Real Data Analysis

Emanuel Bîscă
Denisa Buzan
Răzvan Crăciun
Ciprian Cuc

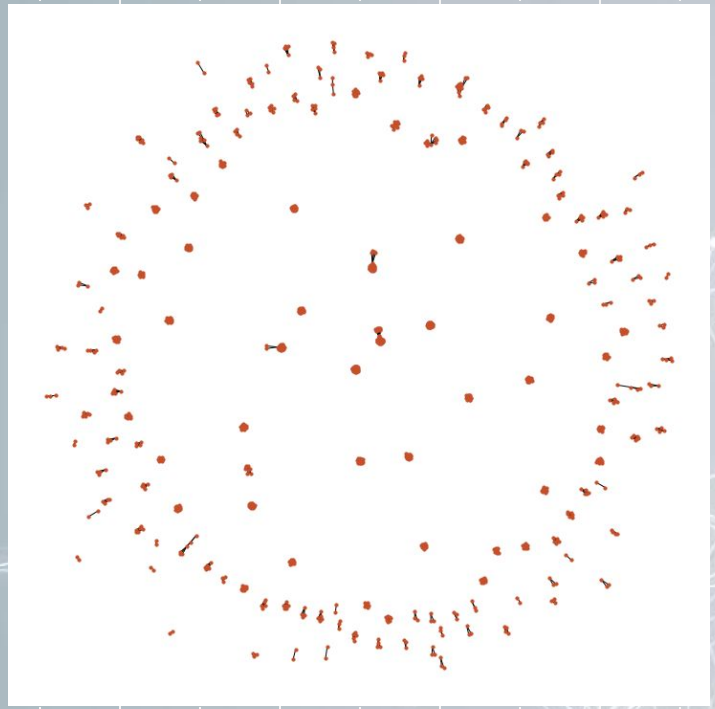
Social Network Analysis
Applied Computational Intelligence

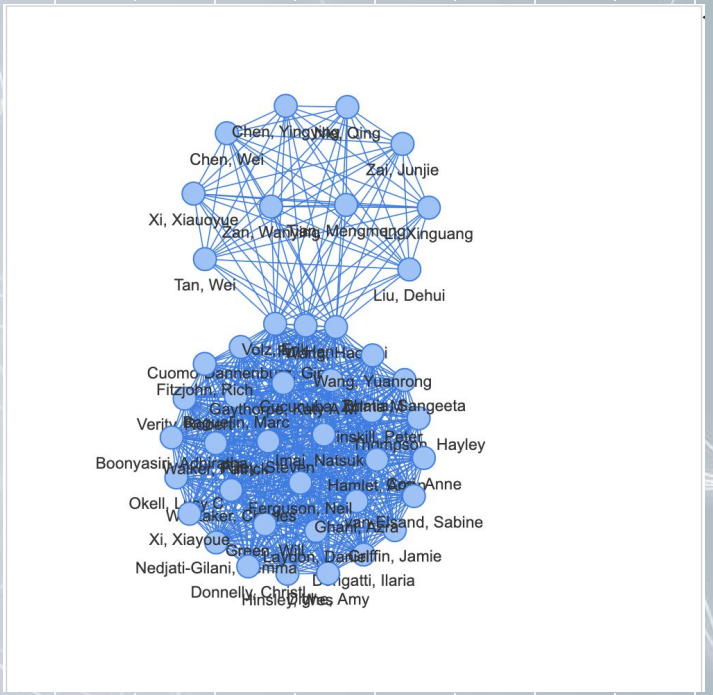
For our real data analysis, we chose to study a collaboration network of scientists

- There are **955** nodes,
- Linked with **4590** edges,
- And **150** connected components.

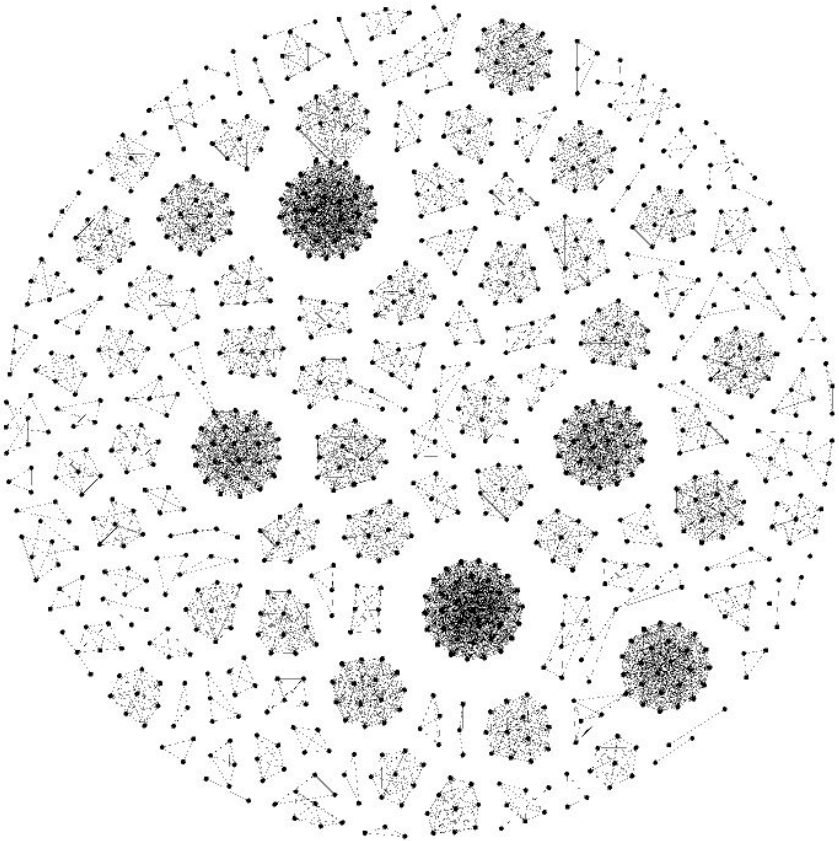
Each authors is considered to be a node in our network, while the links between two nodes exist if the corresponding scientists have collaborated in at least one scientific paper.

