Statistical Evaluation of Spectral Methods for Anomaly Detection in Networks - Supplementary Material

Abstract

The problem of anomaly detection in networks has attracted a lot of attention in recent years, especially with the rise of connected devices and social networks. Anomaly detection spans a wide range of applications, from detecting terrorist cells in counter-terrorism efforts to identifying unexpected mutations during RNA transcription. Fittingly, numerous algorithmic techniques for anomaly detection have been introduced. However, to date, little work has been done to evaluate these algorithms from a statistical perspective. This work is aimed at addressing this gap in the literature, by carrying out statistical evaluation of a suite of popular spectral methods for anomaly detection in networks. Our investigation on statistical properties of these algorithms reveals several important and critical shortcomings, and we make methodological improvements to address such shortcomings. Further, we carry out a performance evaluation of these algorithms using simulated networks, and also extend the methods to count networks.

Keywords: Residual Matrix; Spectral Methods, R-MAT Model, Principal Components

1. Statistical evaluation of the chi-square algorithm

Includes additional figures from section 3

1.1. Comparing the test statistic to the chi-square distribution

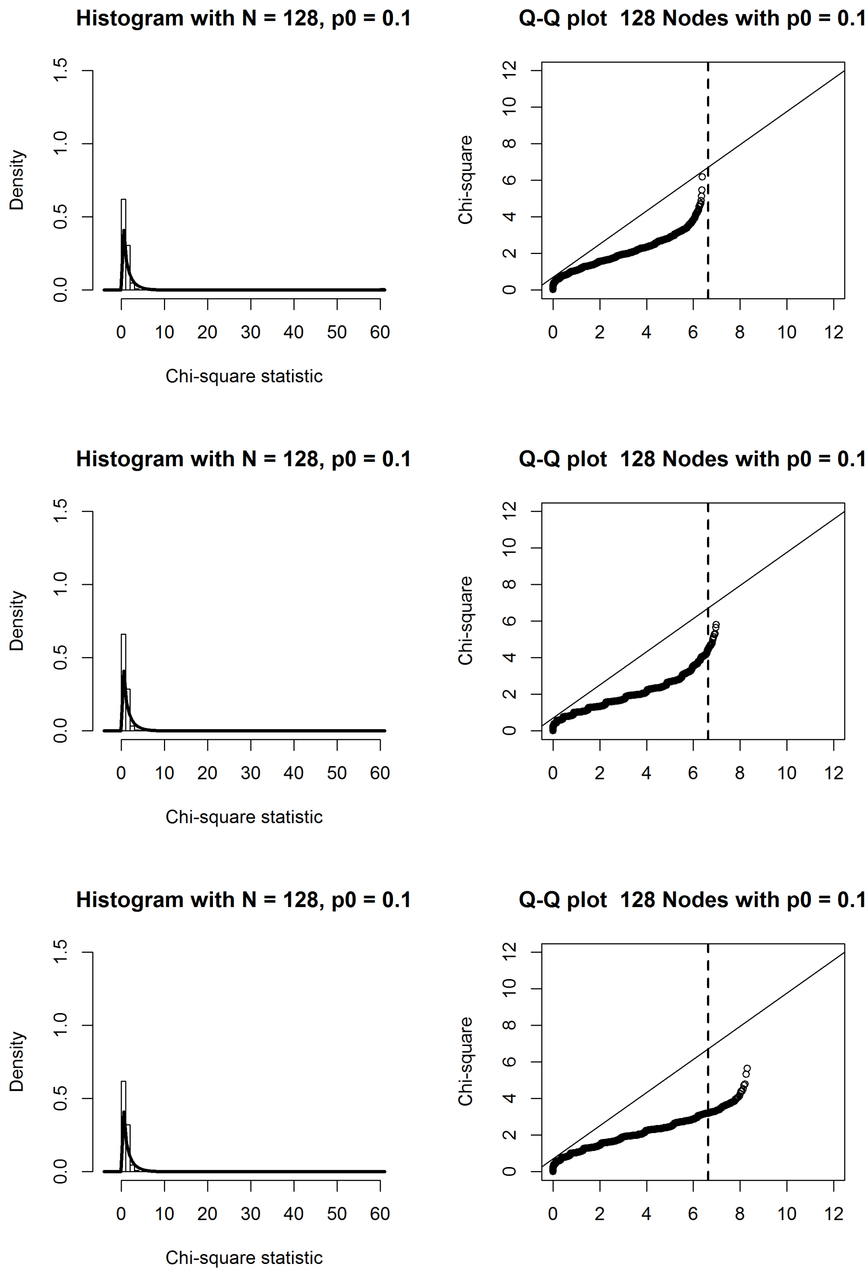

Figure 1: Left figures are histogram density plots of 10,000 simulations with chi-square distribution, $df = 1$, overlaid. $n = 128$ and $p_0 = 0.1$. Right figures are the Q-Q plots of the simulated statistics with the line $y = x$ representing the theoretical χ^2 with $df = 1$.((Top) Erdös-Rényi, (Middle) R-MAT, and (Bottom) Chung-Lu Model)

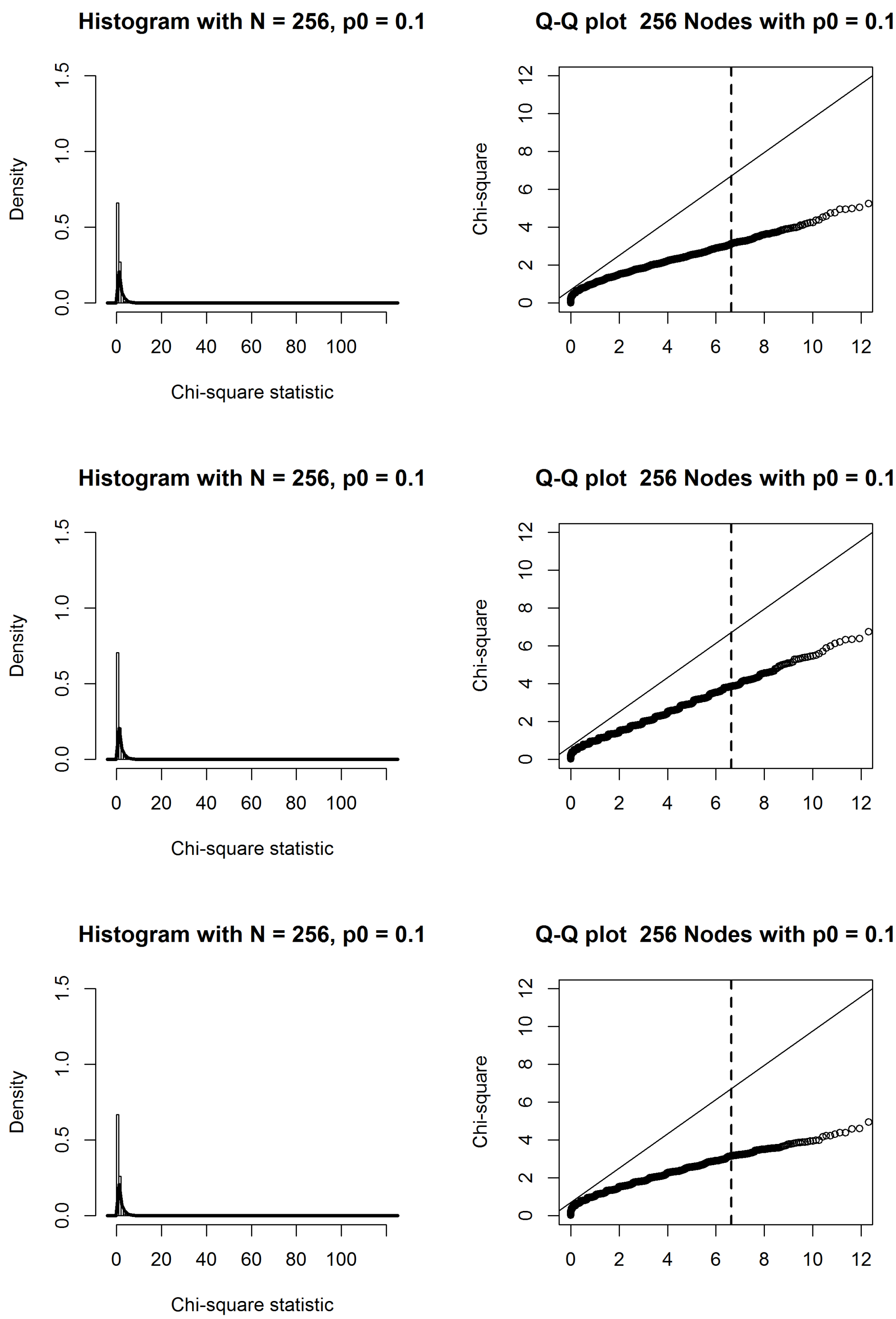

Figure 2: Left figures are histogram density plots of 10,000 simulations with chi-square distribution, $df = 1$, overlaid. $n = 256$ and $p_0 = 0.1$. Right figures are the Q-Q plots of the simulated statistics with the line $y = x$ representing the theoretical χ^2 with $df = 1$.((Top) Erdös-Rényi, (Middle) R-MAT, and (Bottom) Chung-Lu Model)

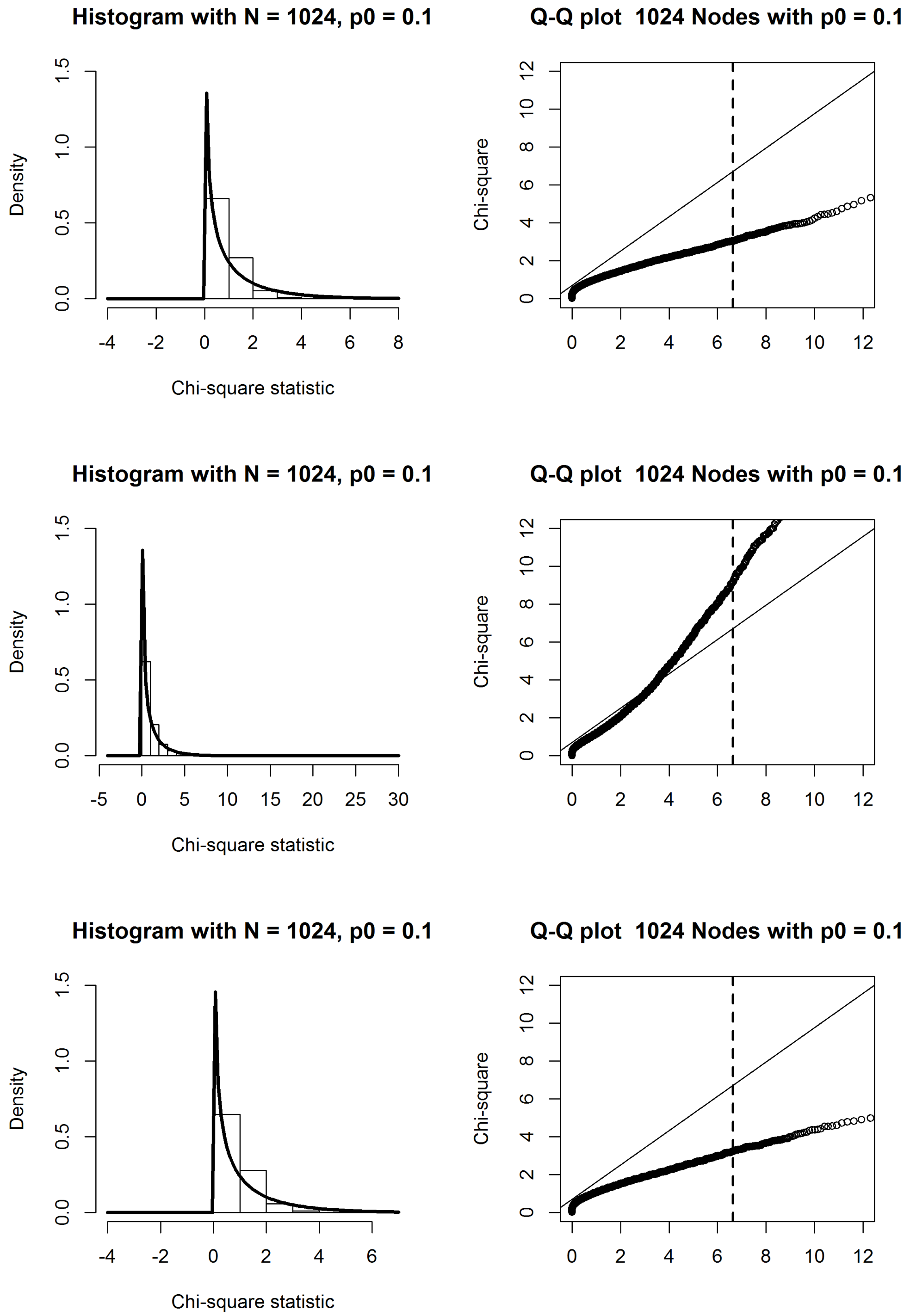

Figure 3: Left figures are histogram density plots of 10,000 simulations with chi-square distribution, $df = 1$, overlaid. $n = 1024$ and $p_0 = 0.1$. Right figures are the Q-Q plots of the simulated statistics with the line $y = x$ representing the theoretical χ^2 with $df = 1$. ((Top) Erdös-Rényi, (Middle) R-MAT, and (Bottom) Chung-Lu Model)

1.2. Improving the chi-square algorithm

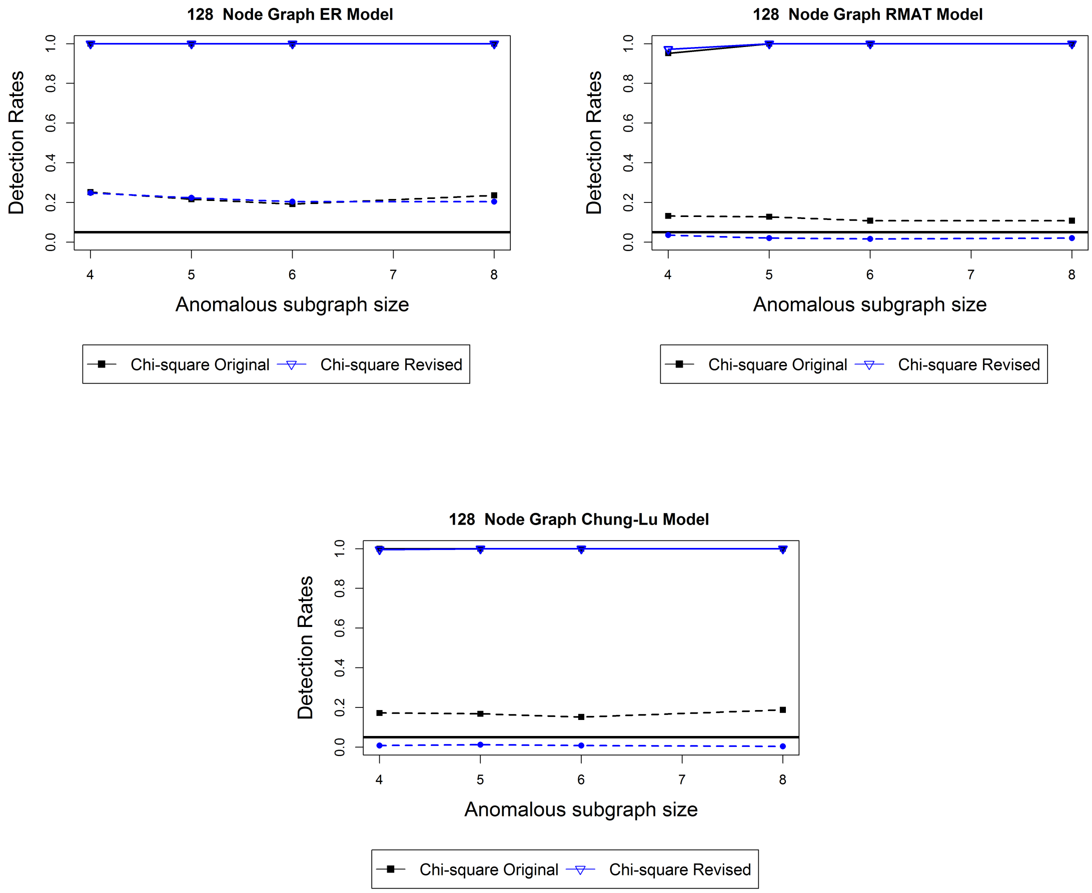

Figure 4: (Erdös-Rényi, R-MAT, and Chung-Lu Model) Number of anomalous subgraph varies from 3%, 4%, 5%, and 6% for $n = 128$. Detection rates are solid lines while false alarm rates are dashed lines. Background connectivity, $p_0 = 0.01$. A comparison of the traditional detection statistic and the improved version

