Exploring Graph Properties, Specifically Fiedler Values, Using Graph Generators

Katherine Turnlund

June 10, 2022

1 Introduction

Graphs are generated randomly based on an algorithm to mimic a given graph with the purpose that the generation encapsulates connectivity information about the given graph. For this paper, I have selected three algorithms to compare and contrast: the LFR Generator, the Hyperbolic Generator (HG), and the Dorogovtsev Mendes Generator (DMG), to examine their generation patterns, as well as their ability to capture and/or replicate the algebraic connectivity of a given graph (vie the Fiedler value). For generation, I used the NetworkIt package in Python along with its sister package for graph manipulation, NetworkX [4].

1.1 LFR Generator

The LFR generator is based off of the graph generation proposed in [3], and is designed to identify communities of nodes within networks. It works best on networks with: nodes that have relatively uniform degree across the network, communities that are all roughly the same size as each other, and on relatively small networks.

1.2 Hyperbolic Generator

The hyperbolic generator[1] works by taking a node count and sets the probability of an edge between two nodes inversely proportional to the distance between them. This model mimics networks best when the networks have strong clustering and varied degree distribution.

1.3 Dorogovtsev Mendes Generator

The Dorogovtsev Mendes Generator (DMG)[2] is based off of a generalized version of the exact solution of the Barabasi-Albert Model, which models growing networks such as scientific papers or social networks. This algorithm is focused on the distribution of the shortest paths of a network or graph, as well as the distribution of the number of connections with a given site.

2 Experimental Setup

Using the matrices found from the SuiteSparse database, 41 graphs across seven different categories were selected to undergo examination. The graphs were selected to create difference between categories, and within each category, a range of sizes were picked with the goal in mind to see what patterns (if any) would emerge as the size of the graph increases. The graphs were also capped in size due to RAM limitations on the experimenter's equipment causing any graph larger than 5,000 nodes to cause the program to stall and the running instance to be killed with no result gleaned.

3 Results

Overall, most graphs considered in this study were successful in their realization of the three different algorithms proposed (see Tables 1 and 2 for results). This implies that the three generators work well within the elected size range.

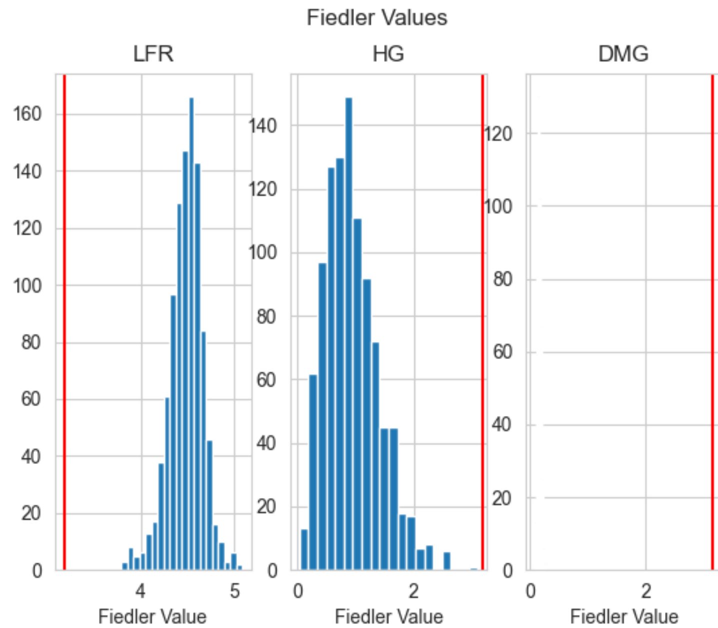


Figure 1: A histogram of the generated Fiedler values for 1,000 generated graphs based on graph 2858, and undirected graph. (Side note: I cannot figure out why the DMG plot is empty. I tried plotting it by itself, examining the input vectors, nothing gave me anything but an empty plot. The Fiedler values exist though, and sit at an average of about 0.15). Note that both the DMG graphs and the hyperbolic graphs seem to undershoot the true Fiedler value, represented by the red line, whereas the LFR graphs seem to overshoot it. This translates the the DMG and HG generated graphs having less connectivity than the expected graph, and LFR having more connectivity than the target value. All three histograms/calculated values give a Gaussian impression of their distribution across 1000 trials.

