

Algoritmos e Estruturas de Dados 2 - EP2

Lourenço Henrique Moinheiro Martins Sborz Bogo

1 Escolhas de Implementação

Primeiro problema que encontrei ao começar a implementar o EP foi: como eu saberia em que vértice uma certa palavra está, de maneira rápida? Para resolver isso, eu decidi usar uma HashTable, onde as chaves são as strings que contêm as palavras e os valores são os nodos nos quais elas estão.

Para implementar a HashTable, decidi fazer minha própria biblioteca de lista ligada, que transformei em uma biblioteca de fila para que pudesse ser usada na BFS também.

Depois disso, implementar quase todas as funções do EP foi muito simples, com exceção das `emCiclo`, que demorei um pouco para pensar como fazer. Decidi implementar quase tudo usando DFS, menos o cálculo da distância onde optei por usar BFS.

Tive que fazer duas funções DFS auxiliares:

`int dfs(int)` Roda uma dfs para o nodo passado como parâmetro e devolve o tamanho de sua componente.

`void dfs(int, int, int, bool, int)` Essa dfs foi a parte mais complicada de implementar do EP. Caso sejam passados só os 3 primeiros parâmetros para ela, ela irá procurar se existe ou não um ciclo que contém o primeiro parâmetro. Caso os 5 parâmetros sejam passados, a função irá fazer quase a mesma coisa, com a condição de que ao invés de buscar um ciclo no primeiro parâmetro, ela irá buscar um ciclo no primeiro parâmetro, que contenha o último parâmetro.

Além dessas funções, implementei mais algumas auxiliares para descobrir a maior componente, a menor componente e o tamanho médio da componente, que são calculadas junto com o número de componentes, ou seja, quando um desses dados é pedido, todos são calculados para ganhar tempo.

Para achar a distância média entre vértices, preferi não usar a função que acha a distância entre dois vértices, pois ficaria muito lerdo. Preferi fazer uma função separada que roda uma BFS para cada nodo e depois tira a média de todas as distâncias encontradas.

2 Como usar

Para usar o EP, deve-se passar como argumento na linha de comando o `k` descrito no enunciado.

Depois disso, basta digitar `help` na prompt, e aparecerá explicações para todos os comandos possíveis.

3 Experimentos

Vamos testar como o tipo do texto, sua língua e seu tamanho influenciam os resultados.

3.1 Romances

3.1.1 Português

k=1

Livro	Vertices	Arestas	Componentes	Maior Comp.	Menor Comp.	Média Comp.	Dist. Média	Denso
cidade-e-as-serras.txt	13405	9598	7444	3248	1	1.799	8.861	False
domCasmurro.txt	9475	7825	5115	2393	1	1.850	8.490	False
memoriasPostumas.txt	10976	8293	6307	2715	1	1.739	8.711	False
quincasBorba.txt	11335	9486	6195	2829	1	1.828	8.740	False
reliquia.txt	14655	10646	7991	3586	1	1.833	9.083	False

k=3

Livro	Vertices	Arestas	Componentes	Maior Comp.	Menor Comp.	Média Comp.	Dist. Média	Denso
cidade-e-as-serras.txt	13311	8915	7448	3151	1	1.786	9.003	False
domCasmurro.txt	9370	6917	5116	2289	1	1.830	8.692	False
memoriasPostumas.txt	10873	7434	6310	2602	1	1.722	8.954	False
quincasBorba.txt	11232	8505	6200	2721	1	1.810	8.947	False
reliquia.txt	14567	9970	7995	3492	1	1.821	9.212	False

k=5

Livro	Vertices	Arestas	Componentes	Maior Comp.	Menor Comp.	Média Comp.	Dist. Média	Denso
cidade-e-as-serras.txt	12527	6582	7488	2195	1	1.672	10.942	False
domCasmurro.txt	8626	4553	5163	1237	1	1.669	10.958	False
memoriasPostumas.txt	10061	4912	6377	1463	1	1.576	11.977	False
quincasBorba.txt	10408	5652	6262	1433	1	1.660	11.708	False
reliquia.txt	13738	7452	8048	2420	1	1.706	11.399	False

