

Social network analysis

Motahhareh Nadimi

PrinFull
CisFull
WOSFull
network



INTRODUCTION

In this project I will analyze the 3 network data.

CisFull network
PrinFull network
WOSFull network

Node count and edge cont per network

1525

2839

5291

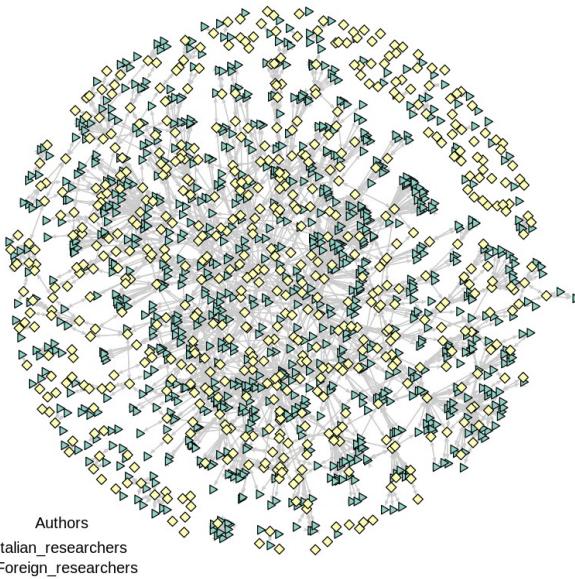
2534

9379

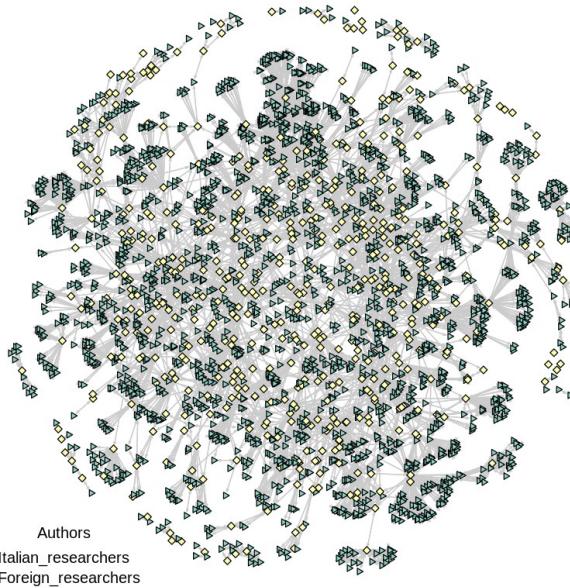
427238

Visualizing large graph layout

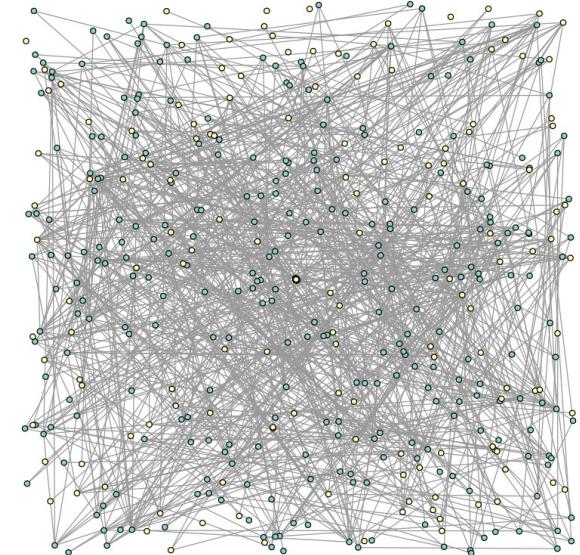
Current Index to Statistics Database Network-CIS



Current Index to Statistics Database Network-PrinFull

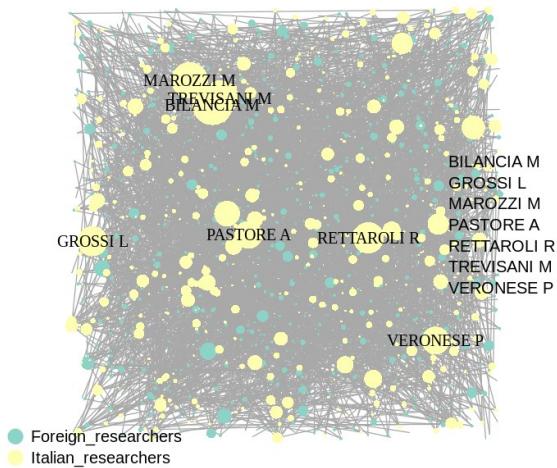


Current Index to Statistics Database Network-WOSFull

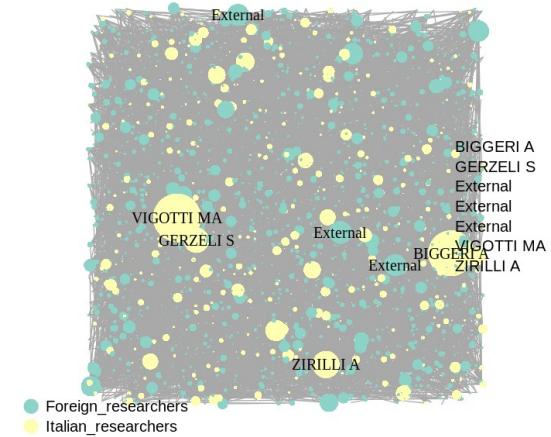


Finding high degree node in graph

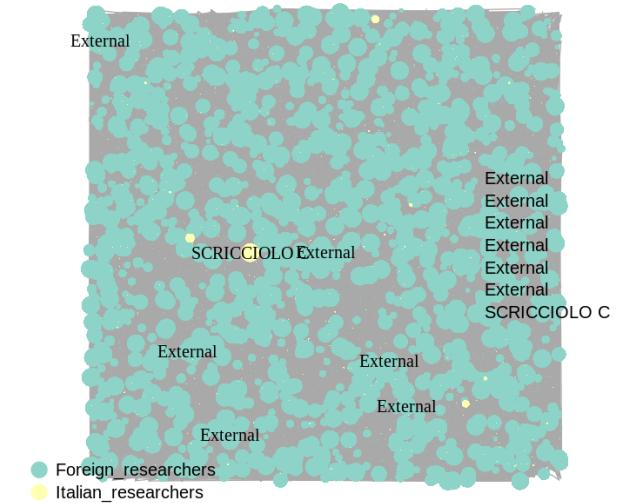
CIS Network-high degree centrality



PRIN Network-high degree centrality



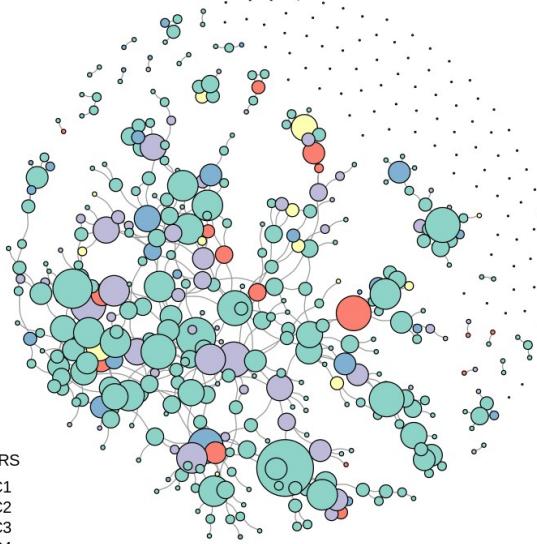
WOS Network-High degree centrality



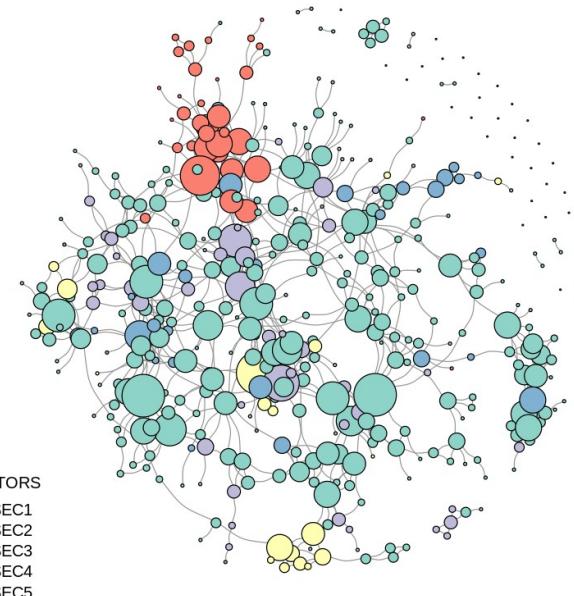
Visualizing based on the Italian researchers

The size of nodes is proportional to degree, most of vertices belongs to sector one, and for other sectors it seems that they are sparse graphs.

