CS-E5740 - Complex Networks
Exercise set 3

Hugues Verlin (584788)
hugues.verlin@aalto.fi

November 28, 2016

1 Degree correlations and assortativity

1. a) Create a scatter plot of the degrees of nodes incident to each edge

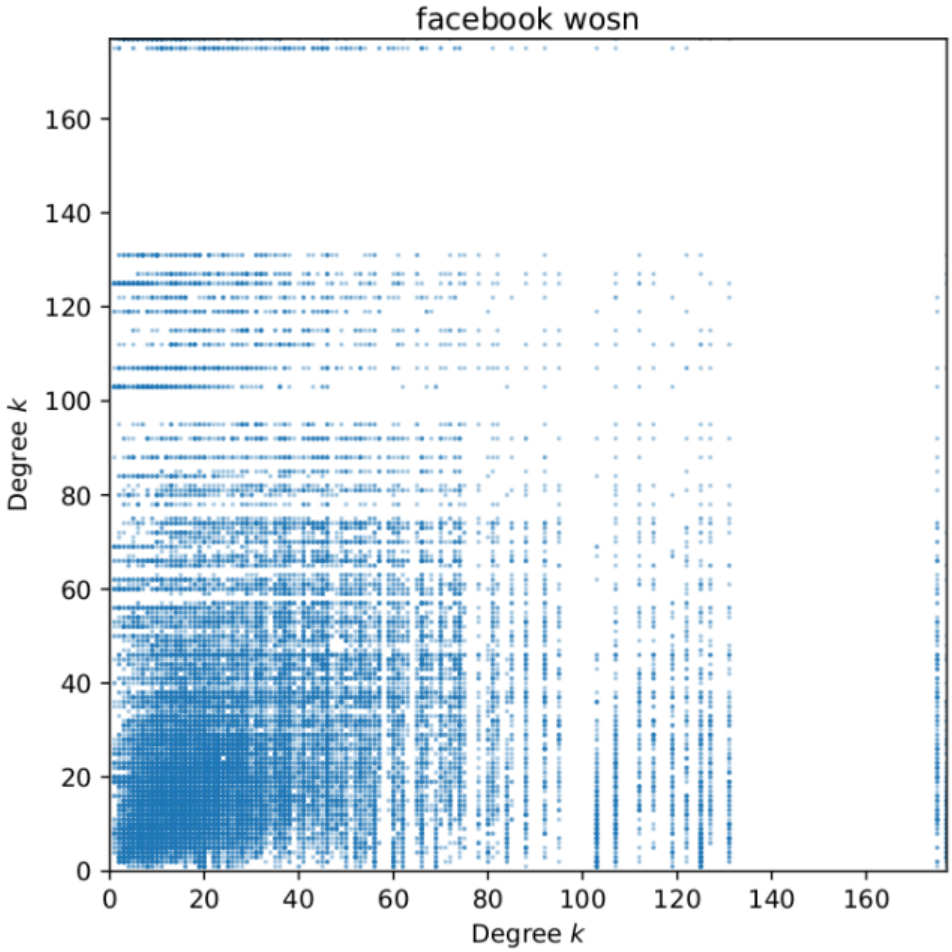

Figure 1.1: Scatter plot - facebook

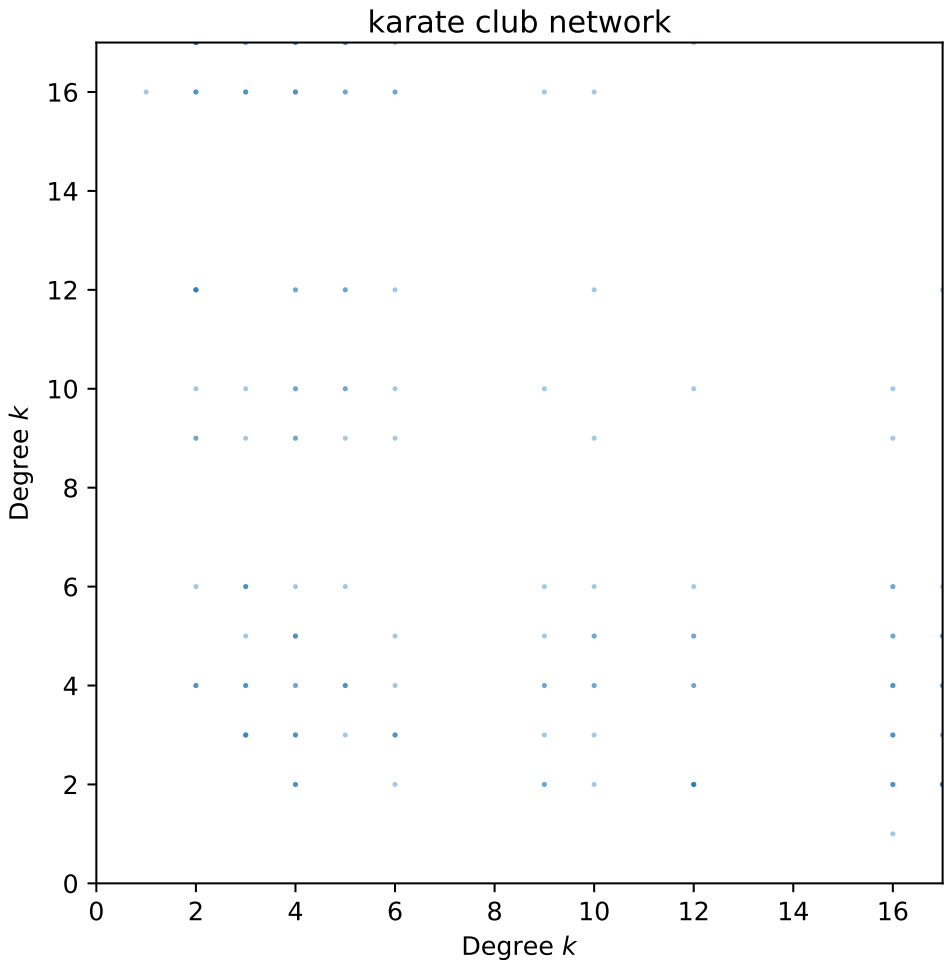

Figure 1.2: Scatter plot - karate club

1. b) Heat Map

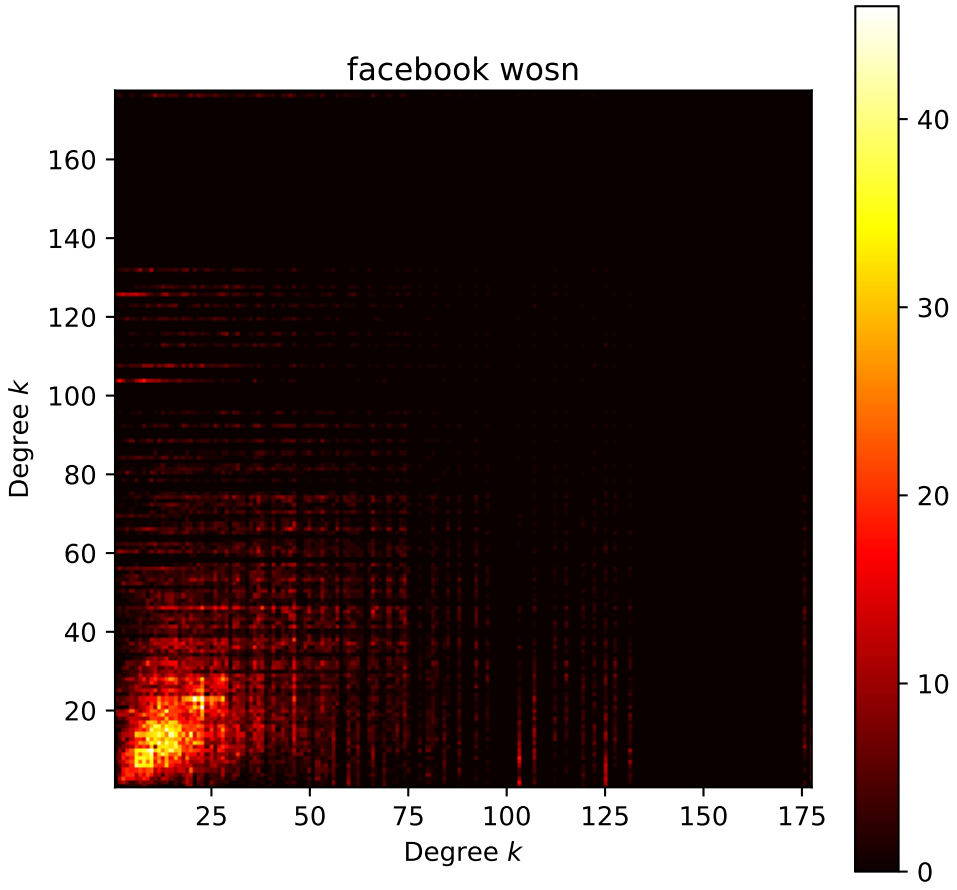

Figure 1.3: Scatter plot - facebook

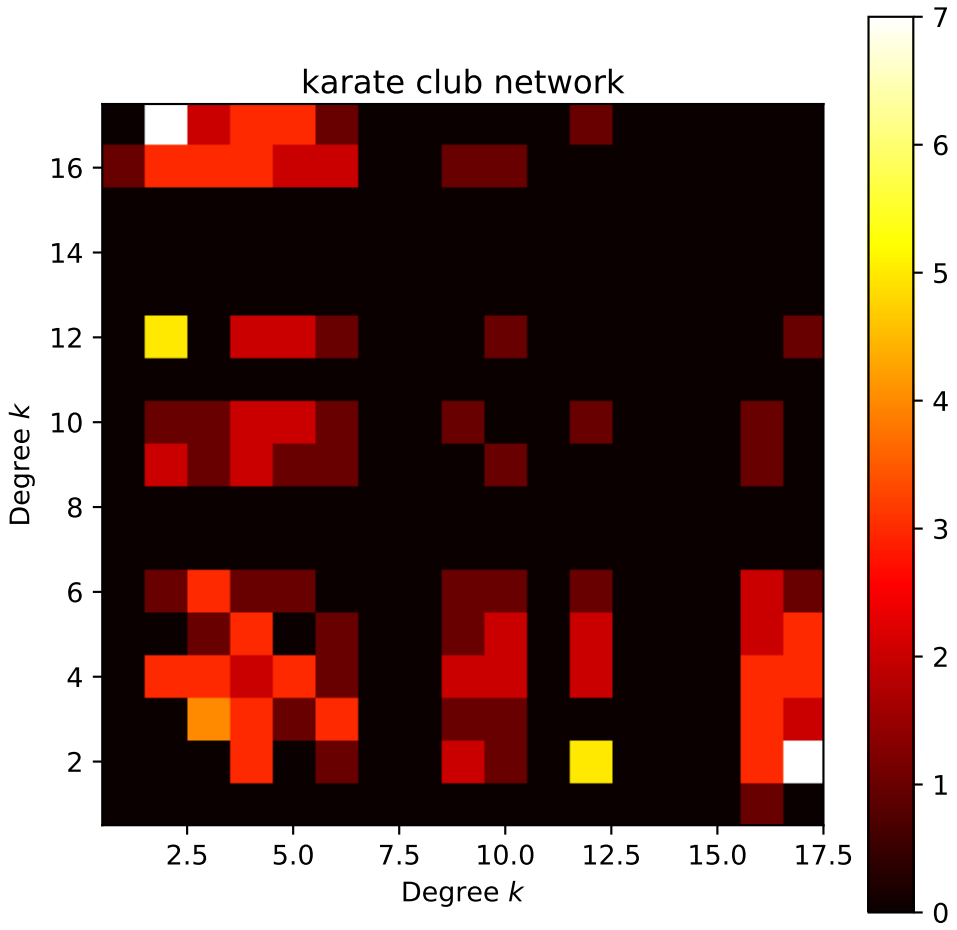

Figure 1.4: Scatter plot - karate club

1. c) Assortativity coefficient

- **Results:**

Facebook

```
Own assortativity for facebook wosn: 0.0559847847659
NetworkX assortativity for facebook wosn: 0.0559847847659
```

Karate Club

```
Own assortativity for karate club network: -0.475613097685
NetworkX assortativity for karate club network: -0.475613097685
```

- **Comments:**

In general, the assortitive value is higher for graph that are social network. The karate club network is not a social friendship, thus it has a low assortative value.

The Facebook network has still a not high value, but it could be explain by the fact that Facebook is not a real social graph, as you can have way more friends on Facebook than in the real life.

