

## Introdução

Neste relatório estão presentes as análises feitas para o trabalho 2 da disciplina de Redes Complexas, com o objetivo de identificar os valores de importâncias de vértices nos grafos. Para o trabalho 1 da disciplina, foram escolhidas redes não-direcionadas. Como alguns algoritmos deste trabalho foram feitos especialmente para tratar de redes direcionadas, foram escolhidas três redes diferentes do trabalho 1, de modo que tais redes sejam direcionadas. De acordo com os repositórios disponíveis, as redes escolhidas foram: “C. Elegans Neural Network” [1], “Political Blogs” [2] e “Gnutella p2p Network - August 8 2002” [3][4]. Para cada rede, foram extraídas informações de quantidade de vértices e arestas, além das medidas de centralidade: grau(entrada e saída), Betweenness [5], Closeness [6], Katz [7] e Pagerank [8]. Para as últimas quatro centralidades, a biblioteca graph-tool [9] foi utilizada para o cálculo, assim como para a plotagem das redes.

## Análise das Redes

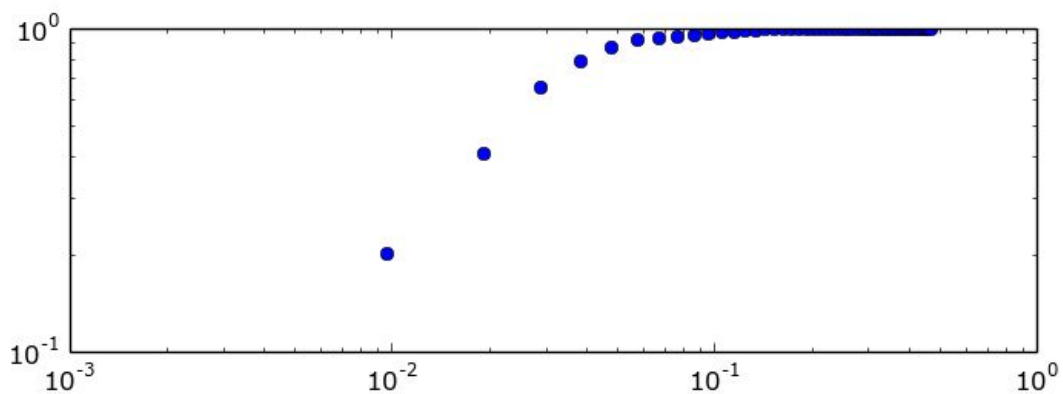
### 1. C. Elegans Neural Network

A “C. Elegans Neural Network” representa a rede neural de uma minhoca da espécie C. Elegans. É uma rede direcionada e com pesos, onde os pesos são valores definidos por Watts [1]. Nesta rede, os neurônios são os vértices e as sinapses entre os neurônios são as arestas. As quantidades de vértices e arestas são:

- Vértices: 297
- Arestas: 2359

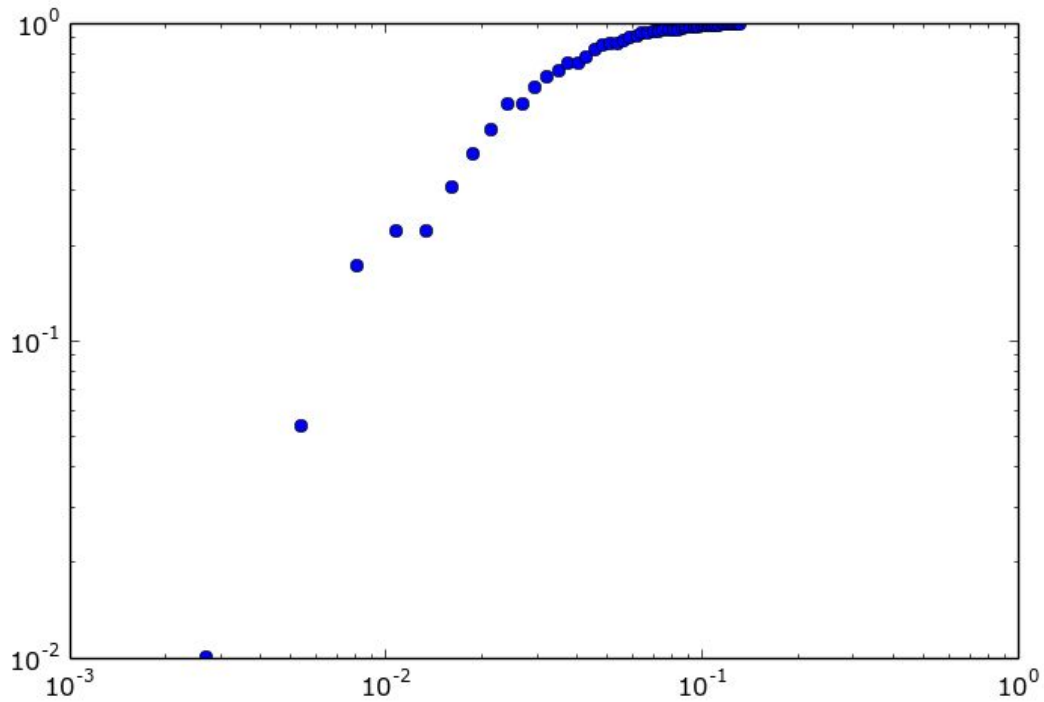
Nas tabelas a seguir estão listados os vértices com maiores e menores valores de centralidade para cada métrica na rede:

C. Elegans - Centralidade de Grau de Entrada										
Maiores										
Vértice	44	12	2	86	84	4	3	190	118	6
Valor	0.4695 94594 6	0.15202 7027	0.13851 35135	0.12162 16216	0.12162 16216	0.11824 32432	0.11148 64865	0.10810 81081	0.10472 97297	0.09121 621622
Menores										
Vértice	260	259	251	243	242	230	143	134	57	54
Valor	0	0	0	0	0	0	0	0	0	0



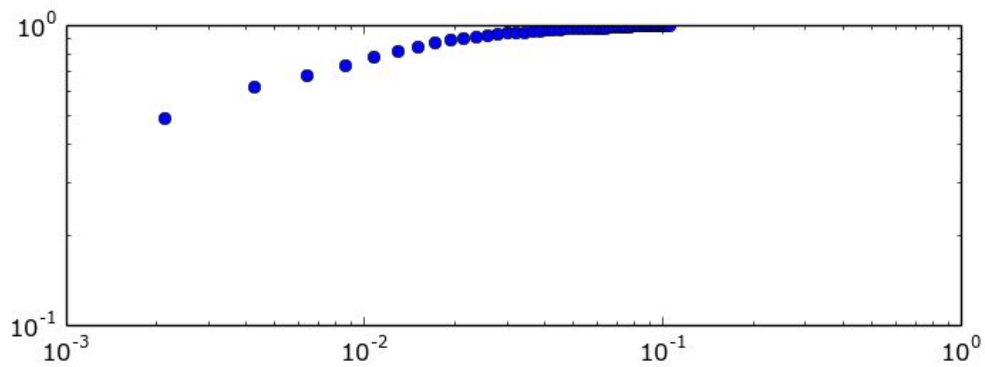
(ECDF centralidade de grau de entrada em escala log/log)