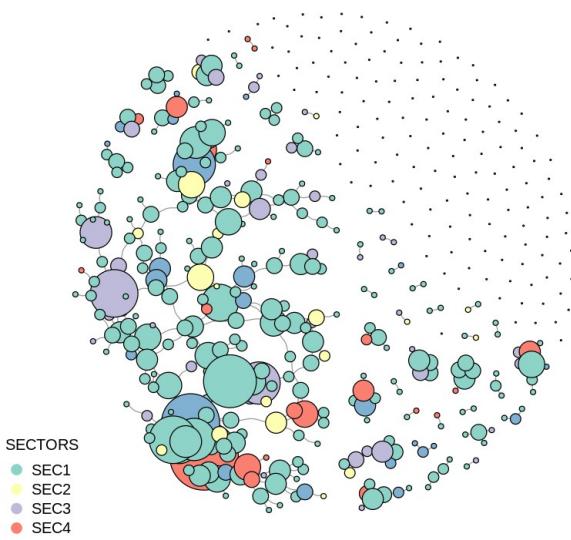
Italian researchers collaboration-CIS



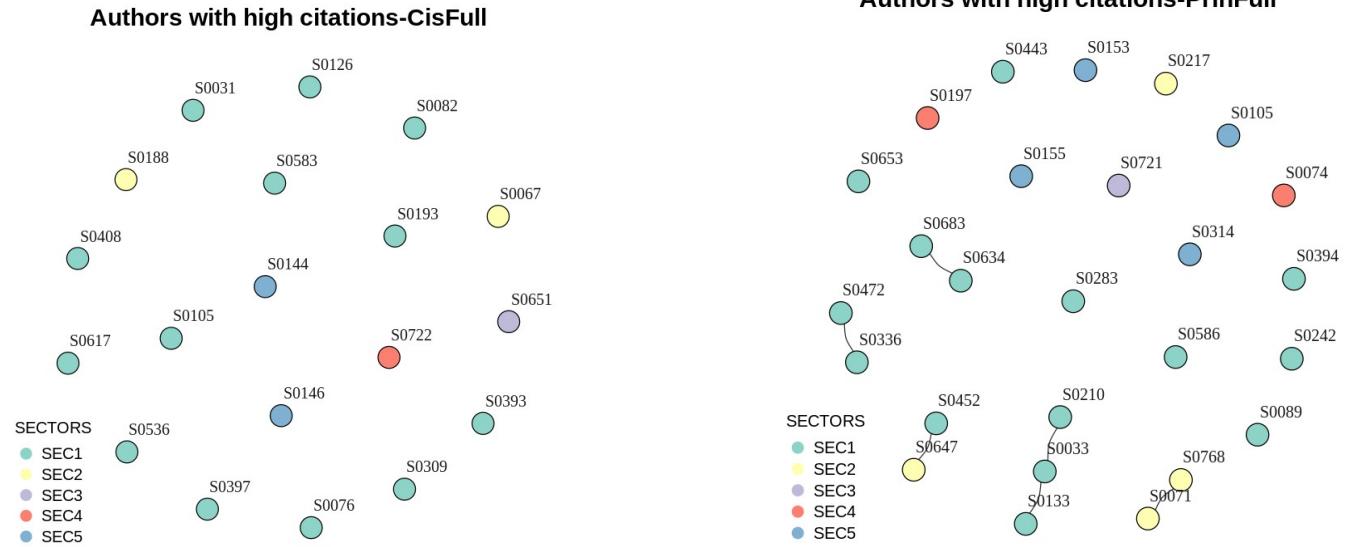
Italian researchers collaboration-PRIN



Italian researchers collaboration-WOS



Visualizing Authors with high citations



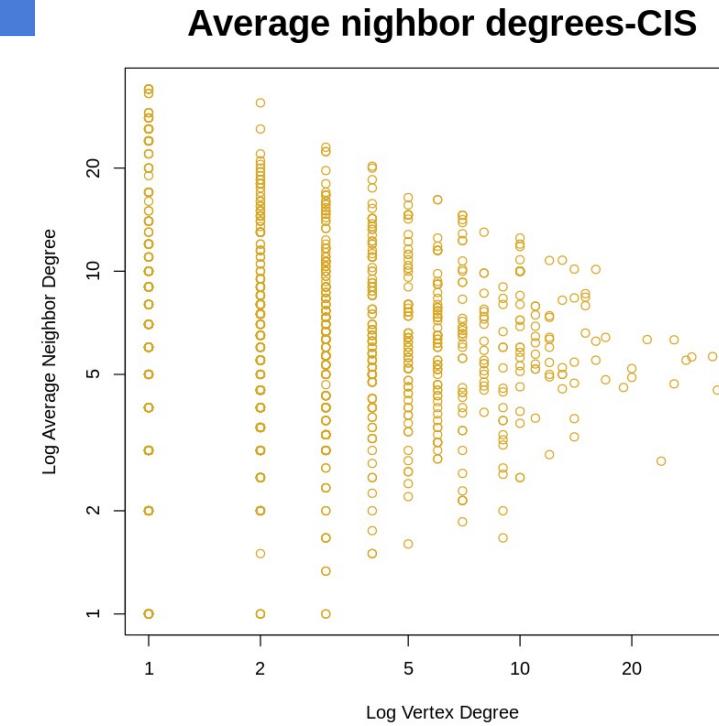
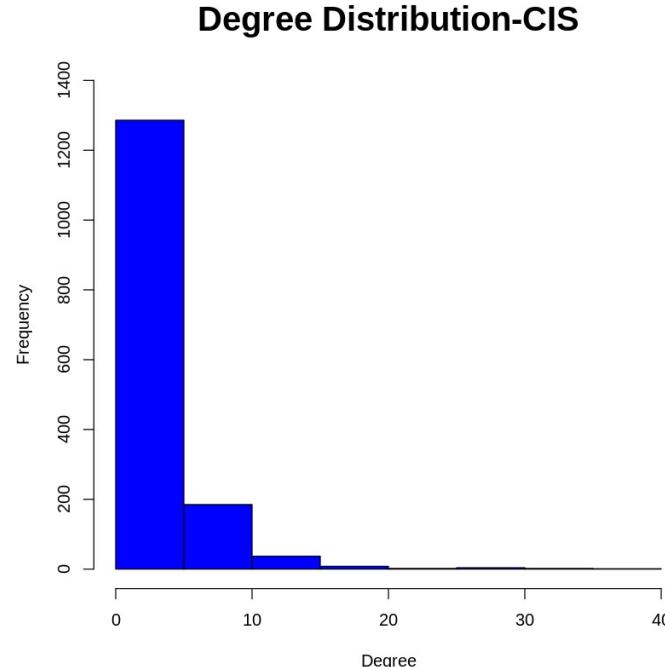
Descriptive analysis

Descriptive analysis	CisFull	PrinFull	WOSFull
size of whole network	1525	2839	5291
size of Italian network	465	440	469
Density	0.002	0.002	0.03
Edge count	2534	9379	426435
Isolated node	60	7	26
The size of largest component	1277	2696	4852
The longest shortest path	19	17	16
Degree centrality	0.02	0.04	0.20
Closeness centrality	0.001	0.0008	0.0003
Eigen centrality	0.99	0.98	0.85
Betweenness centrality	0.09	0.16	
Count of component	114	27	103
Transitivity	0.29	0.53	0.90

Degree Visualizing:

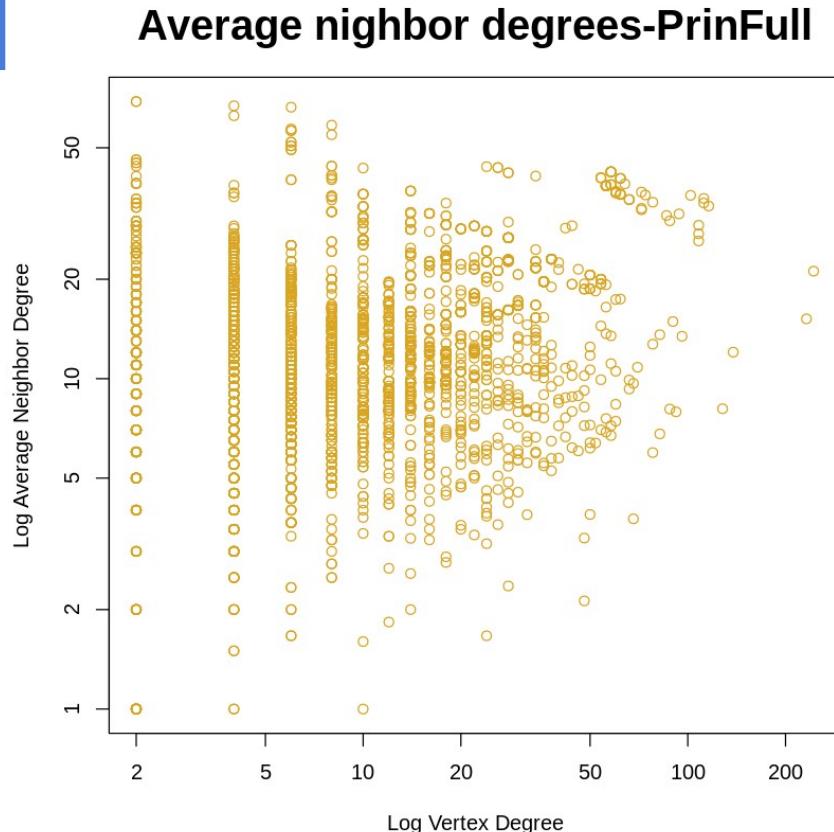
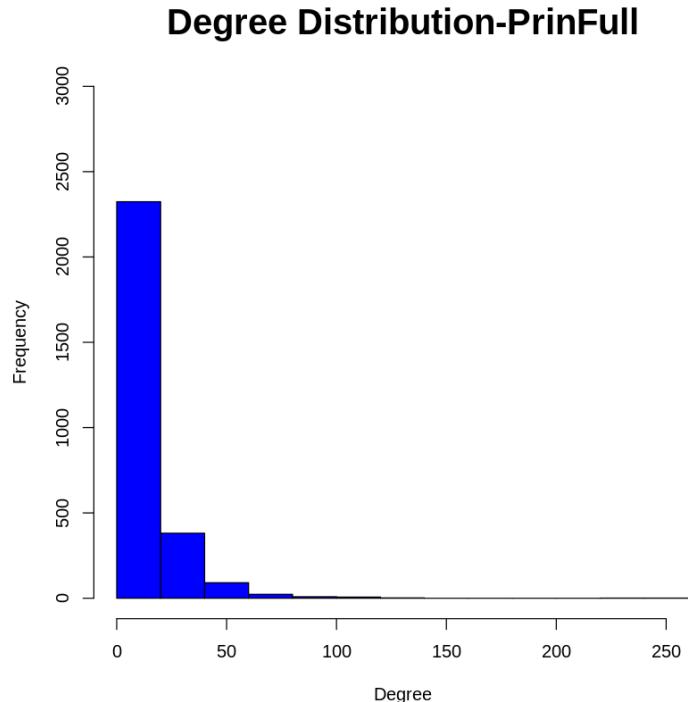
While there is a substantial number of nodes of quite low degree.

