



The Marvel Universe Social Network

Marie BAI, Alex CULPIN, Moussa SIDIBE

AGENDA

1

Describe the dataset

2

Build the graph

3

Describe the two unipartite graphs

4

Find the communities

Nodes

	node	type
0	2001 10	comic
1	2001 8	comic
2	2001 9	comic
3	24-HOUR MAN/EMMANUEL	hero
4	3-D MAN/CHARLES CHAN	hero

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 19090 entries, 0 to 19089
Data columns (total 2 columns):
#   Column  Non-Null Count  Dtype
---  ---
0    node    19090 non-null    object
1    type     19090 non-null    object
dtypes: object(2)
memory usage: 298.4+ KB
```

None

Shape (19090, 2)

	node	type
count	19090	19090
unique	19090	2
top	2001 10	comic
freq	1	12651

Hero network

	hero1	hero2
0	LITTLE, ABNER	PRINCESS ZANDA
1	LITTLE, ABNER	BLACK PANTHER/T'CHAL
2	BLACK PANTHER/T'CHAL	PRINCESS ZANDA
3	LITTLE, ABNER	PRINCESS ZANDA
4	LITTLE, ABNER	BLACK PANTHER/T'CHAL

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 574467 entries, 0 to 574466
Data columns (total 2 columns):
#   Column  Non-Null Count  Dtype
---  ---
0    hero1    574467 non-null    object
1    hero2    574467 non-null    object
dtypes: object(2)
memory usage: 8.8+ MB
```

None

Shape (574467, 2)

	hero1	hero2
count	574467	574467
unique	6211	6173
top	CAPTAIN AMERICA	CAPTAIN AMERICA
freq	8149	8350

Edges

	hero	comic
0	24-HOUR MAN/EMMANUEL	AA2 35
1	3-D MAN/CHARLES CHAN	AVF 4
2	3-D MAN/CHARLES CHAN	AVF 5
3	3-D MAN/CHARLES CHAN	COC 1
4	3-D MAN/CHARLES CHAN	H2 251