C. Elegans - Centralidade de Grau de Saída										
Maiores										
Vértice	12	2	172	125	137	142	118	117	204	226
Valor	0.1317 56756 8	0.13175 67568	0.11486 48649	0.11486 48649	0.10135 13514	0.09797 297297	0.09121 621622	0.09121 621622	0.08783 783784	0.08445 945946
Menores										
Vértice	78	51	49	37	33	25	22	190	44	39
Valor	0.0033 78378 378	0.00337 837837 8	0.00337 837837 8	0.00337 837837 8	0.00337 837837 8	0.00337 837837 8	0.00337 837837 8	0	0	0



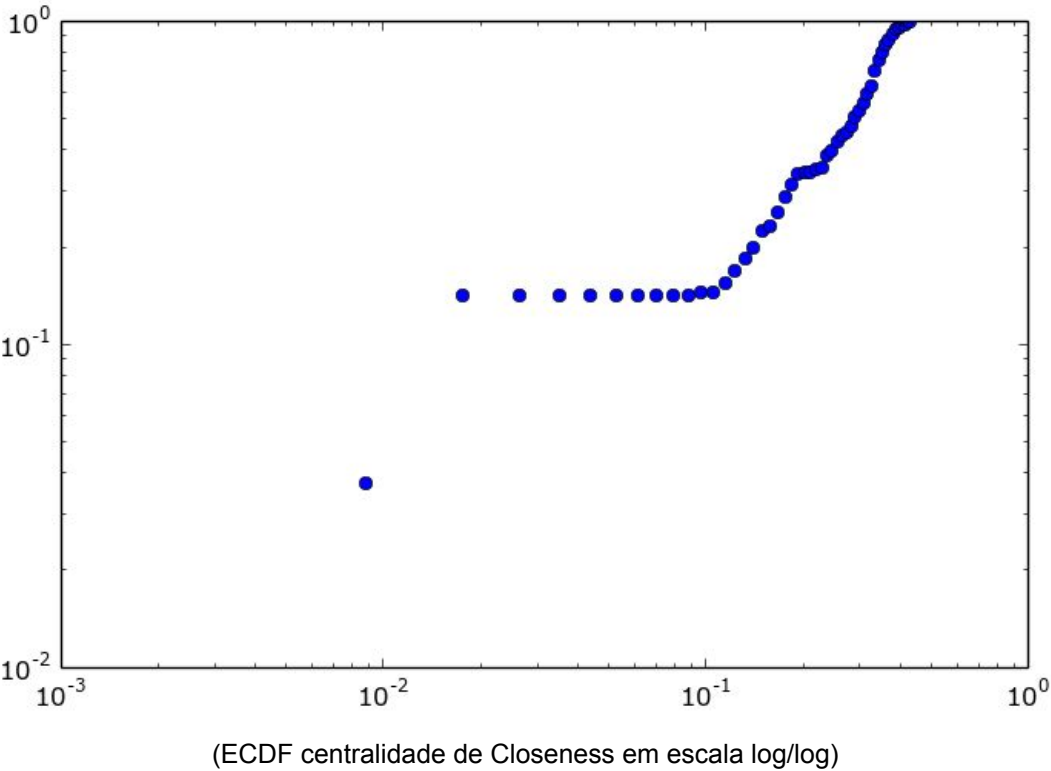
(ECDF centralidade de grau de saída em escala log/log)

C. Elegans - Centralidade de Betweenness										
Maiores										
Vértice	177	142	125	221	172	12	2	165	238	237
Valor	0.1052367245	0.08059589097	0.07871759518	0.07479758024	0.07066858985	0.06575514245	0.05752738146	0.05109010053	0.04628083857	0.04478593556
Menores										
Vértice	231	230	227	190	143	134	57	54	44	39
Valor	0	0	0	0	0	0	0	0	0	0

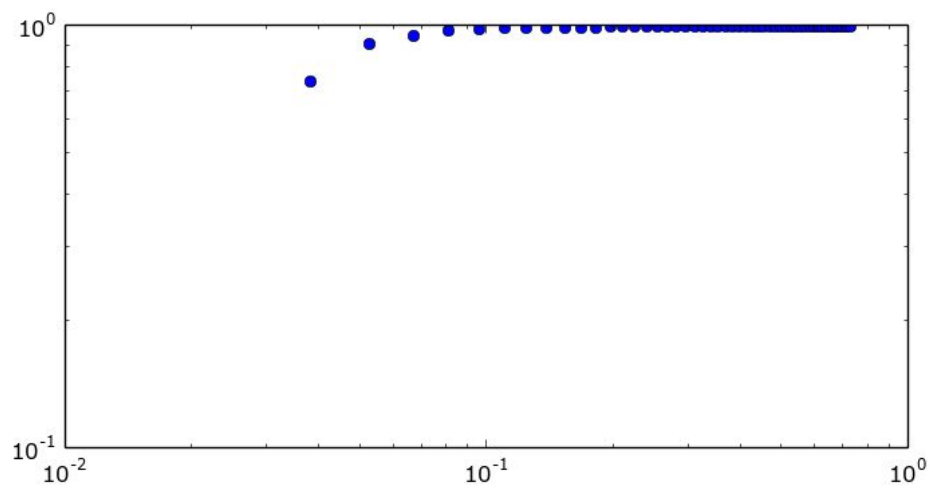


(ECDF centralidade de Betweenness em escala log/log)

C. Elegans - Centralidade de Closeness										
Maiores										
Vértice	125	137	172	142	2	63	12	67	226	111
Valor	0.4307995496	0.4306306306	0.4212274775	0.4204391892	0.4188063063	0.4171734234	0.4063626126	0.4059121622	0.4025900901	0.4009009009
Menores										
Vértice	80	79	78	51	49	37	25	190	44	39
Valor	0.003378378378	0.003378378378	0.003378378378	0.003378378378	0.003378378378	0.003378378378	0.003378378378	0	0	0