The plot "Log Average Neighbor Degree" suggests that while there is a tendency of nodes of higher degrees to link with similar nodes, nodes of lower degree tend to link with nodes of both lower and higher degrees.



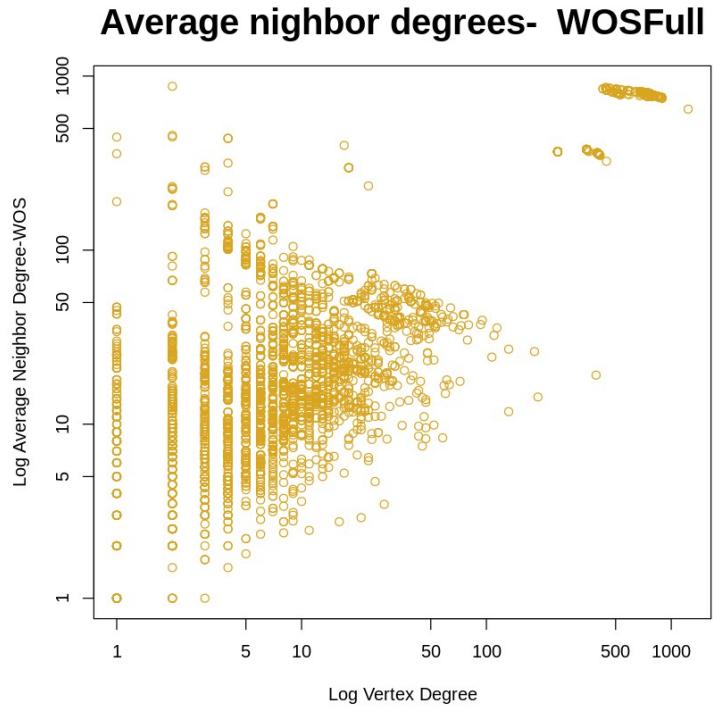
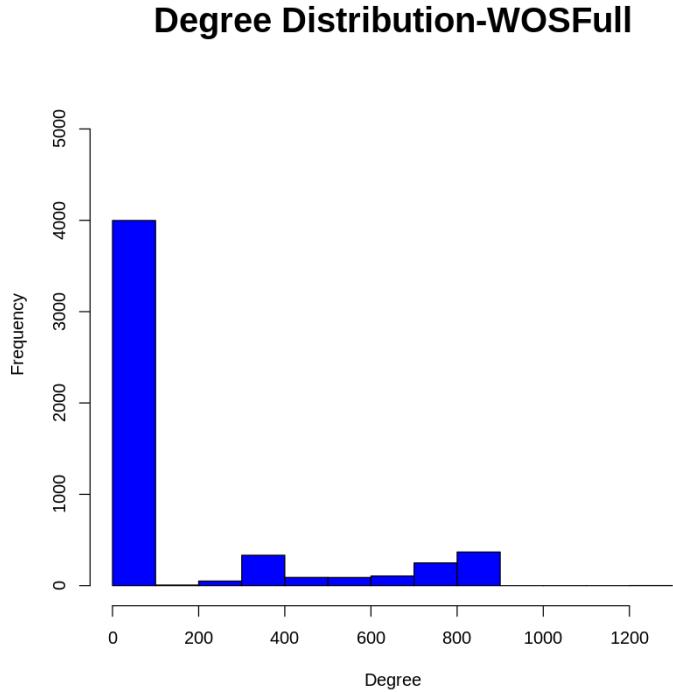
Degree visualization

We have a bit more degree in PrinFull rather than CisFull, the largest degree for PrinFull stands for 122



Degree visualization

We have a bit more degree in PrinFull rather than CisFull, the largest degree for PrinFull stands for 1234 and 325 nodes with 886 degree



Summery of degree

Degree	Min	1 st Quarter	Median	Mean	3rd Qu	Max
CisFull	0	1	2	3.3	4	38
PrinFull	0	2	6.6	8	8	122
WOSFull	0	5	10	161.5	52	1234

clique

The size of the largest clique

CisFull
11

PrinFull
29

WOSFull
679

For PrinFull network: there are 277 nodes (cliques of size one) and 387 (cliques of size two), followed by 132 triangles (cliques of size three).

Clique	1	2	3	4	5
PrinFull	277	387	132	18	1

Clique	1	2	3	4
CisFull	465	524	146	1

Clique	1	2	3	4	5	6	7
WOSFull	469	399	129	41	21	7	1

K CORE

Recall:

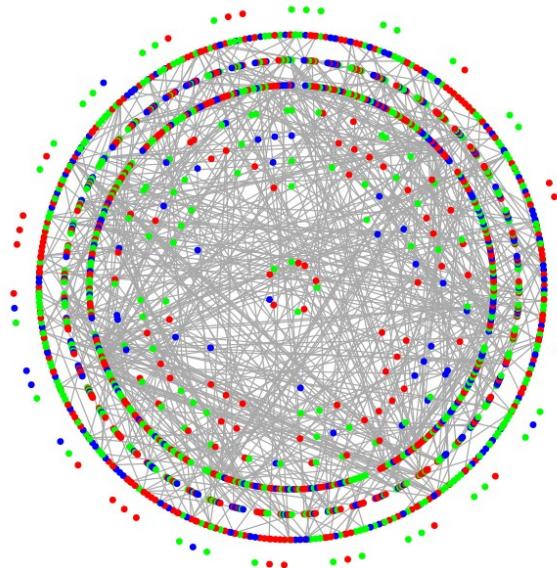
A k-core is a maximal sub graph where each vertex is connected to at least k others.

For example for PrinFull network, the network is made up of the highest k-core is in the center of network in this case core 4 will be in the densest portion of the network.

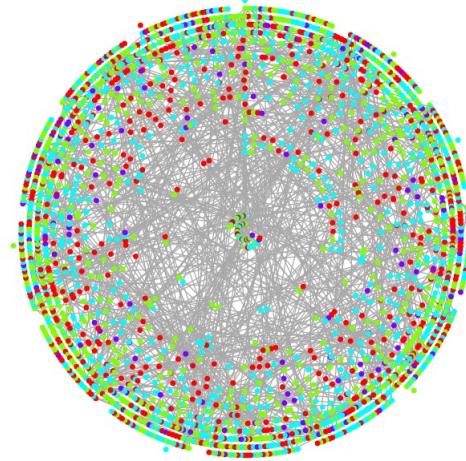
CisFull		PrinFull		WOSFull	
Core	Vertices	core	Vertices	core	Vertices
0	90	0	30	0	141
1	146	1	102	1	146
2	149	2	143	2	149
3	80	3	140	3	26
		4	25	6	7

The Coreness Layout

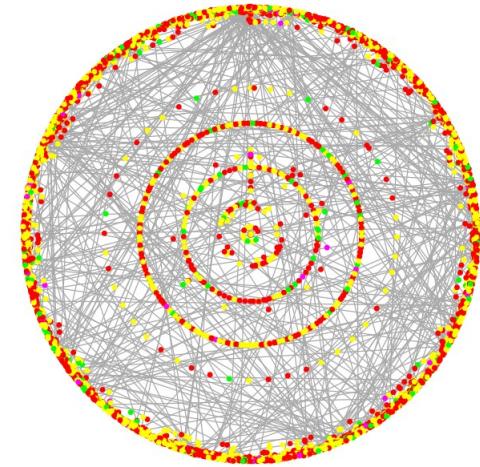
Coreness-CisFull



Coreness-PrinFull



Coreness-WOSFull



Community detection

In all the networks there is a giant component. This component contains more than 90% of the vertices in the network.