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 96104 entries, 0 to 96103
Data columns (total 2 columns):
#   Column  Non-Null Count  Dtype
---  ---
0    hero    96104 non-null    object
1    comic   96104 non-null    object
dtypes: object(2)
memory usage: 1.5+ MB
```

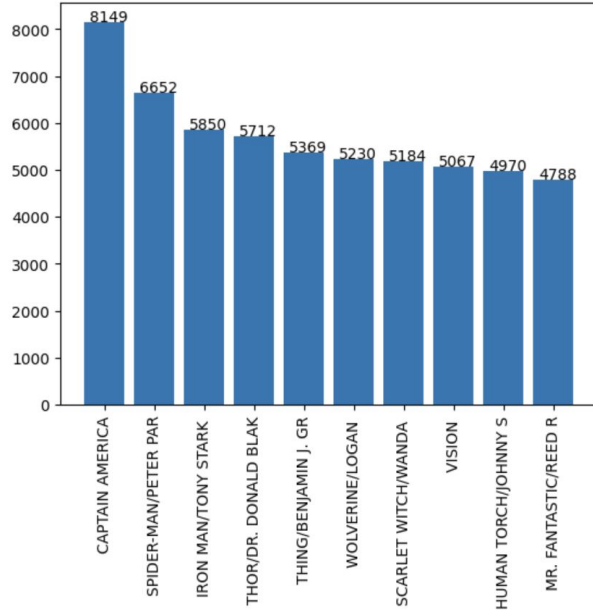
None

Shape (96104, 2)

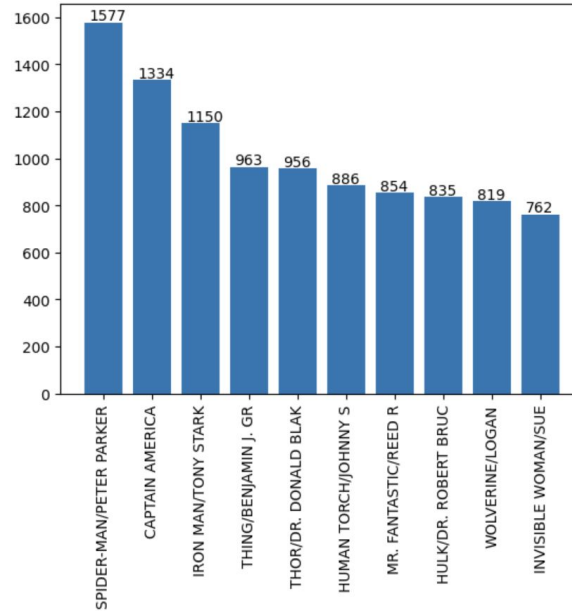
	hero	comic
count	96104	96104
unique	6439	12651
top	SPIDER-MAN/PETER PARKER	COC 1
freq	1577	111

Top heroes

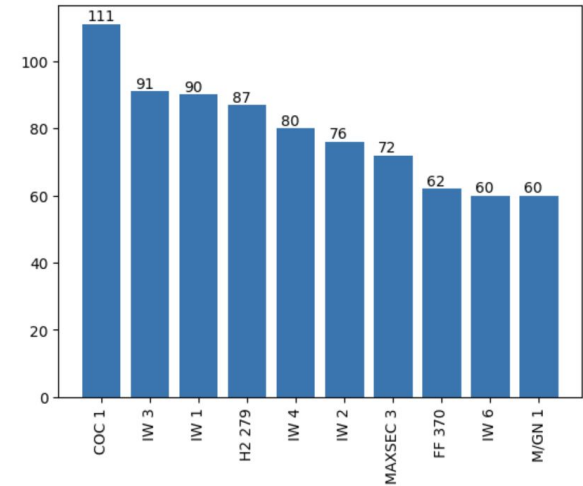