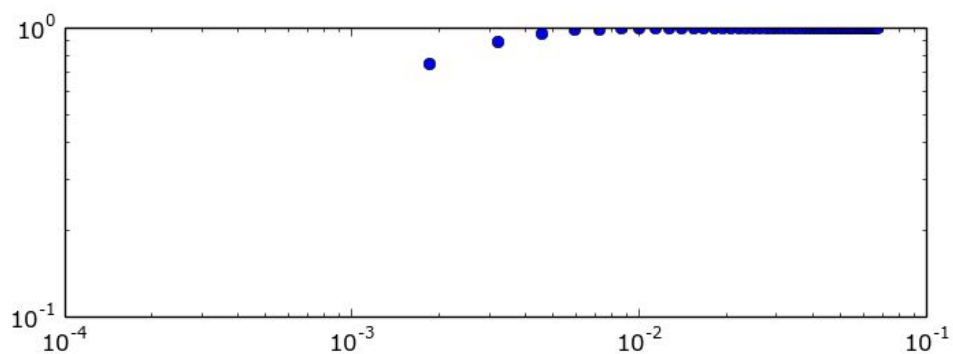


C. Elegans - Centralidade de Katz										
Maiores										
<b>Vértice</b>	44	12	2	172	125	86	137	166	176	162
<b>Valor</b>	0.7273 72766 8	0.18868 88392	0.18637 20192	0.10533 11492	0.10468 34649	0.08586 461501	0.08579 321286	0.08222 747	0.07926 593546	0.07735 85971
Menores										
<b>Vértice</b>	260	259	251	243	242	230	143	134	57	54
<b>Valor</b>	0.0239 99875 32	0.02399 987532	0.02399 987532	0.02399 987532	0.02399 987532	0.02399 987532	0.02399 987532	0.02399 987532	0.02399 987532	0.02399 987532



(ECDF centralidade de Katz em escala log/log)

C. Elegans - Centralidade de PageRank										
Maiores										
<b>Vértice</b>	44	190	6	13	197	12	23	189	200	2
<b>Valor</b>	0.0668 77220 13	0.01442 619271	0.00744 833675 8	0.00665 271327 4	0.00580 883747 2	0.00579 676026	0.00577 683484 6	0.00514 615573	0.00491 901680 9	0.00491 115577 3
Menores										
<b>Vértice</b>	260	259	251	243	242	230	143	134	57	54
<b>Valor</b>	0.0005 05050 5051	0.00050 505050 51	0.00050 505050 51	0.00050 505050 51	0.00050 505050 51	0.00050 505050 51	0.00050 505050 51	0.00050 505050 51	0.00050 505050 51	0.00050 505050 51



(ECDF centralidade de PageRank em escala log/log)

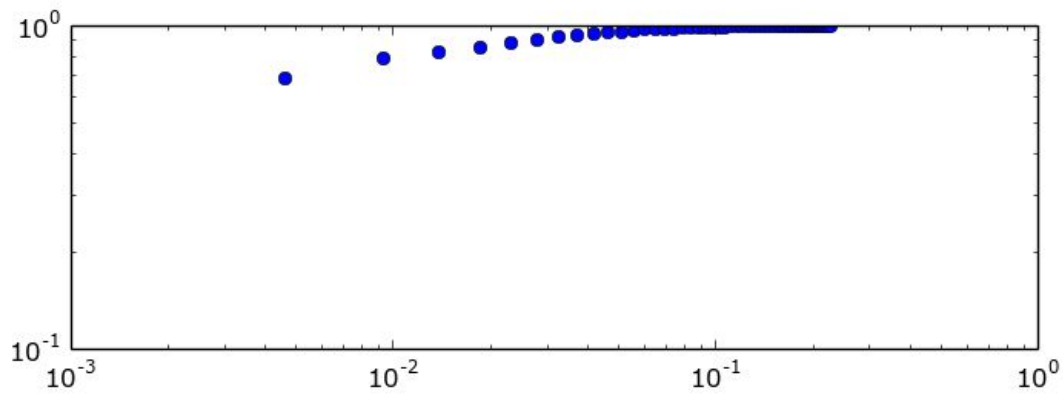
## 2. Political Blogs

A “Political Blogs” é uma rede direcionada que mapeia os hiperlinks entre blogs de assuntos políticos. O snapshot desta rede foi feito em 2005 por Adamic and Glance [2]. As cardinalidades de vértices e arestas na rede são:

- Vértices: 1490
- Arestas: 19090

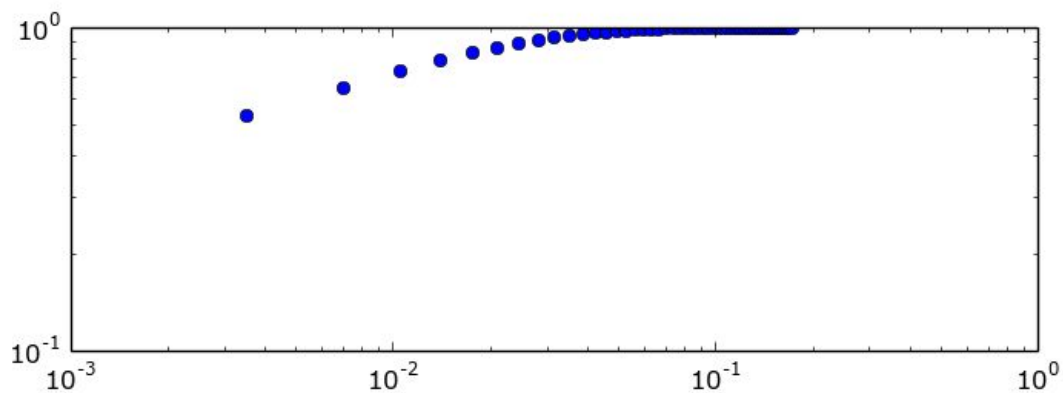
Nas tabelas a seguir estão listados os vértices com maiores e menores valores de centralidade para cada métrica na rede:

Political Blogs - Centralidade de Grau de Entrada										
Maiores										
Vértice	154	1050	640	54	962	1244	854	1152	728	1436
Valor	0.2269 97985 2	0.18603 08932	0.18065 81598	0.17730 02015	0.16118 20013	0.14842 17596	0.14237 74345	0.13498 99261	0.13498 99261	0.12558 76427
Menores										
Vértice	47	42	25	24	19	10	8	5	3	2
Valor	0	0	0	0	0	0	0	0	0	0



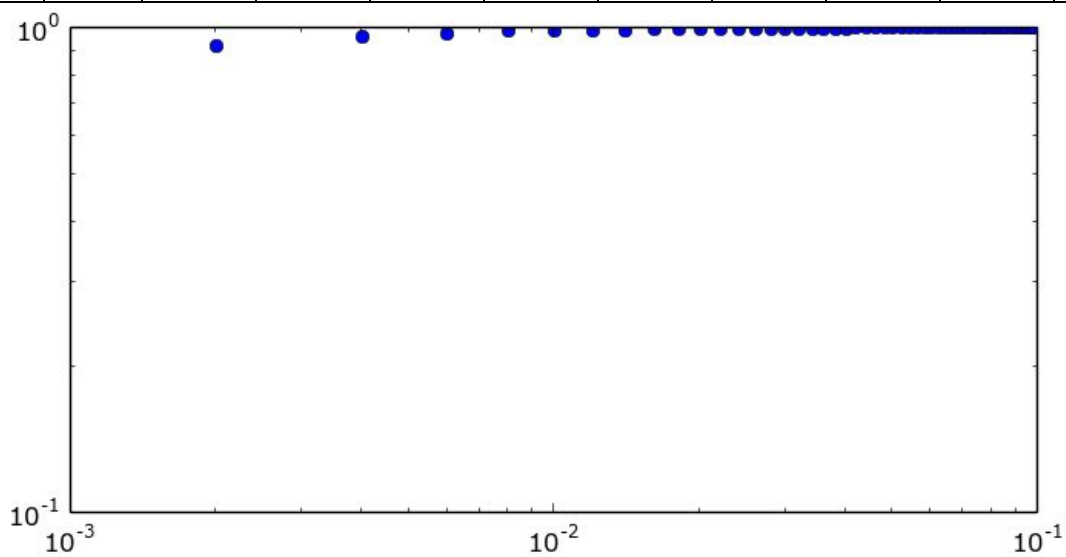
(ECDF centralidade de grau de entrada em escala log/log)

