Emergence of scaling modular structures through random walks rewiring

Bingsheng Chen, Qing Yao, Tim Evans and Kim Christensen

Centre for Complexity and Network Science
Department of Physics
Imperial College London

Dynamics of networks Workshop
@ MoN18
8th Apr 2019

Outline

Motivation

Network is not static but dynamical.

Node will have many different types of labels. Different types of nodes behave differently.

Network has many small world effect, for example, large clustering coefficient, scaling of communities.

Examples: *Shareholder Network*

Shareholder Network Introduction

‘The divorce of ownership from the control of modern corporation has created the ‘quasi-public’ corporation.’
— Berle and Means [5]

Networks in an economics context and complex system have proved useful [2, 11, 1]

In our work, we use complex network methods to study the investment characteristics of different types of shareholders.

Data Source

Data from BvD — Bureau van Dijk
`https://www.bvdinfo.com/en-gb`

> *Amadeus contains comprehensive information on around 21 million companies across Europe. You can use it to research individual companies, search for companies with specific profiles and for analysis.*

Snapshot of the data:
Company name, shareholder's name and shareholder type

	Company name	BvD ID number	Shareholder - BvD ID number	Shareholder - Name	Shareholder - Type	Shareholder - Direct % 12/2014
1.	ÇETINKAYA GIDA PAZARLAMA VE TICARET LIMITED SIRKETI	TR150991F	WW*110007271024	MR AHMET CETINKAYA	One or more named individuals or families	30.00
	ÇETINKAYA GIDA PAZARLAMA VE TICARET LIMITED SIRKETI	TR150991F	WW*110007271022	MR ISMET CETINKAYA	One or more named individuals or families	11.25
	ÇETINKAYA GIDA PAZARLAMA VE TICARET LIMITED SIRKETI	TR150991F	WW*110007271023	MRS GIZEM CETINKAYA	One or more named individuals or families	11.25
	ÇETINKAYA GIDA PAZARLAMA VE TICARET LIMITED SIRKETI	TR150991F	WW*110007271025	MRS AYTEN CETINKAYA	One or more named individuals or families	7.50
2.	ÇETINKAYA KARDESLER OTOMOTIV GIDA NAKLIYE VE TICARET LIMITED SIRKETI	TR13846F				
3.	CETINKAYA KURDELE TEKSTIL URUNLERI URETIM PAZ SAN VE TIC LTD STI	TR253081F	WW*110007551231	MR GOKHAN CETINKAYA	One or more named individuals or families	40.00
	CETINKAYA KURDELE TEKSTIL URUNLERI URETIM PAZ SAN VE TIC LTD STI	TR253081F	WW*110007551233	MR METIN CETINKAYA	One or more named individuals or families	40.00
	CETINKAYA KURDELE TEKSTIL URUNLERI URETIM PAZ SAN VE TIC LTD STI	TR253081F	WW*110007551232	MR NURETTIN CETINKAYA	One or more named individuals or families	20.00

Data Capture

- BvD data extremely expensive to buy
- Used college licence which allows limited numbers of downloads
- Data downloaded in small pieces then joined together
- Focus initially on small countries: Turkey, The Netherlands
  Chosen because of size yet difference in their context
  Networks for larger countries now being constructed

	Company name	BvD ID number	Shareholder - BvD ID number	Shareholder - Name	Shareholder - Type	Shareholder - Direct % 12/2014
1.	ÇETINKAYA GIDA PAZARLAMA VE TICARET LIMITED SIRKETI	TR150991F	WW*110007271024	MR AHMET CETINKAYA	One or more named individuals or families	30.00
	ÇETINKAYA GIDA PAZARLAMA VE TICARET LIMITED SIRKETI	TR150991F	WW*110007271022	MR ISMET CETINKAYA	One or more named individuals or families	11.25
	ÇETINKAYA GIDA PAZARLAMA VE TICARET LIMITED SIRKETI	TR150991F	WW*110007271023	MRS GIZEM CETINKAYA	One or more named individuals or families	11.25
	ÇETINKAYA GIDA PAZARLAMA VE TICARET LIMITED SIRKETI	TR150991F	WW*110007271025	MRS AYTEN CETINKAYA	One or more named individuals or families	7.50
2.	ÇETINKAYA KARDESLER OTOMOTIV GIDA NAKLIYE VE TICARET LIMITED SIRKETI	TR13846F				
3.	CETINKAYA KURDELE TEKSTIL URUNLERI URETIM PAZ SAN VE TIC LTD STI	TR253081F	WW*110007551231	MR GOKHAN CETINKAYA	One or more named individuals or families	40.00
	CETINKAYA KURDELE TEKSTIL URUNLERI URETIM PAZ SAN VE TIC LTD STI	TR253081F	WW*110007551233	MR METIN CETINKAYA	One or more named individuals or families	40.00
	CETINKAYA KURDELE TEKSTIL URUNLERI URETIM PAZ SAN VE TIC LTD STI	TR253081F	WW*110007551232	MR NURETTIN CETINKAYA	One or more named individuals or families	20.00

Data Structure

BvD (Bureau van Dijk) lists changes in share structure.

For each change we have

- Shareholder making change
- Company in which shareholding is changed
  — Companies can also be shareholders
- Time change noted in database
  — may not be time at which transaction occurred

	Company name	BvD ID number	Shareholder - BvD ID number	Shareholder - Name	Shareholder - Type	Shareholder - Direct % 12/2014
1.	ÇETINKAYA GIDA PAZARLAMA VE TICARET LIMITED SIRKETI	TR150991F	WW*110007271024	MR AHMET CETINKAYA	One or more named individuals or families	30.00
	ÇETINKAYA GIDA PAZARLAMA VE TICARET LIMITED SIRKETI	TR150991F	WW*110007271022	MR ISMET CETINKAYA	One or more named individuals or families	11.25
	ÇETINKAYA GIDA PAZARLAMA VE TICARET LIMITED SIRKETI	TR150991F	WW*110007271023	MRS GIZEM CETINKAYA	One or more named individuals or families	11.25
	ÇETINKAYA GIDA PAZARLAMA VE TICARET LIMITED SIRKETI	TR150991F	WW*110007271025	MRS AYTEN CETINKAYA	One or more named individuals or families	7.50
2.	ÇETINKAYA KARDESLER OTOMOTIV GIDA NAKLIYE VE TICARET LIMITED SIRKETI	TR13846F				
3.	CETINKAYA KURDELE TEKSTIL URUNLERI URETIM PAZ SAN VE TIC LTD STI	TR253081F	WW*110007551231	MR GOKHAN CETINKAYA	One or more named individuals or families	40.00
	CETINKAYA KURDELE TEKSTIL URUNLERI URETIM PAZ SAN VE TIC LTD STI	TR253081F	WW*110007551233	MR METIN CETINKAYA	One or more named individuals or families	40.00
	CETINKAYA KURDELE TEKSTIL URUNLERI URETIM PAZ SAN VE TIC LTD STI	TR253081F	WW*110007551232	MR NURETTIN CETINKAYA	One or more named individuals or families	20.00