CisFull

2696

114

PrinFull

1277

27

CisFull

4852

103

Community detection

	average path length in the giant component	the longest paths(diameter) of giant component	vertex connectivity	edge connectivity	CUT vertices
CisFull	7	19	1	1	272
PrinFull	6	17	1	1	262
WOSFull	5	16	1	1	234

In the case of the giant component of the network, the vertex and edge connectivity are both equal to one. Thus it requires the removal of only a single well-chosen vertex or edge in order to break this sub graph into additional components.

In the giant component of the data, almost 10% of the vertices are cut vertices(262) out of all nodes.

Community detection for giant component

CisFull
edge betweenness
31 component
modularity 0.87

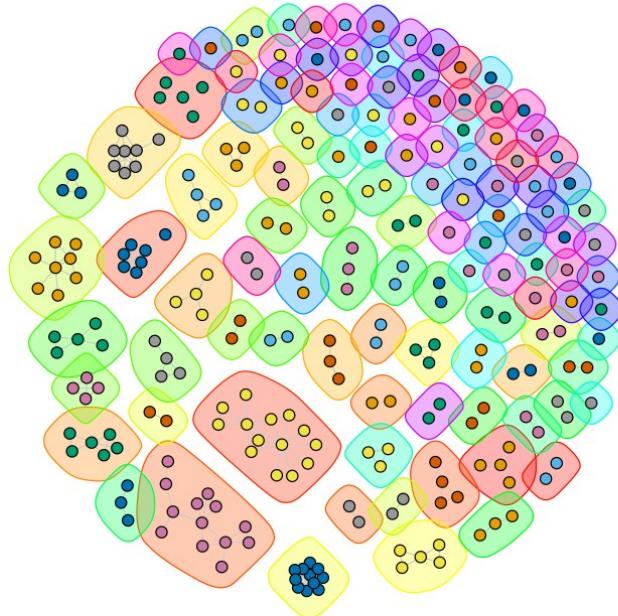
DECIDING BASED ON modularity

PrinFull
edge betweenness
43 component
modularity 0.89

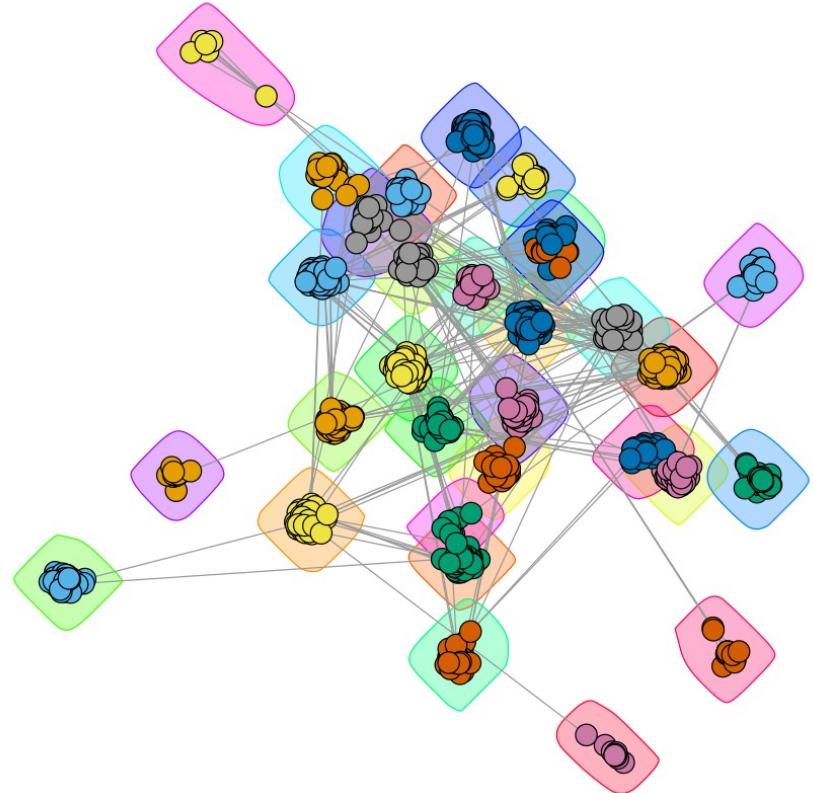
WOSFull
louvain
23 component
modularity 0.36

Community detection- CIS

113 small CIS Network.



Giant component CIS Network.

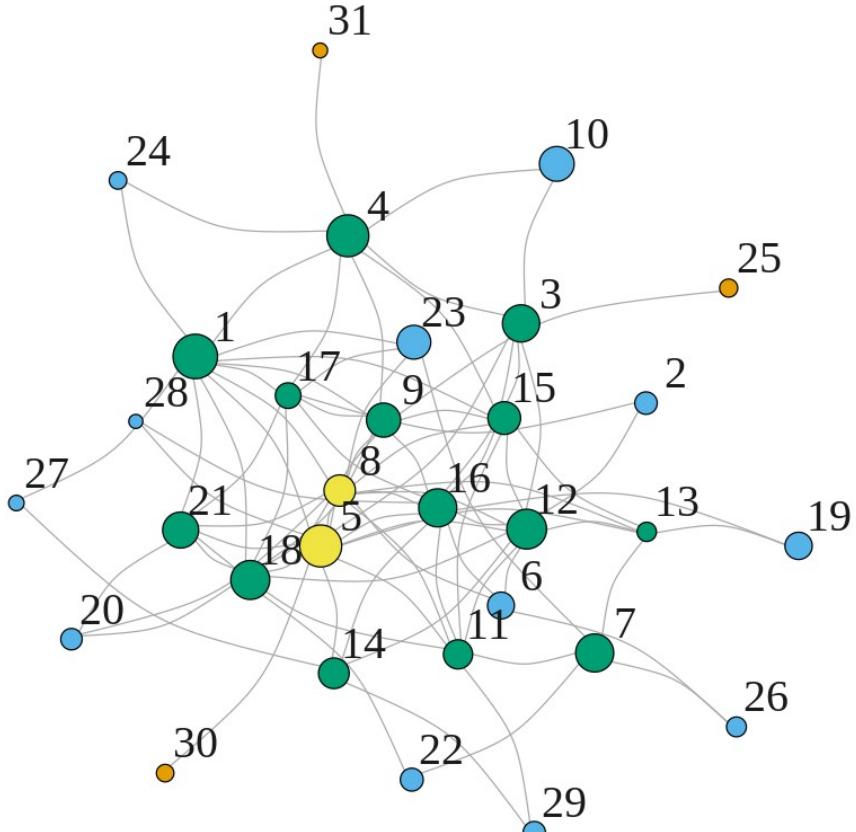


Contract the Communities:

Plotting a single node for each community. Here, I make the area of each community vertex proportional to the number of that community.

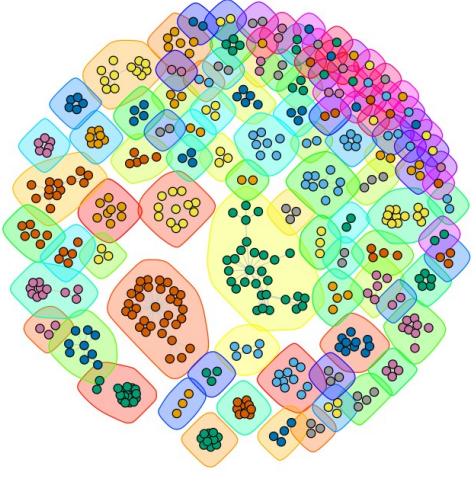
I colored the vertices using a coarse grouping based on their degrees.

Giant component with contracted node-CIS Network.



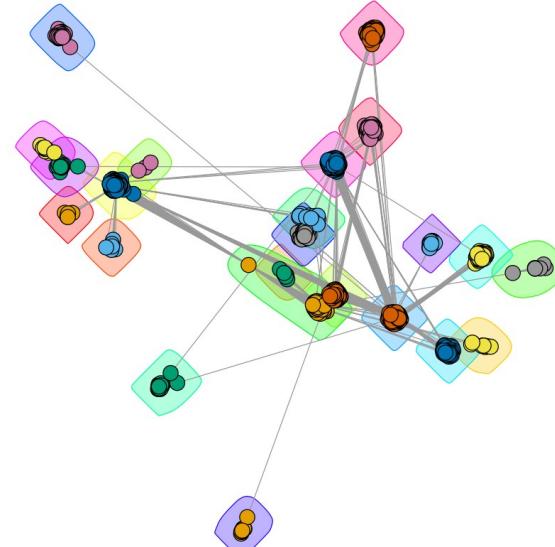
Community detection- WOS

102 small component WOS Network.

