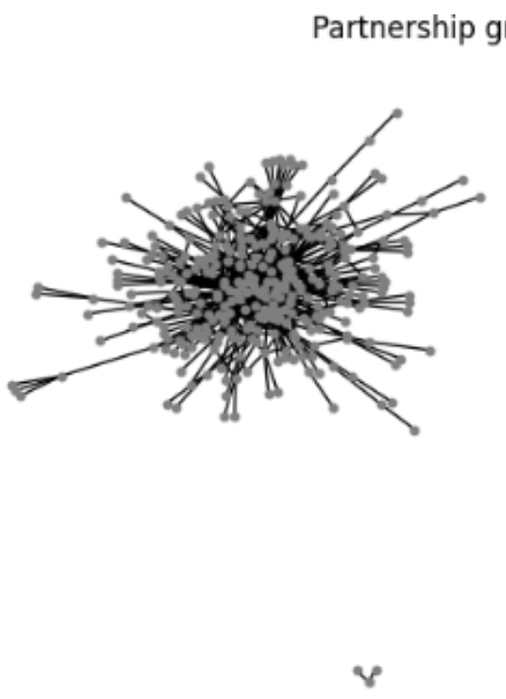
What to say to business applying SNA

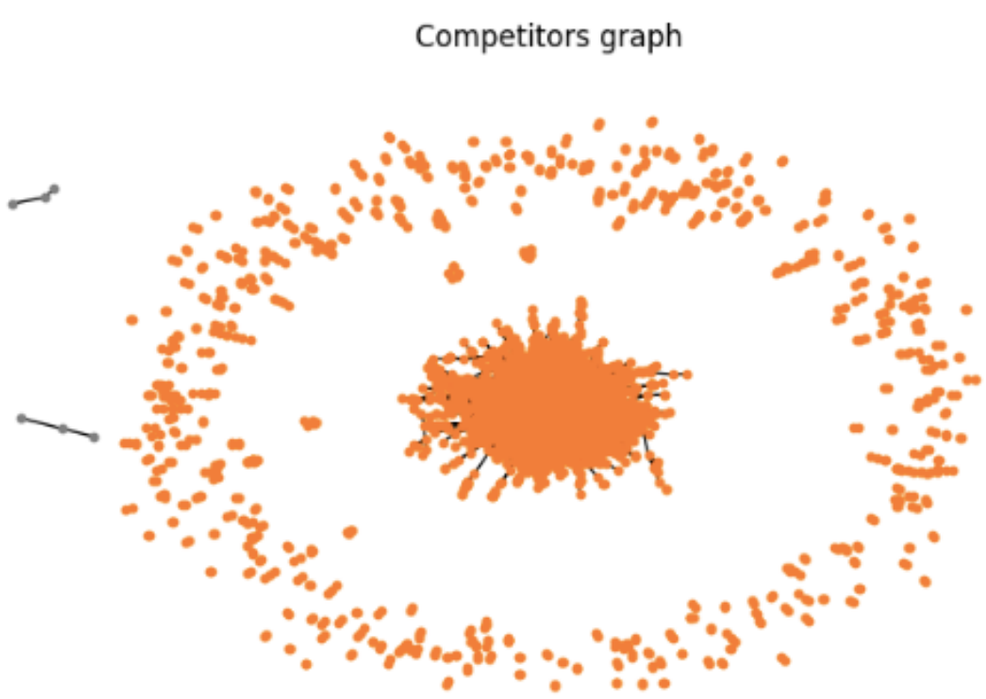
Network Social Analysis
MS Data Science for Economics
Daria Shchrebakova

project framework

- **Puzzle question:** what are the possibilities for business (new partners, competitors, customers)?

- **Data**: companies' undirected networks (filtered by the type of connections - partnership, competitor, customer)

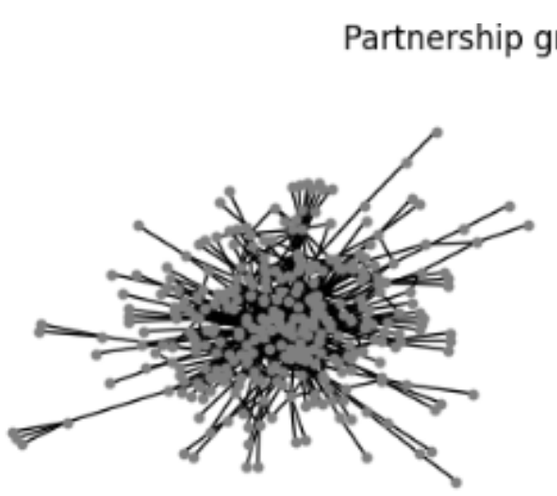


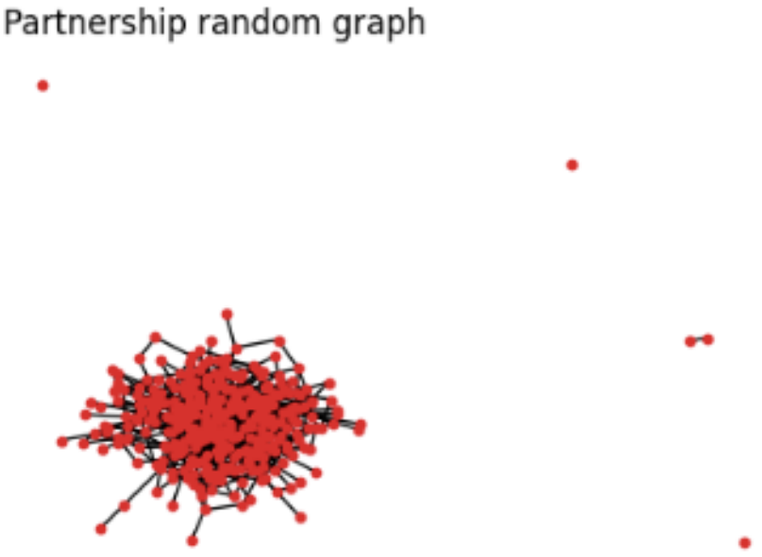


collaboration network

The initial graph with 43065 nodes and 108831 edges has been randomly subsampled
(1000 companies with >1 partner)