3.1.2 Inglês

k=1

Livro	Vertices	Arestas	Componentes	Maior Comp.	Menor Comp.	Média Comp.	Dist. Média	Denso
adventuresHuckleberry.txt	7301	8218	3647	2647	1	1.999	7.237	False
adventuresTomSawyer.txt	8128	6730	4695	2305	1	1.729	7.297	False
frankstein.txt	7173	4512	4479	1460	1	1.599	7.409	False
ladySusan.txt	2978	1435	2058	487	1	1.442	7.560	False
lastMan.txt	13902	10078	8397	3196	1	1.654	8.804	False
prideAndPrejudice.txt	6799	4023	4485	1177	1	1.514	7.268	False

k=3

Livro	Vertices	Arestas	Componentes	Maior Comp.	Menor Comp.	Média Comp.	Dist. Média	Denso
adventuresHuckleberry.txt	7203	7294	3651	2542	1	1.970	7.433	False
adventuresTomSawyer.txt	8073	6327	4699	2238	1	1.716	7.392	False
frankstein.txt	7107	4118	4483	1388	1	1.583	7.674	False
ladySusan.txt	2932	1187	2066	431	1	1.414	8.779	False
lastMan.txt	13835	9536	8401	3125	1	1.646	8.960	False
prideAndPrejudice.txt	6698	3253	4490	1069	1	1.490	7.560	False

k=5

Livro	Vertices	Arestas	Componentes	Maior Comp.	Menor Comp.	Média Comp.	Dist. Média	Denso
adventuresHuckleberry.txt	5963	2630	3812	627	1	1.562	16.407	False
adventuresTomSawyer.txt	6978	2600	4848	565	1	1.437	12.323	False
frankstein.txt	6315	2036	4578	150	1	1.377	7.342	False
ladySusan.txt	2531	511	2069	19	1	1.218	1.963	False
lastMan.txt	12606	5180	8504	1586	1	1.481	14.729	False
prideAndPrejudice.txt	6041	1680	4562	99	1	1.322	7.099	False

3.2 Textos Filosóficos

3.2.1 Alemão

k=1

Livro	Vertices	Arestas	Componentes	Maior Comp.	Menor Comp.	Média Comp.	Dist. Média	Denso
Aphorismen.txt	11099	8924	6577	2270	1	1.686	8.291	False
thePrince_Al.txt	6240	3729	3717	1028	1	1.676	9.765	False

k=3

Livro	Vertices	Arestas	Componentes	Maior Comp.	Menor Comp.	Média Comp.	Dist. Média	Denso
Aphorismen.txt	10910	6699	6595	2024	1	1.653	8.830	False
thePrince_Al.txt	6182	3431	3717	957	1	1.660	10.056	False

k=5

Livro	Vertices	Arestas	Componentes	Maior Comp.	Menor Comp.	Média Comp.	Dist. Média	Denso
Aphorismen.txt	10032	4734	6572	1090	1	1.525	10.259	False
thePrince_Al.txt	5820	2814	3708	617	1	1.567	10.598	False

3.2.2 Inglês

k=1

Livro	Vertices	Arestas	Componentes	Maior Comp.	Menor Comp.	Média Comp.	Dist. Média	Denso
critiquePureReason.txt	7995	4593	5461	952	1	1.462	7.227	False
discourseMethod.txt	2891	1151	2040	372	1	1.412	7.520	False
manifestoCommunist.txt	2257	865	1653	286	1	1.359	7.617	False
thePrince.txt	4774	2788	3107	862	1	1.533	7.114	False

