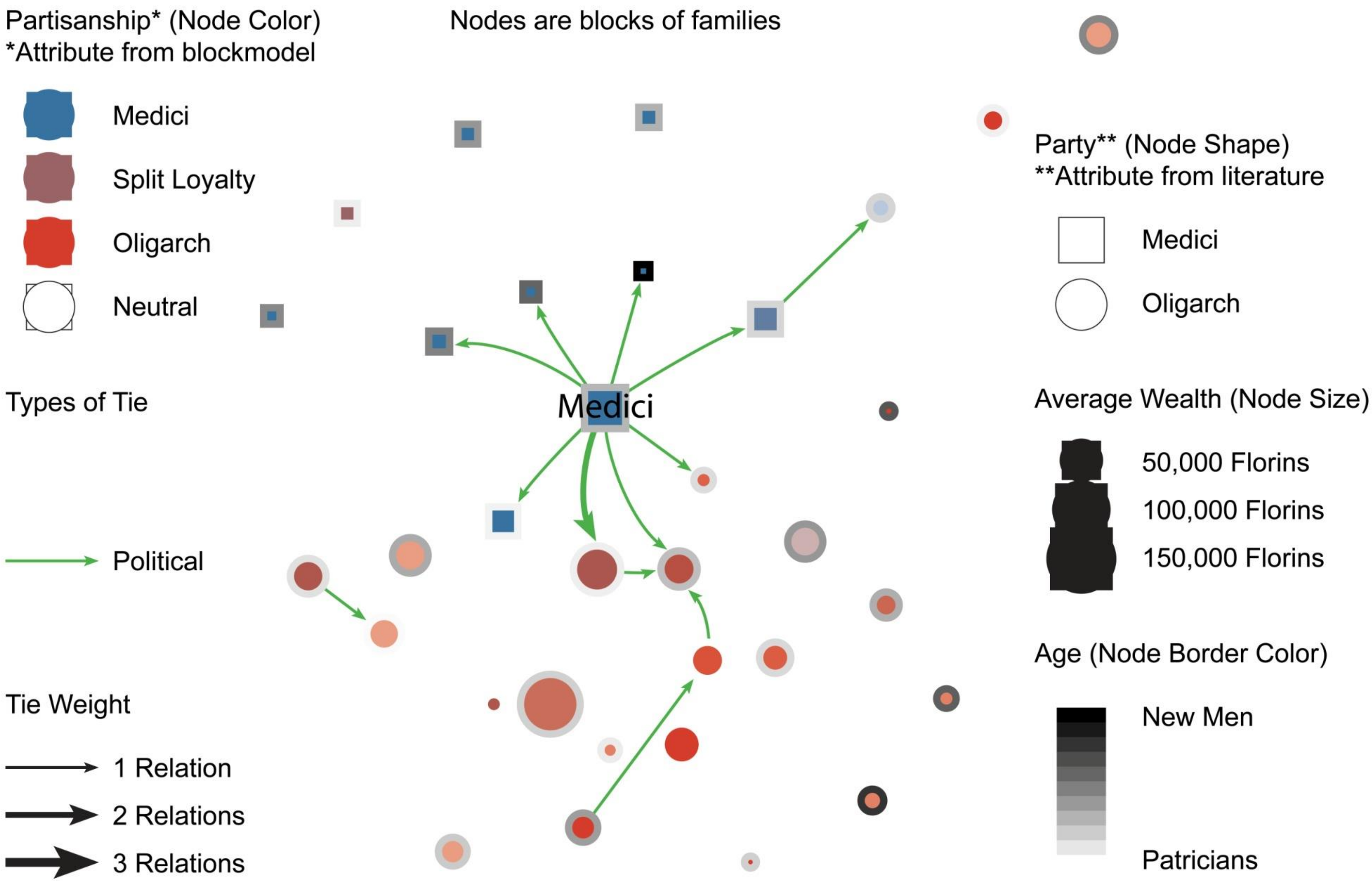
[1] Padgett, J.F. & Ansell, C.K. 1993. *American Journal of Sociology* 98: 1259–1319.

gesis Leibniz Institute for the Social Sciences

The Explanatory Power of Relations

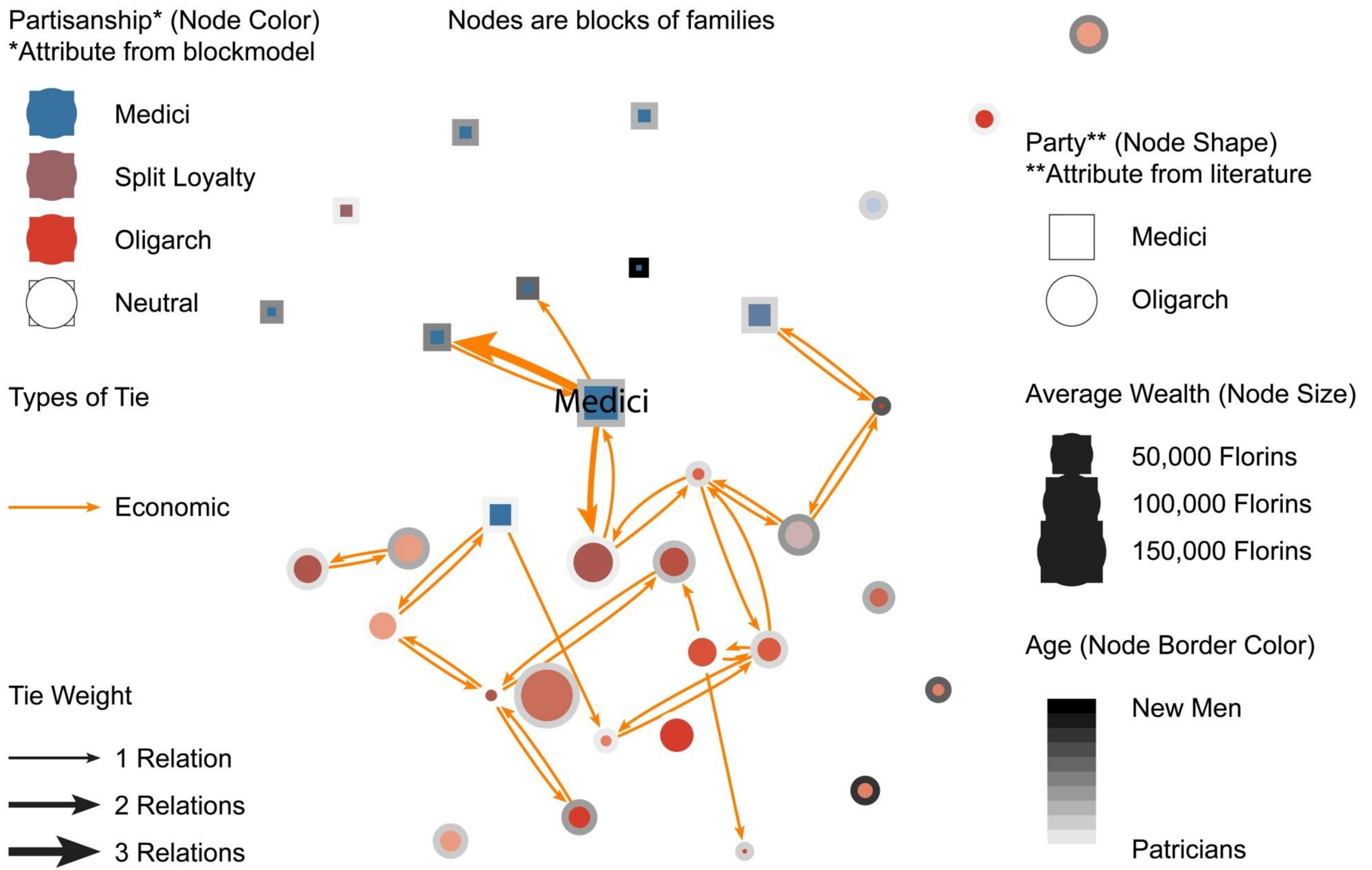


[1] Padgett, J.F. & Ansell, C.K. 1993. *American Journal of Sociology* 98: 1259–1319.



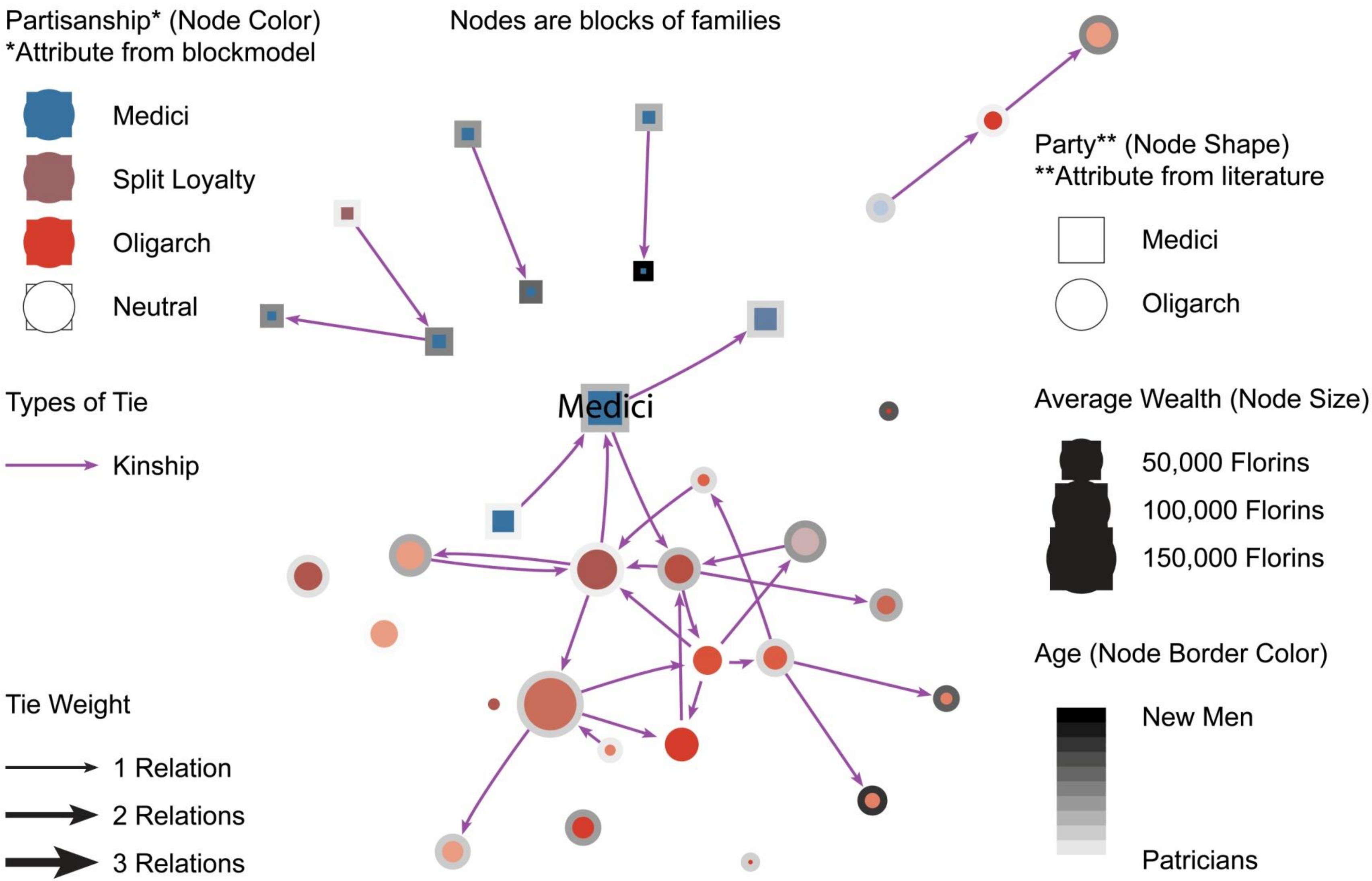
[1] Padgett, J.F. & Ansell, C.K. 1993. *American Journal of Sociology* 98: 1259–1319.

The Explanatory Power of Relations



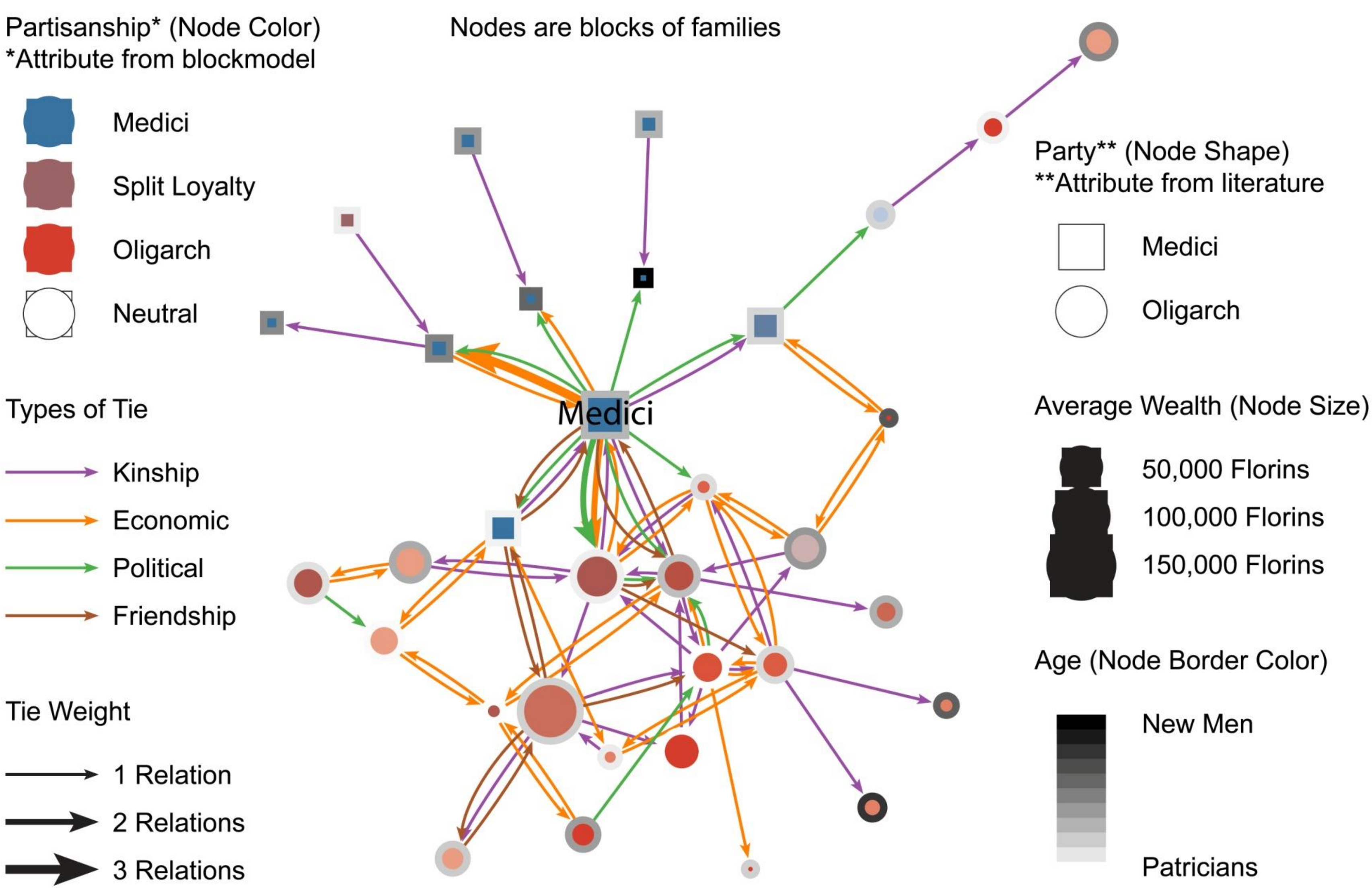
[1] Padgett, J.F. & Ansell, C.K. 1993. *American Journal of Sociology* 98: 1259–1319.