real VS random

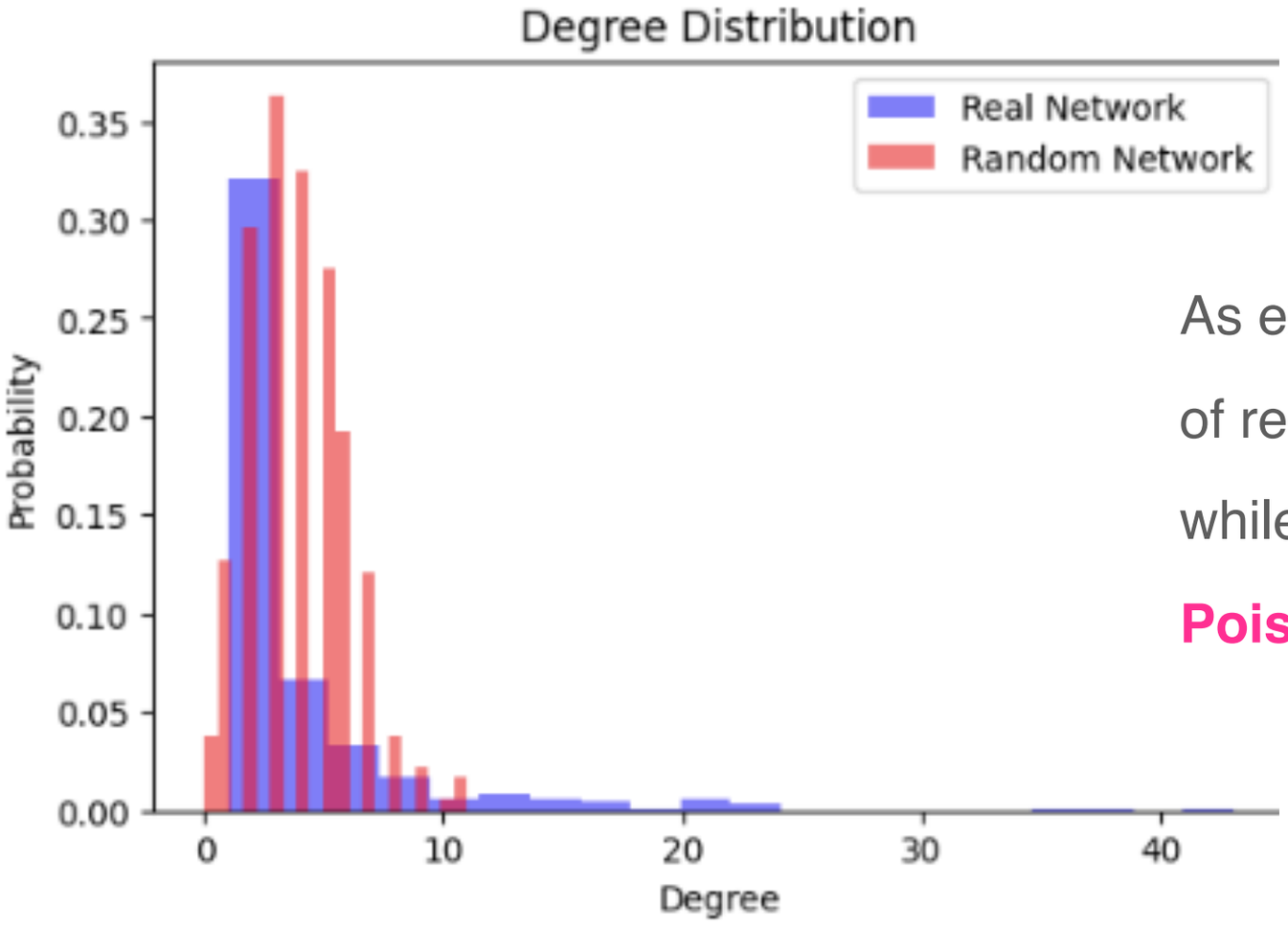




- nodes: 331 , edges: 649
- avg. degree: 3.92 (sd = 5.21)
- density: 0.0119
- connected components: 4
- bridges: 21,6 % out of network

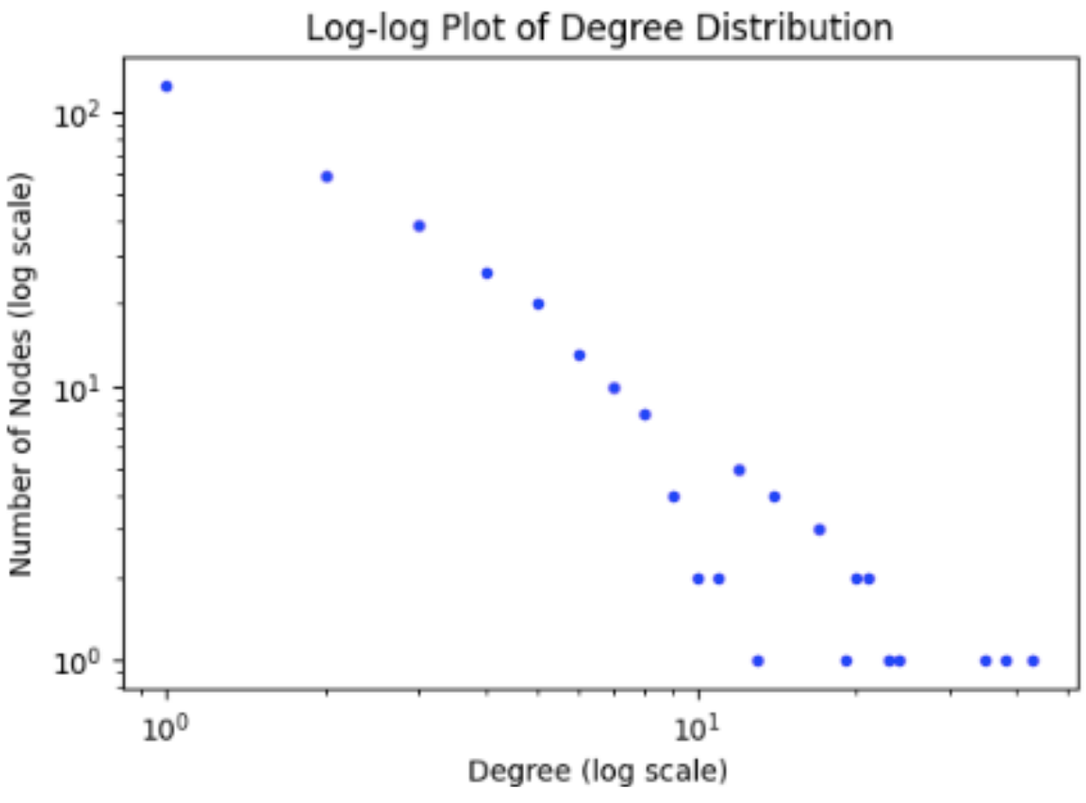
- edges: 657
- avg. degree: 3.97 (sd = 2.04)
- connected components: 9

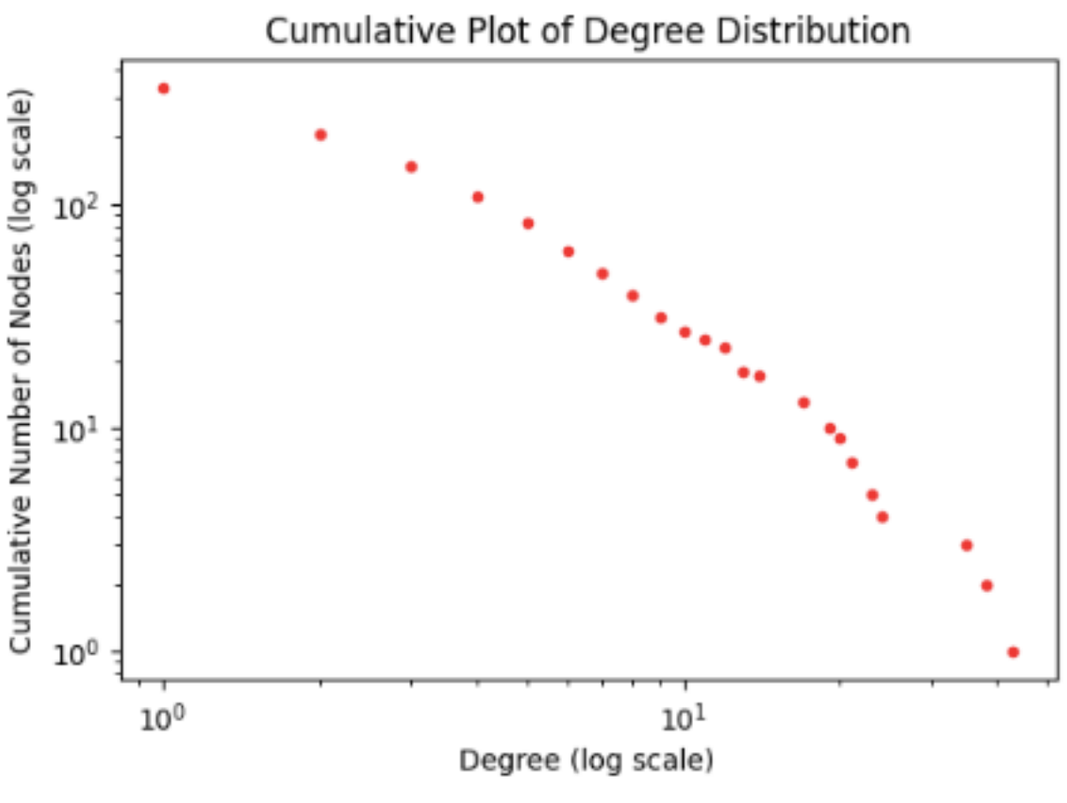
real VS random



As expected, the degree distribution of real network reminds **Power Law**, while random network recalls **Poisson distribution**

just real


Log-log Plot of Degree Distribution
Number of Nodes (log scale)
Degree (log scale)
10^2
10^1
10^0
10^0
10^1


Cumulative Plot of Degree Distribution
Cumulative Number of Nodes (log scale)
Degree (log scale)
10^2
10^1
10^0
10^0
10^1

