Complex and Social Networks
Report 4
Non-linear regression on dependency trees

Sara Montese

UPC - FIB
08/11/2023

1 Introduction

Global syntactic dependency trees represent the hierarchical relationships between words in sentences across multiple languages, encapsulating the syntax and semantics that underlie human communication. In this study, we investigate the significance of network metrics in global syntactic dependency trees to unveil perspectives on language and network theory. Specifically, we will examine the scaling of mean length $< d >$ of the edges as a function of the number of vertices n. Within global syntactic dependency networks, vertices correspond to words and connections between them reflect instances of syntactic dependencies that have been observed in the context of dependency treebanks, as outlined by Ferrer-i Cancho et al [1].
The data include the description of the global syntactic dependency graphs from ten distinct languages, namely Arabic, Basque, Catalan, Chinese, Czech, English, Greek, Hungarian, Italian and Turkish. Each language adds its distinctive linguistic characteristics to enrich our comprehensive study.
We considered following ensemble of models:

- $f(n) = n/3 + 1/3$ (Model 0),
- $f(n) = (n/2)^b$ (Model 1),
- $f(n) = an^b$ (Model 2), a power-law model,
- $f(n) = ae^{cn}$ (Model 3), an exponential model,
- $f(n) = a \log n$ (Model 4), a logarithmic model,
- $f(n) = an^b e^{cn}$ (Model 5, a generalization of model 2).

Additionally, we analysed a generalisation of given models, respectively:

- $f(n) = (n/2)^b + d$ (Model 1+),
- $f(n) = an^b + d$ (Model 2+),
- $f(n) = ae^{cn} + d$ (Model 3+),
- $f(n) = a \log n + d$ (Model 4+),
- $f(n) = an^b e^{cn} + d$ (Model 5+).

2 Results

The summary of relevant statistics from the syntactic dependency trees data is shown in Table 1. We generate initial plots showing the relationships between the number of vertices and mean length of edges, both in normal and log scales. Furthermore, we plot the relations between the number of vertices and their corresponding aggregated mean values. As an example, we present a sample

Language	N	μ_n	σ_n	μ_d	σ_d
Arabic	4099	26.997072	20.654730	2.167945	0.9315586
Basque	2932	11.336289	6.529215	1.961376	0.6914307
Catalan	15052	25.572814	13.618484	2.318709	0.7023050
Chinese	54198	6.246854	3.310788	1.444775	0.4839823
Czech	25015	16.434020	10.724101	2.020050	0.8750689
English	18779	24.046222	11.223216	3.050584	0.8951359
Greek	2951	22.820400	14.381896	2.201016	0.8129454
Hungarian	6424	21.659869	12.566434	3.877771	1.7804100
Italian	4142	18.411154	13.347354	1.972319	0.7683557
Turkish	6016	11.104222	8.290263	1.842759	0.8203996

Table 1: Summary of the properties of the degree sequences. N is the sample size (the number of sentences or dependency trees), μ_n and σ_n are, respectively, the mean and the standard deviation of number n of vertices of a trees, μ_d and σ_d are the mean and the standard deviation of the mean length of edges $< d >$.

plot for the Arabic language (Fig. 1). The preliminary plots for other languages are available in the appendix section.

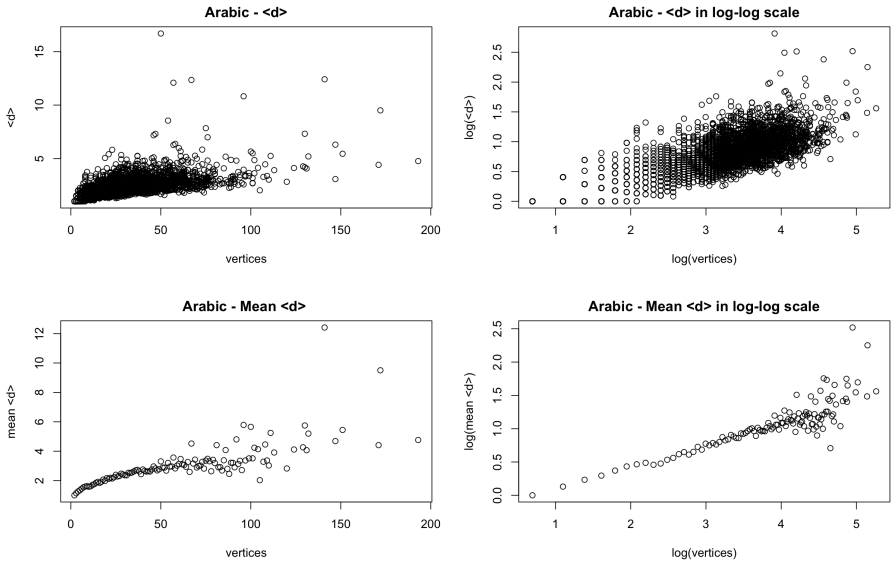


Figure 1: Preliminary plots for Arabic language

After fitting each model in the ensemble, we generate tables presenting essential metrics for model selection, including *AIC* (Akaike information criterion),

s (Residual Standard Error), and ΔAIC. The column names align with the models discussed in Section 1. Table 2 contains the Residual Standard Error (s) for each language and model. This metric is used to measure the goodness of fit of a regression model to the data. In simple terms, it measures the standard deviation of the residuals in a regression model. A regression model that has a small residual standard error will have data points that are closely packed around the fitted regression line.

					s						
Language	**0**	**1**	**2**	**3**	**4**	**5**	**1+**	**2+**	**3+**	**4+**	**5+**
Arabic	22.60	1.02	0.99	0.97	1.09	0.96	1.01	0.97	0.96	1.07	0.96
Basque	6.38	0.48	0.48	0.53	0.52	0.48	0.48	0.48	0.51	0.52	0.49
Catalan	16.62	0.42	0.42	0.45	0.48	0.40	0.42	0.41	0.43	0.47	0.41
Chinese	6.22	1.05	1.02	0.96	1.12	0.97	1.04	0.99	0.96	1.11	0.97
Czech	14.60	4.50	4.18	4.44	4.79	4.04	4.32	4.20	4.26	4.64	3.83
English	14.64	0.83	0.83	0.95	0.95	0.83	0.83	0.83	0.88	0.91	0.84
Greek	14.96	0.86	0.86	0.90	0.89	0.87	0.86	0.87	0.89	0.89	0.87
Hungarian	10.38	0.83	0.83	1.09	1.30	0.84	0.83	0.84	0.98	1.04	0.85
Italian	15.18	0.77	0.74	0.78	0.84	0.73	0.75	0.73	0.74	0.82	0.73
Turkish	8.88	0.38	0.38	0.50	0.40	0.38	0.38	0.38	0.46	0.39	0.38

Table 2: Residual Standard Error (s) for the models under evaluation

To penalize larger models heavily, a common model selection criteria is the Akaike Information Criterion (AIC)[2]. For each language and model, the AIC index obtained by minimizing the error between the model and the empirical data is shown in Table 3.