Network visualization

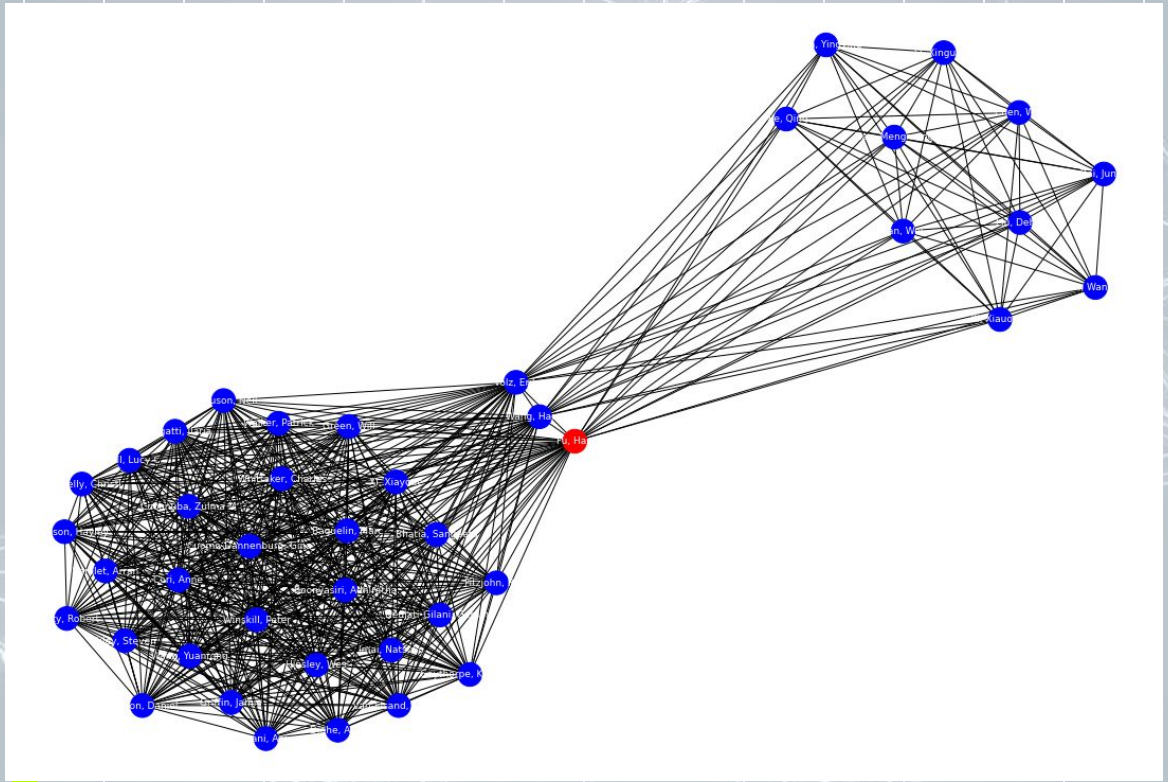




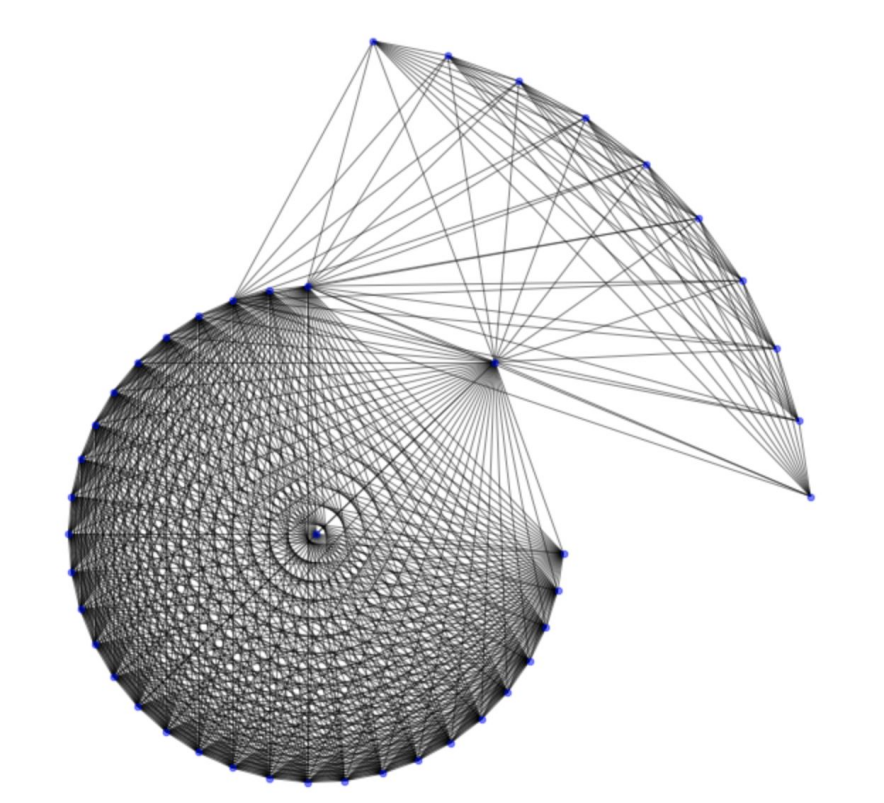
Network visualization



Network visualization - largest component



Network visualization - largest component



Network properties

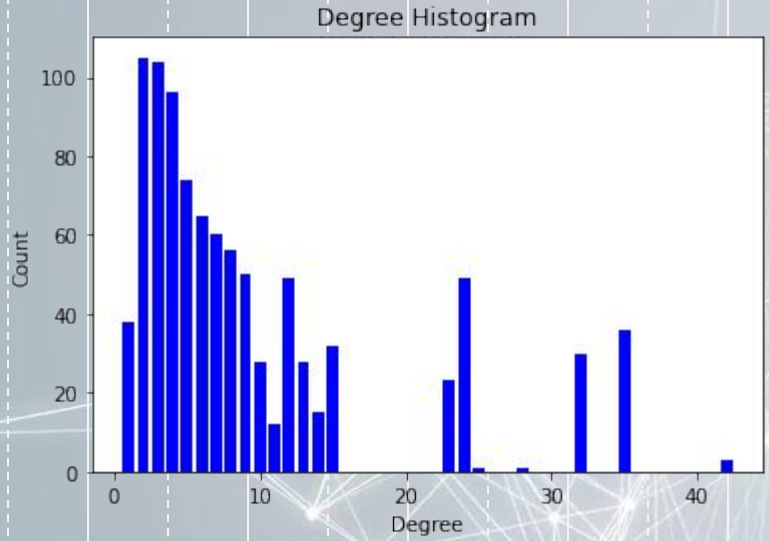
For the whole network

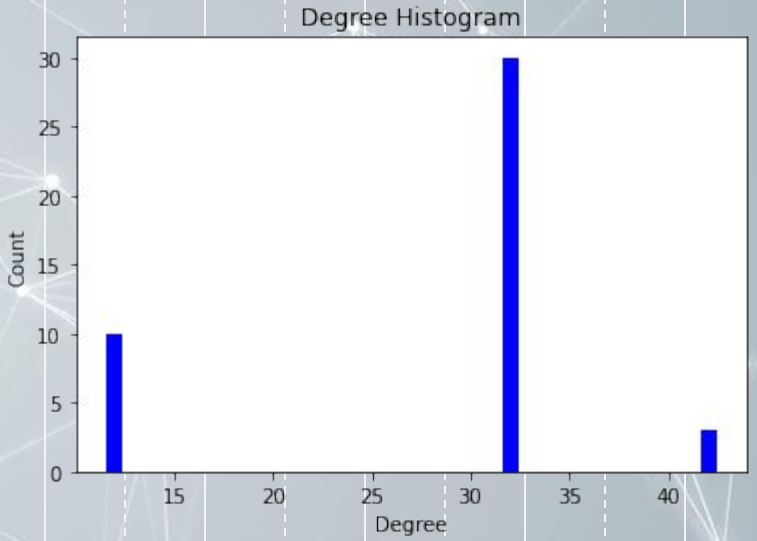
- Number of nodes = 955
