

1 Introdução

Neste trabalho, iremos apresentar 4 redes e extrair as seguintes informações sobre elas: grau, distância, componentes conexas e medidas de centralidade (betweenness, closeness e Katz). Em todas estas informações, serão apresentados a média, o desvio padrão, a mediana, os valores máximo e mínimo e a função de distribuição cumulativa complementar (CCDF) dos graus e da distância. Os gráficos contendo as outras distribuições podem ser encontrados no Github [2].

Para o cálculo da distância foi necessário fazer algumas modificações:

- A biblioteca *graph-tool* define como infinito quando não há um caminho entre 2 vértices; isso faz com que o valor da distância cresça muito quando há muitos vértices que não possuem arestas: este comportamento foi encontrado na Rede 1, que será analisada abaixo
- A biblioteca *graph-tool* define como 0 a distância entre do vértice com ele mesmo

Para os 2 casos acima, os valores foram eliminados na obtenção de informações sobre a distância. Ao fazer isso, sobrarão apenas distâncias de vértices que participam de componentes conexas com mais de 1 vértice.

As redes 1, 2 e 3 foram retirados do site Network Data [1] e a rede 4 foi retirada do site Stanford Network Analysis Project [3]

2 Redes

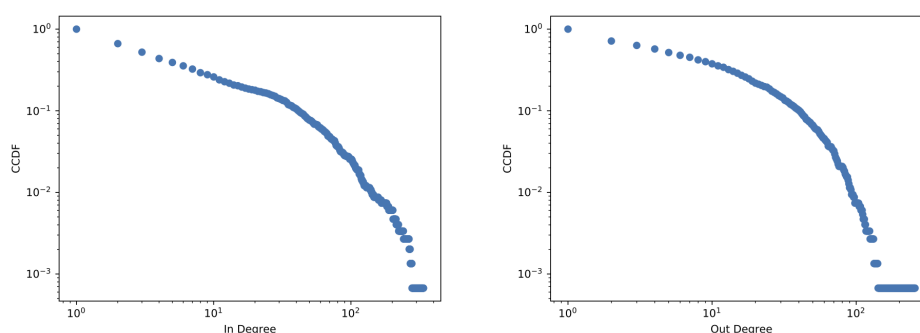
2.1 Rede 1: blogs políticos

Esta rede apresenta blogs políticos dos Estados Unidos. Ela foi coletada em 2005 e mostra uma relação direcionada entre blogs liberais e conservadores. A rede possui 1490 vértices e 19090 arestas e todas estas arestas não possuem peso. A rede foi utilizada no artigo referenciado em [4].

Tabela 1: Métricas para a rede de blogs políticos

		Média	Desvio padrao	Mediana	Máximo	Mínimo
Grau	Entrada	12.81	29.94	2	338	0
	Saída	12.81	20.81	4	256	0
Distância		2.99	1.13	3	9	1
Componentes Conexas		2.16	30.17	1	793	1
Betweenness		0.00071	0.0035	6.32e-7	0.098	0
Closeness		0.23	0.18	0.28	1	0
Katz		0.024	0.0087	0.021	0.11	0.020

Figura 1: Complementary Cumulative Distribution Function para os graus de entrada e de saída



Nesta tabela, podemos observar algumas coisas:

1. valores da distância pequeno: isto ocorre por conta da componente conexa que engloba 793 vértices (mais de 50% dos vértices da rede) e essas distâncias dão maior peso no cálculo da distância da rede.
2. a média dos graus de saída e de entrada são muito próximos
3. o valor mínimo do closeness é igual a 0: isto se dá porque foi necessário assumir valor 0 aos valores de closeness iguais a nan.

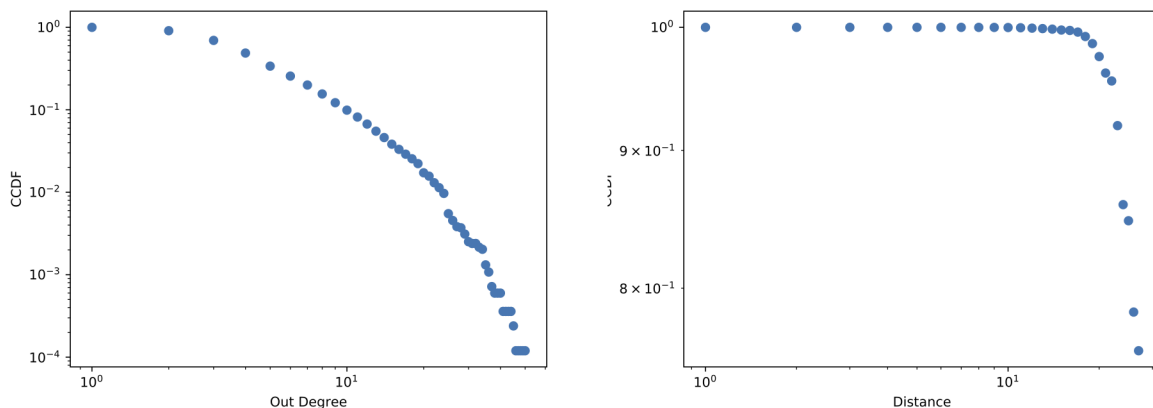
2.2 Rede 2: colaborações em Teoria de Alta Energia

Esta rede apresenta as colaborações entre pesquisadores de Teoria de Alta Energia. Ela é uma rede não-direcionada, com peso nas arestas, mostrando o grau de importância daquela colaboração. A rede possui 8361 vértices e 15751 arestas.