					AIC						
Language	**0**	**1**	**2**	**3**	**4**	**5**	**1+**	**2+**	**3+**	**4+**	**5+**
Arabic	1090.8	348.8	343.6	337.5	364.5	335.3	346.1	337.0	336.6	361.5	337.3
Basque	276.9	60.0	60.9	70.6	67.9	62.7	61.0	62.7	67.1	68.1	64.7
Catalan	814.1	110.8	108.3	121.0	135.3	103.2	109.4	106.6	113.8	132.3	105.0
Chinese	274.7	126.2	125.1	119.8	131.6	121.8	126.1	123.3	121.0	131.6	122.8
Czech	723.6	517.3	505.5	516.0	528.5	500.5	511.5	507.2	510.0	524.1	491.9
English	724.1	219.6	221.4	245.2	243.9	222.7	221.5	222.5	233.8	237.5	224.5
Greek	711.4	221.9	222.7	230.1	227.5	224.5	222.9	224.5	228.1	227.8	226.5
Hungarian	610.9	203.2	204.8	248.3	275.5	206.8	204.8	206.8	231.7	240.9	208.8
Italian	705.7	200.0	194.3	203.4	215.2	192.9	196.3	192.8	195.4	211.6	194.7
Turkish	412.8	54.5	56.5	87.7	61.2	55.8	56.4	57.5	79.3	60.5	56.8

Table 3: Akaike Information Criterion (AIC) for the models under evaluation

AIC scores are mostly useful as a relative metric of goodness of fit (relative to other models). For each language, we set the model with the lowest AIC score as *preferred model* and then, for each model i, we compute:

$$\Delta AIC_i = minAIC - AIC_i$$

Results are shown in Table 4.

	ΔAIC										
Language	**0**	**1**	**2**	**3**	**4**	**5**	**1+**	**2+**	**3+**	**4+**	**5+**
Arabic	755.6	13.5	8.3	2.2	29.2	0.0	10.8	1.8	1.4	26.2	2.0
Basque	216.9	0.0	0.9	10.6	7.9	2.7	1.0	2.7	7.1	8.1	4.7
Catalan	710.9	7.6	5.0	17.8	32.1	0.0	6.1	3.3	10.6	29.1	1.8
Chinese	154.9	6.4	5.2	0.0	11.8	2.0	6.3	3.4	1.2	11.8	3.0
Czech	231.7	25.4	13.6	24.1	36.6	8.5	19.5	15.3	18.1	32.2	0.0
English	504.5	0.0	1.8	25.7	24.3	3.1	1.9	2.9	14.3	18.0	4.9
Greek	489.5	0.0	0.9	8.2	5.7	2.7	1.0	2.6	6.3	6.0	4.6
Hungarian	407.7	0.0	1.7	45.1	72.4	3.6	1.6	3.6	28.5	37.8	5.6
Italian	512.9	7.2	1.6	10.7	22.4	0.1	3.5	0.0	2.6	18.8	2.0
Turkish	358.2	0.0	2.0	33.2	6.6	1.3	1.9	3.0	24.8	6.0	2.3

Table 4: AIC difference (ΔAIC) for each model and language

For each language, we display plots of the best-fitting model from the ensemble, illustrating how it aligns with the data.