1. d) scatter plot of k_{nn} as a function of k

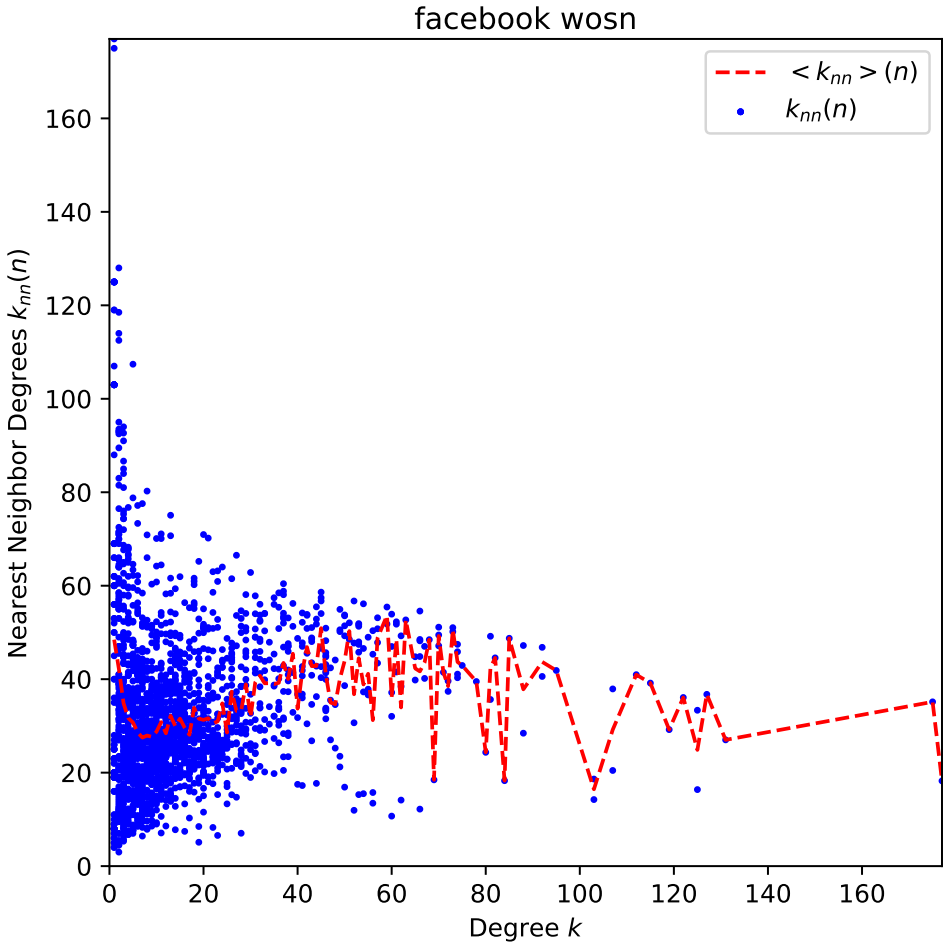

Figure 1.5: Scatter plot - facebook

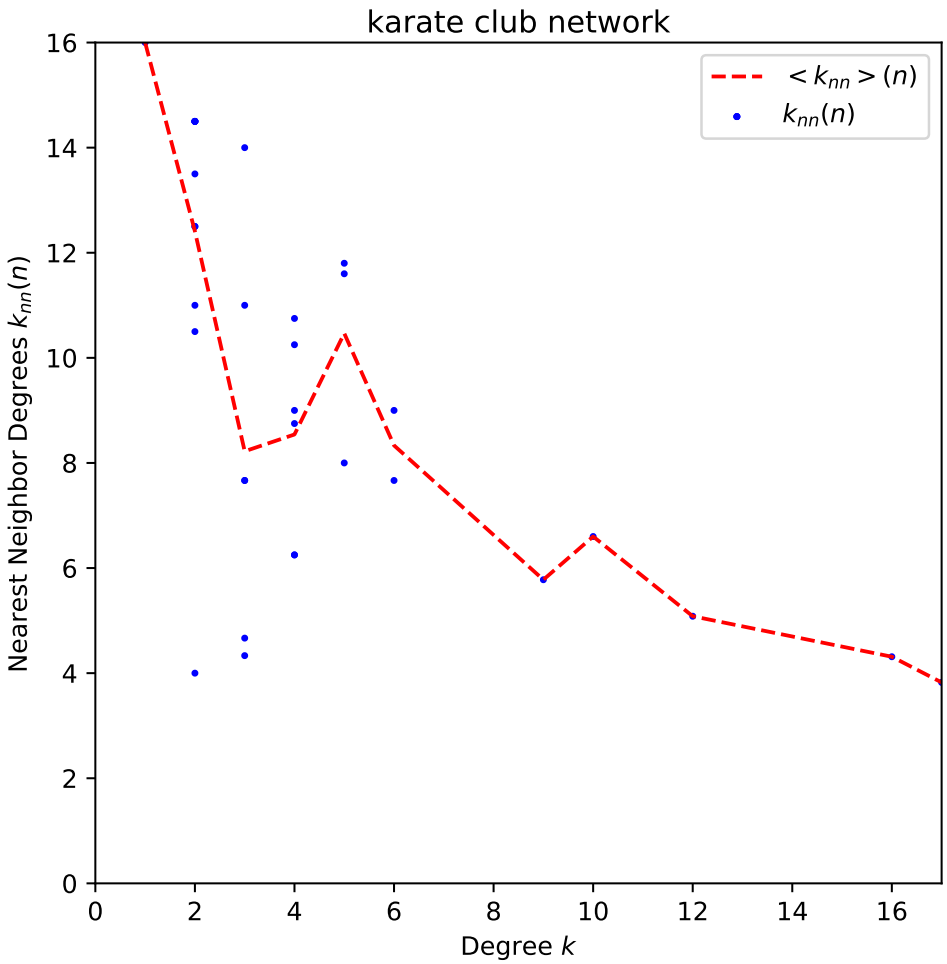

Figure 1.6: Scatter plot - karate club

2 Centrality measures for undirected networks

2. a) Centrality measures

- **Degree**
  - $k(A) = 2$
  - $k\,(B) = 3$
  - $k\,(C) = 4$
  - $k\,(D) = 1$
  - $k\,(E) = 2$

- **Betweenness centrality**

$$\mathrm{bc}(a) = \frac{1}{(N-1)\,(N-2)} \sum_{s \neq a} \sum_{t \neq a} \frac{\sigma_{sat}}{\sigma_{st}}$$

$$= \frac{1}{(5-1)\,(5-2)} \times 0 = 0$$

$$\mathrm{bc}(b) = \frac{1}{12}\left(\frac{1}{2}\right) = \frac{1}{24}$$

$$\mathrm{bc}(c) = \frac{1}{12}\left(1+1+1+\frac{1}{2}\right) = \frac{7}{24}$$

$$\mathrm{bc}(d) = \frac{1}{12}(0) = 0$$

$$\mathrm{bc}(e) = \frac{1}{12}(0) = 0$$

- **Closeness centrality** $C(i)$
  - $C(A) = \frac{5-1}{1+1++2+2} = \frac{2}{3}$
  - $C(B) = \frac{4}{1+1+1+2} = \frac{4}{5}$
  - $C(C) = \frac{4}{4} = 1$
  - $C(D) = \frac{4}{1+2+2+2} = \frac{4}{7}$
  - $C(E) = \frac{4}{1+1+2+2} = \frac{2}{3}$
- **K-shell** $k_s(i)$
  - 1-shell = [A, B, C, D, E]
  - 2-shell = [A, B, C, E]