The Explanatory Power of Relations



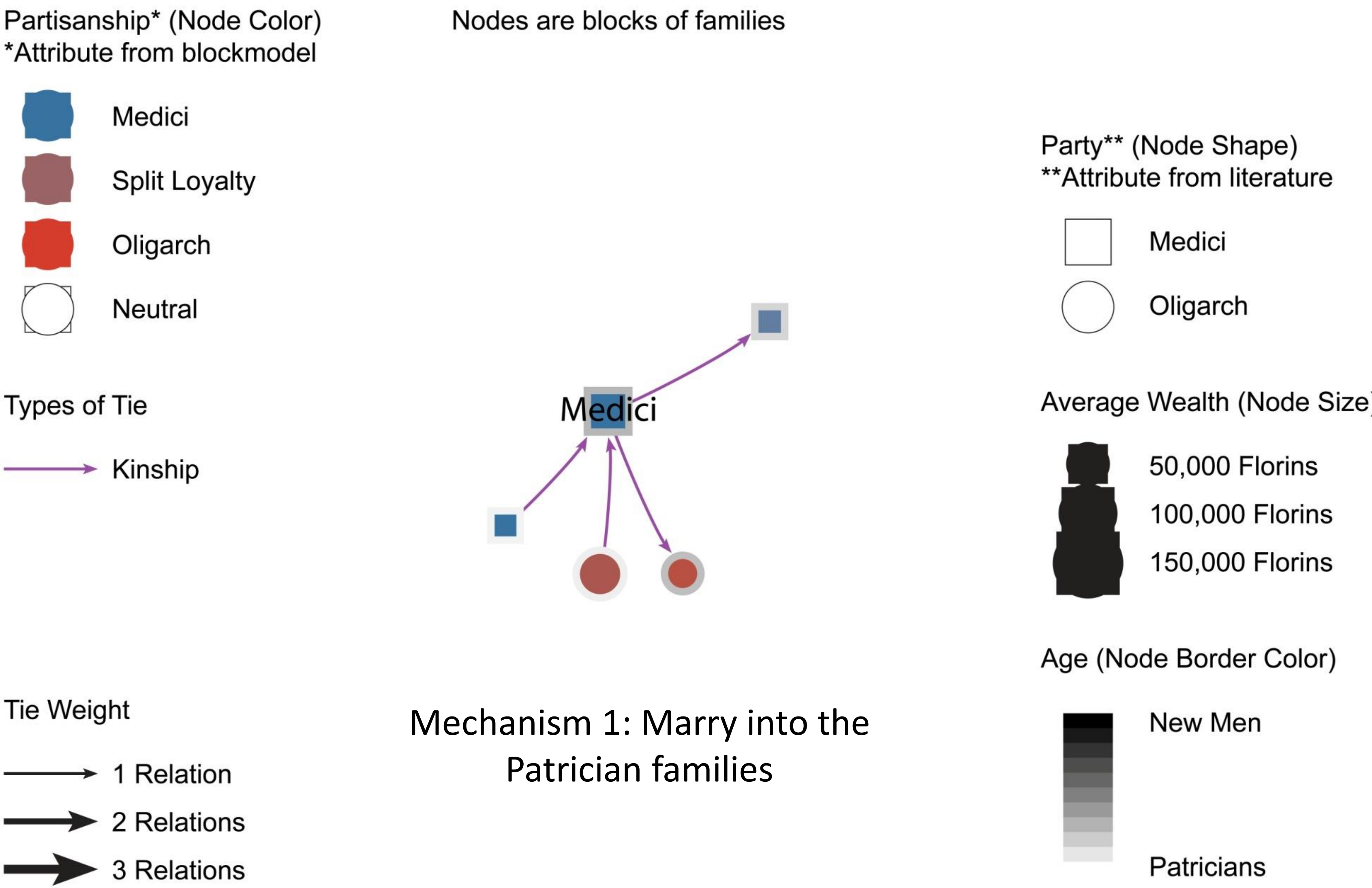
[1] Padgett, J.F. & Ansell, C.K. 1993. *American Journal of Sociology* 98: 1259–1319.

The Explanatory Power of Relations



[1] Padgett, J.F. & Ansell, C.K. 1993. *American Journal of Sociology* 98: 1259–1319.

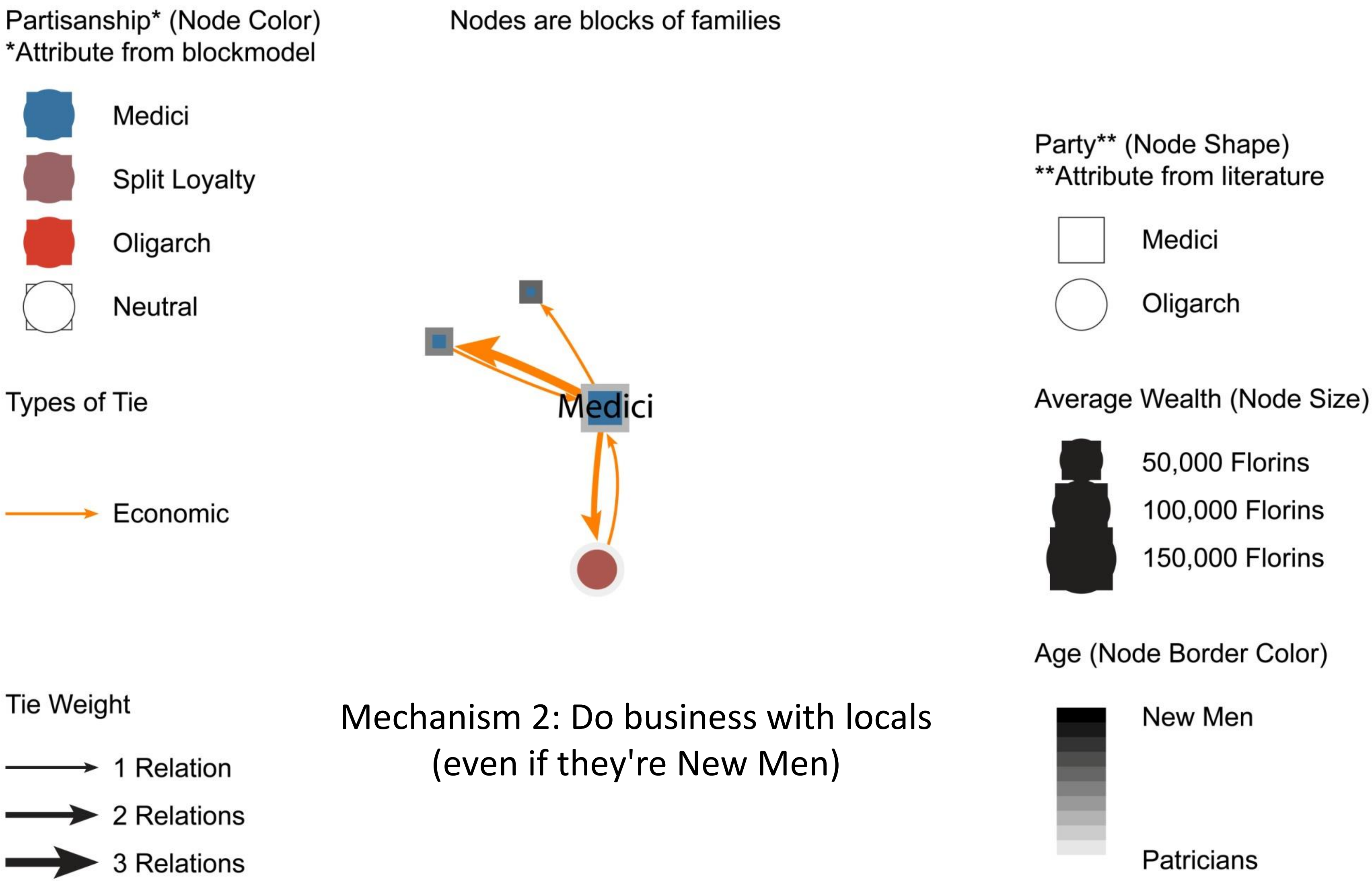
The Explanatory Power of Relations



Mechanism 1: Marry into the Patrician families

[1] Padgett, J.F. & Ansell, C.K. 1993. *American Journal of Sociology* 98: 1259–1319.

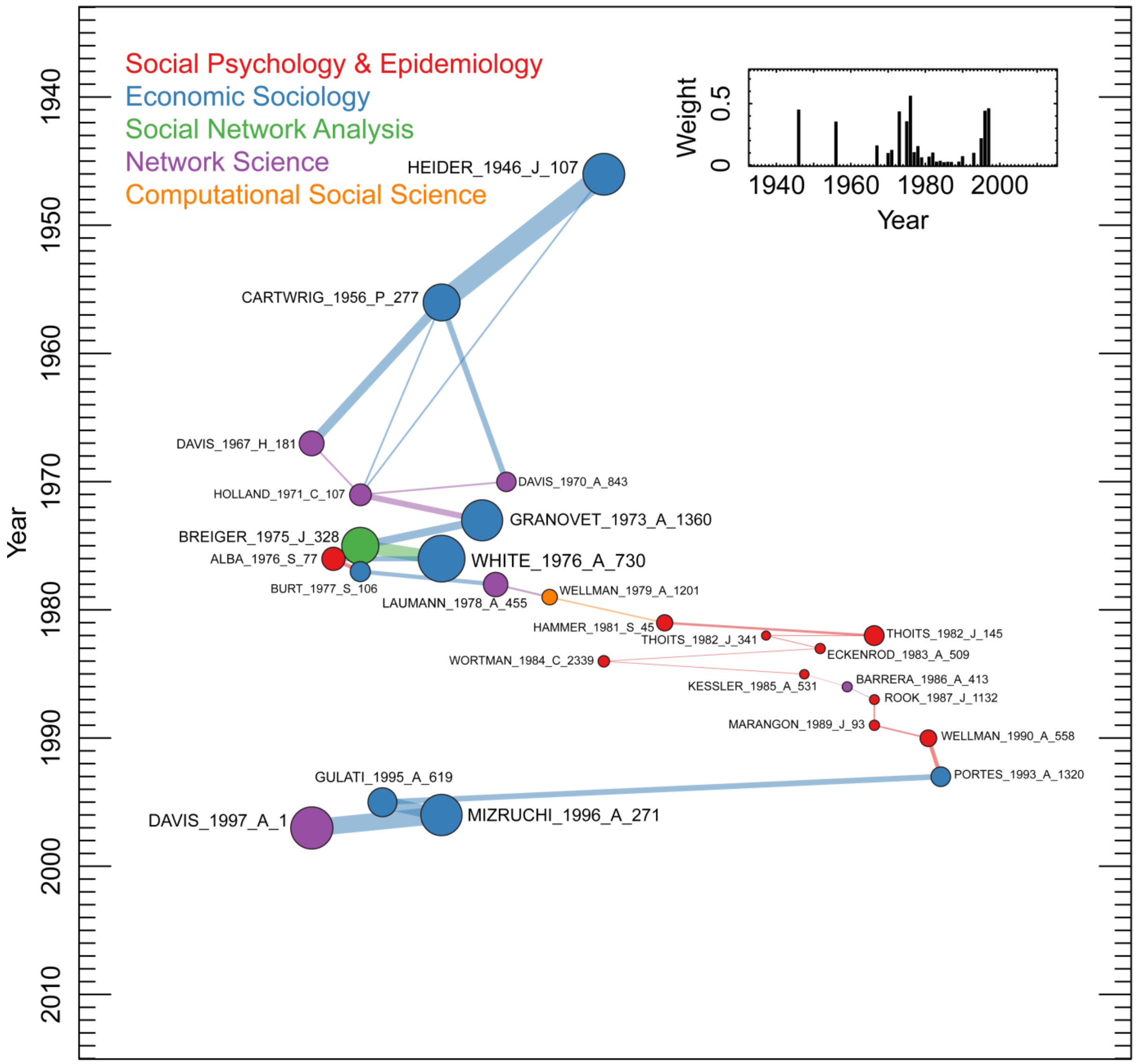
The Explanatory Power of Relations



Mechanism 2: Do business with locals
(even if they're New Men)

[1] Padgett, J.F. & Ansell, C.K. 1993. *American Journal of Sociology* 98: 1259–1319.

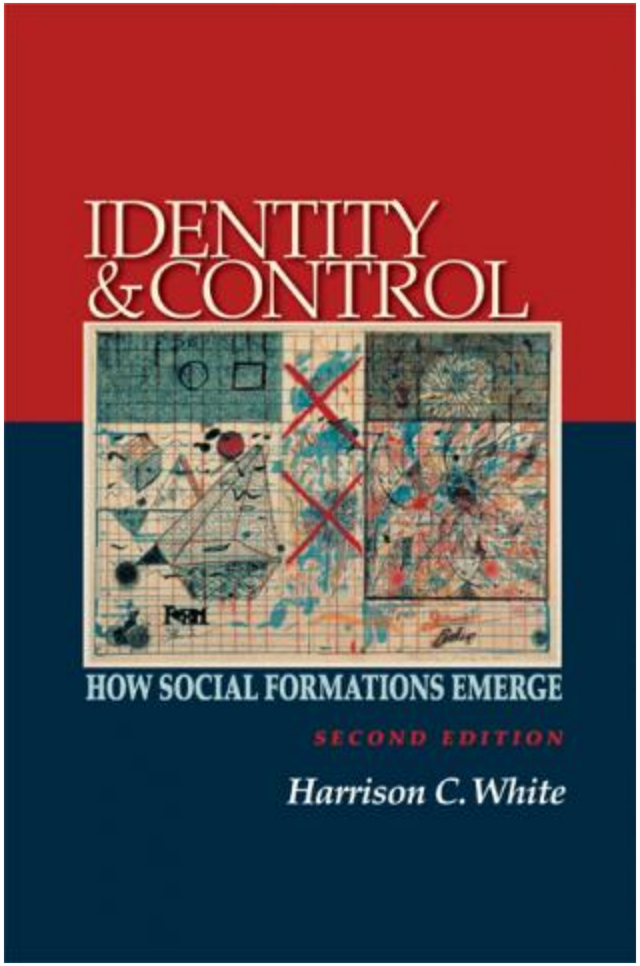
Social Network Analysis



[1] Lietz, H. 2016. *Scale-Free Identity*. Dissertation, University of Duisburg-Essen.

The Harvard Rennaissance

- **Harrison C. White** [1] instrumental in establishing Social Network Analysis as a field: His group „published so much important theory and research focused on social networks that social scientists everywhere, regardless of their field, could no longer ignore the idea.“ [2]



[1] White, H.C. 2008. *Identity and Control*. Princeton University Press.
[2] Freeman, L.C. 2004. *The Development of Social Network Analysis* (pp.127–128). Empirical Press.

The Harvard Rennaissance

- **Harrison C. White** [1] instrumental in establishing Social Network Analysis as a field [2]
- Breakthrough facilitated by combined use of **empirical data**, **visualizations**, and **explicit modeling** [2]
- **Blockmodeling** [3]: „structural ideas to make sense out of field work and… data analyses“ [2]
- At least until 1990: **Systematic cumulative knowledge production** with focus on blockmodeling [4]

TABLE I

Imaginary Data Illustrating Blockmodels, Lean Fit, and Zeroblocks.