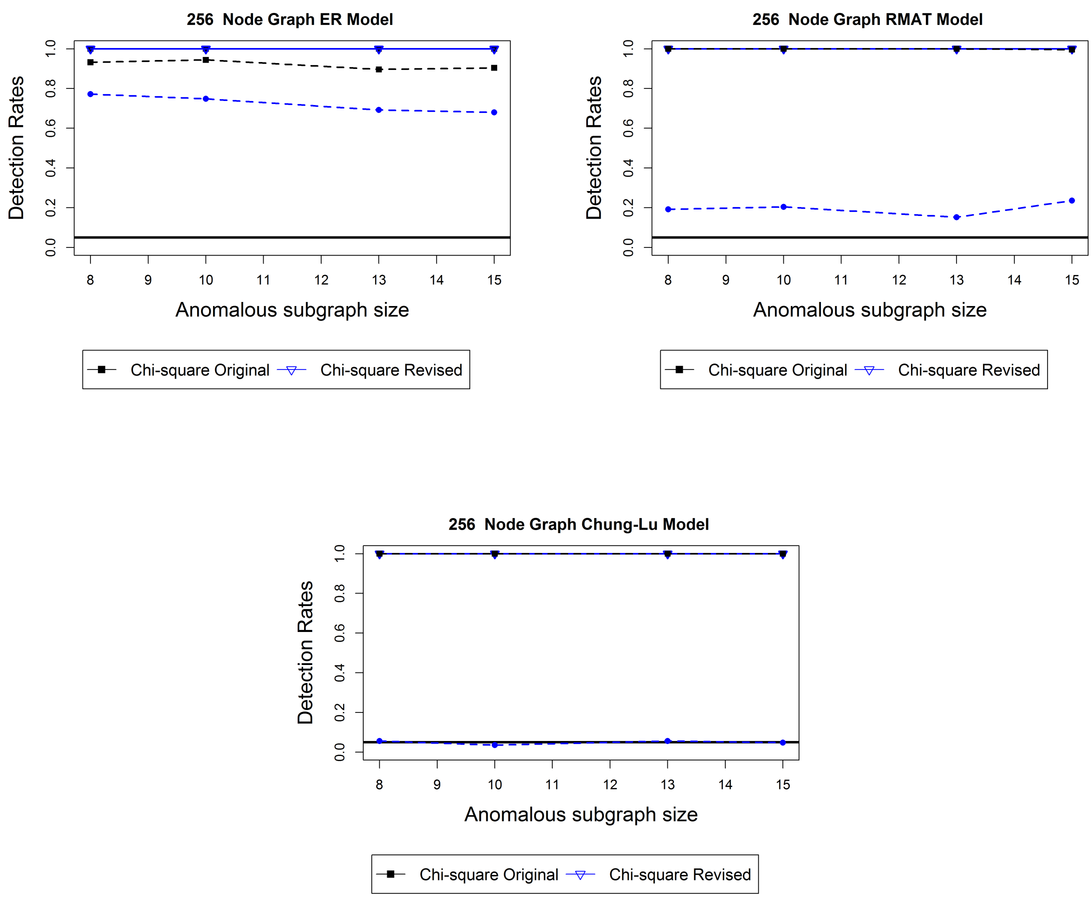

Figure 5: (Erdös-Rényi, R-MAT, and Chung-Lu Model) Number of anomalous subgraph varies from 3%, 4%, 5%, and 6% for $n = 256$. Detection rates are solid lines while false alarm rates are dashed lines. Background connectivity, $p_0 = 0.01$. A comparison of the traditional detection statistic and the improved version

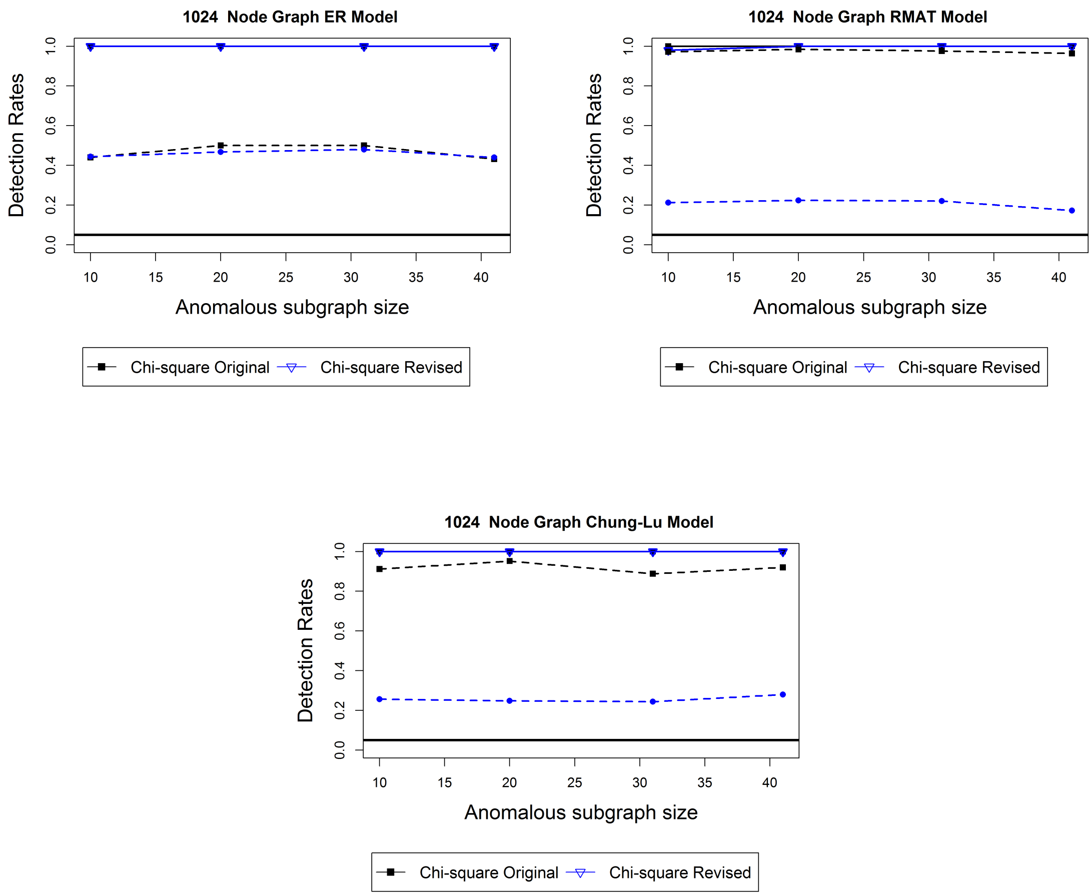

Figure 6: (Erdös-Rényi, R-MAT, and Chung-Lu Model) Number of anomalous subgraph varies from 1%, 2%, 3%, and 4% for $n = 1024$. Detection rates are solid lines while false alarm rates are dashed lines. Background connectivity, $p_0 = 0.01$. A comparison of the traditional detection statistic and the improved version

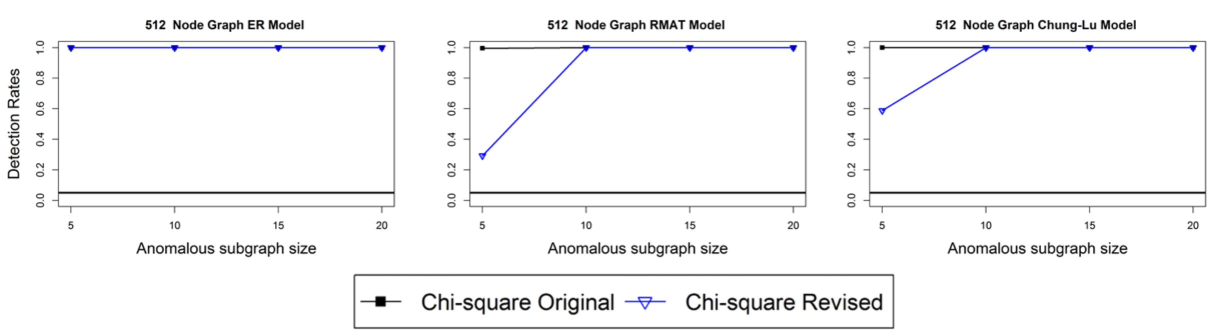

Figure 7: (Erdös-Rényi, R-MAT, and Chung-Lu Model) Number of anomalous subgraph varies from 1%, 2%, 3%, and 4% for $n = 512$. Detection rates are solid lines while false alarm rates are dashed lines. Background connectivity, $p_0 = 0.01$. A comparison of the traditional detection statistic and the improved version

2. Eigenvector L_1 norm algorithm methodology

2.1. Estimating a_m and b_m using historical data and setting $m < n$

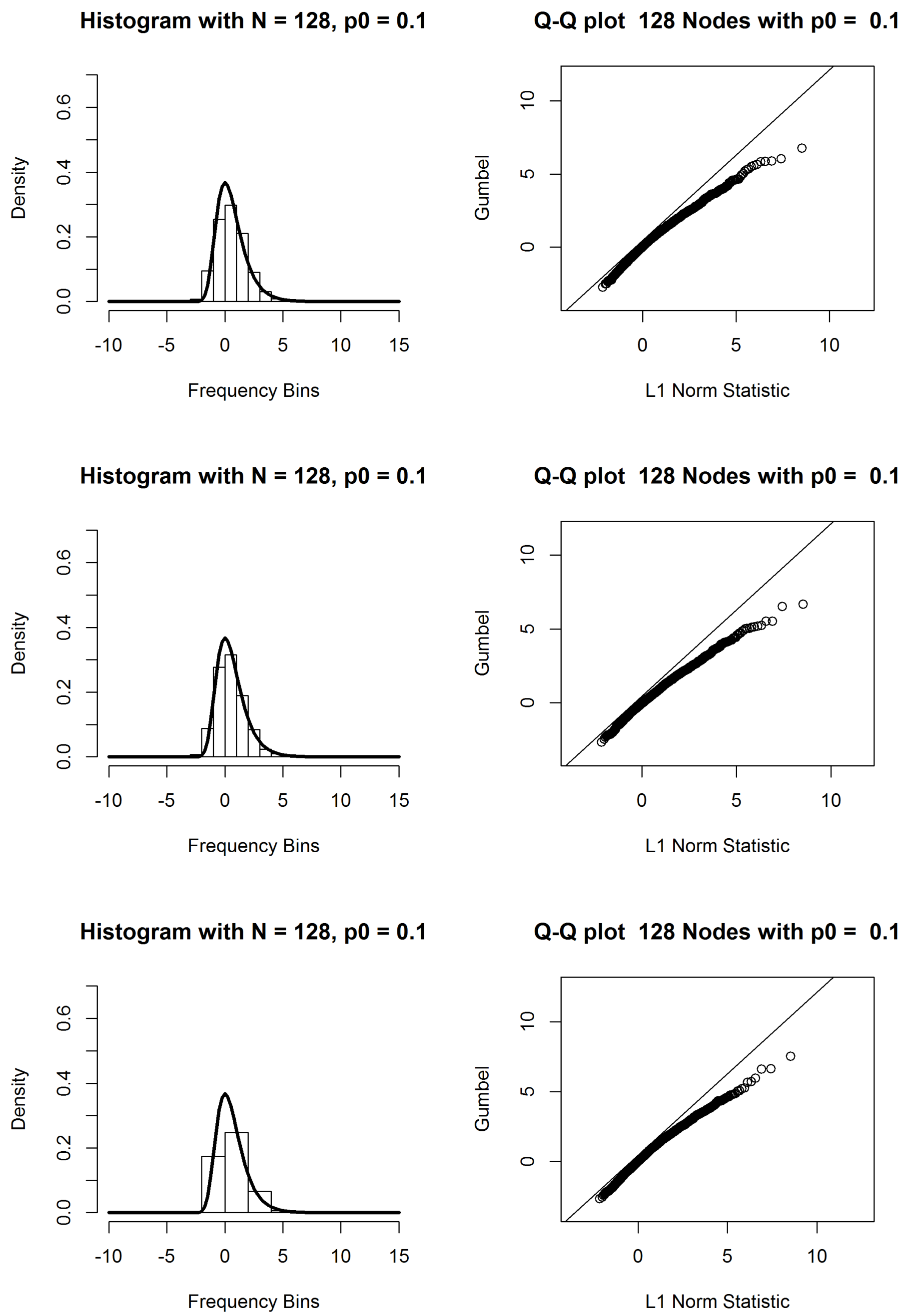

Figure 8: Left figures are histogram density plots when parameters a_m and b_m are estimated using historical data and MOM estimators with $m < n$. Solid black line represents the theoretical Gumbel distribution. Right figures are the Q-Q plots of the simulated statistics with the line $y = x$ representing the theoretical Gumbel distribution. This example is with $n = 128$ and $p_0 = 0.1$. ((Top) Erdös-Rényi, (Middle) R-MAT, and (Bottom) Chung-Lu Model)

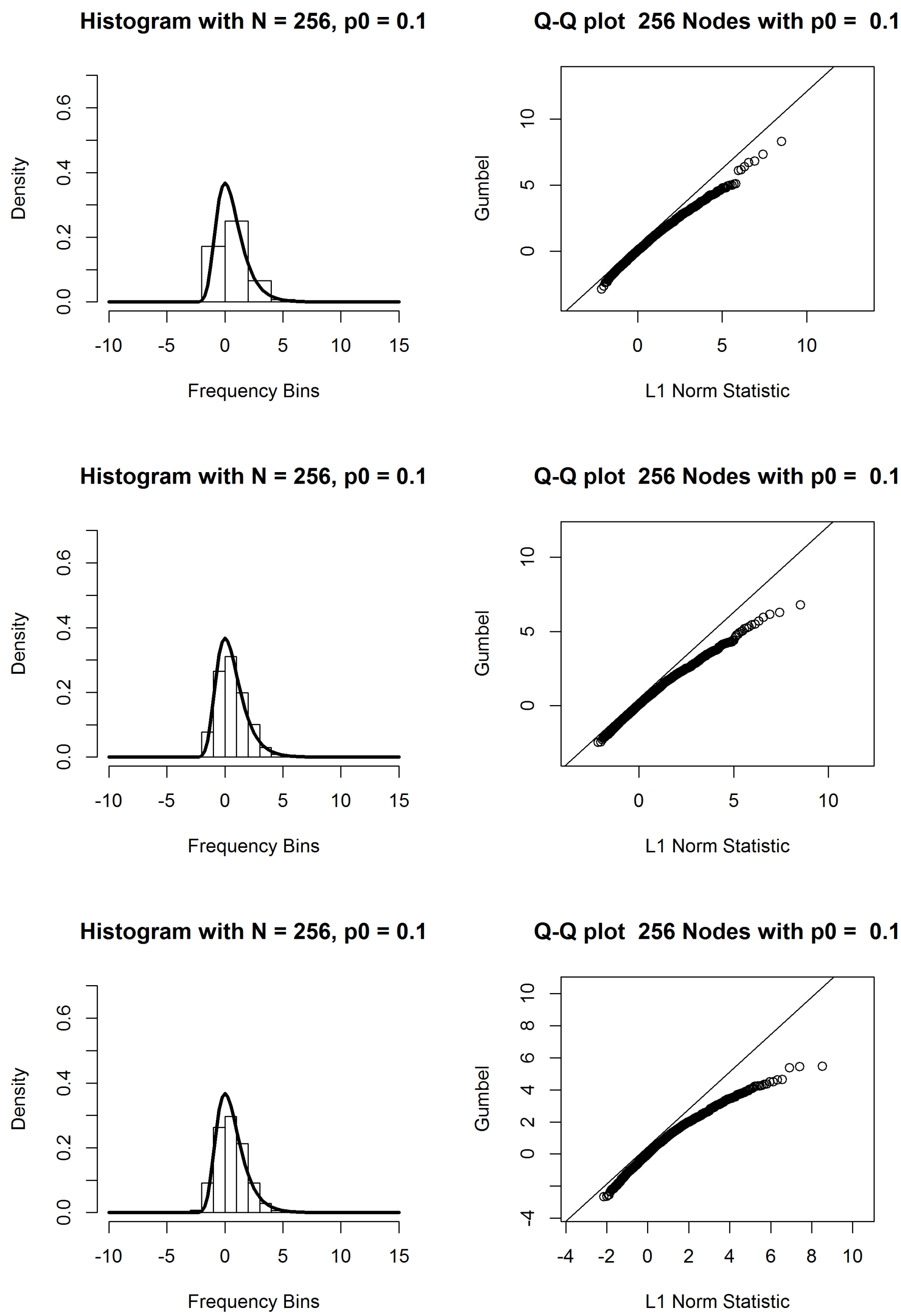

Figure 9: Left figures are histogram density plots when parameters a_m and b_m are estimated using historical data and MOM estimators with $m < n$. Solid black line represents the theoretical Gumbel distribution. Right figures are the Q-Q plots of the simulated statistics with the line $y = x$ representing the theoretical Gumbel distribution. This example is with $n = 256$ and $p_0 = 0.1$. ((Top) Erdös-Rényi, (Middle) R-MAT, and (Bottom) Chung-Lu Model)