- Number of links = 4590
- Average degree = 9.61
- Min degree = 1
- Max degree = 42
- Edge density = 0.01007606

For the largest component

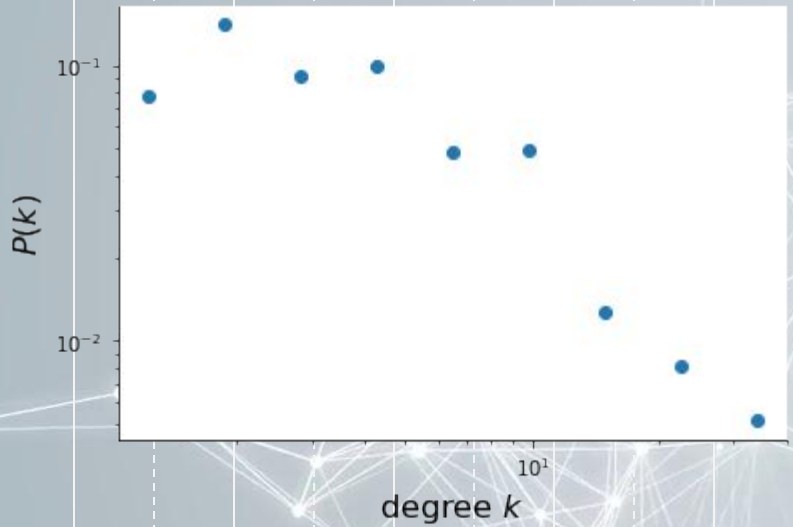
- Number of nodes = 43
- Number of links = 603
- The diameter = 2
- Shortest path length = 1.33
- Average degree = 28.04
- Min degree = 12
- Max degree = 42
- Edge density = 0.667774086378737 6