(a)

1	0	1	0	0	0	0	1	1	0	0
2	0	0	0	1	0	1	1	0	0	0
3	0	0	0	1	0	1	0	0	0	1
4	0	0	1	0	0	1	0	0	0	1
5	0	1	1	0	0	0	1	1	0	1
6	0	0	0	1	0	0	0	0	0	1
7	1	1	0	0	1	0	0	1	1	0
8	0	1	1	0	0	1	1	0	0	1
9	0	1	0	1	0	0	1	1	0	0
10	0	0	1	1	0	0	0	0	0	0

(b)

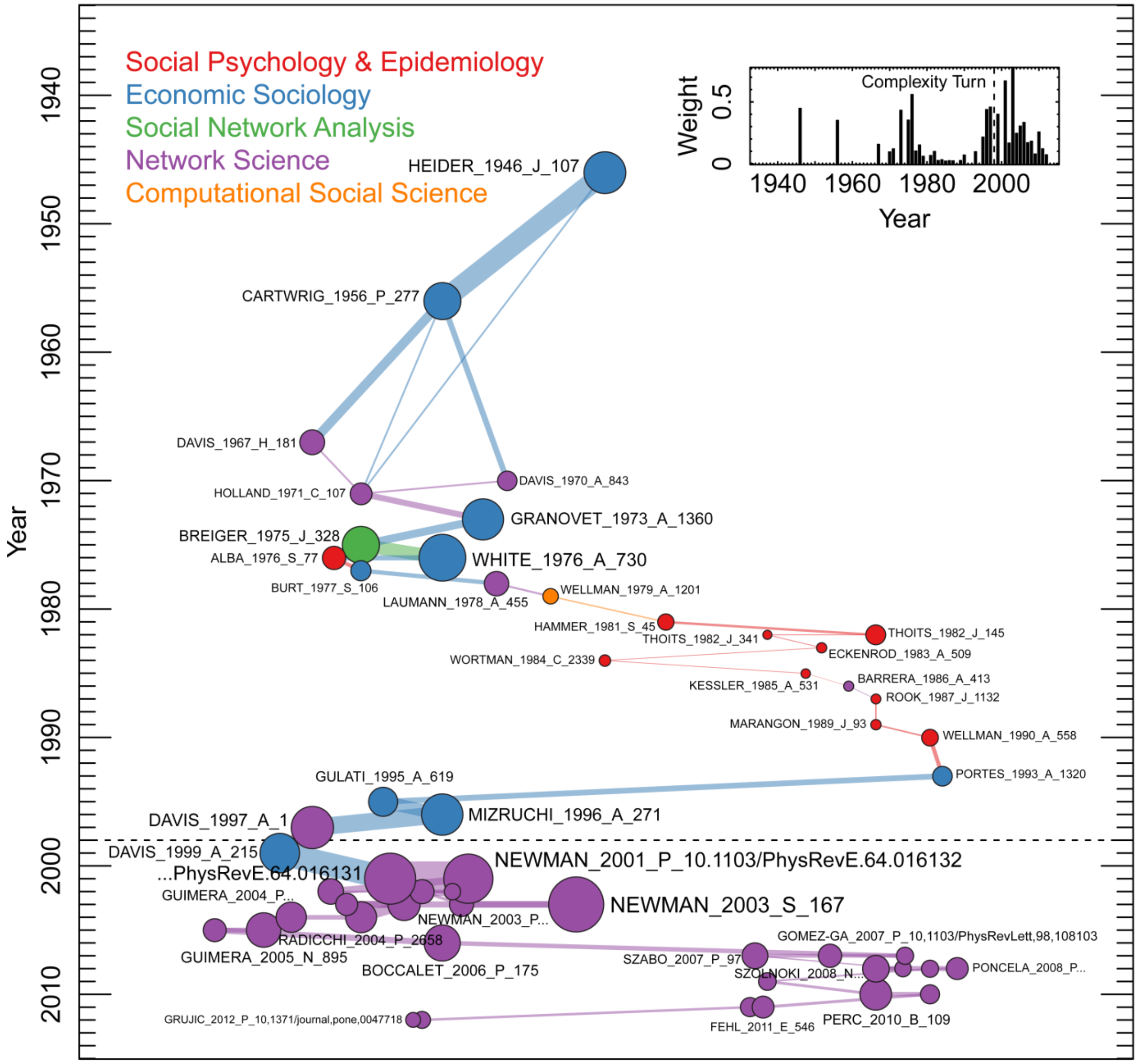
2	0	1	0	0	1	1	0	0	0	0
7	1	0	1	0	0	0	0	1	1	1
8	1	1	0	1	0	1	1	0	0	0
3	0	0	0	0	1	1	1	0	0	0
4	0	0	0	1	0	1	1	0	0	0
6	0	0	0	0	1	0	1	0	0	0
10	0	0	0	1	1	0	0	0	0	0
1	1	1	1	0	0	0	0	0	0	0
5	1	1	1	1	0	0	1	0	0	0
9	1	1	1	0	1	0	0	0	0	0

(c)

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix}$$

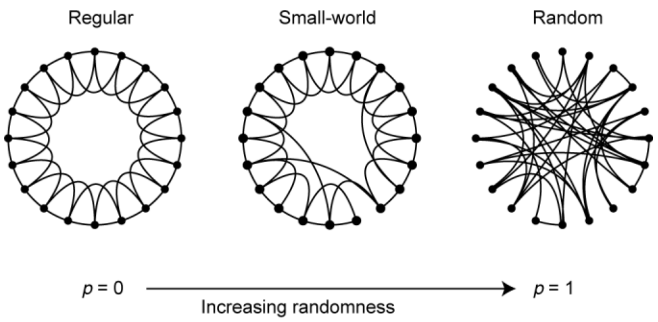
[1] White, H.C. 2008. *Identity and Control*. Princeton University Press.
[2] Freeman, L.C. 2004. *The Development of Social Network Analysis* (pp.127–128). Empirical Press.
[3] Breiger, R.L. *et al.* 1975. *Journal of Mathematical Psychology* 12:328–383.
[4] Hummon, N.P. & Carley, K.M. 1993. *Social Networks* 15:71–106.

The Complexity Turn



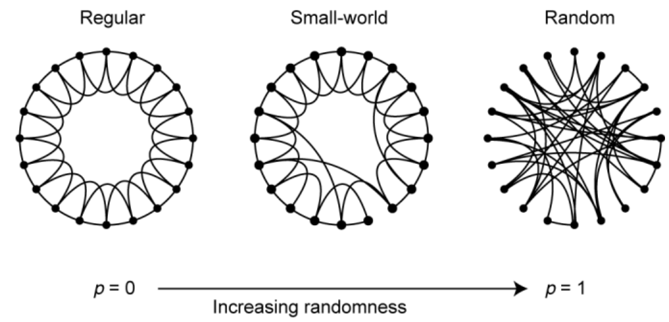
[1] Lietz, H. 2016. *Scale-Free Identity*. Dissertation, University of Duisburg-Essen.

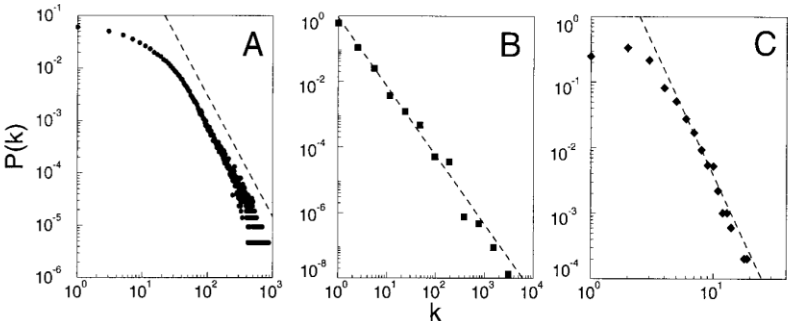
- **Small-world networks** [1]: Networks with two realistic properties (high average clustering coefficient and short average path length) from pure rewiring



[1] Watts, D.J. & Strogatz, S.H. 1998. *Nature* 393:440–442.

- **Small-world networks** [1]: Networks with two realistic properties (high average clustering coefficient and short average path length) from pure rewiring
- **Scale-free networks** [2]: Generative model for networks without a characteristic scale

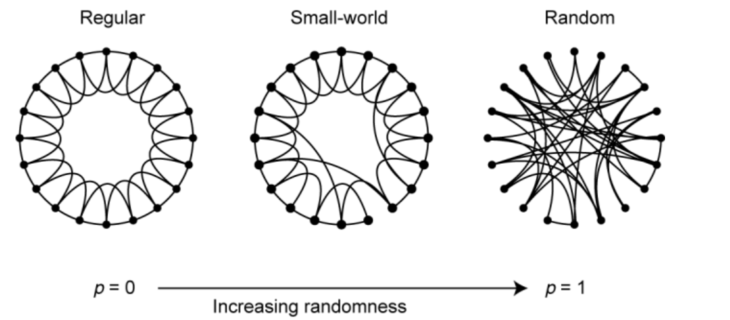


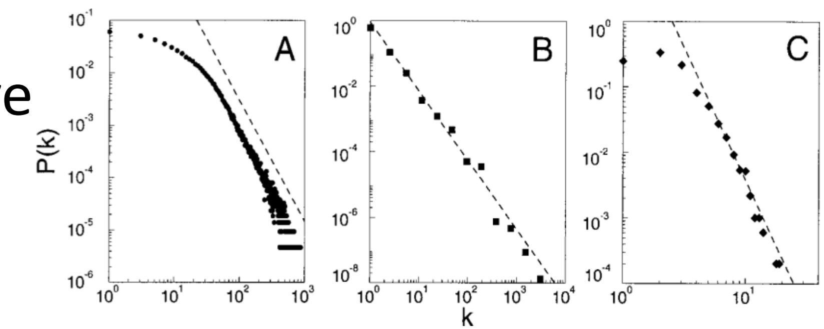


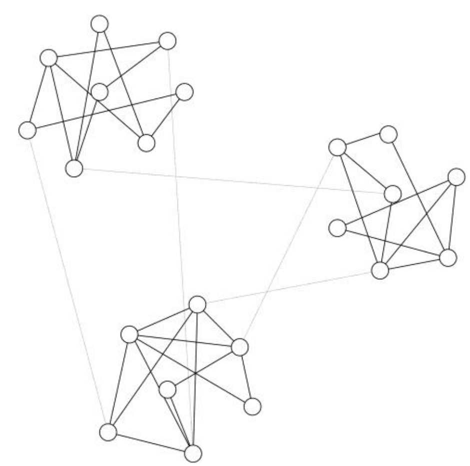
[1] Watts, D.J. & Strogatz, S.H. 1998. *Nature* 393:440–442.
[2] Barabási, A.-L. & R. Albert. 1999. *Science* 286:510–512.

Network Science

- **Small-world networks** [1]: Networks with two realistic properties (high average clustering coefficient and short average path length) from pure rewiring
- **Scale-free networks** [2]: Generative model for networks without a characteristic scale
- **Community structure** [3]: organization of networks in tightly knit groupy







[1] Watts, D.J. & Strogatz, S.H. 1998. *Nature* 393:440–442.
[2] Barabási, A.-L. & R. Albert. 1999. *Science* 286:510–512.
[3] Girvan, M. & Newman, M.E.J. 2002. *Proc. Natl. Acad. Sci.* 99:7821–7826.

	Social Network Analysis	Network Science
Basic Goal	**Understanding** how actors create relations and how relations create actors	**Modeling** complex systems
Basic Scope	**Contextual**, focused on individual data set and on individuals in the data set	Context-free, **universal**
Motivation	Understand human behavior and **social structures**	Understand self-organization in **complex systems**

[1] Zweig, K.A. 2016. *Network Analysis Literacy* (p.40). Springer.

Complex System

A system that consists of many **parts**, the **interactions** of which give rise to a **whole** that is **more than the sum of the parts**

Complexity Science

Patterns

Topics

Tools

Foundations

[1] Shalizi, C.R. 2006. In *Complex Systems Science in Biomedicine* (pp.33-114). Springer.

gesis Leibniz Institute for the Social Sciences

Patterns

Topics

Small-world phenomenon

Scale-freeness

Community structure

…

Tools

Foundations

[1] Shalizi, C.R. 2006. In *Complex Systems Science in Biomedicine* (pp.33-114). Springer.

Patterns

Topics

Formal theory

Nature of organization

Origins of complexity

…

Tools

Foundations

[1] Shalizi, C.R. 2006. In *Complex Systems Science in Biomedicine* (pp.33-114). Springer.

Patterns

Topics

Networks

Equilibrium and turbulence

Evolution of cooperation

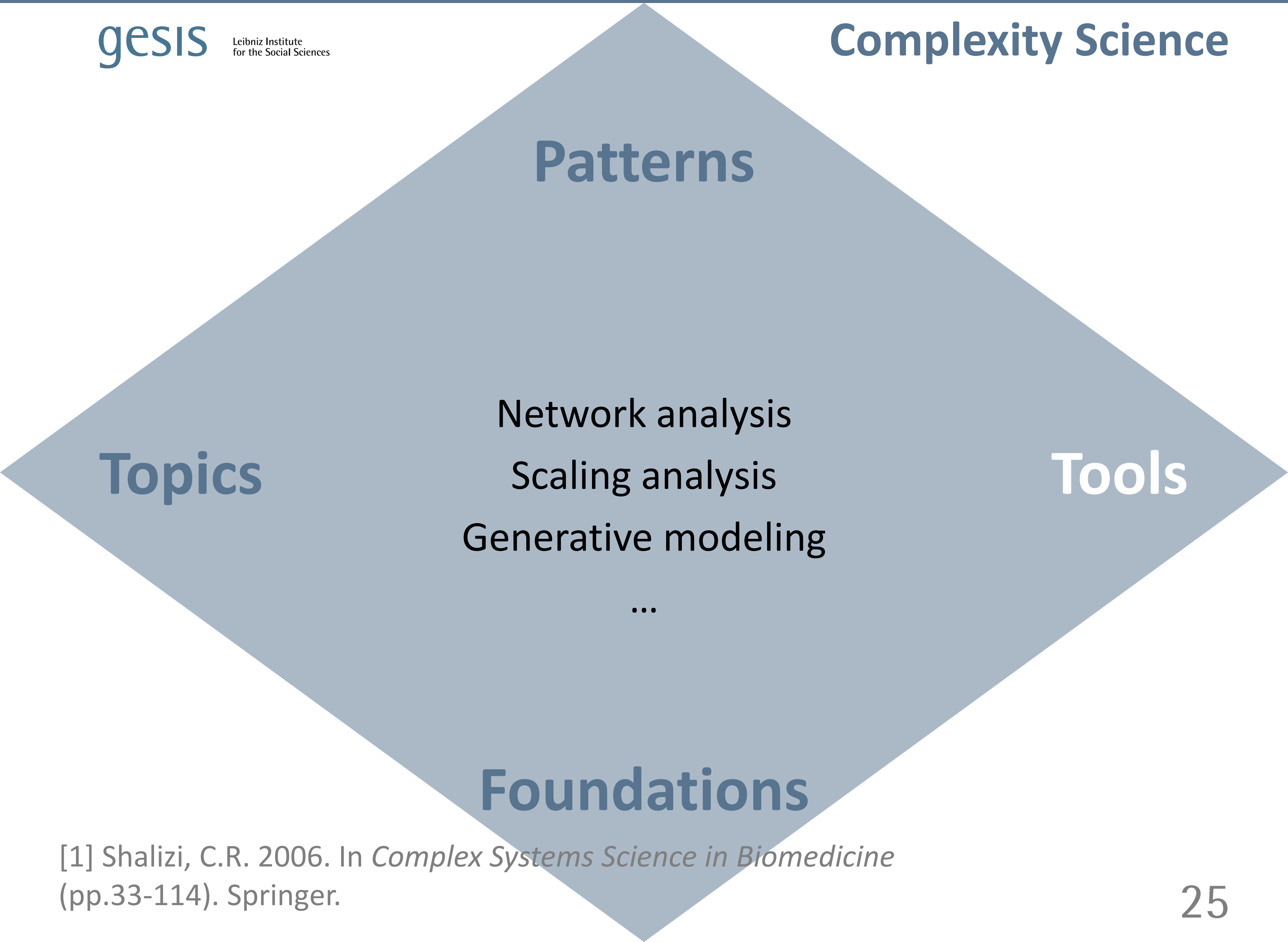
…

Tools

Foundations

[1] Shalizi, C.R. 2006. In *Complex Systems Science in Biomedicine* (pp.33-114). Springer.

Complexity Science

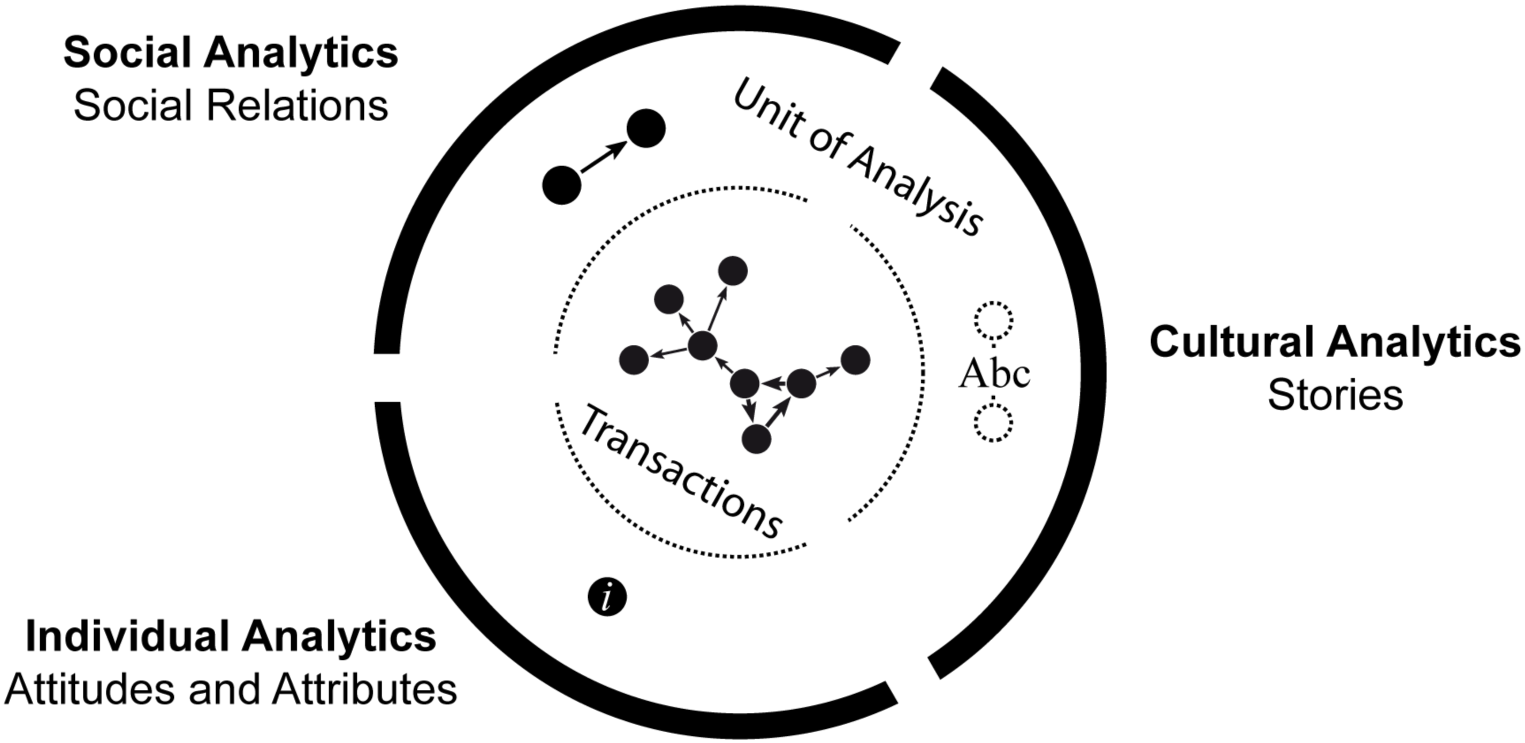


[1] Shalizi, C.R. 2006. In *Complex Systems Science in Biomedicine* (pp.33-114). Springer.

