Amir Imani- ai2335
Lab 3

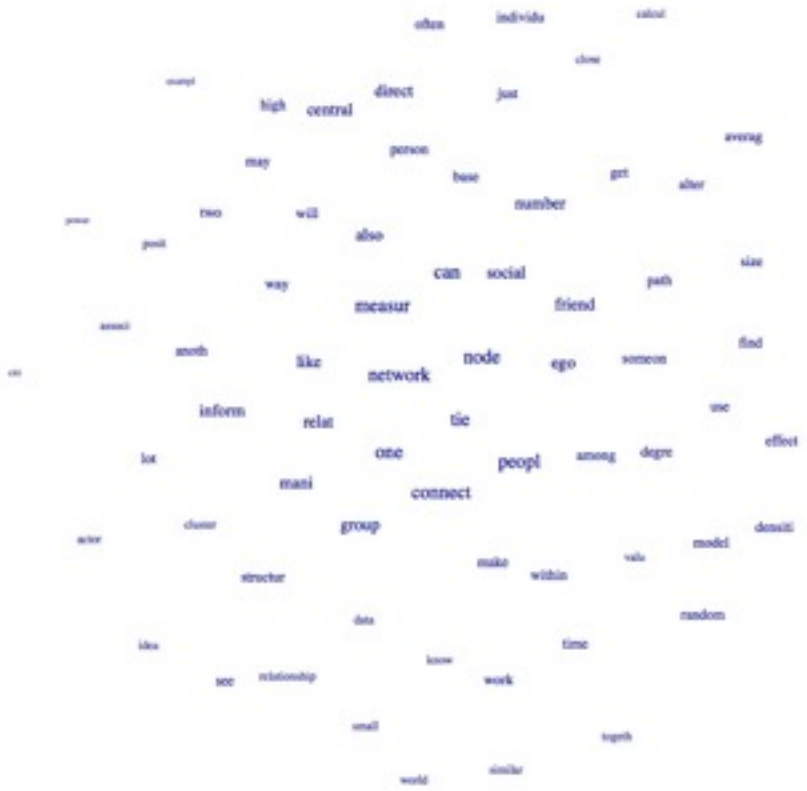
As we discussed_word associations network in the last session, I decided to do the same analysis on Greg's lectures.

I extracted words from presentations, import them in R as a corpus, removed stopwords, numbers, punctuations, etc. Then, I created a Term Document Matrix from the stemmed corpus, and finally created the adjacency matrix based on that.

To make the graph legible, I only used the top 70 words which had a degree higher than 200. The overall topography of the network is as below:

Nodes	Edges	Density
70	1780	0.7162978

Ties in this network represents if two words have been found together in the corpus. The network is shown below.



Nodes with higher degree are larger, and the ones in the centre are more central. Looking at the words, it make sense to have words like “Network”, “Measure” “nodes”, “ties” at the centre and the largest.

After running GirvanNewman and random walk community deletion algorithms, and the results are as below.

GirvanNewman found only one community while random walk with 25 steps found two groups.

Comparing them together yielded the following results:

```
> compare(fgn, fwt, method= c(“nmi"))
0
> compare(fgn, fwt, method= c(“rand"))
0.5859213
> compare(fgn, fwt, method= c(“adjusted.rand"))
0
```

The network with communities found through random walk is shown below. It looks like the blue community is more general concepts (and some sentence fillers like “often”, “may”, etc) comparing to the orange community which is more about network structure.