The most popular hero in terms of network



The most popular hero in comics

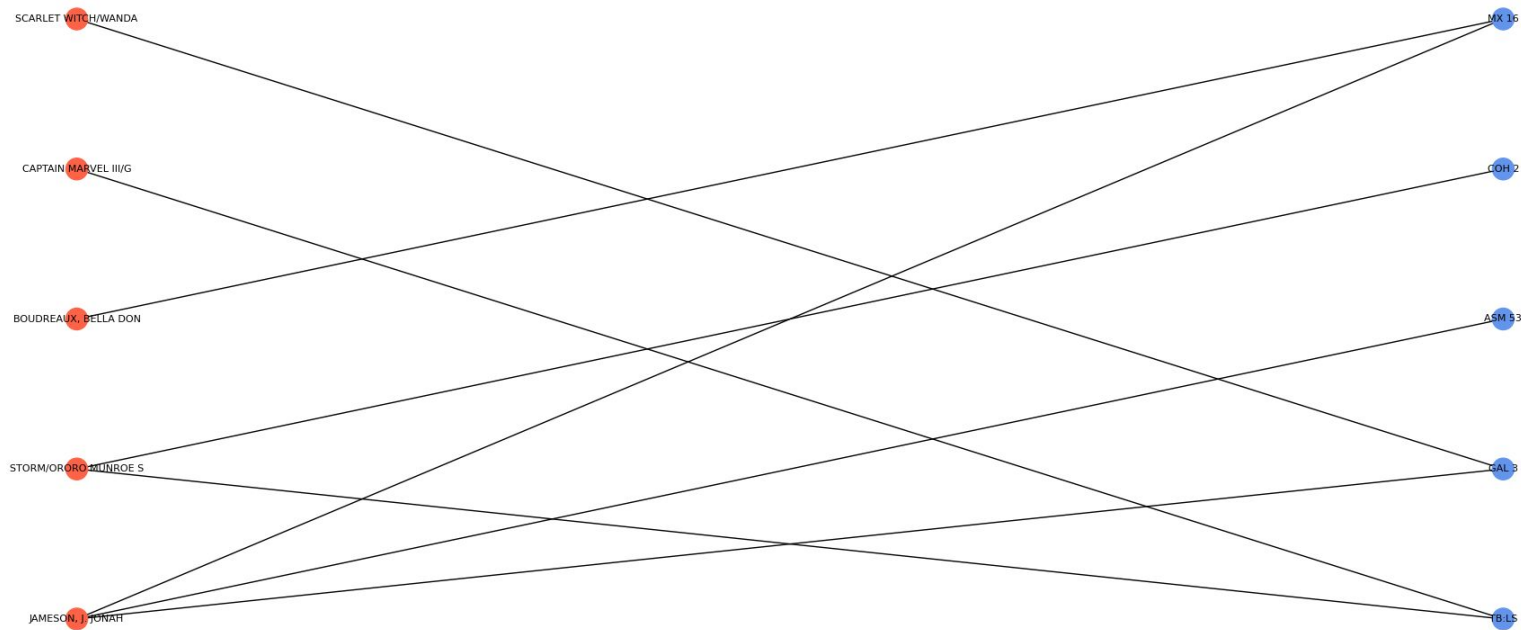


The most participant comics



Graph

mini_bipartite_graph



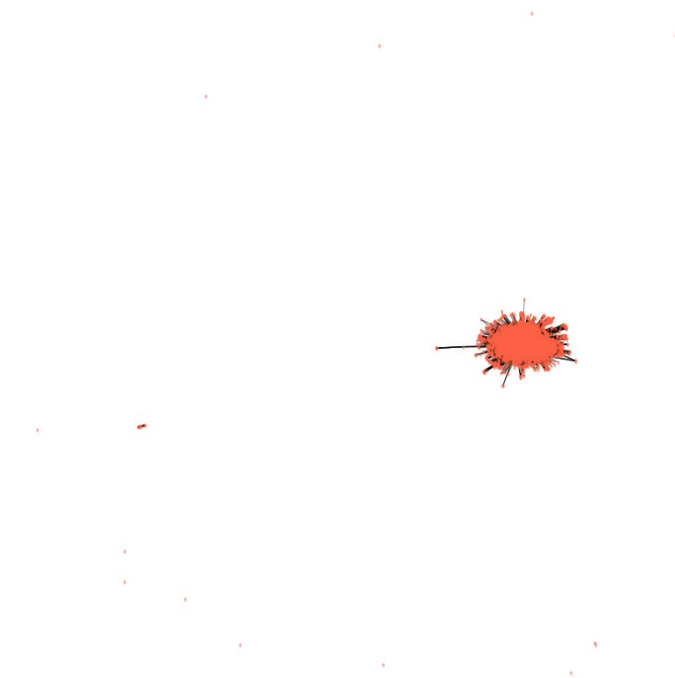
Graph

full_bipartite_graph

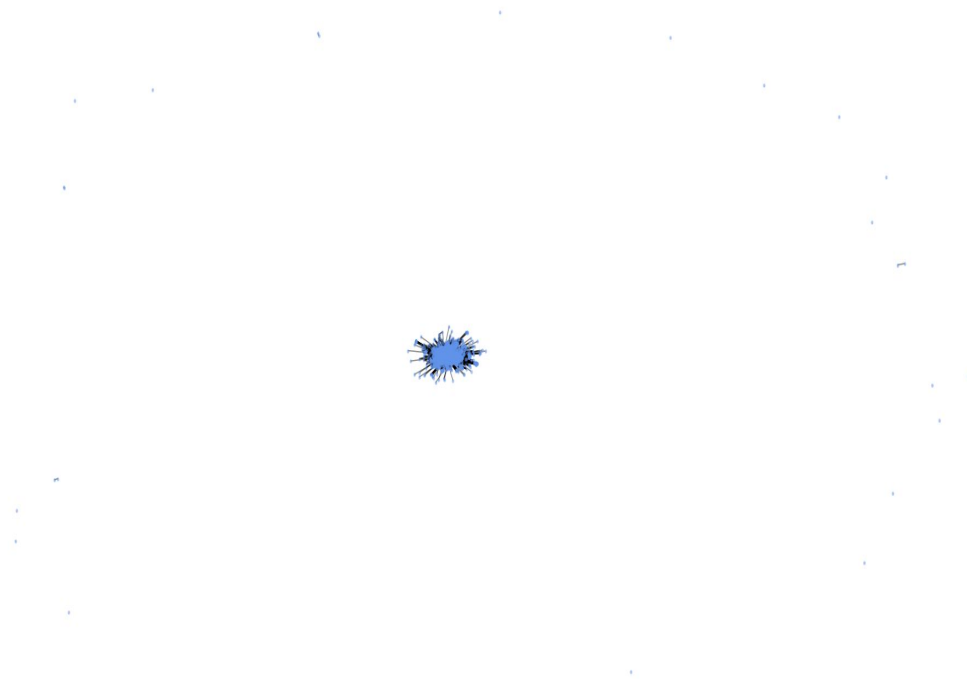