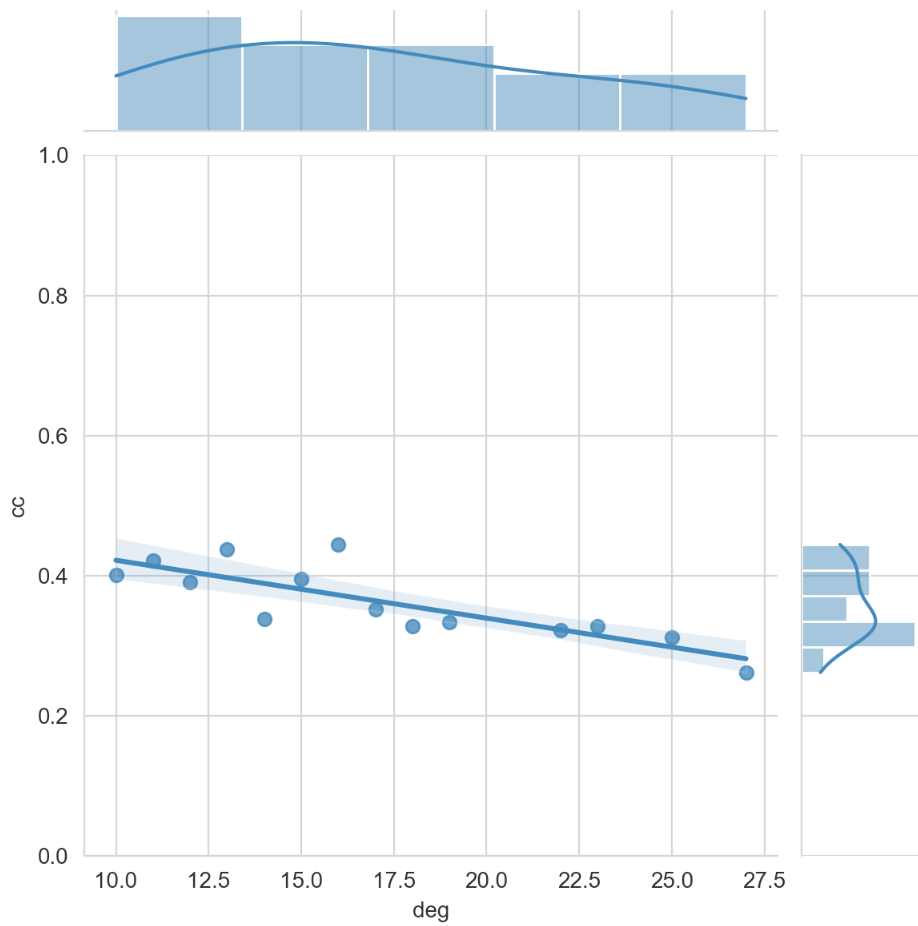


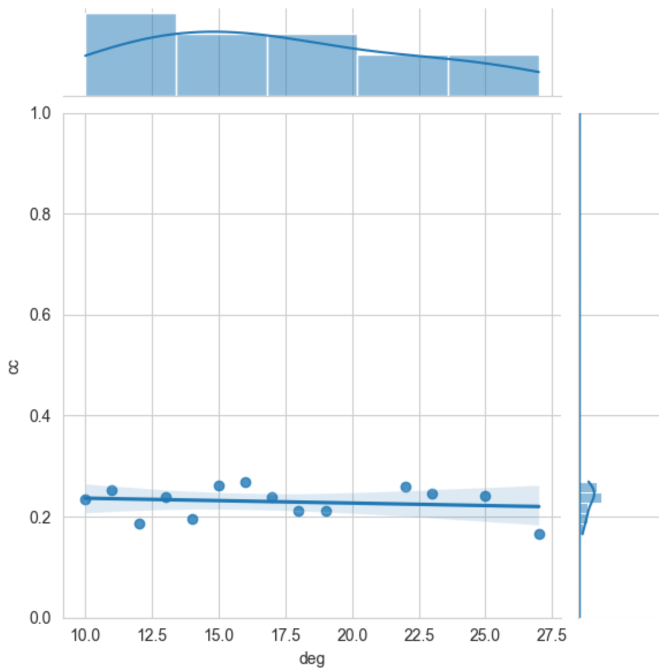
Figure 2: The degree distribution and distribution of clustering coefficient of the original graph of 2858. See Figures 3, 4, and 5 for comparison to generated graphs.

Group	Number	Size	LFR	HG	DMG
leastSquares	11	85	TRUE	TRUE	TRUE
leastSquares	7	85	TRUE	TRUE	TRUE
leastSquares	8	292	TRUE	TRUE	TRUE
leastSquares	9	292	TRUE	TRUE	TRUE
leastSquares	10	292	TRUE	TRUE	TRUE
leastSquares	12	292	TRUE	TRUE	TRUE
electro	296	62	TRUE	TRUE	TRUE
electro	295	398	TRUE	TRUE	TRUE
electro	294	782	TRUE	TRUE	TRUE
electro	293	62	TRUE	TRUE	TRUE
electro	292	398	TRUE	TRUE	TRUE
electro	297	782	TRUE	TRUE	TRUE
circuit	539	2395	FALSE	TRUE	TRUE
circuit	1198	180	FALSE	TRUE	TRUE
circuit	1055	1220	TRUE	TRUE	TRUE
circuit	1106	1220	TRUE	TRUE	TRUE
undirected	2858	106	TRUE	TRUE	TRUE
undirected	2864	2000	TRUE	TRUE	TRUE
undirected	2867	336	TRUE	TRUE	TRUE
undirected	2881	1600	TRUE	TRUE	TRUE
structural	136	87	TRUE	TRUE	TRUE
structural	117	221	TRUE	TRUE	TRUE
structural	119	245	TRUE	TRUE	TRUE
structural	122	310	TRUE	TRUE	TRUE
structural	126	492	TRUE	TRUE	TRUE
structural	134	758	TRUE	TRUE	TRUE
structural	135	869	TRUE	TRUE	TRUE
structural	138	918	TRUE	TRUE	TRUE
structural	221	468	TRUE	TRUE	TRUE
thermal	202	265	TRUE	TRUE	TRUE
thermal	203	406	TRUE	TRUE	TRUE
thermal	204	577	TRUE	TRUE	TRUE

thermal	205	778	TRUE	TRUE	TRUE