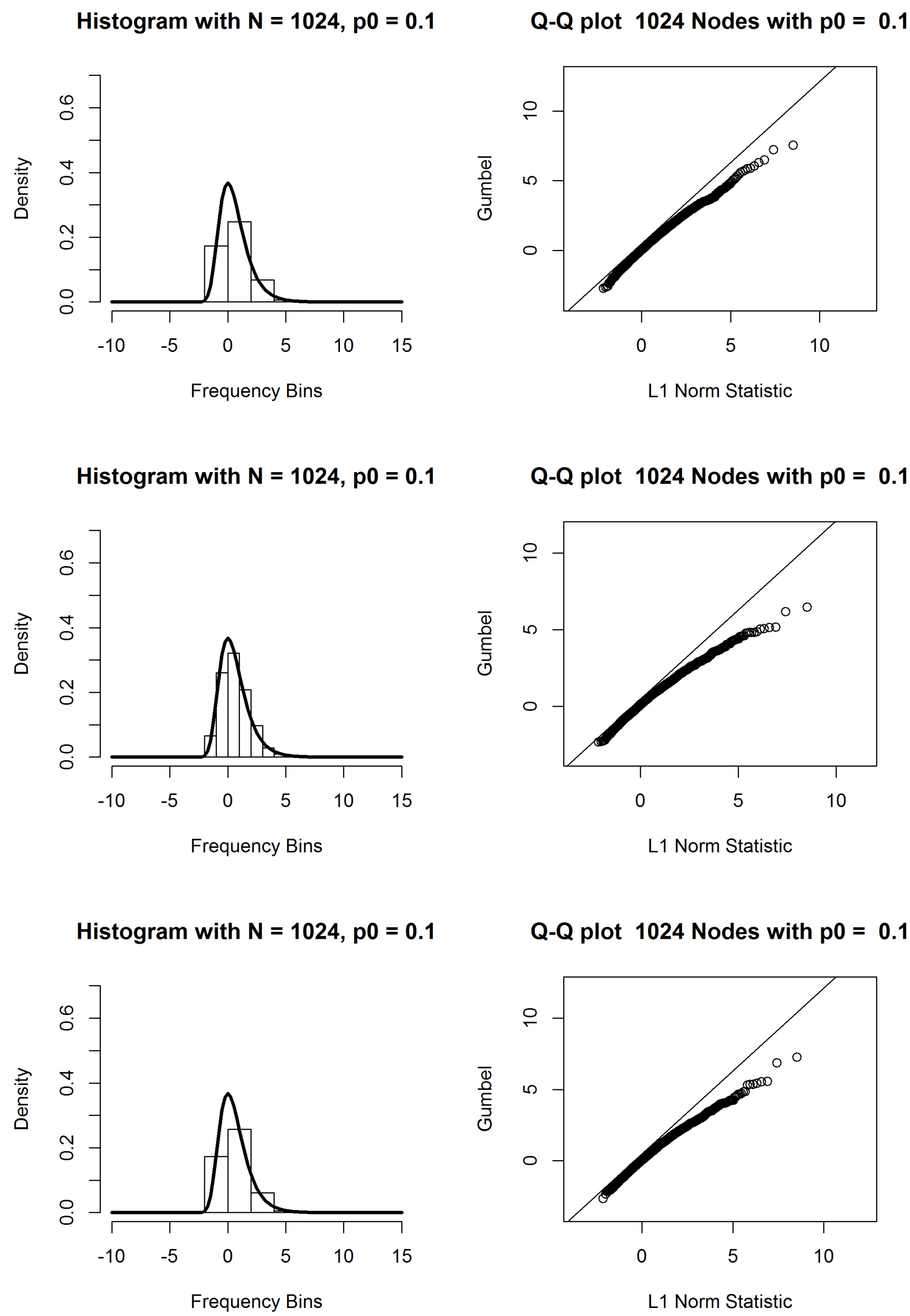

Figure 10: Left figures are histogram density plots when parameters a_m and b_m are estimated using historical data and MOM estimators with $m < n$. Solid black line represents the theoretical Gumbel distribution. Right figures are the Q-Q plots of the simulated statistics with the line $y = x$ representing the theoretical Gumbel distribution. This example is with $n = 1024$ and $p_0 = 0.1$. ((Top) Erdös-Rényi, (Middle) R-MAT, and (Bottom) Chung-Lu Model)

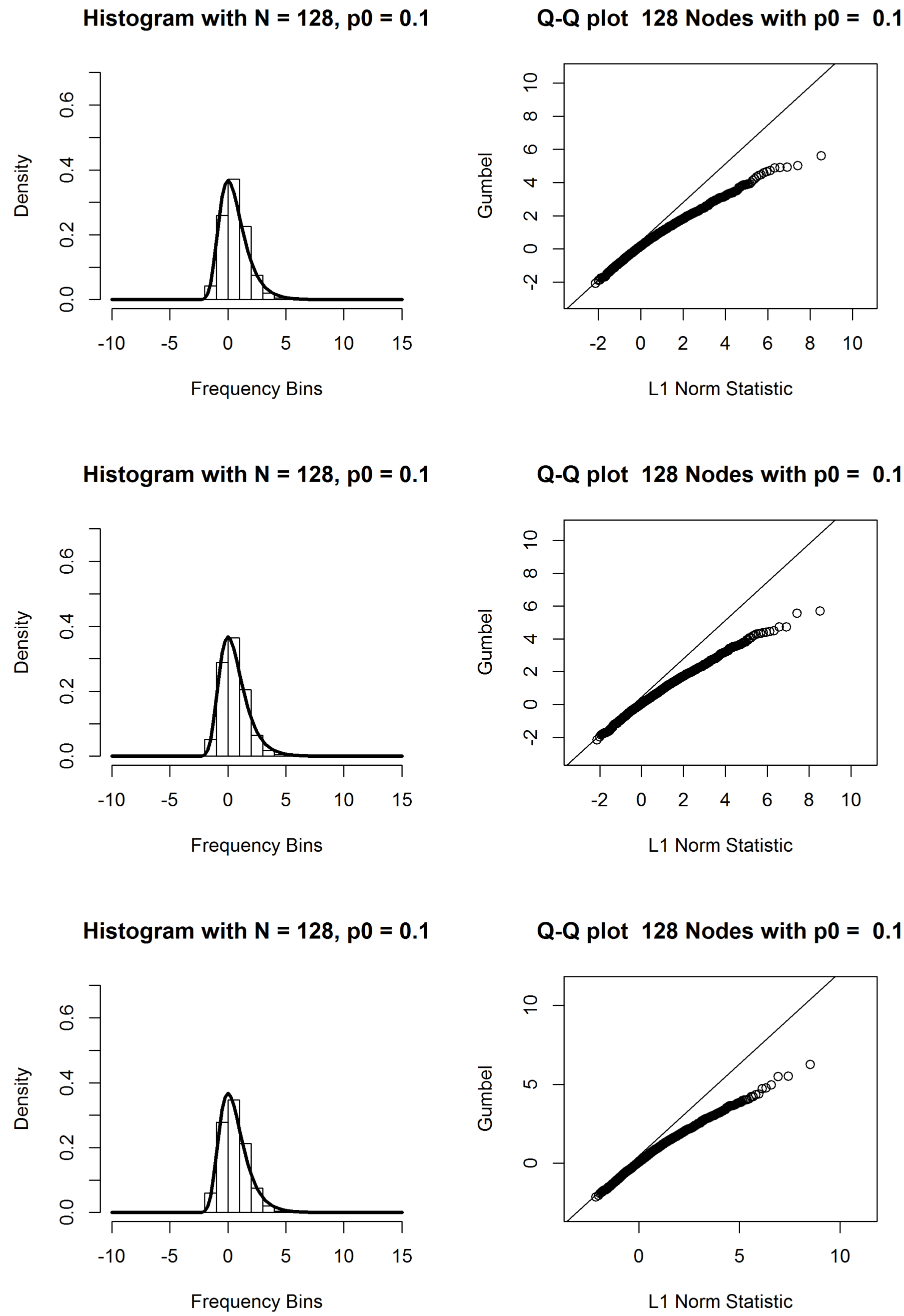

Figure 11: Left figures are histogram density plots when parameters a_m and b_m are estimated using the Extreme Value Theorem with $m < n$. Solid black line represents the theoretical Gumbel distribution. Right figures are the Q-Q plots of the simulated statistics with the line $y = x$ representing the theoretical Gumbel distribution. This example is with $n = 128$ and $p_0 = 0.1$. ((Top) Erdös-Rényi, (Middle) R-MAT, and (Bottom) Chung-Lu Model)

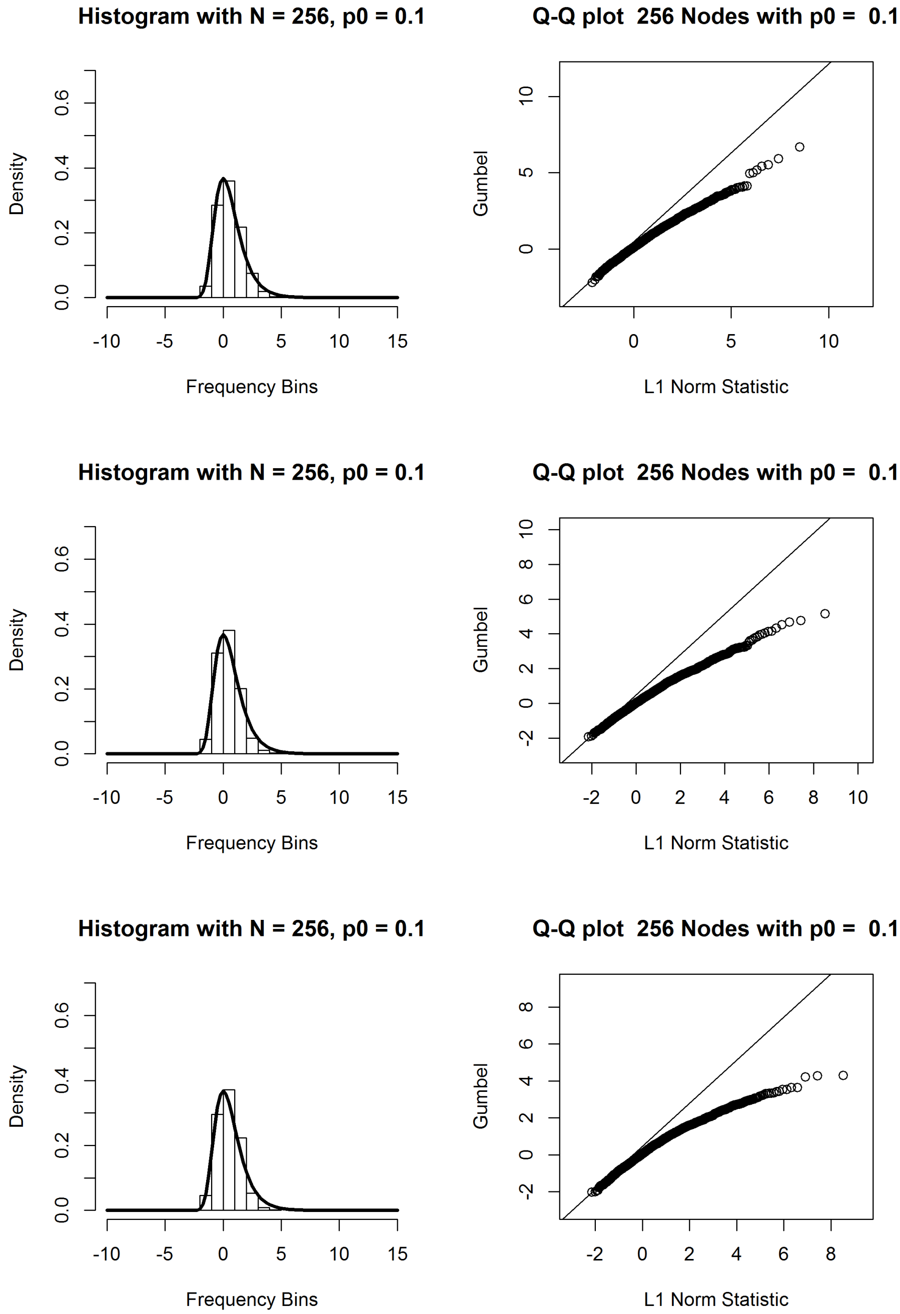

Figure 12: Left figures are histogram density plots when parameters a_m and b_m are estimated using the Extreme Value Theorem with $m < n$. Solid black line represents the theoretical Gumbel distribution. Right figures are the Q-Q plots of the simulated statistics with the line $y = x$ representing the theoretical Gumbel distribution. This example is with $n = 256$ and $p_0 = 0.1$. ((Top) Erdös-Rényi, (Middle) R-MAT, and (Bottom) Chung-Lu Model)

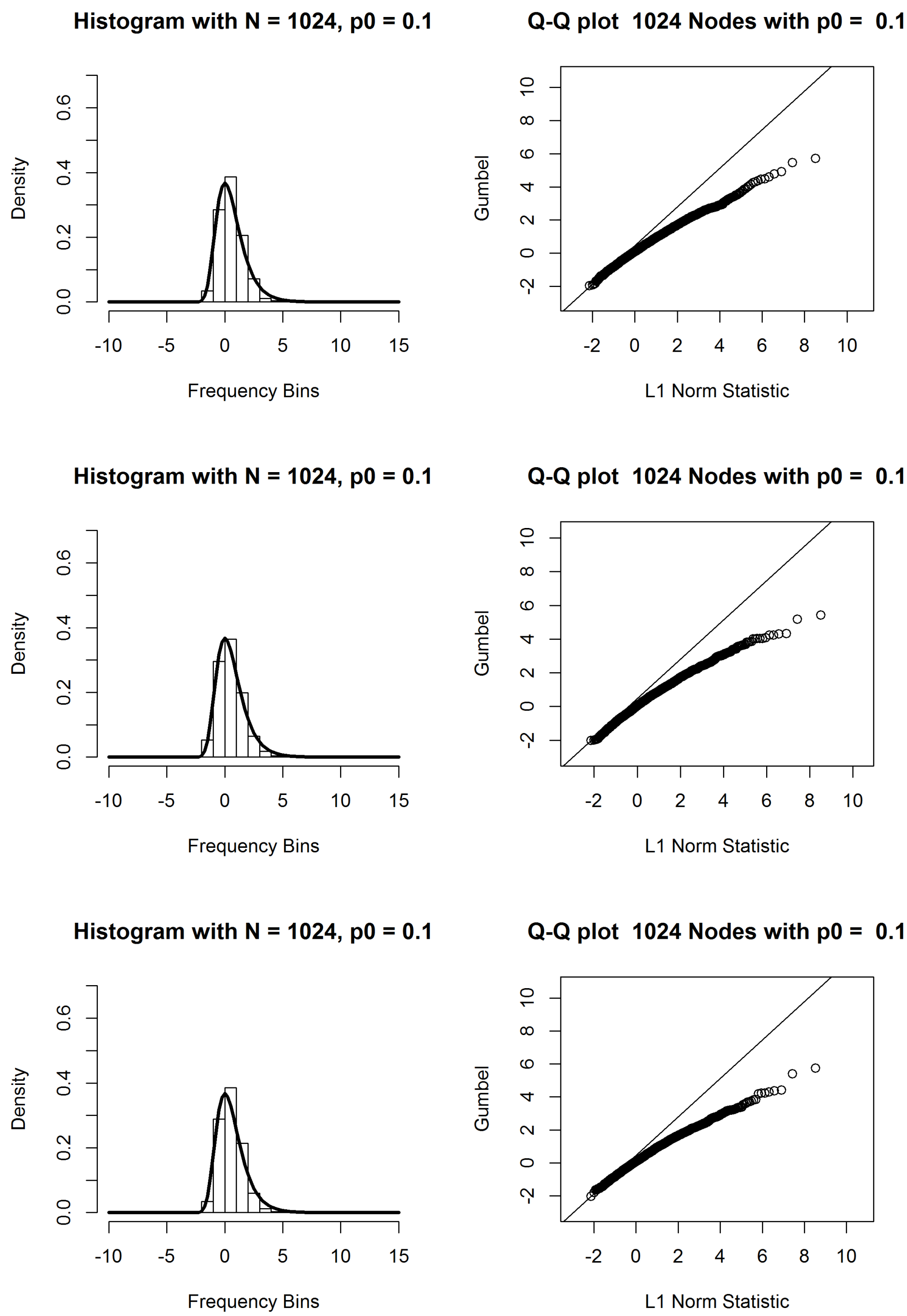

Figure 13: Left figures are histogram density plots when parameters a_m and b_m are estimated using the Extreme Value Theorem with $m < n$. Solid black line represents the theoretical Gumbel distribution. Right figures are the Q-Q plots of the simulated statistics with the line $y = x$ representing the theoretical Gumbel distribution. This example is with $n = 1024$ and $p_0 = 0.1$. ((Top) Erdös-Rényi, (Middle) R-MAT, and (Bottom) Chung-Lu Model)