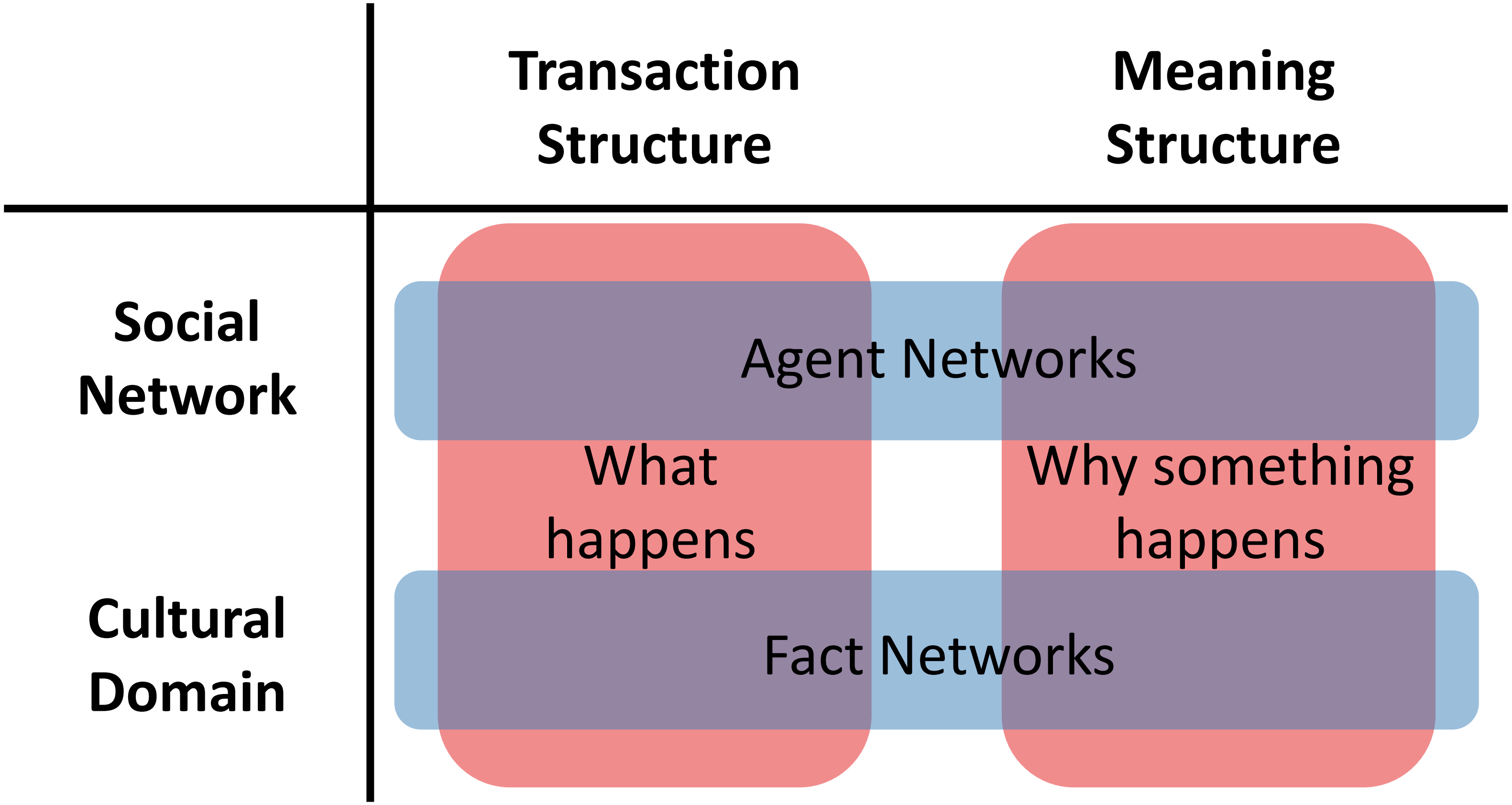
Teaching Goals

- **Empirical**
  - Study small-scale to large-scale networks
  - Draw from Social Network Analysis and Network Science
  - Be literate
- **Visualization**
  - Communicate information through graphs
- **Modeling**
  - Understand social networks as models of complex systems
  - Put methods into perspective

All **traces of human behavior** that are created through the use of digital technology or made accessible by it

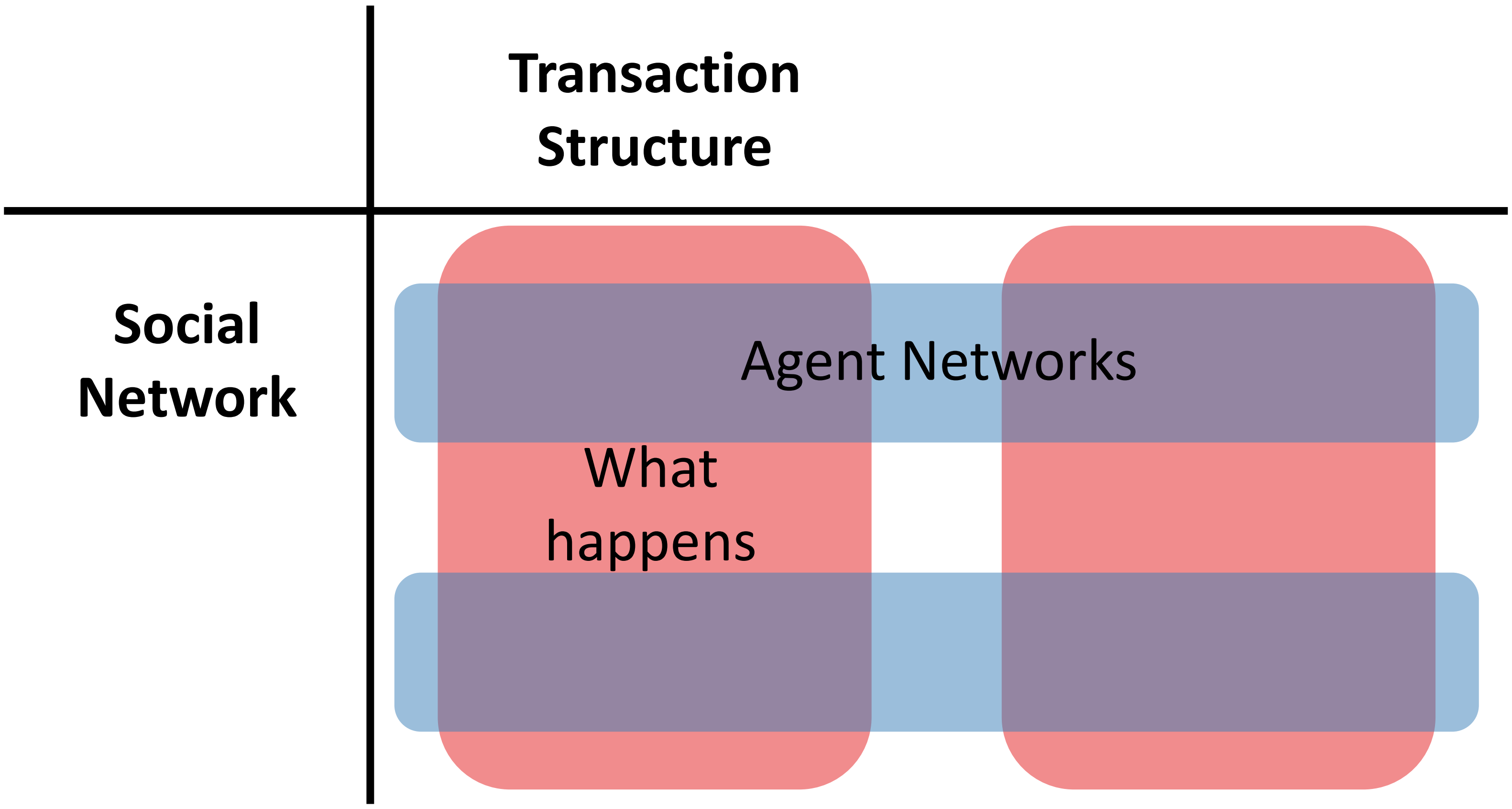


Network Taxonomy



[1] Fuchs, S. (2005). *Against Essentialism*. Harvard University Press.

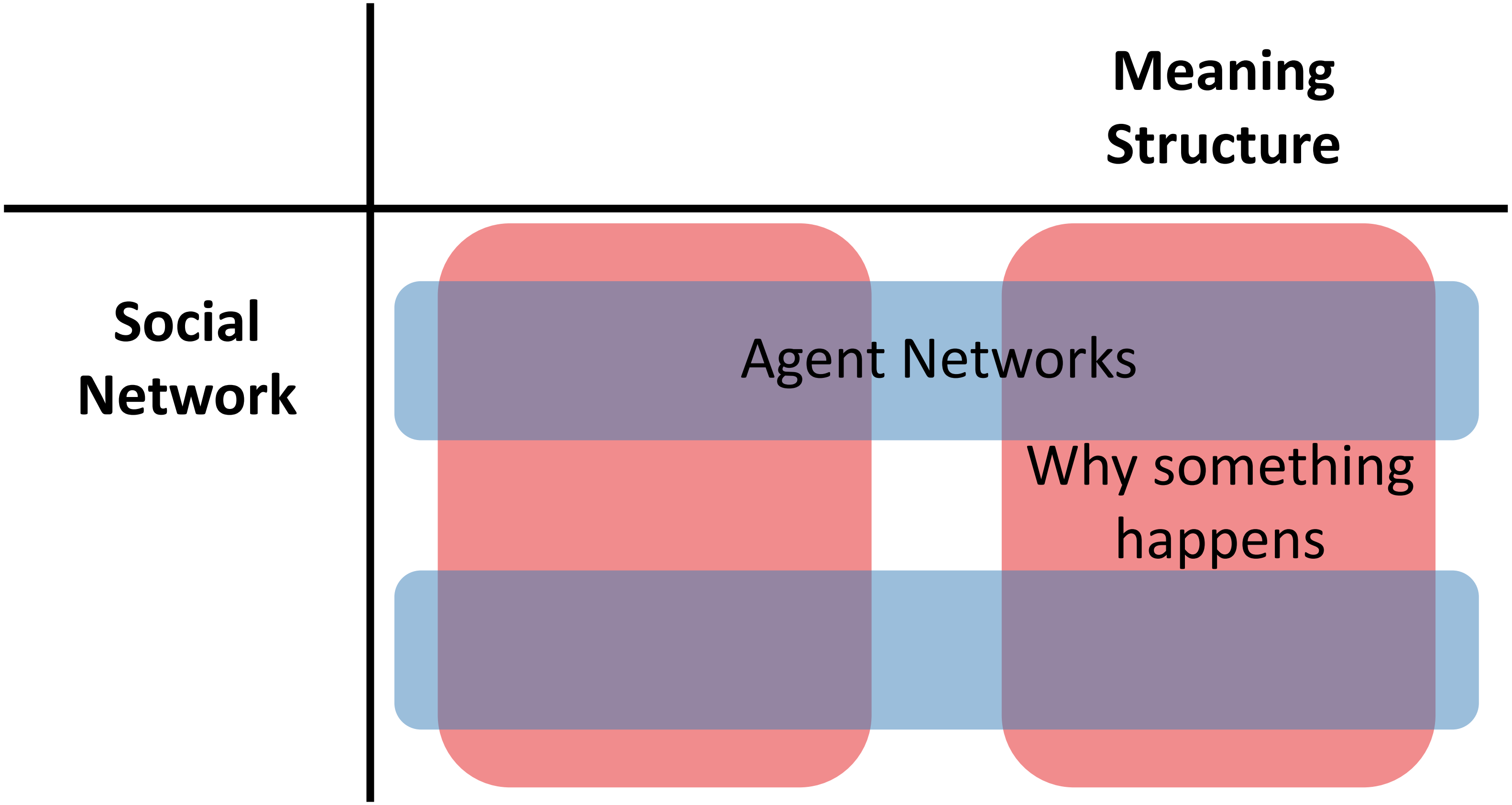
Network Taxonomy



e.g., communication network (1st order observation)

[1] Fuchs, S. (2005). *Against Essentialism*. Harvard University Press.

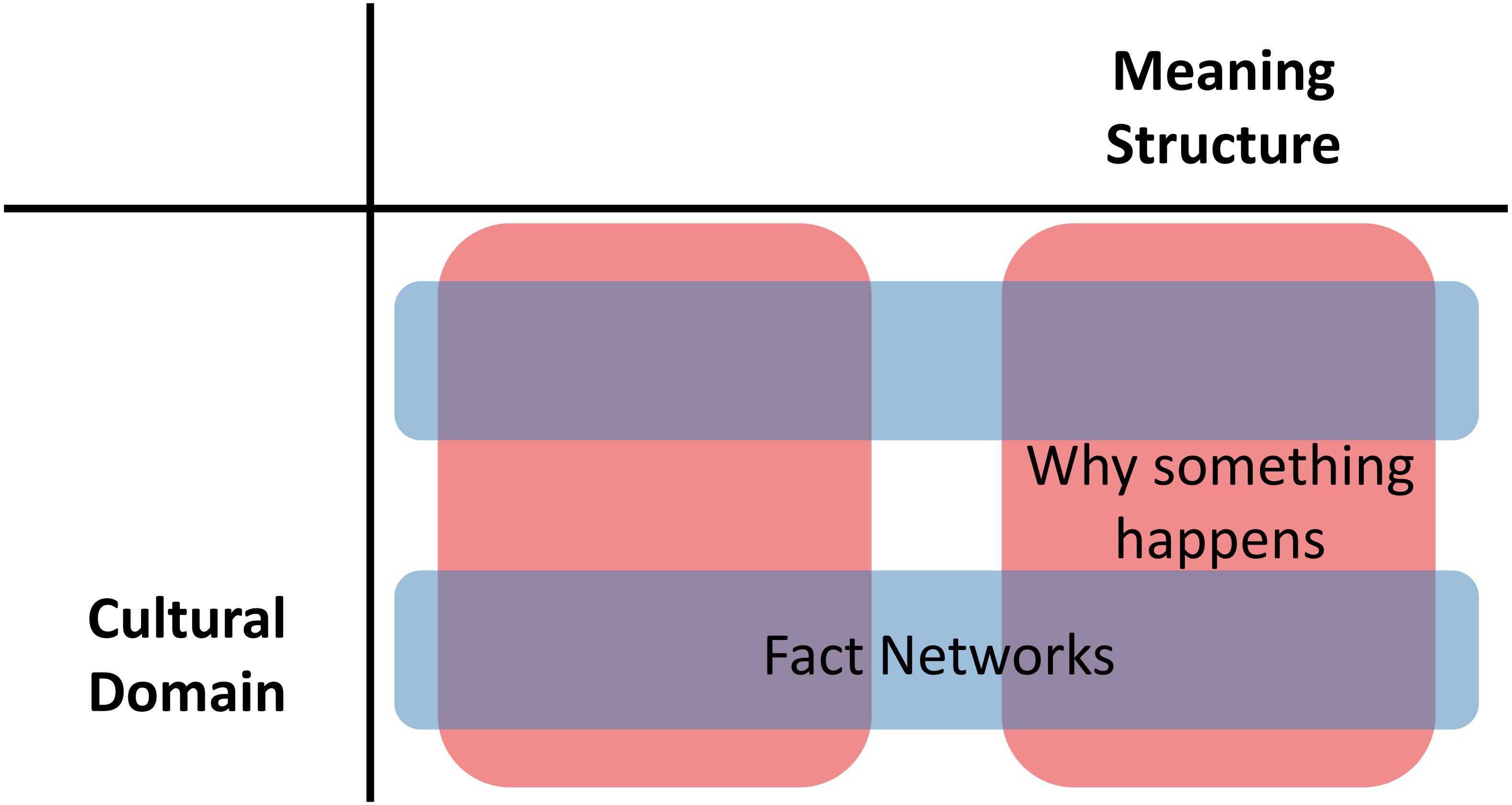
Network Taxonomy



e.g., friendship network (observed in 2nd order observation)

[1] Fuchs, S. (2005). *Against Essentialism*. Harvard University Press.