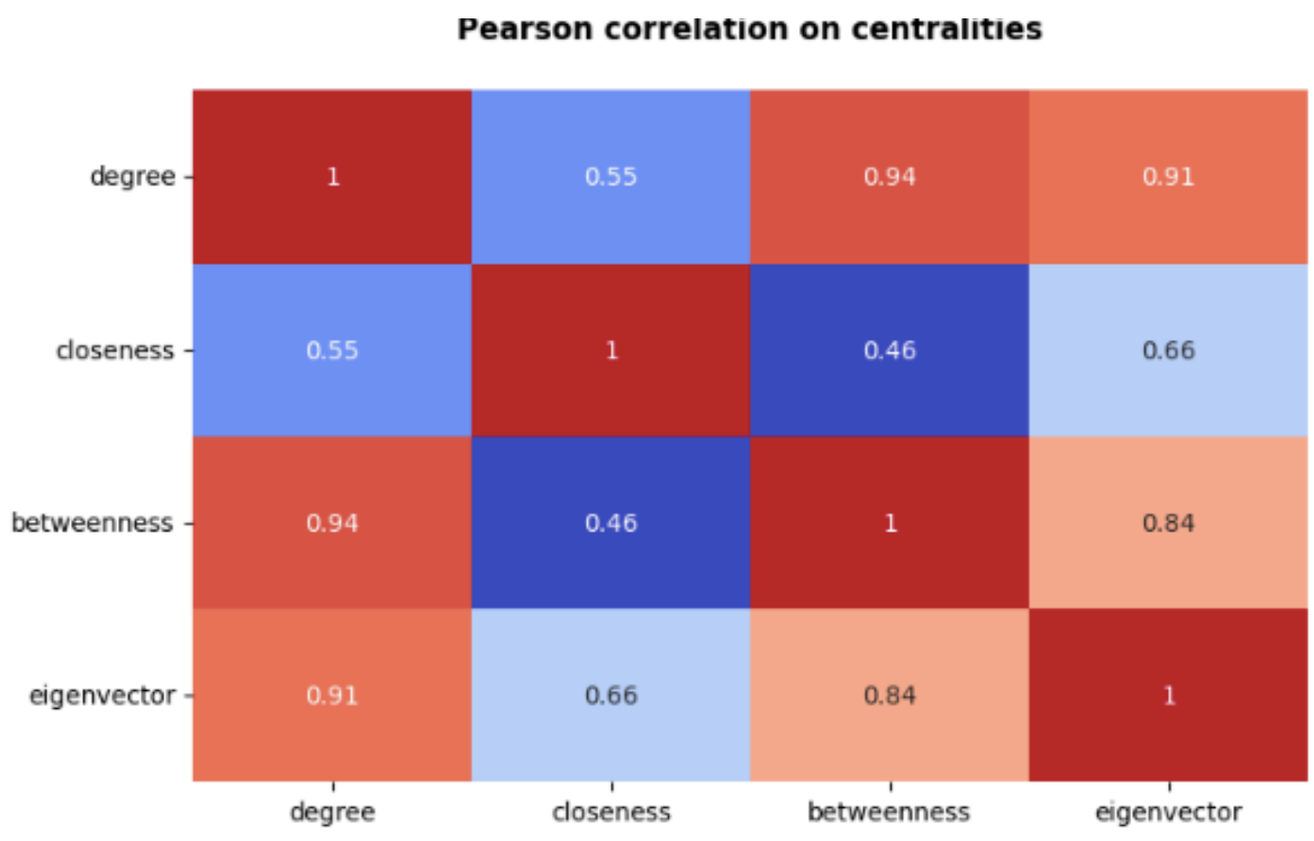
the most central is

Degree: SIEMENS citrix NetApp Deloitte. HSN

Closeness: SIEMENS citrix NetApp Deloitte. DELTA

Betweenness: SIEMENS citrix NetApp Deloitte. DELTA

Eigenvector: SIEMENS citrix NetApp Qualcomm Deloitte.



👉connectivity👈

The graph is not connected, there are 4 connected components in the network.

The largest component (98% of the network):

– nodes: 322

– edges: 643

– Avg. shortest length path: 3.77

– diameter: 9

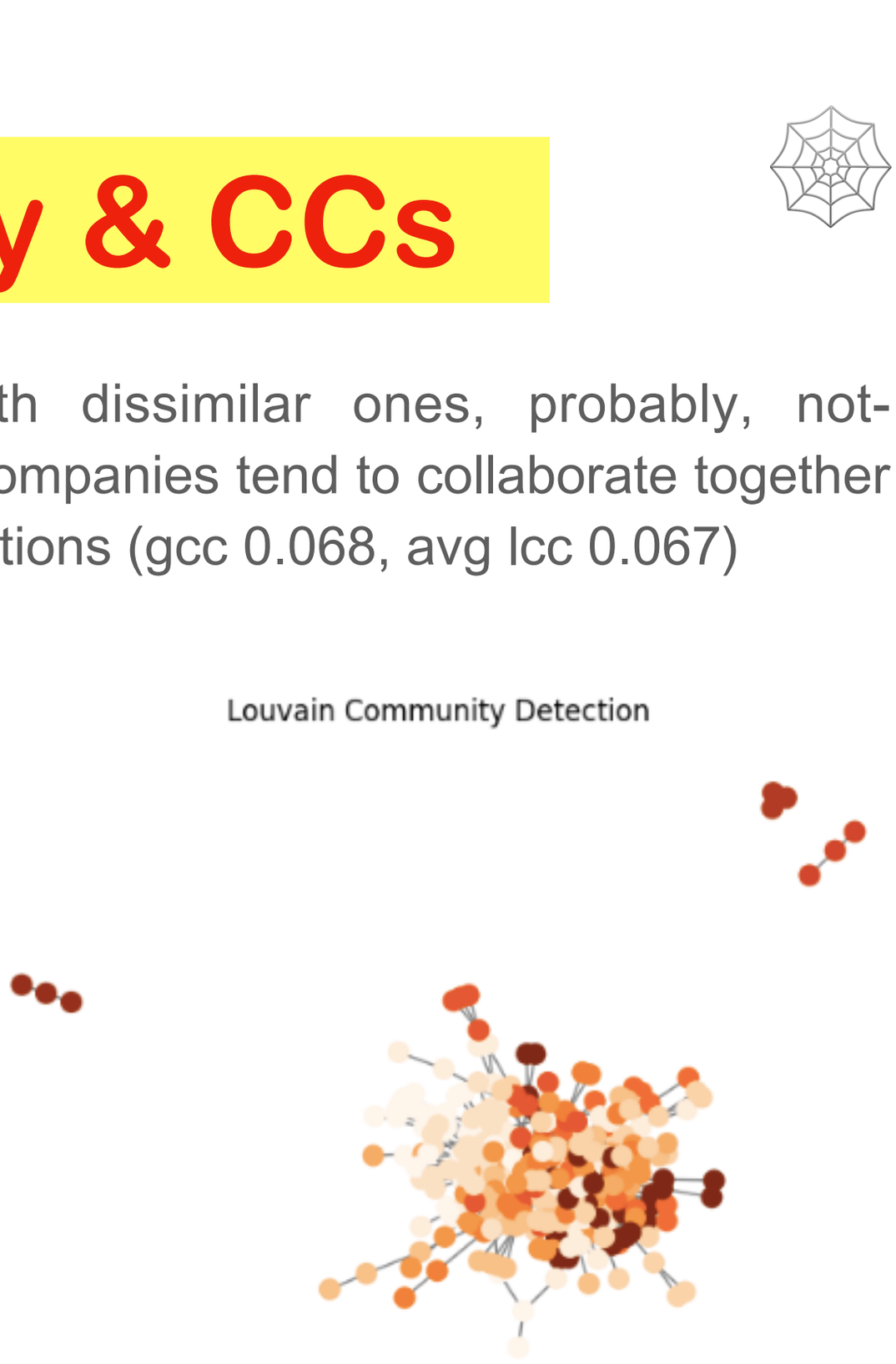
transitivity & CCs

Companies tend to collaborate with dissimilar ones, probably, not-competitors (assortativity -0.12), and companies tend to collaborate together but don’t create the cluster of collaborations (gcc 0.068, avg lcc 0.067)

Greedy Modularity Communities

Network Modularity = 0.531
 # communities = 17

Louvain Community Detection



Network Modularity = 0.525
 # communities = 14

With whom can we collaborate?

Based on structural equivalence measures, the link prediction model has been applied (logistic regression)

Interpretation: with increase in Jaccard distances, RAI, AAI, Pref (separately) the probability of being customer between 2 companies increases (by 0.053, -1.012, 1.302, 47.214)

Examples: Wolters Kluwer & Change.org; Panasonic & Northrop Grumman; Panasonic & Jetstar

Model quality: accuracy 75%, predicted 68% of non-partners classified are non-partners and 90% for partners;
93% of non-partners and 58% of partners are classified correctly.

With whom can we collaborate?

