

Atividade 3 - Visualização de Redes

SCC0252 - Visualização Computacional

15/10/2017

Hugo Moraes Dzin 8532186

Profª Maria Cristina

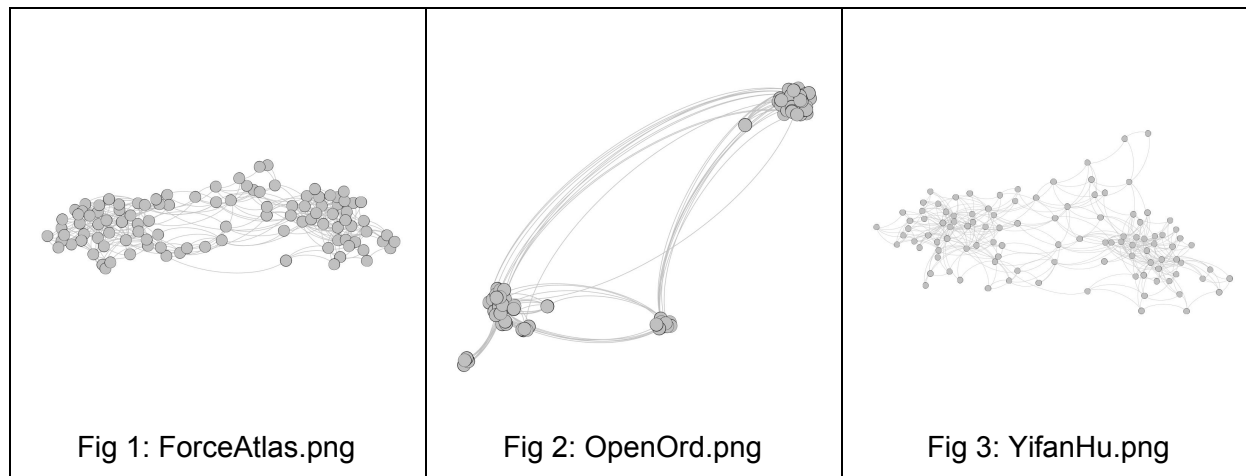
Introdução

O objetivo desta atividade é analisar uma rede utilizando o Gephi e PCA, extraindo medidas como degree e clustering. A rede utilizada foi a *Books about US politics*, disponível em <http://www-personal.umich.edu/~mejn/netdata/>. Ela é um grafo bidirecional, sem pesos, com classes nas arestas, que representa compras de livros. Cada aresta é um livro, que pertence a uma entre três classes: conservador, neutro ou liberal. As arestas ligam livros que foram comprados pela mesma pessoa.

Para executar o script, é necessário executar `python src/t3.py` na **pasta raiz** da atividade. Todos os arquivos referentes ao Gephi (projeto, csv, imagens) estão em `gephi/`. As imagens geradas pelo script estão em `plot/`.

Visualizações da Rede

Dentre os layouts disponíveis, foram testados *Yifan Hu*, *OpenOrd* e *Force Atlas*. Os três conseguem mostrar os principais clusters de nós, mas *OpenOrd* e *Force Atlas* tendem a comprimir demais os grupos de nós, o que dificulta um pouco a visualização. Por isso, nos passos seguintes, apenas *Yifan Hu* foi utilizado.



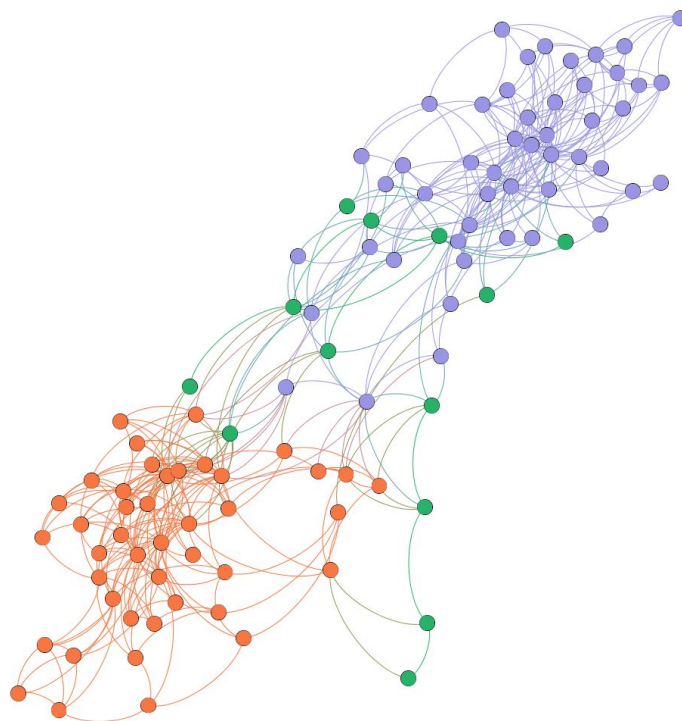
Em seguida, foram extraídas quatro medidas: Grau, Betweenness, Closeness e Clustering. Também foi utilizado a classe dos nós (rotulada como "value"). Foram geradas três visualizações:

