

Structural Analysis of Relevance Propagation Models

Supplementary Material

Values of Centrality and Connectivity Measures

Table SM-1: Clustering coefficients values of DMOZ models

Model	Average CC	Number of nodes with CC=0	Number of nodes with CC=1
T	0	571148	0
S	0.0096	548,312	131
R	0.0484	496,157	2,040
G1	0.0531	372,272	935
G2	0.0754	372,272	4,103
G3	0.4823	0	0
G4	0.4078	0	0
G5	0.1092	363,764	323
G6	0.0897	496,157	41,273
G7	0.0179	489,025	30
G8	0.4261	0	0
G9	0.0548	464,308	8,881
G10	0.0810	464,308	22,280
G11	0.4652	0	0
G12	0.4660	0	0
G13	0.3975	0	0
G14	0.4352	0	0
G15	0.2420	0	0
G16	0.2785	0	0
G17	0.2793	0	0
G18	0.3956	0	0
G19	0.4406	0	0
G20	0.4064	0	0

Table SM-2: Scaling exponents of in-degree, out-degree and degree (for underlying undirected graphs) distributions for DMOZ models

Model	γ^i	γ^o	γ
T	-	-	-
S	3.215	2.03	-
R	2.012	3.873	-
G ₁	2.242	2.265	2.339
G ₂	2.302	2.221	2.339
G ₃	-	0.911	0.958
G ₄	1.987	0.997	1.047
G ₅	1.434	0.936	-
G ₆	2.121	2.121	-
G ₇	2.652	2.220	-
G ₈	2.372	0.980	-
G ₉	2.168	2.109	-
G ₁₀	2.225	2.101	-
G ₁₁	1.990	0.989	-
G ₁₂	2.080	1.028	-
G ₁₃	1.543	0.917	-
G ₁₄	2.134	1.063	1.155
G ₁₅	1.934	0.727	0.929
G ₁₆	1.351	0.594	-
G ₁₇	1.346	0.594	0.825
G ₁₈	2.221	0.924	1.042
G ₁₉	2.181	0.944	1.099
G ₂₀	2.017	1.023	-

Betweenness centrality

The Betweenness Centrality measure helps to identify influential nodes in a network. Such assessment of influence is carried on by counting the paths that a node is involved into. The average BC values as well as the number of nodes with BC=0 and the highest BC value of each RPM are shown in Table SM-4, and Fig. SM-1 shows a chart with the highest BC values for the most salient models.

Table SM-3: Betweenness centrality values of DMOZ models

Model	Average BC	Number of nodes with BC=0	Highest BC
T	22	435,641	529,605
S	44	512,596	963,578
R	25,121	517,718	105,026,375
G ₁	2,708,969	233,362	25,192,567,132
G ₂	2,538,287	233,061	14,572,372,601
G ₃	0	571,148	0
G ₄	1,731,073	264,413	58,235,337,597
G ₅	0	571,148	0
G ₆	852,643	481,642	3,191,455,243
G ₇	26,705	369,333	374,575,132
G ₈	22,456	413,026	368,145,533
G ₉	784,937	393,117	2,749,380,502
G ₁₀	1,183,409	382,971	3,678,132,867
G ₁₁	947,509	504,098	34,582,303,776
G ₁₂	1,185,520	443,436	39,545,756,469
G ₁₃	4,823	486,903	356,494,435
G ₁₄	1,274,742	369,141	43,544,454,702
G ₁₅	1,050,380	285,249	7,910,607,336
G ₁₈	1,274,819	274,422	31,917,718,286
G ₁₉	1,078,864	411,080	29,338,006,924
G ₂₀	1,732,976	258,773	55,680,921,822