Tabela 2: Métricas para a rede de colaboração entre pesquisadores da área de Alta Energia

	Média	Desvio padrao	Mediana	Máximo	Mínimo
Grau	3.76	4.30	2	50	0
Distância	4.46	1.97	4.23	27.38	0.043
Componentes Conexas	6.27	159.77	1	5835	1
Betweeness	0.00047	0.0016	0.0	0.041	0.0
Closeness	0.46	1.31	0.24	22.99	0.0
Katz	0.010	0.00069	0.0106	0.024	0.0104

Figura 2: Complementary Cumulative Distribution Function para os graus e distância



A distância, diferente das outras redes, é um valor decimal. Isto ocorre porque as arestas não possuem apenas valor 1. Abaixo, na Figura 3, podemos ver a distribuição dos pesos das arestas.

Nas redes a seguir não foram computadas algumas métricas das seguintes métricas :

- A distância foi calculada utilizando uma amostra dos vértices. O cálculo desta métrica é feito entre um vértice *source* e um vértice *target*. Para cada uma das redes seguintes, foram escolhidos aleatoriamente 1000 vértices *source* e para cada *source*, foram escolhidos aleatoriamente 100 vértices *target*.
- Não foi calculada a medida de centralidade de closeness. No *graph-tool*, não há como limitar a quantidade de computações que o closeness faz.
- Para a medida de centralidade de betweeness, foi utilizado um conjunto de 1000 vértices, escolhidos aleatoriamente entre todos os vértices do grafo.
- Para o cálculo da medida de Katz, foi definido o parâmetro *max_iter* igual a 10000, oferecido pelo *graph-tool*, para limitar o número de iterações feitas.

Todas estas mudanças ocorreram porque o computador que foi utilizado nesta tarefa não possui poder computacional suficiente para conseguir lidar com grafos muito grandes.

Figura 3: Distribuição dos pesos das arestas na rede

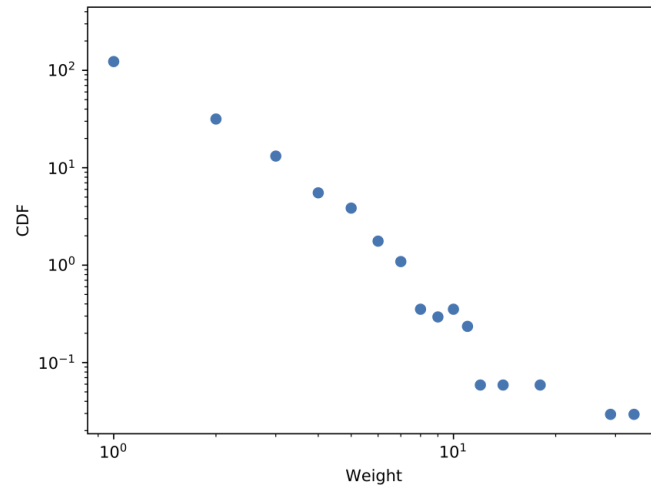


Tabela 3: Métricas para a rede de sistemas autônomos da internet

	Média	Desvio padrao	Mediana	Máximo	Mínimo
Grau	4.21	32.94	2	2390	1
Distância	0.0014	0.89	4	9	1
Componentes Conexas	22963.0	0	22963.0	22963	22963
Betweenness	0.00012	0.0021	0.0	0.13	0.0
Katz	0.0054	0.0036	0.0047	0.20	0.0042

2.3 Rede 3: sistemas autônomos da internet

Esta rede possui uma parte da Internet, constituída de sistemas autônomos e coletada a partir de 2006. Possui 22963 vértices e 48436 arestas não-direcionadas, sem peso.

É interessante notar que, a rede possui apenas uma componente conexa com todos os vértices - visto na tabela. Isto significa que todos os vértices estão conectados entre si. Isso pode ser visto também no grau mínimo da rede, que é 1: todos os vértices possuem, pelo menos, uma aresta ligada a eles.

2.4 Rede 4: páginas do Facebook

Esta rede foi coletada do Facebook, sobre páginas do site que foram verificadas. É uma rede não-direcionada, com vértices representando as páginas e as arestas, os likes mútuos entre 2 páginas. O dataset inicial foi dividido em diferentes categorias de páginas, sendo alguns deles: programas de TV, páginas de políticos, atletas, artistas. Para nossa análise, coletamos a parcela do dataset que corresponde às páginas da categoria artista. Esta rede possui 50515 e 819303 arestas, sendo todas as arestas sem peso.