Based on Node2Vec technique the recommendation system provides possible candidates for collaborations

Siemens could've consider collaboration with **Cypress Semiconductor, LogLogic, Banner Health, Rittal**

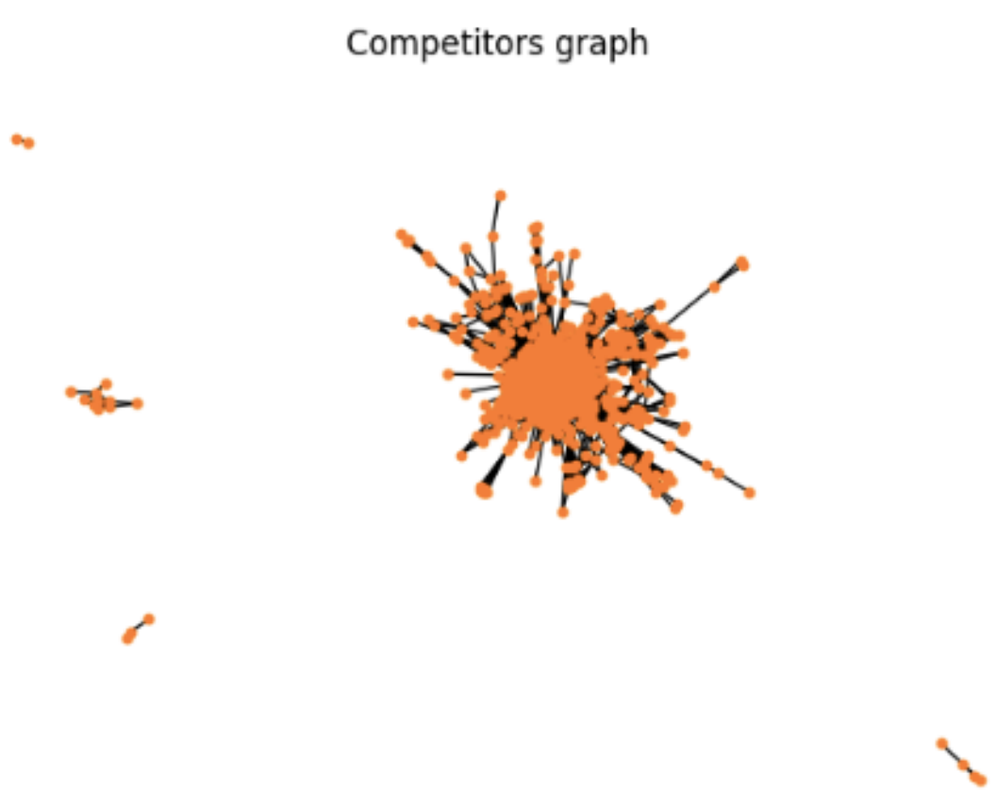
Microsoft may collaborate with **Sanlam, Duracell, DreamFactory Software, Laboratory Corporation of America, American Water**

Boeing may collaborate with **Reliant Energy, SyCom, Entergy, LoJack, Rittal**

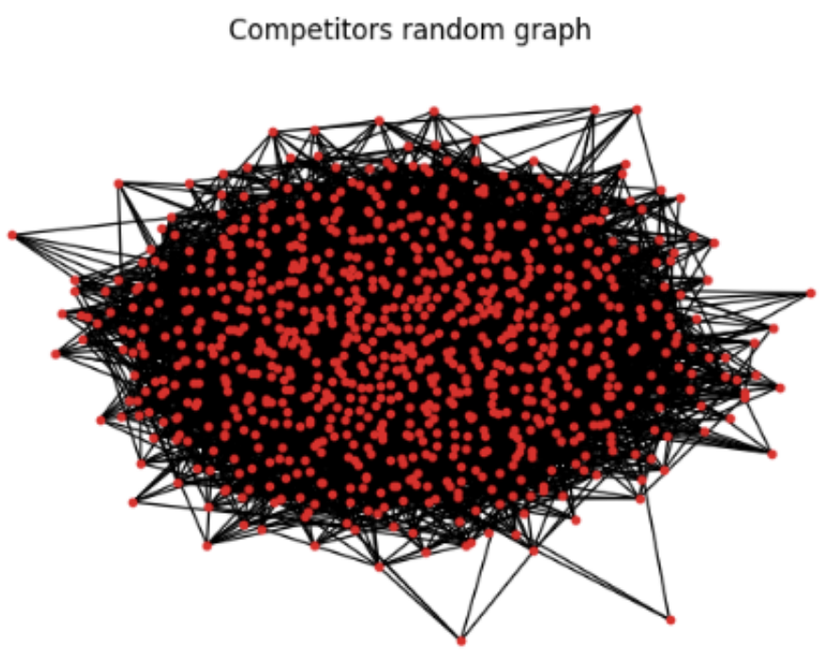
🥊competitors network🥊

The initial graph with 8596 nodes and 23654 edges has been subsampled
(companies with > 10 competitors)

real VS random

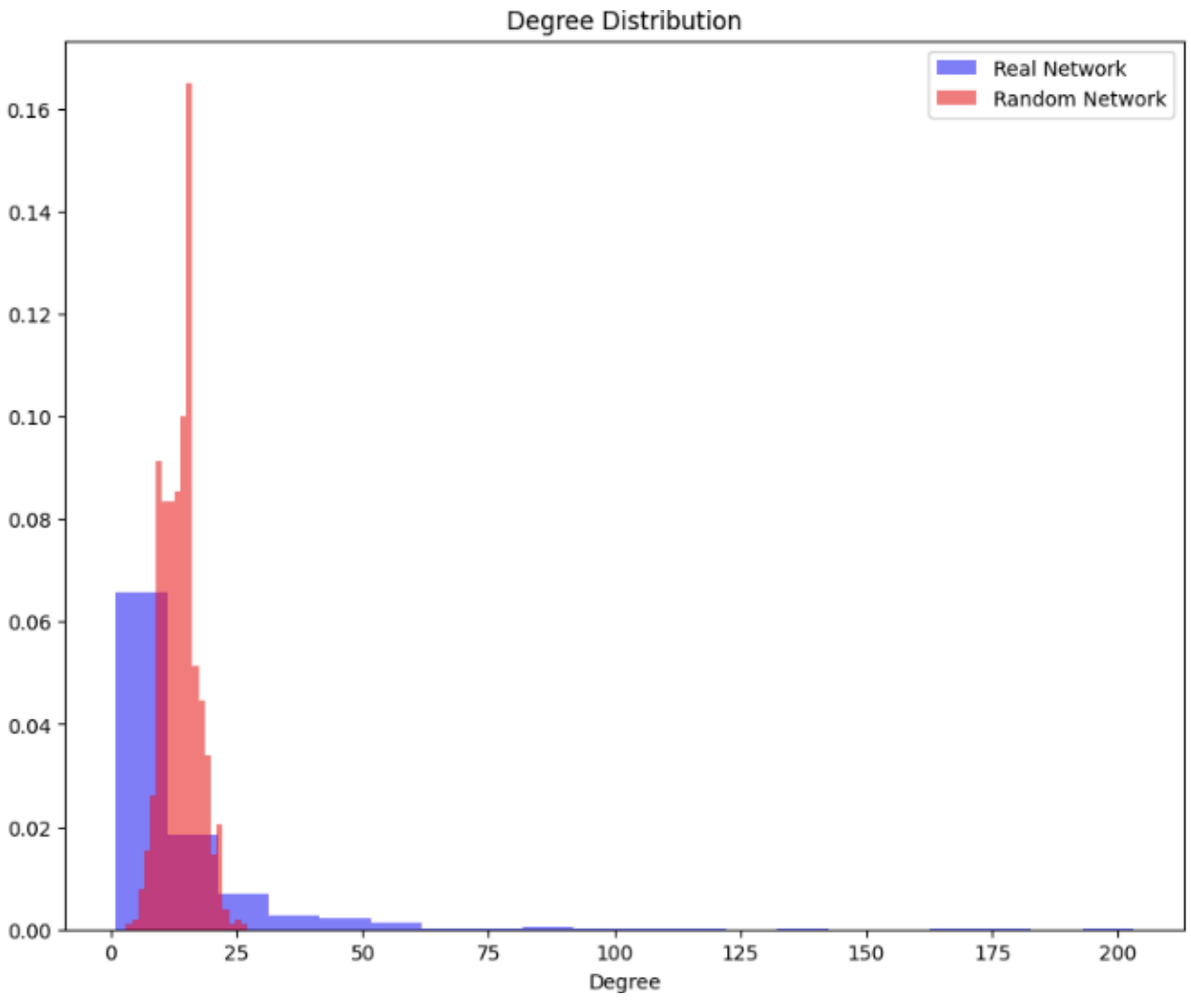


- nodes: 859 , edges: 5 807
- avg. degree: 13.52 (sd = 17.96)
- density: 0.0158
- connected components: 5
- bridges: 38 (0.006 %)



