# The Structural Dimension of Cooperation

## Cooperation Networks as Cohesive Small Worlds
## PhD Thesis Defense

Jordi Torrents

Department of Sociology
University of Barcelona

April 11, 2017

# Large Scale Cooperation

The last half of twentieth century has witnessed a key shift in knowledge intensive production processes.

**The rising importance of collective research and cooperation (Wuchty et al., 2007)**

- During most of the last century important papers or inventions were mostly developed by a single author
- Since 2000s top cited papers or inventions are developed by a team.

# Large Scale Cooperation

The last half of twentieth century has witnessed a key shift in knowledge intensive production processes.

## The rising importance of collective research and cooperation (Wuchty et al., 2007)

- During most of the last century important papers or inventions were mostly developed by a single author
- Since 2000s top cited papers or inventions are developed by a team.

## **Socialization** concept drawn from the Marxian tradition (Adler, 2007)

- In Political Science: transfer of ownership from the private to the public sphere.
- In Psychology: process whereby people new to a culture internalize its knowledge, norms and values.
- Marx's use was broader than either and encompasses both:

*The [...] socialized (i.e. collective) labour come into being through cooperation, division of labour within the workshop, the use of machinery, and in general, the transformation of production by the conscious use of the sciences, of mechanics, chemistry, etc. for specific ends, technology, etc. and similarly, through the enormous increase of scale corresponding to such developments (for it is only socialized labour that is capable of applying the general products of human development, such as mathematics, to the immediate process of production; [...]). (Marx, 1990, 1024)*

## Large Scale Cooperation

The last half of twentieth century has witnessed a key shift in knowledge intensive production processes.

**The rising importance of collective research and cooperation (Wuchty et al., 2007)**

- During most of the last century important papers or inventions were mostly developed by a single author
- Since 2000s top cited papers or inventions are developed by a team.

**Socialization concept drawn from the Marxian tradition (Adler, 2007)**

- In Political Science: transfer of ownership from the private to the public sphere.
- In Psychology: process whereby people new to a culture internalize its knowledge, norms and values.
- Marx's use was broader than either and encompasses both:

*The [...] socialized (i.e. collective) labour come into being through cooperation, division of labour within the workshop, the use of machinery, and in general, the transformation of production by the conscious use of the sciences, of mechanics, chemistry, etc. for specific ends, technology, etc. and similarly, through the enormous increase of scale corresponding to such developments (for it is only socialized labour that is capable of applying the general products of human development, such as mathematics, to the immediate process of production; [...]). (Marx, 1990, 1024)*

Wuchty, Jones, and Uzzi (2007) show that until 1950s the likelihood that an important —ie wildly cited— paper or invention was developed by a single author was bigger than it was developed by a team. But this trend has experimented a shift in the last four decades. The rising importance of collective research and cooperation is illustrated by the fact that top cited papers are mostly created by teams in 2000s.

# Theoretical Approaches to Cooperation

## Macro level approach

Cooperation as a macro level phenomenon in which the center of analysis is the collective or group (Marx, 1990; Adler and Heckscher, 2006; Adler, 2015).

- Focus on large organizations and groups: Collaborative Communities
  - Shared values and goals
  - Generalized trust
  - Authority forms

# Theoretical Approaches to Cooperation

## Macro level approach

Cooperation as a macro level phenomenon in which the center of analysis is the collective or group (Marx, 1990; Adler and Heckscher, 2006; Adler, 2015).

- Focus on large organizations and groups: Collaborative Communities
  - Shared values and goals
  - Generalized trust
  - Authority forms

## Micro level approach

Cooperation as a micro level phenomenon in which the center of analysis is the dyad (Axelrod and Hamilton, 1981; Watts, 1999; Eguíluz et al., 2005).

- Reductionist approach: Cooperation as an atomic process.
  - Strategic dyadic interactions.
  - Agent-based models.
  - Payoffs of different strategies.

