

Redes Complexas

Projecto 1



Miguel Martins Duarte
nº 85016

Introdução

Neste trabalho pretendeu-se, partindo de uma implementação em java para cálculo da betweenness, fazer algumas análises dessa métrica em alguns datasets.

O primeiro objectivo estipulado foi o de avaliar qual o impacto quando se procura obter os nós com maior betweenness, executando o algoritmo de BFS apenas para parte dos vértices da rede. O problema coloca-se pois, como sabemos, a complexidade temporal de $O(N * E)$ é muito pesada para redes de tamanho comum, com $N > 10^6$, por exemplo.

De seguida, fez-se uma análise do impacto do grau de um nó na sua betweenness.

Os datasets utilizados, referidos como a rede eswiki (rede da wikipédia espanhola em 2013) e hollywood (rede de actores de hollywood, em 2009).

Relativamente à implementação

O algoritmo utilizado é proposto por Ulrik Brandes¹. O algoritmo foi implementado em paralelo, funcionando em múltiplos cores e em várias máquinas simultaneamente, já que devido ao tempo que este demorava a executar, se decidiu que seria boa ideia correr o algoritmo de forma distribuída no cluster do IST.

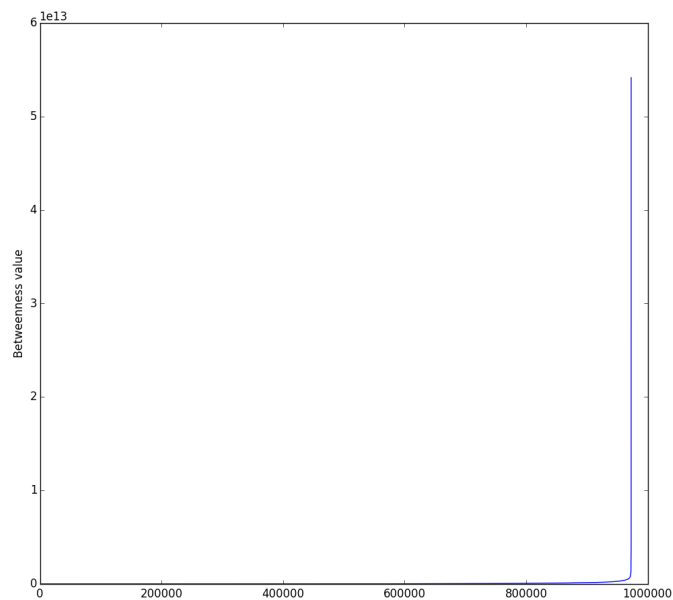
O código procurou utilizar o webgraph na medida do possível. Para correr, deve-se utilizar `./run basename p`, sendo `p` a proporção de nós que se pretende utilizar para o algoritmo de BFS e o `basename` é o nome do grafo como é entendido pelo webgraph. Para correr noutras máquinas, deve-se correr com a flag `-c` no fim do comando, adaptar os ips no código e correr o programa `./runService` nas restantes máquinas do cluster.

A complexidade do algoritmo paralelizado é semelhante, $O(\frac{N \times E}{k} + N)$, com `k` igual ao total de cores em todas as máquinas, sendo que apenas a soma final precisa de correr no mesmo processador.

¹ <http://algo.uni-konstanz.de/publications/b-fabc-01.pdf>

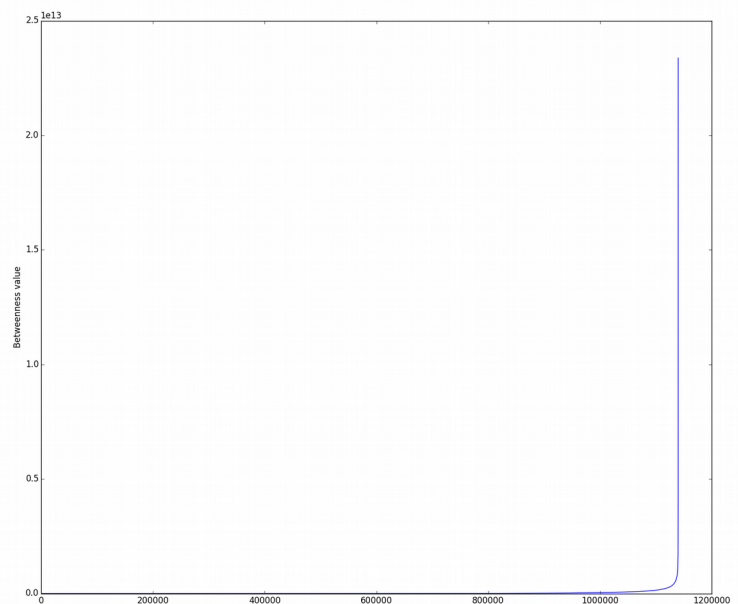
Distribuição de betweenness

Começo por apresentar uma distribuição da betweenness para a rede eswiki e da rede hollywood. Os nós estão ordenados por essa métrica.