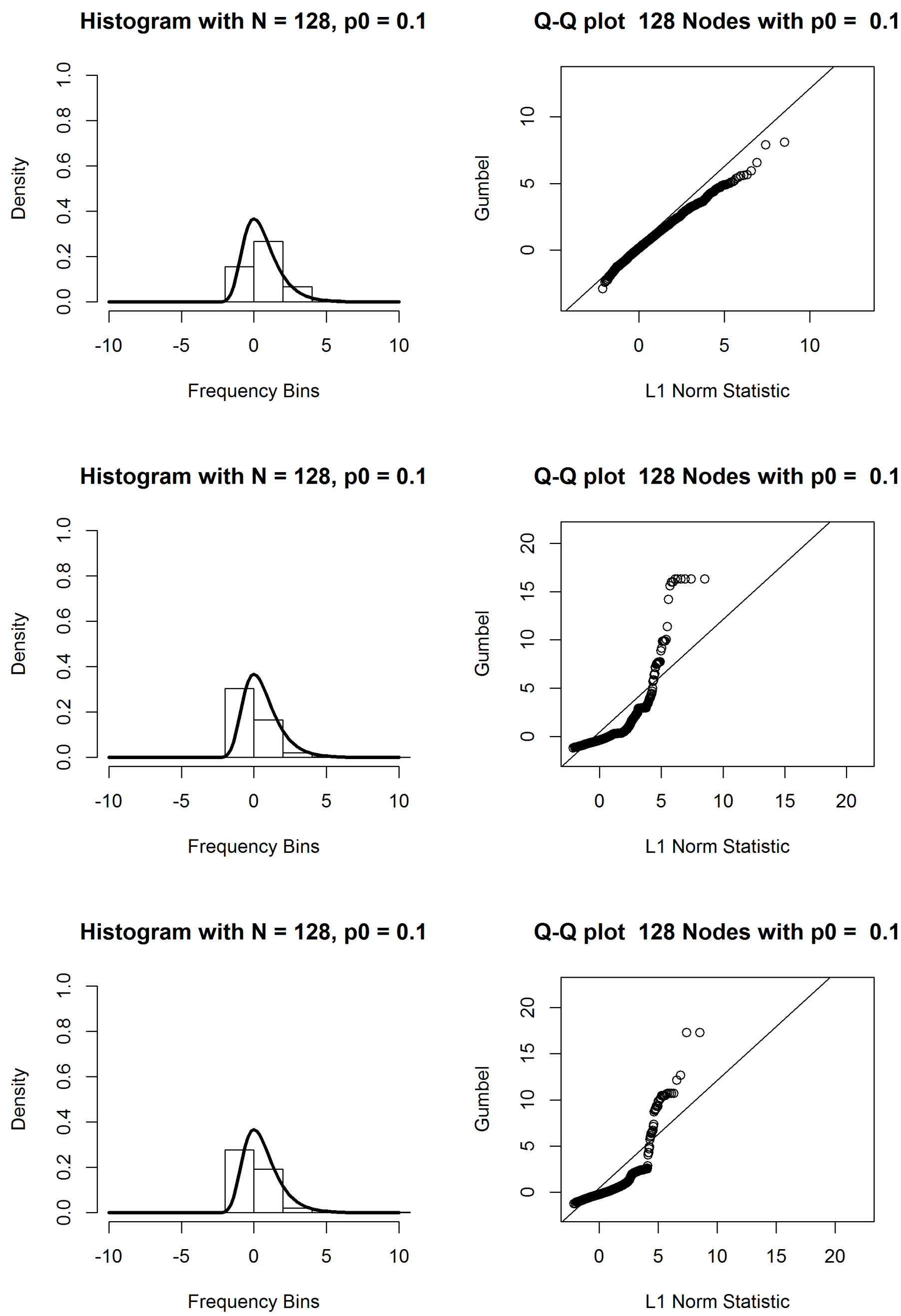

Figure 14: Left figures are histogram density plots when parameters a_m and b_m are estimated using historical data and MOM estimators with $m = n$. Solid black line represents the theoretical Gumbel distribution. Right figures are the Q-Q plots of the simulated statistics with the line $y = x$ representing the theoretical Gumbel distribution. This example is with $n = 128$ and $p_0 = 0.1$. ((Top) Erdös-Rényi, (Middle) R-MAT, and (Bottom) Chung-Lu Model)

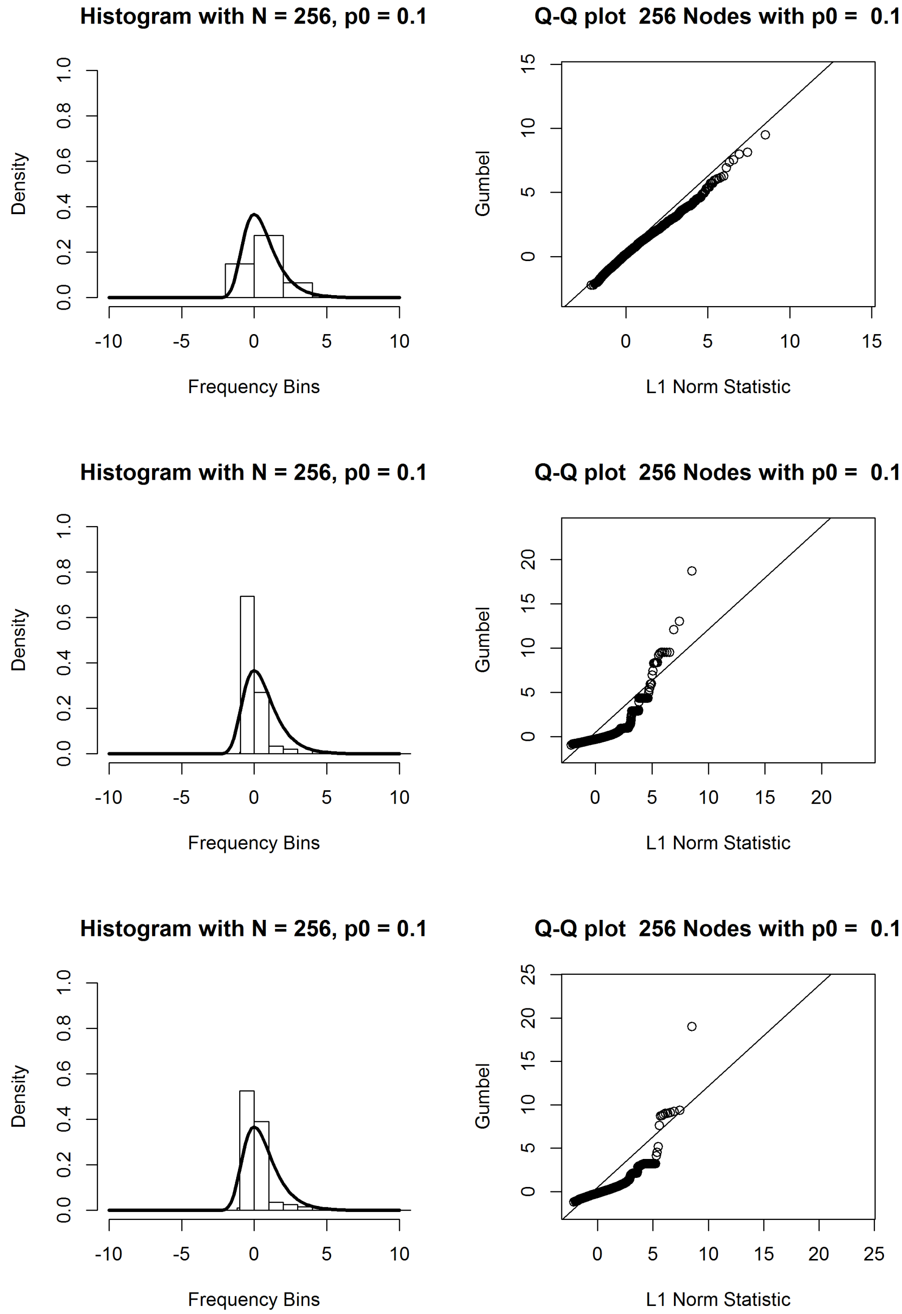

Figure 15: Left figures are histogram density plots when parameters a_m and b_m are estimated using historical data and MOM estimators with $m = n$. Solid black line represents the theoretical Gumbel distribution. Right figures are the Q-Q plots of the simulated statistics with the line $y = x$ representing the theoretical Gumbel distribution. This example is with $n = 256$ and $p_0 = 0.1$. ((Top) Erdös-Rényi, (Middle) R-MAT, and (Bottom) Chung-Lu Model)

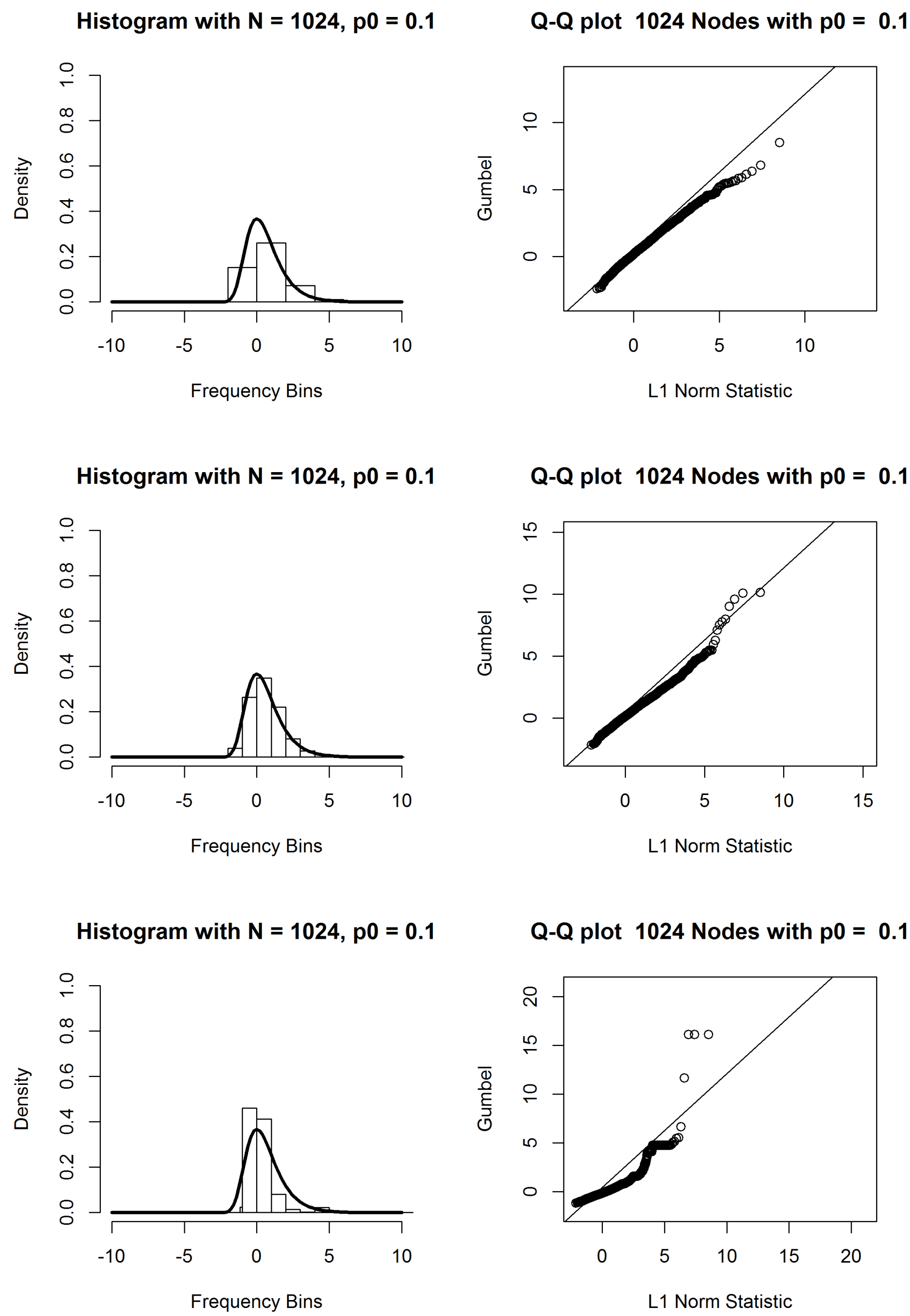

Figure 16: Left figures are histogram density plots when parameters a_m and b_m are estimated using historical data and MOM estimators with $m = n$. Solid black line represents the theoretical Gumbel distribution. Right figures are the Q-Q plots of the simulated statistics with the line $y = x$ representing the theoretical Gumbel distribution. This example is with $n = 1024$ and $p_0 = 0.1$. ((Top) Erdös-Rényi, (Middle) R-MAT, and (Bottom) Chung-Lu Model)

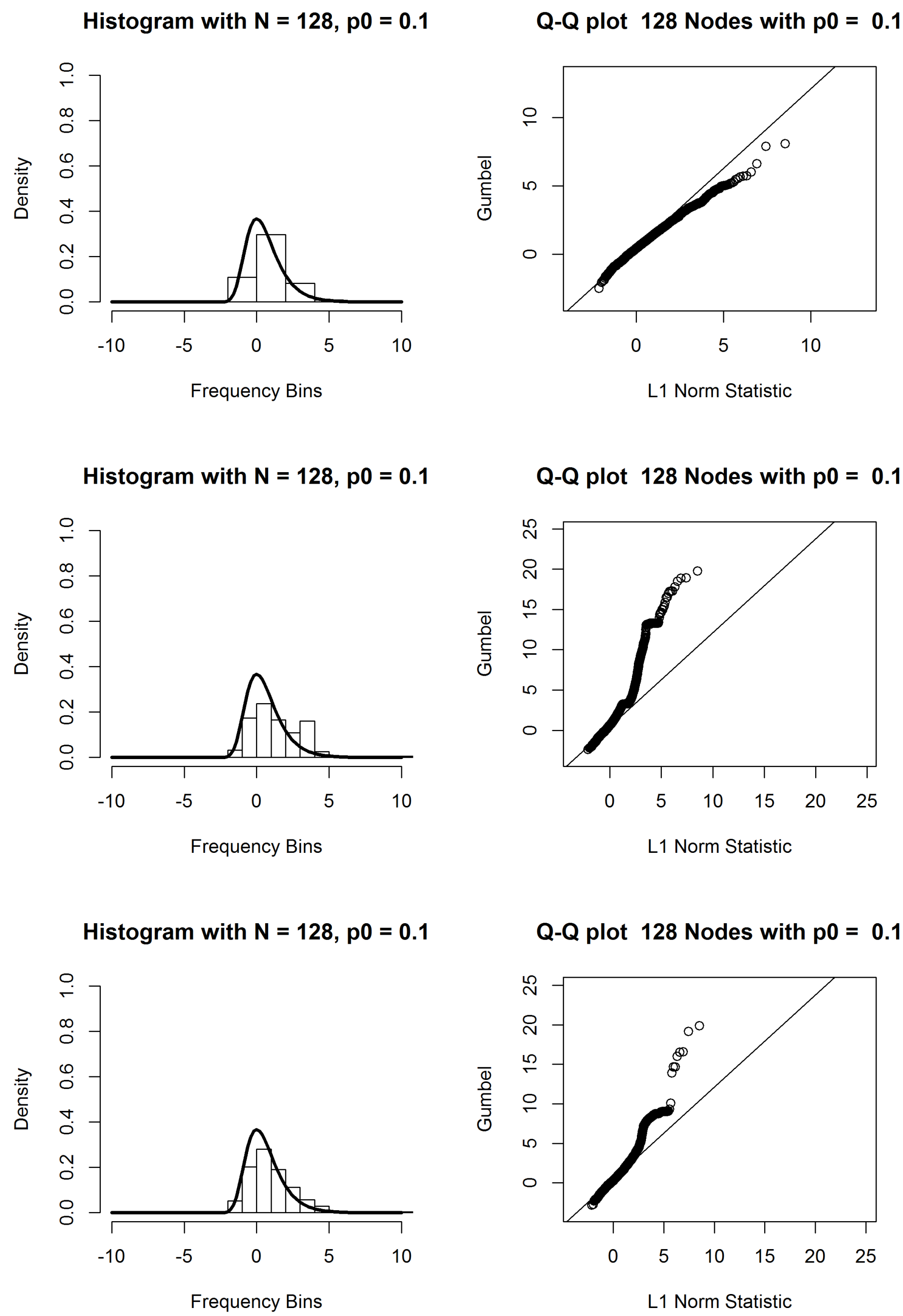

Figure 17: Left figures are histogram density plots when parameters a_m and b_m are estimated using the Extreme Value Theorem with $m = n$. Solid black line represents the theoretical Gumbel distribution. Right figures are the Q-Q plots of the simulated statistics with the line $y = x$ representing the theoretical Gumbel distribution. This example is with $n = 128$ and $p_0 = 0.1$. ((Top) Erdös-Rényi, (Middle) R-MAT, and (Bottom) Chung-Lu Model)