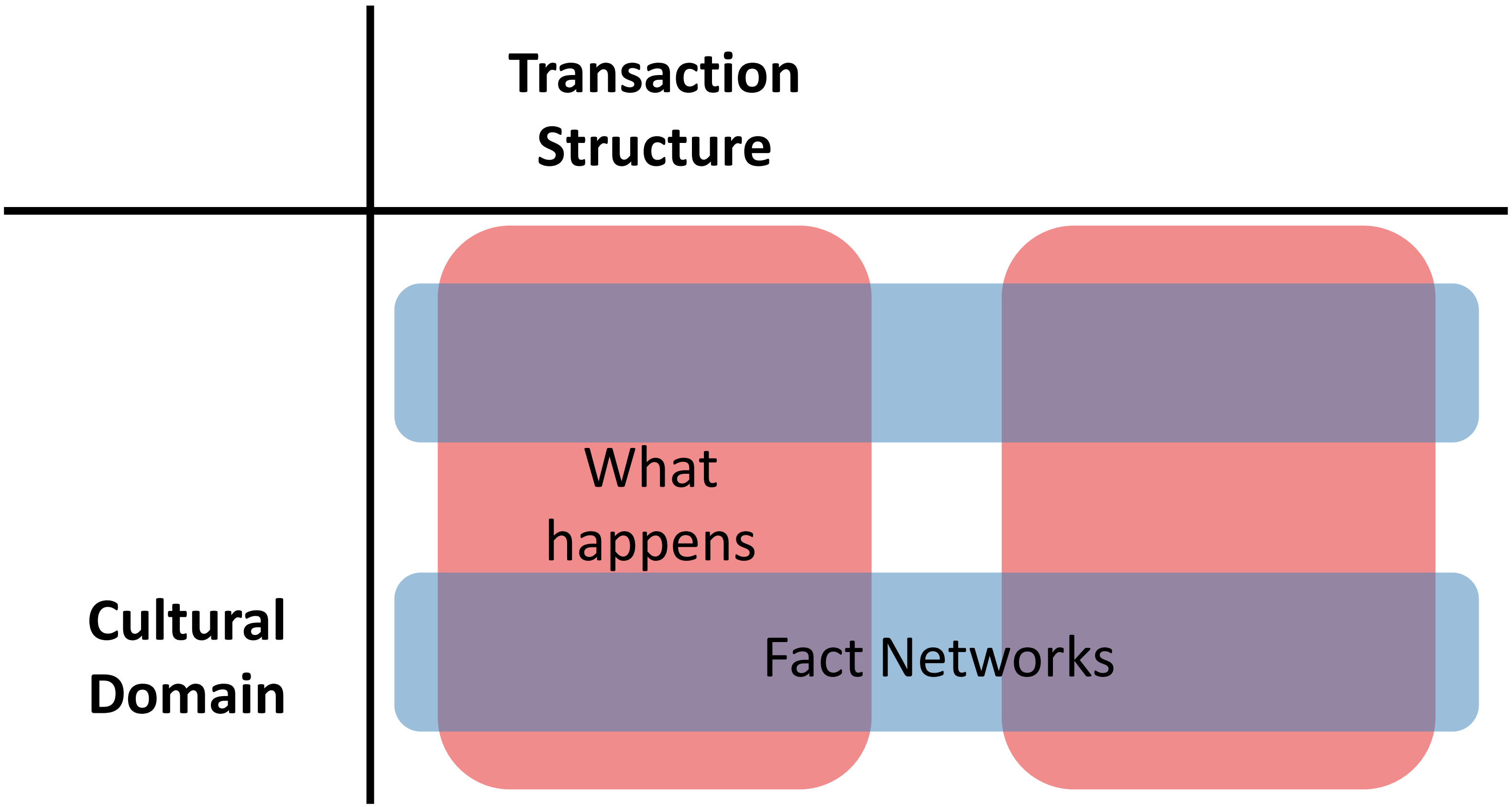
Network Taxonomy



e.g., concept co-occurrence network (observed in 2nd order observation)

[1] Fuchs, S. (2005). *Against Essentialism*. Harvard University Press.

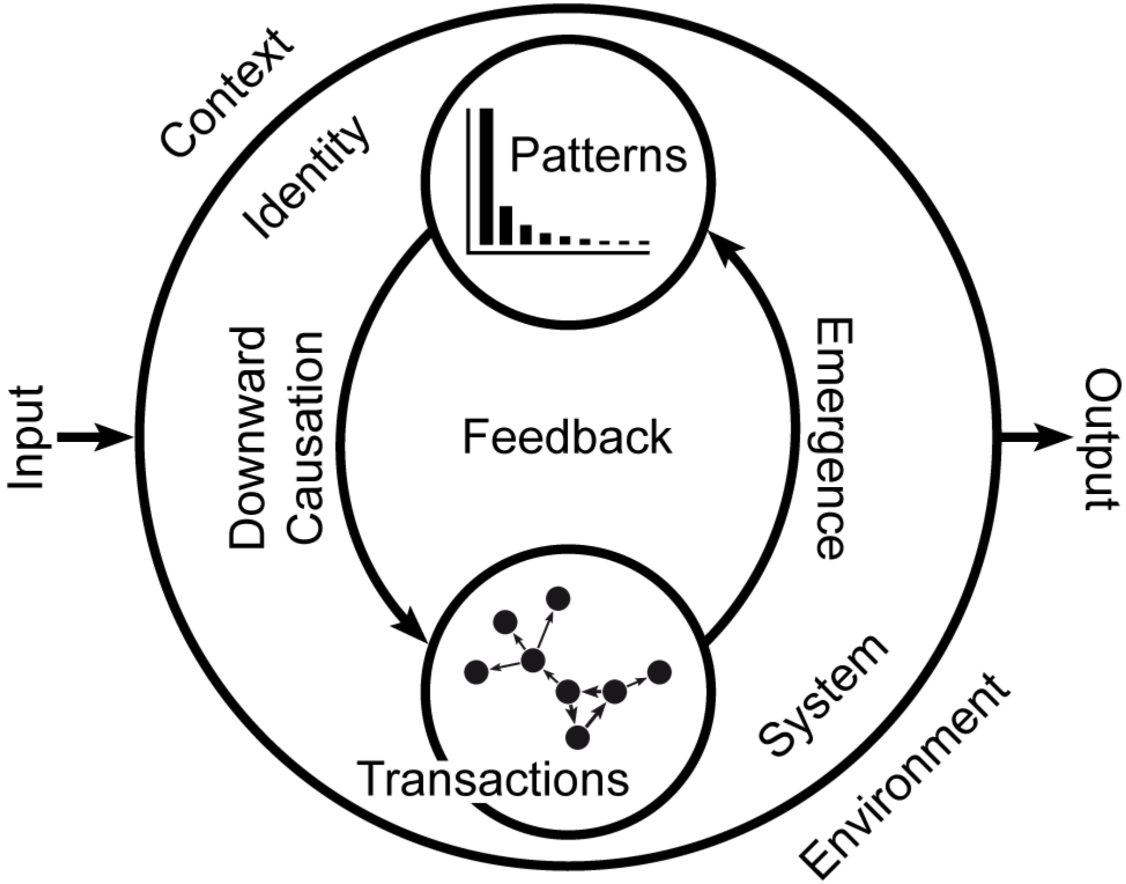
Network Taxonomy



Non-existent (no 1st order observation by entities without agency)

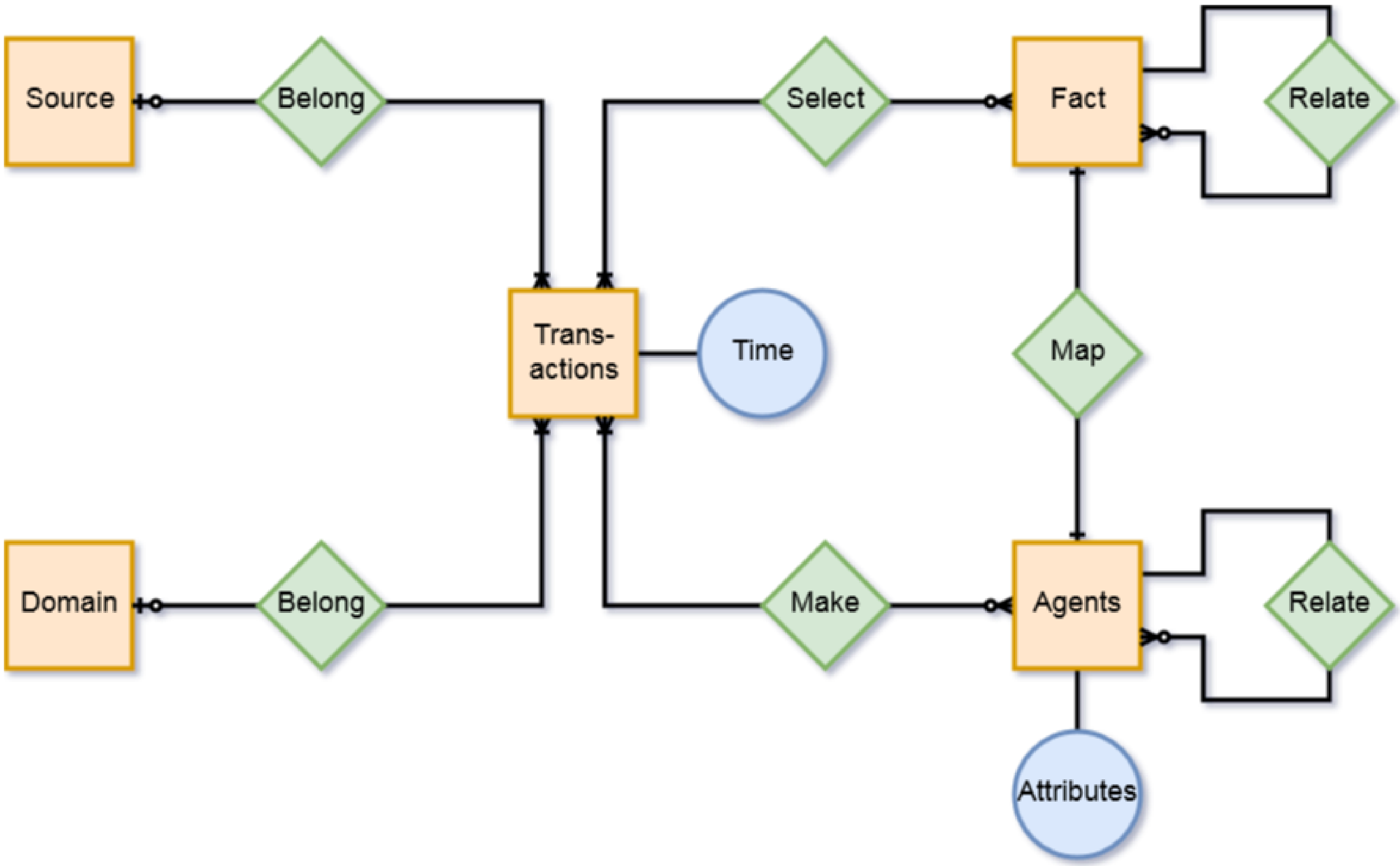
[1] Fuchs, S. (2005). *Against Essentialism*. Harvard University Press.

Identity Model

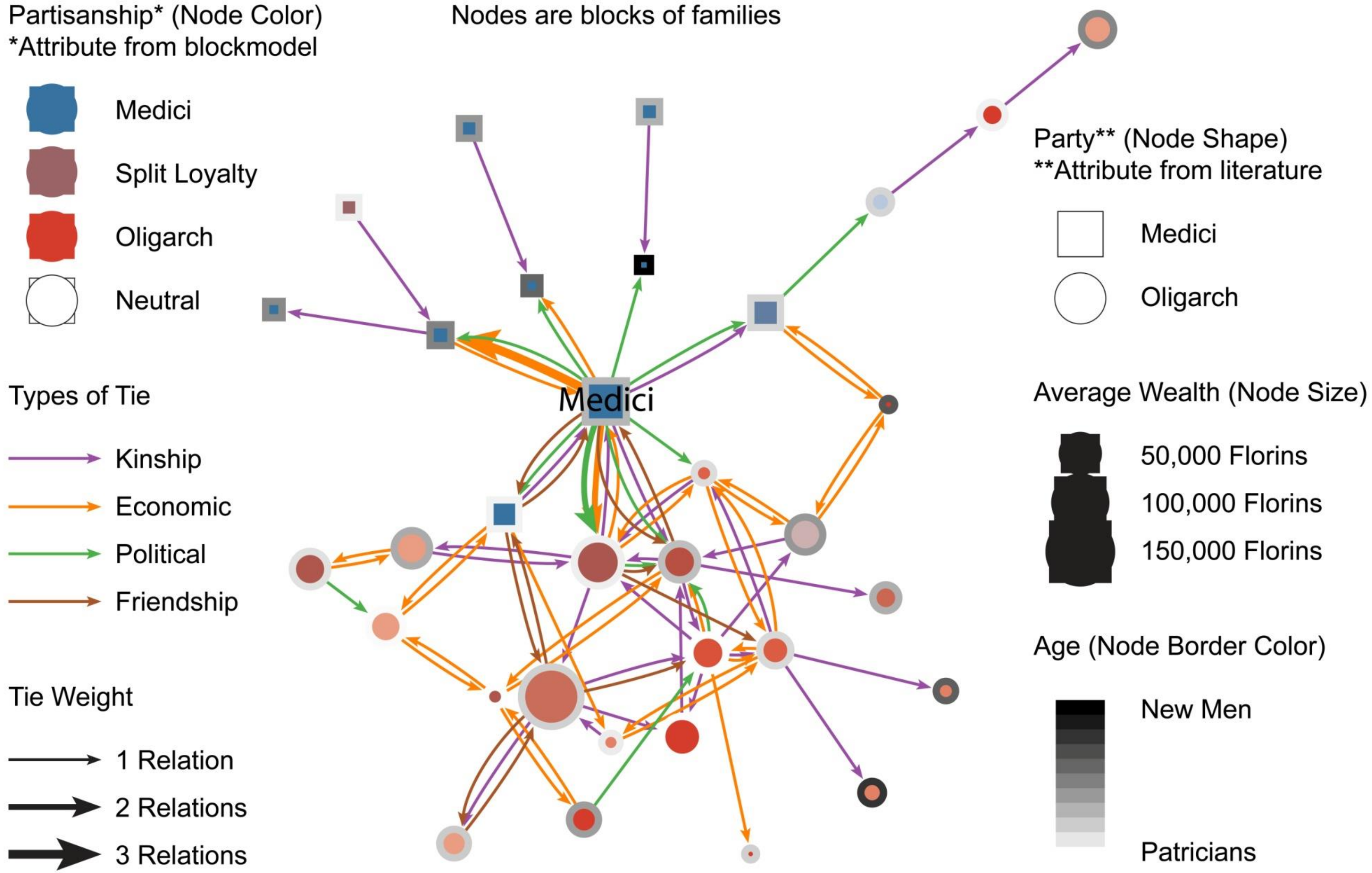


[1] Mohr (1998). *Annual Review of Sociology* 24:345–370.
[2] Fuhse (2009). *Sociological Theory* 27:51–73.
[3] Page (2015). *Annual Review of Sociology* 41:21–41.