Analysis


Degree Histogram
Count
Degree
0
20
40
60
80
100
0
10
20
30
40

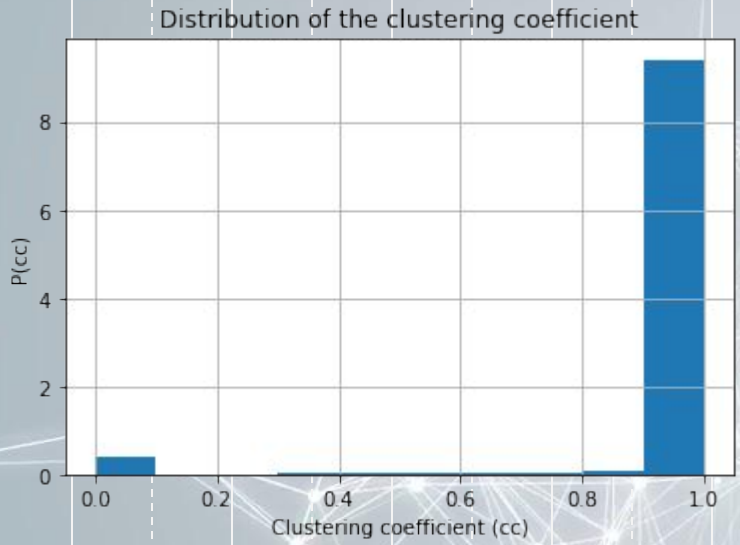

Degree Histogram
Count
Degree
0
5
10
15
20
25
30
15
20
25
30
35
40

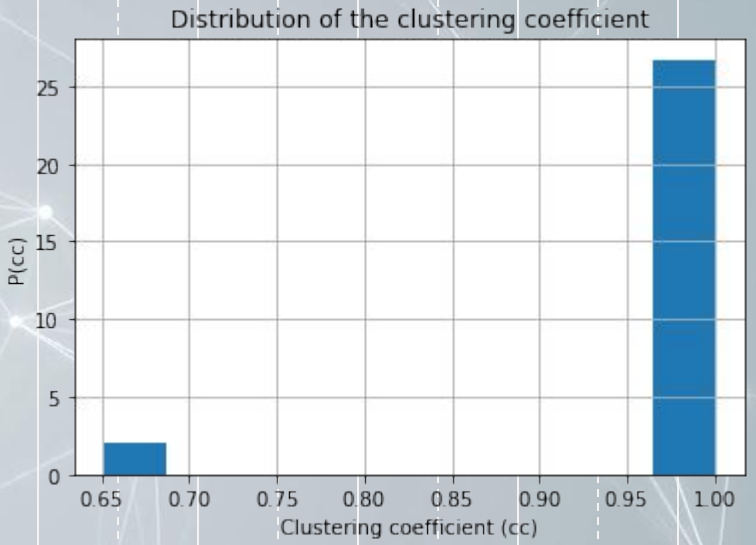
Analysis


10^{-1}
10^{-2}
$P(k)$
10^{1}
degree k

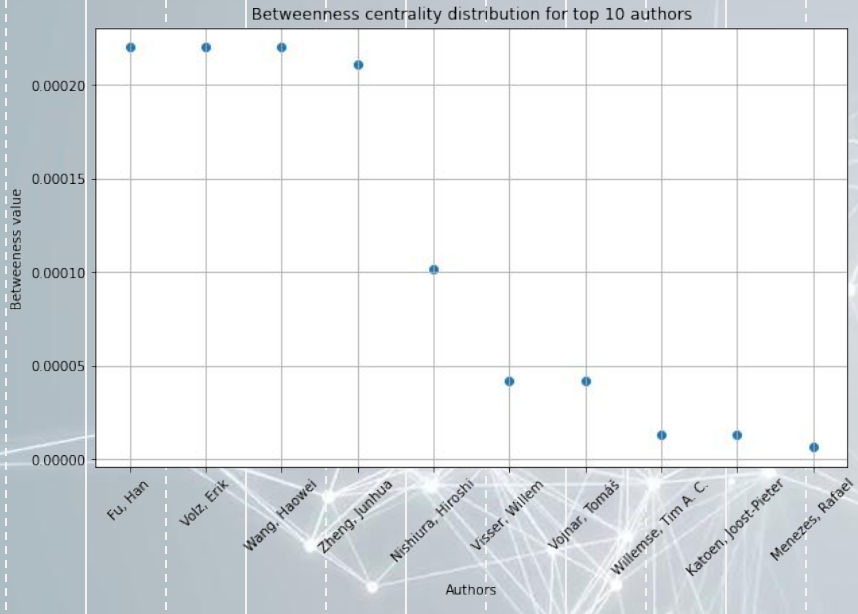

10^{-1}
$P(k)$
2×10^{1}
3×10^{1}
4×10^{1}
degree k

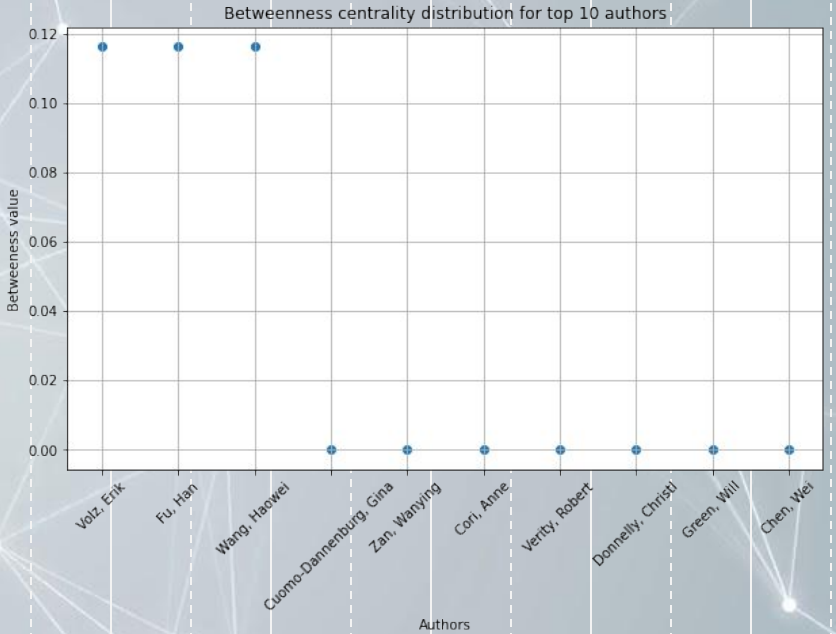
Analysis


Distribution of the clustering coefficient
P(cc)
0
2
4
6
8
0.0
0.2
0.4
0.6
0.8
1.0
Clustering coefficient (cc)

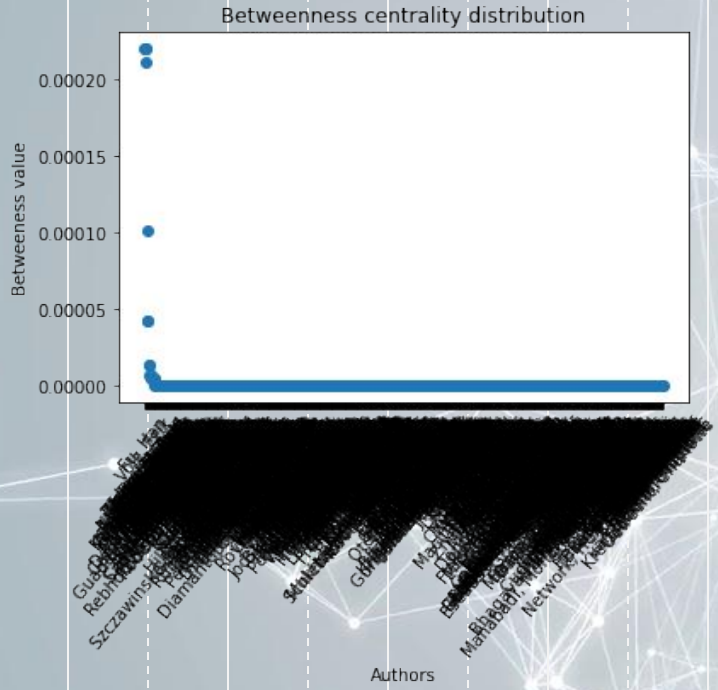

Distribution of the clustering coefficient
P(cc)
0
5
10
15
20
25
0.65
0.70
0.75
0.80
0.85
0.90
0.95
1.00
Clustering coefficient (cc)