Projected Networks

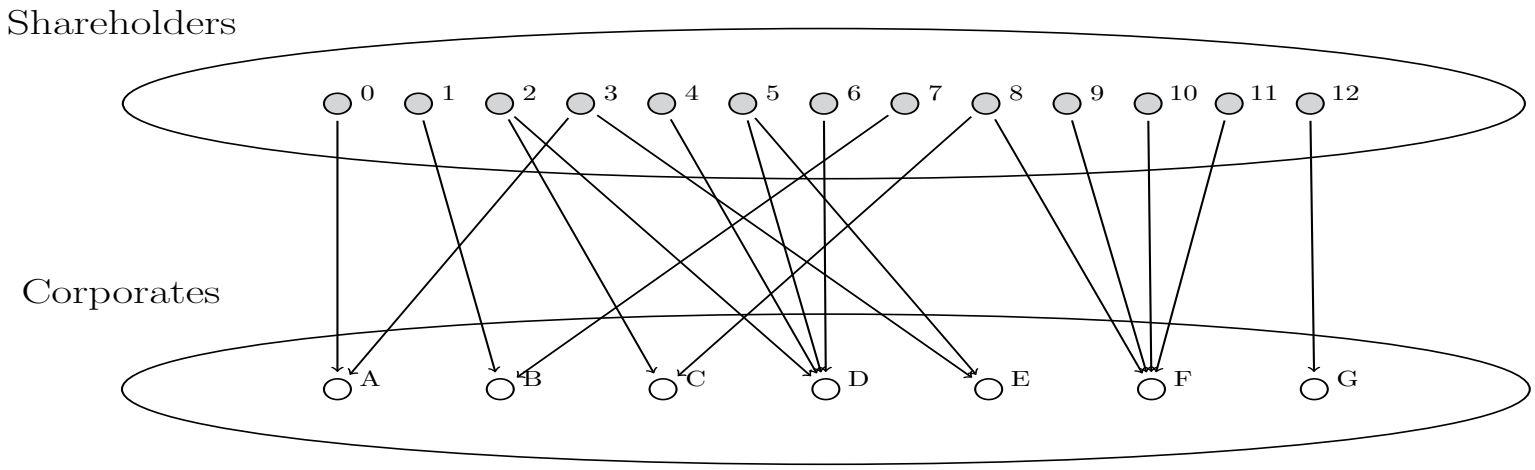
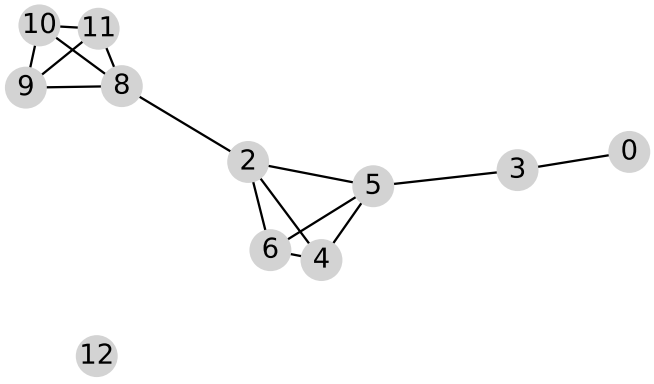


Figure 1: Bipartite corporate network



The projection of the network above onto a network of owners

Degree Distributions

$$P(k) = \frac{N(k)}{N}, \tag{1}$$

The powerlaw asymptotics $P(k) \sim k^{-\gamma}$ is associated with a *scale-free networks*, where γ denotes the slope of a linear fit for binned data in a log-log plot.

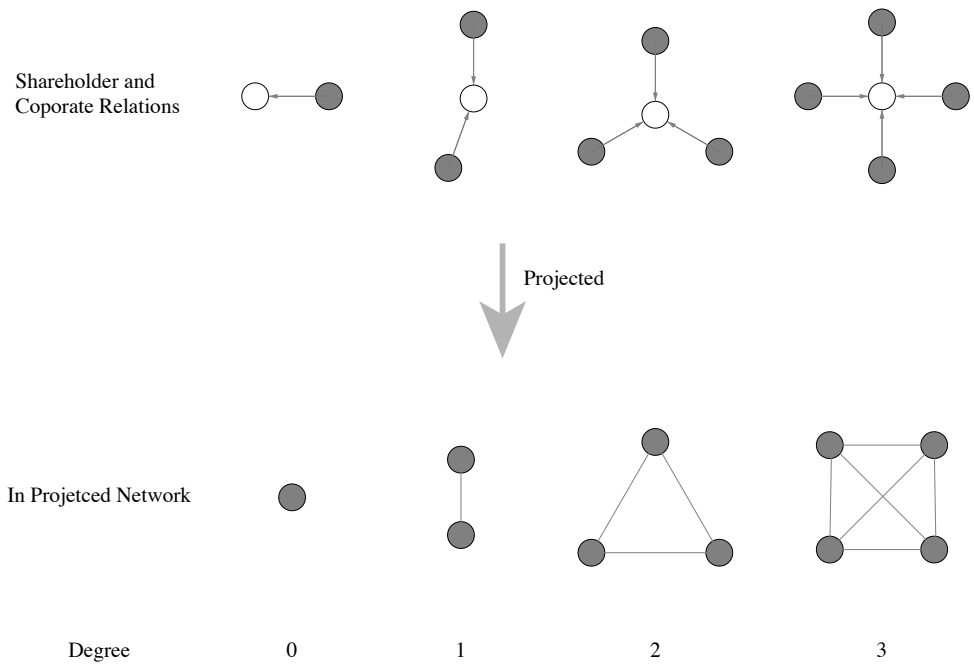
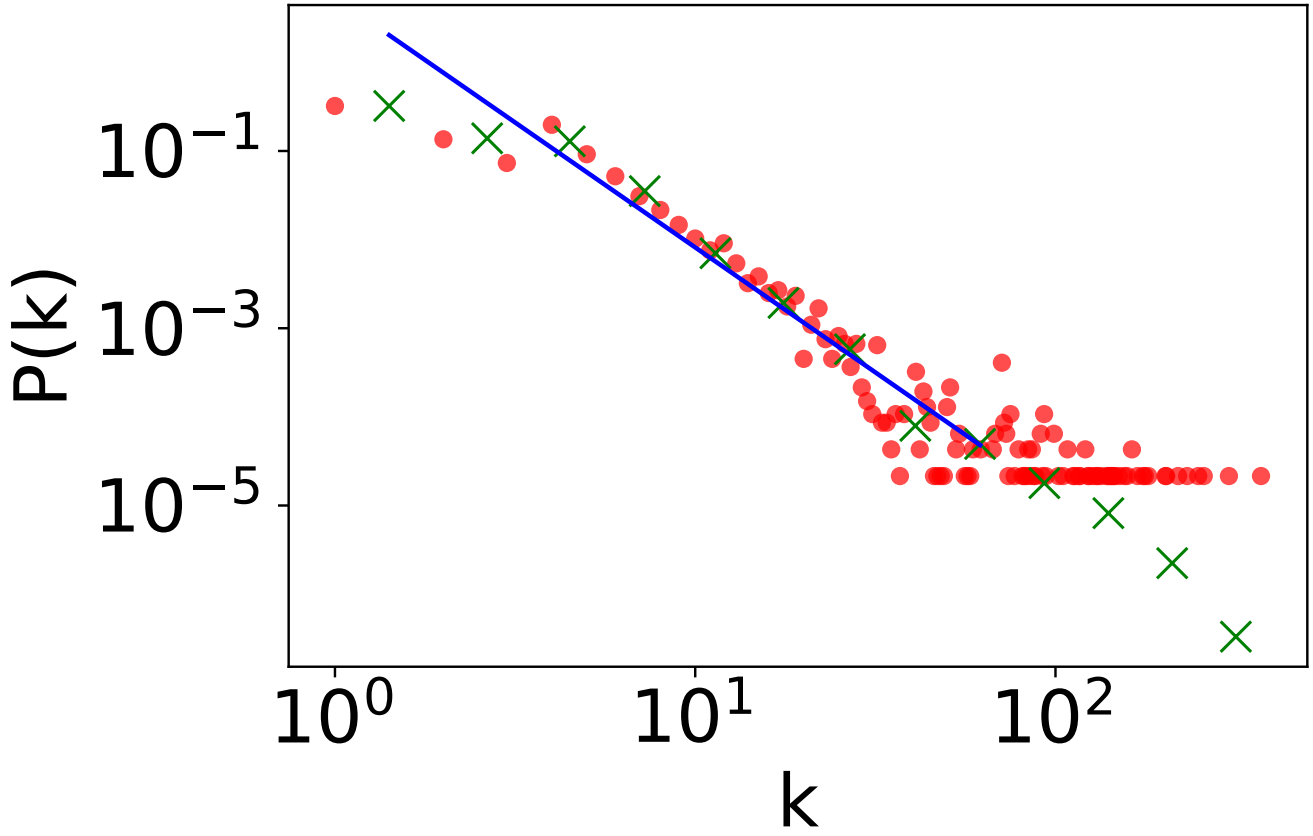


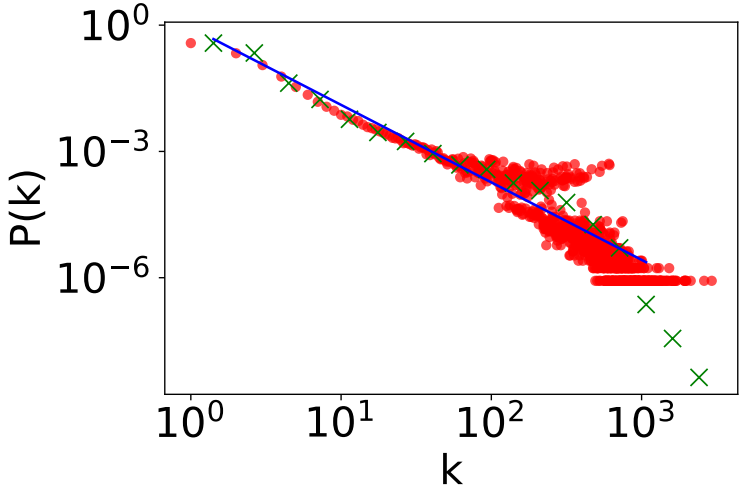
Figure 2: Illustration of small structures of the network