2.1 Best-fitting models

2.1.1 Arabic

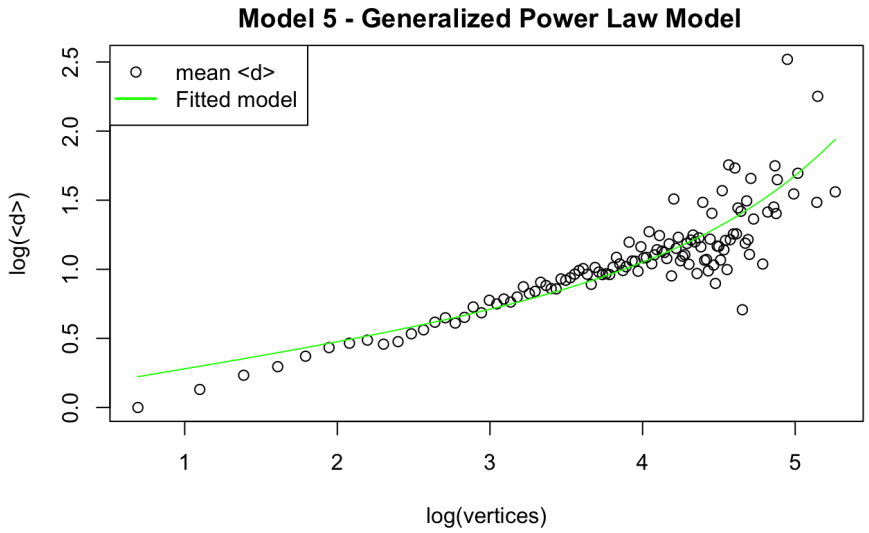


Figure 2: Best model for Arabic language - Model 5

2.1.2 Basque

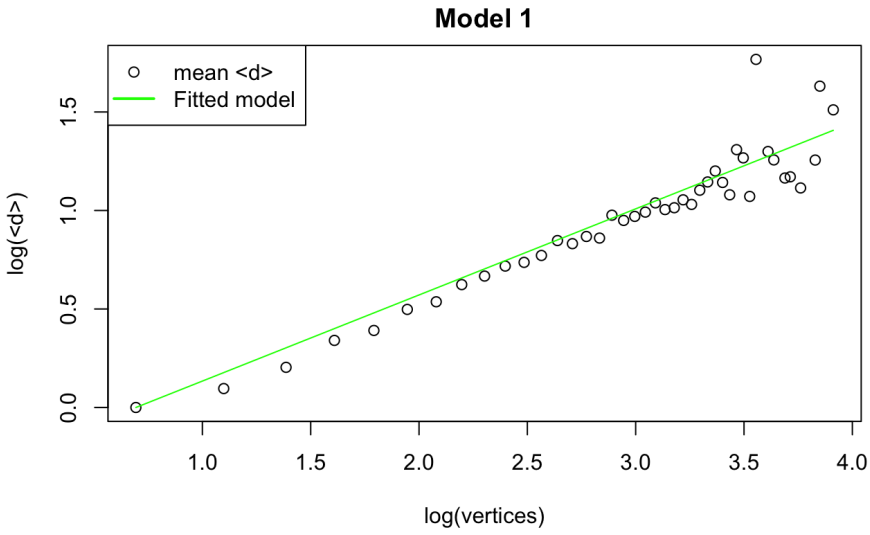


Figure 3: Best model for Basque language - Model 1

2.1.3 Catalan

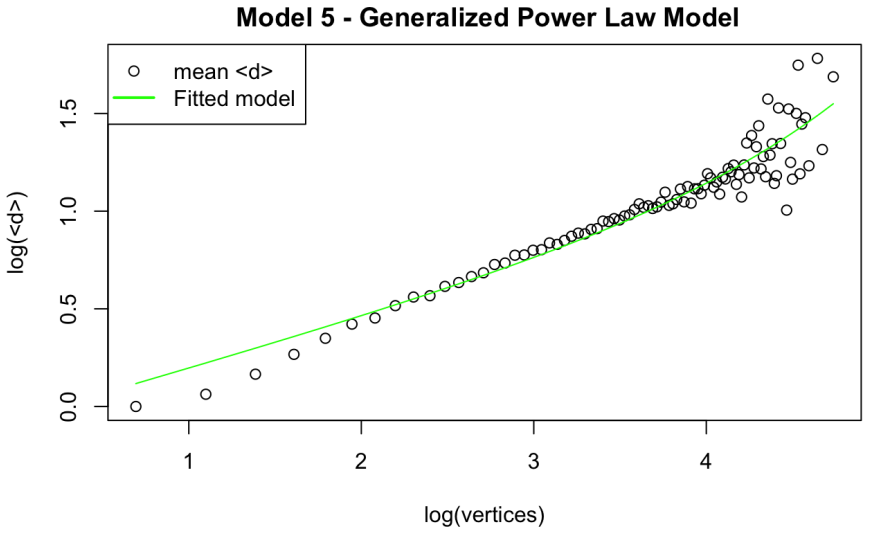


Figure 4: Best model for Catalan language - Model 5

2.1.4 Chinese

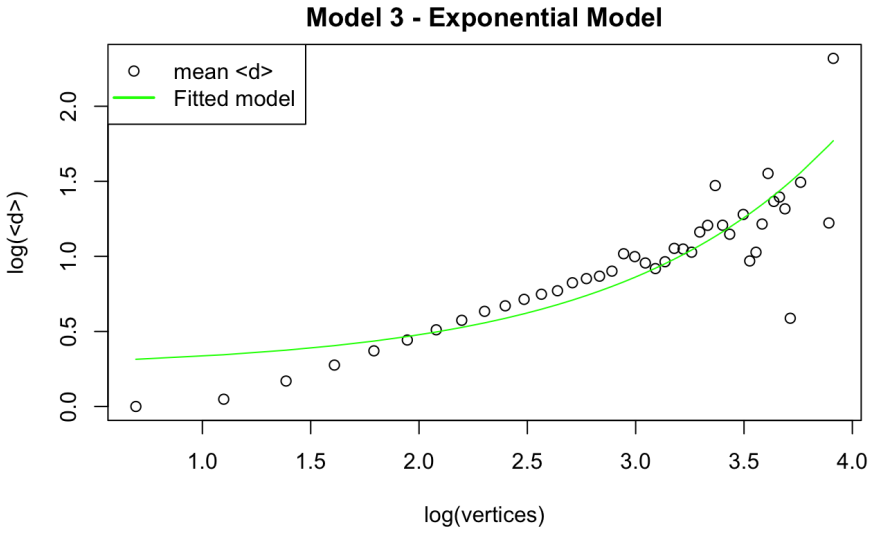


Figure 5: Best model for Chinese language - Model 3

2.1.5 Czech

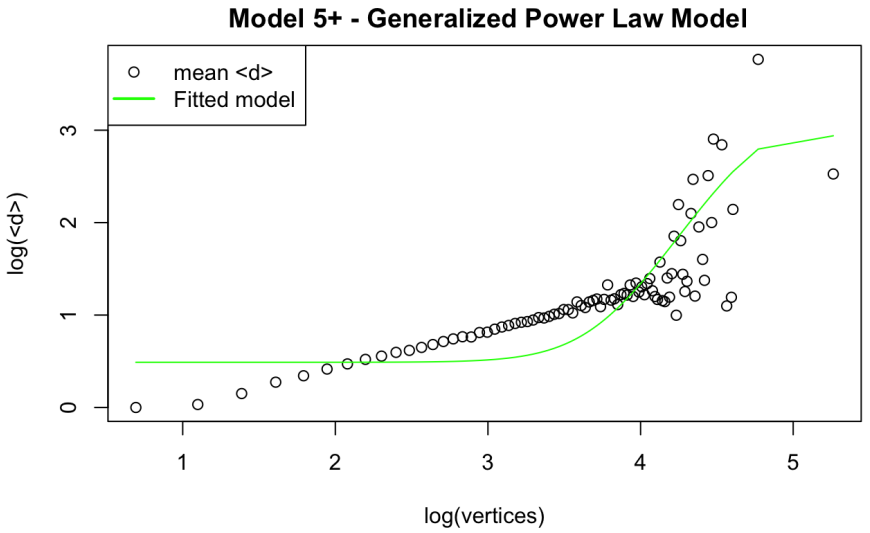


Figure 6: Best model for Czech language - Model 5+

2.1.6 English

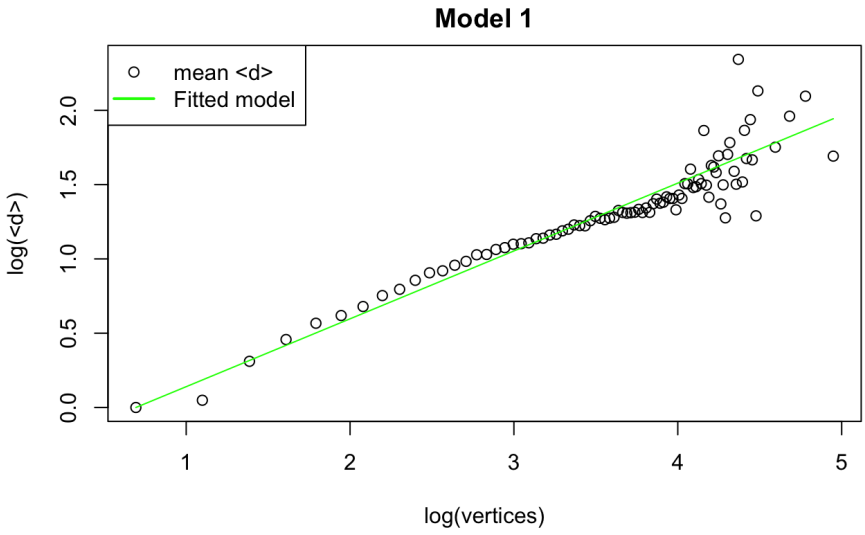


Figure 7: Best model for Basque language - Model 1

2.1.7 Greek

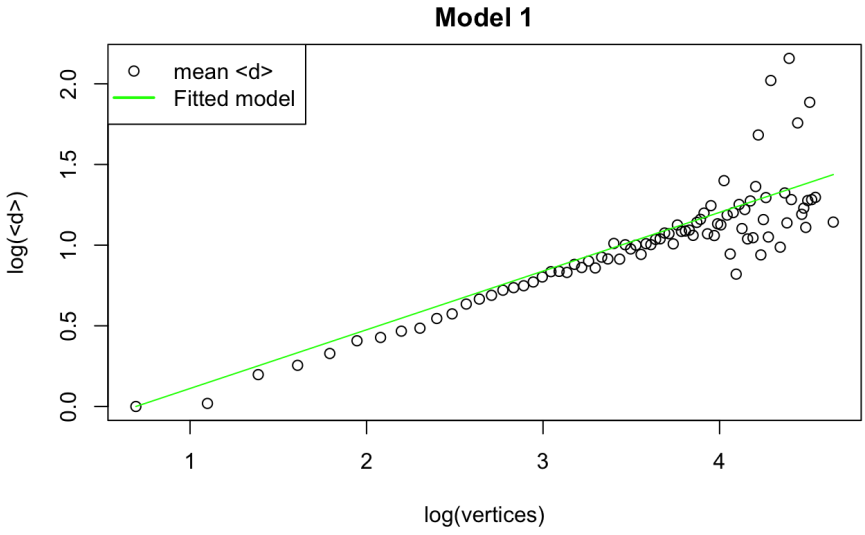


Figure 8: Best model for Greek language - Model 1

2.1.8 Hungarian

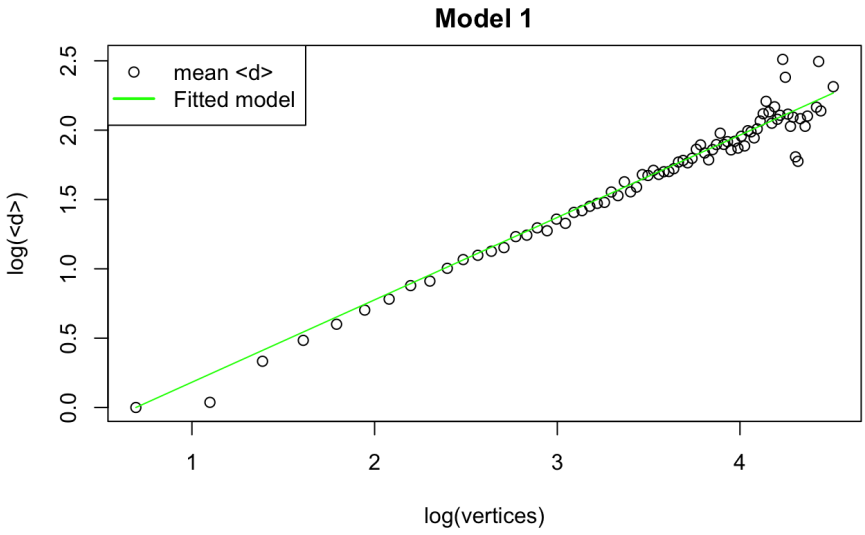


Figure 9: Best model for Hungarian language - Model 1

2.1.9 Italian

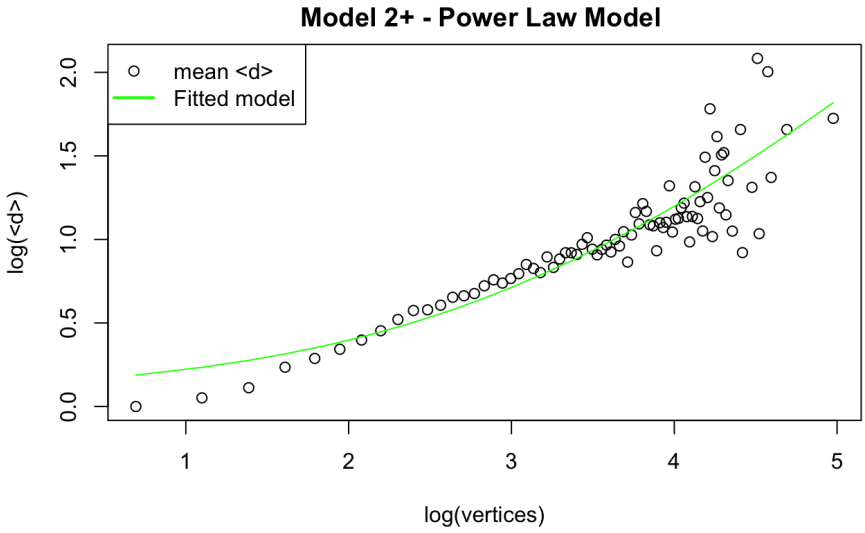


Figure 10: Best model for Italian language - Model 2+

2.1.10 Turkish

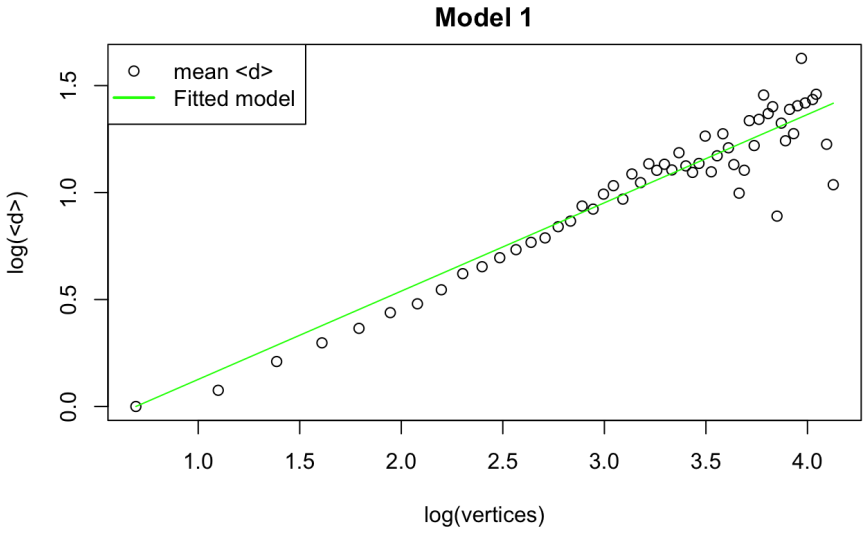


Figure 11: Best model for Turkish language - Model 1

3 Discussion

Regarding homoscedasticity, the absolute residuals plots (Fig. 13) reveal a systematic widening or narrowing of the spread as fitted values increase. This observation strongly suggests the presence of heteroscedasticity in each language and model under analysis. It is worth noting that the absolute residuals of the null model for each language exhibit a substantially smaller order of magnitude compared to the ones of other models. However, it's essential to keep in mind that its low residuals do not indicate better predictive performance. By examining the variance of data points in relation to the number of vertices (Fig.12) and keeping into account the different y-limits of the plots, some discernible pattern emerge. Across several languages, including Arabic, Catalan, Chinese, Czech, English, Greek, Hungarian, Italian, and Turkish, the variability of data points is not consistently constant, indicating the presence of heteroscedasticity. In the case of Basque, although the variance appears almost constant, a subtle pattern is observable. For Hungarian, Greek and Czech, certain points exert disproportionate influence on the variance, suggesting the potential presence of outliers. Further investigation is warranted to better understand and address these outlier points. Therefore, analyzing the absolute residuals reveals a departure from the homoscedasticity assumption across all languages and models. This observation is further supported by the variance analysis. In response to these findings, our approach involves smoothing the original data by adopting