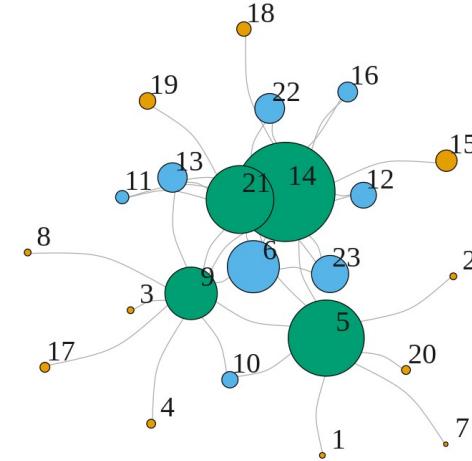


cluster louvain Algorithm

The giant component -WOS

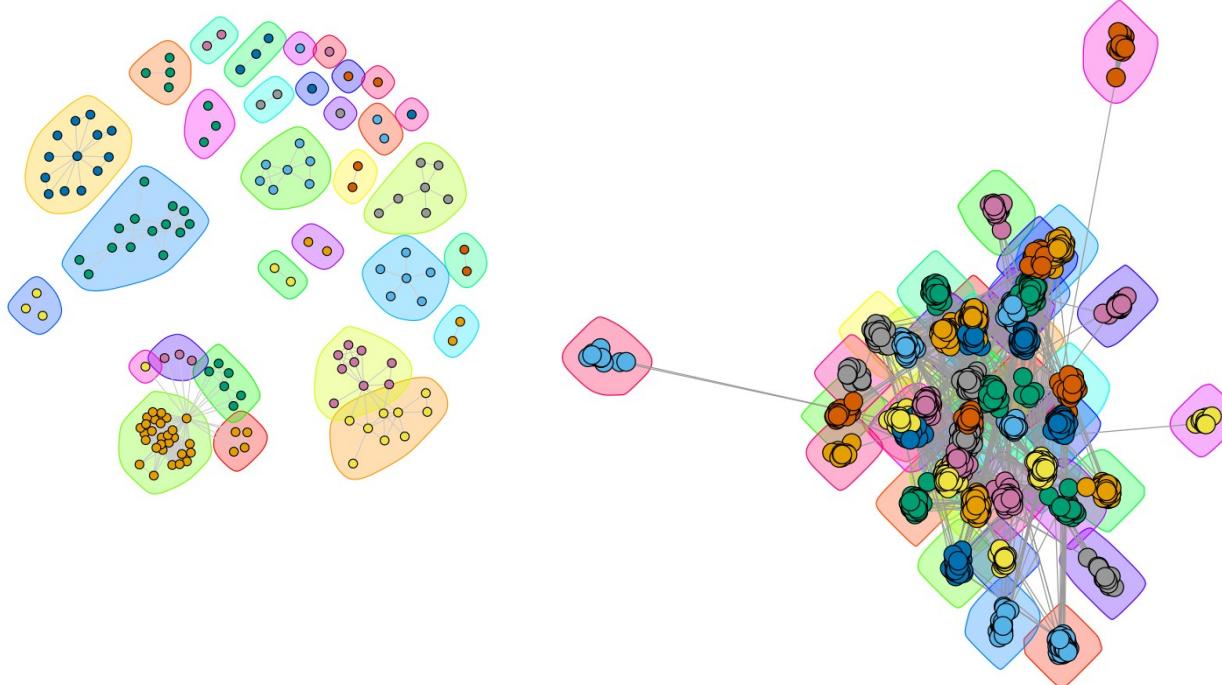


Giant component with contracted node-WOS Network.

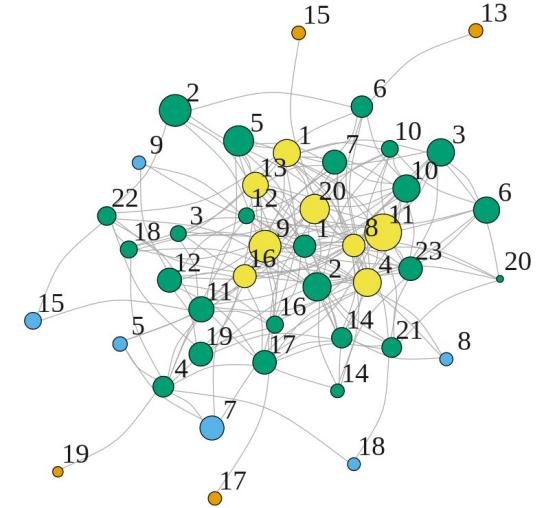


Community detection PrinFull

26 small PRIN Network-PrinFull Network. The giant component-PrinFull network



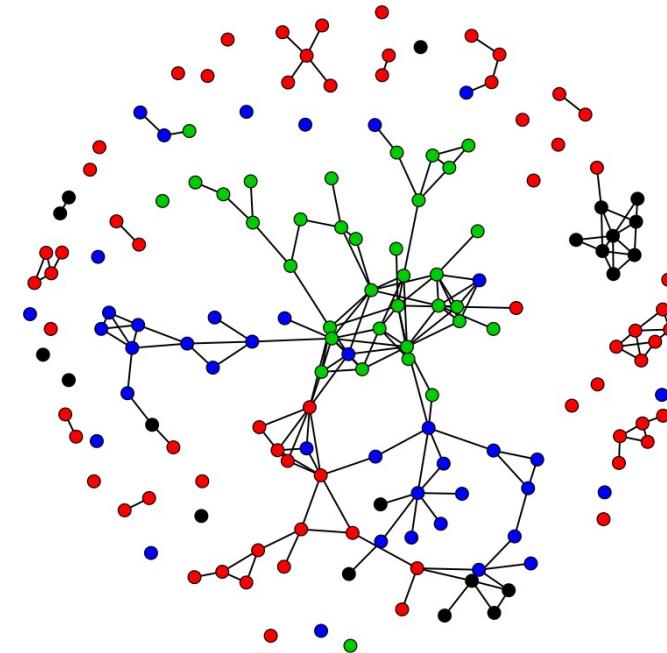
The giant component with contracted node



Exponential Random Graph Model : PrinFull

I defined an induced sub graph for Italian researchers who work in sector 2,3,4,5. So the size of network seems good with 163 vertices and total edges is equal to 187. The transitivity of this sub-graph is **36%** so I didn't fit the model with transitivity. First I created a NULL model that includes only the edges term. For the NULL model I got a negative edge parameter(**-4.242**), since the network is rather sparse. And conditional probability of having a tie is equal to **0.01** that is equal to density.

homophily based on Sector-PrinFull

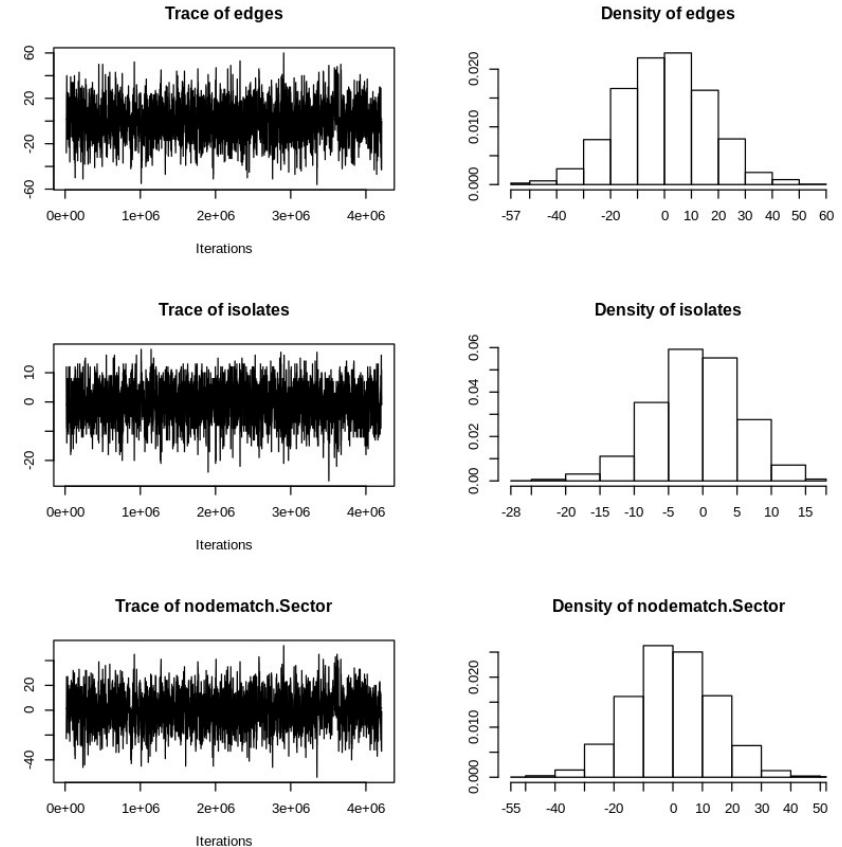


Exponential Random Graph Model : PrinFull

The final model is like below :

```
ergm(l2 ~ edges + isolates +  
nodematch('Sector'))
```

there is a good estimation around 80%
with error :0.23 for isolated node.
Also for citation the estimate is
positive with low error.
we are aiming for the trace to look like
a hairy caterpillar. And there shouldn't
be a burn-in in the plot.