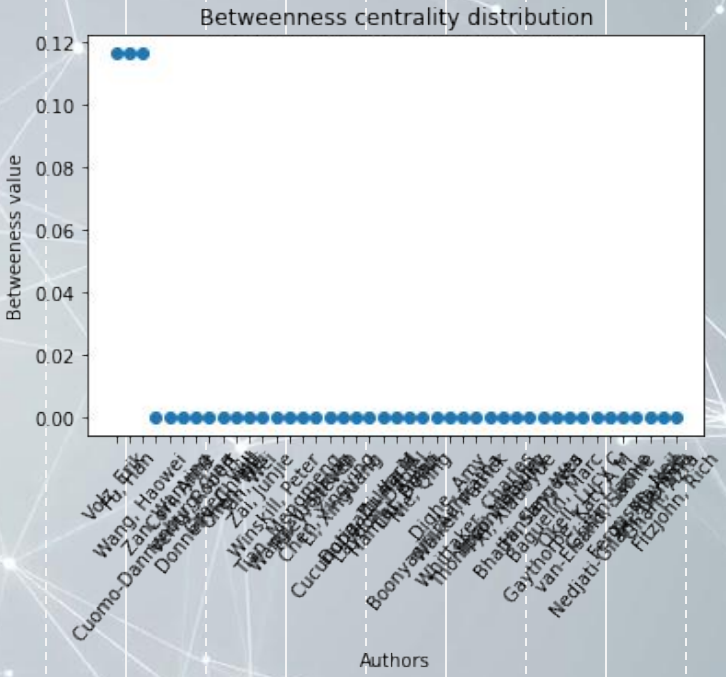
Centrality measures



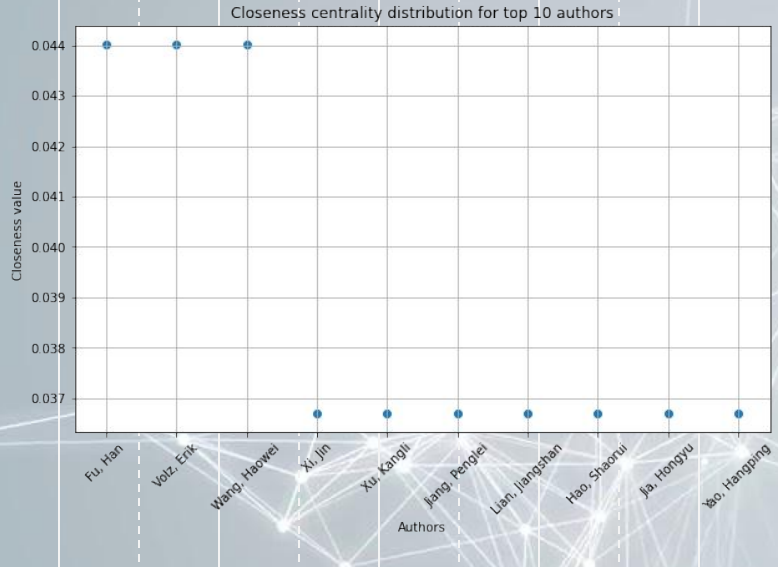


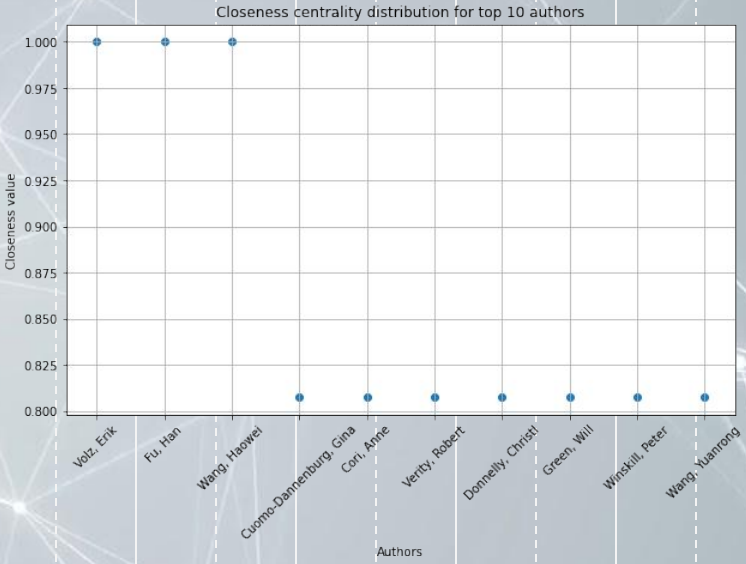
Centrality measures



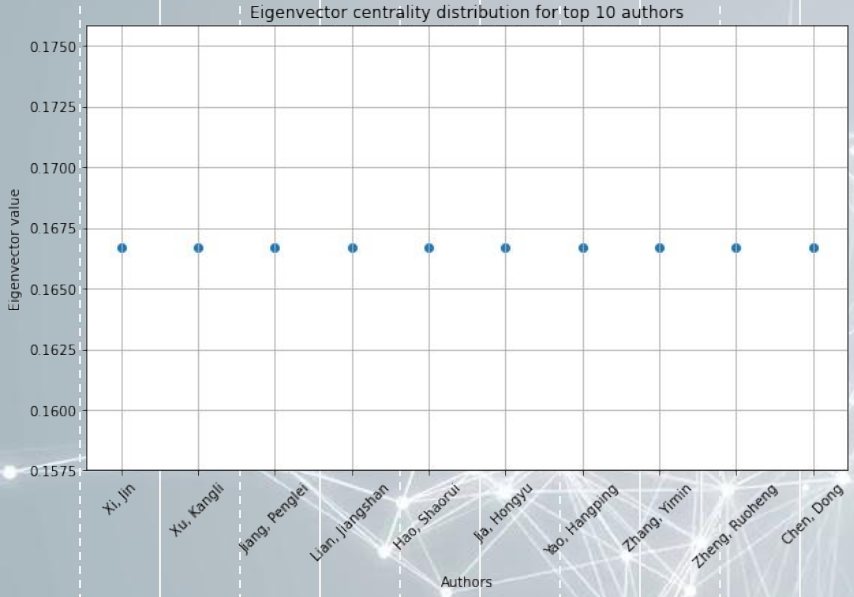


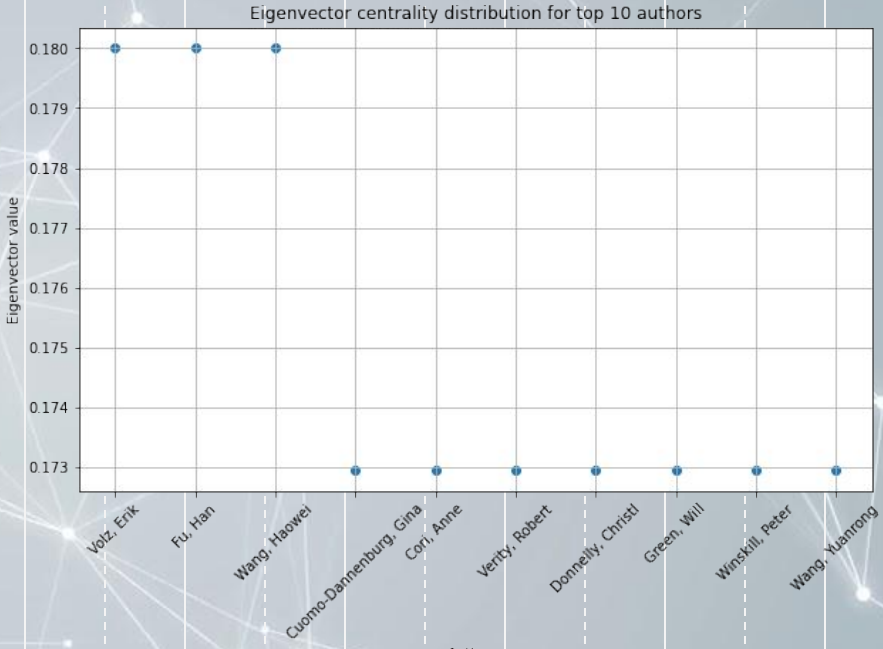
Centrality measures



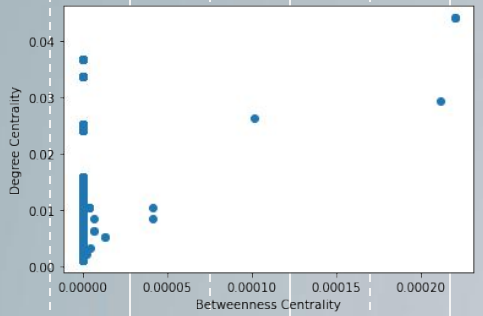


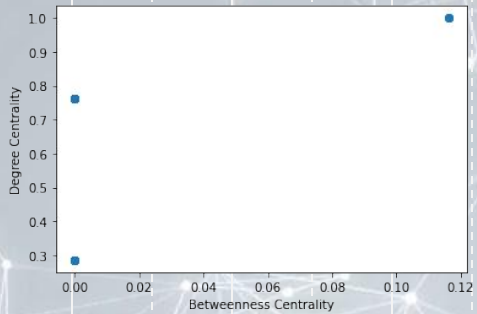
Centrality measures



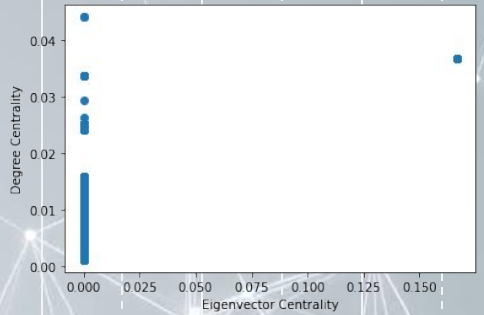


Centrality measures - comparison

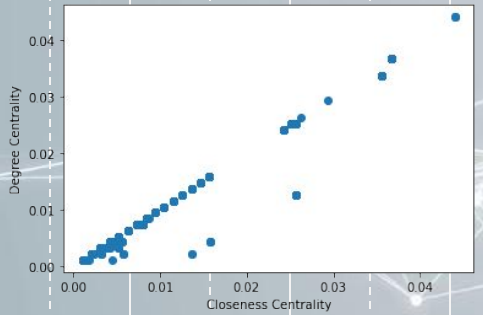




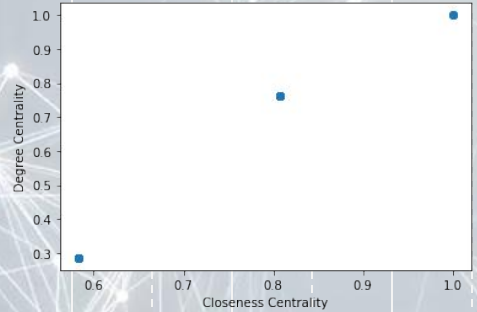
The whole network



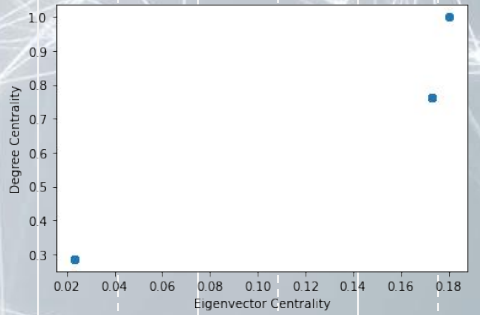
The whole network



The largest component



The largest component



The most important nodes

```
Most important 5 nodes according to degree centrality are:
('Fu, Han', 0.04402515723270441)
('Volz, Erik', 0.04402515723270441)
('Wang, Haowei', 0.04402515723270441)
('Xi, Jin', 0.03668763102725367)
('Xu, Kangli', 0.03668763102725367)
```

```
Most important 5 nodes according to betweenness centrality are:
('Fu, Han', 0.00021998279734524755)
('Volz, Erik', 0.00021998279734524755)