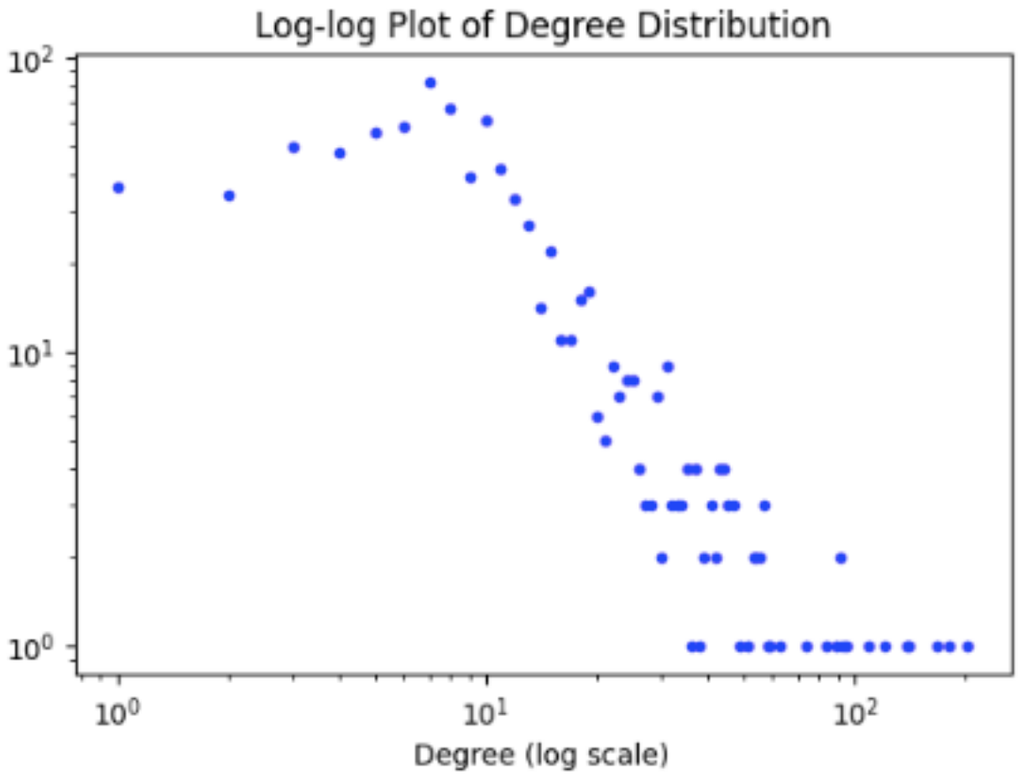
- nodes: 859
- edges: 5 904
- avg. degree: 13.75
  (sd = 3.5)
- connected components: 1

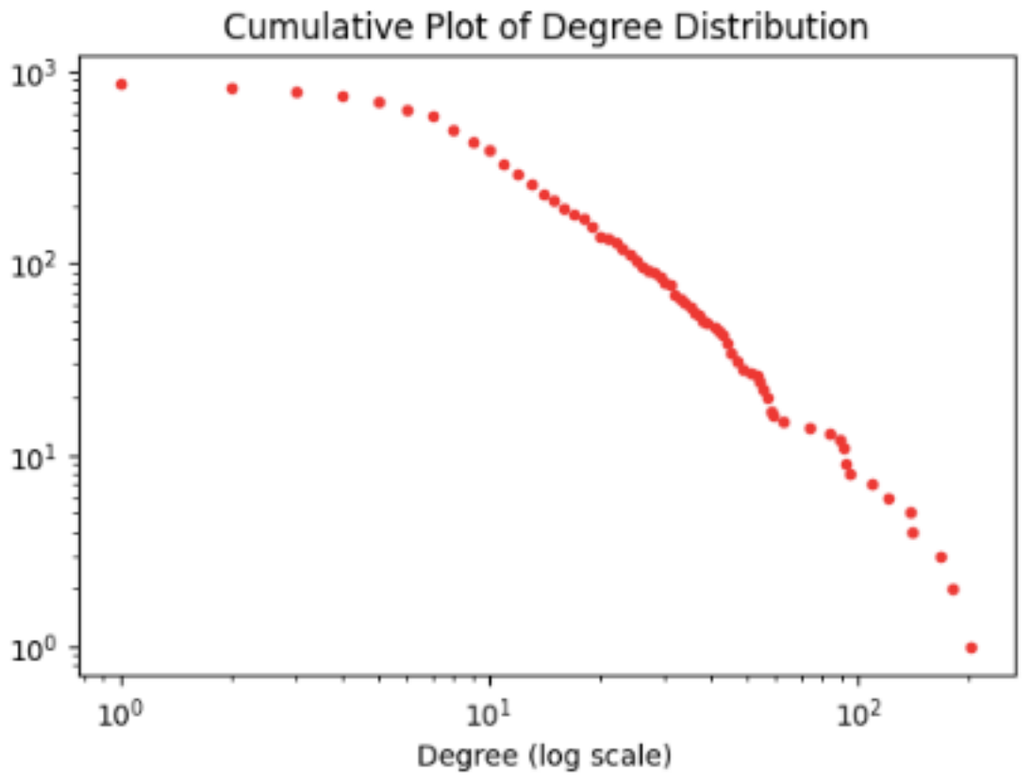
real VS random



As expected, the degree distribution of real network reminds **Power Law**, while random network recalls **Poisson distribution**

just real


Log-log Plot of Degree Distribution
10^2
10^1
10^0
10^0
10^1
10^2
Degree (log scale)


Cumulative Plot of Degree Distribution
10^3
10^2
10^1
10^0
10^0
10^1
10^2
Degree (log scale)

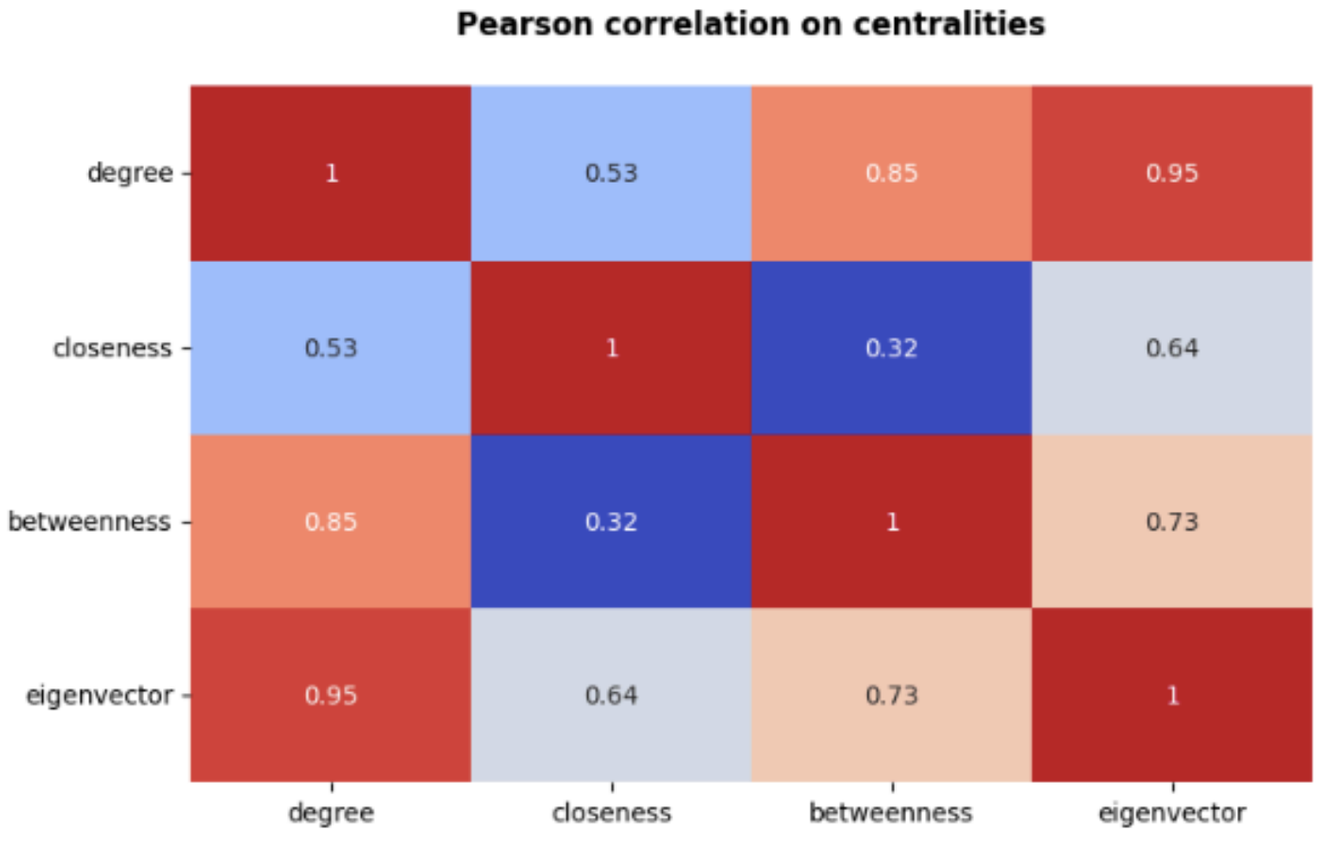
the most central is

Degree:

Closeness:

Betweenness:

Eigenvector:



👉connectivity👈

The graph is not connected, there are 5 connected components in the network.