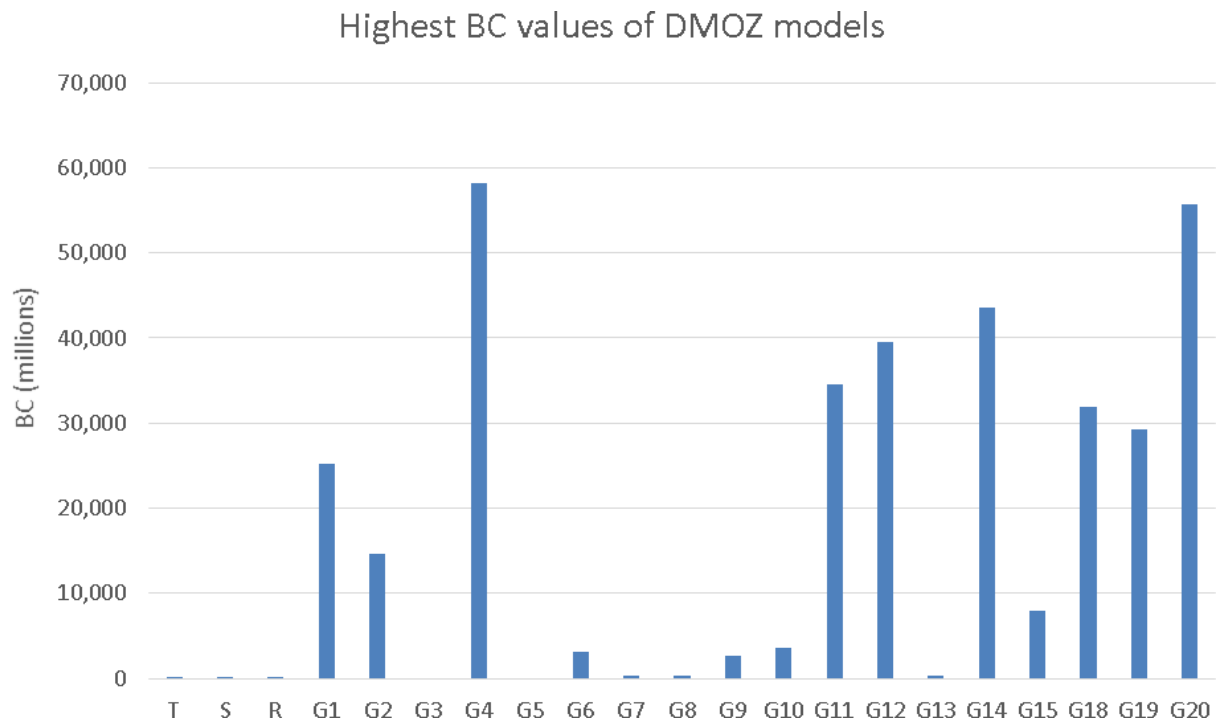


Figure SM-1: Highest BC values in millions for the most important DMOZ RPM's.

The calculation of BC values for some RPM's is computationally very expensive. For instance, it can be observed that model G_4 has a node that participates in more than 58,000 millions of shortest paths. As seen in Fig. SM-2, the analysis of the BC distribution for many RPM's indicates that all of them are heavy-tailed but only the BC values for S are consistent with a power law distribution. Another interesting question about the BC measure in RPM's is the very high magnitude of the difference between the average and highest value. Besides, it is important to highlight the existence of many alternative paths between some pairs of nodes in a network having the shortest length.

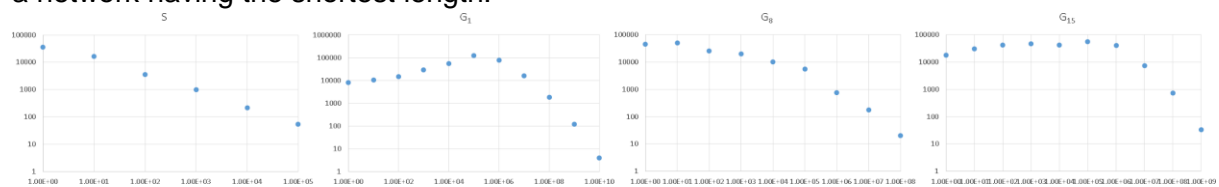


Figure SM-2: BC distribution in log-log scale for some representative DMOZ RPM's.

Alternative Measures of Centrality

This section reports alternative measures computed on the RPM's of DMOZ, with the purpose of assessing their connectivity patterns from another point of view. Table SM-5 shows the mean and max values for the measures of Closeness Centrality, Harmonic Centrality and Lin's Index described in the Background section.

Table SM-4: Average values of Closeness Centrality, Harmonic Centrality and Lin's Index for DMOZ models.