2. b) Centrality measures

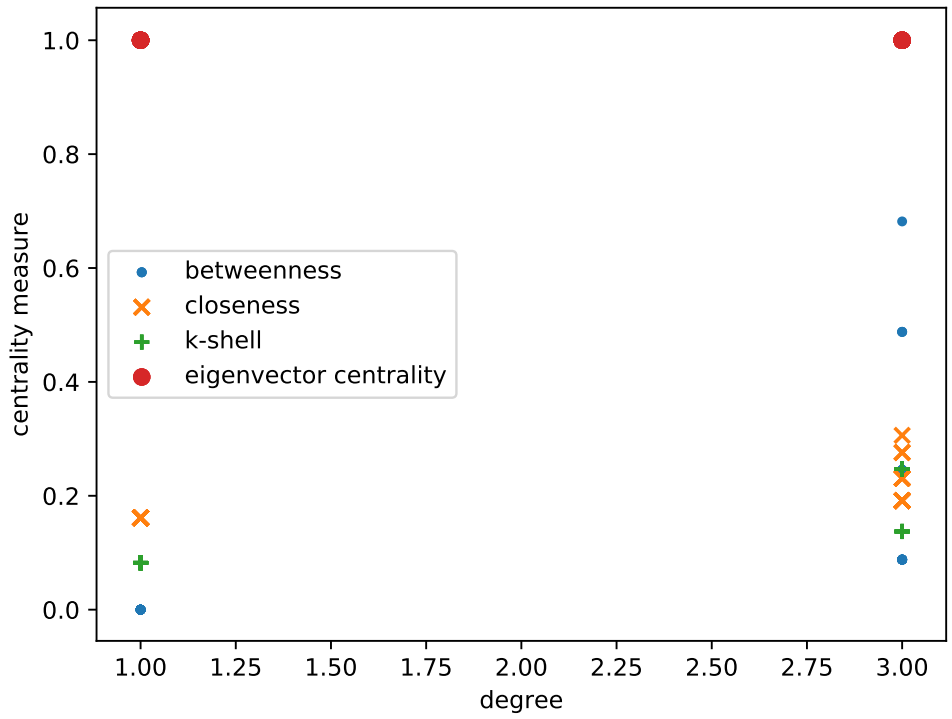

Figure 2.1: Cayley tree

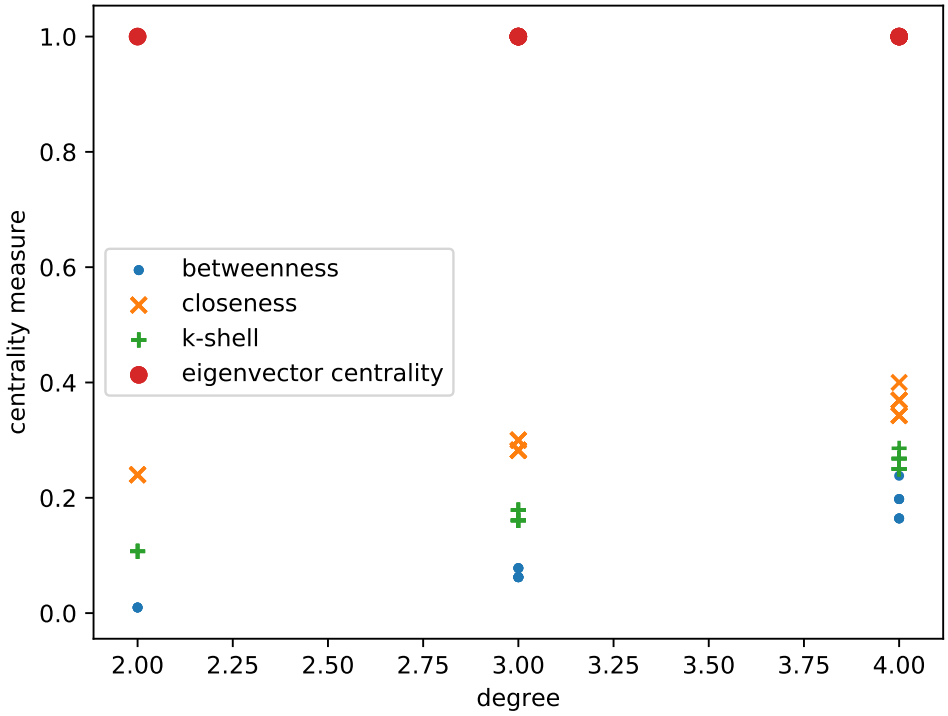

Figure 2.2: lattice

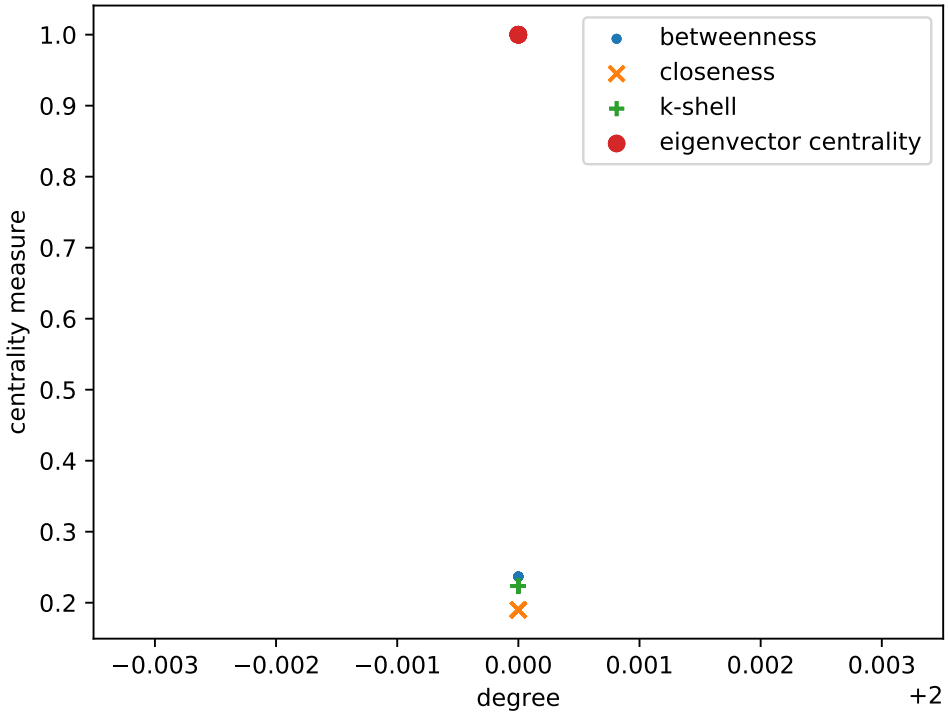

Figure 2.3: Ring

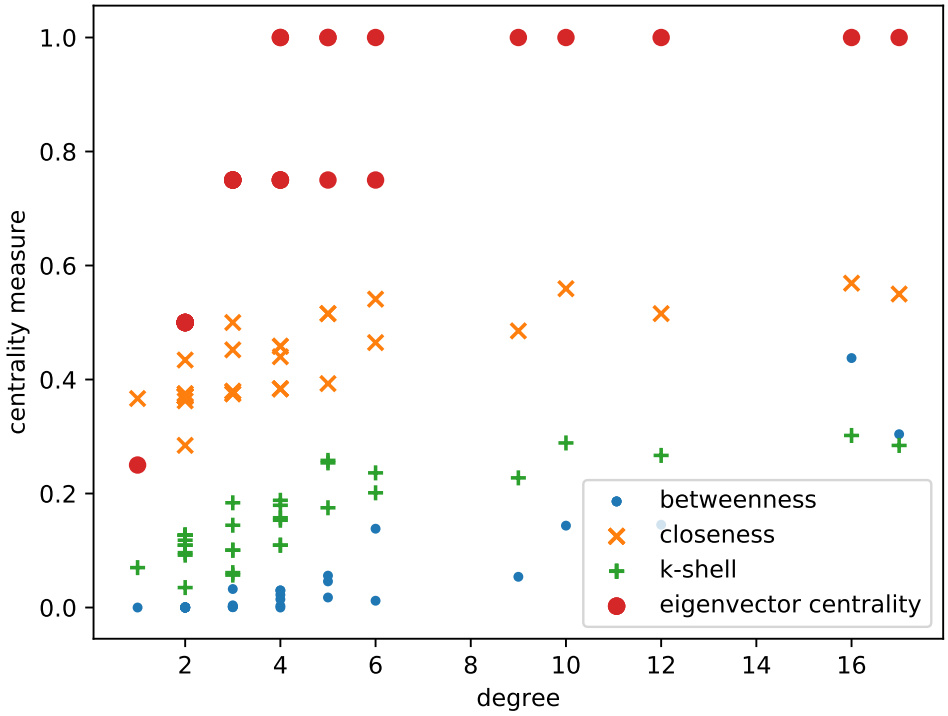

Figure 2.4: Karate club network