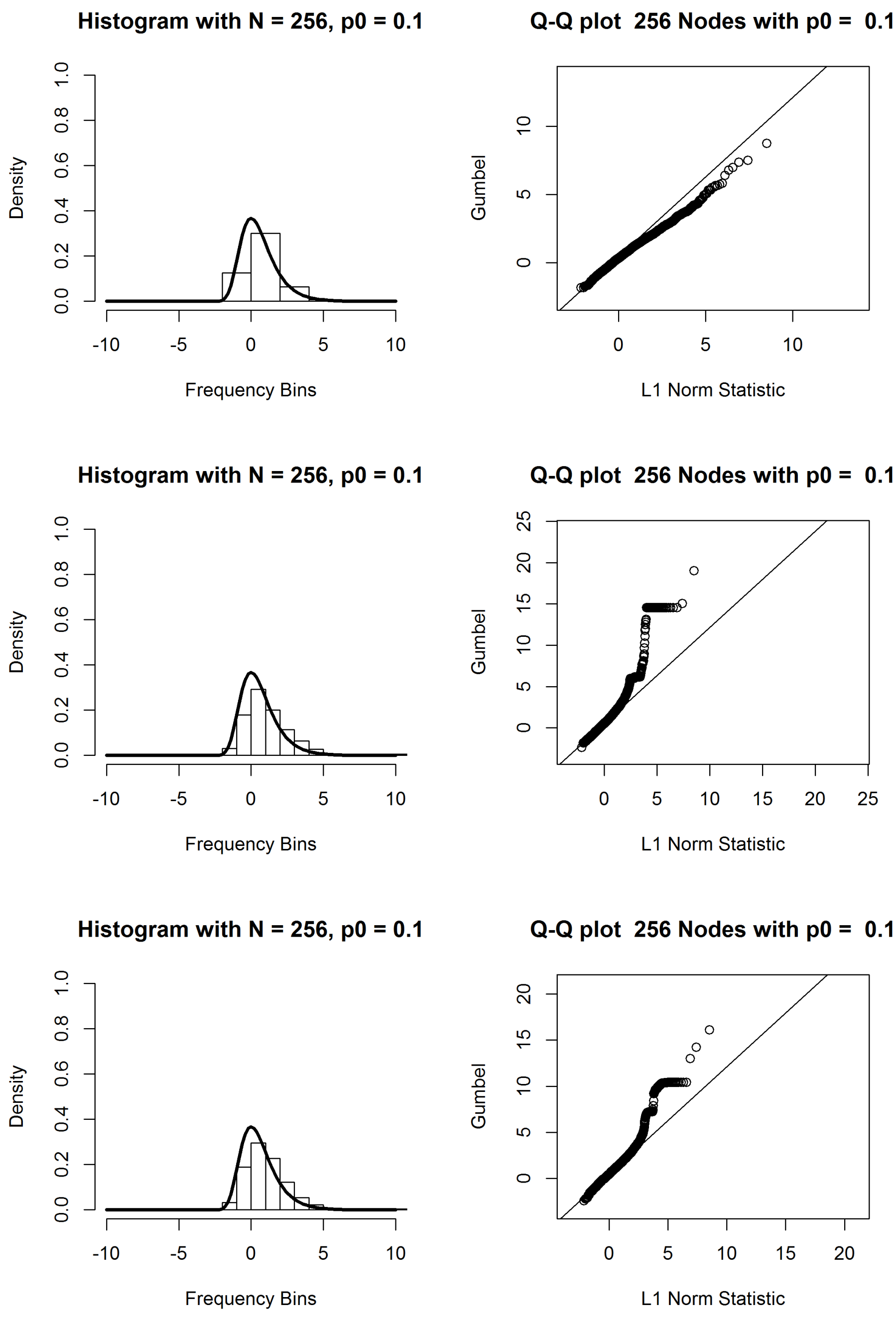

Figure 18: Left figures are histogram density plots when parameters a_m and b_m are estimated using the Extreme Value Theorem with $m = n$. Solid black line represents the theoretical Gumbel distribution. Right figures are the Q-Q plots of the simulated statistics with the line $y = x$ representing the theoretical Gumbel distribution. This example is with $n = 256$ and $p_0 = 0.1$. ((Top) Erdös-Rényi, (Middle) R-MAT, and (Bottom) Chung-Lu Model)

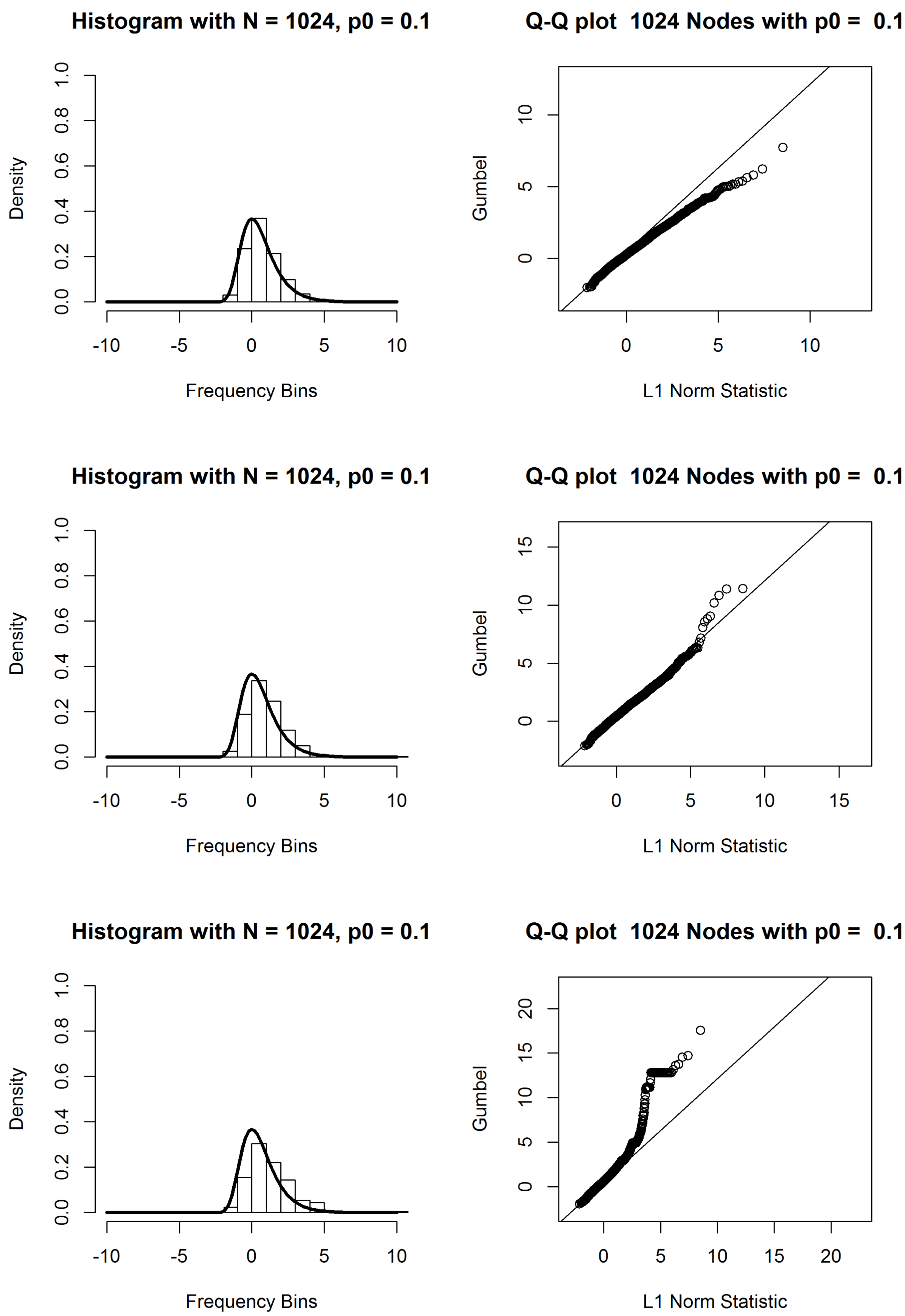

Figure 19: (Left figures are histogram density plots when parameters a_m and b_m are estimated using the Extreme Value Theorem with $m = n$. Solid black line represents the theoretical Gumbel distribution. Right figures are the Q-Q plots of the simulated statistics with the line $y = x$ representing the theoretical Gumbel distribution. This example is with $n = 1024$ and $p_0 = 0.1$. ((Top) Erdös-Rényi, (Middle) R-MAT, and (Bottom) Chung-Lu Model)

2.5. Improving the L_1 norm algorithm

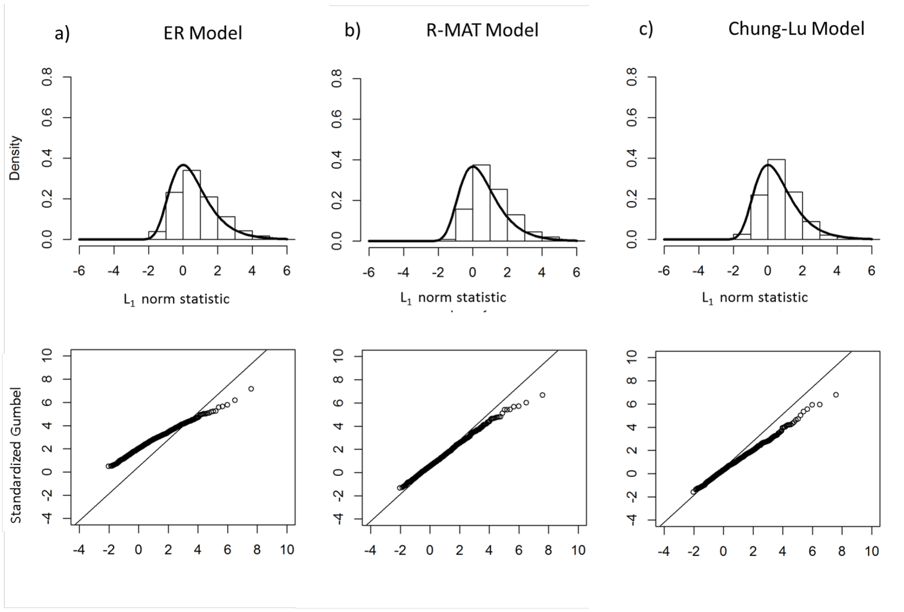

Figure 20: Top figures Histogram density plots of 10,000 simulations using inter-quantile range, IQR, and the median, M to standardize detection statistic. Bottom figures are the Q-Q plots of the simulation. $n = 512$ and $p_0 = 0.1$. ((a) Erdös-Rényi, (b) R-MAT, and (c) Chung-Lu Model)

Table 1: (L_1 norm, $m < n$, Median and IQR) 10,000 in-control simulations are run and the results compared to the theoretical Gumbel distribution when m = 30 for n = 128, 256 and m = 50 for n = 512, 1024.

		ER Model					R-MAT Model					Chung-Lu Model				
Network Size	p_0	95%	96%	97%	98%	99%	95%	96%	97%	98%	99%	95%	96%	97%	98%	99%
128	0.050	6.55	6.90	7.52	8.56	10.09	3.81	4.05	4.45	4.71	5.14	4.25	4.49	4.92	5.39	6.36
128	0.100	4.30	4.50	5.03	5.86	6.74	3.16	3.39	3.80	4.01	4.49	3.58	3.78	4.07	4.27	5.17
128	0.300	3.26	3.48	3.88	4.26	5.31	3.51	3.70	4.02	4.46	5.40	3.35	3.52	3.88	4.26	5.03
256	0.010	7.37	7.60	8.28	8.79	9.77	4.26	4.41	4.71	5.13	6.14	4.69	4.96	5.30	5.82	6.62
256	0.100	4.13	4.35	4.65	5.06	6.11	2.41	2.56	2.77	3.03	3.86	3.35	3.53	3.76	4.12	4.88
256	0.300	3.53	3.87	4.34	4.90	5.87	3.31	3.43	3.79	4.33	4.77	3.60	4.01	4.26	4.72	5.44
512	0.010	10.53	10.73	11.53	12.18	13.16	3.14	3.27	3.41	3.59	3.83	5.43	5.69	5.94	6.52	7.02
512	0.100	3.17	3.39	3.68	4.08	4.68	3.50	3.68	3.93	4.35	4.76	2.77	2.98	3.26	3.66	4.22
512	0.300	3.18	3.47	3.79	4.16	4.87	4.24	4.57	4.71	5.01	5.76	3.14	3.27	3.51	4.08	4.74
1024	0.010	8.91	9.70	10.21	11.15	13.36	1.48	1.58	1.68	1.82	2.01	3.85	3.98	4.22	4.60	5.02
1024	0.100	3.44	3.65	3.92	4.36	5.09	8.05	8.81	9.42	9.99	10.71	2.45	2.60	2.81	3.36	3.84
1024	0.300	3.29	3.58	3.83	4.30	4.78	7.73	7.98	8.19	8.89	9.43	3.15	3.41	3.67	4.06	4.56
Gumbel quantiles		**2.97**	**3.20**	**3.49**	**3.90**	**4.60**	**2.97**	**3.20**	**3.49**	**3.90**	**4.60**	**2.97**	**3.20**	**3.49**	**3.90**	**4.60**

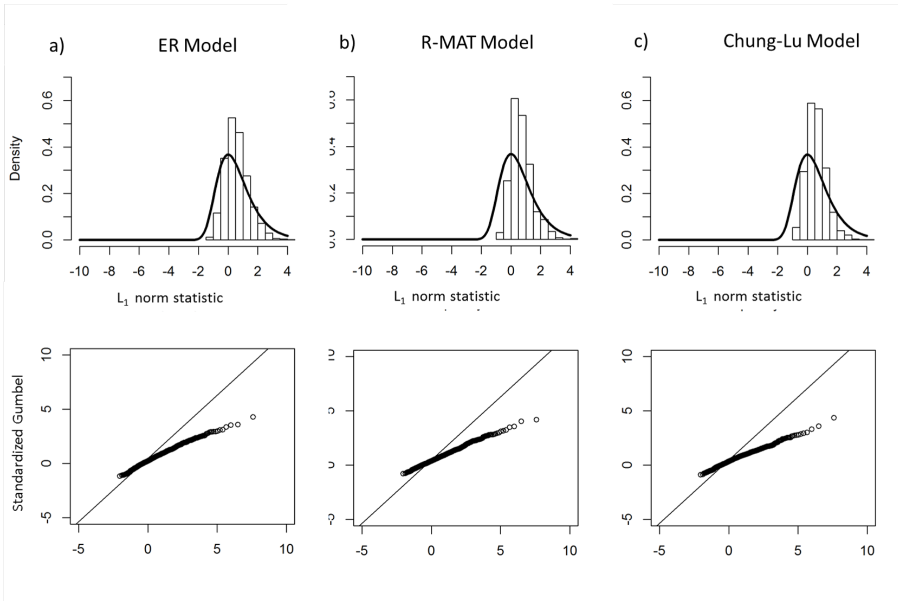

Figure 21: Top figures Histogram density plots of 10,000 simulations using mean, μ, and the standard deviation, σ, to standardize detection statistic. Bottom figures are the Q-Q plots of the simulation. $n = 512$ and $p_0 = 0.1$. ((a) Erdös-Rényi, (b) R-MAT, and (c) Chung-Lu Model)

Table 2: (L_1 norm, $m < n$, Mean and SD) 10,000 in-control simulations are run and the results compared to the theoretical Gumbel distribution when m = 30 for n = 128, 256 and m = 50 for n = 512, 1024.