Political Blogs - Centralidade de Grau de Saída										
Maiores										
Vértice	854	453	511	386	879	362	1100	999	523	143
Valor	0.1719 27468 1	0.09402 283412	0.08797 850907	0.08797 850907	0.08260 577569	0.07723 304231	0.07588 985897	0.07387 508395	0.07320 349228	0.07118 871726
Menores										
Vértice	48	47	43	38	36	29	24	6	3	2
Valor	0	0	0	0	0	0	0	0	0	0



(ECDF centralidade de grau de saída em escala log/log)

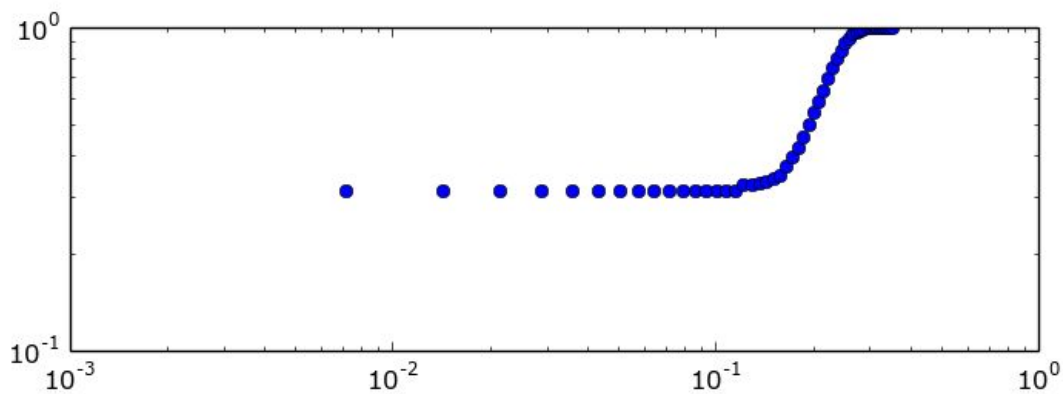
Political Blogs - Centralidade de Betweenness										
Maiores										
<b>Vértice</b>	854	54	1050	154	453	386	1478	1100	1040	728
<b>Valor</b>	0.09860876896	0.04107703165	0.03438191723	0.02474202084	0.02071054637	0.02030867029	0.01833861322	0.01628263839	0.01544597667	0.01474595201
Menores										
<b>Vértice</b>	29	25	24	19	10	8	6	5	3	2
<b>Valor</b>	0	0	0	0	0	0	0	0	0	0



(ECDF centralidade de Betweenness em escala log/log)

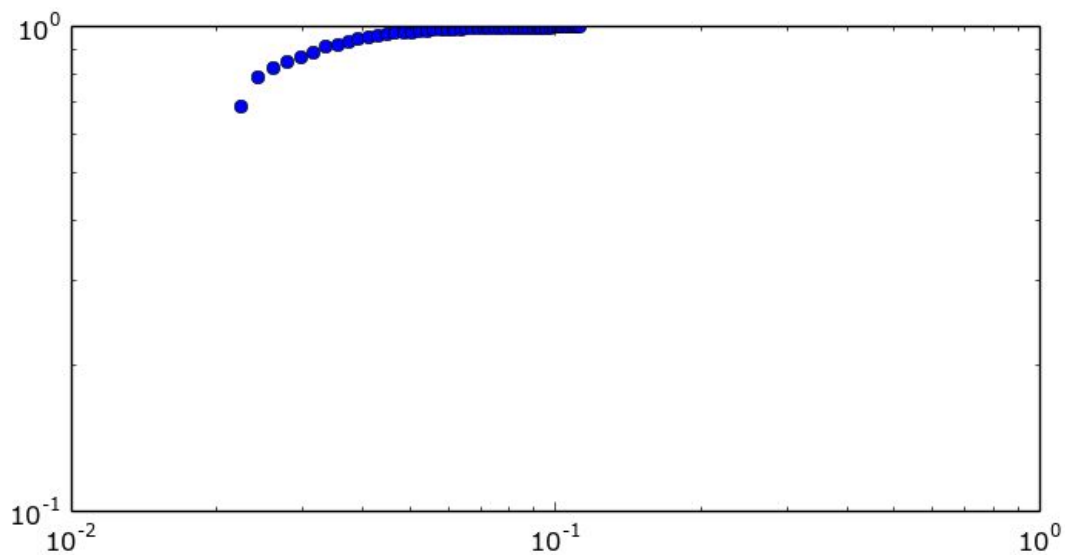
Political Blogs - Centralidade de Closeness										
Maiores										
<b>Vértice</b>	854	879	386	934	511	1100	855	453	523	1426
<b>Valor</b>	0.35161182	0.3136445041	0.3094134766	0.3055630177	0.2999104544	0.2990389843	0.2980859637	0.297526304	0.2959816432	0.2929370942
Menores										
<b>Vértice</b>	48	47	43	38	36	29	24	6	3	2
<b>Valor</b>	0	0	0	0	0	0	0	0	0	0





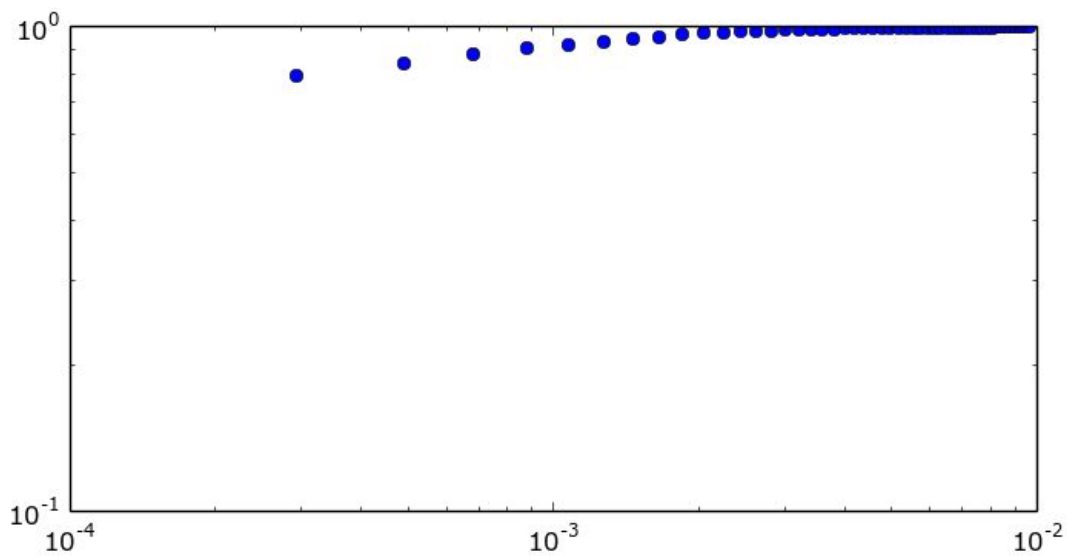
(ECDF centralidade de Closeness em escala log/log)