- Mapeando cor à classe do nó
- Mapeando cor ao grau, tamanho ao betweenness
- Mapeando cor ao clustering, tamanho ao closeness

No primeiro mapeamento (figura 4) podemos notar que o layout mapeia bem as classes de nó: as classes conservador e liberal tendem a ficar separados, e a classe neutro tende a ficar no meio do caminho entre os dois clusters.

O mapeamento de grau e betweenness (figura 5) nos indica que, em geral, nós com grau alto (coloração azul) tendem a ter betweenness alto, assim como os nós intermediários entre os dois clusters.

Já no mapeamento 3 (figura 6), podemos ver que clustering alto corresponde à closeness menor. Além disso, também é perceptível que os nós intermediários são os que tem o clustering mais baixo e o closeness maior.



Legenda

- Conservador
- Neutro
- Liberal

Fig 4: Visualização das classes dos nós
value.png

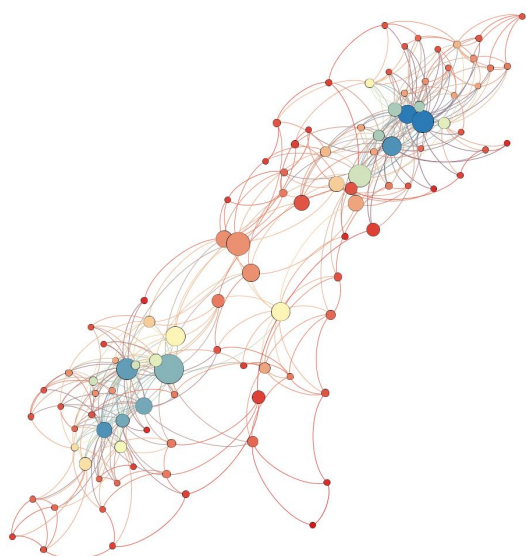


Fig 5: cor ⇒ grau; tamanho ⇒ betweenness

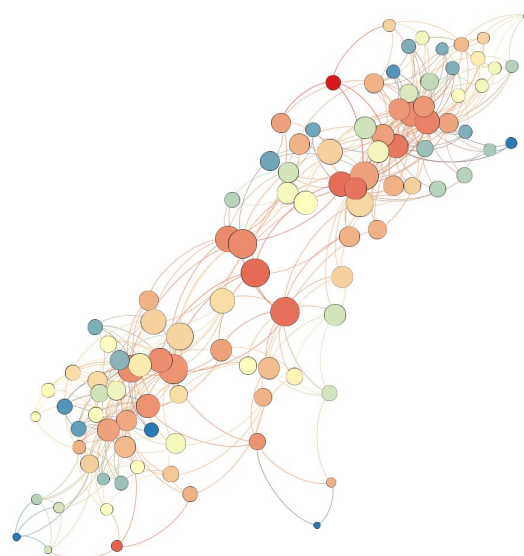


Fig 6: cor ⇒ clustering; tamanho ⇒ closeness

Análise de Componente Principal

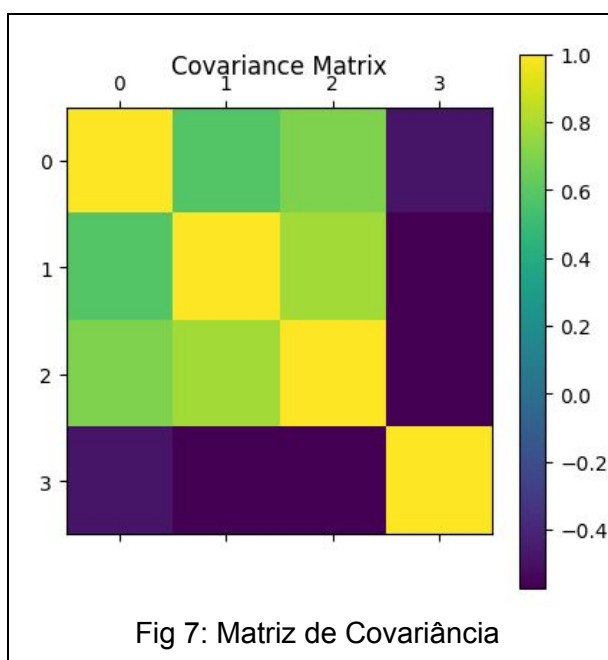


Fig 7: Matriz de Covariância

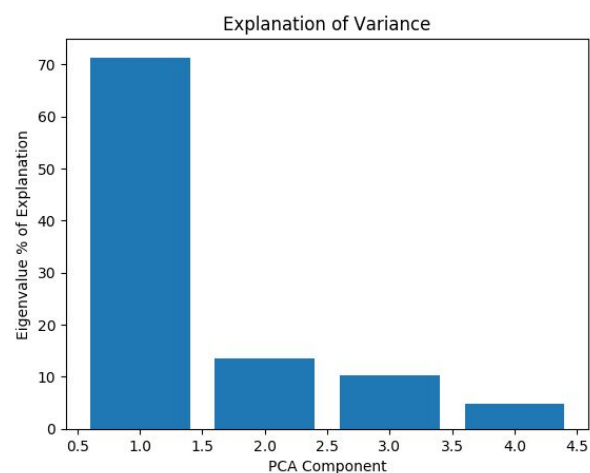


Fig 8 : Peso dos autovalores do PCA

Pela matriz de covariância, notamos que há 3 variáveis muito correlacionadas: degree (0), closeness (1) e betweenness (2); a última, clustering (3), não correlaciona bem com nenhuma das outras. Na figura 8, podemos ver que o primeiro componente principal já explica quase toda a variância nos dados, talvez devido à correlação descrita anteriormente.

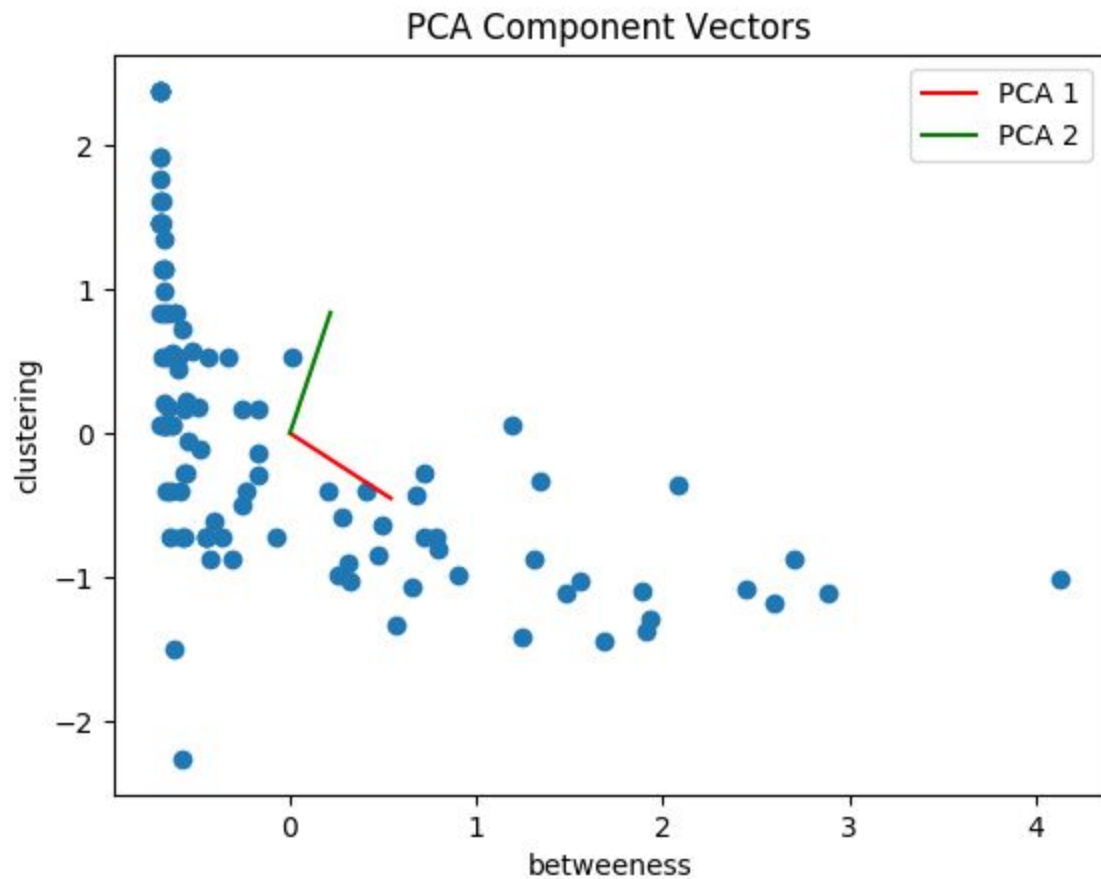


Fig 8: Vetores do PCA vs dados da rede