# Theoretical Approaches to Cooperation

## Macro level approach

Cooperation as a macro level phenomenon in which the center of analysis is the collective or group (Marx, 1990; Adler and Heckscher, 2006; Adler, 2015).

- Focus on large organizations and groups: Collaborative Communities
  - Shared values and goals
  - Generalized trust
  - Authority forms

## Micro level approach

Cooperation as a micro level phenomenon in which the center of analysis is the dyad (Axelrod and Hamilton, 1981; Watts, 1999; Eguíluz et al., 2005).

- Reductionist approach: Cooperation as an atomic process.
  - Strategic dyadic interactions.
  - Agent-based models.
  - Payoffs of different strategies.

## A Meso level approach to Cooperation

Focus on Cooperation networks: patterns of relations that direct producers establish in the production process.

- Structural approach, that is, a network approach.
  - Sub-groups that are more connected internally than with the rest of the network.
  - Longitudinal analysis of the formation and dissolution of these groups.
  - Key mechanisms to explain and understand large scale cooperation.

# Main Topic and Research Question

The central topic of this thesis is to understand and explain under which conditions and through which social mechanisms large scale cooperation operates.

## Free and Open Source Software as a case study

**Free Software** is computer software that allows users to run, copy, distribute, study, change and improve it.

- Availability of detailed data about the production process.
- Impressive increment of scale in the last two decades.
- Allows to analyze the development methods of FOSS from the perspective of a knowledge based production process.
- focusing on the patterns of relations between direct producers we can analyze key mechanisms that enable and foster large scale cooperation.

## Research question

How does large scale cooperation work in knowledge intensive and technically complex production processes developed in new organizational environments, such as Free and Open Source Software projects, where loosely coupled individuals that rarely meet face to face have to coordinate through internet in order to produce world class software products, without relying on hierarchy or market as coordinating mechanisms.

# Group Cohesion in the sociological literature

Central concept that has a long and illustrious history in sociology. Its use in most sociological research has been ambiguous at best (Moody and White, 2003):

1. sloppy definitions of cohesion with lack of generality.
2. grounded mostly in intuition and common sense.

**Group cohesion (Doreian and Fararo, 1998) can be divided analytically into:**

Ideational component
: based on the members' identification with a collectivity.

Relational component
: based on the patterns of connections among members.

# Group Cohesion in the sociological literature

Central concept that has a long and illustrious history in sociology. Its use in most sociological research has been ambiguous at best (Moody and White, 2003):

1. sloppy definitions of cohesion with lack of generality.
2. grounded mostly in intuition and common sense.

**Group cohesion (Doreian and Fararo, 1998) can be divided analytically into:**

- Ideational component: based on the members' identification with a collectivity.
- Relational component: based on the patterns of connections among members.

The relational component of group cohesion has been the focus of Social Network Analysis.

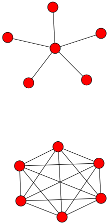
**Network theory measures used to define group cohesion**

- Classical measures: cliques, clans, clubs, *k*-cores, lambda sets, ...
- Community algorithms: detect groups of nodes more densely connected among them than with the rest of the network.

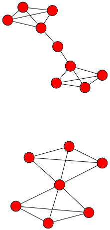
Neither the classical approaches nor new developments in community analysis work well in empirical analysis of group cohesion.

# Key properties that a cohesion measures should have

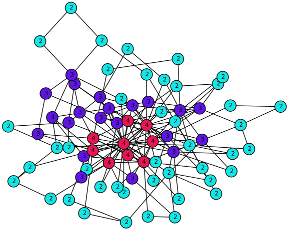
**Robustness** its qualification as a group should not be dependent on the actions of a single individual.



**Overlap** a cohesion measure should allow some actors to be part of more than one cohesive subgroup.



**Hierarchical** highly cohesive subgroups are nested inside less cohesive ones.



# The structural cohesion model

The structural cohesion model (White and Harary, 2001; Moody and White, 2003) is based on two mathematically equivalent definitions of cohesion.