the average mean length of the vertices, as opposed to relying solely on the original mean length. This adjustment is designed to alleviate the identified heteroscedasticity and elevate the overall reliability of our conclusions.
Our attention shifts to the Residual Standard Error (s), and Table 2 provides an overview for each language and model. For each language, the null model (Model 0) shows a high value of s, meaning that it fits the data worse than other models, as expected. The rest of the models have low values of s (typically close to 1 or smaller), indicating better-fitting models. In the ensemble, Czech language stands out by showing slightly higher values of s for non-null models. This suggests that the data points for this language spread more widely around the regression line than others. From Table 2, it is clear that larger models fit better and so have smaller s, but use more parameters. Thus, the best choice of model will balance fit with model size. Table 3 highlights that the models' generalizability may vary across languages. The null model can be seen as a foundational baseline for our analysis. It is evident from the AIC values associated with each language that this basic model is inadequate in effectively capturing the inherent data patterns. Models ranging from 1 to 5 exhibit significantly lower AIC values, and their variants incorporating an additional parameter d in some cases demonstrate more favorable fit to the data, despite the increase of complexity in the model (i.e. Model 5/5+). AIC is less stringent in penalizing model complexity. If we aim the model selection criterion to strongly penalize model complexity and prefer simpler models, an alternative can be the Bayesian Information Criterion (BIC). For Czech language, we notice that the data is less amenable to the chosen ensemble of models, yet leading to higher AIC values compared to other languages. On the other hand, for Basque and Turkish languages, the chosen ensemble of models seems to be more appropriate, given the lower magnitude of the AIC values. From Table 3, we can determine the best-fit model for each language. In Section 2.1, we showcase models demonstrating the best fit to the data. The appendix contains a comprehensive display of all fitted models. Model 1, marked by its simplicity with just one parameter, offers a strong fit for the majority of languages. Additionally, the more complex models 5 and 5+ exhibit a favorable fit. Upon visual inspection of each plot, a consistent observation emerges: in log-log scale, the data reveals a more sparse tail, while the remaining portion typically conforms to a straight line. This pattern holds true across all languages. Specifically, the best models for Arabic, Basque, Catalan, English, Hungarian, Greek, and Italian closely resemble a straight line or a subtle curve, demonstrating a proper fit to the data. However, it's worth noting that they tend to underfit in the final part due to the sparse nature of the data. In the case of Czech, however, the best model doesn't seem to accurately describe the data. On the other hand, Model 3+ visually seems to fit better. For each language, the comprehensive analysis indicates a logarithmic relationship between the mean edge length and the size of the tree.

4 Methods

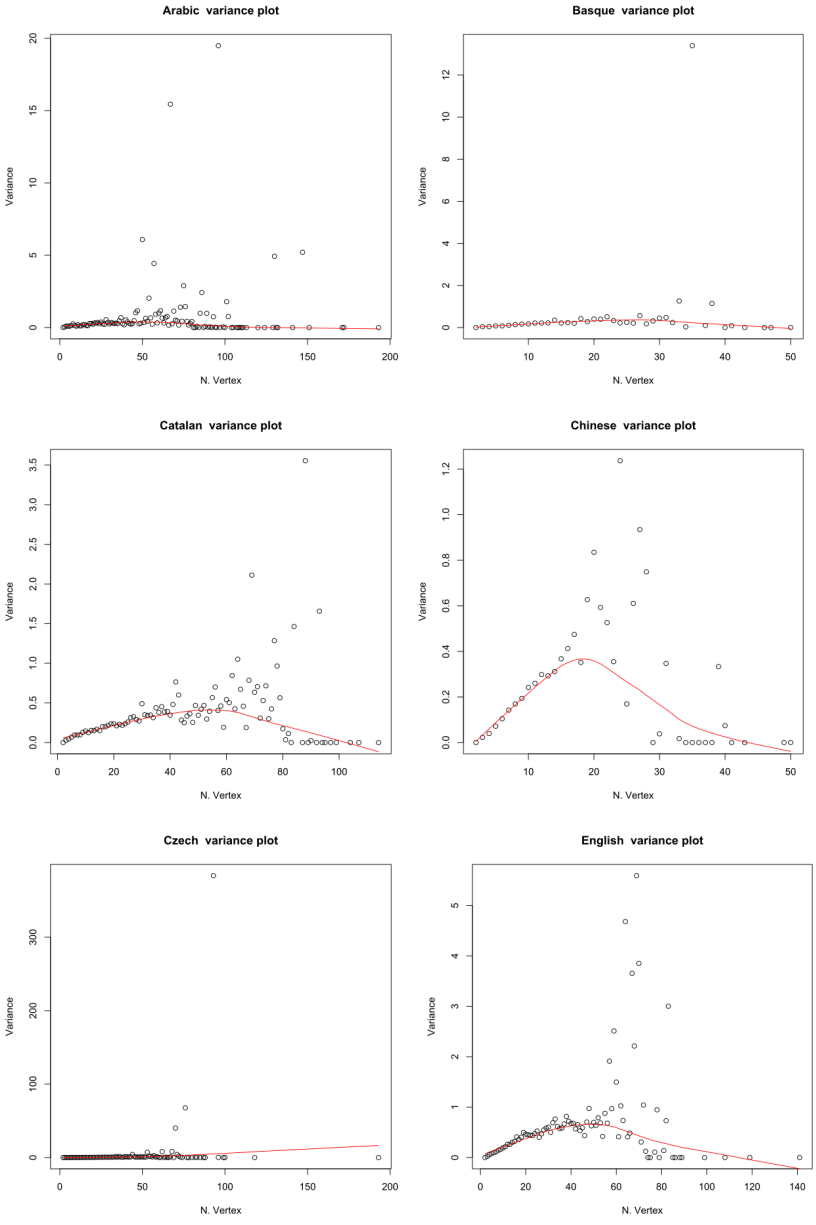
4.1 Homoscedasticity Assessment

The way we invoke nls()[3] for nonlinear regression, in general, assumes homoscedasticity. Evaluating homoscedasticity in the context of nonlinear regression poses challenges due to the inapplicability of standard tests designed for linear regression assumptions. Nevertheless, visual methods offer a viable alternative for assessing homoscedasticity in nonlinear regression. Our approach incorporates the following techniques:

- Variance Analysis: We employ visualizations to show the variance of points concerning the number of vertices in each language dependency tree (Fig. 12). If the variance remains roughly constant across all vertices, it indicates homoscedasticity.
- Residual Analysis: By visualizing plots of absolute residuals versus fitted values (Fig. 13), we conduct a residual analysis. Homoscedasticity is indicated if the spread of absolute residuals remains relatively constant across all levels of fitted values. The analysis encompassed Model 0 through Model 5, given that models from 1+ to 5+ represent a generalization from the preceding ones.

In addition to these visual methods, another technique for assessing homoscedasticity involves conducting a Levene's test, though it was not performed in this instance since visual inspection proved sufficient for our analysis.