('Wang, Haowei', 0.00021998279734524755)
('Zheng, Junhua', 0.0002111834854514377)
('Nishiura, Hiroshi', 0.00010119208677881389)
```

```
Most important 5 nodes according to closeness centrality are:
('Fu, Han', 0.0440251572327044)
('Volz, Erik', 0.0440251572327044)
('Wang, Haowei', 0.0440251572327044)
('Xi, Jin', 0.03668763102725367)
('Xu, Kangli', 0.03668763102725367)
```

```
Most important 5 nodes according to eigenvector centrality are:
('Xi, Jin', 0.16666640626296628)
('Xu, Kangli', 0.16666640626296628)
('Jiang, Penglei', 0.16666640626296628)
('Lian, Jiangshan', 0.16666640626296628)
('Hao, Shaorui', 0.16666640626296628)
```

```
Most important 5 nodes according to degree centrality are:
('Volz, Erik', 1.0)
('Fu, Han', 1.0)
('Wang, Haowei', 1.0)
('Cuomo-Dannenburg, Gina', 0.7619047619047619)
('Cori, Anne', 0.7619047619047619)
```

```
Most important 5 nodes according to betweenness centrality are:
('Volz, Erik', 0.11614401858304306)
('Fu, Han', 0.11614401858304306)
('Wang, Haowei', 0.11614401858304306)
('Cuomo-Dannenburg, Gina', 0.0)
('Zan, Wanying', 0.0)
```