The largest component (99% of the network):

– nodes: 840

– edges: 5 783

– Avg. shortest length path: 3.14

– diameter: 9

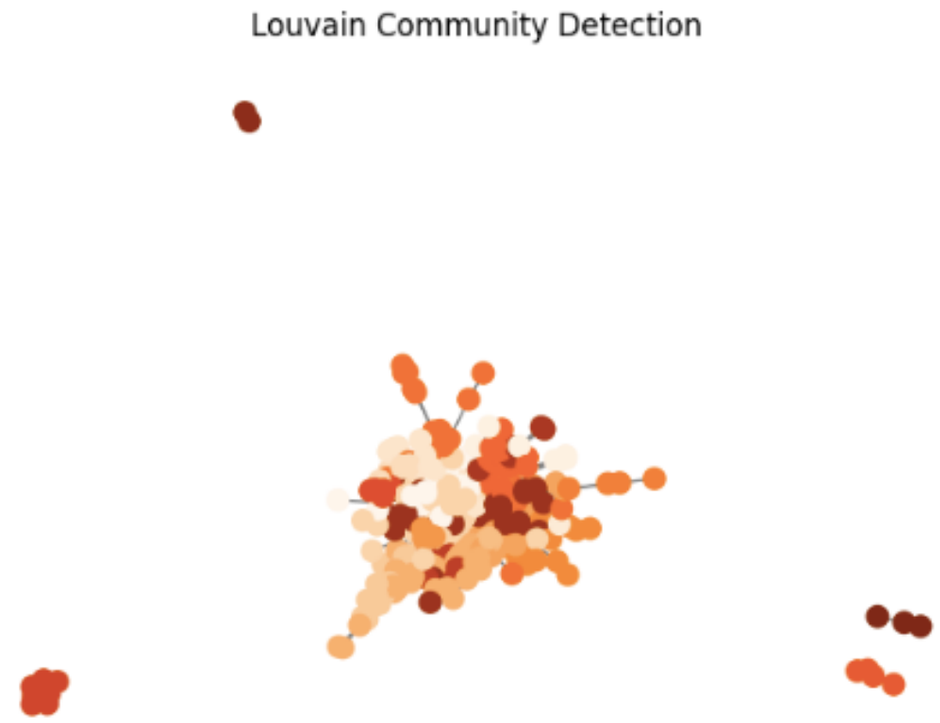
Insight: this subgraph of competitors is somehow connected and the information flows pretty fast (through ~ 4 links between different companies).

For instance, the information of company's innovation can reach the competitor just in 4 steps.

transitivity & CCs

Unexpectedly, companies tend to be competitors with dissimilar ones (assortativity -0.027), and companies slightly tend to create the cluster of competitors (gcc 0.155, avg lcc 0.31)

Network Modularity = 0.367
communities = 20



Network Modularity = 0.39
communities = 23

Who might be my competitor?

Based on structural equivalence measures, the link prediction model has been applied (logistic regression)

Interpretation: with increase in Jaccard distances, RAI, AAI, Pref (separately) the probability of being competitors between 2 companies increases (by 8.6, 16.56, 9.56 , 81.87)

Examples: Canon vs Panasonic; Intel vs inContact; Intel vs LG

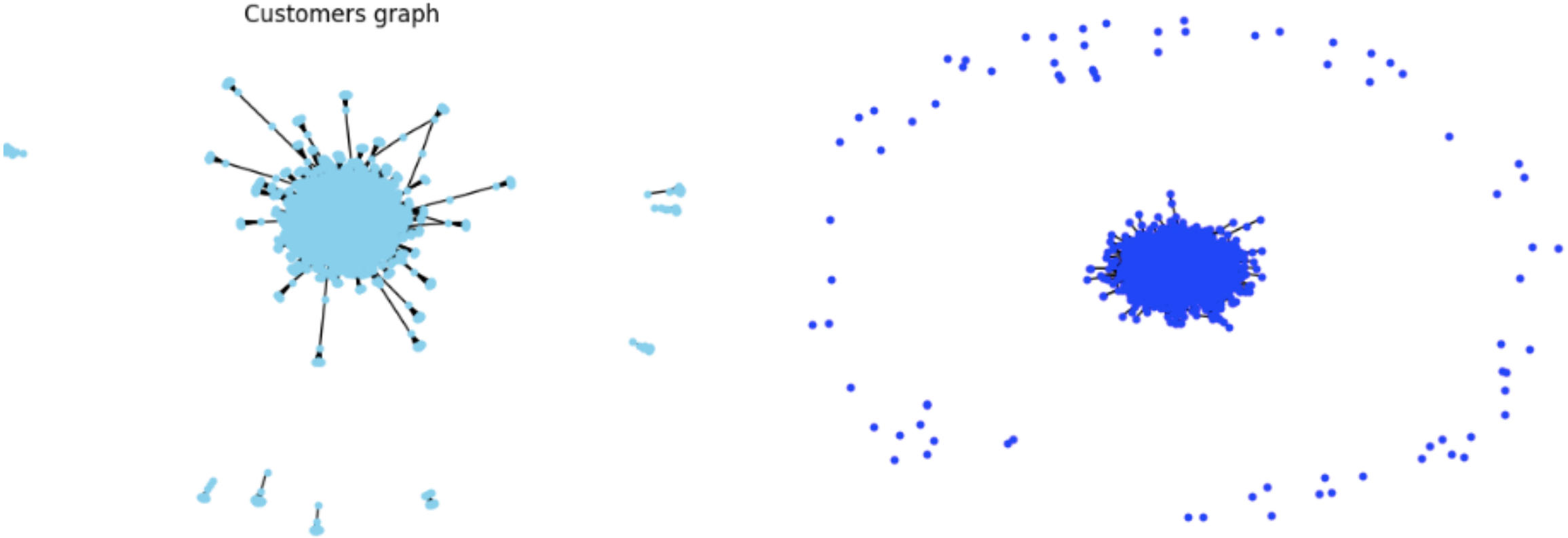
Model quality: accuracy 80%, predicted 74% of non-competitors classified are non-competitors and 88% for competitors;

90% of non competitors and 70% of competitors are classified correctly.

customers network

The initial graph with 19653 nodes and 77601 edges has been subsampled
(companies with >=9 and <= 15 customers)