Degree Distributions

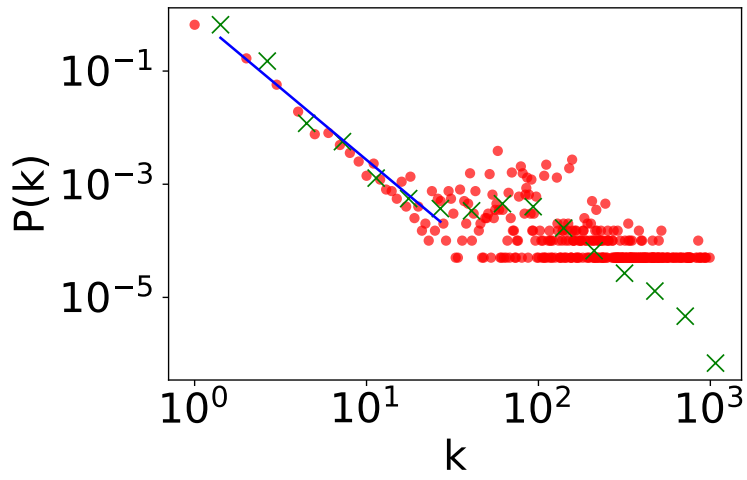


(a) Turkey $\gamma \approx 2.8$

Degree Distributions



(b) Germany $\gamma \approx 1.84$



(c) Netherlands $\gamma \approx 2.5$

Community Analysis

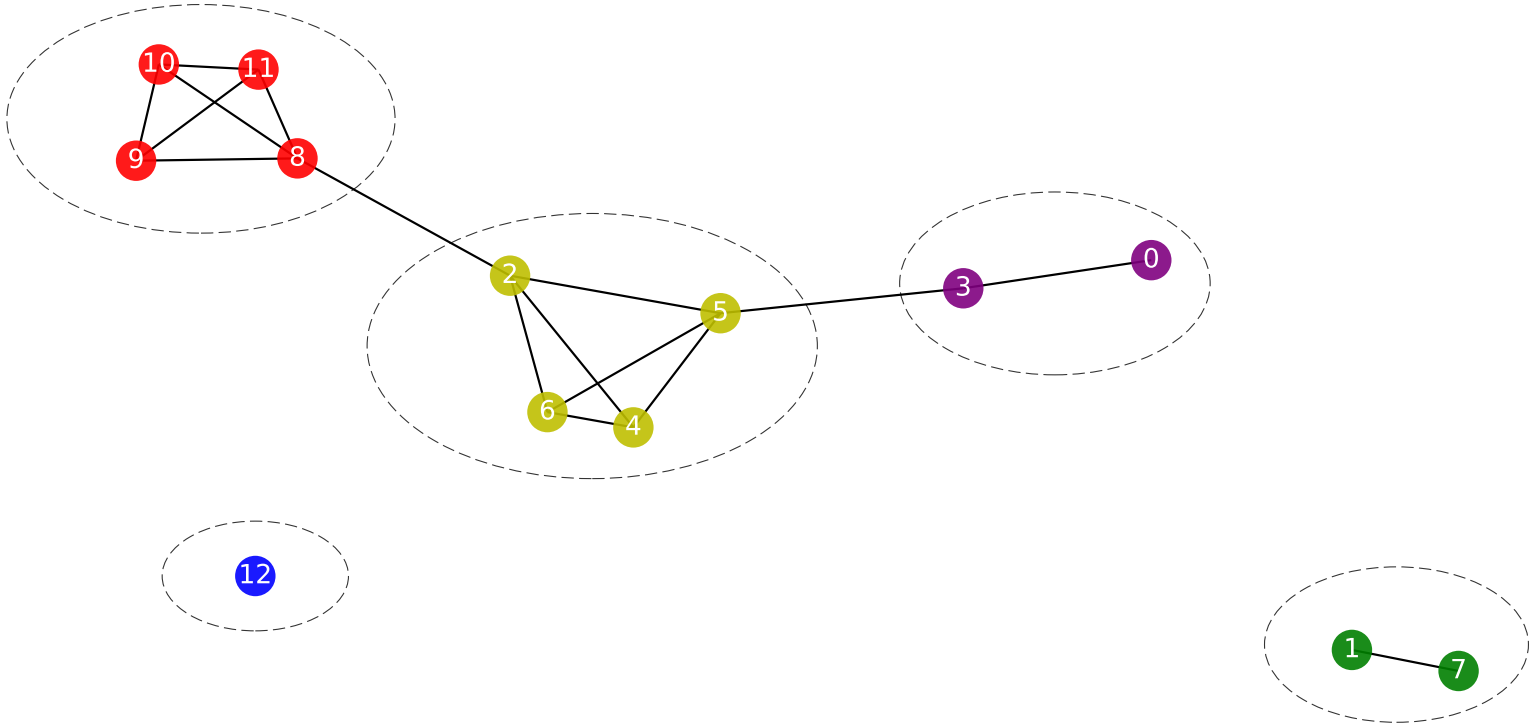
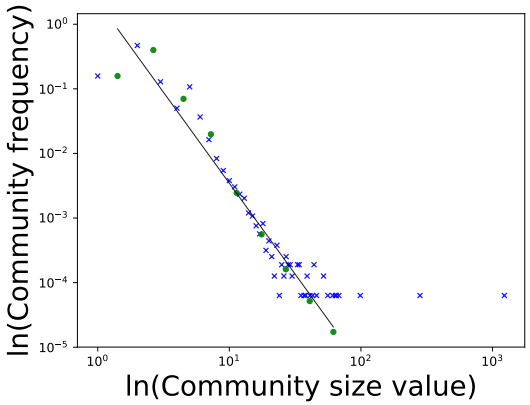
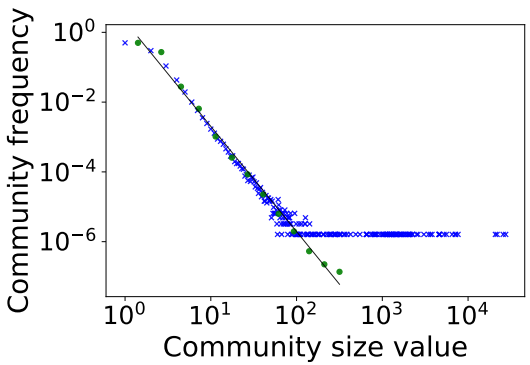


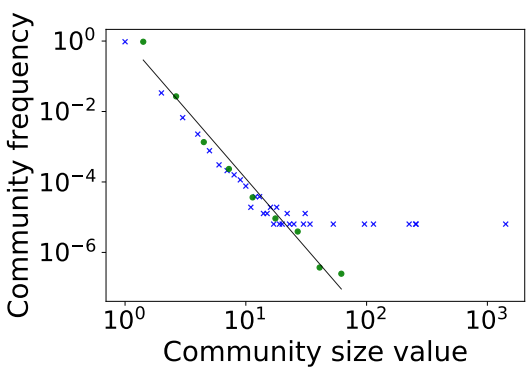
Figure 3: The same projected graph as in Figure 8 from the network graph shown in Figure 1.