**Precise definition of group cohesion based on node connectivity**

- a group's structural cohesion is equal to the minimum number of actors who, if removed from the group, would disconnect the group.
- a group's structural cohesion is equal to the minimum number of node independent paths linking each pair of actors in the group.

This equivalence relation has a deep sociological meaning because it allows to define structural cohesion in terms of:

- the difficulty to pull a group apart by removing actors.
- multiple relations between actors that keep a group together.

# Measures of structural cohesion

## Node connectivity $\kappa(G)$

**Local** given two nodes $u$ and $v$, $\kappa_G(u, v)$ is the minimum number of nodes that must be removed to destroy all paths that join $u$ and $v$.

**Global** the minimum number of nodes that must be removed in order to disconnect a graph $G$.

$$\kappa = min\{\kappa_G(u, v) : u, v \in V(G)\}$$

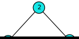
## Average node connectivity $\bar{\kappa}(G)$

the sum of local node connectivity between all pairs of different nodes of $G$ divided by the number of distinct pairs of nodes. (Beineke, Oellermann, and Pippert, 2002).

$$\bar{\kappa}(G) = \frac{\sum_{u,v} \kappa_G(u, v)}{\binom{n}{2}}$$

## $k$-components

A $k$-component is a maximal subgraph that has, at least, node connectivity $k$: we need to remove at least $k$ nodes to break it into more components.



# Structural cohesion in empirical analysis

Structural cohesion is a powerful and robust explanatory factor for a wide variety of interesting empirical social phenomena:

- likelihood of building alliances and partnerships among biotech firms (Powell et al., 2005)
- school attachment and academic performance of young people and the tendency of firms to enroll in similar political activity behaviors (Moody and White, 2003)
- emerging trust relations among neighborhood residents and the hiring relations among top level US graduate programs (Grannis, 2009)

## Some problems with the structural cohesion model

Despite all its merits, the structural cohesion model has not been widely applied to empirical analysis because:

- it is not practical to compute it for networks with more than a few hundreds nodes and edges due to its computational complexity.
- it is not implemented in most popular network analysis software packages.

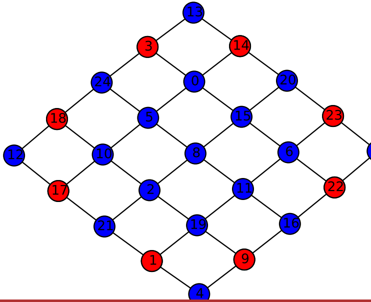
# Exact Algorithm

Moody and White (2003, appendix A) provide an algorithm for identifying $k$-components in a network, which is based on the Kanevsky (1993) algorithm for finding all minimum-size node cut-sets of a graph.

1. Identify the node connectivity, $k$, of the input graph.
2. Identify all $k$-cutsets at the current level of connectivity.
3. Generate new graph components based on the removal of these cutsets.
4. If the graph is neither complete nor trivial, return to 1; otherwise end.

```
import networkx as nx
from networkx import convert_node_labels_to_inte

G = cnlti(nx.grid_2d_graph(5,5))
list(nx.all_node_cuts(G))
[{1, 9}, {22, 23}, {17, 18}, {3, 14}]
# To compute k-components
kcomponents = nx.k_components(G)
```



# Heuristics

The heuristics we developed are based on:

## Approximation to local node connectivity

White and Newman (2001) fast approximation algorithm for finding good lower bounds of the number of node independent paths between two nodes.

```
import networkx as nx
from networkx.algorithms import approximation as apxa
# Exact computation of node connectivity
exact = nx.node_connectivity(G)
# Lower bound approximation
aprox = apxa.node_connectivity(G)
```