Political Blogs - Centralidade de Katz										
Maiores										
Vértice	154	1050	54	640	728	1244	962	1152	854	1111
Valor	0.1123 20881 6	0.09897 767205	0.09802 948042	0.09799 177777	0.08174 382702	0.08170 687398	0.07930 067695	0.07472 100978	0.07348 604789	0.07131 058512
Menores										
Vértice	47	42	25	24	19	10	8	5	3	2
Valor	0.0204 88812 15	0.02048 881215	0.02048 881215	0.02048 881215	0.02048 881215	0.02048 881215	0.02048 881215	0.02048 881215	0.02048 881215	0.02048 881215



(ECDF centralidade de Katz em escala log/log)

Political Blogs - Centralidade de PageRank										
Maiores										
Vértice	154	54	1050	854	640	1152	962	728	1244	797
Valor	0.009622162746	0.008166089646	0.006770472708	0.006698937902	0.006667666401	0.005850895949	0.005744324855	0.005655185233	0.00479164694	0.004619209989
Menores										
Vértice	47	42	25	24	19	10	8	5	3	2
Valor	0.0001006711409	0.0001006711409	0.0001006711409	0.0001006711409	0.0001006711409	0.0001006711409	0.0001006711409	0.0001006711409	0.0001006711409	0.0001006711409



(ECDF centralidade de PageRank em escala log/log)

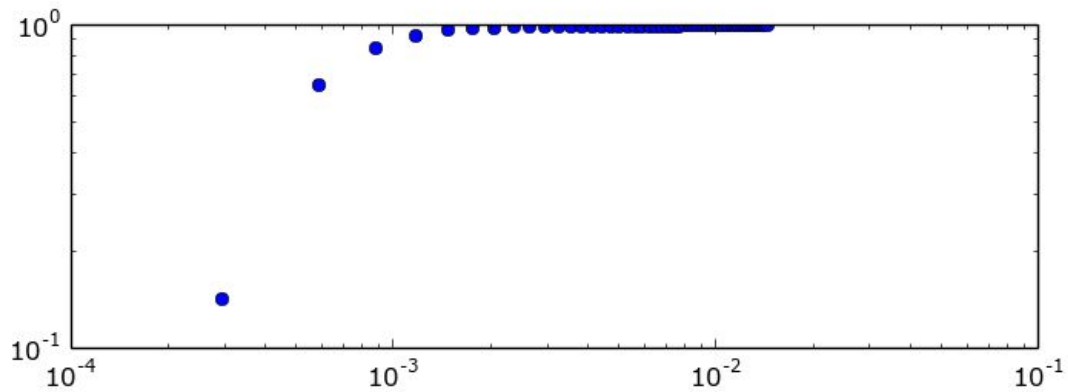
### 3. Gnutella p2p Network - August 8 2002

A “Gnutella p2p Network” é um snapshot da rede peer-to-peer de compartilhamento de arquivos do Gnutella, extraída do dia 8 de Agosto de 2002. É uma rede direcionada onde os hosts são representados por vértices e as conexões entre hosts são representados por arestas. As quantidades de vértices e arestas são:

- Vértices: 6301
- Arestas: 20777

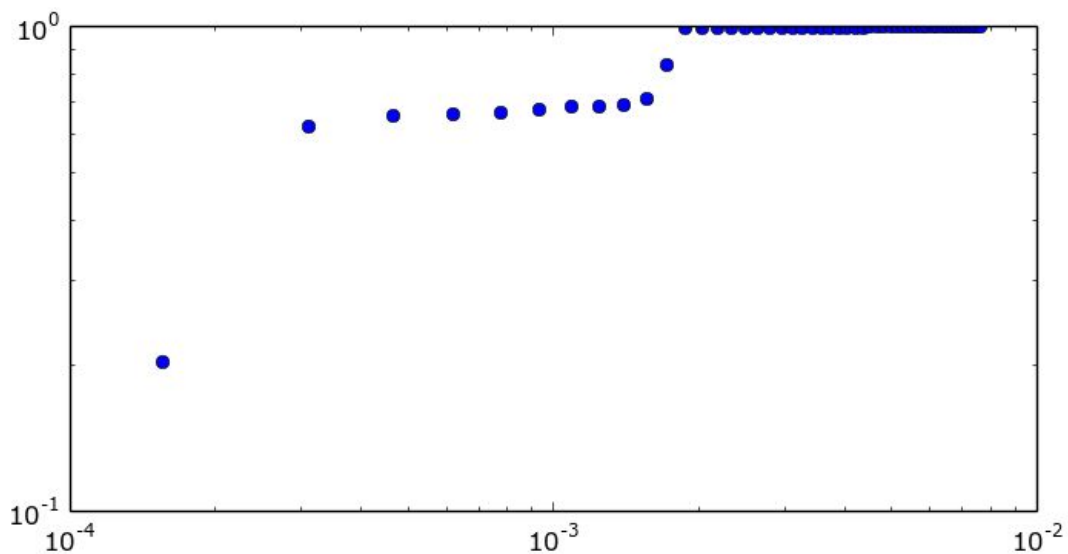
Nas tabelas a seguir estão listados os vértices com maiores e menores valores de centralidade para cada métrica na rede:

Gnutella - Centralidade de Grau de Entrada										
Maiores										
Vértice	266	123	367	127	424	249	427	264	251	145
Valor	0.014444444444444444	0.01380952381	0.01365079365	0.01349206349	0.01317460317	0.01301587302	0.01285714286	0.01285714286	0.01285714286	0.01285714286
Menores										
Vértice	366	310	236	192	109	98	66	44	22	0
Valor	0	0	0	0	0	0	0	0	0	0



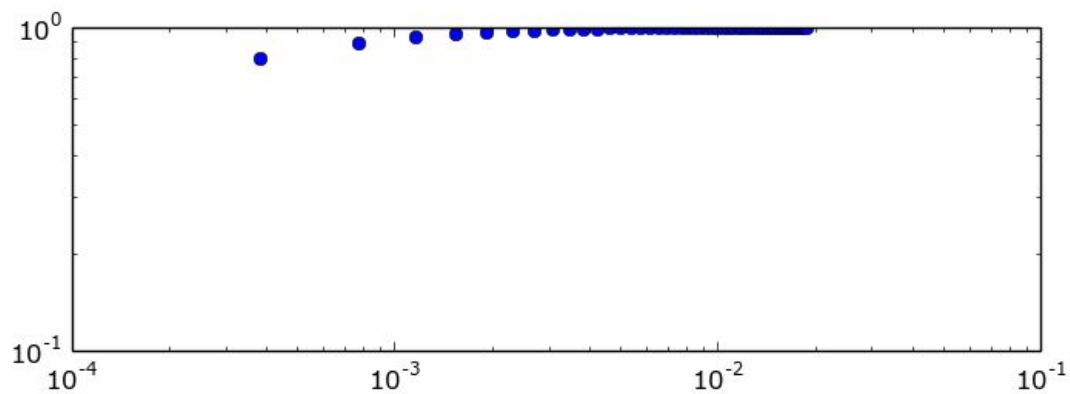
(ECDF centralidade de grau de entrada em escala log/log)