Graph

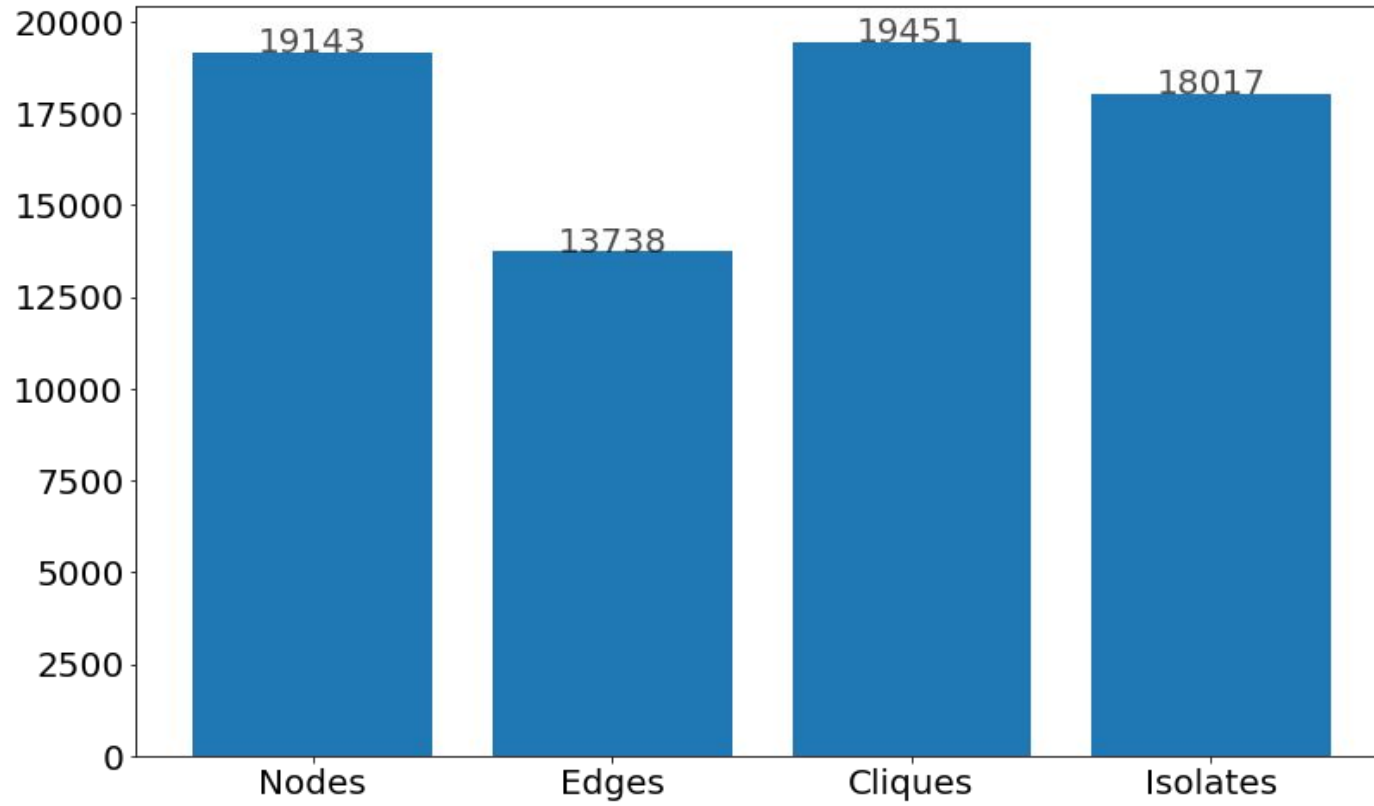
hero_unipartite_graph



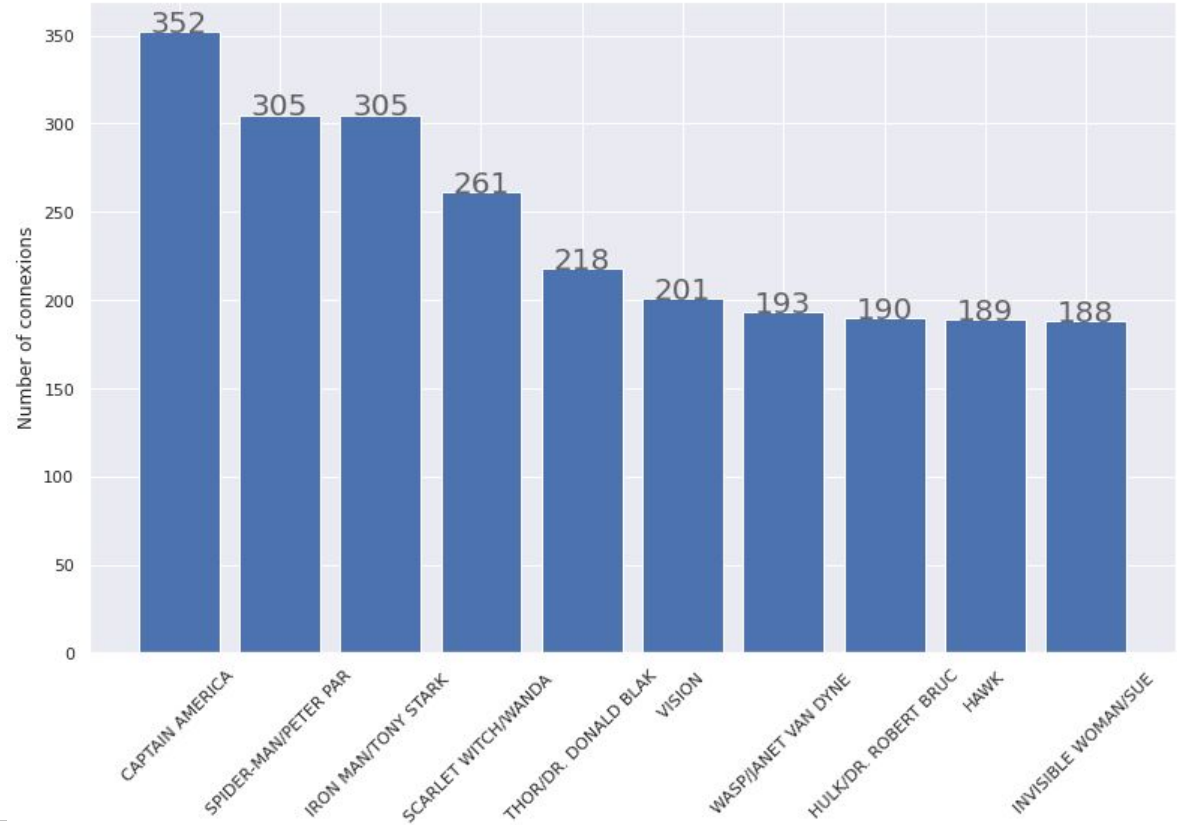
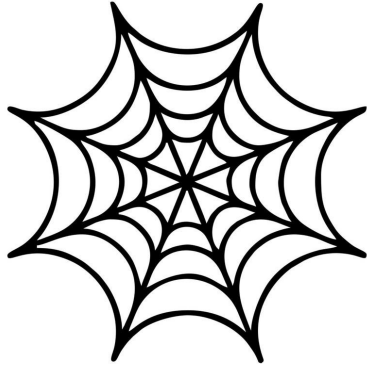
commics_unipartite_graph



Components of the graph



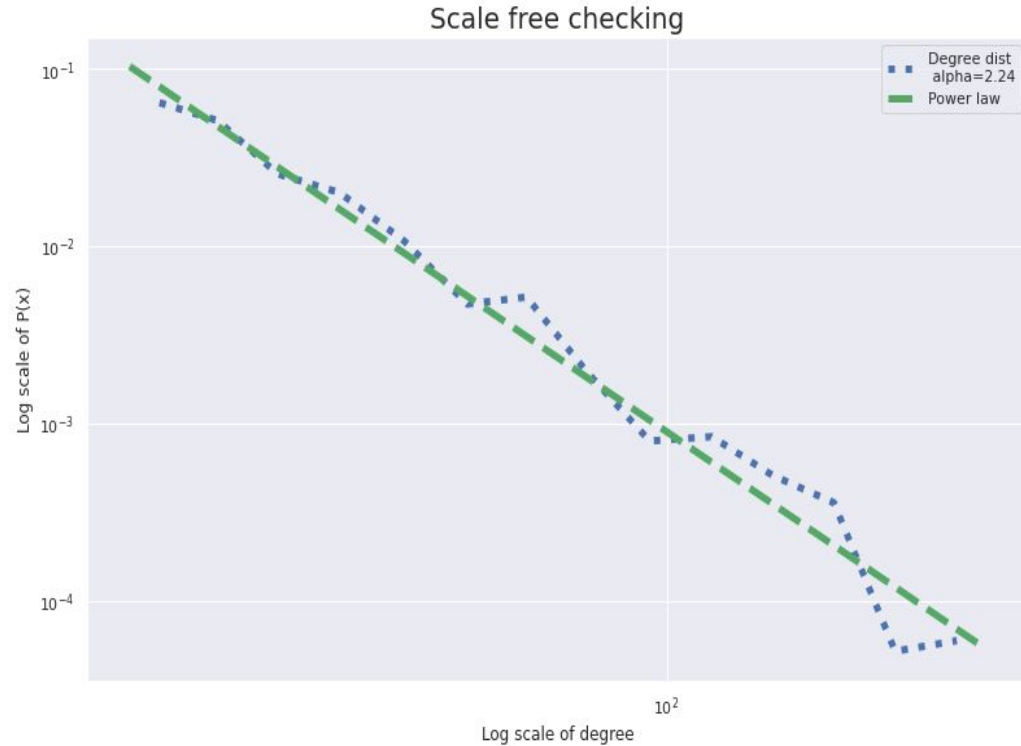