## Whitney's theorem

inclusion relation among node connectivity $\kappa(G)$, edge connectivity $\lambda(G)$ and minimum degree $\delta(G)$ for any graph $G$:

$$\kappa(G) \leq \lambda(G) \leq \delta(G)$$

"*k*-cores can be regarded as seedbeds, within which we can expect highly cohesive subsets to be found" Seidman (1983, 281)

# Heuristics

## Main logic of the heuristics

1. repeatedly applying fast algorithms for $k$-cores (Batagelj and Zaveršnik, 2011) and biconnected components (Tarjan, 1972) in order to narrow down the number of pairs of different nodes over which we have to compute their local node connectivity.
2. Build an auxiliary graph $H$ in which two nodes are linked if they have at least $k$ node independent paths connecting them.
3. Complete subgraphs in $H$ have a one to one correspondence with subgraphs of $G$ in which each node is connected to every other node in the subgraph for at least $k$ node independent paths. That is, $k$-components.

## Trade Accuracy for Speed

| | Bipartite | | | | Unipartite | | | |
|---|---|---|---|---|---|---|---|---|
| Network | # nodes | # edges | Av. degree | Time(s) | # nodes | # edges | Av. degree | Time(s) |
| Debian Lenny | 13,121 | 20,220 | 3.08 | 1,105.2 | 1,383 | 5,216 | 7.54 | 204.7 |
| High Energy (theory) | 26,590 | 37,566 | 2.81 | 3,105.7 | 9,767 | 19,331 | 3.97 | 7,136.0 |
| Nuclear Theory | 10,371 | 15,969 | 3.08 | 1,205.2 | 4,827 | 14,488 | 6.00 | 3,934.1 |

# Collaborative Communities

A new form of community, qualitatively different from the traditional *Gemeinschaft* and the modern *Gesellschaft* (Tönnies, 1974).

### Collaborative Communities (Adler and Heckscher, 2006)

Novel organizational form —both inside and outside large capitalist corporations— strongly grounded on large scale cooperation which defy the traditional dichotomy between **hierarchy** and **market** as coordinating mechanisms.

- Generalized trust based on other's contributions towards a shared end.
- Conscious cooperation and high individual interdependence.
- Shared values and value-rational basis for legitimate authority.

So far scarce attention is given to the structural features of collaborative communities.

**Question** explore the network structure that lead to their emergence and effectiveness in the production and diffusion of knowledge.

**Contribution** I suggest that a unique network structure undergirds collaborative communities and build a model to understand its key mechanisms: the **cohesive small world** model.

# knowledge-based production processes

Collaborative Communities are especially well suited to deal with the challenges of knowledge-based production processes because, hierarchy and market have proved ineffective, at best, at managing knowledge.

### Hierarchy as a coordinating mechanism

- Knowledge is treated as a scare resource.
- Centralized at the higher levels where key decisions are taken.
- Rigidity prevents flexibility to deal with unanticipated problems.
- Difficulty to foster innovation and generation of new knowledge.

### Market as coordinating mechanism

- Fails to optimize at the same time production and allocation of knowledge.
- Knowledge is a public good that grow rather than diminish with use.
- Strong property rights incentive knowledge production, but
- they block socially optimal allocation (free access).

### Collaborative Communities that excel at knowledge-based production

- Free and Open Source Communities.
- Novel forms of professional work organization (Adler, Kwon, and Heckscher, 2008).
- Knowledge-intensive production processes in corporations (Adler and Heckscher, 2006).

# A network model for Collaborative Communities

## The Small World Model

Networks characterized by a high level of local density of social ties and short average distances among nodes.

- $L$: average number of intermediaries between any two nodes.
- $CC$: mean probability that two nodes that are neighbors of the same other node will themselves be neighbors.
- They foster the flow of information and ideas.

## The Structural Cohesion Model

Networks characterized by the presence of increasingly cohesive groups nested inside each other.

- Cohesion definition based on the graph-theoretic property of connectivity.
- $k$-components as the key subgroups of the network.
- They foster the development of trust and social cohesion.