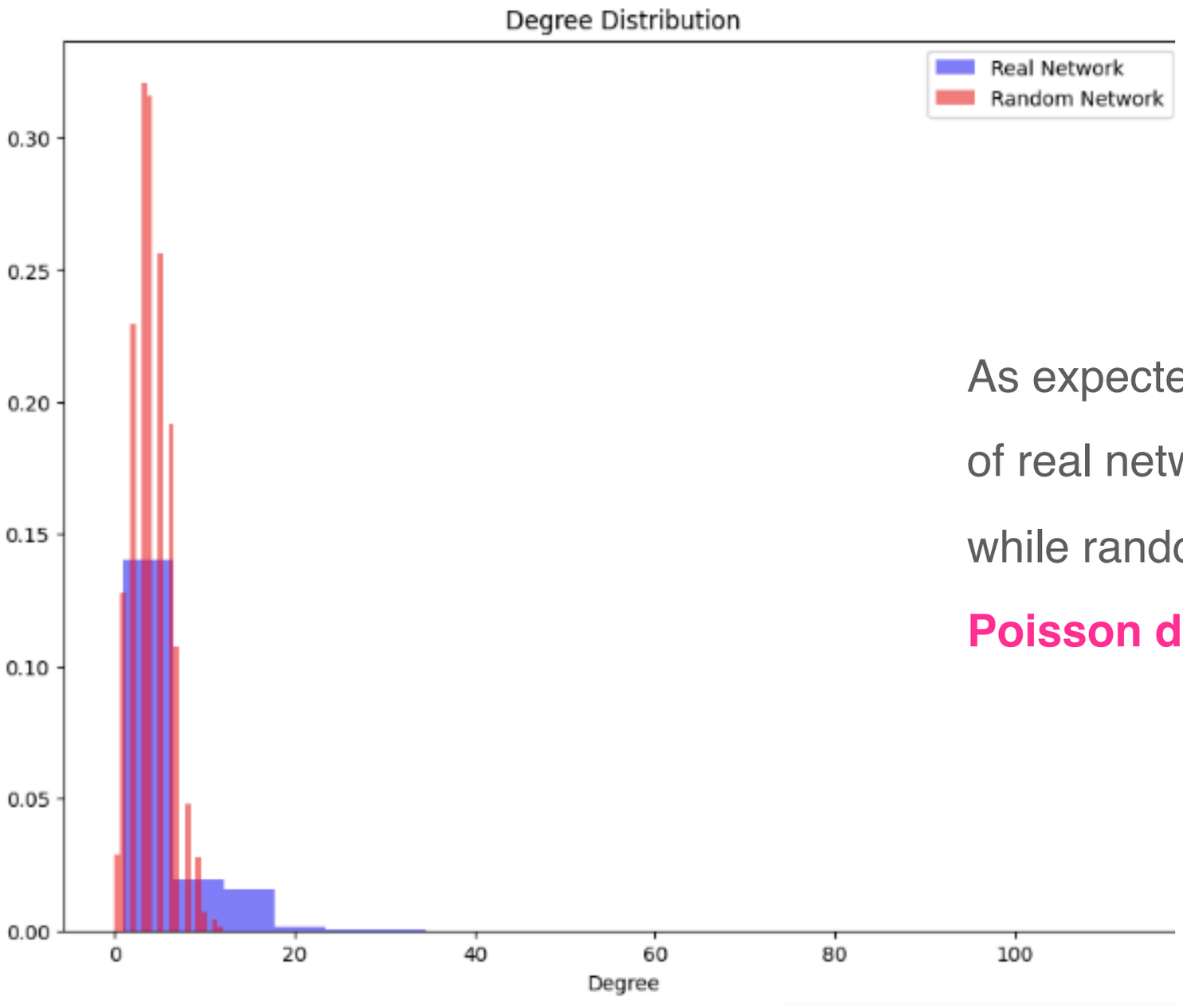
real VS random



- nodes: 3 762 , edges: 7 600
- avg. degree: 4.04 (sd = 6.22)
- density: 0.0011
- connected components: 9
- bridges: 29% of the network

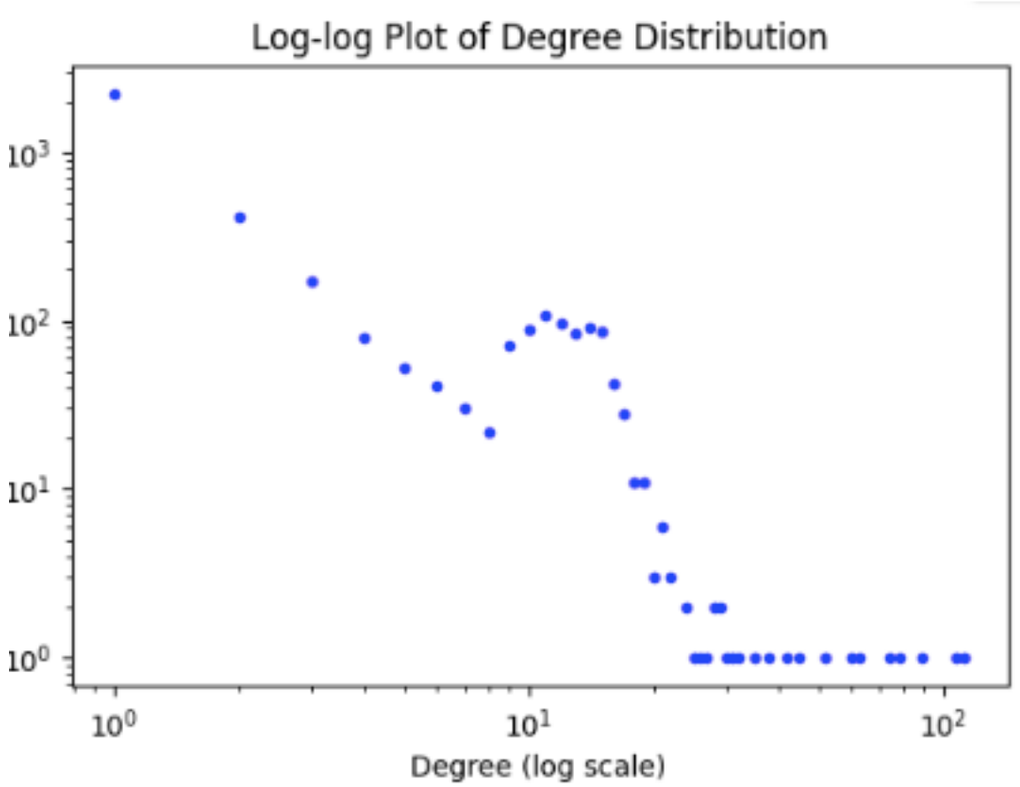
- edges: 7635
- avg. degree: 4.05
  (sd = 2.02)
- connected components: 71

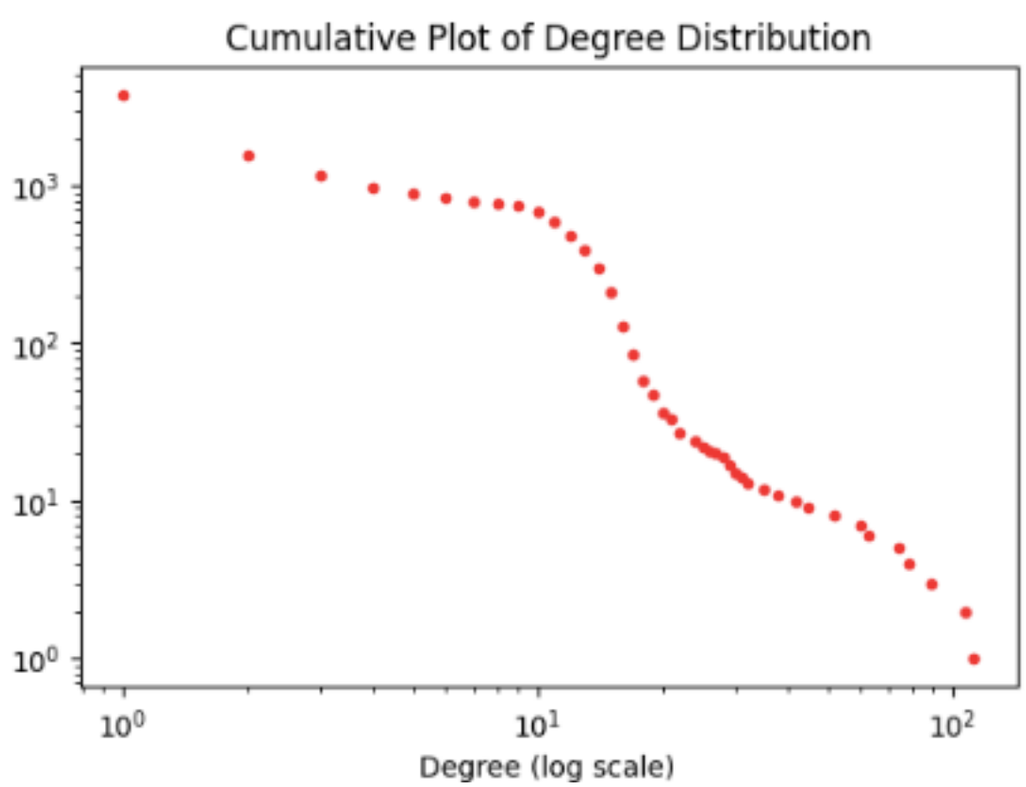
real VS random



As expected, the degree distribution of real network reminds **Power Law**, while random network recalls **Poisson distribution**

just real


Log-log Plot of Degree Distribution
10^3
10^2
10^1
10^0
10^0
10^1
10^2
Degree (log scale)


Cumulative Plot of Degree Distribution
10^3
10^2
10^1
10^0
10^0
10^1
10^2
Degree (log scale)

