Statistical test for power law

The following table outlines the results of our power law distribution statistical testing in Chapter 4. We first test with the first three hours of our collected retweet data, using the technique described by Clauset [1]. When the p-value is greater than 0.1, the null hypothesis, that the dataset is a power law, is not rejected. Datasets with less than 50 retweets were not processed and are consequently not included in the table.

Dataset	Data points in region	α (MLE)	p-value
1	98	0.841	0.03
2	181	0.913	0.01
4	175	0.975	0.01
5	247	0.886	0.0
6	117	0.888	0.01
8	193	0.664	**0.31**
10	483	0.824	0.0
11	318	0.747	0.01
12	93	0.821	**0.64**
13	106	0.781	0.02
14	270	0.886	0.01
16	248	0.984	0.0
19	499	1.028	0.0
20	628	0.68	0.0
23	562	0.854	0.0
24	431	0.917	0.0
31	227	1.309	0.0
32	79	1.366	**0.78**
33	898	0.746	0.0
35	135	0.861	**0.34**
36	162	0.932	0.0
37	154	0.936	**0.16**
38	195	0.857	**0.35**
43	169	0.931	0.01
44	89	1.053	0.01
45	100	0.964	0.0

52	216	1.071	0.01
53	274	1.138	0.0
59	485	1.024	0.0
62	140	1.148	0.01
63	103	0.874	0.08
64	82	0.781	**0.68**
70	1130	0.504	**0.88**
72	127	0.971	**0.13**
73	215	0.699	**0.42**
77	159	0.806	**0.16**
79	179	1.072	0.0
82	424	0.864	**0.11**
84	131	1.033	0.1
85	88	0.876	**0.22**
87	132	0.983	**0.21**
88	301	1.025	0.0
90	83	0.884	**0.14**
91	126	0.863	0.0
93	94	0.997	**0.29**
99	205	0.807	0.0
101	141	0.75	0.0
102	70	0.958	**0.71**
103	265	0.895	**0.53**
105	128	0.98	0.0
106	97	0.983	**0.35**
107	131	0.891	0.03
110	191	0.952	**0.29**
113	166	0.812	**0.48**
115	92	0.95	**0.95**
117	96	0.764	**0.16**
120	115	0.91	**0.42**
123	114	1.194	**0.19**
124	111	0.912	0.0
126	260	0.691	**0.79**
127	632	0.86	0.0

129	114	0.879	0.0
130	104	0.94	0.01
132	155	0.904	0.0
136	113	0.921	0.02
137	106	0.708	0.02
142	221	0.482	0.09
143	109	0.846	**0.11**
144	90	0.841	**0.61**
146	184	0.79	0.06
147	215	0.708	**0.52**
148	91	0.941	0.0
149	200	0.854	0.0
150	103	0.902	0.01
153	197	0.789	0.01
154	463	0.836	**0.16**
155	208	0.98	0.0
156	94	0.983	0.02
158	186	0.82	0.0
160	1320	0.564	0.0
161	182	0.81	0.0
164	313	0.854	**0.12**
165	157	0.84	**0.19**
168	248	0.965	0.01
169	190	0.874	**0.35**
170	104	1.087	0.0
171	119	0.934	**0.86**
173	153	0.671	0.0
174	207	0.805	0.02
176	283	0.721	0.01
177	456	0.889	0.0
178	133	0.912	0.0
180	135	0.971	0.0
181	90	0.749	**0.7**
185	253	0.968	**0.13**
186	98	0.976	0.0

189	354	0.91	0.0
190	477	0.498	**0.34**
193	104	0.944	**0.23**
194	146	1.128	0.0
195	93	1.29	**0.42**
196	72	0.898	**0.17**
203	205	0.867	0.0
205	140	0.842	0.05
206	110	0.711	**0.46**
207	256	0.929	**0.83**
208	383	0.446	**0.59**
214	227	0.624	**0.24**
215	130	0.835	**0.16**
218	107	0.928	**0.86**
219	318	0.251	0.01
220	361	0.601	**0.97**
224	282	0.826	0.0
227	85	0.746	**0.9**
229	113	0.974	0.01
231	316	0.896	0.0
233	104	0.935	0.0
234	87	0.848	0.03
235	201	1.039	0.0
237	247	1.015	0.0
238	136	0.826	**0.14**
239	109	0.824	0.0
242	159	0.795	0.0
243	415	0.952	0.0
246	91	0.853	**0.81**
247	116	0.945	0.01
249	107	1.035	**0.12**
252	118	0.84	**0.16**
253	375	0.929	0.0
255	176	0.96	0.0
256	71	1.043	**0.44**

257	92	1.02	0.09
258	75	0.975	**0.15**
259	673	0.812	0.07
262	144	0.871	**0.26**
263	113	0.951	0.0
273	445	0.998	0.0
274	292	0.803	0.0
279	135	1.078	0.04
280	147	1.013	0.01
282	235	0.956	**0.29**
283	93	0.765	**0.36**
284	214	0.829	**0.72**
285	132	0.923	0.02
286	185	0.81	**0.98**
288	105	0.708	**0.62**
289	140	0.804	**0.34**
292	126	0.851	0.06
293	148	0.637	**0.17**
296	105	0.984	**0.29**
297	323	0.679	0.0
298	138	0.801	**0.2**
301	160	0.81	**0.35**
302	649	0.687	0.0
304	150	0.913	0.0
307	307	0.788	**0.18**
308	446	0.748	0.03

Of the 157 datasets tested, 66 (42.0%) passed the statistical test for power law.

The following table outlines the same power law test with one hour of data.