Exponential Random Graph Model : PrinFull

SIMULATION

The result for simulation is as below:
Simulated network:

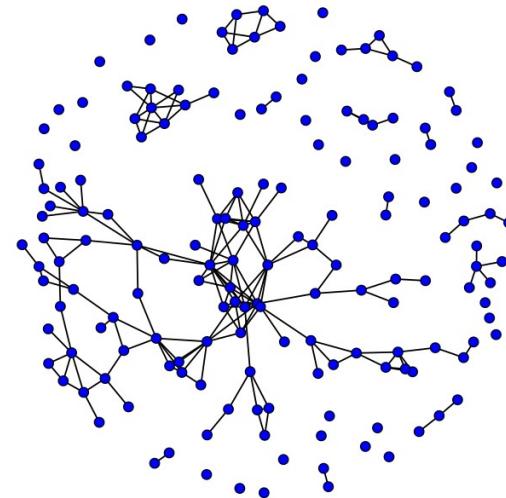
edges: -5.32

isolates: 0.8

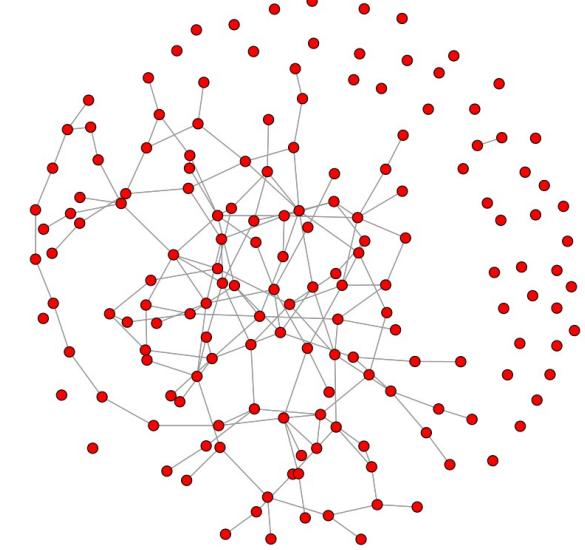
node match Sector:2.32
with 0.18 error.

The density of observed
and simulated
is 0.01.

Italian researcher in sector 2:5-PrinFull

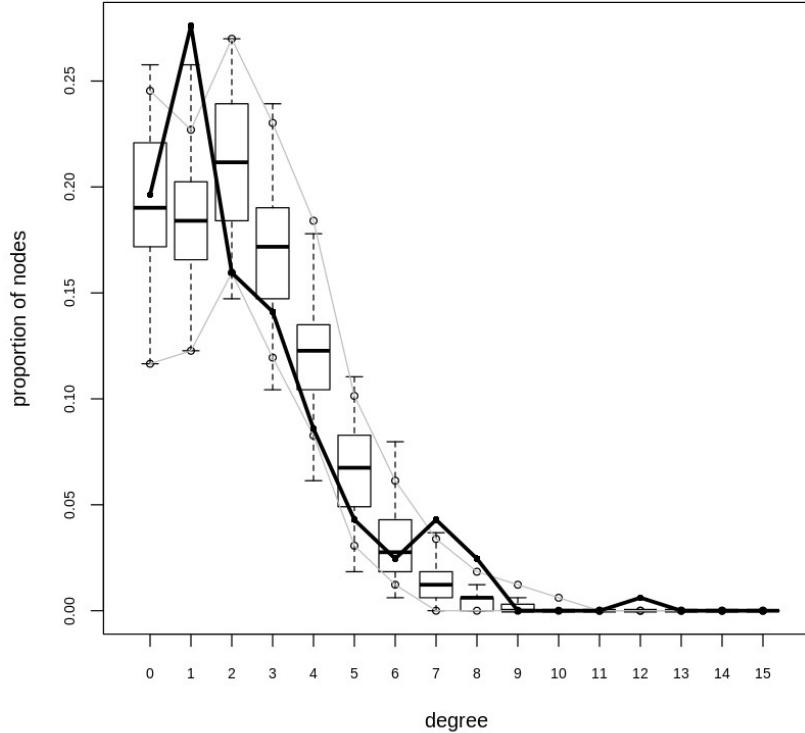
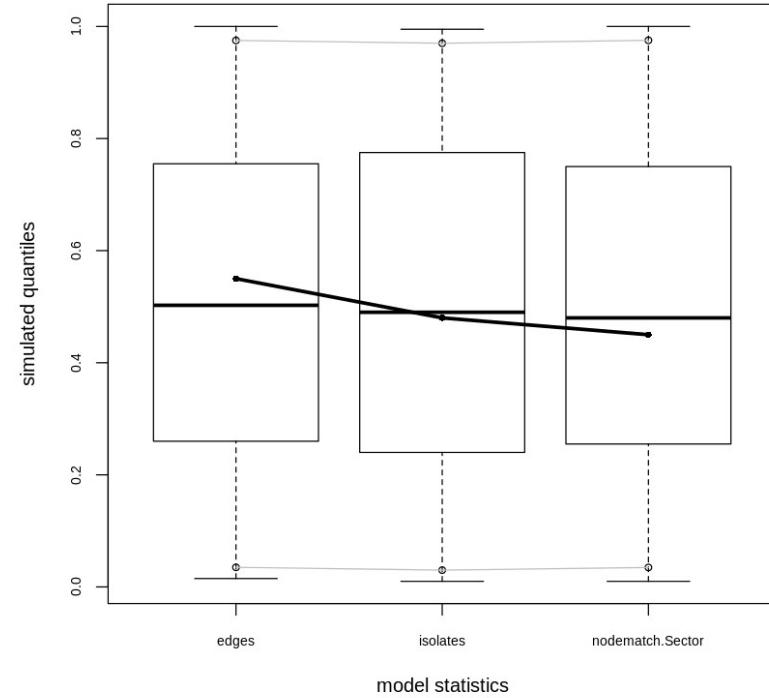


simulated network-PrinFull



Exponential Random Graph Model : PrinFull

Goodness-of-fit diagnostics

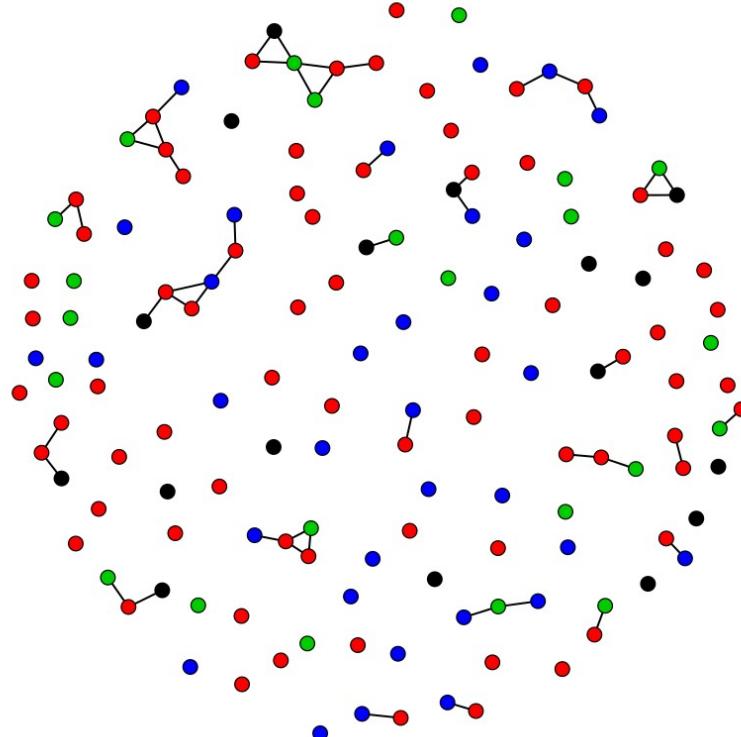


Exponential Random Graph Model : CisFull

The subset of data is the same as PrinFull, I created a NULL model that includes only the edges term, for the NULL model I got a negative edge parameter (-5.3) since the network is rather sparse. And conditional probability of having a tie is equal edges: **0.004** that is equal to density. Then I visualize the relation between degree and node attribute (citations), there is no particular association,

so when I fit model based on nodematch sector I get very bad result as :-**0.65**, and the result for nodecov citations is **0.0003** with very low error.

homophily based on Sector-ClisFull



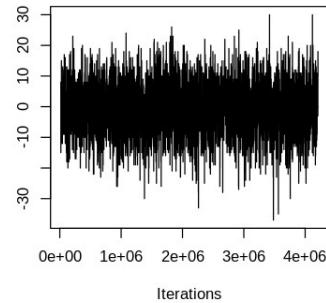
Exponential Random Graph Model : CisFull

The final model is like below :

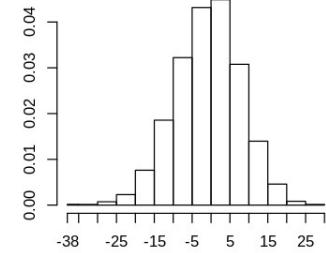
```
ergm(l2 ~ edges + isolates)
```

we are aiming for the trace to look like a hairy caterpillar. And there shouldn't be a burn-in in the plot.