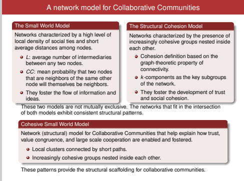
These two models are not mutually exclusive. The networks that fit in the intersection of both models exhibit consistent structural patterns.

## Cohesive Small World Model

Network (structural) model for Collaborative Communities that help explain how trust, value congruence, and large scale cooperation are enabled and fostered.

- Local clusters connected by short paths.
- Increasingly cohesive groups nested inside each other.

These patterns provide the structural scaffolding for collaborative communities.




On the one hand, the generation of trust and congruent values among heterogeneous individuals are fostered by structurally cohesive groups in the connectivity hierarchy of cooperation networks because individuals embedded in these structures are able to compare independent perspectives on each other through a variety of relations that flow through distinct sets of intermediaries, which provides multiple independent sources of information about each other. Thus, the perception of an individual embedded in such structures of the other members of the group to whom she is not directly linked is filtered by the perception of a variety of others whom she trusts because is directly linked to them. This mediated perception of the group generates trust at a global scale. On the other hand, the existence of dense local clusters connected between them by relative short paths allows successful cooperation among heterogeneous individuals with common interests and, at the same time, fosters the flow of information between these clusters preventing the local clusters to be trapped in echo chambers of like minded collaborators.

## Pure Structural Cohesion

- Robust to node removal
- Not necessary short average distance
- Not necessary local clustering

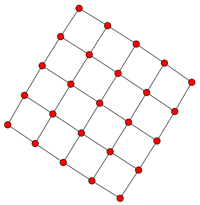
## Cohesive Small World

- Intersection of the two models
- Short average distance
- High local cluster coefficient
- Robust to node removal

## Pure Small World

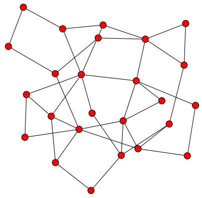
- Short average distance
- High local cluster coefficient
- Not necessary robust to node removal

Pure Structural Cohesion



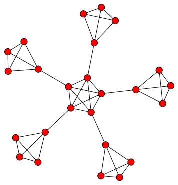
n=25, m=40, CC=0.00, APL=3.33, SWI=0.00, nodes in GBC=100%

Cohesive Small World



n=25, m=40, CC=0.19, APL=2.67, SWI=1.57, nodes in GBC=100%

Pure Small World



n=25, m=45, CC=0.82, APL=3.38, SWI=4.23, nodes in GBC=20%

# Free and Open Source Projects: Debian and Python

Free Software, broadly defined, is computer software that allows users to run, copy, distribute, study, change and improve it.

**The Debian Project: 1999-2012**

- A free Operating System
- 392 developers in 1999, 1435 in 2012
- 2876 programs in 1999, 10469 in 2012
- Widely used in servers (google), desktops (Ubuntu) and embedded devices (raspberry pi)

**The Python project: 1999-2014**

- A free Programming Language
- 9 developers in 1999, 62 in 2014
- 1137 files in 1999, 2134 in 2014
- Widely used in web development (reddit, youtube) and scientific computing

# Cooperation Networks and Null Models

My modeling strategy to capture the patterns of relations among developers in these two projects is to focus on the actual contributions of each developer to the project.

## Building Cooperation Networks

- Focus on the patterns of relations among developers in the productive process.
- Cooperation networks are bipartite or two mode; node sets are developers and programs/files and edges only link nodes from opposite sets.
- A developer is linked to the package (in Debian) or source code file (in Python) that she works on.
- We have complete electronic records of all package uplads to Debain and all sorce code file edits in Python.

## Building Suitable Null models

- We need to compare the statistical measures obtained from the actual networks with a suitable null model in order to assert that what we observe is not the result of pure chance.