2. c) Visualizations

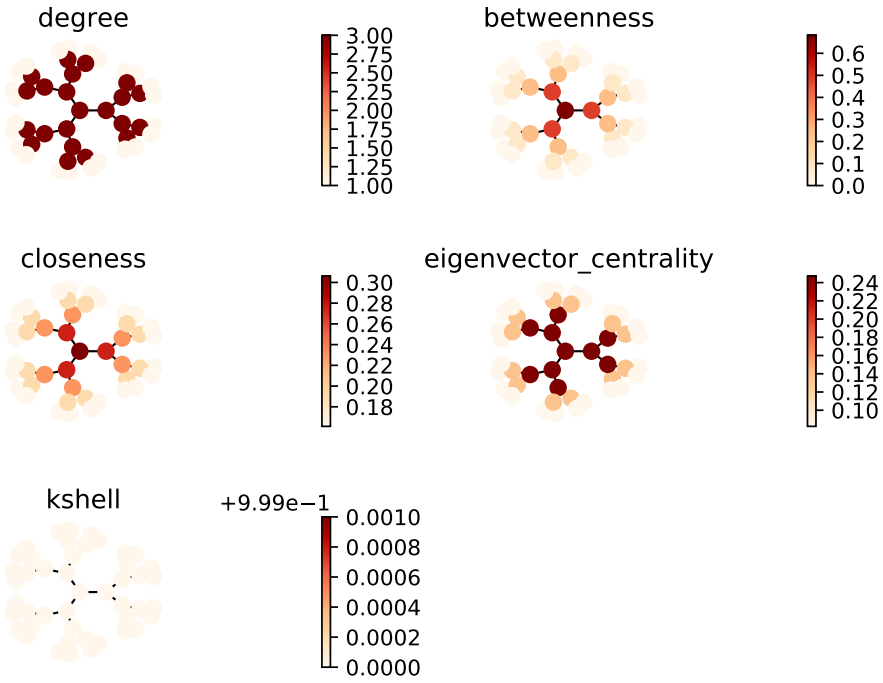

Figure 2.5: Cayley tree

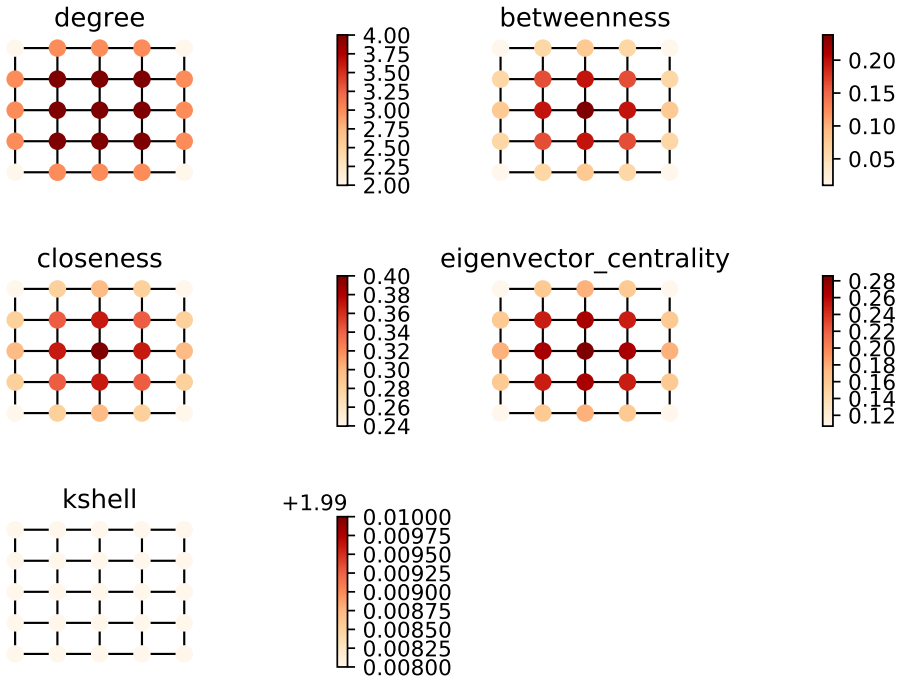

Figure 2.6: lattice

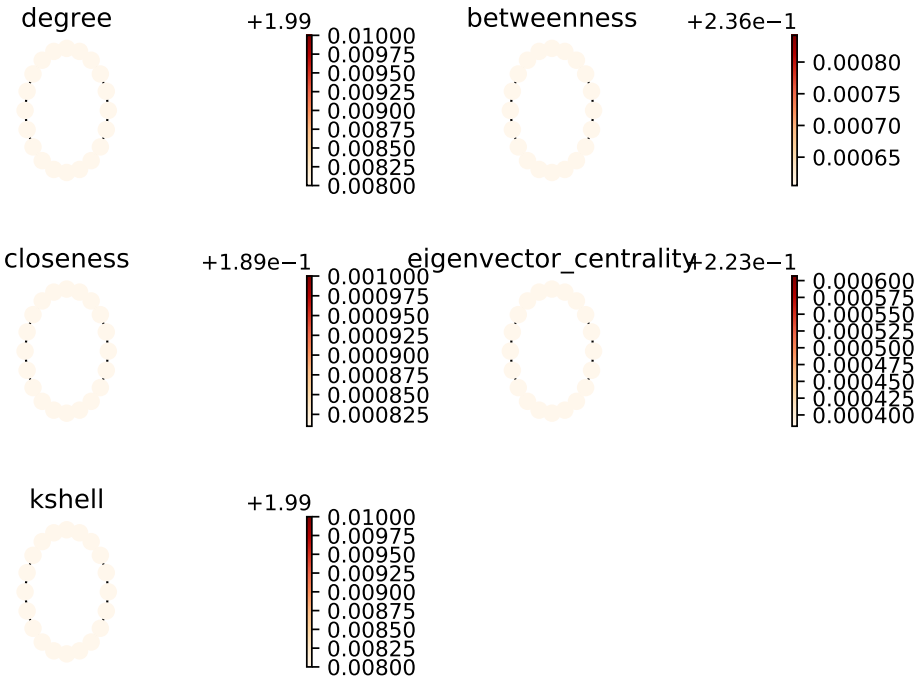

Figure 2.7: Ring

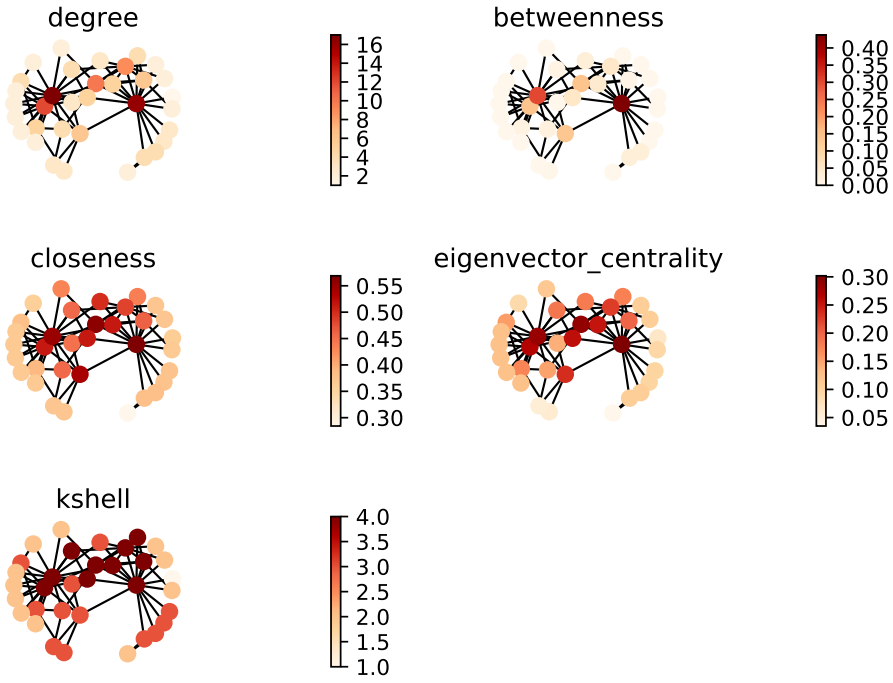

Figure 2.8: Karate club network