Community Analysis
Community size frequency

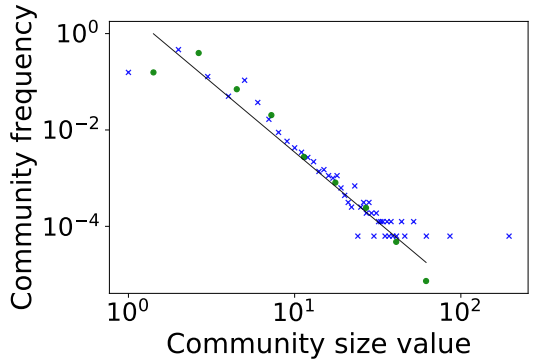
Louvain

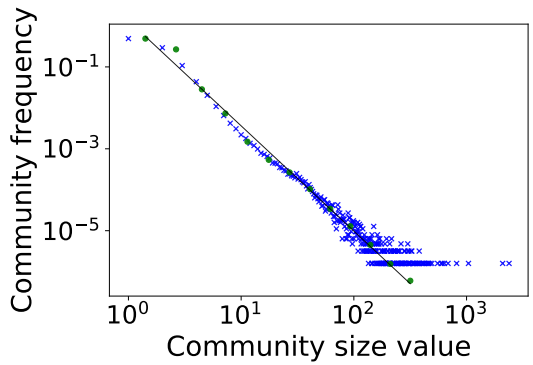


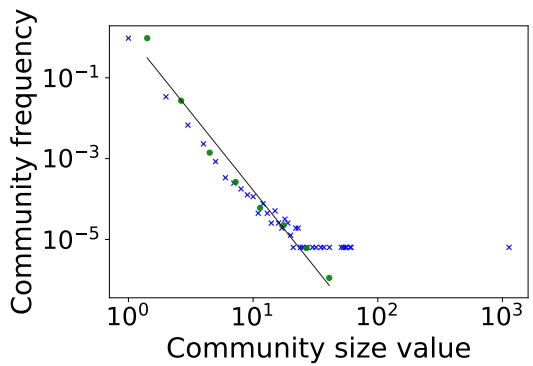




Infomap





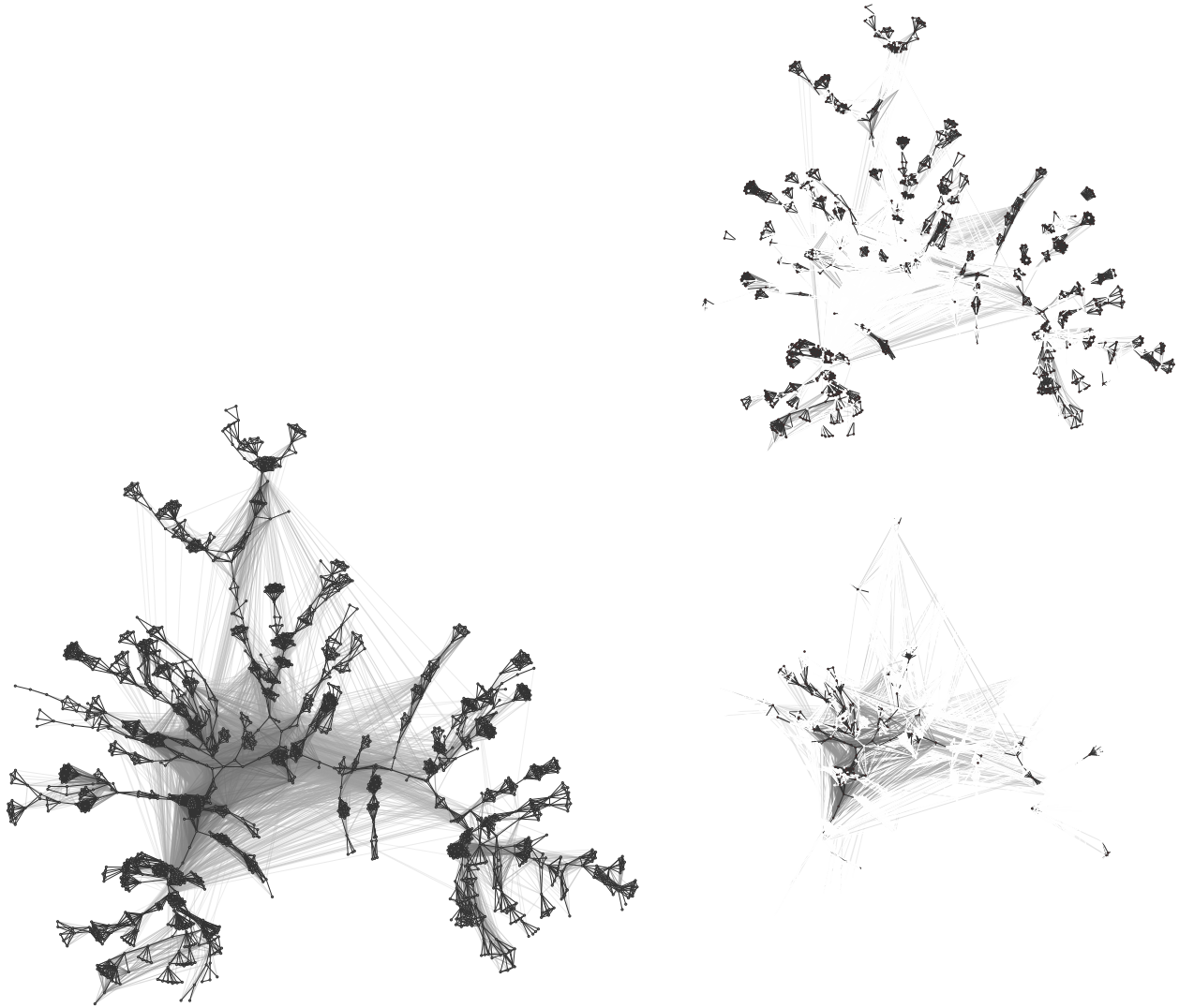


Types of Owners

This classification is retrieved from the BvD database

Owner Type ID	Owner Type
1	Employees/Managers/Directors
2	Venture capital
3	Other unnamed shareholders aggregated
4	Financial company
5	One or more named individuals or families
6	Public (publicly listed companies)
7	Public authority State Government
8	Hedge funds
9	Insurance company
10	Self ownership
11	Private Equity firms
12	Industrial company
13	Mutual & Pension Fund/Nominee/Trust/Trustee
14	Bank
15	Foundation/Research Institute

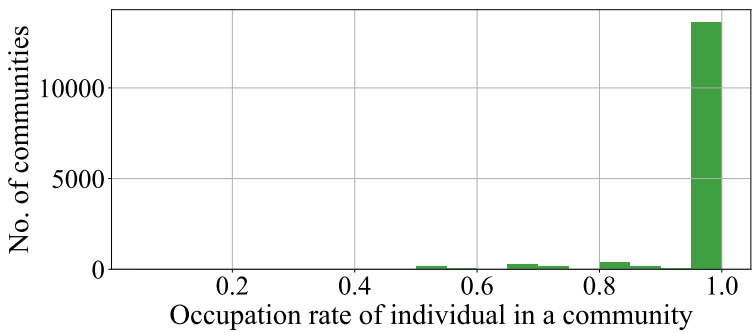
Types of Owners

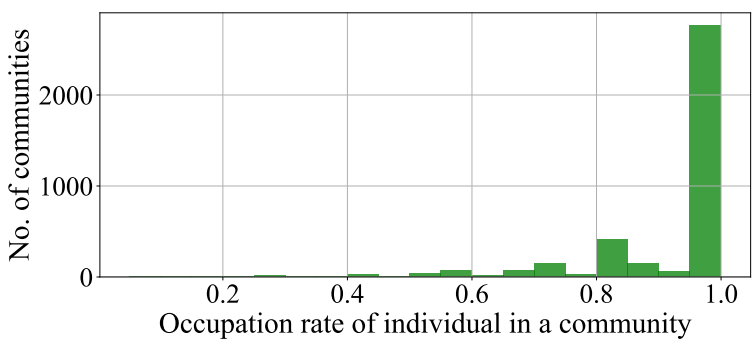


Family Owners in Turkey
Louvain

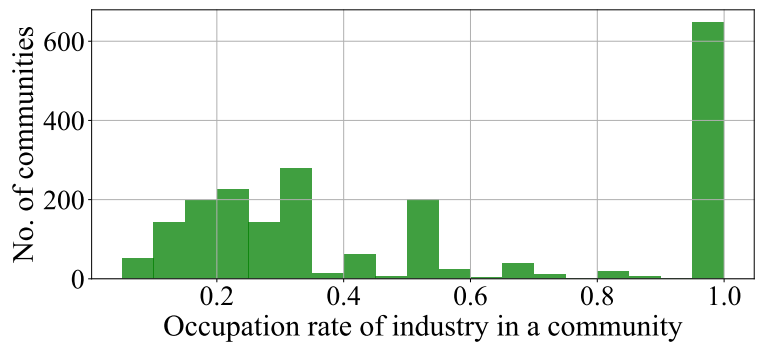
Occupation rate for one type of investor inside a community: $r_i = \frac{n_i}{\sum_j n_j}$, where n_i is the number of type i investor and $\sum_j n_j$ is the number of all types investors inside a community