thermal	196	1561	TRUE	TRUE	TRUE
thermal	197	1882	TRUE	TRUE	TRUE
thermal	287	1794	TRUE	TRUE	TRUE
economic	80	1794	TRUE	TRUE	TRUE
economic	208	1258	TRUE	TRUE	TRUE
economic	224	2529	TRUE	TRUE	TRUE
economic	275	2529	TRUE	TRUE	TRUE
economic	5	2529	TRUE	TRUE	TRUE

Table 1: Continuation of Table 1. This lists the realization results of the given graph and its category for the LFR, HG, and DMG graph generators. Graphs that worked for none of the above were removed from contention; specifically, four graphs from the undirected catgory, and two from the Economics category. Failing to be realized overall does seem to be correlated with being a significantly larger graph (10,000+ nodes), although for those thaat were realized by some graph generation methods and not others, size does not seem to be a factor, and category seems to matter more. Also, the only generation metehod to fail where others succeeded was the LFR generator. The LFR generator does highly prefer smaller graphs, though the two failures in the "Circuit Design Simulation" category seem to neither be particularly large, especially

Figure 3: The degree distribution of the LFR generated graph. Note that this LFR graph shows a mush tighter clustering distribution than the true graph, but a very similar-looking degree distribution. This makes sense, asa the generator routinely overs=estimates algebraic connectivity for a graph.