k=3

Livro	Vertices	Arestas	Componentes	Maior Comp.	Menor Comp.	Média Comp.	Dist. Média	Denso
critiquePureReason.txt	7859	3285	5467	805	1	1.436	8.171	False
discourseMethod.txt	2861	1042	2047	327	1	1.393	9.040	False
manifestoCommunist.txt	2218	687	1664	198	1	1.327	8.825	False
thePrince.txt	4708	2357	3113	784	1	1.509	7.813	False

k=5

Livro	Vertices	Arestas	Componentes	Maior Comp.	Menor Comp.	Média Comp.	Dist. Média	Denso
critiquePureReason.txt	7272	2099	5513	20	1	1.317	1.951	False
discourseMethod.txt	2500	508	2037	15	1	1.222	1.679	False
manifestoCommunist.txt	1917	305	1636	7	1	1.166	1.375	False
thePrince.txt	4066	1066	3128	37	1	1.297	3.496	False

3.3 Manual

3.3.1 Inglês

k=1

Livro	Vertices	Arestas	Componentes	Maior Comp.	Menor Comp.	Média Comp.	Dist. Média	Denso
cat.txt	133	129	95	32	1	1.295	2.423	False
gcc.txt	11781	11984	7083	2267	1	1.662	6.543	False
ghc.txt	1702	797	1307	131	1	1.295	4.380	False
grep.txt	962	860	672	132	1	1.417	4.108	False
sort.txt	317	368	234	52	1	1.312	2.235	False
wget.txt	2400	1591	1669	374	1	1.432	6.545	False

k=3

Livro	Vertices	Arestas	Componentes	Maior Comp.	Menor Comp.	Média Comp.	Dist. Média	Denso
cat.txt	106	9	97	3	1	0.990	1.182	False
gcc.txt	11472	7512	7107	1917	1	1.613	8.532	False
ghc.txt	1638	383	1322	16	1	1.231	2.241	False
grep.txt	886	229	682	21	1	1.284	2.611	False
sort.txt	275	42	239	4	1	1.109	1.212	False
wget.txt	2308	759	1683	241	1	1.365	10.290	False

k=5

Livro	Vertices	Arestas	Componentes	Maior Comp.	Menor Comp.	Média Comp.	Dist. Média	Denso
cat.txt	65	0	65	1	1	0.846	-nan	False
gcc.txt	9951	4165	7171	79	1	1.386	3.049	False
ghc.txt	1405	175	1235	4	1	1.130	1.069	False
grep.txt	713	107	610	7	1	1.152	1.297	False
sort.txt	194	12	182	2	1	1.011	1.000	False
wget.txt	1926	354	1606	9	1	1.193	1.312	False

4 Conclusão

Com todos os testes, pude perceber o seguinte:

- O tamanho faz uma diferença muito grande entre os textos, o que era de se esperar. É mais fácil usar palavras diferentes em textos pequenos do que em textos grandes e isso faz com que quanto maior um texto fica, mais difícil é aumentar o seu número de vértices, e quando uma palavra diferente é adicionada, mais possível é que ela se conecte em diversas partes do grafo. Isso fica bem claro ao ver que o número de arestas em textos grandes é bem mais próximo do número de vértices, e em textos pequenos, não chega nem perto.
- O autor faz uma certa diferença nos textos também, mas não é tão grande. Apesar do fato de diferentes autores terem diferentes vocabulários, o número de conexões que acontecem entre as palavras em seus textos acabam ficando bem parecidos, caso as outras condições também sejam parecidas.
- O gênero influencia os textos também, por dois motivos. Textos de diferentes gêneros tendem a ter tamanhos diferentes e, obviamente, o jeito de se expressar neles têm que ser diferente.
- A língua faz uma certa diferença também, apesar de não ser tão grande.
- A última coisa que pude notar é que variar o k influenciou os diversos textos da mesma maneira: a maioria dos dados diminuiu um pouco, com exceção do tamanho da maior componente e do número de arestas, que diminuiu **drasticamente**.