Gnutella - Centralidade de Grau de Saída										
Maiores										
Vértice	5831	1675	5202	6139	1890	6039	1382	2852	989	1994
Valor	0.007619047619	0.00746031746	0.007301587302	0.006507936508	0.005396825397	0.004920634921	0.004603174603	0.004444444444	0.003968253968	0.00380952381
Menores										
Vértice	24	23	19	18	16	12	10	6	2	1
Valor	0	0	0	0	0	0	0	0	0	0



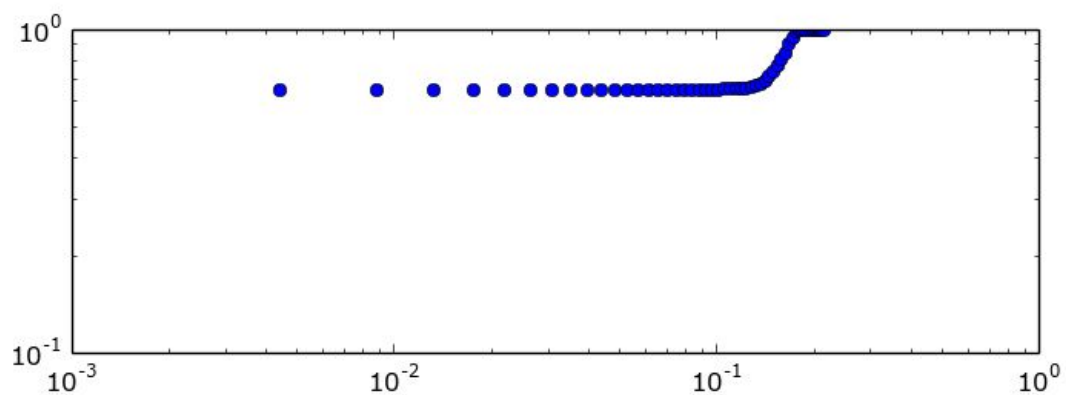
(ECDF centralidade de grau de saída em escala log/log)

Gnutella - Centralidade de Betweenness										
Maiores										
Vértice	1317	3	146	390	175	559	1534	250	700	264
Valor	0.01886870555	0.01723520025	0.01461101845	0.01406802759	0.01400021904	0.01188855744	0.01171035446	0.01117484081	0.01031638159	0.01015777467
Menores										
Vértice	22	19	18	16	12	10	6	2	1	0
Valor	0	0	0	0	0	0	0	0	0	0



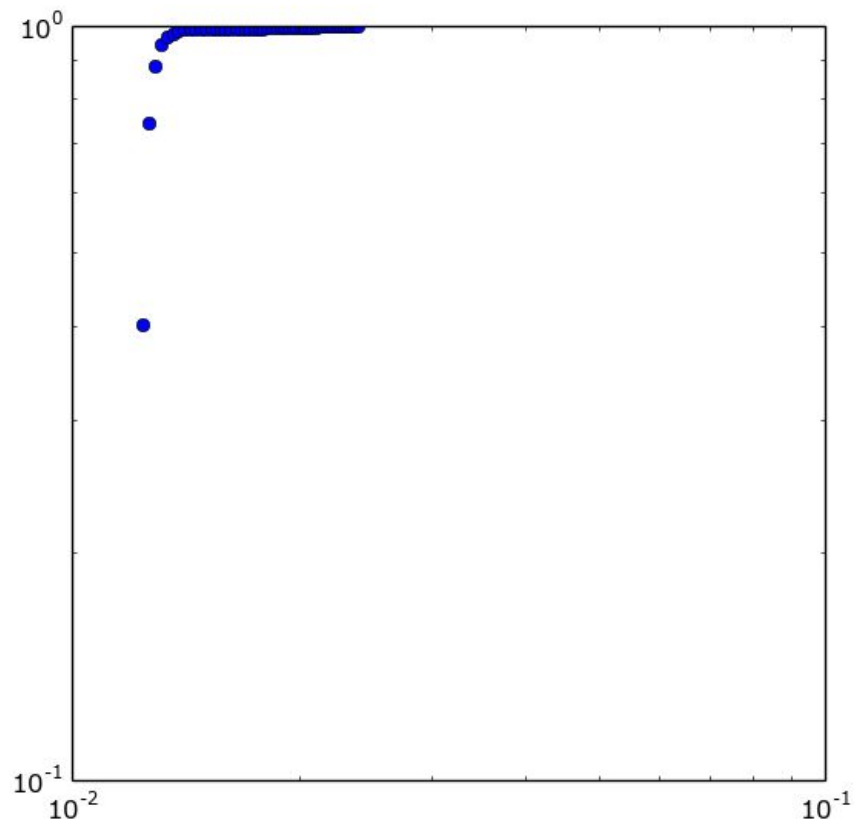
(ECDF centralidade de Betweenness em escala log/log)

Gnutella - Centralidade de Closeness										
Maiores										
Vértice	5831	2614	5202	1382	1675	2852	989	1534	4533	1136
Valor	0.2155 23794 3	0.19708 09926	0.19618 85971	0.19443 50376	0.19268 50165	0.18958 61352	0.18761 38284	0.18732 7815	0.18528 0862	0.18294 86774
Menores										
Vértice	24	23	19	18	16	12	10	6	2	1
Valor	0	0	0	0	0	0	0	0	0	0