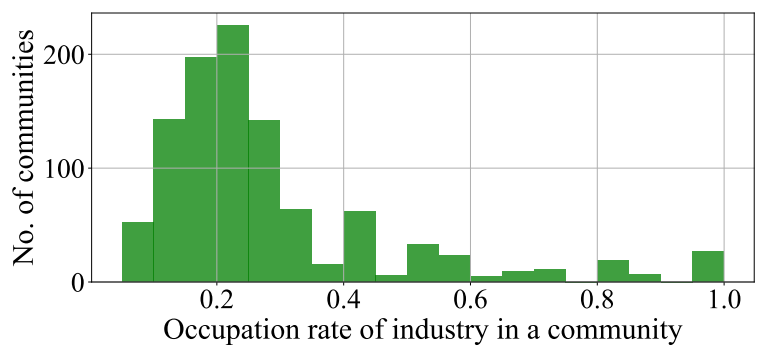
Individuals or families





Industrial company

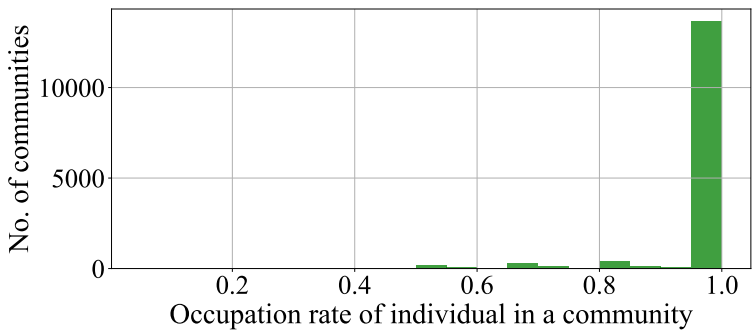


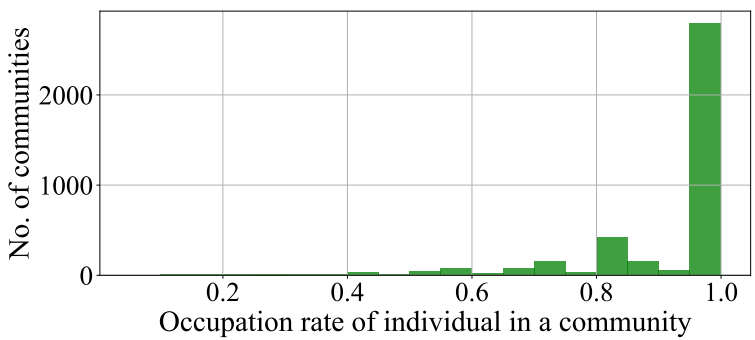


Families in Turkey
Infomap

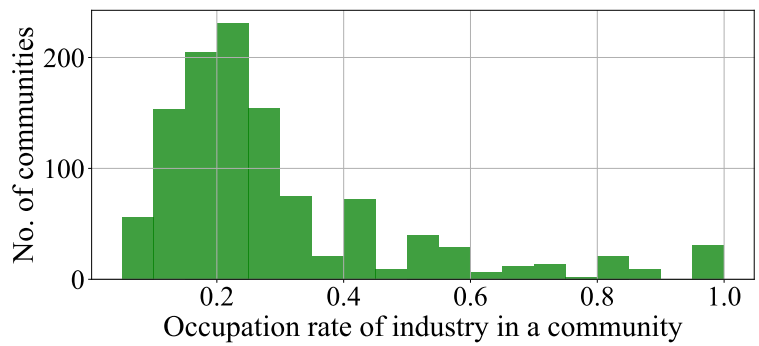
If the structure of communities is established well enough, the two should be able to give similar results, see [17]

Individuals or families





Industrial company



Motivation for making models

Why model?

- Node label considered(heterogeneity of the networks)
- Mimic the dynamical process of the network(but with some quantity conserved)
- To reproduce the small world effect, large clustering coefficient and the emergence of a communities

Why random walk?

- Mimic local researching behaviour

Random Walks

i Initialisation with a random directed graph D

ii Randomly pick up a node o and an edge (o, v);

iii Let a random walker start from the o walk to the next neighbor vertex b

iv Reverse the graph and let random walker to continue walk from b to a neighbour, p

v Reverse the graph again the let the random walker to continue walk from p to a neighbour, c

vi Check the new edge(o, c) whether exists in the graph. Delete the starting edge (o, v), only when edge(o, a) does not exist in the graph. Thus the new directed edge (o, c) has been created. If the edge already exists in the graph, make node c as the starting node continue 3, Until a new edge is found or exceed the maximum trial (100).

vii Go back to 4.

Random Walks

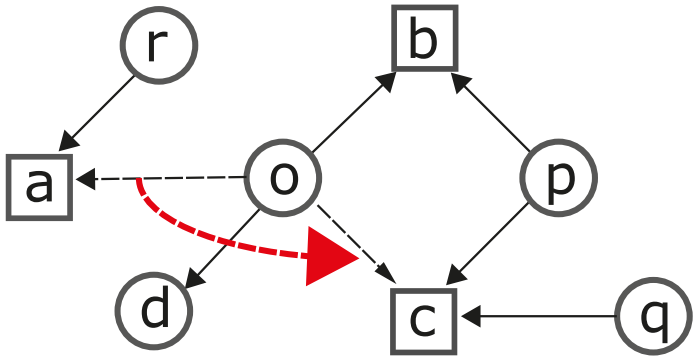


Figure 4: The illustration of rewiring based on the random walk on a directed network. The directed edge (o, a) is rewired to (o, c) based on the random walk starting from o.

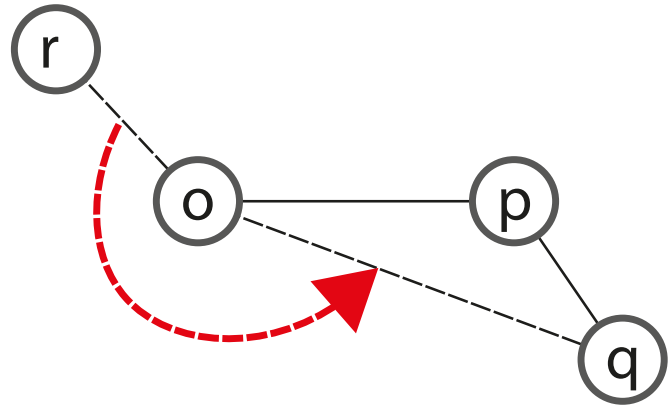


Figure 5: The illustration of the undirected networks projected from the rewired directed networks. After rewiring, the edge undirected edge $(r, 0)$ is rewired to (o, q). A triangle is created.

Two Types of Investors

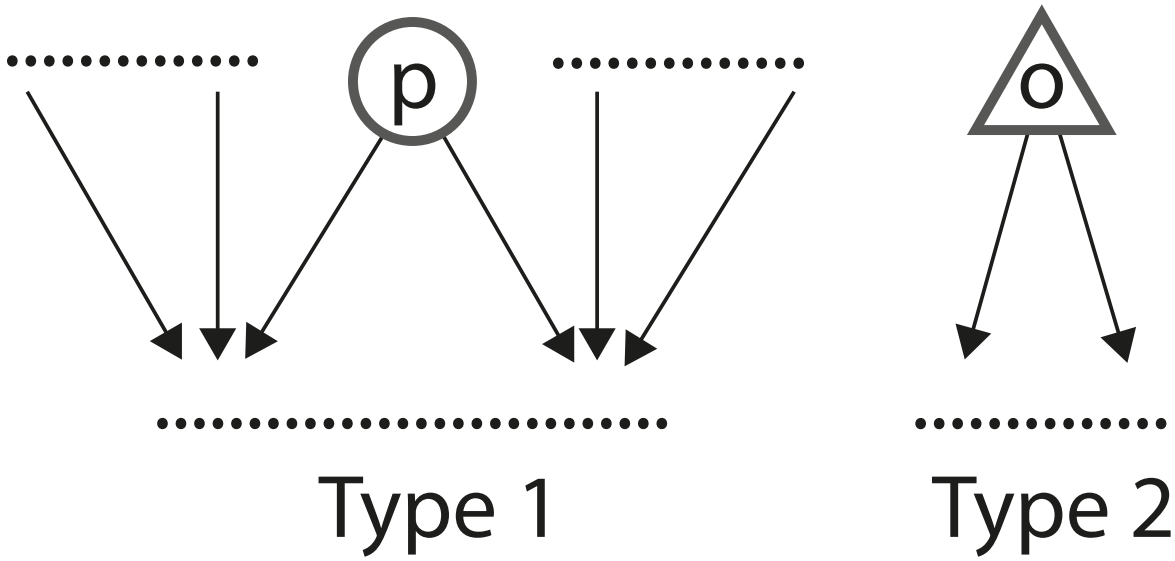


Figure 6: Illustration for two types of vertex, type 1 prefer to attaching to targets with lots of other predecessors; type 2 prefer to attaching to targets with no more two predecessors