		ER Model					R-MAT Model					Chung-Lu Model				
Network Size	p_0	95%	96%	97%	98%	99%	95%	96%	97%	98%	99%	95%	96%	97%	98%	99%
128	0.050	3.08	3.21	3.36	3.59	3.81	1.88	1.97	2.06	2.15	2.41	2.05	2.12	2.18	2.25	2.56
128	0.100	2.31	2.38	2.51	2.80	3.16	1.69	1.82	1.94	2.07	2.37	1.90	2.01	2.10	2.25	2.46
128	0.300	1.87	1.97	2.08	2.23	2.33	1.85	1.94	2.07	2.23	2.36	1.72	1.82	2.04	2.21	2.45
256	0.010	3.20	3.30	3.44	3.62	3.83	2.23	2.37	2.48	2.68	3.20	2.36	2.47	2.54	2.71	3.02
256	0.100	2.07	2.20	2.46	2.61	2.89	1.37	1.44	1.51	1.67	1.88	1.72	1.79	1.90	2.12	2.28
256	0.300	1.91	2.00	2.16	2.34	2.65	1.38	1.48	1.67	1.84	2.35	1.72	1.84	2.05	2.27	2.50
512	0.010	4.85	5.00	5.13	5.30	5.80	1.69	1.76	1.83	1.93	2.12	2.97	3.04	3.10	3.20	3.37
512	0.100	2.09	2.19	2.36	2.51	2.91	2.20	2.27	2.36	2.65	2.84	1.77	1.85	2.07	2.40	2.67
512	0.300	2.02	2.10	2.23	2.62	3.12	0.66	0.73	0.82	0.91	1.07	1.84	1.97	2.13	2.27	2.74
1024	0.010	4.84	5.00	5.15	5.43	5.98	1.19	1.24	1.37	1.45	1.54	2.47	2.53	2.67	2.85	3.12
1024	0.100	2.26	2.42	2.49	2.70	3.11	4.25	4.37	4.51	4.64	4.83	1.65	1.73	1.82	1.96	2.22
1024	0.300	2.00	2.12	2.31	2.53	2.84	0.57	0.60	0.63	0.69	0.74	1.92	2.11	2.34	2.43	2.73
Gumbel quantiles		**2.97**	**3.20**	**3.49**	**3.90**	**4.60**	**2.97**	**3.20**	**3.49**	**3.90**	**4.60**	**2.97**	**3.20**	**3.49**	**3.90**	**4.60**

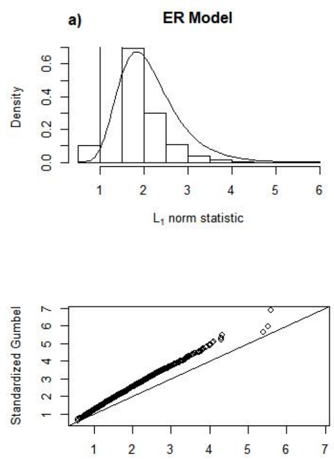

Figure 22: ((a) Erdös-Rényi, and (c) Chung-Lu Model) Top figures Histogram density plots of 10,000 simulations using inter-quantile range, IQR_m, and the median, M_m to estimate parameters, μ_m and σ_m. Bottom figures are the Q-Q plots of the simulation for Count Networks.

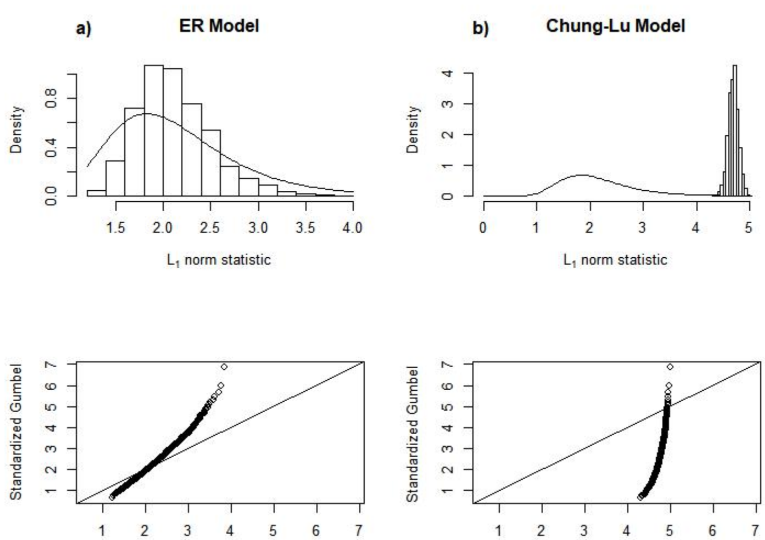

Figure 23: ((a) Erdös-Rényi, and (c) Chung-Lu Model) Top figures Histogram density plots of 10,000 simulations using standard deviation and the mean to estimate parameters, μ_m and σ_m. Bottom figures are the Q-Q plots of the simulation for Count Networks.

Table 3: (L_1 norm, $m < n$, *median* and *IQR*) 10,000 in-control simulations are run and the results compared to the theoretical Gumbel distribution when m = 30 for n = 128, 256 and m = 50 for n = 512, 1024 for count networks.

		ER Model					Chung-Lu Model					
Network size	λ_0	95%	96%	97%	98%	99%	η	95%	96%	97%	98%	99%
128	0.2	4.913	4.941	4.984	5.024	5.093	0.133	4.913	4.941	4.984	5.024	5.093
128	1	5.964	5.980	5.991	6.012	6.041	0.333	5.964	5.980	5.991	6.012	6.041
128	3	6.115	6.122	6.130	6.146	6.166	1	6.115	6.122	6.130	6.146	6.166
256	0.2	5.091	5.113	5.133	5.168	5.234	0.133	5.091	5.113	5.133	5.168	5.234
256	1	5.530	5.553	5.577	5.602	5.631	0.333	5.530	5.553	5.577	5.602	5.631
256	3	6.061	6.069	6.079	6.090	6.109	1	6.061	6.069	6.079	6.090	6.109
512	0.2	8.263	8.274	8.295	8.318	8.350	0.133	8.263	8.274	8.295	8.318	8.350
512	1	7.885	7.902	7.932	7.958	8.035	0.333	7.885	7.902	7.932	7.958	8.035
512	3	8.890	8.912	8.934	8.966	9.010	1	8.890	8.912	8.934	8.966	9.010
1024	0.2	9.026	9.036	9.048	9.069	9.105	0.133	9.026	9.036	9.048	9.069	9.105
1024	1	7.859	7.880	7.901	7.922	7.976	0.333	7.859	7.880	7.901	7.922	7.976
1024	3	9.283	9.295	9.307	9.324	9.354	1	9.283	9.295	9.307	9.324	9.354
Gumbel quantiles		**2.970**	**3.199**	**3.491**	**3.902**	**4.600**		**2.970**	**3.199**	**3.491**	**3.902**	**4.600**

Table 4: (L_1 norm, $m < n$, *mean* and *SD*) 10,000 in-control simulations are run and the results compared to the theoretical Gumbel distribution when m = 30 for n = 128, 256 and m = 50 for n = 512, 1024 for count networks.

		ER Model					Chung-Lu Model					
Network size	λ_0	95%	96%	97%	98%	99%	η	95%	96%	97%	98%	99%
128	0.2	2.583	2.722	2.872	3.115	3.516	0.133	4.913	4.941	4.984	5.024	5.093
128	1	1.926	2.021	2.154	2.309	2.645	0.333	5.964	5.980	5.991	6.012	6.041
128	3	1.838	1.924	2.030	2.186	2.531	1	6.115	6.122	6.130	6.146	6.166
256	0.200	2.184	2.321	2.480	2.633	2.887	0.133	5.091	5.113	5.133	5.168	5.234
256	1	1.851	1.965	2.114	2.302	2.572	0.333	5.530	5.553	5.577	5.602	5.631
256	3	1.852	1.947	2.099	2.301	2.515	1	6.061	6.069	6.079	6.090	6.109
512	0.2	2.274	2.393	2.515	2.716	3.137	0.133	8.263	8.274	8.295	8.318	8.350
512	1	2.045	2.152	2.279	2.503	2.827	0.333	7.885	7.902	7.932	7.958	8.035
512	3	1.985	2.103	2.257	2.414	2.900	1	8.890	8.912	8.934	8.966	9.010
1024	0.200	2.053	2.195	2.335	2.503	2.890	0.133	9.026	9.036	9.048	9.069	9.105
1024	1	1.938	2.051	2.177	2.431	2.742	0.333	7.859	7.880	7.901	7.922	7.976
1024	3	2.070	2.199	2.373	2.552	2.764	1	9.283	9.295	9.307	9.324	9.354
Gumbel quantiles		**2.970**	**3.199**	**3.491**	**3.902**	**4.600**		**2.970**	**3.199**	**3.491**	**3.902**	**4.600**

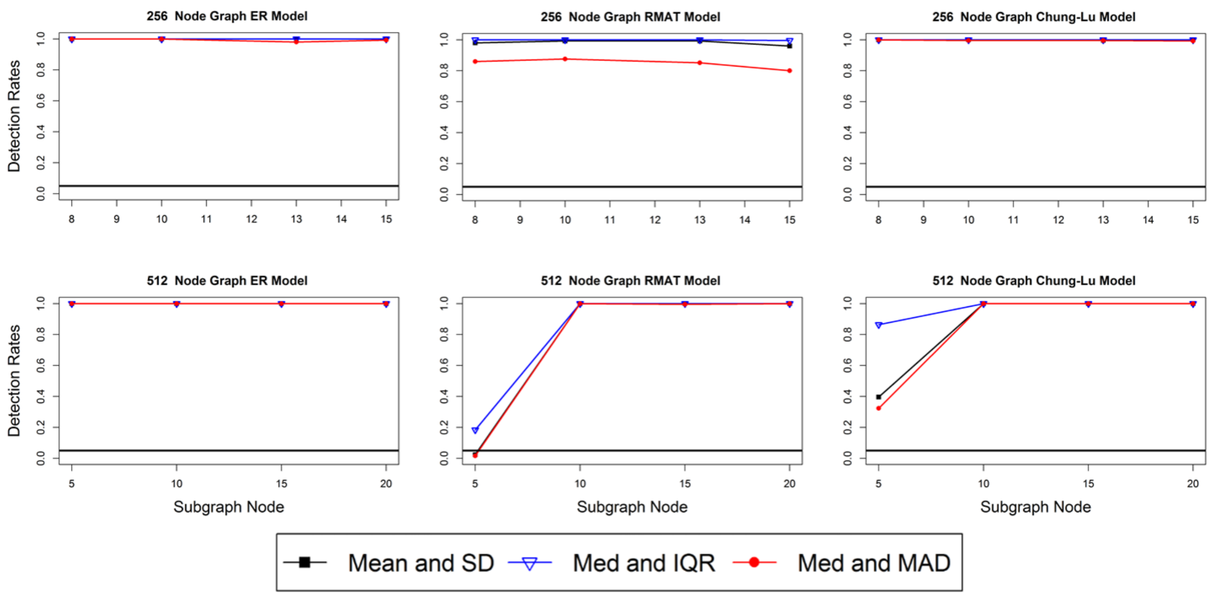

Figure 24: Binary network. Detection and False alarm rates with $n = 256$ and 512 and $m < n$. Number of anomalous subgraph varies from 3%, 4%, 5%, and 6% for $n = 256$ and 3%, 4%, 5%, and 6% for $n = 512$. Detection rates are solid lines while false alarm rates are dashed lines. Background connectivity, $p_0 = 0.01$. (Erdös-Rényi, R-MAT, and Chung-Lu Model)

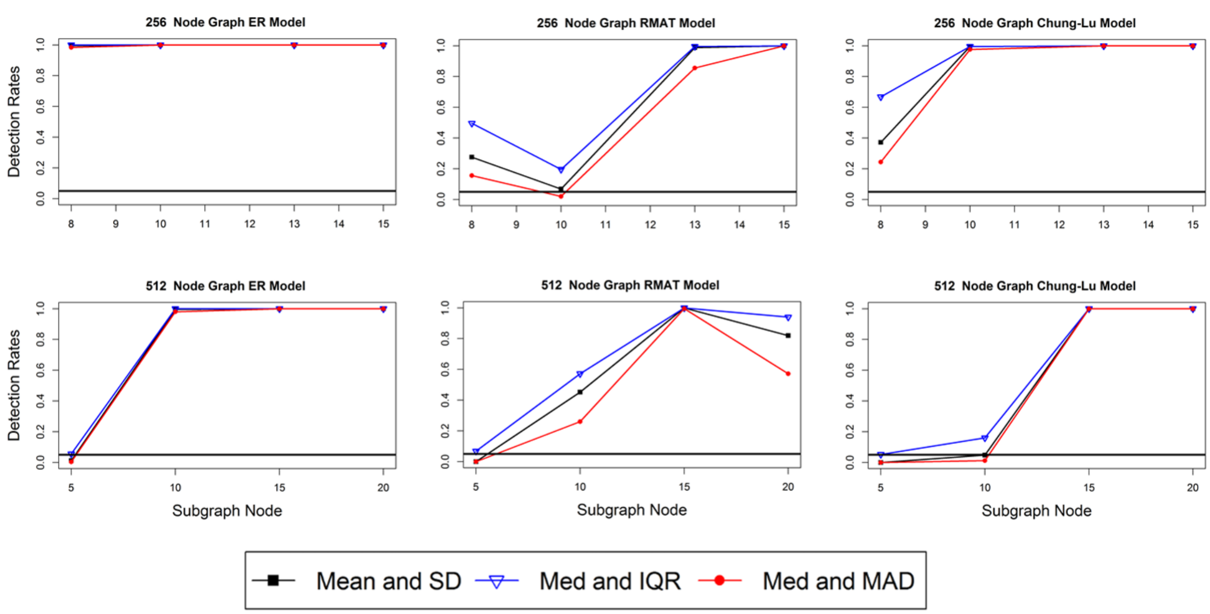

Figure 25: Binary network. Detection and False alarm rates with $n = 256$ and 512 and $m < n$. Number of anomalous subgraph varies from 3%, 4%, 5%, and 6% for $n = 256$ and 3%, 4%, 5%, and 6% for $n = 512$. Detection rates are solid lines while false alarm rates are dashed lines. Background connectivity, $p_0 = 0.1$. (Erdös-Rényi, R-MAT, and Chung-Lu Model)

3. Evaluating algorithm performance

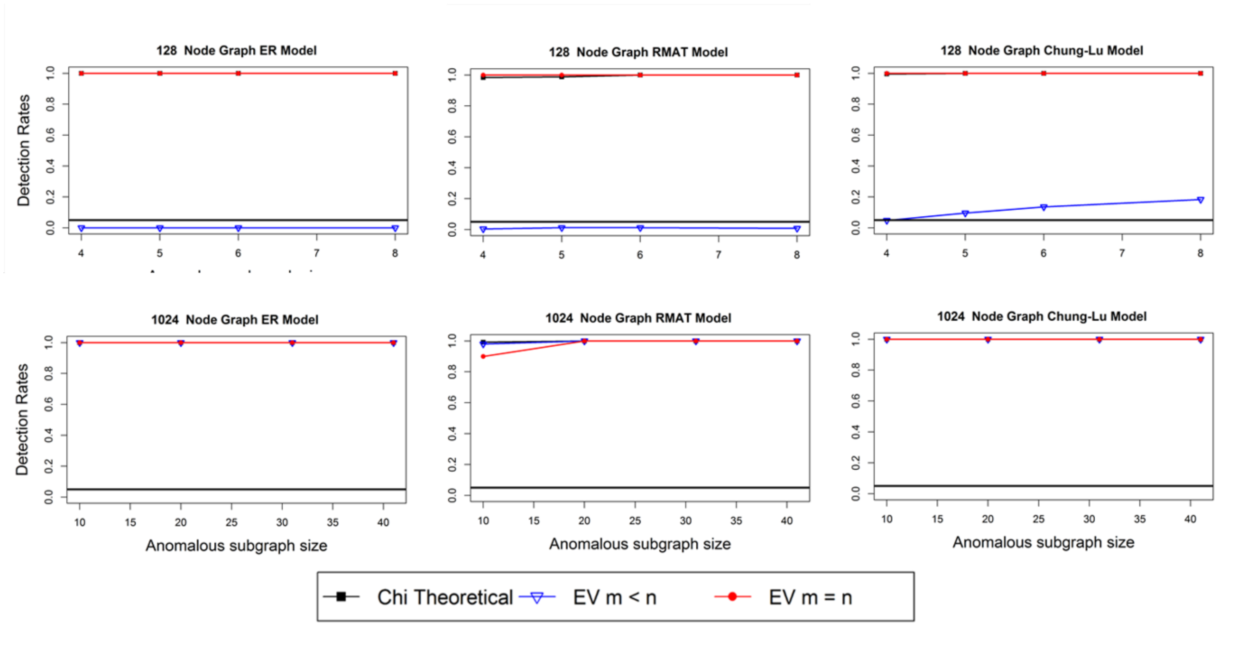

Figure 26: (Erdös-Rényi, R-MAT, and Chung-Lu Model) Detection and False alarm rates with $n = 128$ and 1024. Number of anomalous subgraph nodes varies from 3%, 4%, 5%, and 6% of the network size for $n = 256$ and 3%, 4%, 5%, and 6% of the network size for $n = 1024$. Detection rates are solid lines while false alarm rates are dashed lines. Background connectivity, $p_0 = 0.01$

4. Applying anomaly detection algorithms to Count Networks

4.1. Evaluating statistical properties of the algorithms in count networks when there is no anomaly

4.1.1. Statistical properties of the Chi-square algorithm

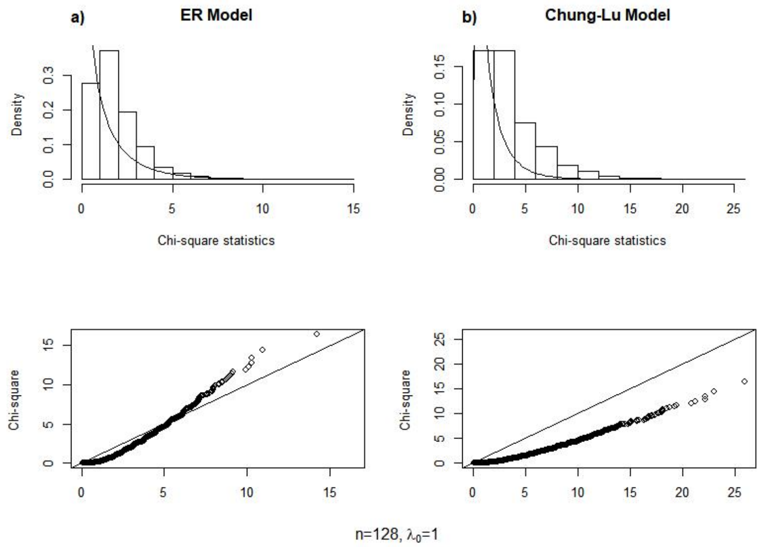

Figure 27: Top figures are histogram density plots of the chi-square statistic based on 10,000 simulations with chi-square distribution overlaid. $n = 128$ and $\lambda_0 = 1$. Bottom figures are the corresponding Q-Q plots.((a) Erdös-Rényi, and (b) Chung-Lu Models)