3 PageRank (directed network)

3. a) Display the network

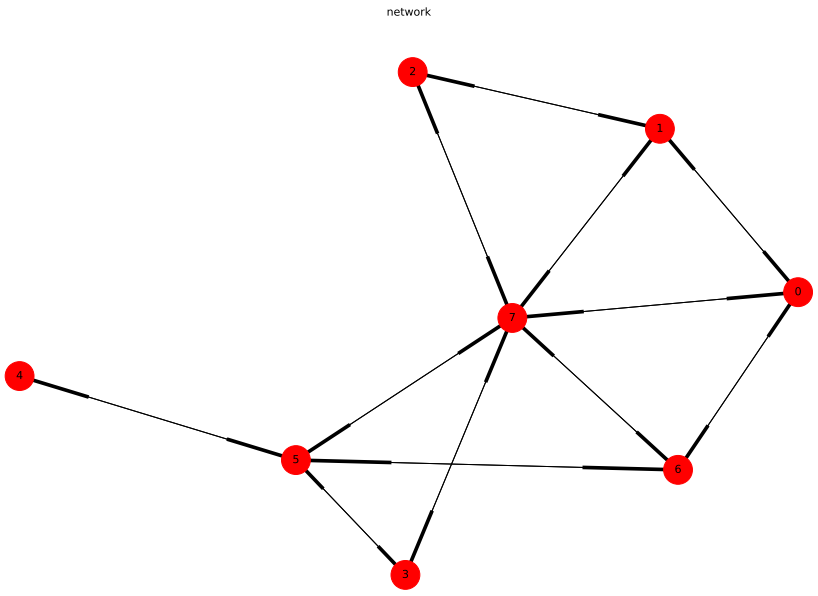

Figure 3.1: page rank - test network

3. b) Naive PageRank

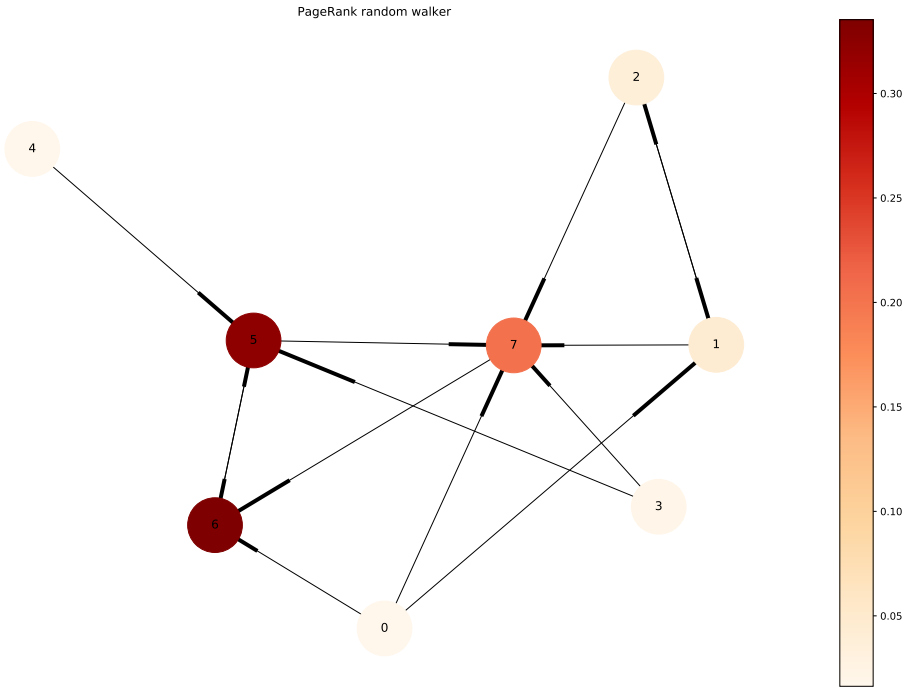

Figure 3.2: naive page rank - $d = 0.85$ $N_{\text{steps}} = 10\,000$ on the test network