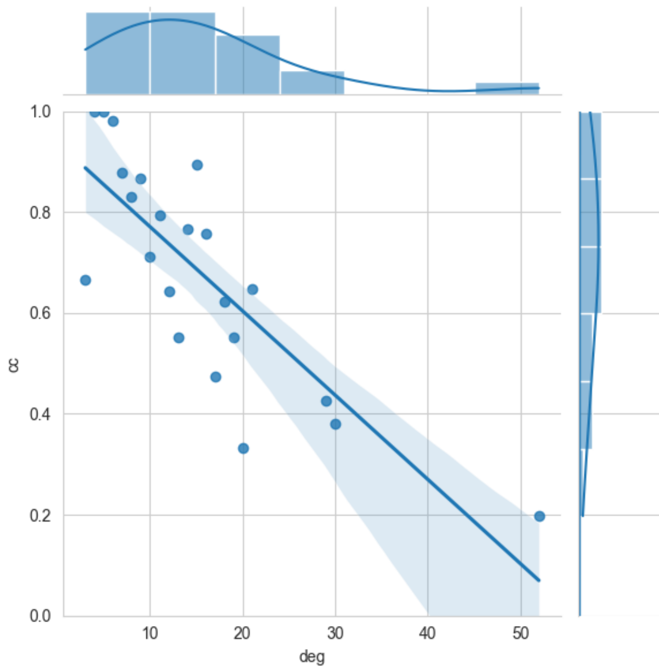


4 Conclusion

Overall unfortunately the data did not show any clear trends along the graph size when examining the Fiedler values like originally expected. Throughout the analysis, we've seen that the LFR and DMG generators seem to perform closer to the true graph than the hyperbolic generator. This makes sense, as the LFR and DMG generators take more information from the true graph than the hyperbolic generator does: hyperbolic only take the number of nodes contained within the graph, LFE and DMG take significantly more than that.

All three seem to be performing decently within our size range for selected graphs, although in future work (and hopefully with equipment that can better handle the required computation) a push into much much larger graphs would be nice to test the boundaries of functionality of the generators.

Figure 4: The degree distribution of the hyperbolic generated graph. This graph has a much wider distribution the clustering coefficient than the true graph, but a narrower degree distribution. This graph generator not lining up with this particular graph isn't hugely surprising, as the hyperbolic generator is based on node count only.



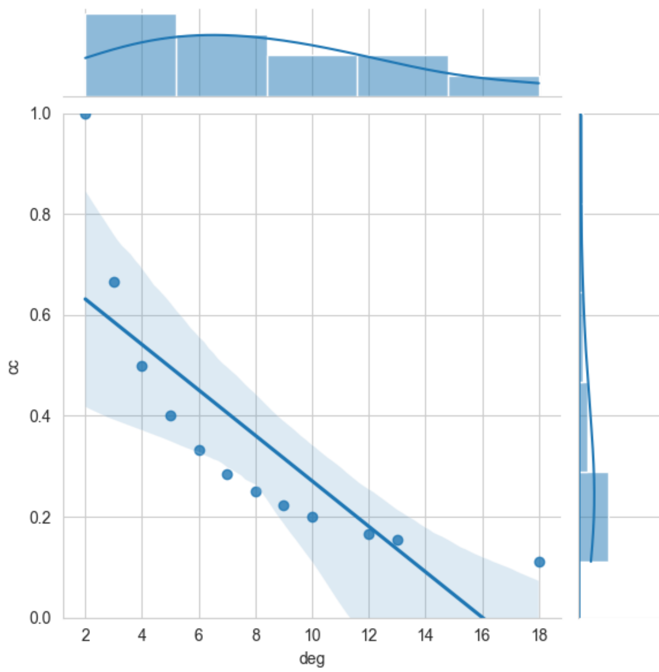
References

[1] Rodrigo Aldecoa, Chiara Orsini, and Dmitri Krioukov. Hyperbolic graph generator. *Computer Physics Communications*, 196:492–496, nov 2015.