Dataset	Data points in region	α (MLE)	p-value
1	80	0.616	**0.21**
2	150	0.73	**0.39**
4	146	0.843	0.1
5	200	0.74	**0.93**

8	119	0.713	**0.22**
10	365	0.702	**0.32**
11	219	0.672	**0.65**
12	68	0.748	**0.71**
13	77	0.633	**0.68**
14	212	0.763	**0.21**
16	210	0.839	0.0
19	433	0.878	0.0
20	384	0.702	**0.65**
23	433	0.723	0.03
24	340	0.82	0.1
28	113	1.256	0.0
31	206	1.28	0.0
32	73	1.34	**0.63**
33	625	0.67	0.0
35	98	0.861	0.1
36	134	0.764	0.0
37	125	0.809	**0.45**
38	144	0.796	**0.7**
43	135	0.819	**0.56**
44	83	0.805	**0.85**
45	82	0.838	0.0
52	189	0.941	**0.21**
53	237	1.082	**0.11**
59	447	0.769	0.0
62	122	1.086	0.01
63	80	0.787	**0.36**
64	59	0.699	**0.62**
70	620	0.5	**0.97**
72	97	1.0	0.04
73	141	0.688	**0.74**
77	110	0.787	**0.2**
79	161	0.899	0.0
82	316	0.803	**0.43**
84	110	0.951	**0.37**

85	64	0.855	0.03
87	106	0.935	**0.2**
88	244	0.995	0.0
90	62	0.827	0.03
91	101	0.694	**0.77**
93	71	1.088	**0.93**
99	147	0.797	0.01
102	55	0.946	**0.5**
103	203	0.851	**0.59**
105	105	0.875	0.06
106	79	0.89	**0.67**
107	104	0.752	0.06
110	150	0.901	**0.37**
113	122	0.759	**1.0**
115	70	0.967	**0.85**
117	72	0.616	**0.67**
120	89	0.863	**0.45**
123	102	1.143	**0.32**
124	92	0.727	**0.64**
125	95	1.112	0.02
126	168	0.691	**0.57**
127	430	1.017	0.01
129	96	0.627	**0.11**
130	83	0.848	0.05
132	124	0.803	**0.11**
136	90	0.815	**0.11**
137	75	0.603	**0.18**
142	112	0.6	0.01
143	81	0.756	**0.35**
144	65	0.821	**0.89**
146	134	0.678	**0.53**
147	133	0.782	0.08
148	80	0.7	0.01
149	157	0.701	0.01
150	82	0.76	**0.13**

153	150	0.594	0.01
154	339	0.776	**0.31**
155	174	0.846	**0.24**
156	84	0.745	**0.6**
158	142	0.686	**0.17**
160	806	0.463	0.01
161	139	0.648	**0.11**
164	236	0.778	**0.8**
165	115	0.787	**0.24**
168	201	0.865	**0.25**
169	135	0.921	0.02
170	94	0.907	0.0
171	94	0.86	**0.91**
173	104	0.536	0.01
174	152	0.709	**0.11**
176	203	0.592	**0.44**
177	363	0.746	**0.11**
178	105	0.788	**0.12**
180	112	0.856	0.03
181	63	0.673	**0.8**
185	190	1.005	0.01
189	291	0.731	0.0
190	250	0.558	**0.19**
193	81	0.893	**0.11**
194	123	1.087	0.01
196	60	0.707	0.08
203	160	0.73	**0.13**
205	106	0.746	0.05
206	74	0.689	0.1
207	200	0.888	**0.79**
208	202	0.401	**0.27**
214	137	0.663	**0.79**
215	98	0.75	**0.11**
218	82	0.917	**0.74**
219	135	0.31	**0.76**

220	209	0.649	**0.96**
224	210	0.72	0.0
227	57	0.714	**0.84**
229	96	0.802	**0.6**
231	253	0.747	0.01
233	86	0.78	0.03
234	71	0.6	**0.62**
235	173	0.91	0.06
237	209	0.882	**0.13**
238	97	0.813	**0.15**
239	83	0.692	0.03
243	332	0.856	0.0
246	65	0.864	**0.58**
247	89	0.919	0.02
249	87	0.998	**0.15**
252	84	0.799	**0.32**
253	315	0.76	0.01
255	148	0.783	**0.81**
256	60	0.972	**0.54**
257	76	0.943	**0.31**
258	63	0.827	**0.68**
259	486	0.761	**0.22**
262	111	0.783	**0.63**
263	97	0.762	0.01
273	369	0.894	0.01
274	185	0.934	**0.38**
279	117	0.966	**0.57**
280	123	0.908	0.07
282	183	0.93	**0.2**
283	57	0.907	**0.35**
284	152	0.83	**0.4**
285	109	0.749	0.1
286	127	0.844	**0.85**
288	68	0.704	**0.33**
289	105	0.691	**0.53**

292	96	0.735	**0.2**
293	93	0.564	**0.11**
296	80	1.023	**0.15**
297	227	0.477	0.0
298	94	0.783	**0.32**
301	109	0.871	0.06
302	453	0.517	0.06
304	118	0.827	0.0
305	115	0.844	**0.71**
307	210	0.823	**0.2**
308	304	0.688	**0.71**
310	207	0.691	0.0
313	95	0.722	**0.11**
314	153	0.707	0.02

Of the 158 datasets tested, 100 (63.3%) passed the statistical test for power law. This is significantly more than the 42.0% for the three hour time window.

References

[1] Aaron Clauset, Cosma Rohilla Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Rev.*, 51(4):661–703, November 2009.