Figure 12: Variance of points as a function of the number of vertices of each language



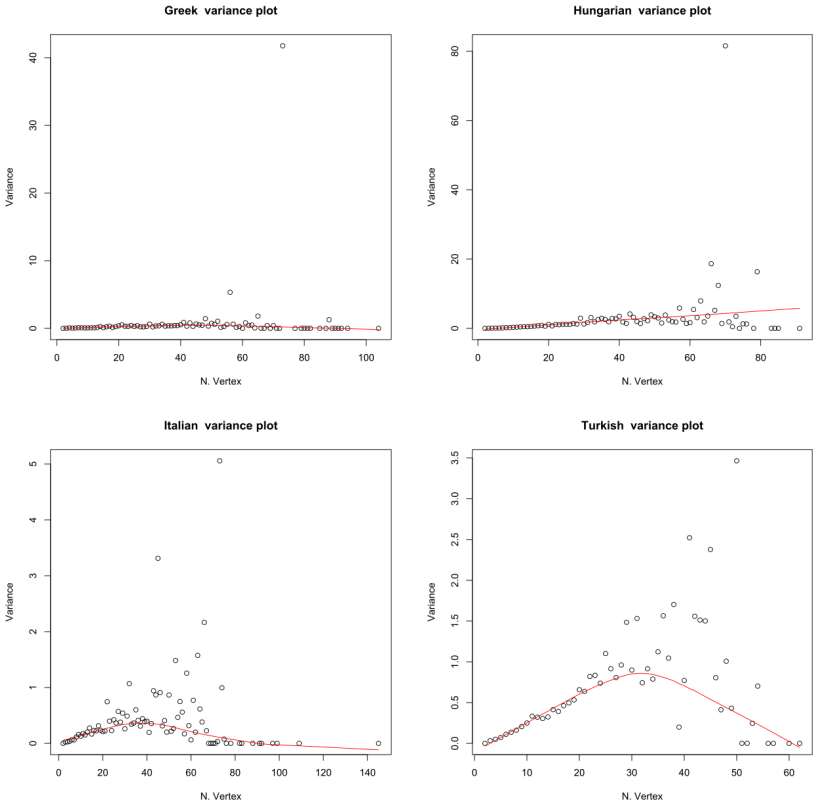
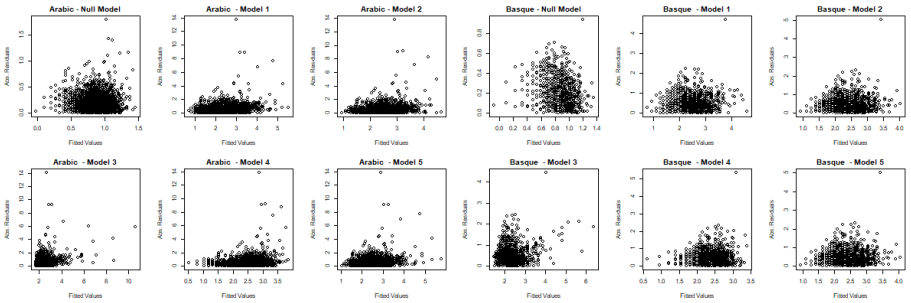
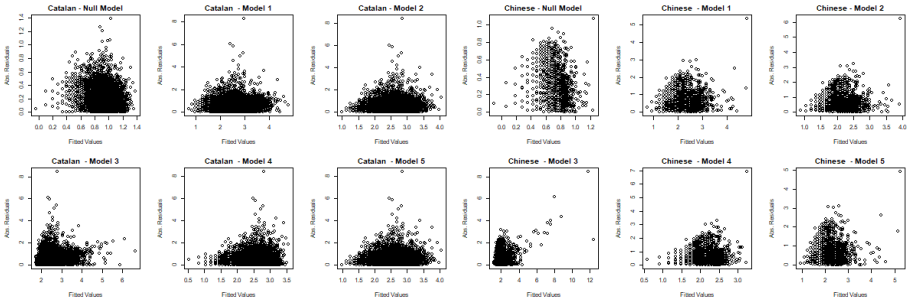

Greek variance plot
Variance
N. Vertex
0
10
20
30
40
0
20
40
60
80
100
Hungarian variance plot
Variance
N. Vertex
0
20
40
60
80
0
20
40
60
80
Italian variance plot
Variance
N. Vertex
0
1
2
3
4
5
0
20
40
60
80
100
120
140
Turkish variance plot
Variance
N. Vertex
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
0
10
20
30
40
50
60

Figure 13: Fitted valued vs Absolute residuals for each model and language



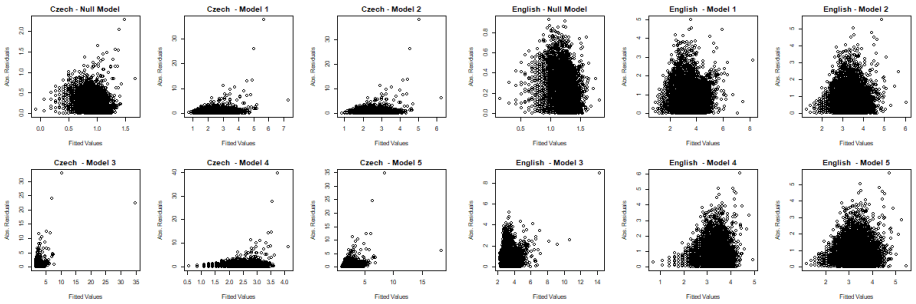
(a) Arabic

(b) Basque



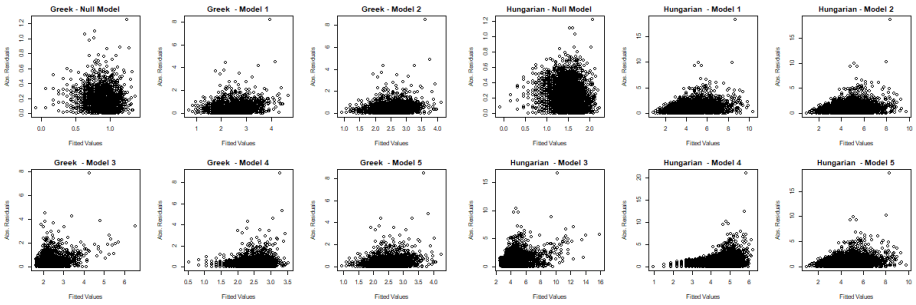
(c) Catalan

(d) Chinese



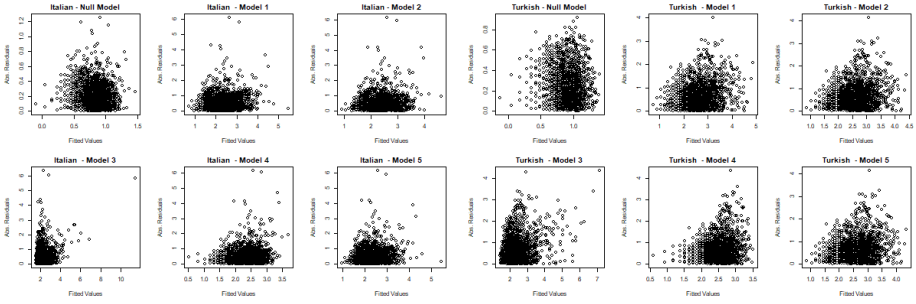
(e) Czech

(f) English



(g) Greek

(h) Hungarian



(i) Italian

(j) Turkish

4.2 Tuning of parameters

For models 1 to 5, the initial parameters are derived from the parameters that best fit their respective linear models, and the optimization algorithm employed is the Gauss-Newton algorithm (*default*). For models 1+ to 5+, the initial parameters are determined based on the best-fit parameters obtained from simplified versions of the nonlinear models, which are models 1 to 5. It's important to note that optimizing models 1+ to 5+ requires a careful selection of the optimization algorithm and the parameter bounds. In such cases, the *port* algorithm is preferred, as the default algorithm may fail to converge or encounter singularities. To ensure optimal convergence, a trial-and-error process was carried out to establish upper and lower bounds for the parameters of models 2+, 3+, and 5+ as follows:

- Model 2+: $-2 \leq p \leq 10$ for each parameter $p \in \{a, b, d\}$
- Model 3+: $-10 \leq p \leq 10$ for each parameter $p \in \{a, c, d\}$
- Model 5+: $-5 \leq p \leq 5$ for each parameter $p \in \{a, b, c, d\}$

Initial and final values of models 1-5 are shown in Tables 5 and 6. Initial and final values of the models 1+-5+ are shown in Tables 7 and 8. Model 0 has no parameters.

	1	2		3		4		5	
Language	b	a	b	a	c	a	a	b	c
Arabic	0.3644179	0.7026980	0.3644179	1.755183	0.0077017	-0.9772333	0.9113379	0.2501367	0.0028339
Basque	0.4621184	0.6681424	0.4621184	1.415740	0.0256485	-0.4116374	0.7012874	0.4283459	0.0021539
Catalan	0.3737014	0.7173181	0.3737014	1.667200	0.0103915	-0.4744520	0.7984216	0.3205660	0.0017326
Chinese	0.4958233	0.6028050	0.4958233	1.317919	0.0285604	-0.9200616	0.7322247	0.3597364	0.0087134
Czech	0.5685368	0.4360963	0.5685368	1.533582	0.0163292	-5.3690887	0.8608720	0.2381483	0.0105487
English	0.4379761	0.7704005	0.4379761	2.110912	0.0118926	-1.2164984	0.8009561	0.4185703	0.0006494
Greek	0.3788344	0.7072904	0.3788344	1.635487	0.0111616	-0.5278215	0.7252466	0.3659615	0.0004542
Hungarian	0.6128427	0.6103065	0.6128427	2.264730	0.0195826	-3.3418391	0.5883283	0.6324287	-0.0007495
Italian	0.4114133	0.6341535	0.4114133	1.584659	0.0119050	-0.8988036	0.7746301	0.3107956	0.0034337
Turkish	0.4326759	0.6959606	0.4326759	1.560304	0.0186894	-0.3452013	0.6285258	0.4959577	-0.0032694

Table 5: Initial values of parameters of nls()[3] for Models 1-5

	1	2		3		4		5	
Language	b	a	b	a	c	a	a	b	c
Arabic	0.3512487	0.4323499	0.4866883	1.872914	0.0072093	0.8161159	1.0972607	0.1736616	0.0048328
Basque	0.4370843	0.6230338	0.4871953	1.599748	0.0216222	0.9509584	0.7015671	0.4245260	0.0029139
Catalan	0.3565486	0.6297550	0.4088009	1.799790	0.0092372	0.8173598	0.9381678	0.2506688	0.0037396
Chinese	0.4593331	0.3632897	0.6631397	1.288822	0.0303259	0.9900542	1.2456681	0.0174740	0.0295642
Czech	0.5248320	0.0443474	1.1699411	2.494960	0.0112172	1.2980523	0.0000042	3.7197036	-0.0222724
English	0.4566473	0.6804634	0.4732807	2.527430	0.0090444	1.1312747	0.8242994	0.4019784	0.0015002
Greek	0.3637845	0.6186193	0.4203148	1.804814	0.0098503	0.8287195	0.7183969	0.3603857	0.0014851
Hungarian	0.5940305	0.6150513	0.6124458	2.947576	0.0147270	1.7149655	0.5770948	0.6373196	-0.0006090
Italian	0.3726801	0.4540470	0.5034952	1.831949	0.0096928	0.8454177	0.7487875	0.3184867	0.0038532
Turkish	0.4127669	0.7352859	0.4186454	1.809590	0.0148177	0.9241477	0.5391134	0.5675458	-0.0056560

Table 6: Final values of parameters of nls()[3] for Models 1-5

	1+		2+			3+			4+		5+			
Language	b	d	a	b	d	a	c	d	a	d	a	b	c	d
Arabic	0.3512487	0	0.4323499	0.4866883	0	1.872914	0.0072093	0	0.8161159	0	1.0972607	0.1736616	0.0048328	0
Basque	0.4370843	0	0.6230338	0.4871953	0	1.599748	0.0216222	0	0.9509584	0	0.7015671	0.4245260	0.0029139	0
Catalan	0.3565486	0	0.6297550	0.4088009	0	1.799790	0.0092372	0	0.8173598	0	0.9381678	0.2506688	0.0037396	0
Chinese	0.4593331	0	0.3632897	0.6631397	0	1.288822	0.0303259	0	0.9900542	0	1.2456681	0.0174740	0.0295642	0
Czech	0.5248320	0	0.0443474	1.1699411	0	2.494960	0.0112172	0	1.2980523	0	0.0000042	3.7197036	-0.0222724	0
English	0.4566473	0	0.6804634	0.4732807	0	2.527430	0.0090444	0	1.1312747	0	0.8242994	0.4019784	0.0015002	0
Greek	0.3637845	0	0.6186193	0.4203148	0	1.804814	0.0098503	0	0.8287195	0	0.7183969	0.3603857	0.0014851	0
Hungarian	0.5940305	0	0.6150513	0.6124458	0	2.947576	0.0147270	0	1.7149655	0	0.5770948	0.6373196	-0.0006090	0
Italian	0.3726801	0	0.4540470	0.5034952	0	1.831949	0.0096928	0	0.8454177	0	0.7487875	0.3184867	0.0038532	0
Turkish	0.4127669	0	0.7352859	0.4186454	0	1.809590	0.0148177	0	0.9241477	0	0.5391134	0.5675458	-0.0056560	0