What hero are most connected?



This graph is it scale free ???

Power Law or Pareto law

$$P(\text{deg} = x) \propto x^{-\alpha}$$



Centrality analysis

There are many algorithm for centrality analysis in a community:

LIKE:

- PageRank
- Closeness centrality
-

In this case they return the same result

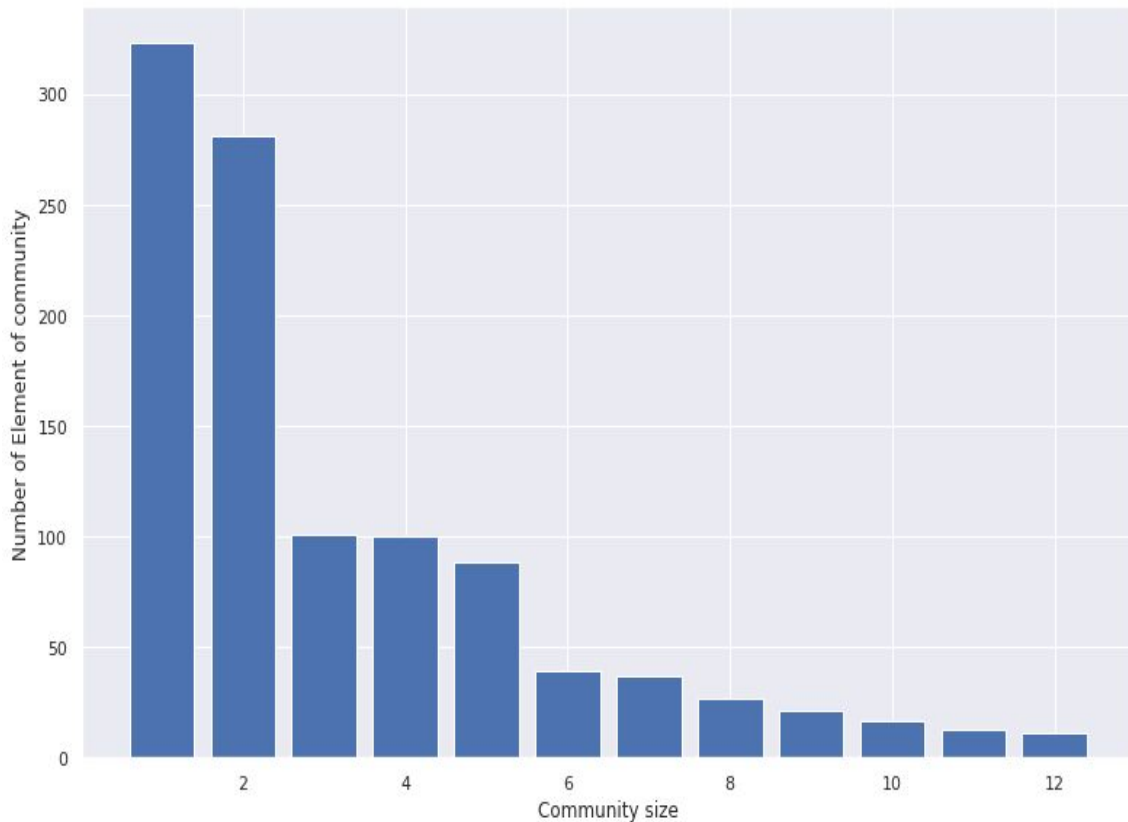
It inform about the influence or importance of a node in a community

Community distribution

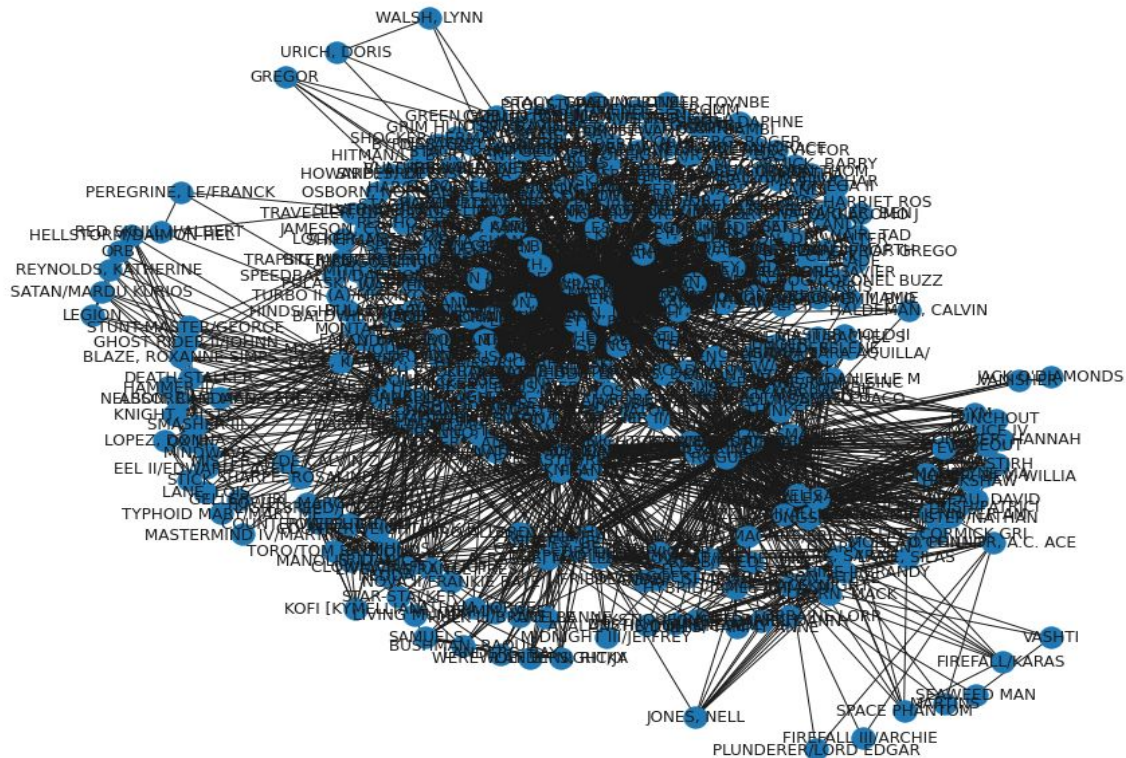
Number of communities: 20

Biggest community: 323 nodes and 3141 edges as central hero Spider Man