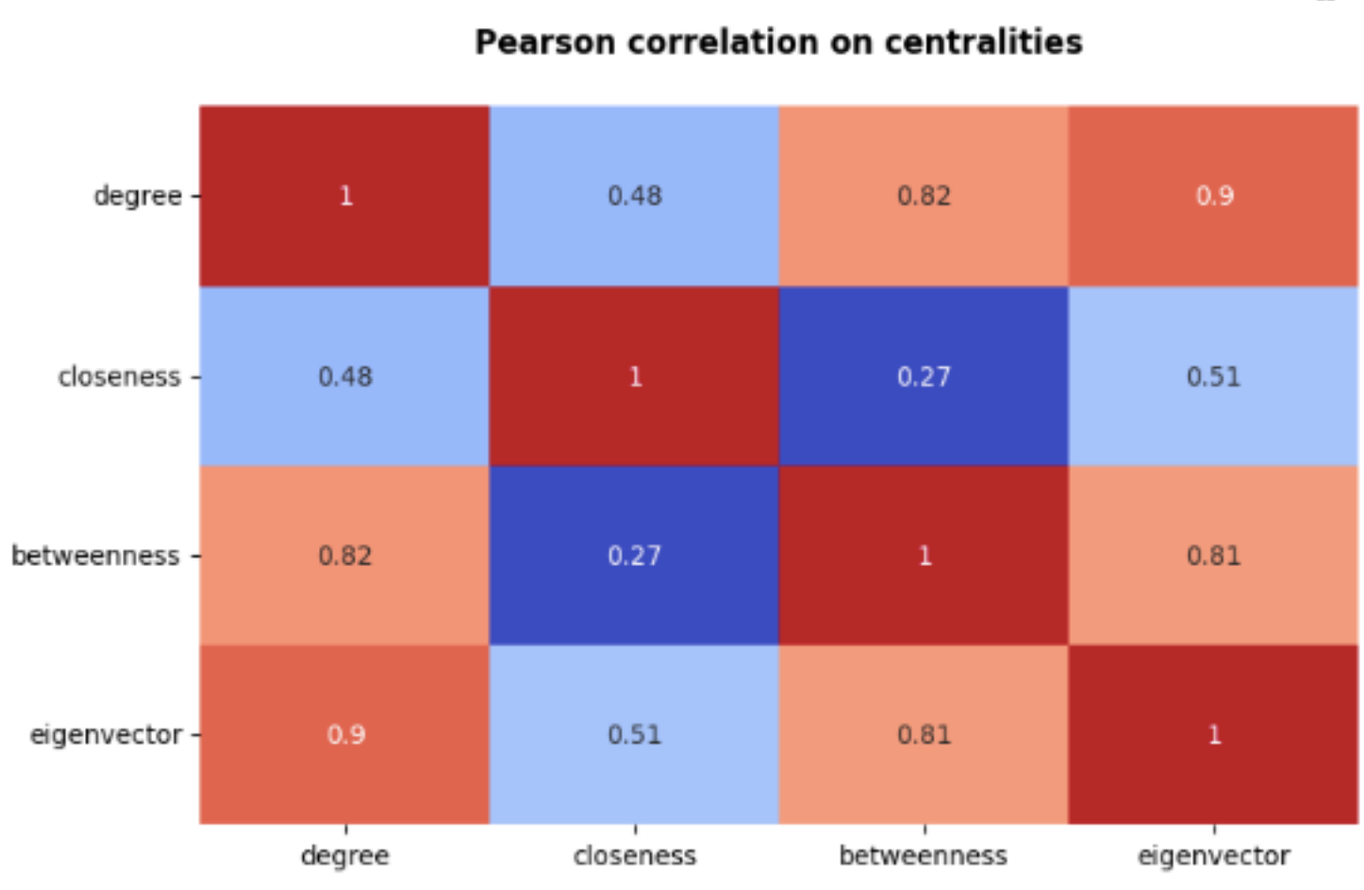
the most central is

Degree:

Closeness:

Betweenness:

Eigenvector:



👉connectivity👈

The graph is not connected, there are 9 connected components in the network.

The largest component:

– nodes: 3 658

– edges: 7 504

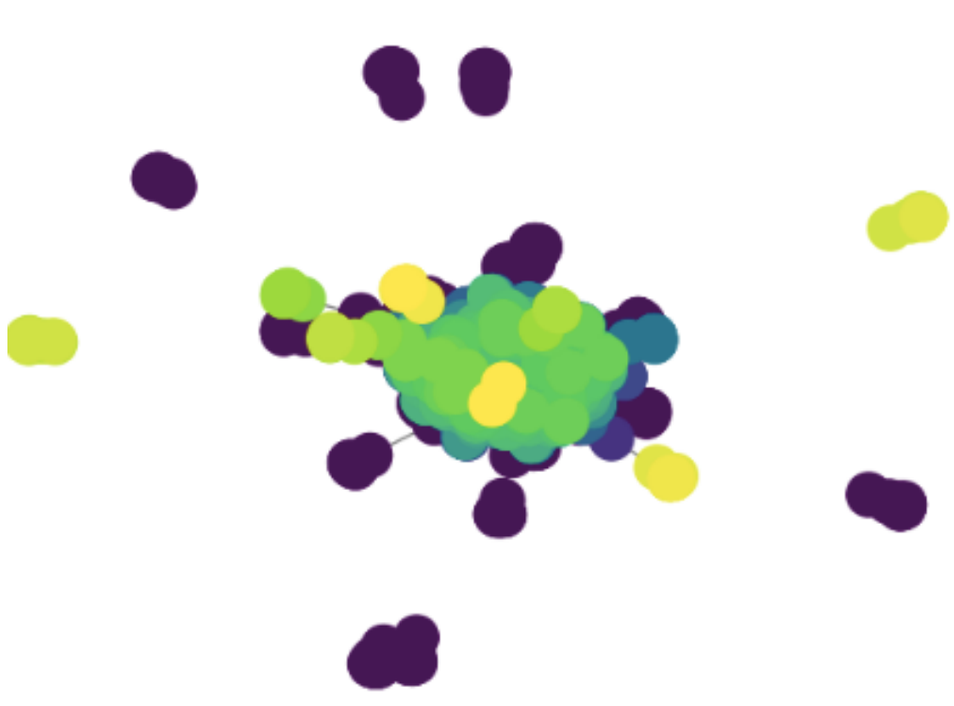
– Avg. shortest length path: 4.72

Insight: in this customers' subgraph the information flows via 5 links between different companies.

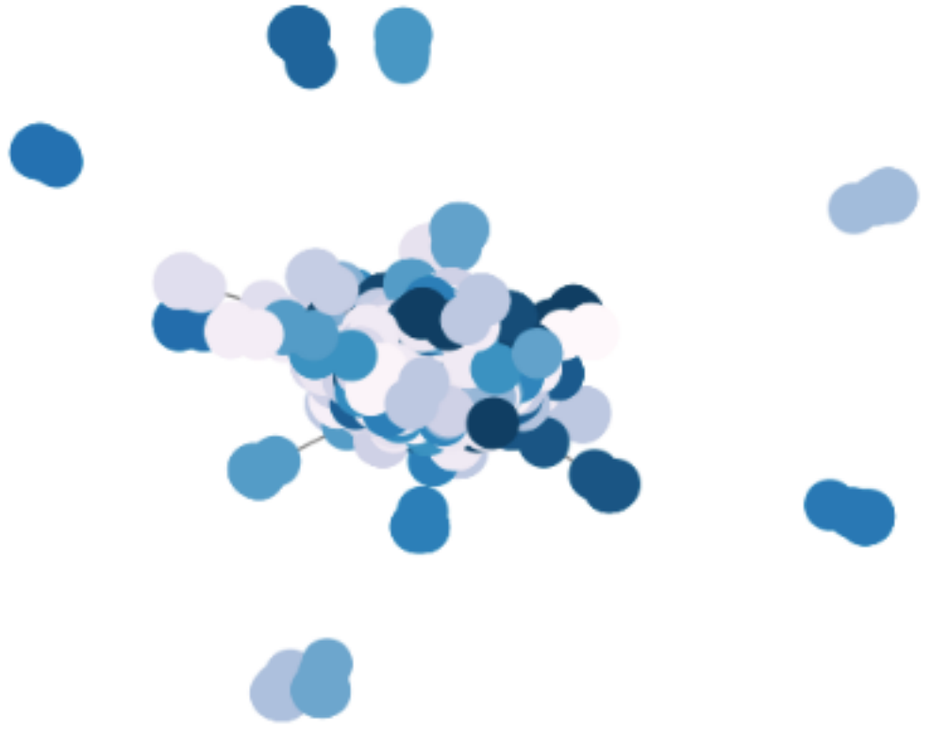
For instance, the information of company's new product can reach the customer in 5 steps.

transitivity & CCs

Companies tend to be customers of dissimilar companies, probably, not-competitors (assortativity ~0), and companies don't tend to create the cluster of customers (gcc 0.007, avg lcc 0.005)



Network Modularity = 0.534
 # communities = 42



Network Modularity = 0.537
 # communities = 40

Who might be my customer?

Based on structural equivalence measures, the link prediction model has been applied (logistic regression)

Interpretation: with increase in Jaccard distances, RAI, AAI, Pref (separately) the probability of being customer between 2 companies increases (by -0.62, 18.69, -15.28, 60.81)

Examples: Citibank & Xiotech; WeChat & Bell Mobility; GoPro & IEC Electronics

Model quality: accuracy 75%, predicted 69% of non-customers classified are non-customers and 87% for customers;
92% of non-customers and 58% of customers are classified correctly.

TBD

- Optimal companies (sample) selection: choose particular sphere(s) for detailed analysis

- Weight network (~ company profit) & add attributes

- Clustering method including possible other attributes