



Aula 3 - Análise estrutural de uma amostra do Twitter

José Devezas <jld@fe.up.pt>

Faculdade de Engenharia da Universidade do Porto

Gestão da Informação em Redes Sociais

Mestrado em Ciência da Informação



Conteúdos

- Introdução
- Conceitos
- Conjunto de dados do Twitter
- 1ª avaliação



Introdução

- Apresentação de novos conceitos úteis à análise de redes (p.e., coeficiente médio de agrupamento).
- Caso de estudo: apresentação de uma rede social com base em interações (menções e *retweets*).
 - Método de recolha e amostragem.
 - Características básicas da rede.
 - Início dos trabalhos individuais e visão geral do trabalho de grupo.



Conceitos

- **Densidade do grafo** \Rightarrow fração de arestas presentes no grafo, face ao número máximo de arestas possíveis.
 - Um grafo 100% denso é um grafo completo (todos os nós ligam a todos os outros nós; p.e., um clique).



Conceitos

- **Coeficiente médio de agrupamento** \Rightarrow coeficiente de agrupamento de um nó é igual à fração de arestas entre os vizinhos de um nó, face ao número máximo de arestas possíveis.
 - O coeficiente de agrupamento é 100% para um nó quando todos os seus vizinhos estão ligados entre si.
 - **Intuição:** mede a densidade da vizinhança.
 - O coeficiente médio de agrupamento para um grafo é igual à média dos coeficientes de agrupamento de todos os nós do grafo.
 - Considera-se que se verifica o *small-world effect* num grafo quando o coeficiente médio de agrupamento é significativamente maior do que num grafo aleatório equivalente (i.e., um grafo com o mesmo conjunto de nós e arestas geradas aleatoriamente).



Conceitos

- **Componentes fortemente e fracamente ligados** \Rightarrow num componente ligado, existe sempre um caminho entre qualquer par de nós.
 - Num grafo dirigido, um componente é fortemente ligado quando esse caminho é dirigido ou fracamente ligado quando se considera a versão não-dirigida.
 - Num grafo não-dirigido apenas existem componentes fracamente ligados.



Conceitos

- **Atingibilidade (*Reachability*)** \Rightarrow refere-se à habilidade de chegar de um vértice a outro vértice no grafo.
 - Um vértice t é atingível a partir de um vértice s , se existir uma sequência de vértices adjacentes (i.e., um caminho no grafo), que começa em s e acaba em t .
 - No Gephi, a atingibilidade pode ser testada através do ícone do avião (quando não existe caminho mais curto, também não é possível ligar dois vértices).



Conjunto de dados do Twitter

Amostragem

- **2** *seed users*: @realDonaldTrump e @elonmusk.
- Recolha feita com base na API¹ de pesquisa do Twitter (p.e. pesquisando “@elonmusk”).
- Amostragem aleatória de **5** utilizadores mencionados nos tweets e de **5** utilizadores autores de tweets.
- Repete-se **3** vezes o processo para cada um dos 10 (**5+5**) utilizadores amostrados.

Number of seed users	2
Maximum distance	3
Sample size	5
Twitter Search API Rate Limit	450
Predicted Time in Hours	1:06:40
Predicted Time in Days	0.05

¹ Application Program Interface (API)



Conjunto de dados do Twitter

Caraterização

- Num. Vértices 39.129
- Num. Arestas 273.434 (108.041, removendo múltiplas arestas)
 - *mentioned* 238.551
 - *retweeted* 34.883
- Densidade do grafo 0,0001785937 (ignorando *loops*, i.e., ligações para o próprio)
- Diâmetro 25 (15 na versão não-dirigida)
- Tam. caminho médio 10,73 (6,23 na versão não-dirigida)

Nota: Os valores aqui apresentados foram calculados sobre o grafo original. No Gephi, ao carregar um grafo, são removidas múltiplas arestas, ignorando as *labels*, isto é, uma relação *mentioned* e uma relação *retweeted* entre o mesmo par de nós conta apenas uma vez como uma interação.



1ª avaliação

- Individualmente (ou em grupo, caso não haja computadores para todos):
 - Para a caracterização global do grafo e a detecção de comunidades podem trabalhar em grupo.
 - Cada aluno do grupo deve extrair um subgrafo diferente, referente a uma comunidade, para estudar.
 -
- Individualmente:
 - Estudar a estrutura social do grafo para a comunidade extraída.
 - Anotar estatísticas e observações numa folha de cálculo ou num documento.
 - Escrever um mini-relatório (uma ou duas páginas), descrevendo a estrutura social da comunidade do Twitter estudada, utilizando uma linguagem acessível (não matemática), e assumindo um destinatário não-técnico e que desconhece as métricas e o vocabulário da análise de redes.
- Entrega:
 - Pode ser feita até à próxima aula.