Rede eswiki

Rede hollywood

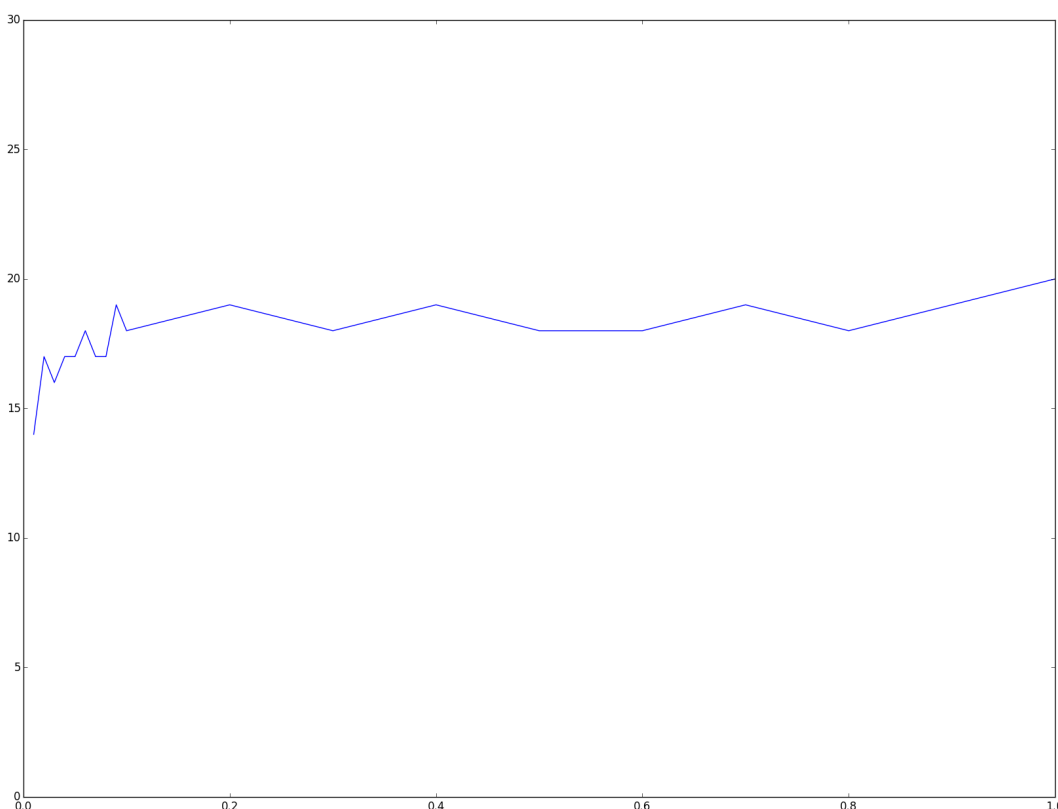


Como podemos ver, há muitos nós com betweenness relativamente baixa ou igual a zero, enquanto que outros têm um valor muito alto. O que significa numa pequena parte dos vértices passam quase todos os caminhos mais curtos.

Sampling de nós para cálculo aproximado da betweenness

Procurou-se observar o efeito de calcular quais os 20 elementos da rede com maior betweenness, utilizando pesquisa em largura para número variável de nós, de 0.01% a 0.09% e de 0.1% a 1% da rede. Esse número variável de nós foi obtido escolhendo, para cada execução, nós de forma totalmente aleatória.

Com base nesses dados, elaborou-se o seguinte gráfico, que representa o número de elementos encontrados em comum com o encontrado com 1% da rede. A comparação poderia ter sido feita relativamente ao cálculo da betweenness com 100% dos nós da rede, mas tal iria requerer muito tempo. Ainda assim, pela estabilização que o gráfico aparenta atingir, o resultado real deve ser muito perto do obtido para o 1% da rede. Utilizou-se a rede eswiki.



Com um pouco mais de tempo, teria sido interessante explorar outras técnicas de sampling, nomeadamente a apresentada neste documento².

2 <http://matteo.rionda.to/papers/RiondatoKornaropoulos-BetweennessSampling-DMKD.pdf>

Tentativa de correlacionar grau com betweenness do nó

Procurou-se observar se existe alguma relação entre o grau de um nó e a sua betweenness. Para isso, fez-se a média desta medida para os vértices de cada grau. Não se conseguiu atingir uma correlação nesses dados, pelo que podemos assumir que a betweenness é relativamente independente do grau. Se o grau de um nó for elevado, poderá indicar que a betweenness é elevada, mas se o grau for pequeno não se pode concluir nada.