Tabela 4: Métricas para a rede de páginas do Facebook

	Média	Desvio padrao	Mediana	Máximo	Mínimo
Grau	32.43	63.47	13.0	1469	0
Distância	0.00028	0.77	4.0	9	1
Componentes Conexas	25257.5	25256.5	25257.5	50514	1
Betweenness	5.28e-5	0.00036	2.77e-6	0.032	0.0
Katz	0.0	0.0	0.0	0.0	0.0

Pela tabela, podemos ver que existem componentes conexas muito grandes, assim como existem componentes conexas com apenas 1 vértice. A maior componente conexa possui 49% dos vértices totais do grafo. O grau dos vértices também varia bastante, possuindo um desvio padrão maior do que a média.

Referências

[1] Network data. <http://www-personal.umich.edu/~mejn/netdata/>. (Accessed on 10/19/2019).

Figura 4: Complementary Cumulative Distribution Function para os graus e distância

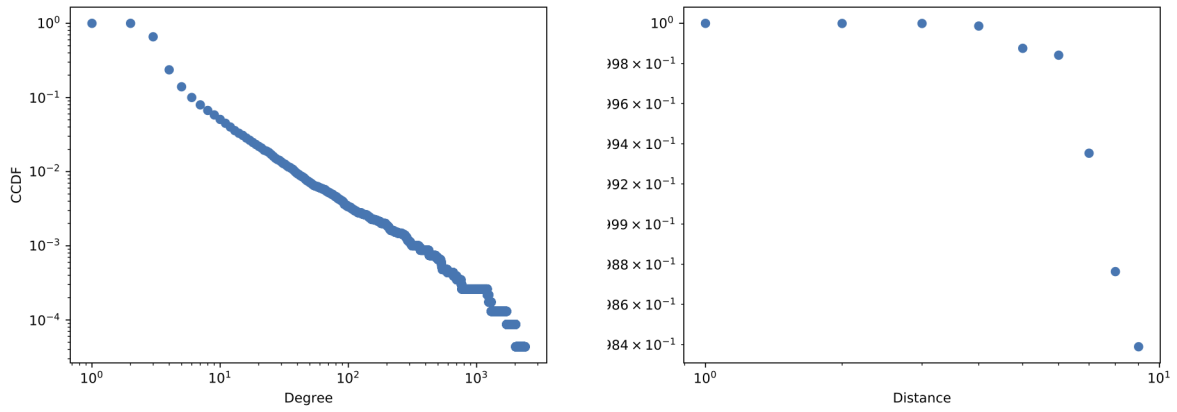
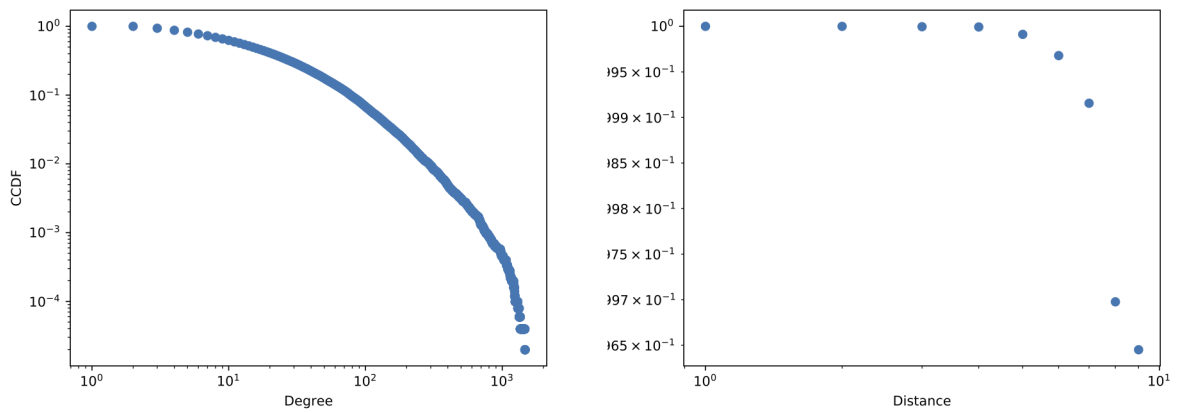


Figura 5: Complementary Cumulative Distribution Function para os graus e distância



- [2] network-science/practice_1 at master · lefreire/network-science. https://github.com/lefreire/network-science/tree/master/practice_1. (Accessed on 10/19/2019).
- [3] Snap: Network datasets: Graph embedding with self clustering: Facebook and deezer. <http://snap.stanford.edu/data/gemsec-Facebook.html>. (Accessed on 10/19/2019).
- [4] ADAMIC, L. A., AND GLANCE, N. The political blogosphere and the 2004 u.s. election: Divided they blog. In *Proceedings of the 3rd International Workshop on Link Discovery* (New York, NY, USA, 2005), LinkKDD '05, ACM, pp. 36–43.