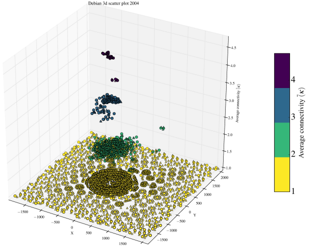
Configuration Model: assigns at random developers to packages, or developers to source code files, maintaining the concrete skewed distribution of packages by developer and files by developer observed in the actual networks. That is, the degree distribution.

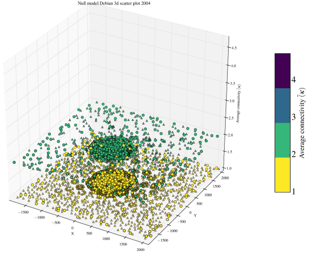
A network fits the small world model if it is more locally clustered ($CC$) than its random network counterpart but has approximately the same average distance ($L$) between nodes.

$$Q = \frac{CC_{ratio}}{L_{ratio}} \qquad \textit{Where:} \qquad CC_{ratio} = \frac{CC_{actual}}{CC_{random}} \qquad L_{ratio} = \frac{L_{actual}}{L_{random}} \tag{1}$$

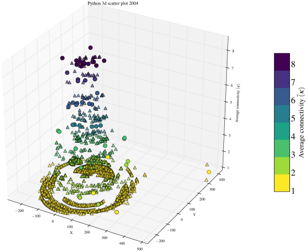
# Structural Cohesion Analysis

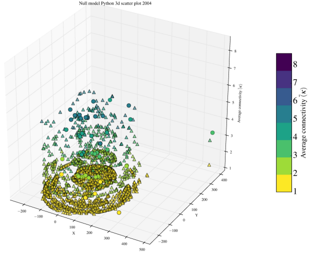
## Debian Cooperation Network 2004




## Python Cooperation Network 2004

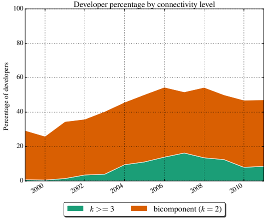


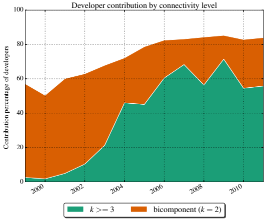



# Iron law of the Oligarchy or an Open Elite?

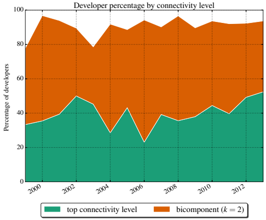
The literature on FOSS, especially from software engineering and computer science, have stressed that only a small fraction of the developers is doing most of the work.

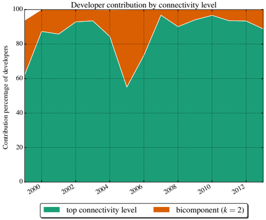
## Debian Contributions



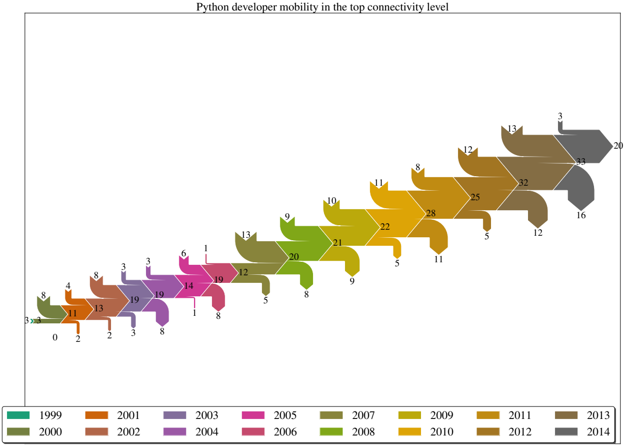




## Python Contributions

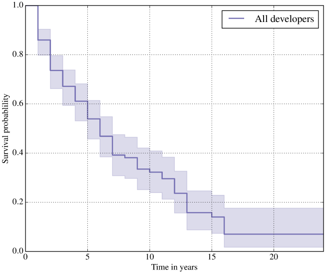


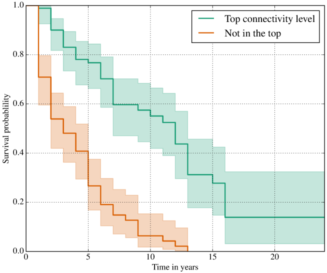
Sankey diagram of Python developer mobility in the top connectivity level.