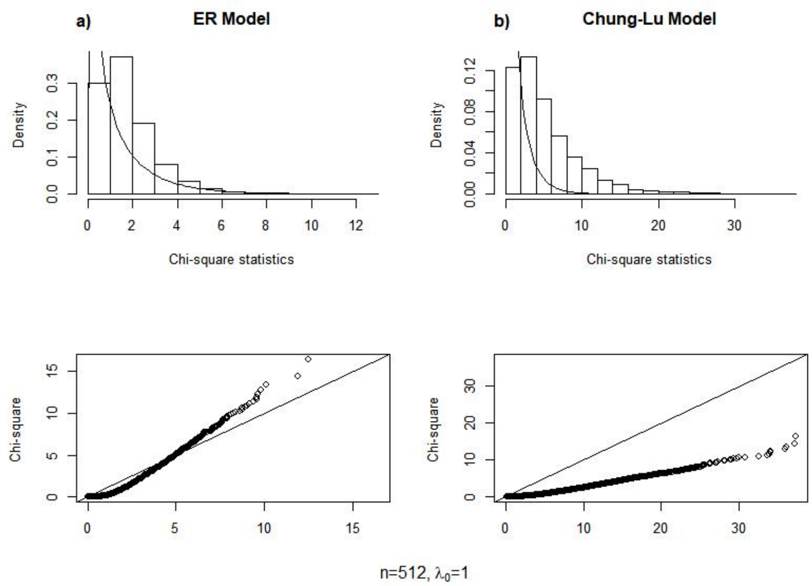

Figure 28: Top figures are histogram density plots of the chi-square statistic based on 10,000 simulations with chi-square distribution overlaid. $n = 512$ and $\lambda_0 = 1$. Bottom figures are the corresponding Q-Q plots.((a) Erdös-Rényi, and (b) Chung-Lu Models)

4.1.2. Statistical properties of the L_1 norm algorithm

Estimating a_m and b_m using historical data

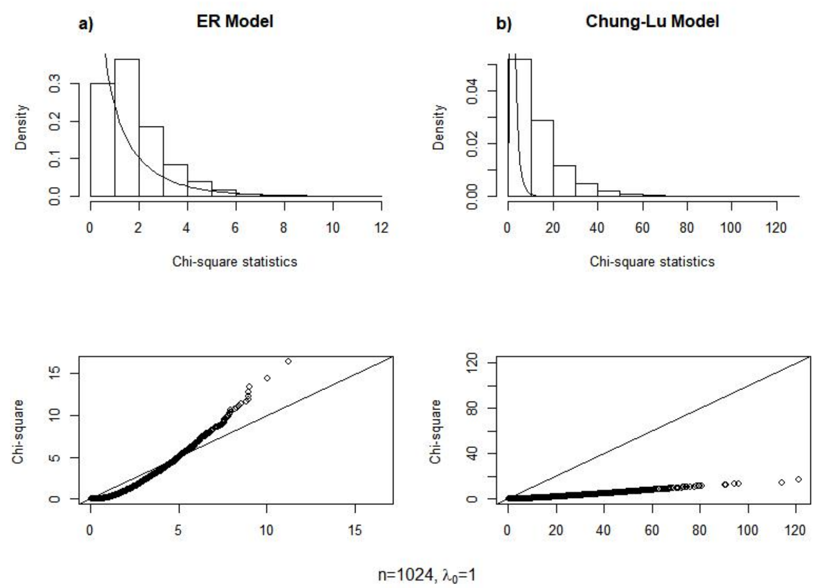

Figure 29: Top figures are histogram density plots of the chi-square statistic based on 10,000 simulations with chi-square distribution overlaid. $n = 1024$ and $\lambda_0 = 1$. Bottom figures are the corresponding Q-Q plots.((a) Erdös-Rényi, and (b) Chung-Lu Models)

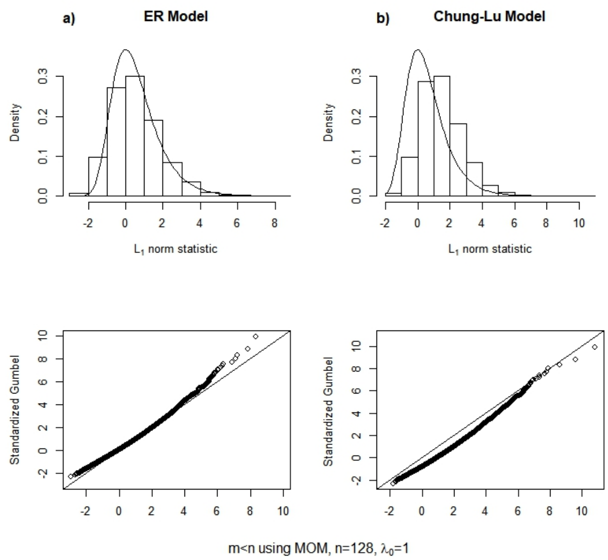

Figure 30: Top figures are histogram density plots of the L1 norm statistics based on 10,000 simulations using MOM estimation and $m < n$ with Gumbel distribution overlaid. $n = 128$ and $\lambda_0 = 1$. Bottom figures are the corresponding Q-Q plots ((a) Erdös-Rényi, and (b) Chung-Lu Models).

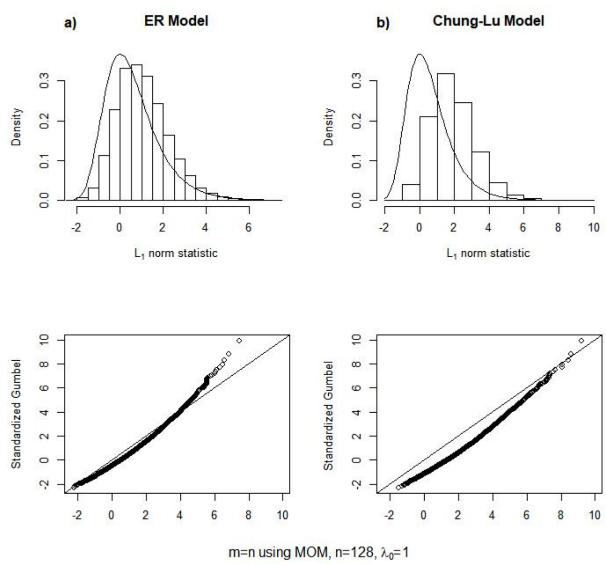

Figure 31: Top figures are histogram density plots of the L1 norm statistics based on 10,000 simulations using MOM estimation and $m = n$ with Gumbel distribution overlaid. $n = 128$ and $\lambda_0 = 1$. Bottom figures are the corresponding Q-Q plots ((a) Erdös-Rényi, and (b) Chung-Lu Models)

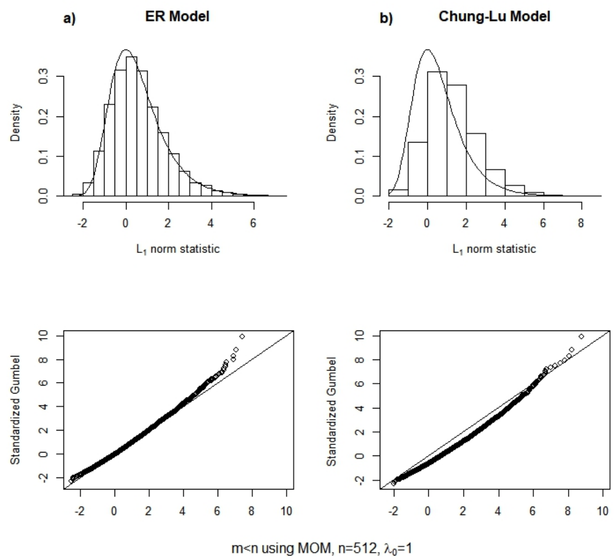

Figure 32: Top figures are histogram density plots of the L1 norm statistics based on 10,000 simulations using MOM estimation and $m < n$ with Gumbel distribution overlaid. $n = 512$ and $\lambda_0 = 1$. Bottom figures are the corresponding Q-Q plots ((a) Erdös-Rényi, and (b) Chung-Lu Models).

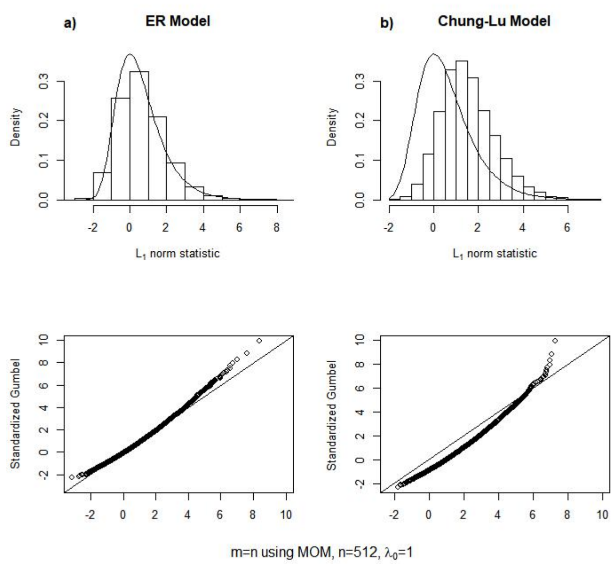

Figure 33: Top figures are histogram density plots of the L1 norm statistics based on 10,000 simulations using MOM estimation and $m = n$ with Gumbel distribution overlaid. $n = 512$ and $\lambda_0 = 1$. Bottom figures are the corresponding Q-Q plots ((a) Erdös-Rényi, and (b) Chung-Lu Models)

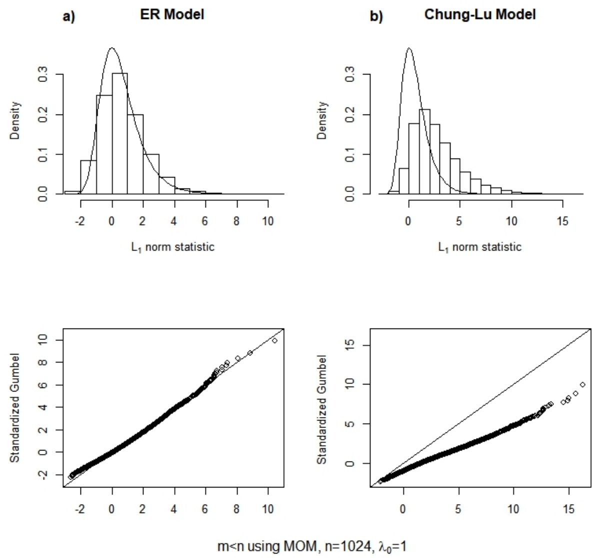

Figure 34: Top figures are histogram density plots of the L1 norm statistics based on 10,000 simulations using MOM estimation and $m < n$ with Gumbel distribution overlaid. $n = 1024$ and $\lambda_0 = 1$. Bottom figures are the corresponding Q-Q plots ((a) Erdös-Rényi, and (b) Chung-Lu Models).

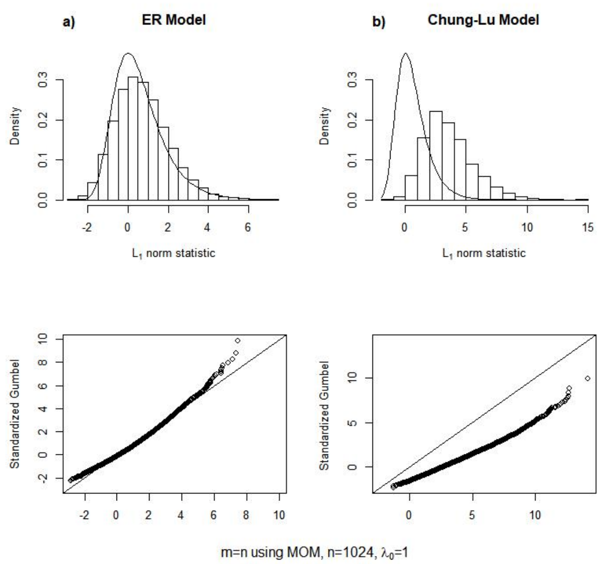

Figure 35: Top figures are histogram density plots of the L1 norm statistics based on 10,000 simulations using MOM estimation and $m = n$ with Gumbel distribution overlaid. $n = 1024$ and $\lambda_0 = 1$. Bottom figures are the corresponding Q-Q plots ((a) Erdös-Rényi, and (b) Chung-Lu Models)

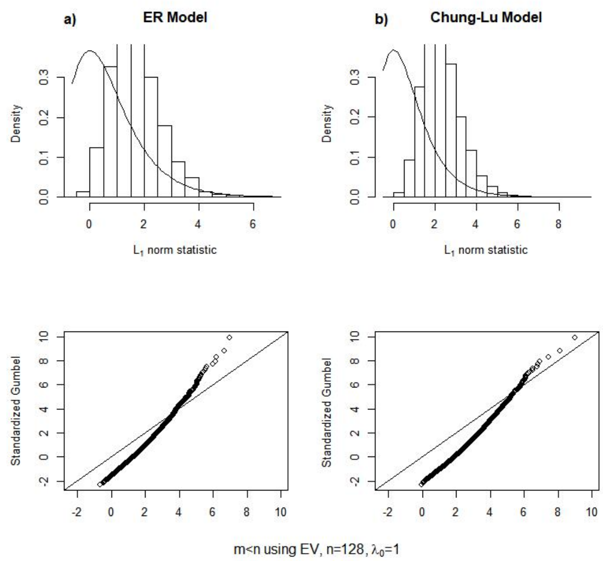

Figure 36: Top figures are histogram density plots of the L1 norm statistics based on 10,000 simulations using the Extreme Value Theorem and $m < n$ with Gumbel distribution overlaid. $n = 128$ and $\lambda_0 = 1$. Bottom figures are the corresponding Q-Q plots ((a) Erdös-Rényi, and (b) Chung-Lu Models).

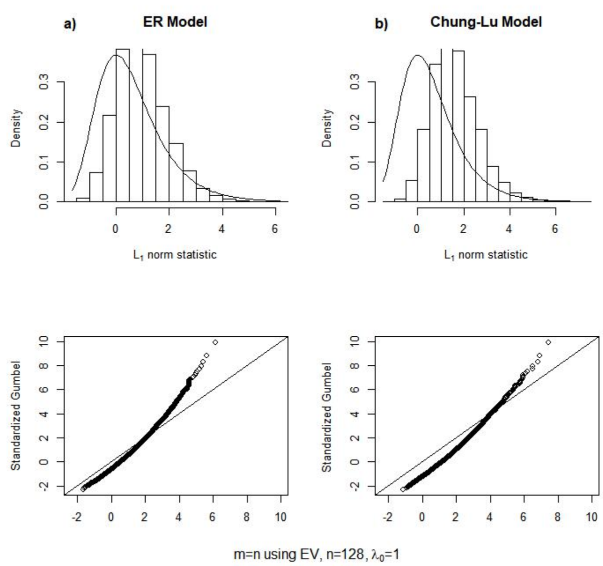

Figure 37: Top figures are histogram density plots of the L1 norm statistics based on 10,000 simulations using the Extreme Value Theorem and $m = n$ with Gumbel distribution overlaid. $n = 128$ and $\lambda_0 = 1$. Bottom figures are the corresponding Q-Q plots ((a) Erdös-Rényi, and (b) Chung-Lu Models)

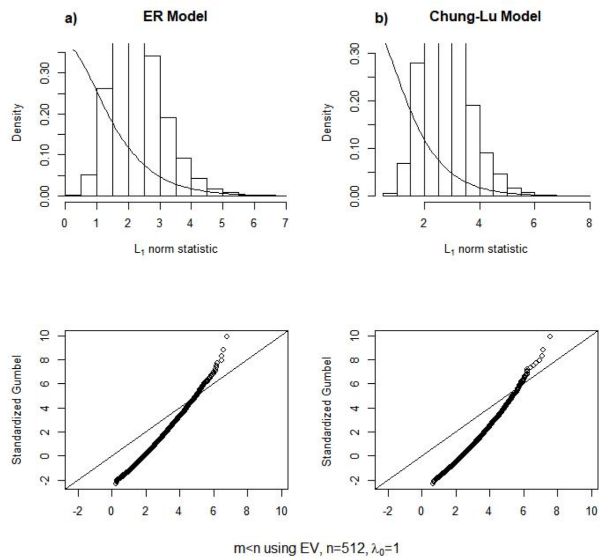

Figure 38: Top figures are histogram density plots of the L1 norm statistics based on 10,000 simulations using the Extreme Value Theorem and $m < n$ with Gumbel distribution overlaid. $n = 512$ and $\lambda_0 = 1$. Bottom figures are the corresponding Q-Q plots ((a) Erdös-Rényi, and (b) Chung-Lu Models).