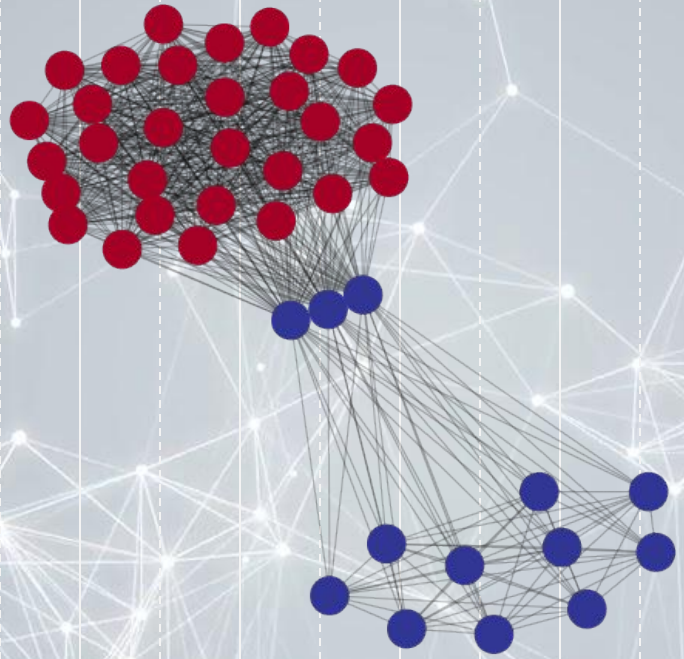
```
Most important 5 nodes according to closeness centrality are:
('Volz, Erik', 1.0)
('Fu, Han', 1.0)
('Wang, Haowei', 1.0)
('Cuomo-Dannenburg, Gina', 0.8076923076923077)
('Cori, Anne', 0.8076923076923077)
```

```
Most important 5 nodes according to eigenvector centrality are:
('Volz, Erik', 0.1800008251843376)
('Fu, Han', 0.1800008251843376)
('Wang, Haowei', 0.1800008251843376)
('Cuomo-Dannenburg, Gina', 0.17294937959892778)
('Cori, Anne', 0.17294937959892778)
```

Compute Erdosi number

```
For Volz, Erik, Erdosi number = 4
For Wang, Haowei, Erdosi number = None
For Fu, Han, Erdosi number = 5
For Walker, Patrick, Erdosi number = 5
For Cucunuba, Zulma M, Erdosi number = None
```

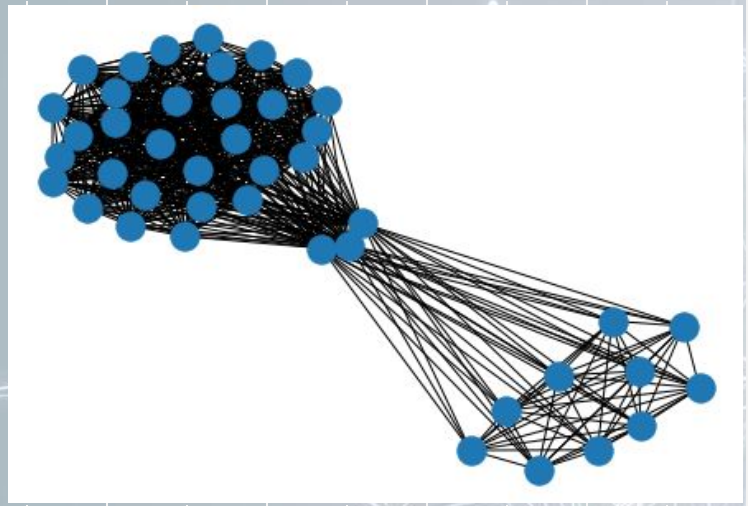
Community detection

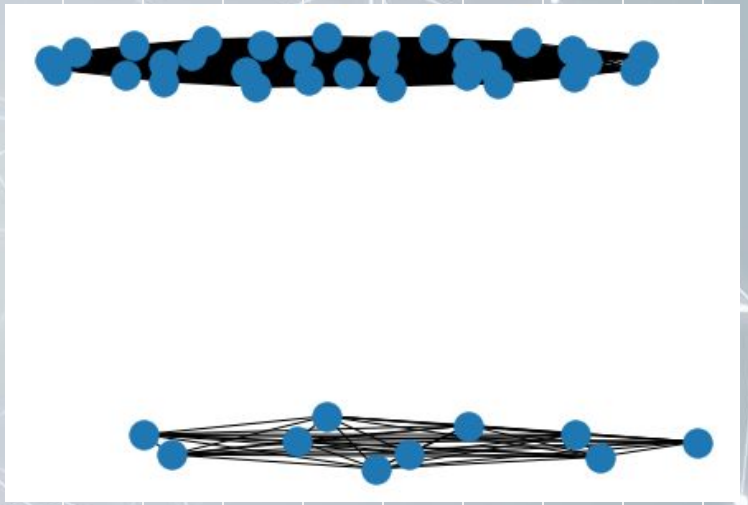


Node removal

Authors to be removed:

- Fu Han
- Wang Haowei
- Volz Erik





Node removal

Degree centrality

```
[('Cuomo-Dannenburg, Gina', 0.7435897435897436),
 ('Cori, Anne', 0.7435897435897436),
 ('Verity, Robert', 0.7435897435897436),
 ('Donnelly, Christl', 0.7435897435897436),
 ('Green, Will', 0.7435897435897436),
 ('Winskill, Peter', 0.7435897435897436),
 ('Wang, Yuanrong', 0.7435897435897436),
 ('Riley, Steven', 0.7435897435897436),
 ('Cucunuba, Zulma M', 0.7435897435897436),
 ('Dorigatti, Ilaria', 0.7435897435897436)]
```