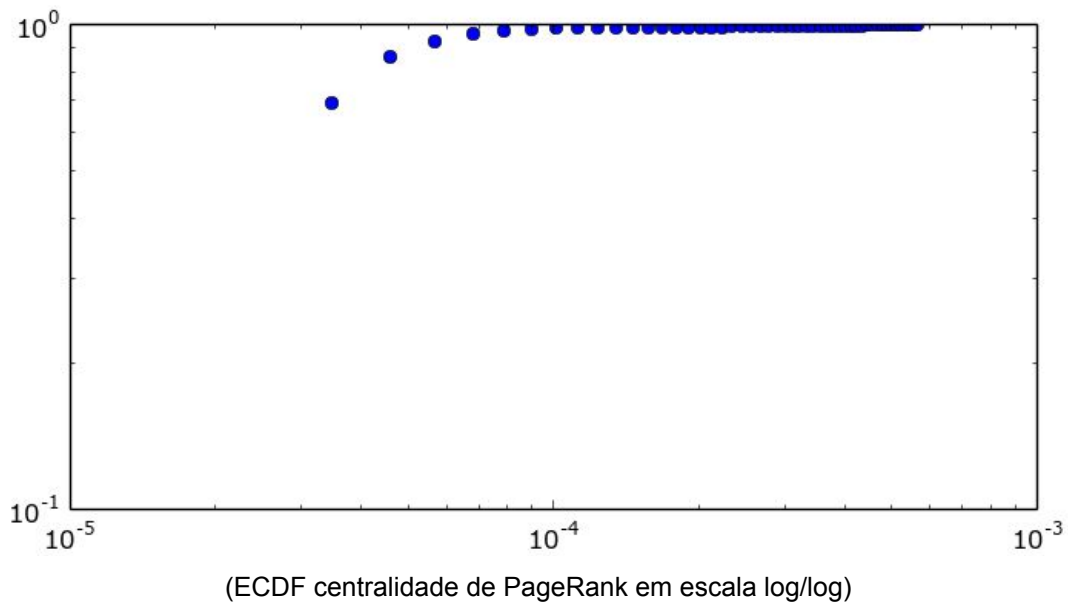
(ECDF centralidade de Closeness em escala log/log)

Gnutella - Centralidade de Katz										
Maiores										
Vértice	266	367	123	127	249	145	264	427	424	251
Valor	0.0239 17510 55	0.02377 504854	0.02351 088011	0.02319 665282	0.02312 253353	0.02290 48228	0.02280 718278	0.02279 07923	0.02271 332176	0.02268 36225
Menores										
Vértice	366	310	236	192	109	98	66	44	22	0
Valor	0.0121 49400 08	0.01214 940008	0.01214 940008	0.01214 940008	0.01214 940008	0.01214 940008	0.01214 940008	0.01214 940008	0.01214 940008	0.01214 940008



(ECDF centralidade de Katz em escala log/log)

Gnutella - Centralidade de PageRank										
Maiores										
<b>Vértice</b>	367	249	145	264	266	123	127	122	1317	5
<b>Valor</b>	0.0005652103537	0.000517060507	0.0004864389858	0.0004731539811	0.0004647773711	0.0004411001754	0.0004403969707	0.0004386915879	0.0004364033165	0.0004334557094
Menores										
<b>Vértice</b>	366	310	236	192	109	98	66	44	22	0
<b>Valor</b>	2.38E-05	2.38E-05	2.38E-05	2.38E-05	2.38E-05	2.38E-05	2.38E-05	2.38E-05	2.38E-05	2.38E-05



## Considerações Finais

A partir das medidas de centralidade é possível obter informações sobre importância dos vértices em suas respectivas redes. Cada medida possui uma abordagem diferente e podem capturar aspectos diferentes nos grafos.

A lista de vértices mais importantes pode variar para cada métrica, porém, alguns vértices podem ser captados como importantes em todas elas. Por exemplo, os vértices 2 e 12 da rede neural da *C. Elegans* está presente nos dez vértices mais importantes das seis métricas utilizadas. Isso pode representar dois neurônios essenciais para as características da *C. Elegans*. No caso do grafo Political Blogs, o vértice sempre presente entre os mais importantes é o 854. Este vértice pode representar um blog de referência no debate político devido sua importância em todas as métricas. E no caso do grafo Gnutella, nenhum vértice aparece em todas as métricas. O que mais aparece em repetidas métricas como importante é o vértice 264 - apenas em 4 das 6 métricas. Esse fato é interessante pois podemos inferir que como é uma rede de compartilhamento peer-to-peer a ideia é ter as importâncias realmente divididas pela rede para aumentar a eficiência do compartilhamento.

A análise interessante a se fazer em relação aos vértices de menor importância é apontar o fato de que as centralidades de Katz e PageRank não zeraram seus valores de importância. Isso acontece devido à uma importância intrínseca que cada vértice possui na rede, definido no algoritmo como o parâmetro *beta*. Para as outras quatro métricas, as redes analisadas possuem na maioria das vezes mais que 10 vértices com valores de centralidade iguais a zero.

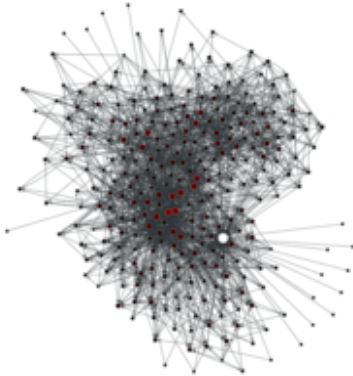
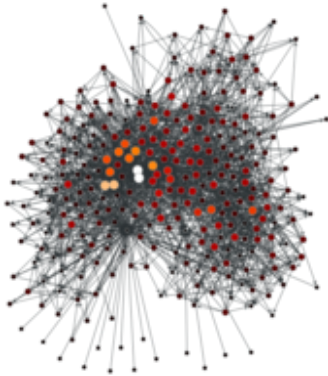
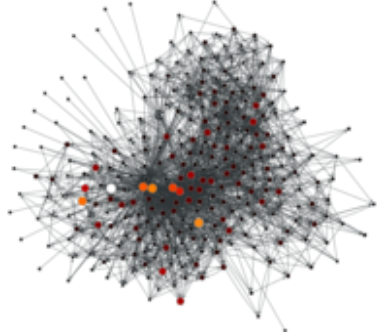
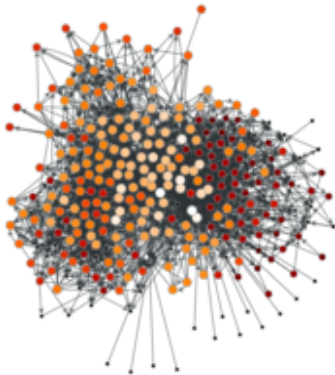
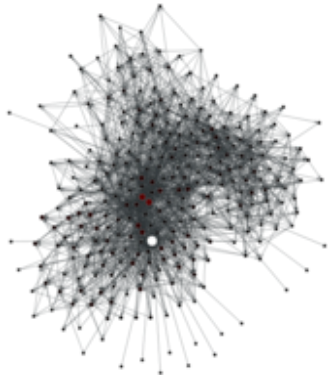
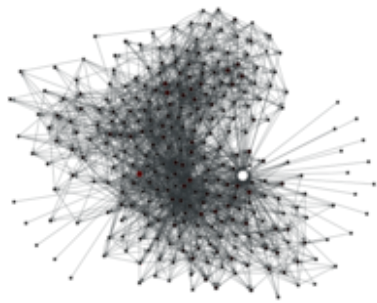
A representação visual dos grafos e suas importâncias podem ser verificadas no Apêndice A, que contém uma plotagem das redes para cada métrica onde a cor do vértice varia com sua importância: quanto mais claro (mais branco) mais importante e quanto mais escuro (mais preto) menos importante. Além disso, a aplicação desenvolvida para a elaboração deste trabalho, assim como este relatório, pode ser acessado através do GitHub [10].