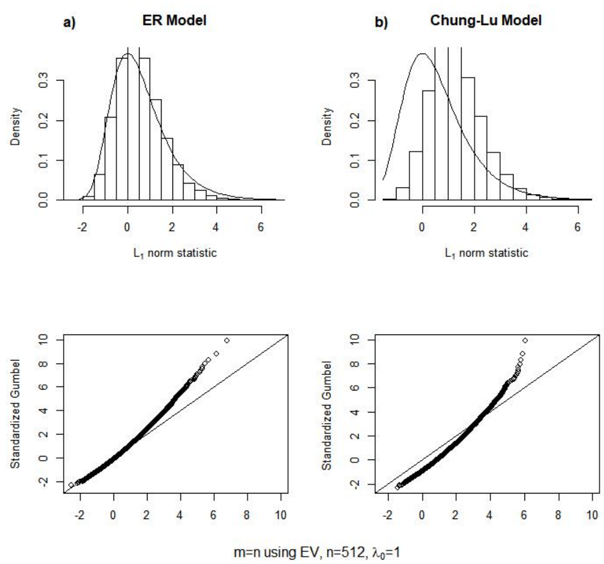

Figure 39: Top figures are histogram density plots of the L1 norm statistics based on 10,000 simulations using the Extreme Value Theorem and $m = n$ with Gumbel distribution overlaid. $n = 512$ and $\lambda_0 = 1$. Bottom figures are the corresponding Q-Q plots ((a) Erdös-Rényi, and (b) Chung-Lu Models)

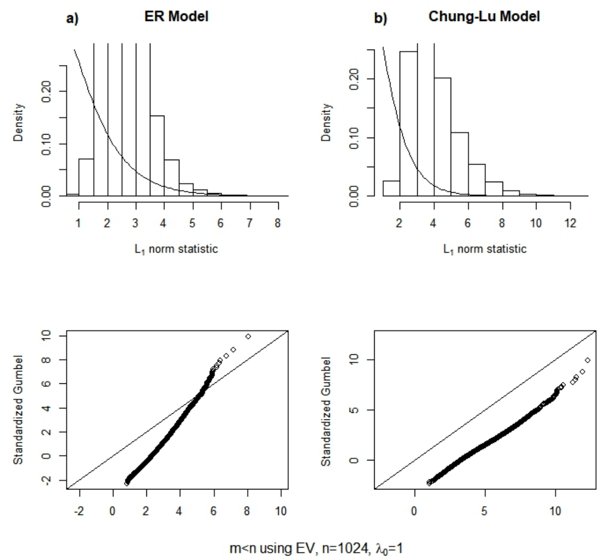

Figure 40: Top figures are histogram density plots of the L1 norm statistics based on 10,000 simulations using the Extreme Value Theorem and $m < n$ with Gumbel distribution overlaid. $n = 1024$ and $\lambda_0 = 1$. Bottom figures are the corresponding Q-Q plots ((a) Erdös-Rényi, and (b) Chung-Lu Models).

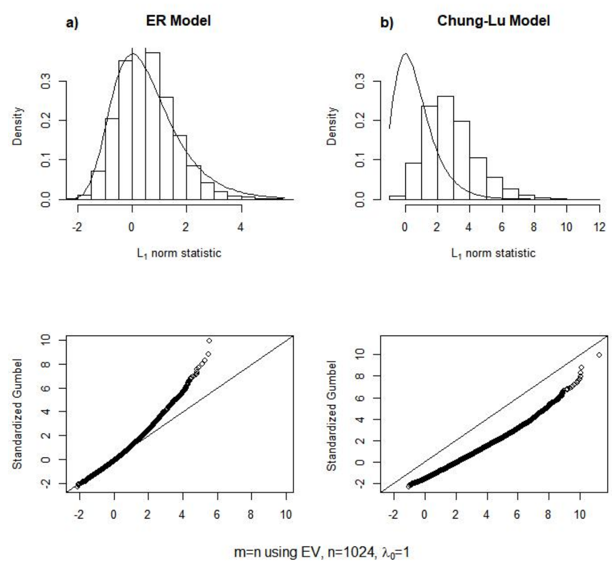

Figure 41: Top figures are histogram density plots of the L1 norm statistics based on 10,000 simulations using the Extreme Value Theorem and $m = n$ with Gumbel distribution overlaid. $n = 1024$ and $\lambda_0 = 1$. Bottom figures are the corresponding Q-Q plots ((a) Erdös-Rényi, and (b) Chung-Lu Models)

4.2. Evaluating the performance of the chi-square and L_1 norm algorithms in count networks

False alarms rates

Table 5: Comparison of false alarm rates for the three cases, χ^2 algorithm, L_1 norm using Extreme Value Theorem and $m < n$, and L_1 norm using Extreme Value Theorem and $m = n$. 95^{th} percentile was used for signaling threshold.

Network order	chi-square			L_1 norm $m < n$			L_1 norm $m = n$		
ER model	$\lambda_0 = 0.2$	$\lambda_0 = 1$	$\lambda_0 = 3$	$\lambda_0 = 0.2$	$\lambda_0 = 1$	$\lambda_0 = 3$	$\lambda_0 = 0.2$	$\lambda_0 = 1$	$\lambda_0 = 3$
128	0.11	0.06	0.08	0.04	0.02	0.04	0.06	0.04	0.05
256	0.10	0.06	0.07	0.04	0.02	0.01	0.08	0.07	0.04
512	0.08	0.05	0.08	0.01	0.02	0.00	0.06	0.04	0.04
1024	0.07	0.07	0.11	0.01	0.01	0.01	0.04	0.06	0.04
Chung-Lu model	$\eta_0 = 0.133$	$\eta_0 = 0.333$	$\eta_0 = 1$	$\eta_0 = 0.133$	$\eta_0 = 0.333$	$\eta_0 = 1$	$\eta = 0.133$	$\eta_0 = 0.333$	$\eta_0 = 1$
128	0.36	0.35	0.47	0.00	0.04	0.01	0.09	0.06	0.04
256	0.55	0.63	0.28	0.03	0.16	0.01	0.09	0.16	0.02
512	0.45	0.46	0.65	0.01	0.16	0.00	0.03	0.19	0.00
1024	0.66	0.70	0.53	0.01	0.01	0.08	0.03	0.01	0.25

Detection rates for the ER model

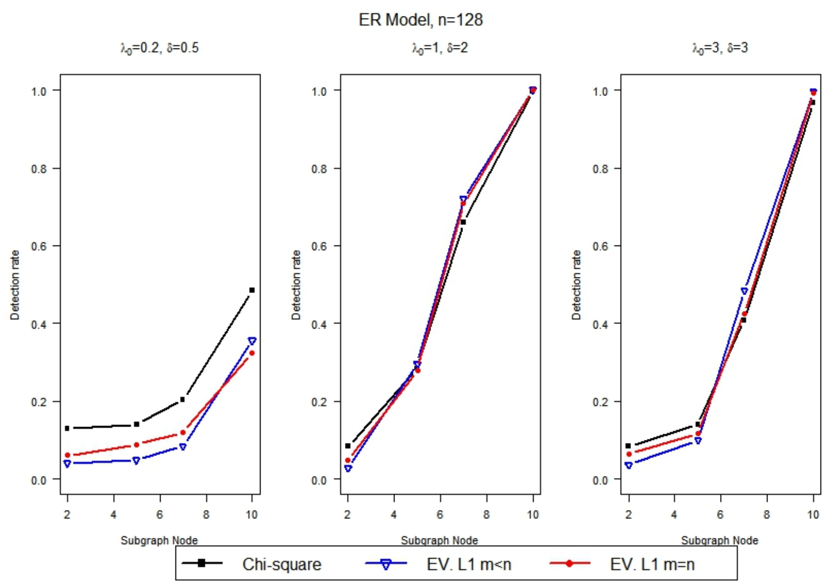

Figure 42: Detection rates for count networks with $n = 128$. Number of anomalous subgraph varies from 2% to 10% of $n = 128$. (Erdös-Rényi Model)

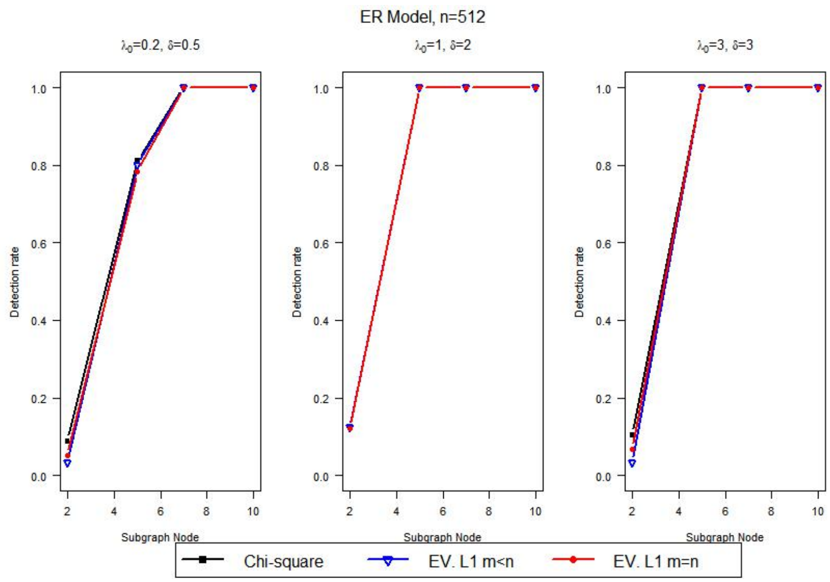

Figure 43: Detection rates for count networks with $n = 512$. Number of anomalous subgraph varies from 2% to 10% of $n = 512$. (Erdös-Rényi Model)

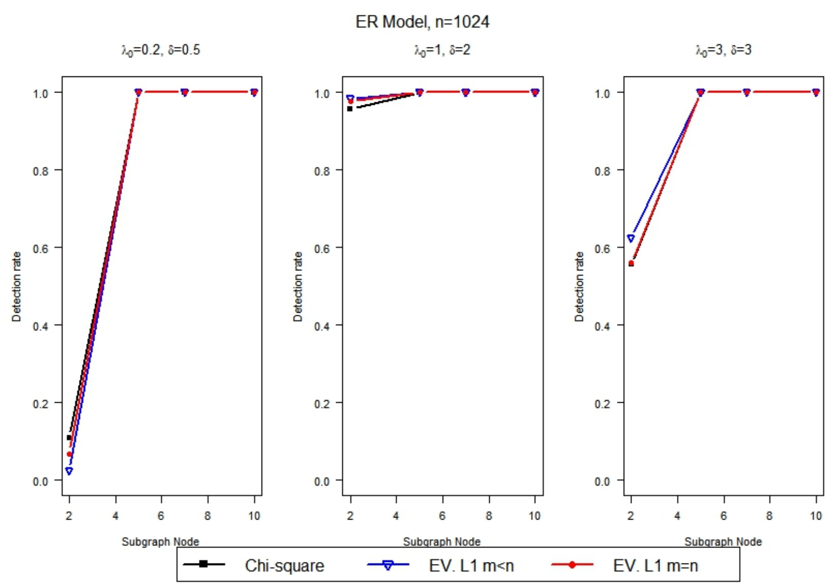

Figure 44: Detection rates for count networks with $n = 1024$. Number of anomalous subgraph varies from 2% to 10% of $n = 1024$. (Erdös-Rényi Model)

Detection rates for the Chung-Lu model

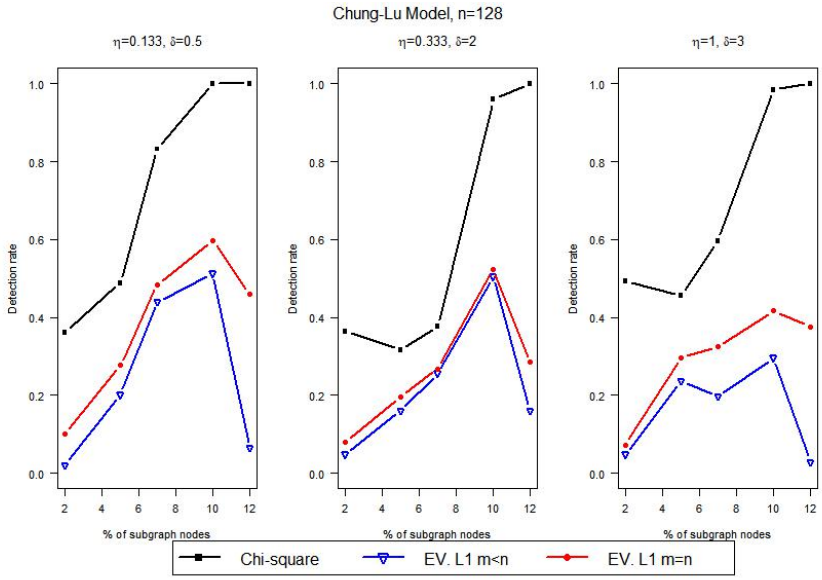

Figure 45: Detection rates for count networks with $n = 128$. Number of anomalous subgraph varies from 2% to 12% of $n = 128$. (Chung-Lu Model)

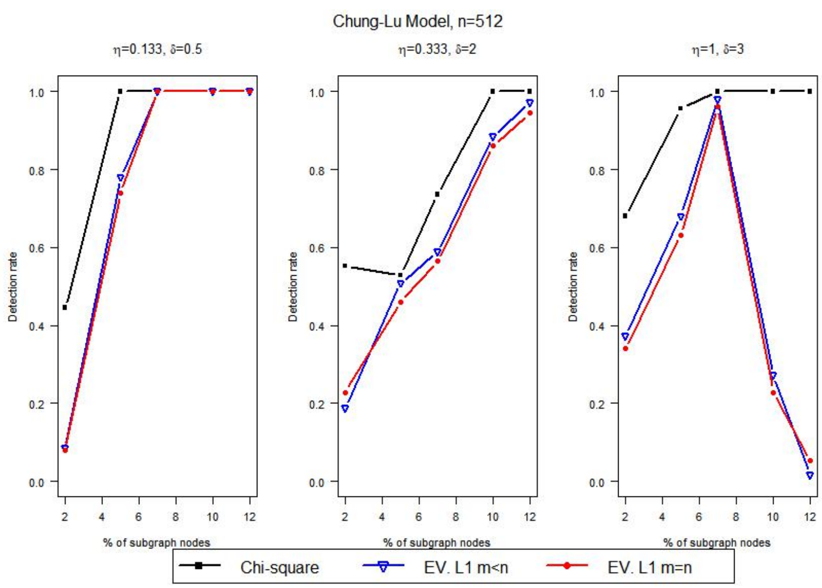

Figure 46: Detection rates for count networks with $n = 512$. Number of anomalous subgraph varies from 2% to 12% of $n = 512$. (Chung-Lu Model)

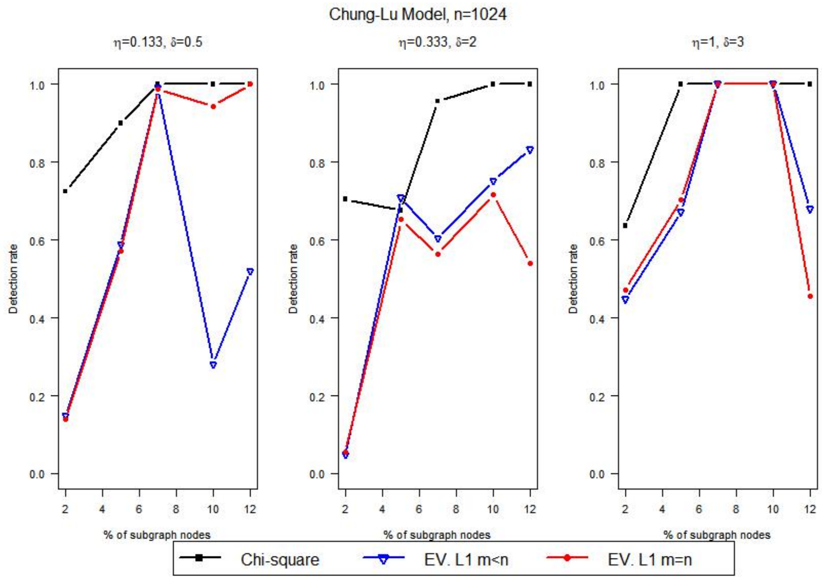

Figure 47: Detection rates for count networks with $n = 1024$. Number of anomalous subgraph varies from 2% to 12% of $n = 1024$. (Chung-Lu Model)