Data Model


Source
Belong
Select
Fact
Relate
Trans-
actions
Time
Map
Domain
Belong
Make
Agents
Relate
Attributes

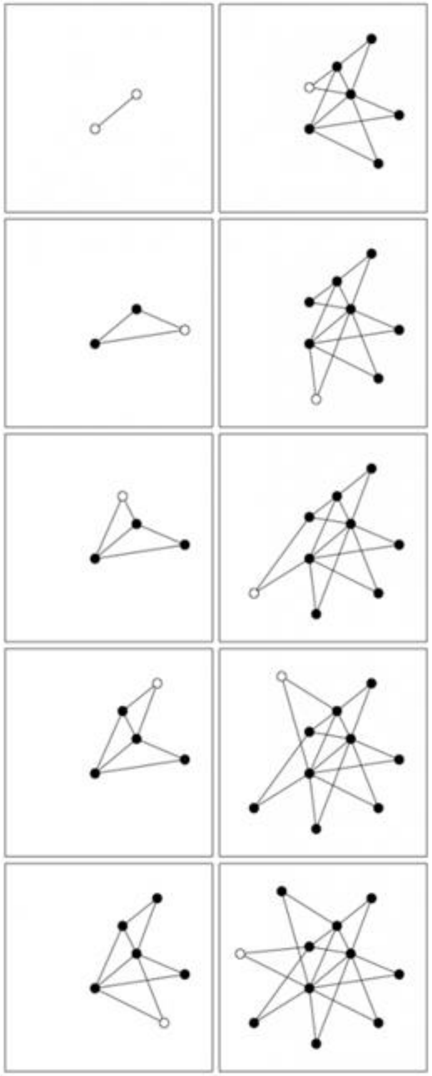
Plotting As Information Visualization



[1] Krempel, L. (2011). In Scott, J. & Carrington, P. J. (eds.), *The Sage Handbook of Social Network Analysis* (pp.558–577). Sage.

Two Schools of Modeling

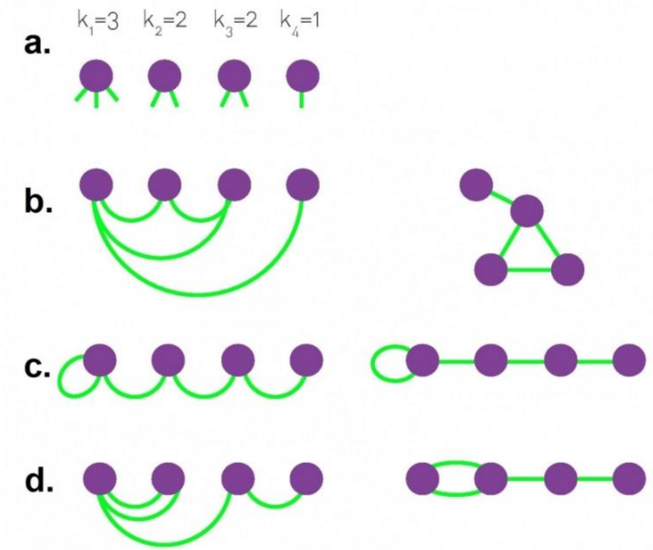
- **Mechanistic modeling**
  1. Identify microscopic mechanisms
  2. Define dynamic model
  3. Generate emergent patterns of real systems [1]
  - *e.g.*, Barabási–Albert Model [2]
  - Akin to kinetic theory and agent-based Analytical Sociology [3]



[1] Holme, P. & Liljeros, F. 2015. *Frontiers in Physics* 3:78.
[2] Barabási, A.-L. & R. Albert. 1999. *Science* 286:510–512.
[3] Hedström, P. & Bearman, P. 2011. *The Oxford Handbook of Analytical Sociology*. Oxford University Press.

Two Schools of Modeling

- **Mechanistic modeling**
- **Ensemble approach**
  1. Identify static properties of system
  2. Obtain ensembles of random graphs that have the same properties
  3. Use their parameters to summarize the system [1]
  - *e.g.*, Exponential Random Graph Models [2], Stochastic Blockmodels [3]
  - Akin to Statistical Mechanics



[1] Cimini, G. *et al.* 2019. *Nature Reviews Physics* 1:58–71.
[2] Robins, G. et al. 2007. *Social Networks* 29: 173–191.
[3] Holland, P.W. *et al.* 1983. *Social Networks* 5:109–137.

gesis Leibniz Institute for the Social Sciences

Course Program

Monday, 15.07.19: Introduction and Micro-Scale Analysis
11:00-12:00 Lecture: Introduction and Micro-Scale Analysis
12:00-13:00 Lunch
13:00-13:30 Demo: Repetition of Basic Python Concepts
13:30-15:00 Demo: Network Construction and Centrality
15:00-15:30 Coffee Break
15:30-17:00 Exercise: Network Construction and Centrality

Tuesday, 16.07.19: Meso-Scale Analysis
09:00-09:30 Lecture: Meso-Scale Analysis
09:30-11:00 Demo: Community Detection
11:00-11:30 Coffee Break
11:30-13:00 Exercise: Community Detection
13:00-14:00 Lunch
14:00-15:30 Demo: Stochastic Blockmodeling
15:30-17:00 Exercise: Stochastic Blockmodeling
17:00 Fingerfood & Drinks

Course Program

Wednesday, 17.07.19: Macro-Scale Analysis
09:00-09:30 Lecture: Macro-Scale Analysis
09:30-11:00 Demo: Small-World Networks
11:00-11:30 Coffee Break
11:30-13:00 Exercise: Small-World Networks
13:00-14:00 Lunch
14:00-15:30 Demo: Scale-Free Networks
15:30-17:00 Exercise: Scale-Free Networks
17:00 Geocaching and Brauhaus Visit

Thursday, 18.07.19: Social Simulation
09:00-09:30 Lecture: Social Simulation
09:30-11:00 Demo: Mechanistic Modeling
11:00-11:30 Coffee Break
11:30-13:00 Exercise: Mechanistic Modeling
13:00-14:00 Lunch
14:00-15:30 Demo: Dynamics on Networks
15:30-17:00 Exercise: Dynamics on Networks

Course Program

Friday, 19.07.19: Project Work
09:00-13:00 Project Work
13:00-14:00 Lunch
14:00-14:30 Project Work: Presentation of Results
14:30-15:30 Summary and Evaluation

- Project work
  - Empirical, visualization, modeling
  - Your data, given data
- Individual study
- Repeat sessions

39. Methodenseminar: Big Data Module II

Any questions before we go on?

Leibniz Institute
for the Social Sciences

Leibniz
Association

Micro-Scale Analysis

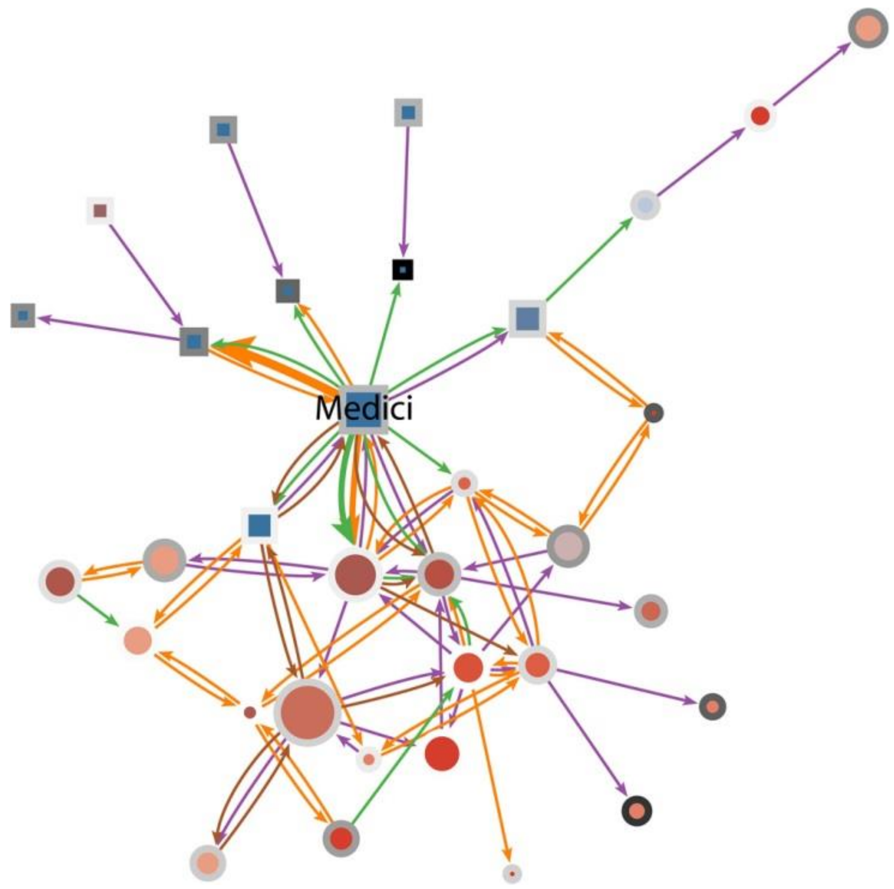
Network Construction and Centrality

Graph Theoretical Notation

- In maths, a network is called a **graph**, denoted by G
- A graph consists of a **set of vertices** V (nodes) and a **set of edges** E: $G = (V, E)$
- n is the **number of vertices**: $V = \{v_1, v_2, \dots, v_n\}$
- m is the **number of edges**: $E = \{e_1, e_2, \dots, e_m\}$
- **Undirected edges** are represented by unordered pairs of nodes
- **Directed edges** are represented by ordered pairs of nodes, and the graph is called a **Digraph**
- If edges are **weighted**, a real value is assigned to them through a weight function w: $G = (V, E, w)$

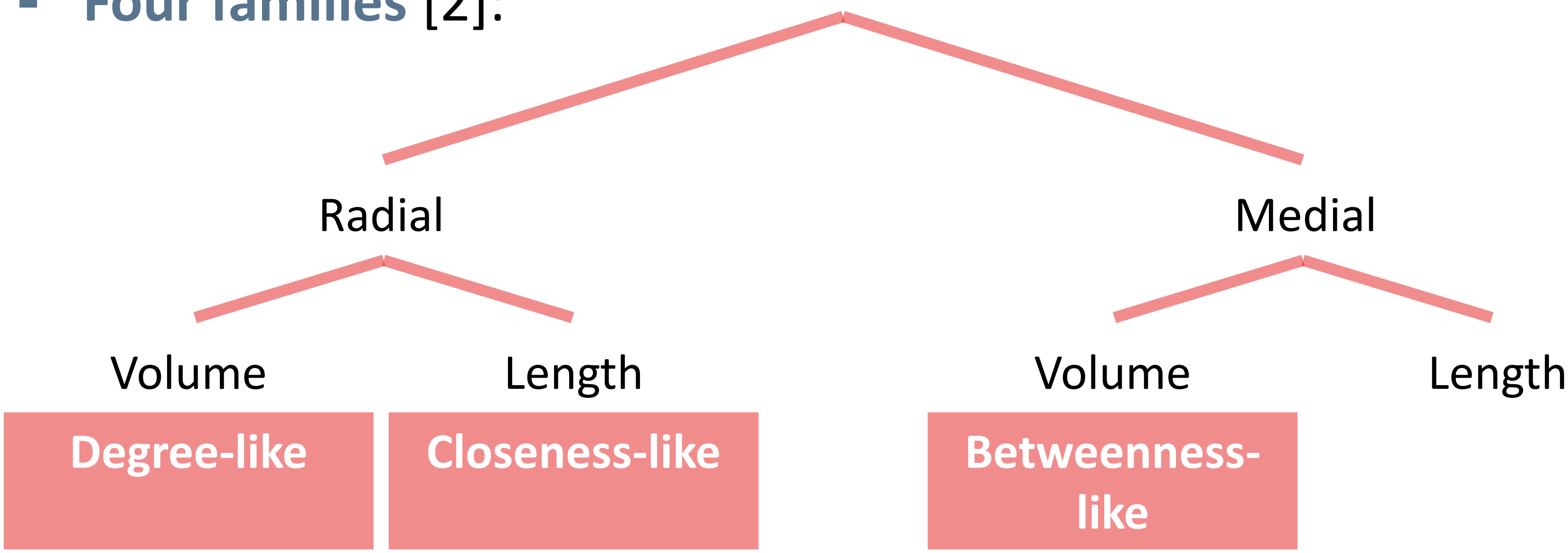
Graph Theoretical Notation

- If multiple (parallel) edges exist among a node pair, G is a multiplex graph or **multigraph**



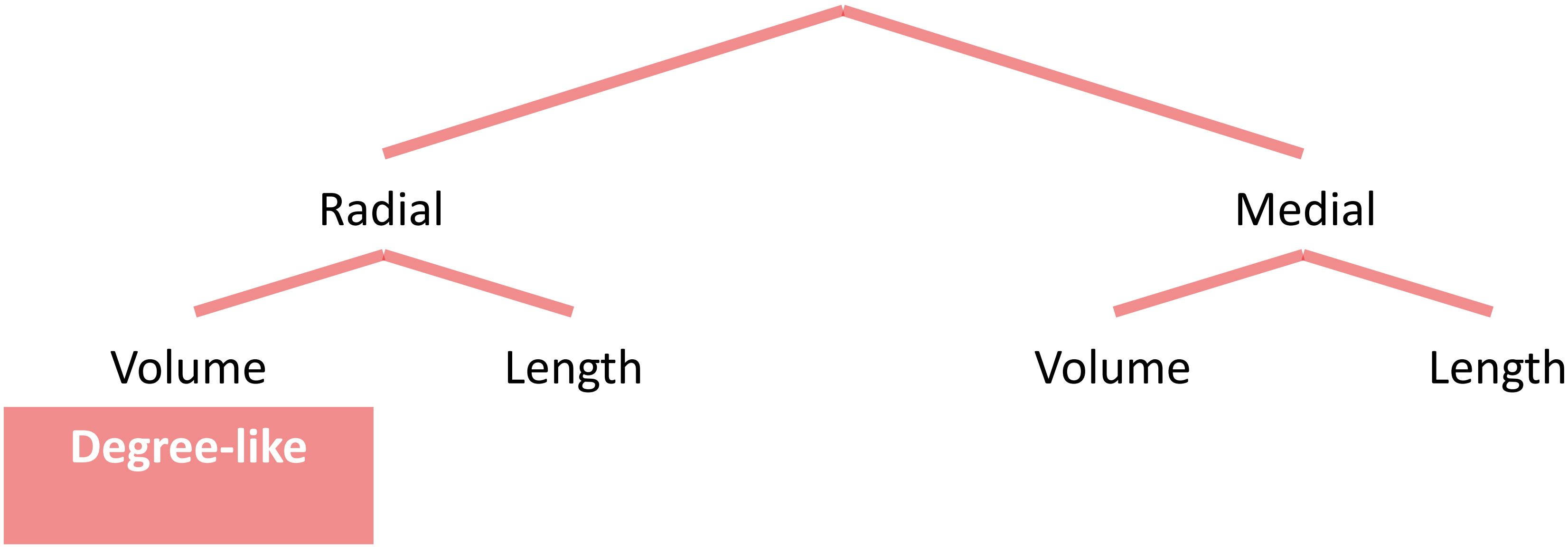
Centrality Ranking

- „[A] **centrality index** is a real-valued function on the nodes of a graph, i.e., it assigns a number to all nodes. This value is only depending on the structure of the graph, not on external parameters associated with the nodes.“ [1]
- They quantify a node‘s involvement in a given set of **walks**
- **Four families** [2]:



[1] Zweig, K.A. 2016. *Network Analysis Literacy* (p.245). Springer.
[2] Borgatti, S.P. & Everett, M.G. 2006. *Social Networks* 28:466–484.