Closeness centrality

```
[('Cuomo-Dannenburg, Gina', 0.7435897435897436),
 ('Cori, Anne', 0.7435897435897436),
 ('Verity, Robert', 0.7435897435897436),
 ('Donnelly, Christl', 0.7435897435897436),
 ('Green, Will', 0.7435897435897436),
 ('Winskill, Peter', 0.7435897435897436),
 ('Wang, Yuanrong', 0.7435897435897436),
 ('Riley, Steven', 0.7435897435897436),
 ('Cucunuba, Zulma M', 0.7435897435897436),
 ('Dorigatti, Ilaria', 0.7435897435897436)]
```

Betweenness centrality

```
[('Cuomo-Dannenburg, Gina', 0.0),
 ('Zan, Wanying', 0.0),
 ('Cori, Anne', 0.0),
 ('Verity, Robert', 0.0),
 ('Donnelly, Christl', 0.0),
 ('Green, Will', 0.0),
 ('Chen, Wei', 0.0),
 ('Tan, Wei', 0.0),
 ('Zai, Junjie', 0.0),
 ('Winskill, Peter', 0.0)]
```

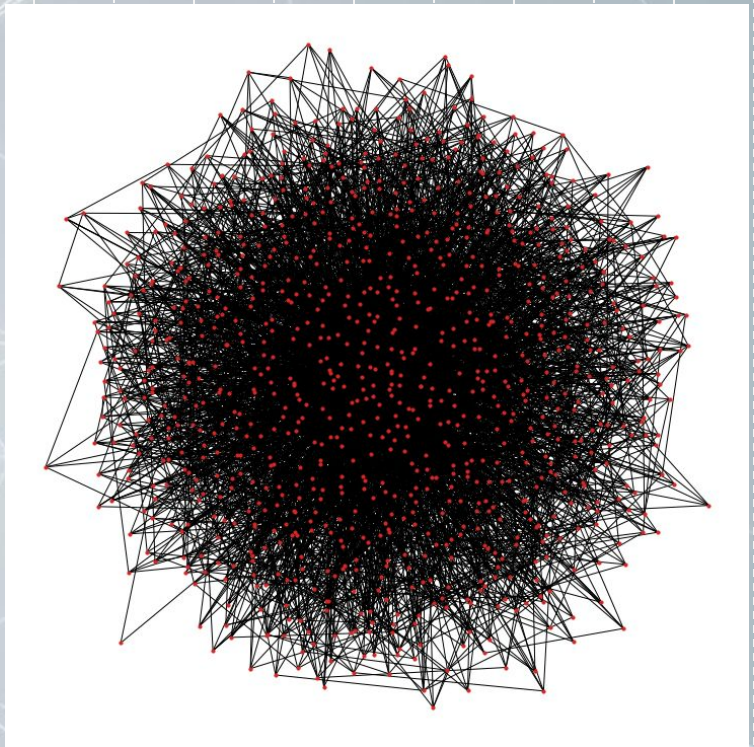
Eigenvector centrality

```
[('Cuomo-Dannenburg, Gina', 0.18257418583408572),
 ('Cori, Anne', 0.18257418583408572),
 ('Verity, Robert', 0.18257418583408572),
 ('Donnelly, Christl', 0.18257418583408572),
 ('Green, Will', 0.18257418583408572),
 ('Winskill, Peter', 0.18257418583408572),
 ('Wang, Yuanrong', 0.18257418583408572),
 ('Riley, Steven', 0.18257418583408572),
 ('Cucunuba, Zulma M', 0.18257418583408572),
 ('Dorigatti, Ilaria', 0.18257418583408572)]
```

Scale free network

Properties

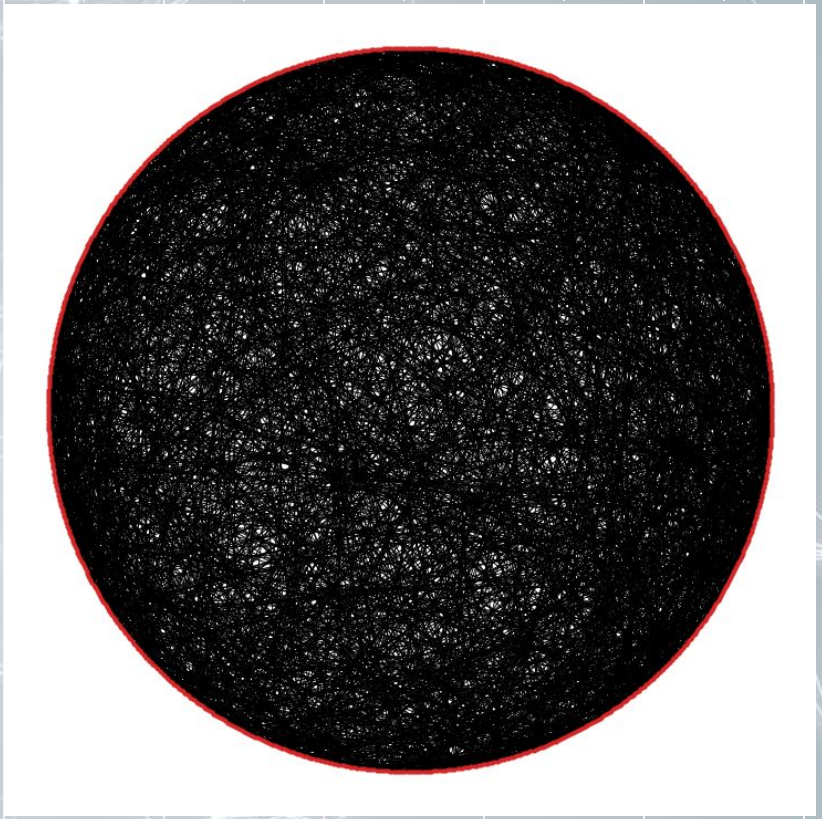
- Number of nodes = **955**
- Number of links = **4750**
- Average degree = **9.947**
- Min degree = **2**
- Max degree = **111**
- Edge density = **0.01042**



Random network

Properties

- Number of nodes = **955**
- Number of links = **2291**
- Average degree = **4.79**
- Min degree = **0**
- Max degree = **13**
- Edge density = **0.005**



THANKS!

Any questions?