Two Types of Investors

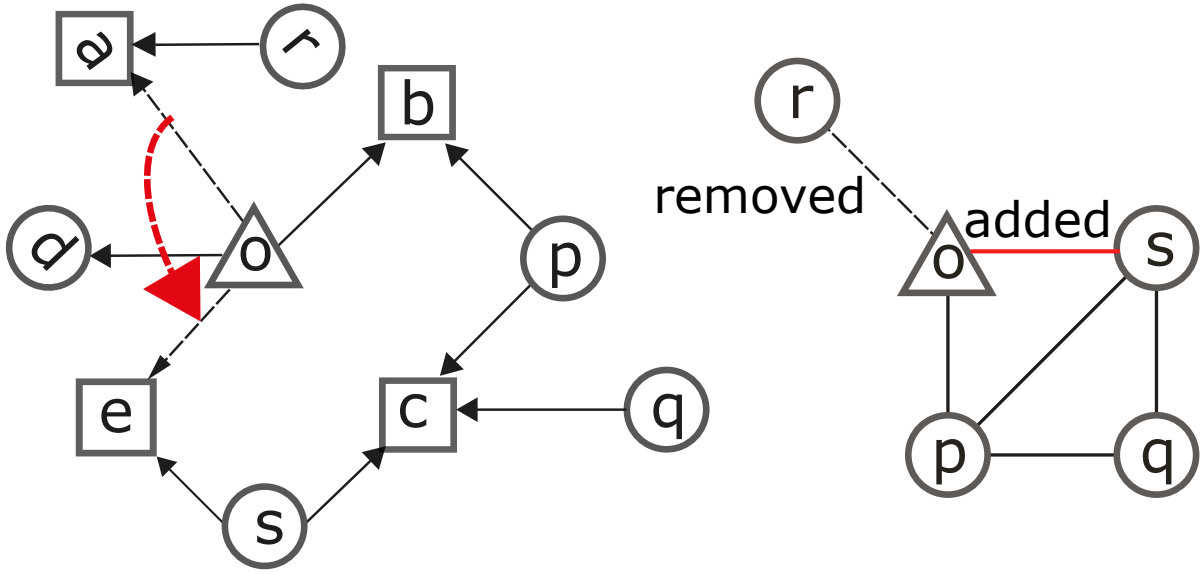


Figure 7: Illustration of rewiring based on the random walk on a directed network with labelled nodes.

Simulation Results

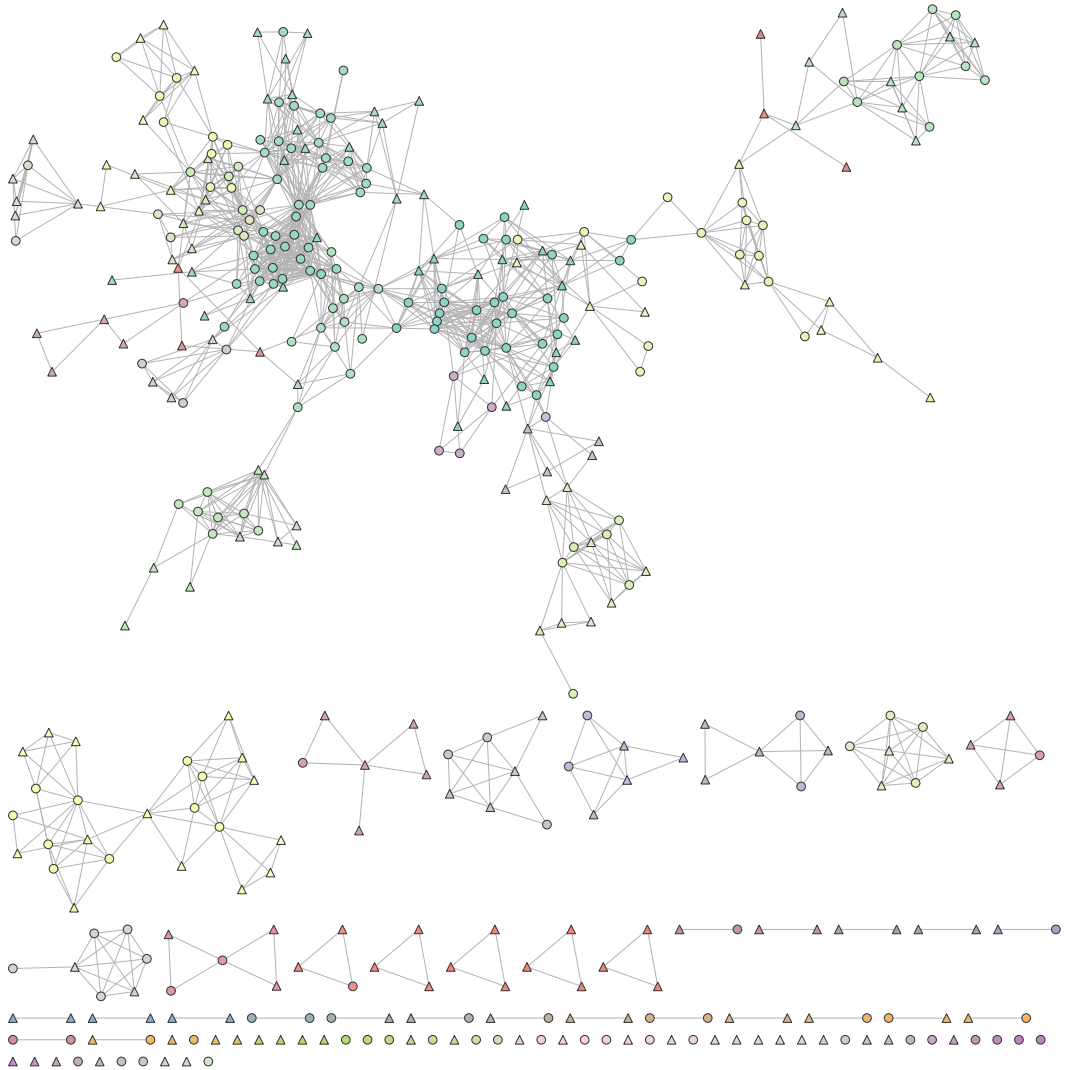
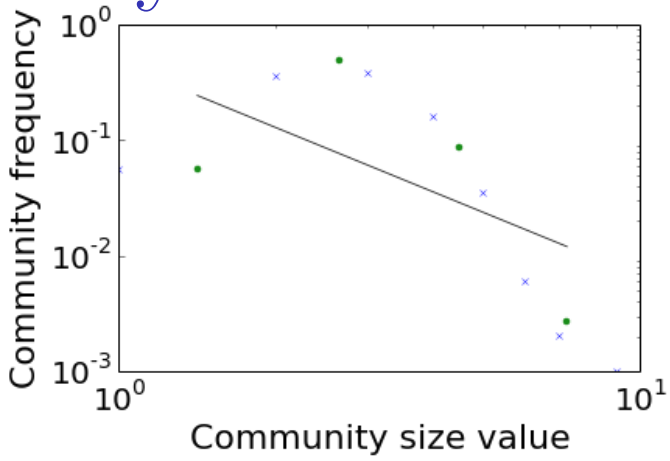
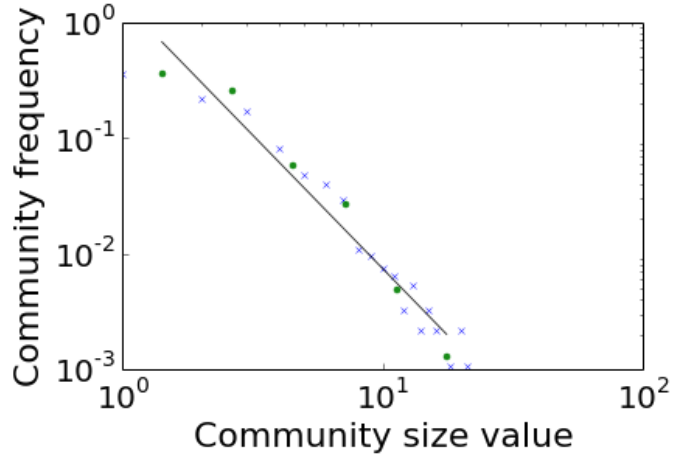


Figure 8: The projected network with 200 nodes and 1315 edges. Triangle stands Type 2 and circle stands for Type 1. Colours represented different communities detected using infomap.