Degree Centrality

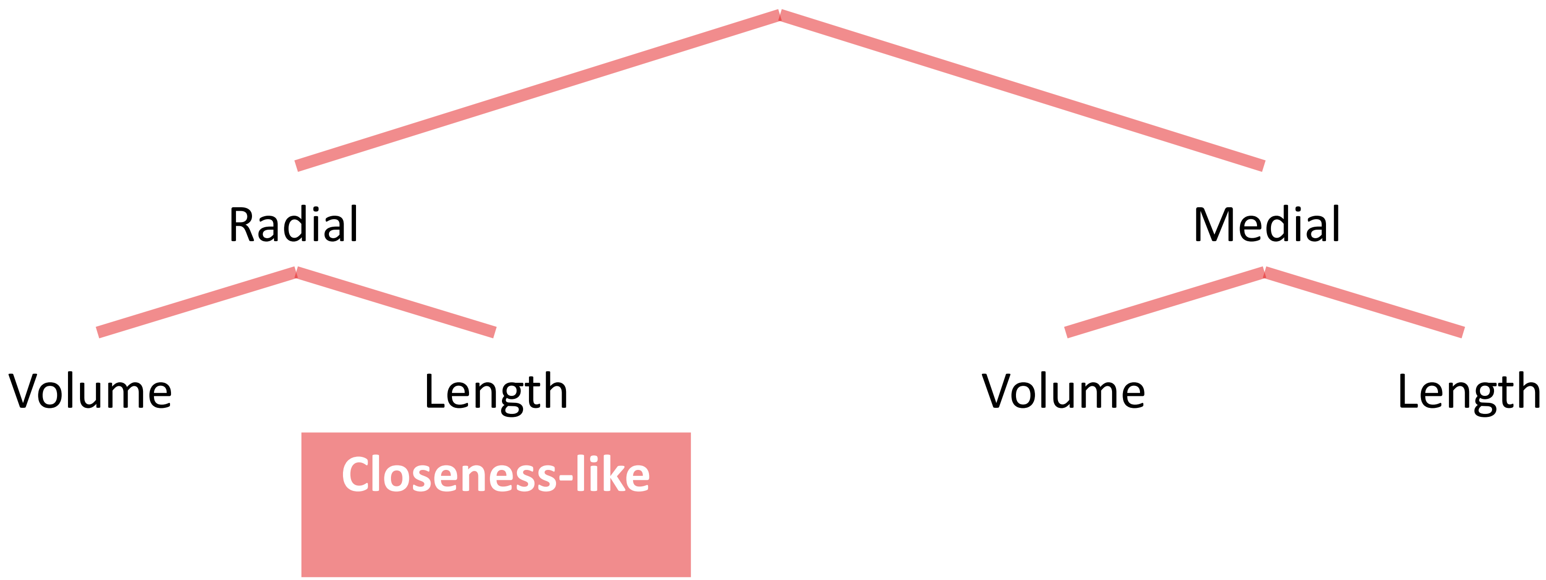


- Volume: Counting walks with a given constraint (**distance 1**)
- Radial: Only counting walks that **start or end** at node x
- $C_D(x) = \frac{deg(x)}{n-1}$
- Fraction of nodes connected to node x

[1] Zweig, K.A. 2016. *Network Analysis Literacy* (pp.261–263). Springer.

gesis Leibniz Institute for the Social Sciences

Closeness Centrality

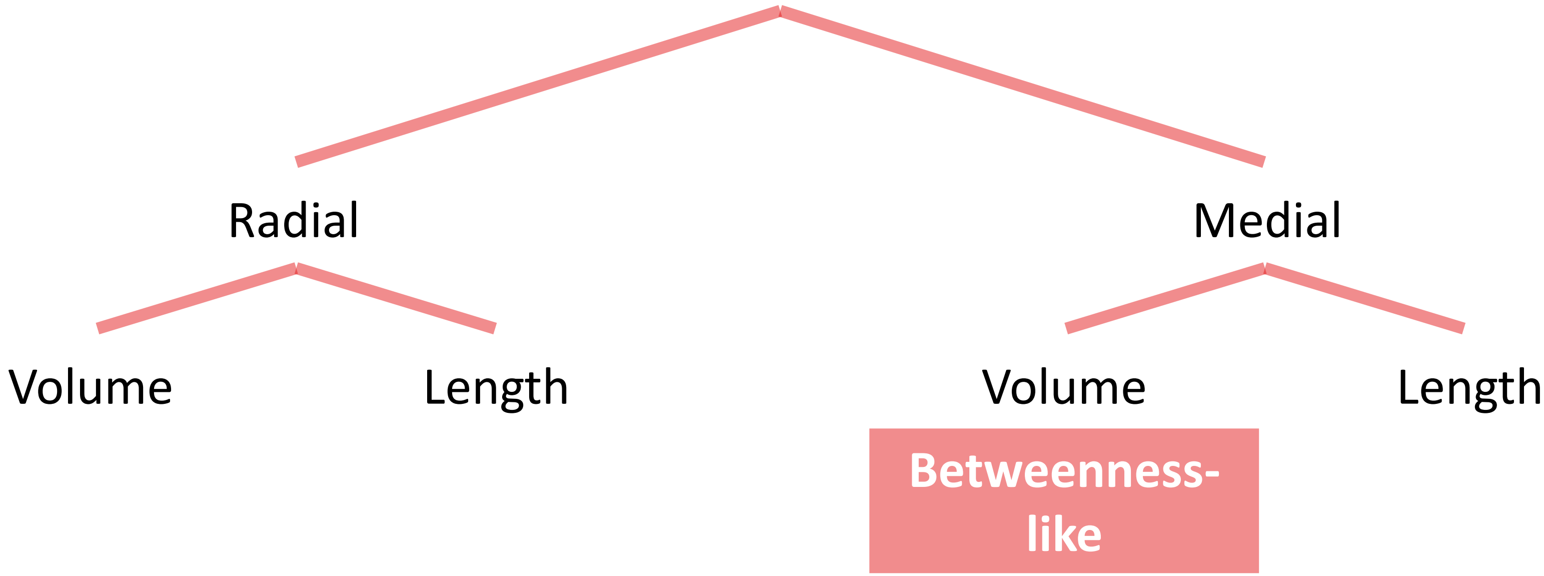


- Length: Counting walks and considering their **length**
- Radial: Only counting walks that **start or end** at node x
- $C_C(x) = 1/\sum_{v \in V(G)} d(x,v)$
- Inverse of the sum of a node x's distance d to other nodes

[1] Freeman, L.C. 1979. *Social Networks* 1:215-239.
[2] Zweig, K.A. 2016. *Network Analysis Literacy* (pp.261–263). Springer.

gesis Leibniz Institute for the Social Sciences

Betweenness Centrality



- Volume: Counting walks with a given constraint (**shortest walks**)
- Medial: Only counting walks that **pass through** node x
- $C_B(x) = \sum_{s \in V(G)} \sum_{t \in V(G)} \frac{\sigma_{st}(x)}{\sigma_{st}}$
- Sum of the fraction of all-pairs shortest paths that pass through x

[1] Freeman, L.C. 1977. *Sociometry* 40:35–41.
[2] Zweig, K.A. 2016. *Network Analysis Literacy* (pp.261–263). Springer.

Brokerage and Closure

- There is a **trade-off** between open and closed structures [1]
- Operationalization of open structure through the **effective size** [2]
  - $E(x) = \sum_{v \in N(x)} (1 - \sum_{w \in N(v)} p_{xw} m_{vw})$
    - $N(x)$: Set of x's neighbors
    - p_{xw}: normalized mutual weight of x's edges
    - m_{vw}: normalized mutual weight of alteri's edges
  - The extent to which a node x's **neighbors are not connected** among themselves (have information that is not redundant)

[1] Latora, V. *et al.* 2013. *Journal of Statistical Physics* 151: 745–764.
[2] Burt, R.S. 1995. *Structural Holes*. Harvard University Press.

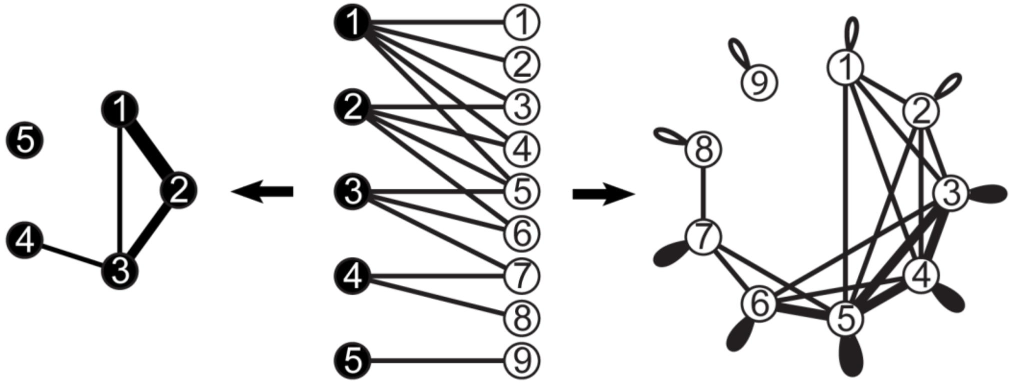
Brokerage and Closure

- There is a **trade-off** between open and closed structures [1]
- Operationalization of open structure through the **effective size** [2]
- Operationalization of closed structure through the **clustering coefficient** [3]
  - $CC(x) = {}^{m_x}/_{n_x(n_x-1)/2}$, for $n_x \geq 2$, otherwise 0
    - m_x: Number of edges in x's ego network
    - n_x: Number of x's neighbors
  - The extent to which a node x's **neighbors are connected** among themselves

[1] Latora, V. *et al.* 2013. *Journal of Statistical Physics* 151: 745–764.
[2] Burt, R.S. 1995. *Structural Holes*. Harvard University Press.
[3] Watts, D.J. & Strogatz, S.H. 1998. *Nature* 393:440–442.

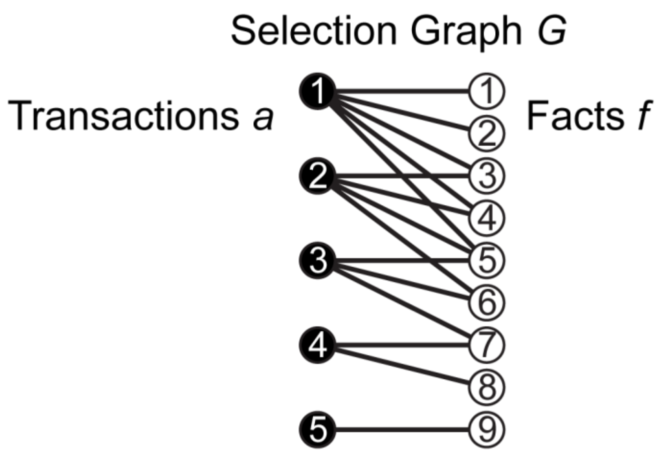
Graph Theoretical Notation

- If the nodes can be partitioned into two modes such that there are only inter-mode edges (no intra-mode edges), G is a **bipartite graph**
- A bipartite graph can be **projected** to two unipartite graphs



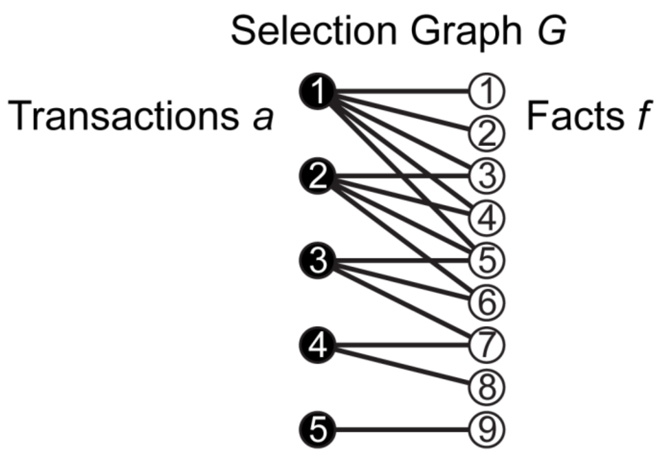
[1] Breiger, R.L. 1974. *Social Forces* 53: 181–190.

- To anchor data analysis in theory we use the terminology of the **identity model**



Sociological Terminology

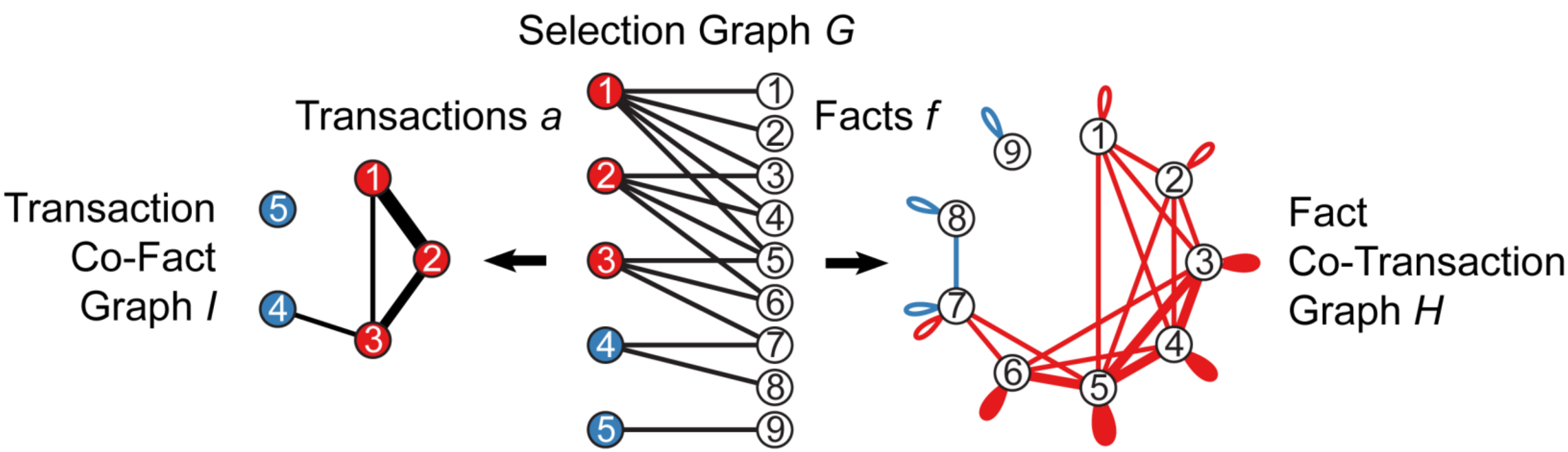
- A **transaction** is a communicative event among at least one sender and at least one receiver (think of scholarly publications or tweets)
- In a transaction, the sending side **selects**, and is influenced by, at least one fact



- **Facts** are either the receiving agents or cultural symbols

Duality of Node and Link Communities

- To facilitate multiplex analysis, we make use of the algebraic fact that **node communities** on one side of the selection graph resemble **link communities** on the other side



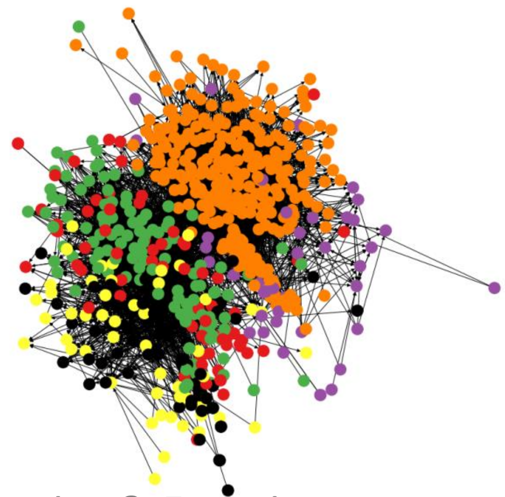
[1] Breiger, R.L. 1974. *Social Forces* 53: 181–190.
[2] Ahn, Y.-Y. et al. 2010. *Nature* 466: 761–765.

Force-Directed Layouting

- **Spring Embedding**: Algorithms for arranging nodes assuming forces among the sets of vertices; vertices **attract** each other if connected by edge, they **repell** each other if not connected
- **Fruchterman–Reingold** algorithm [1]: Equilibrium not guaranteed (max. number of iterations can be set by parameter); attractive force can be set

[1] Fruchterman, T.M.J. & Reingold, E.M. 1991. *Software – Practice & Experience* 21:1129–1164.

Force-Directed Layouting

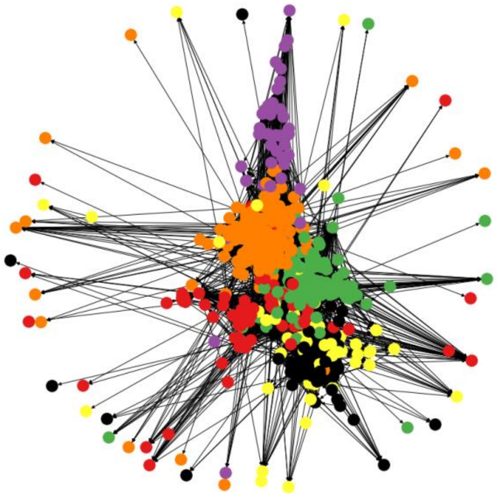


- **Spring Embedding**: Algorithms for arranging nodes assuming forces among the sets of vertices; vertices **attract** each other if connected by edge, they **repell** each other if not connected
- **Fruchterman–Reingold** algorithm [1]: Equilibrium not guaranteed (max. number of iterations can be set by parameter); attractive force can be set by parameter *k*
- **Kamada–Kawai** algorithm [2]: Faster; more prone to get stuck in local minimum

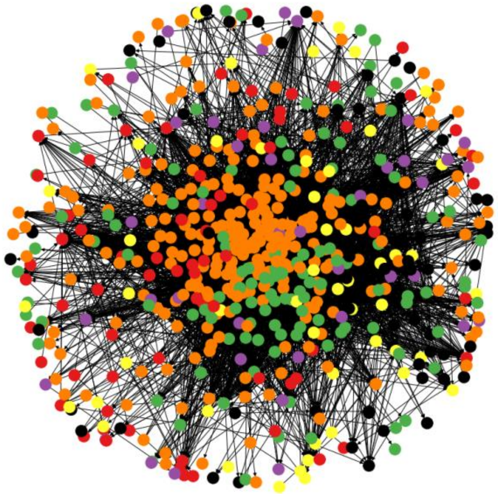
[1] Fruchterman, T.M.J. & Reingold, E.M. 1991. *Software – Practice & Experience* 21:1129–1164.
[2] Kamada, T. & Kawai, S. 1989. *Information Processing Letters* 31:7–15.

