

Homework 2

Diogo Bogas
201202482
Network Science 18/19
May 2019

1 Exercise 1

1.1 item a

1.1.1 point i

The produced network has undirected edges, and after all the initial processing, I found out the following about the network composition:

1. Nodes: 916 nodes
2. Edges: 1350 edges
3. Average degree: 2,948
4. Average weighted degree: 0.692
5. Number of connected components: 3
6. Size of the giant component 908 nodes and 1344 edges

1.1.2 point ii

Only for the giant component we then measured:

1. Diameter (normalized centralities): 8
2. Average path lenght: 4.1957
3. Top 3 nodes (betweenness centrality): English, French and India
4. Top 3 nodes (closeness centrality): English, United States, Germany

1.1.3 point iii

Follows a screenshot of the derived network, with the nodes colored by group. Rose is group 1, which reflects languages, and green is group 0, that reflects countries. 72.6% of the nodes belong to group 1 and the remaining 27.4% belongs to group 0.

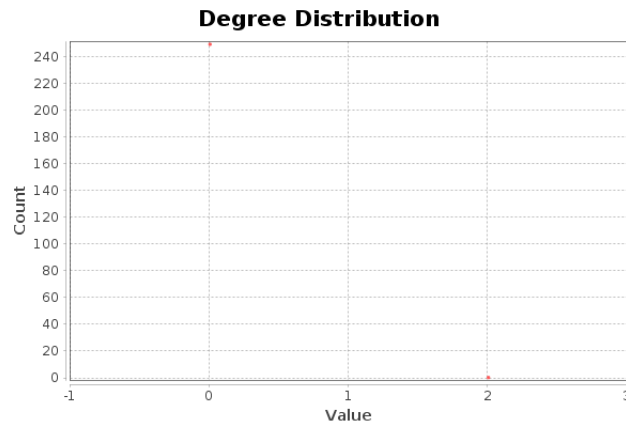


1.2 item b

To plot both degree distributions, I used Gephi. To isolate the projections, I used a filter, by a property called `bipartite_group`, that I decided to add. We begin with group 0(countries) in the bipartition, colored in green: It is a scarce

Results:

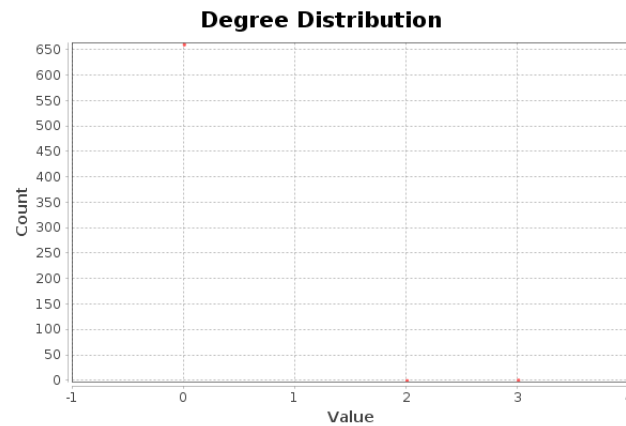
Average Degree: 0,008



distribution, and a very low average degree. As we are only looking at 1 of the two projections this is normal. It would even be normal to have avg. degree 0. The avg. degree not being 0 means that a few nodes from every projection are connected among themselves, which is totally fine, or there was an error while parsing the raw data. Lets check the degree distribution for the other projection(languages), to see if the same is observed:

Results:

Average Degree: 0,012



As the same conclusions can be drawn, the very low average degree is justified by the fact that the majority of the edges exist between projections. Finally, to clarify, by sparse projection, one is to take away that with only 3 or 4 points in a plot, it is possible to show the degree distribution for a network with 900+ nodes.

1.3 item c

See unicode.filtered.gephi.

1.4 item d

To answer to this item i used Gephi.

2 Exercise 2

2.1 item a

The command run for this item was:

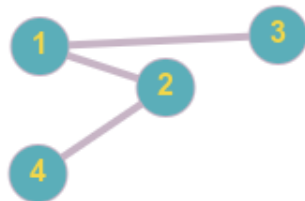
```
./gtrieScanner_src_01/gtrieScanner -s 4 -m esu -g network.txt -f simple
```

And after parsing the results.txt generated, I got the following table

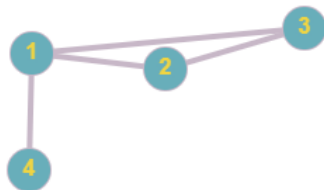
Graph	Org_Freq
a	21
b	5
c	4
d	1
e	1

And the motifs follow:

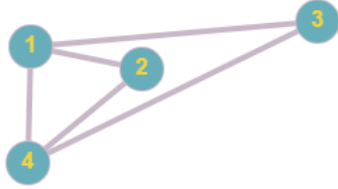
Motif a:



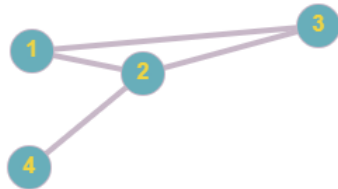
Motif b:



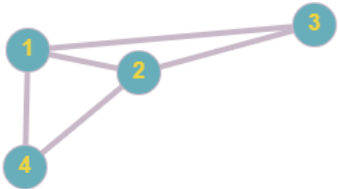
Motif c:



Motif d:



Motif e:



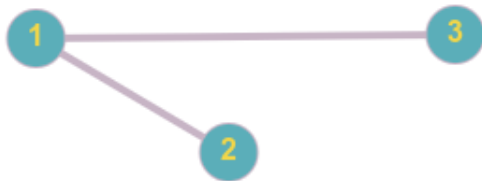
2.2 item b

2.3 item c

After running:

```
./gttrieScanner -s 3 -m gttrie dir3.gt -d -g ecoli.txt -r 500 -raw
```

I observed that the most overrepresented subgraph is:



Its z-score is -7.03, the frequency on the original network is 250 (Org_Freq), the average number of occurrences(Rnd_Avg +/-) is 345.19 +/- and standard deviation on the randomized networks.(Rnd_Dev) is 14.25.

Honestly, after a quick review of both articles I found 0 references to the

subgraph described, be it by it not being reported, or being distracted. As it is fairly simple to look at, ignoring the numbers on the nodes, it is quite clearly a chain. The nodes are not participating in a directed graph, so the chain can be viewed between 3 and 2 in any order.