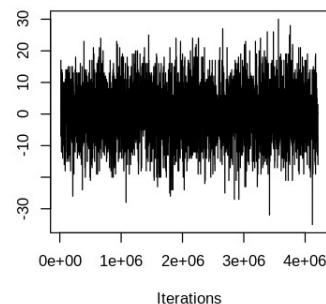
Trace of edges



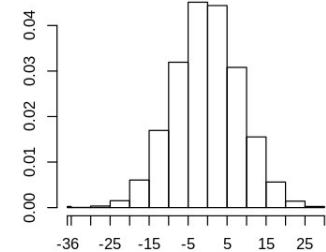
Density of edges



Trace of isolates



Density of isolates



Exponential Random Graph Model : CisFull

SIMULATION

The result for simulation is as below:

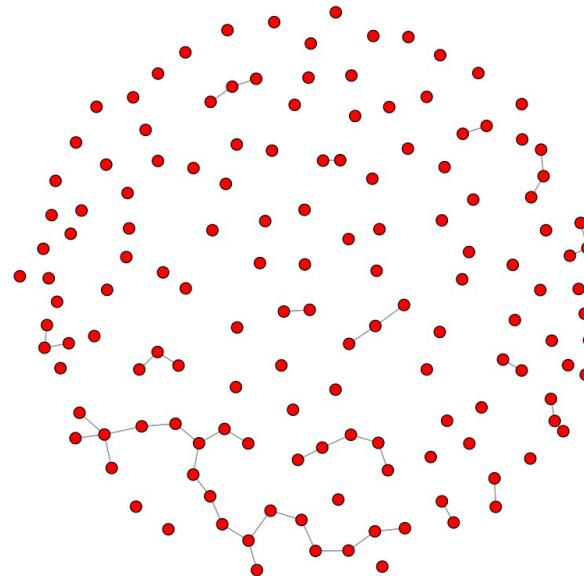
Simulated network:

Edges: -4.8

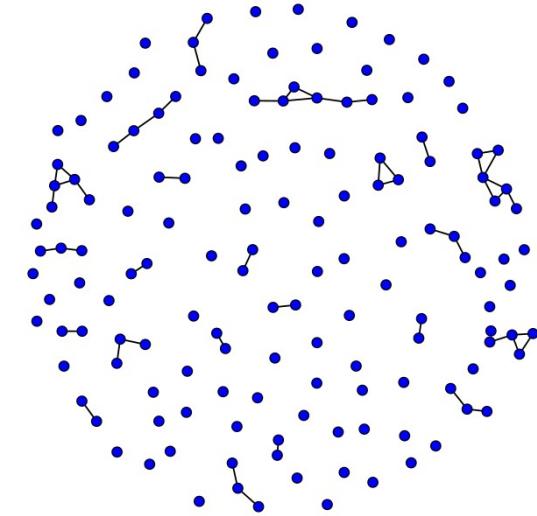
Isolates: 0.53

The density of observed and simulated is 0.004

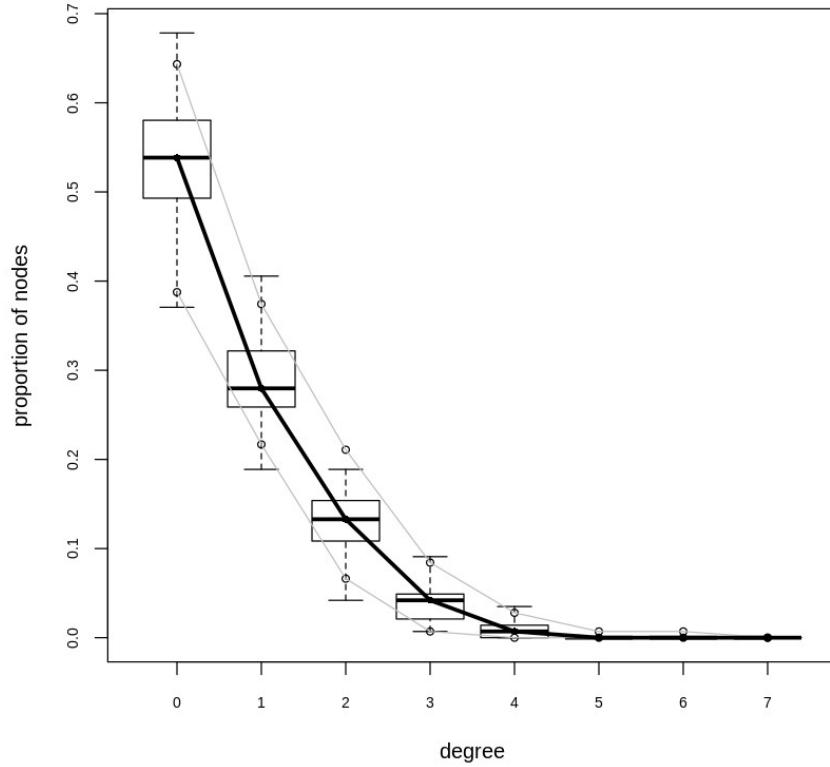
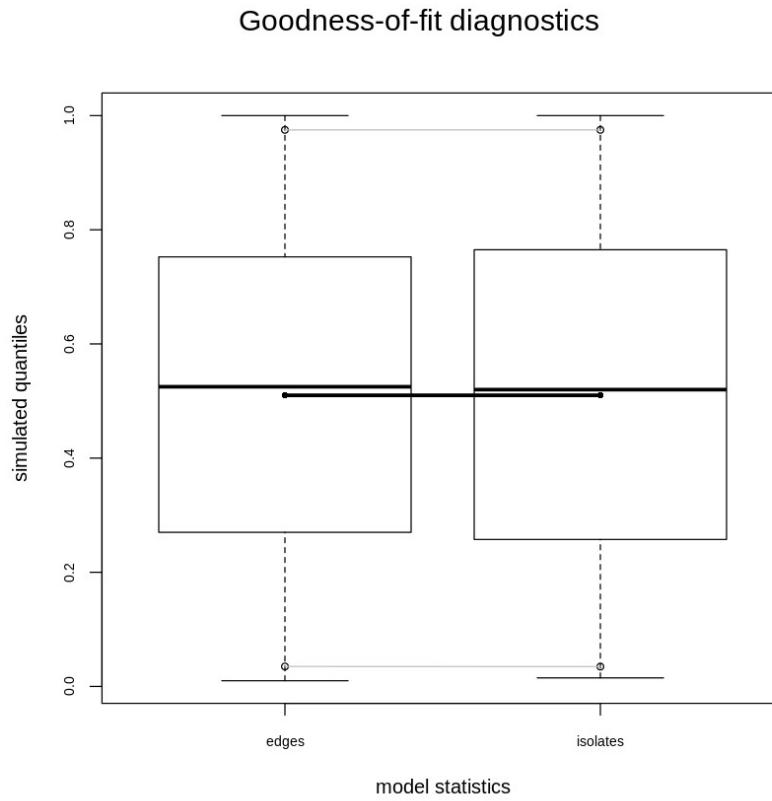
Simulated network- CisFull



italian researcher in sector 2:4-CisFull



Exponential Random Graph Model : CisFull



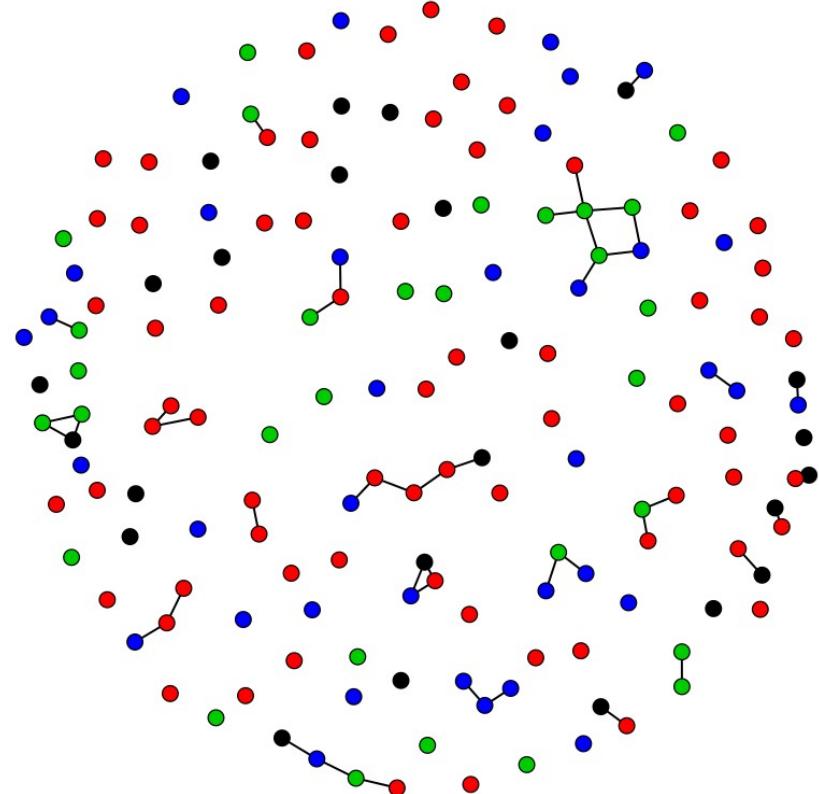
Exponential Random Graph Model : WOSFull

I created a NULL model that include only the edges term, for the NULL model I got a negative edge parameter **(-5.6)** And conditional probability of having a tie is equal to **0.003** that is equal to density.