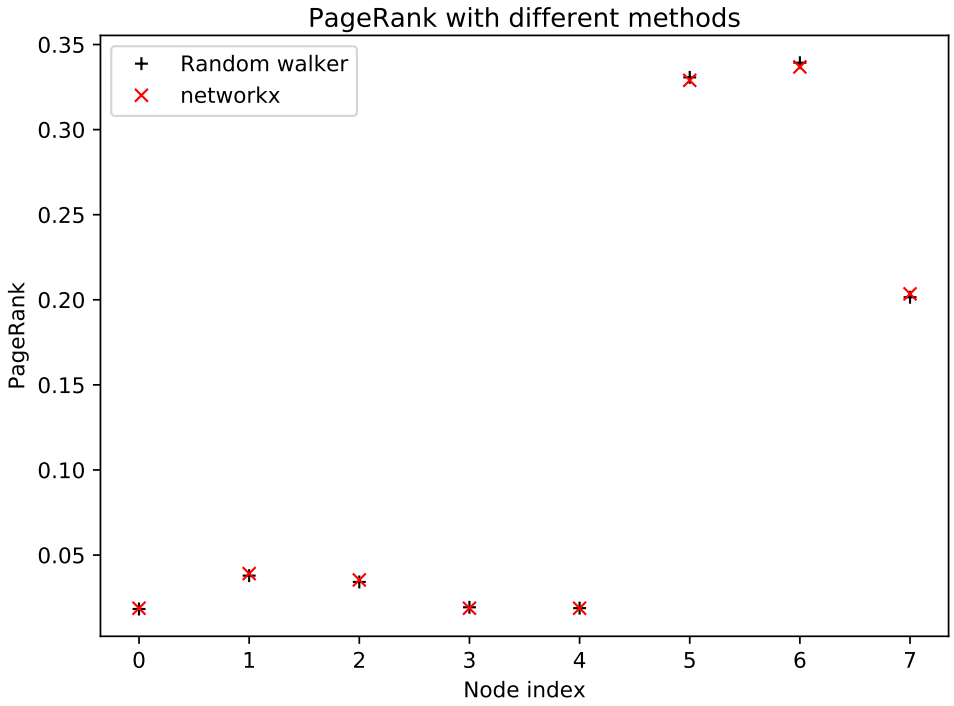

Figure 3.3: naive page rank - $d = 0.85$ $N_{\text{steps}} = 10\,000$ on the test network

3. c) Power iterations PageRank

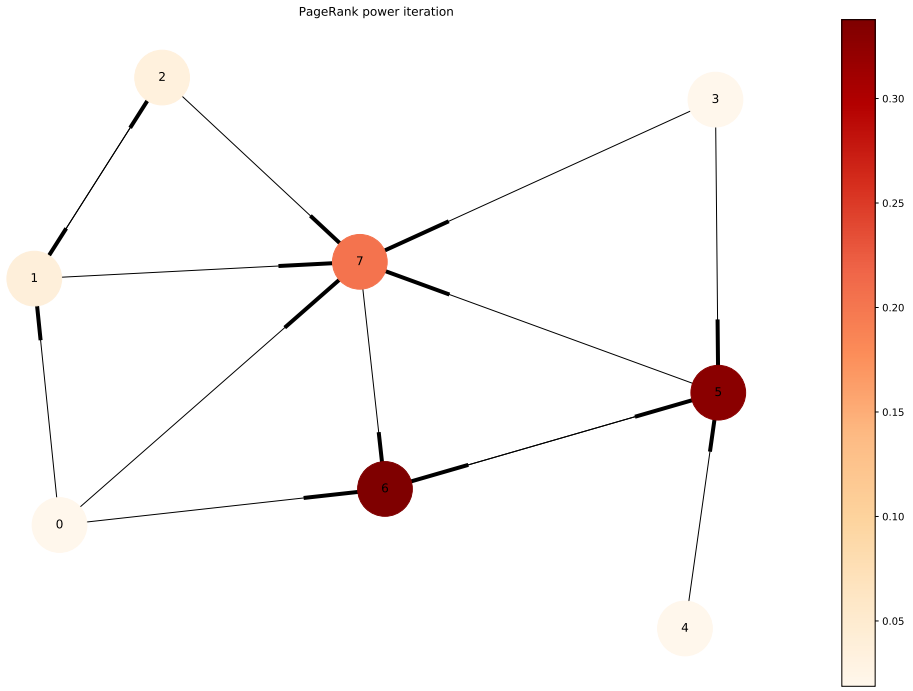

Figure 3.4: power iteration page rank - $d = 0.85$ $N_{\text{steps}} = 10\,000$ on the test network