Model	Closeness Centrality		Harmonic Centrality		Lin's Index	
	Avg	Max	Avg	Max	Avg	Max
T	0.042939	1	2.57	3.25	1.74	1.87
S	0.256610	1	1.77	589.92	1.51	520.00
R	0.077657	1	114.28	16179.53	109.03	14441.91
G ₁	0.000142	1	25409.22	43242.16	24430.69	39754.73
G ₂	0.000002	1	29271.97	47432.96	28145.28	43589.30
G ₃	0.152510	1	7.03	14.00	7.03	14.00
G ₄	0.000355	1	38280.34	47609.34	36027.50	42373.23
G ₅	0.284470	1	5.68	747.00	5.68	747.00
G ₆	0.034762	1	7488.22	31691.50	7152.99	29411.45
G ₇	0.010165	1	100.50	2244.37	85.97	1816.23
G ₈	0.023856	1	121.80	2311.87	99.50	1912.68
G ₉	0.151940	1	4429.86	31747.85	4246.06	28761.93
G ₁₀	0.088250	1	11790.05	38631.06	11304.51	35755.26
G ₁₁	0.000631	1	39473.65	51527.90	37285.34	46391.14
G ₁₂	0.000002	1	47103.86	60620.70	44200.87	54468.84

G_{13}	0.026235	1	386.70	5592.87	353.83	4903.17
G_{14}	0.000369	1	49582.84	64048.37	46655.90	57775.42
G_{18}	0.000002	1	53696.74	67034.06	50387.85	60696.77
G_{19}	0.000002	1	61123.58	81575.45	57629.46	71544.55
G_{20}	0.000002	1	41128.94	50922.54	38894.88	45593.31

As denoted in Table SM-5 and explained in (Xamena et al 2017), all the RPM's exhibit a maximum Closeness Centrality value of 1. This is caused by the children of the root node in the main taxonomy, that have a unique co-reachable node, in fact, the root node. The remaining RPM's reflect the presence of, on the one hand, nodes of the taxonomy with the mentioned feature or, on the other hand, nodes that only have one co-reachable node in the S and R components. In the charts of Figure SM-3 and SM-4, very different average values are immediately identified. For the case of Harmonic centrality and Lin's index, many instances have maximum values that are one or two orders of magnitude higher than the mean values. This can be caused by the fact that the corresponding networks have a small number of central nodes, according to those measures. A deeper study over specific central nodes can shed light on the importance of the corresponding DMOZ topics in each RPM.