[2] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin. Structure of growing networks with preferential linking. *Physical Review Letters*, 85(21):4633–4636, nov 2000.

[3] Andrea Lancichinetti, Santo Fortunato, and Filippo Radicchi. Benchmark graphs for testing community detection algorithms. *Physical Review E*, 78(4), oct 2008.

Figure 5: The degree distribution of the DMG generated graph. Again, similar to the hyperbolic graph, there is a wider range of clustering coefficients, although the degree distribution of the DMG generated graph mimics that of the true graph much closer than the hyperbolic graph. This is interesting, because based on the histogram of Fiedler value distributions the generator is routinely underestimating algebraic connectivity.



[4] Christian Staudt, Aleksejs Sazonovs, and Henning Meyerhenke. Networkit: An interactive tool suite for high-performance network analysis. *CoRR*, abs/1403.3005, 2014.

Group	Number	Size	LFR eig	HG eig	DMG eig
leastSquares	11	85	0.507	1.21E-16	0.14
leastSquares	7	292	0.482	0.858379991	0.144
leastSquares	8	292	0.402	1.74E-32	0.074
leastSquares	9	292	0.401	3.22E-34	0.042
leastSquares	10	292	0.377	2.66E-16	0.055
leastSquares	12	62	0.401	9.16E-24	0.044
electro	296	398	0.697	0.705922776	0.231
electro	295	782	0.525	2.42E-16	0.042
electro	294	62	0.503	3.58E-16	0.024
electro	293	398	0.333	2.84E-16	0.424
electro	292	782	0.359	1.31E-32	0.057
electro	297	2395	0.393	8.06E-21	0.026
circuit	539	180	0	-3.07E-16	0.009
circuit	1198	1220	0	9.43E-17	0.064
circuit	1055	1220	0.045	-8.67E-21	0.019
circuit	1106	106	0.066	-7.64E-19	0.013
undirected	2858	2000	4.26	0.928148577	0.158
undirected	2864	336	0.323	0.000978925	0.007
undirected	2867	1600	1.583	0.269919202	0.055
undirected	2881	87	0.38	0.003270861	0.01
structural	136	221	0.343	0.989791694	0.197
structural	117	245	0.289	9.49E-38	0.066
structural	119	310	0.07	4.17E-32	0.092
structural	122	492	0.49	6.86E-33	0.049
structural	126	758	0.199	2.45E-17	0.036
structural	134	869	0.176	5.48E-17	0.016
structural	135	918	0.254	9.03E-15	0.032
structural	138	468	0.322	6.10E-18	0.015
structural	221	265	1.554	2.47E-16	0.037
thermal	202	406	0.358	0.383469177	0.047
thermal	203	577	0.343	3.27E-15	0.042
thermal	204	778	0.374	8.20E-34	0.028

thermal	205	1561	0.341	1.11E-18	0.022
thermal	196	1882	0.261	-7.05E-17	0.015
thermal	197	1794	0.263	-2.08E-16	0.009
thermal	287	1794	0.054	2.65E-16	0.013
economic	80	1258	0.039	-1.11E-16	0.008
economic	208	2529	0.357	6.58E-17	0.014
economic	224	2529	1.937	5.93715386	0.008
economic	275	2529	1.921	7.128665225	0.009
economic	5	2529	1.696	8.81270949	0.006

Table 2: Continuation of above, listing the Fiedler values found for each matrix. Note that a zero entered for a Fiedler value indicates that the graph failed to be realized when going through generation. Note that based on graph size the Fiedler values do not seem to follow any specific pattern, within the graph groups or overall.