Calcular as medidas de centralidade na rede pode ajudar a entendermos a organização estrutural entre outros aspectos do grafo. Porém, não podemos definir uma “melhor métrica”. É preciso analisar cada caso e avaliar o uso das métricas nas redes dependendo do objetivo desejado.

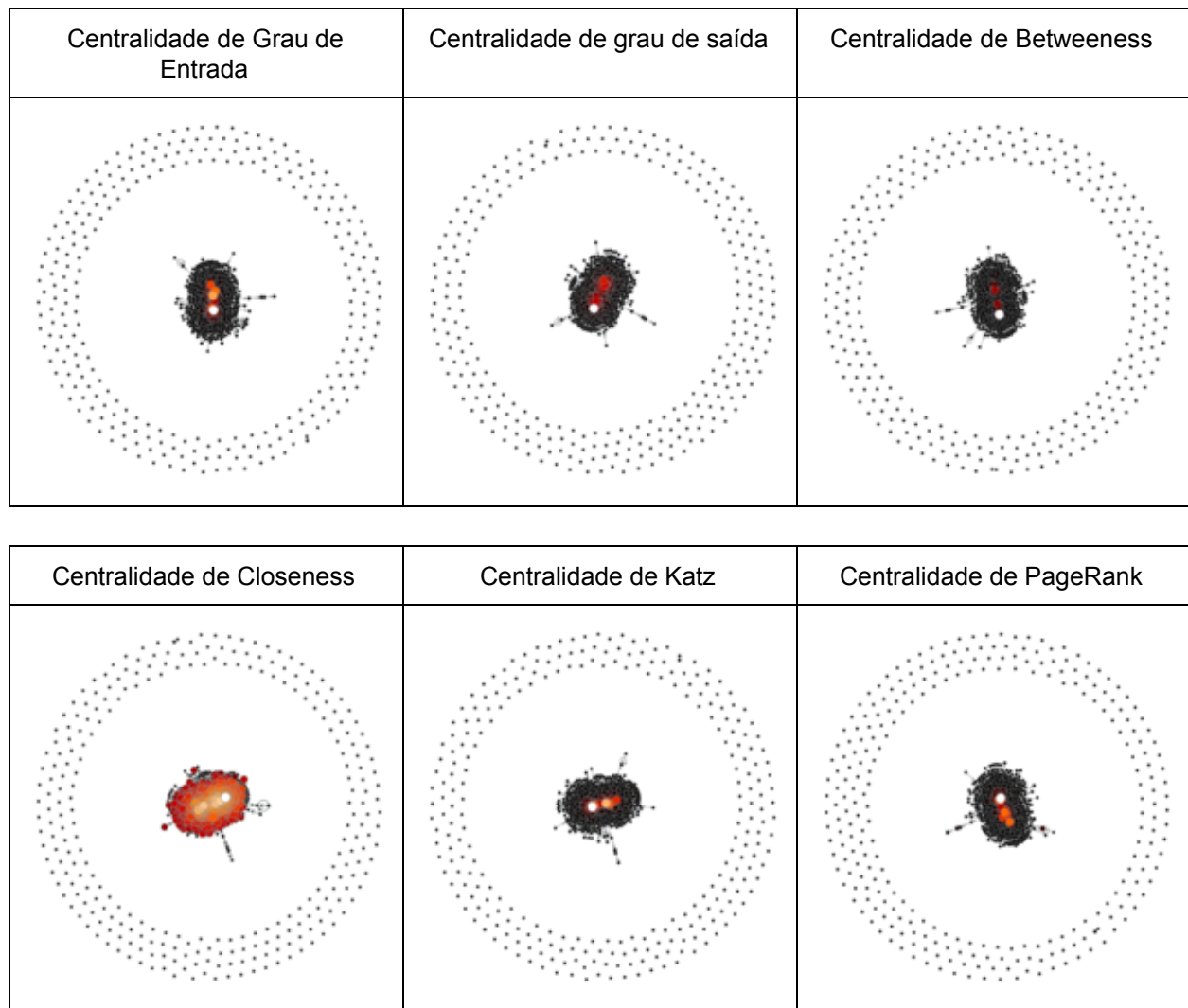


## Apêndice A - Representação Visual das Medidas de Centralidade

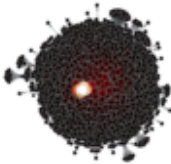
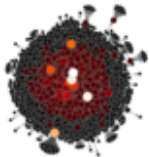
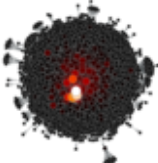
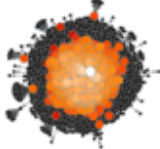
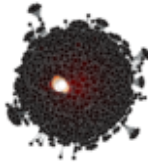
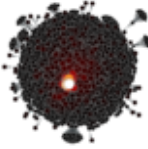
### A.1 - C. Elegans

Centralidade de Grau de Entrada	Centralidade de grau de saída	Centralidade de Betweenness
		
Centralidade de Closeness	Centralidade de Katz	Centralidade de PageRank
		

## A. 2 - Political Blogs



### A. 3 - Gnutella

Centralidade de Grau de Entrada	Centralidade de grau de saída	Centralidade de Betweenness
		
Centralidade de Closeness	Centralidade de Katz	Centralidade de PageRank
		

## Referências

- [1] D. J. Watts and S. H. Strogatz, Nature 393, 440-442 (1998).
- [2] L. A. Adamic and N. Glance, "The political blogosphere and the 2004 US Election", in Proceedings of the WWW-2005 Workshop on the Weblogging Ecosystem (2005).
- [3] J. Leskovec, J. Kleinberg and C. Faloutsos. Graph Evolution: Densification and Shrinking Diameters. ACM Transactions on Knowledge Discovery from Data (ACM TKDD), 1(1), 2007.
- [4] M. Ripeanu and I. Foster and A. Iamnitchi. Mapping the Gnutella Network: Properties of Large-Scale Peer-to-Peer Systems and Implications for System Design. IEEE Internet Computing Journal, 2002.
- [5] U. Brandes, "A faster algorithm for betweenness centrality", Journal of Mathematical Sociology, 2001, DOI: 10.1080/0022250X.2001.9990249
- [6] Opsahl, T., Agneessens, F., Skvoretz, J., "Node centrality in weighted networks: Generalizing degree and shortest paths". Social Networks 32, 245-251, 2010 DOI: 10.1016/j.socnet.2010.03.006
- [7] L. Katz, "A new status index derived from sociometric analysis", Psychometrika 18, Number 1, 39-43, 1953, DOI: 10.1007/BF02289026
- [8] P. Lawrence, B. Sergey, M. Rajeev, W. Terry, "The pagerank citation ranking: Bringing order to the web", Technical report, Stanford University, 1998
- [9] graph-tool - Efficient network analysis, disponível em: <https://graph-tool.skewed.de/>, acessado em: 22/03/2017
- [10] GitHub - Complex Network Analysis, disponível em: <https://github.com/chriiscardozo/complexNetworkCentrality>, acessado em: 11/04/2017