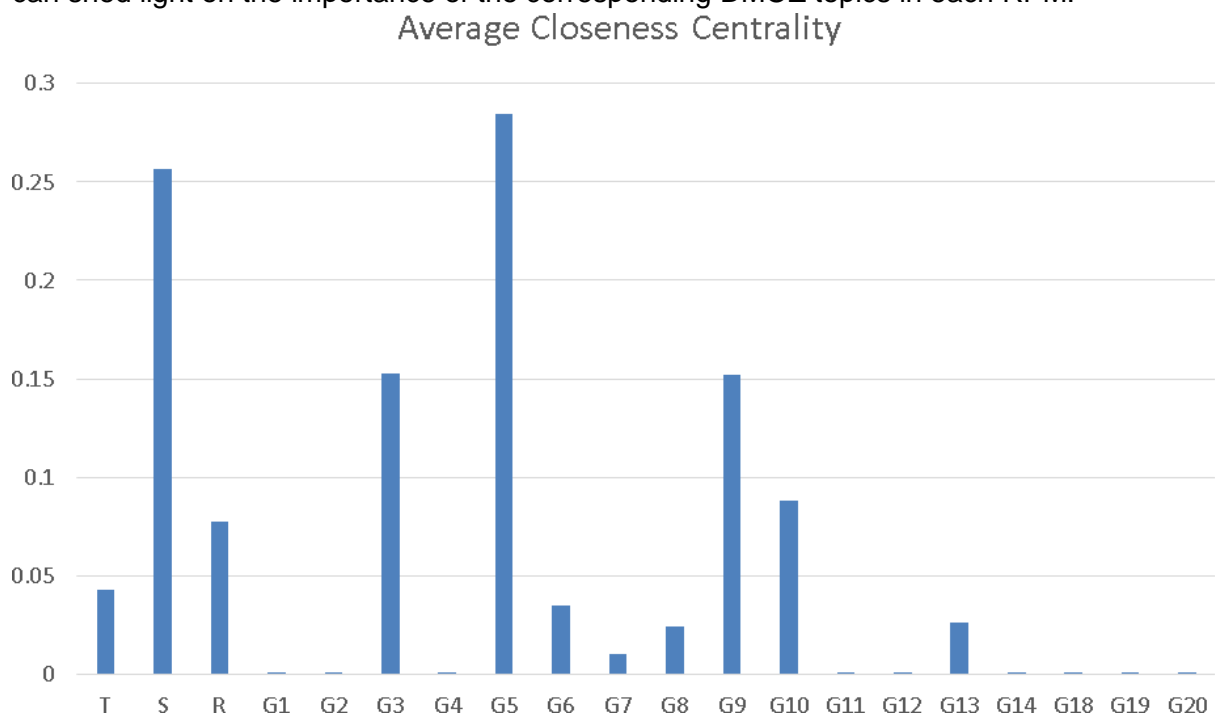


Figure SM-3: Average Closeness Centrality of DMOZ RPM's.

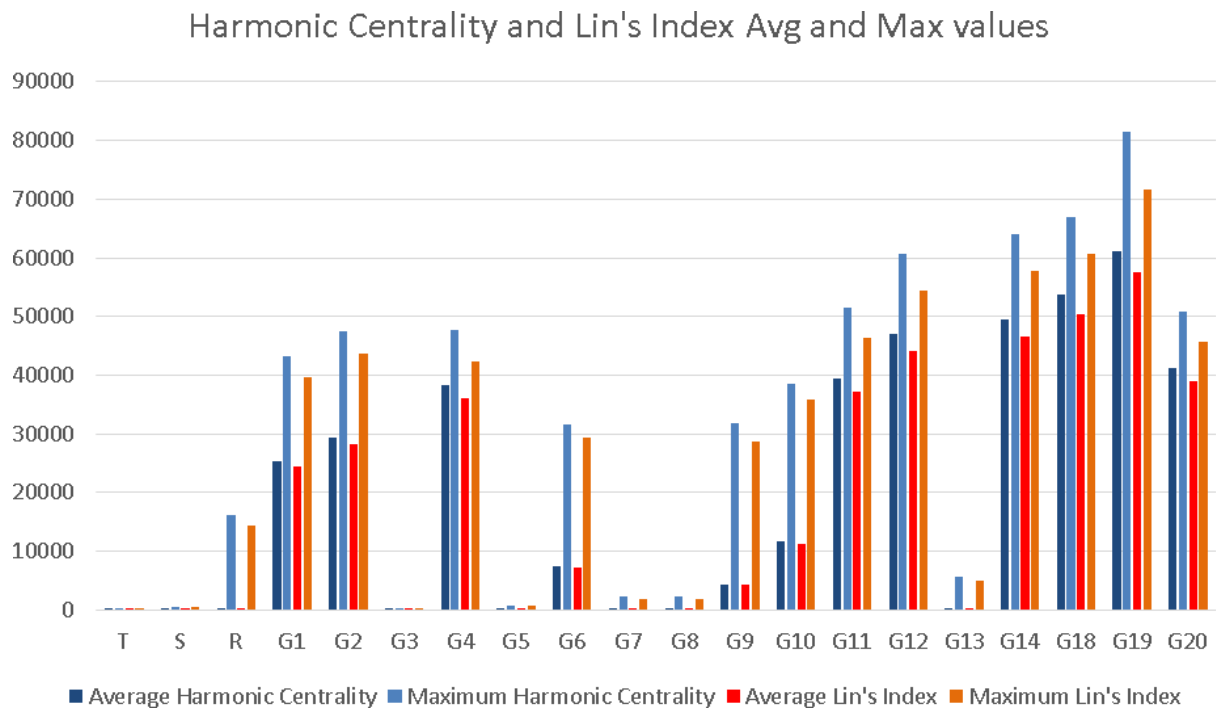


Figure SM-4: Average and maximum values of Harmonic Centrality and Lin's Index of DMOZ RPM's.