I started with model `ergm(l2 ~ edges + isolates + nodematch('Sector'))`
Edges : **-5.2108** er :**0.3761**
Isolates **0.4481** er :**0.3155**
Nodematch Sector: **0.1549** er: **0.3252**

As we can see the error is high, because plot there is no homophily in the graph based on sector. I fit model as `ergm(l2 ~ edges + isolates)`, I didn't get a good result on density of simulated, then I add `nodecov` citations to model, so I got better result.

homophily based on Sector-WOSFull



Exponential Random Graph Model : WOSFull

The edges of observed model is 43

```
ergm(l2 ~ edges + isolates )
```

	Estimate	Std.	Error
edges	-5.1733	0.3622	
Isolates	0.4396	0.3190	

Simulated :

vertices = 158

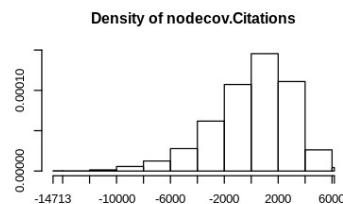
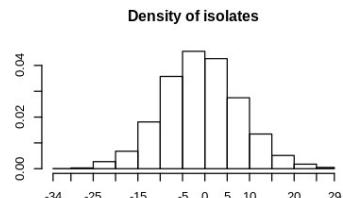
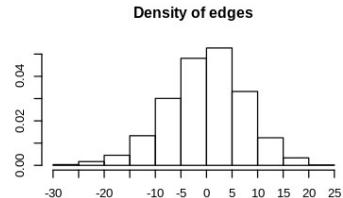
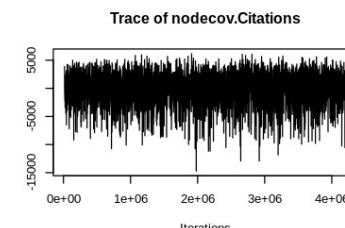
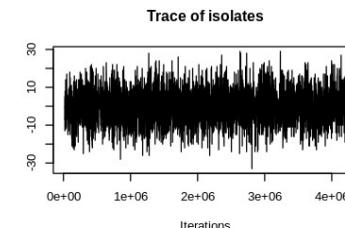
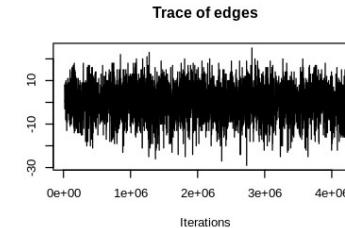
total edges= 55

The final model is like below :

```
ergm(l2 ~ edges +  
isolates+nodecov('Citations' ) )
```

vertices = 158

total edges= 36



Exponential Random Graph Model : WOSFull

SIMULATION

The result for simulation is as below:

Simulated network:

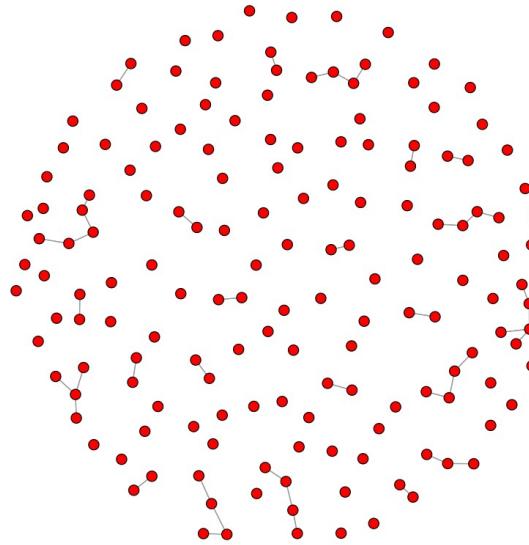
Edges: -5.1

Isolates: 0.448

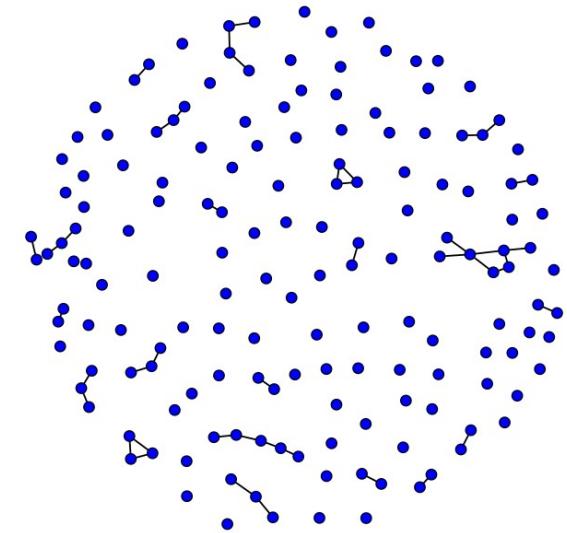
Nodecov Citations
0.0001

The density of simulated is 0.002, but the density of observed is 0.003.

simulated network-WOSFull

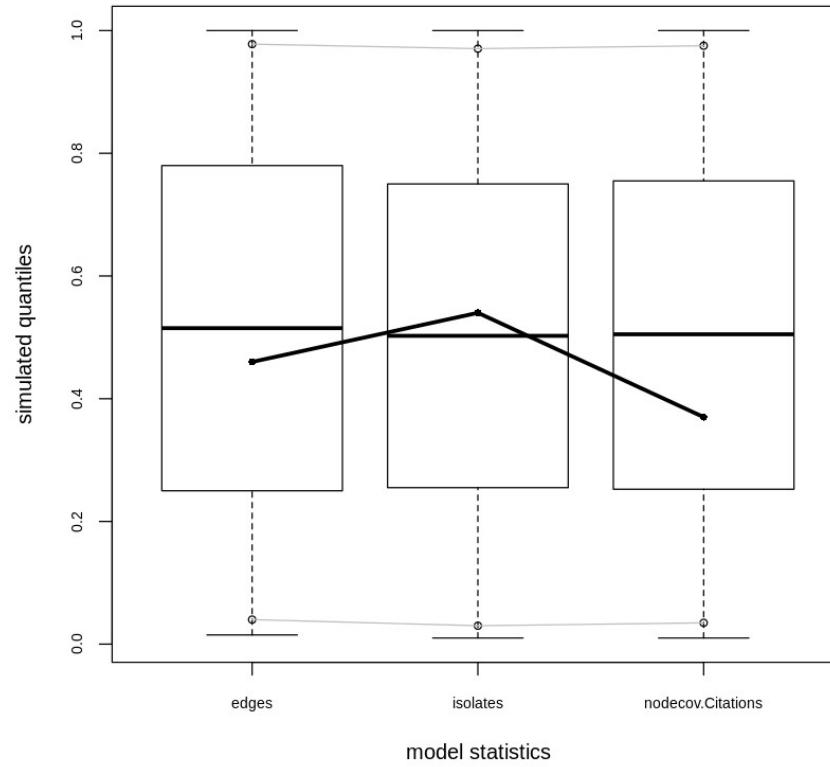
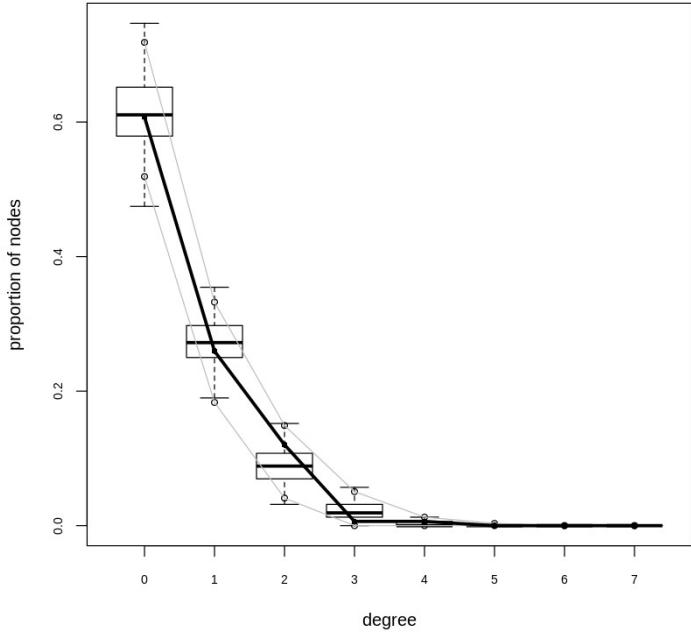


italian researcher in sector 2:5-WOSFull



Exponential Random Graph Model : CisFull

Goodness-of-fit diagnostics





Thank you