# Modeling robustness as median active live of individuals in the project




(a) Survival Function for all developers




(b) Survival Function for developers in the top connectivity level

Figure: Estimation of the survival function using the Kaplan-Meier estimate. The median survival time of a developer in the community, defined as the point in time where on average half of the population has abandoned the community, is 6 years if I consider all developers (left). But if I consider separately the developers in the top level of the connectivity hierarchy (right), their median survival time is 12 years; but only 3 years for the developers that are not on the top of the connectivity hierarchy.

Adler, P. (2007). The future of critical management studies: A paleo-marxist critique of labour process theory. *Organization Studies 28*(9), 1313.

Adler, P. and C. Heckscher (2006). *The Firm as a Collaborative Community: Reconstructing Trust in the Knowledge Economy*. Oxford University Press, Oxford, UK.

Adler, P., S. Kwon, and C. Heckscher (2008). Professional work: The emergence of collaborative community. *Organization Science 19*(2).

Adler, P. S. (2015). Community and innovation: from tönnies to marx. *Organization Studies 36*(4), 445–471.

Axelrod, R. and W. Hamilton (1981). The evolution of cooperation. *Science 211*(4489), 1390–1396.

Batagelj, V. and M. Zaveršnik (2011). Fast algorithms for determining (generalized) core groups in social networks. *Advances in Data Analysis and Classification 5*(2), 129–145.

Beineke, L., O. Oellermann, and R. Pippert (2002). The average connectivity of a graph. *Discrete mathematics 252*(1-3), 31–45.

Doreian, P. and T. Fararo (1998). *The problem of solidarity: theories and models*. Routledge.

Eguíluz, V., M. Zimmermann, C. Cela-Conde, and M. Miguel (2005). Cooperation and the Emergence of Role Differentiation in the Dynamics of Social Networks 1. *American journal of sociology 110*(4), 977–1008.

Grannis, R. (2009). Paths and semipaths: reconceptualizing structural cohesion in terms of directed relations. *Sociological Methodology 39*(1), 117–150.

Kanevsky, A. (1993). Finding all minimum-size separating vertex sets in a graph. *Networks 23*(6), 533–541.

Marx, K. (1990). *Capital: a critique of political economy. Volume 1*. Penguin Books in association with New Left Review.

Moody, J. and D. White (2003). Social cohesion and embeddedness: A hierarchical conception of social groups. *American Sociological Review 68*(1), 103–28.

Powell, W., D. White, K. Koput, and J. Owen-Smith (2005). Network dynamics and field evolution: The growth of interorganizational collaboration in the life sciences. *American journal of sociology 110*(4), 1132–1205.

Seidman, S. (1983). Network structure and minimum degree. *Social networks 5*(3), 269–287.

Tarjan, R. (1972). Depth-first search and linear graph algorithms. In *Switching and Automata Theory, 1971., 12th Annual Symposium on*, pp. 114–121. IEEE.

Tönnies, F. (1974). *Community and association:(Gemeinschaft und Gesellschaft)*. Taylor & Francis.

Watts, D. (1999). *Small Worlds: The Dynamics of Networks Between Order and Randomness*. Princeton University Press.

White, D. and F. Harary (2001). The cohesiveness of blocks in social networks: Node connectivity and conditional density. *Sociological Methodology*, 305–359.

White, D. and M. Newman (2001). Fast approximation algorithms for finding node-independent paths in networks. *Santa Fe Institute Working Papers Series*.