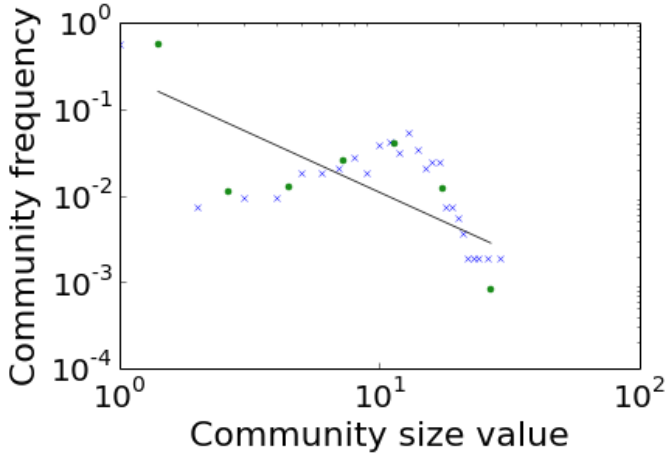
Community size distribution

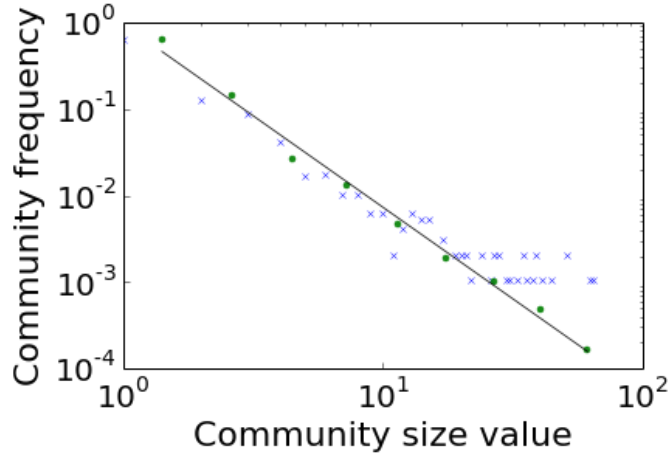
Louvain





Infomap

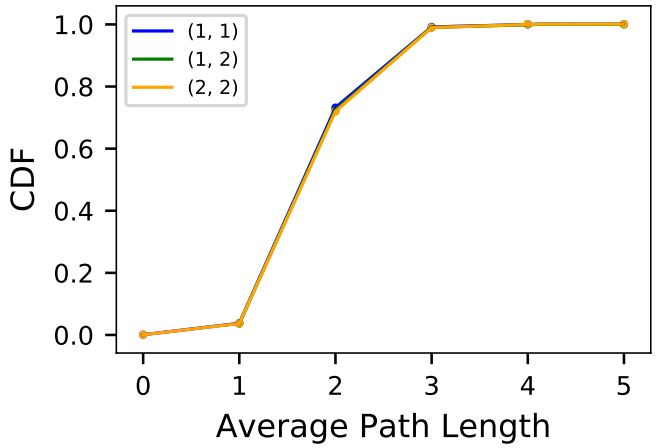




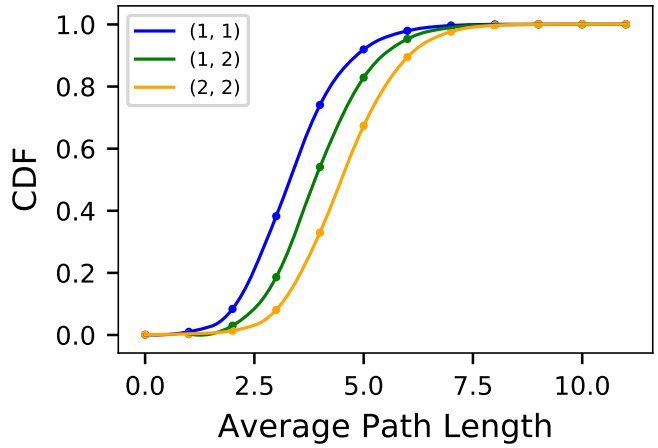
Rewiring process with $p_{rw} = 0.98$ and $E_{th} = 2$. 3000 nodes and 8660 edges. The number of communities of louvain method is 931 compared with the number of Infomap is 961. And its average clustering coefficient is 0.53.