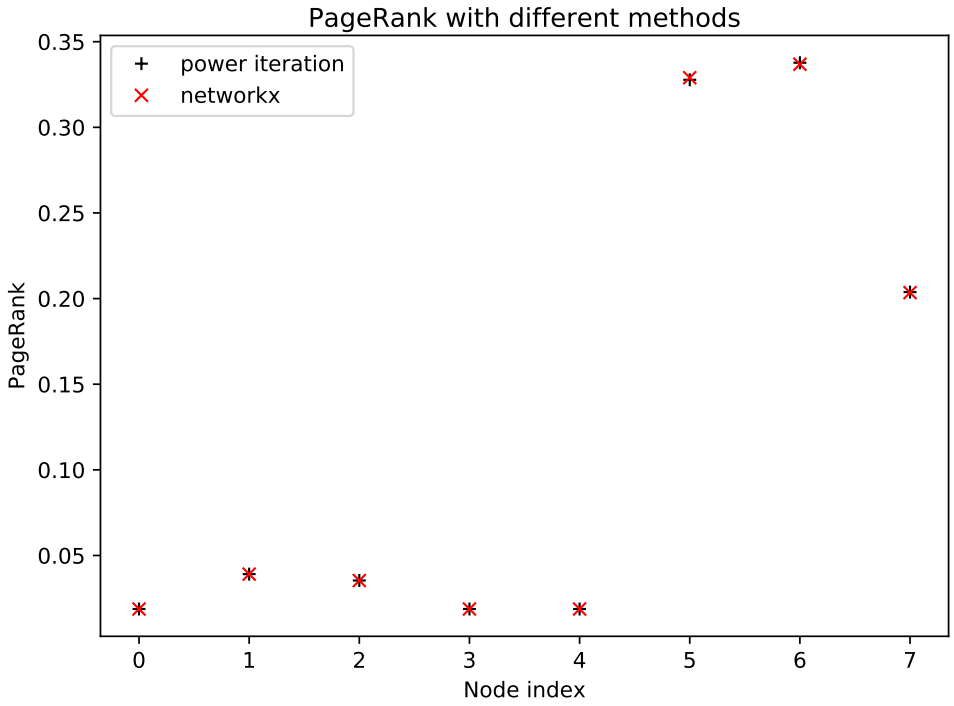

Figure 3.5: power iteration page rank - $d = 0.85$ $N_{\text{steps}} = 10\,000$ on the test network

3. d) Estimating PageRank computation time

```
power iteration (k5 size = 10 ** 3) = 0.7829688139972859
random walker (k5 size = 10**3) = 92.53090116499516
```

Then, it we assume that the computation time is linear with respect to the size of the network, then power iteration should take 20 357.18 secondes ($\sim$ 5hours) on the full network. The random walk would take $\sim$ 668.27 hours.

3. e) How the network's structure relates to PageRank

- What is the connection between degree k or in-degree k in and PageRank?

The degree seems more or less correlated with the page rank. A high in-degree node has a good probability to yield a good page rank.

- How does PageRank change if the node belongs to a strongly connected component?

The page rank improves if the node belongs to a strongly connected component because it is more likely that the walker will stay in the component.

- How could this information be used in improving the power iteration algorithm given in part c)?

We could use this value to provide a better values for the starting point of the algorithm. We should take the degree into account, thus it could converge faster.

3. f) Dampling factor

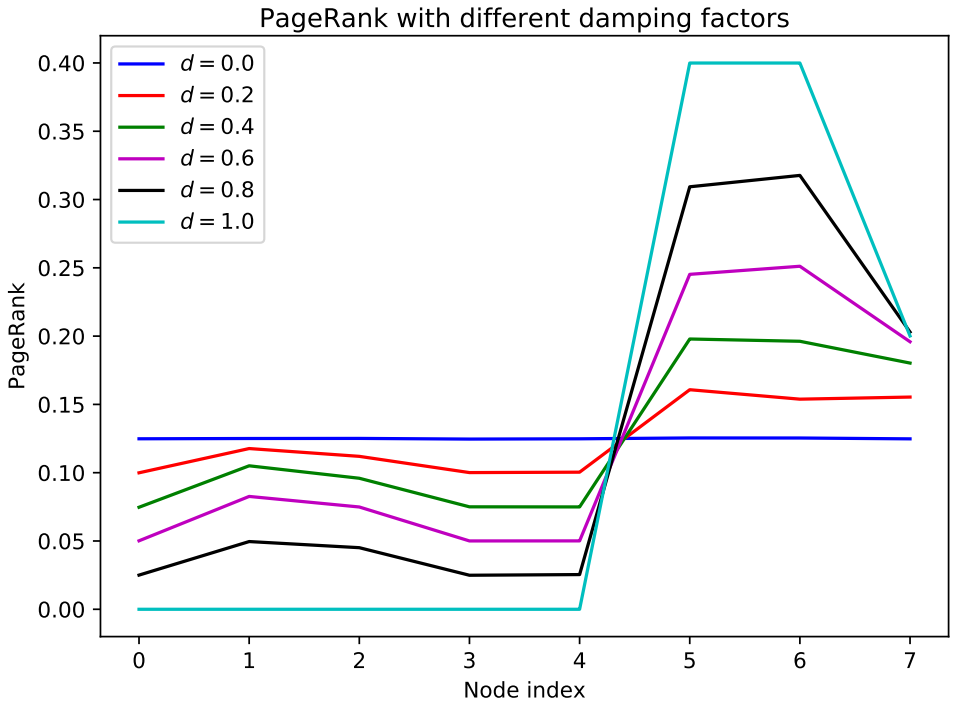

Figure 3.6: Comparaison of different dampling factor

- How does the change of d affect the rank of the nodes and the absolute PageRank values?

The dampling factor d seems to influence the contrast between high rankgpage node value and low rankpage node value. If it too small, then we cannot see very weel how the node of each node is with respect to the others. If it is too high, then the high page rank value is kind of "too saturated", so you can not make a difference any more..

3. g) Page rank wikipedia

Results

- Highest PageRank:
  - 0.03519319071432259 : Graph_theory
  - 0.020361350619844686 : Social_network
  - 0.01677151139830182 : Mathematics
  - 0.01646208363207607 : Social_network_analysis
  - 0.014703296264824407 : Social_networking_service
- Highest In-degree:
  - 82 : Social_network
  - 73 : Social_network_analysis
  - 63 : Small_world_experiment
  - 62 : Social_networking_service
  - 62 : Orkut
- Highest Out-degree:
  - 140 : Network_science
  - 82 : Social_network
  - 73 : Social_network_analysis
  - 67 : Small-world_network
  - 65 : Sexual_network

Comment the differences and similarities between the three lists of most central pages

- "Graph therory" and "Mathematics" have a high page rank, but they suprisingly don't have belong to the tops highest in-degree. For example, "social network" is in the top three each time.
- "Network science" lead to many pages, but not there are not a lot of pages that link back to it.
- "social network" and "social network analysis" have both correlated subject, that's why they have a good ranking as it can improve this way.