Fruchterman–Reingold Layouting

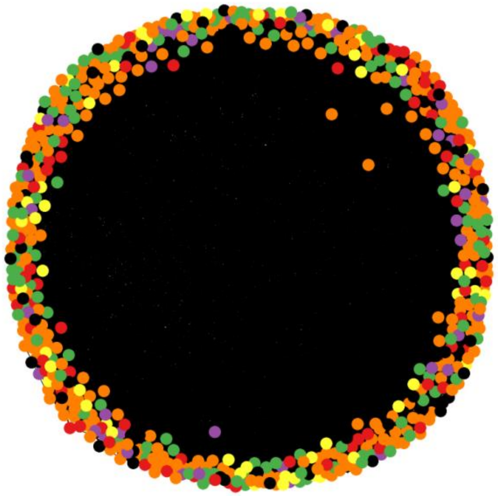
iterations=50 (default)



k=0.05 (default)



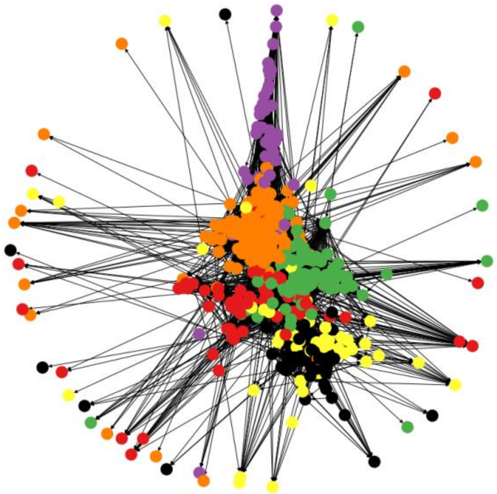
k=0.5



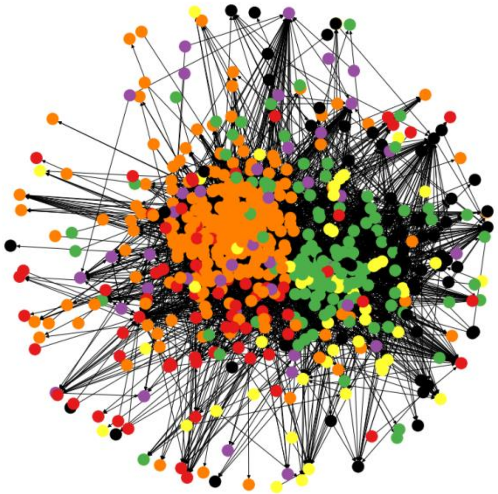
k=5

Fruchterman–Reingold Layouting

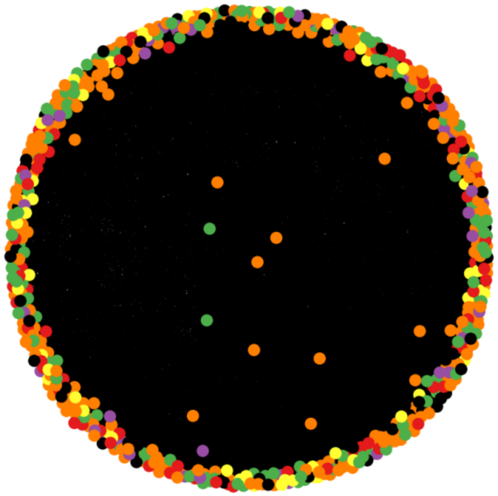
iterations=100



k=0.05 (default)



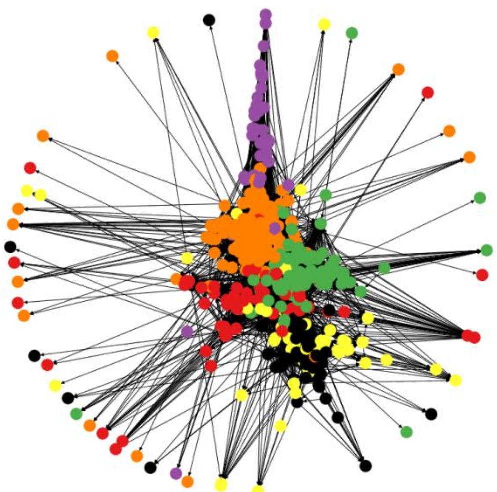
k=0.5



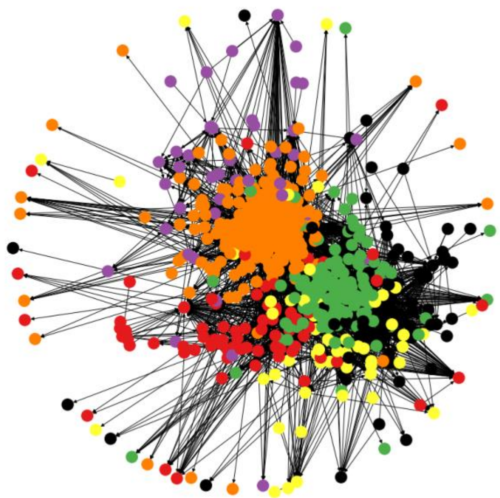
k=5

Fruchterman–Reingold Layouting

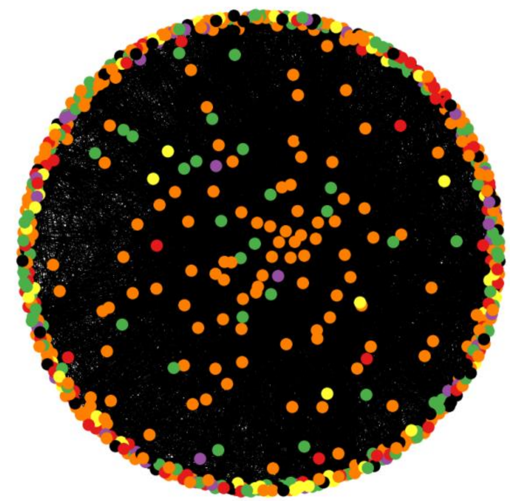
iterations=200



k=0.05 (default)



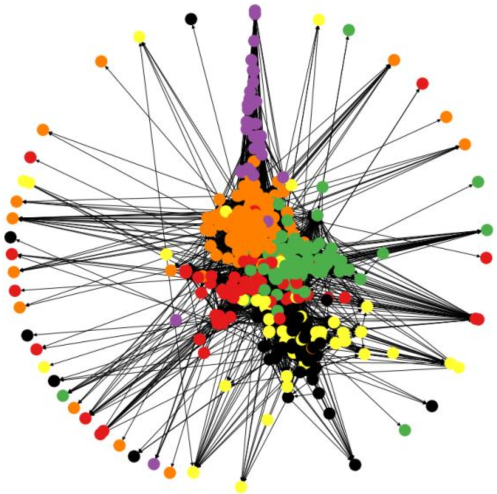
k=0.5



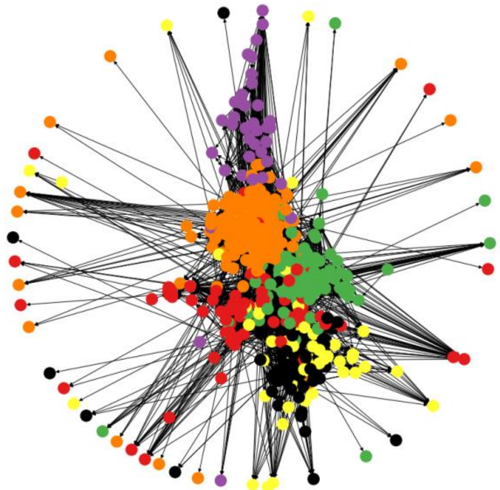
k=5

Fruchterman–Reingold Layouting

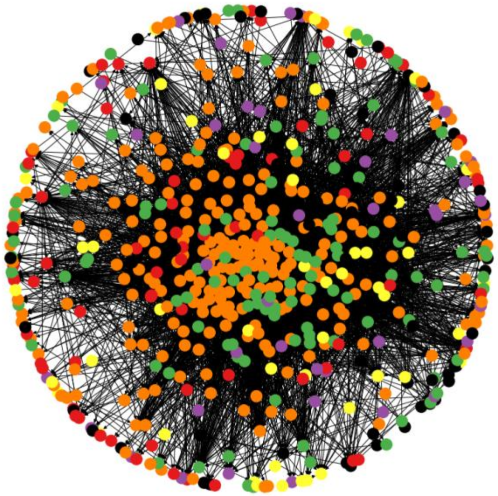
iterations=500



k=0.05 (default)



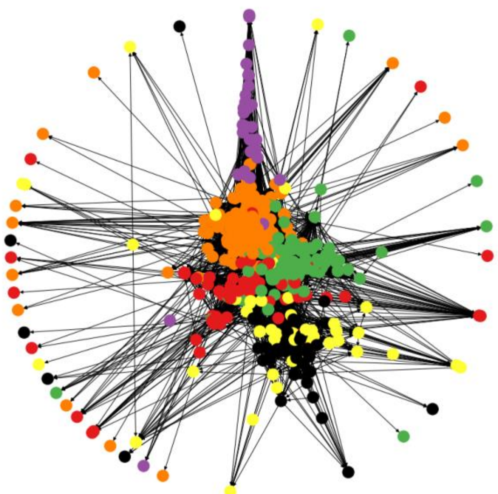
k=0.5



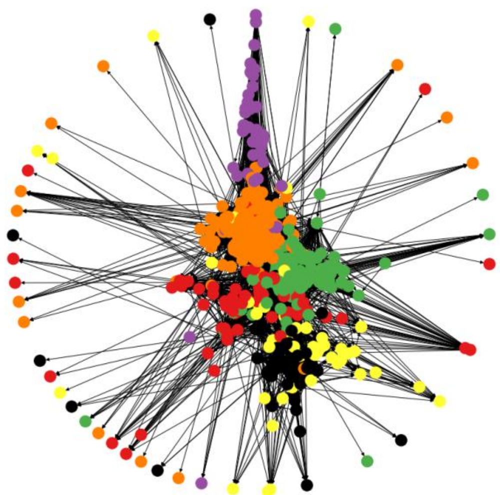
k=5

Fruchterman–Reingold Layouting

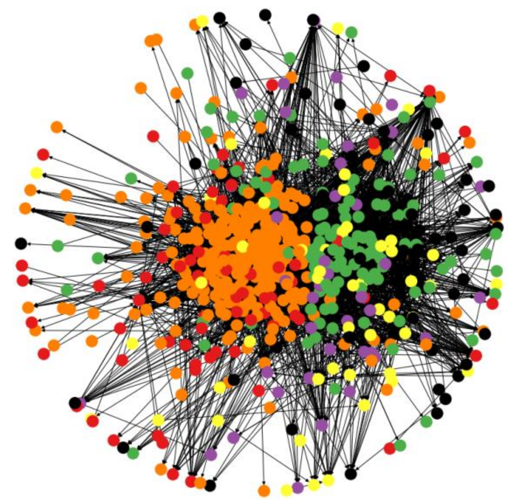
iterations=1000



k=0.05 (default)



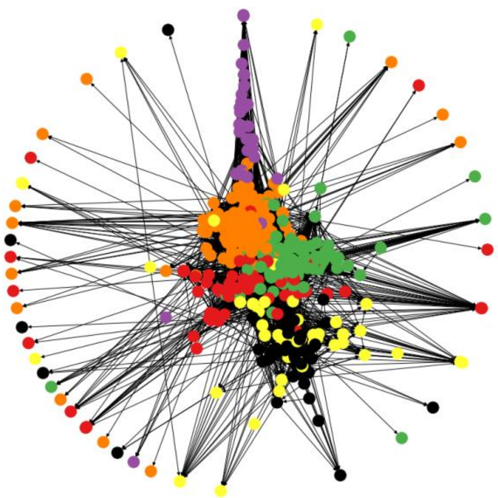
k=0.5



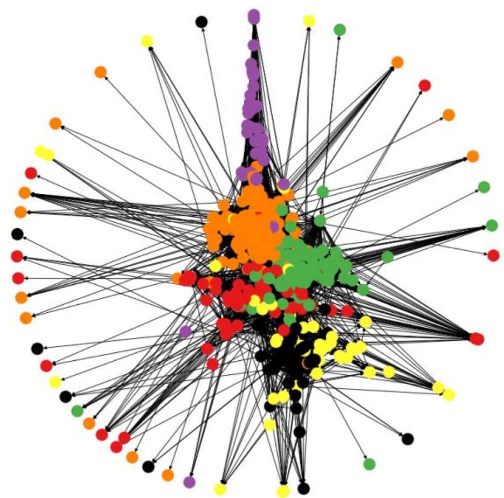
k=5

Fruchterman–Reingold Layouting

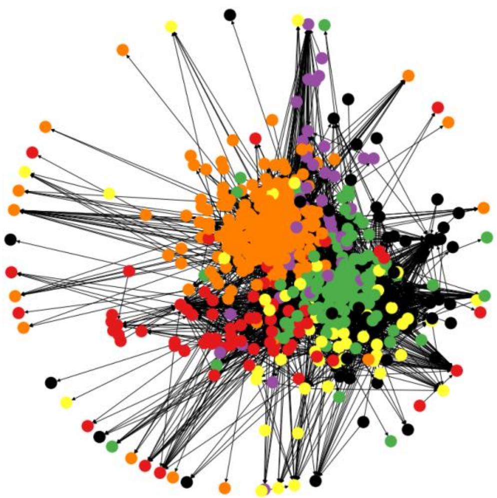
iterations=2000



k=0.05 (default)



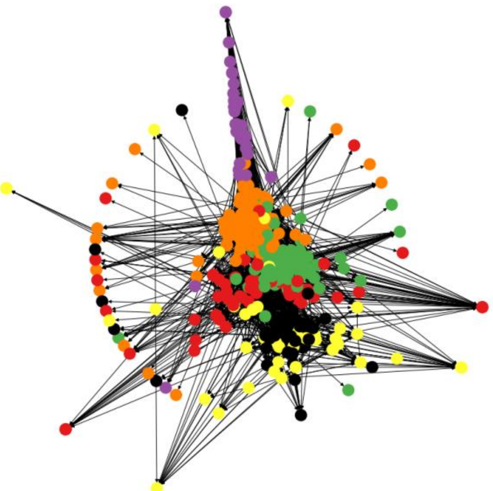
k=0.5



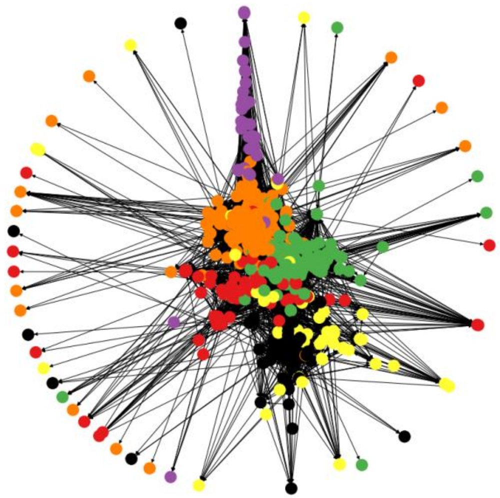
k=5

Fruchterman–Reingold Layouting

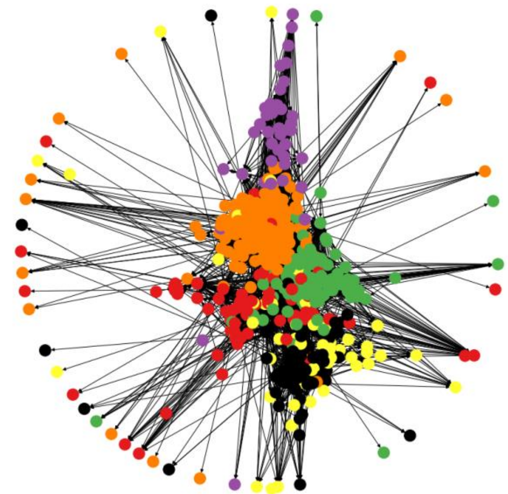
iterations=5000



k=0.05 (default)



k=0.5



k=5

Next: Repetition of Basic Python Concepts (Demo)

Leibniz Institute
for the Social Sciences

Leibniz
Association