Average path length

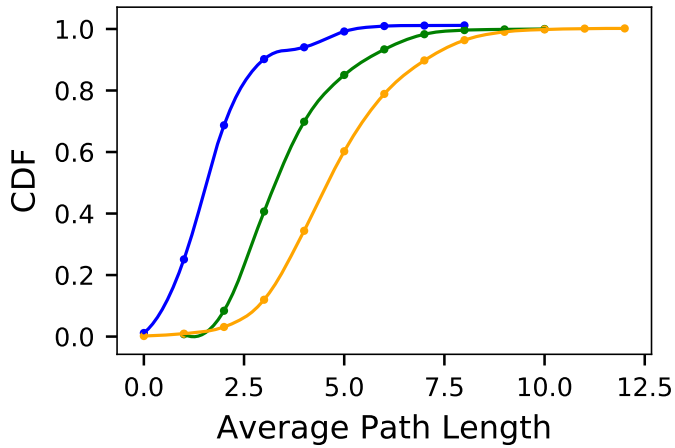
(a.1) Random walk without labels



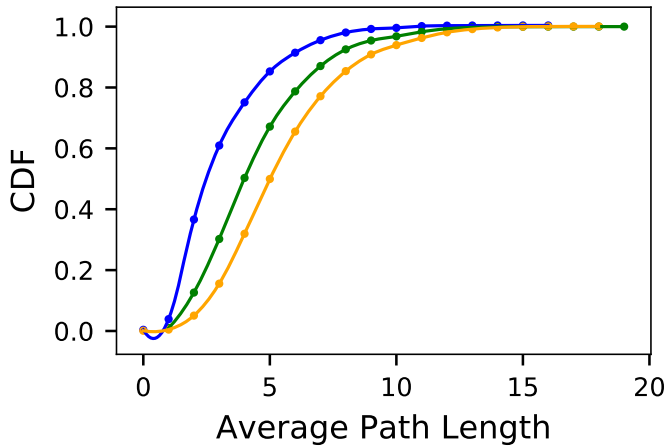
(a.2) Random walk with labels



(b.1) LCC of Turkey Shareholder Network

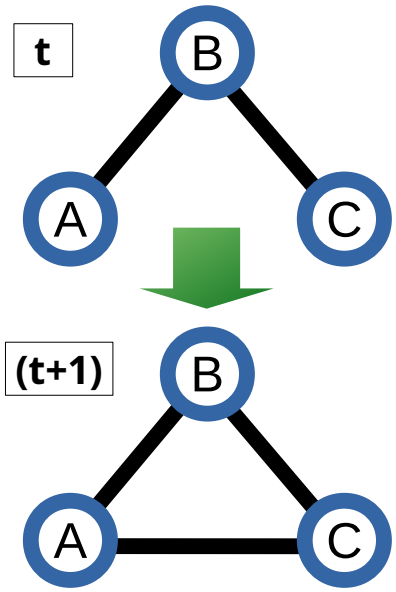


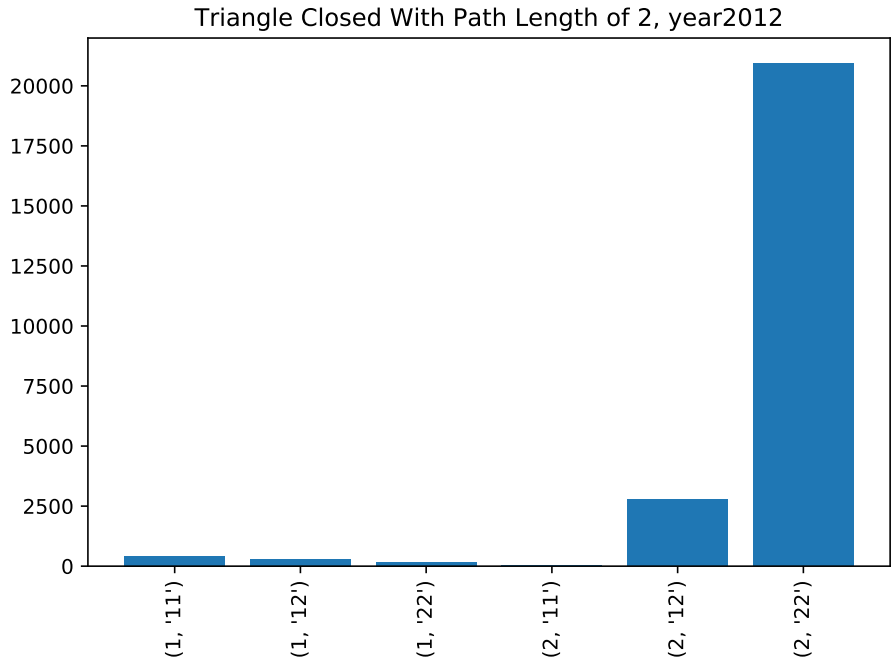
(b.2) LCC of Netherlands Shareholder Network



Understanding the local structure

Observation: Triadic Closure for different type of nodes.

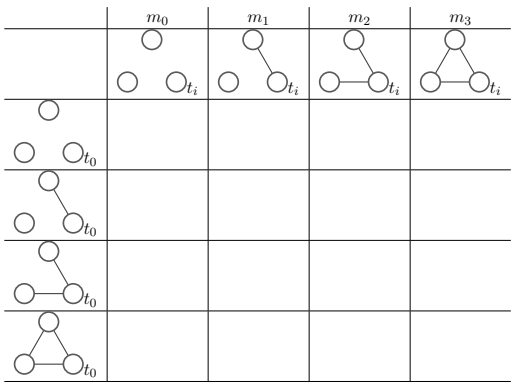


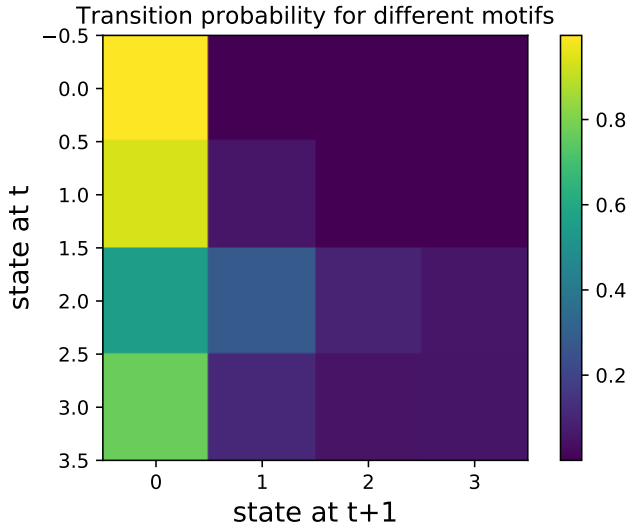


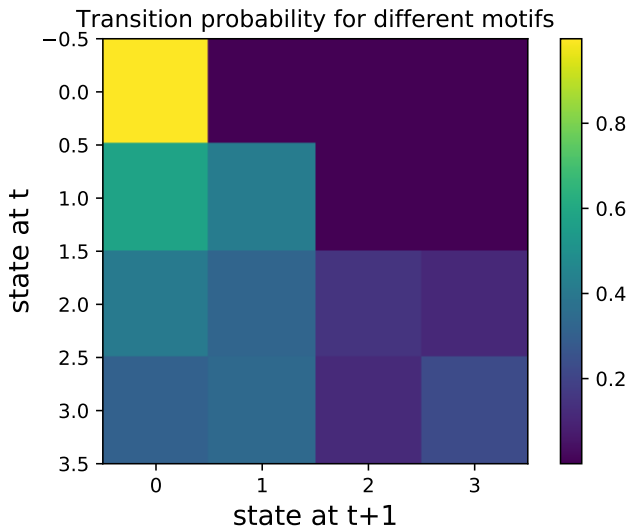
Representation?

Understanding the local structure

Motif transition







Financial crisis so most motif become m_0, i.e. many people no longer invest some common assets

Summary

Our model

- is able to explain the emergence of scaling modular structures through random walks rewiring and many small world properties, such as the average shortest.
- is different from the previous models since it generated small components and isolated nodes, not just the connected component.
- is qualitatively matched of the statistics, maybe to extended by careful parameter estimation.

Summary

Thank you for your attention!

Bibliography I

[1] D. Acemoglu, U. Akcigit, and W. Kerr. Networks and the macroeconomy: An empirical exploration. *NBER Macroeconomics Annual*, 30(1):273–335, 2016.

[2] W. B. Arthur. Complexity and the economy. *science*, 284(5411):107–109, 1999.

[3] S. Y. Auyang. *Foundations of complex-system theories: in economics, evolutionary biology, and statistical physics*. Cambridge University Press, 1999.

[4] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[5] A. A. Berle and G. G. C. Means. *The modern corporation and private property*. Transaction publishers, 1991.

[6] V. D. Blondel, J.-l. Guillaume, and E. Lefebvre. Fast unfolding of communities in large networks. pages 1–12.

[7] K. Christensen and N. R. Moloney. *Complexity and criticality*, volume 1. Imperial College Press, 2005.

[8] R. M. Cyert, J. G. March, et al. A behavioral theory of the firm. *Englewood Cliffs, NJ*, 2, 1963.

[9] S. Dorogovtsev, J. Mendes, and A. Samukhin. Size-dependent degree distribution of a scale-free growing network. *Physical Review E*, 63(6):062101, 2001.

[10] T. Evans and J. Saramäki. Scale-free networks from self-organization. *Physical Review E*, 72(2):026138, 2005.

[11] J. D. Farmer, M. Gallegati, C. Hommes, A. Kirman, P. Ormerod, S. Cincotti, A. Sánchez, and D. Helbing. A complex systems approach to constructing better models for managing financial markets and the economy. *Eur. Phys. J. Spec. Top.*, 214:295–324, 2012.

[12] S. Fortunato. Community detection in graphs. *Physics Reports*, 486(3-5):75–174, 2010.

[13] R. F. Fox and Y.-n. Lu. Emergent collective behavior in large numbers of globally coupled independently stochastic ion channels. *Physical Review E*, 49(4):3421, 1994.

[14] L. C. Freeman. Centrality in social networks conceptual clarification. *Social networks*, 1(3):215–239, 1978.

Bibliography II

[15] J. Haleblian, C. E. Devers, G. McNamara, M. A. Carpenter, and R. B. Davison. Taking stock of what we know about mergers and acquisitions: A review and research agenda. *Journal of Management*, 2009.

[16] P. L. Krapivsky, S. Redner, and F. Leyvraz. Connectivity of growing random networks. *Physical review letters*, 85(21):4629, 2000.

[17] A. Lancichinetti, S. Fortunato, and F. Radicchi. Benchmark graphs for testing community detection algorithms. *Physical review E*, 78(4):046110, 2008.

[18] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(1):2, 2007.

[19] M. Newman. *Networks: an introduction*. OUP Oxford, 2010.

[20] M. E. Newman. Power laws, pareto distributions and zipf's law. *Contemporary physics*, 46(5):323–351, 2005.

[21] R. Porta, F. Lopez-de Silanes, and A. Shleifer. Corporate ownership around the world. *The journal of finance*, 54(2):471–517, 1999.

[22] R. L. Porta, F. Lopez-de Silane, A. Shleifer, and R. W. Vishny. Law and finance. Technical report, National Bureau of Economic Research, 1996.

[23] H. A. Simon. A behavioral model of rational choice. *The quarterly journal of economics*, pages 99–118, 1955.

[24] H. A. Simon. On a class of skew distribution functions. *Biometrika*, 42(3/4):425–440, 1955.

[25] S. H. Strogatz. Exploring complex networks. *Nature*, 410(6825):268–276, 2001.

[26] S. Vitali, J. B. Glattfelder, and S. Battiston. The network of global corporate control. *PloS one*, 6(10):e25995, 2011.

[27] M. Zeitlin. Corporate ownership and control: The large corporation and the capitalist class. In *Classes, Power, and Conflict*, pages 196–223. Springer, 1982.