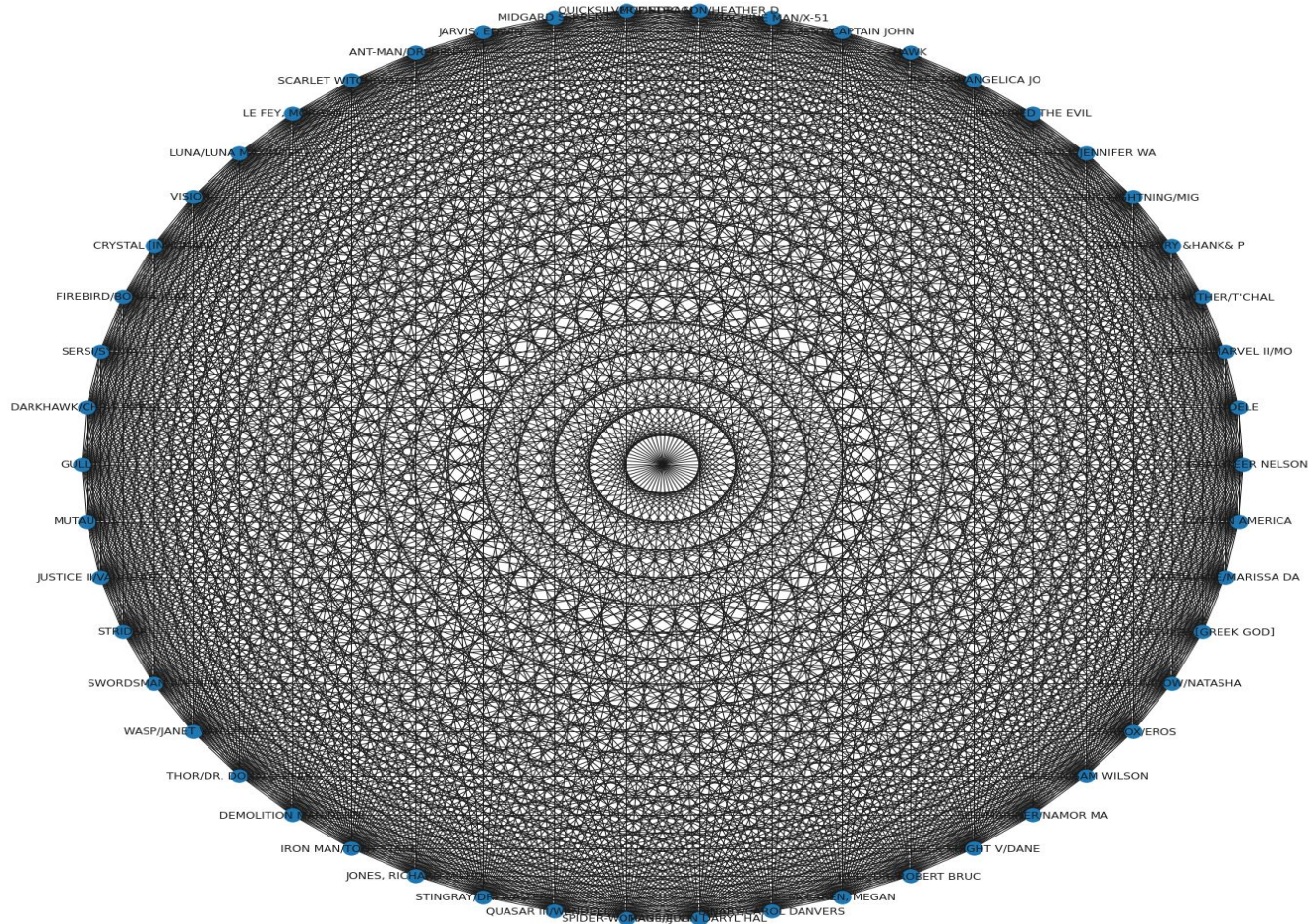
Second and most connexe: 281 nodes with 4235 edges as CAPTAIN AMERICA as central hero.



Subgraph with 323 elements



Kamada kawai Graph on subgraph with 283 elements





Thank you for your attention