Table 7: Initial values of parameters of nls()[3] for Models 1+-5+

	1+		2+			3+			4+		5+			
Language	b	d	a	b	d	a	c	d	a	d	a	b	c	d
Arabic	0.3851794	-0.4773642	0.0133009	1.1293945	1.5961888	8.3359400	0.0025422	-6.763678	1.0552890	-0.9772334	1.0833596	0.1754859	0.0048388	0.0146326
Basque	0.4545042	-0.1537018	0.4085778	0.5746192	0.3714965	10.0000000	0.0056072	-8.676988	1.0831010	-0.4116374	0.7443813	0.4098439	0.0031304	-0.0524410
Catalan	0.3729499	-0.1926650	0.1643019	0.6511584	0.9535907	10.0000000	0.0025128	-8.413512	0.9404799	-0.4744520	1.8808676	0.1520208	0.0034538	-1.0320674
Chinese	0.5094033	-0.4900687	0.0023797	1.9112938	1.5847255	0.3353893	0.0544454	1.298515	1.2854609	-0.9200616	0.1638938	0.2674641	0.0471940	1.3723949
Czech	0.6286149	-2.6106650	0.0255161	1.2737787	0.5837535	10.0000000	0.0059261	-9.169251	2.7124553	-5.3690886	0.0000000	5.0000000	-0.0306847	1.6307107
English	0.4598128	-0.0506465	0.3223360	0.6086192	0.7953141	10.0000000	0.0035591	-7.904060	1.4525040	-1.2164984	0.2434365	0.6771541	-0.0006425	0.9497538
Greek	0.3826869	-0.2156273	0.3220508	0.5361624	0.5532834	10.0000000	0.0027353	-8.439830	0.9691032	-0.5278215	0.2453769	0.6037454	-0.0007653	0.6759144
Hungarian	0.5988579	-0.1145586	0.6638207	0.5978169	-0.1228305	10.0000000	0.0067897	-7.697413	2.6202004	-3.3418391	0.4723294	0.6890437	-0.0012344	0.2028433
Turkish	0.4165570	-0.0371493	1.6134574	0.2864992	-1.1892814	10.0000000	0.0041099	-8.438760	1.0265662	-0.3452013	0.1211272	1.0061351	-0.0146602	0.7668771

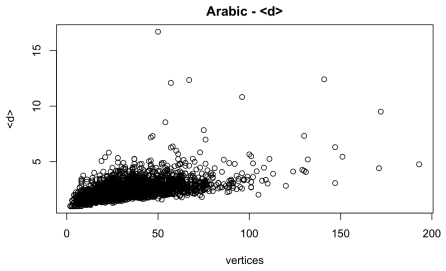
Table 8: Final values of parameters of nls()[3] for Models 1+-5+

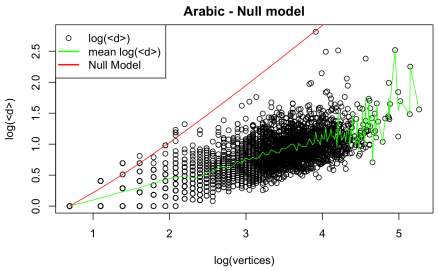
It's worth highlighting that in many cases, the final value of parameter a in model 3+ coincides with its upper bound. An interesting avenue for further investigation would be to explore the impact of setting a higher upper bound for this parameter on the optimization process.

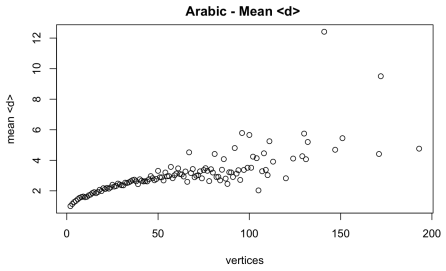
5 Appendix

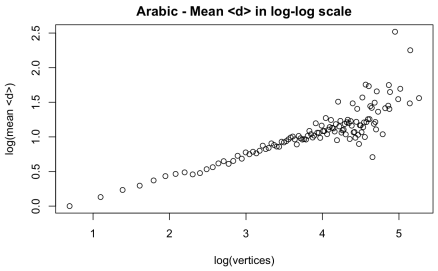
In this section, we will present both preliminary plots and graphs that depict the model fitting for each language.

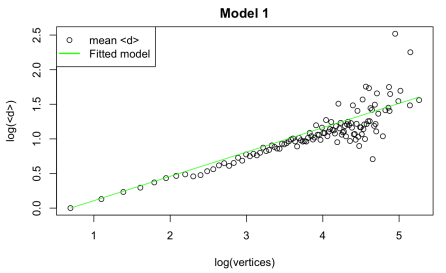
5.1 Arabic

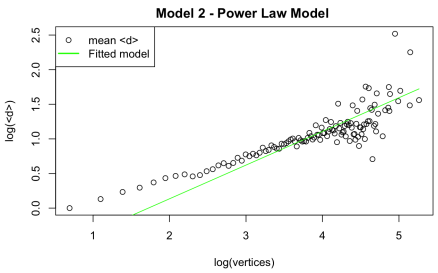


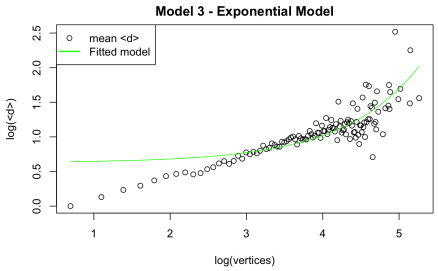


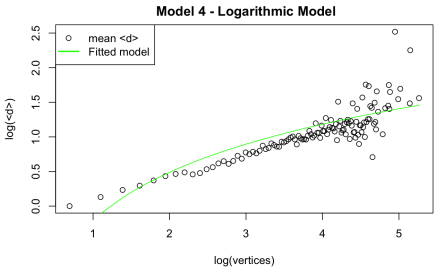


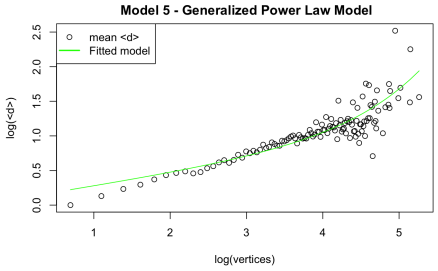


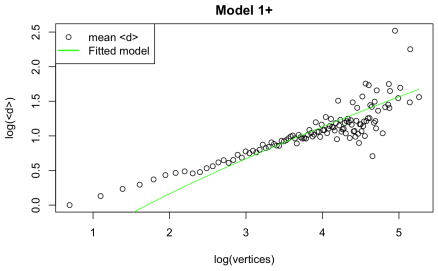


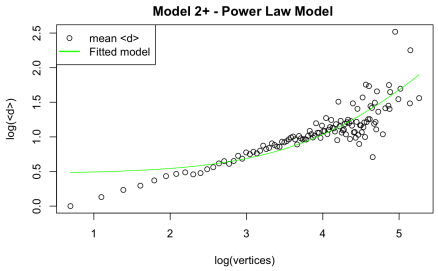


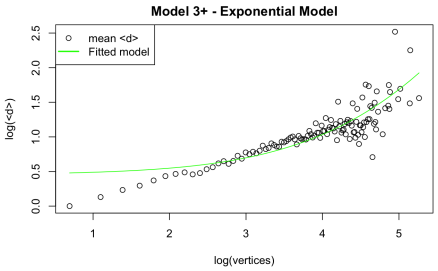

Model 3 - Exponential Model
mean <d>
Fitted model
log(<d>)
log(vertices)

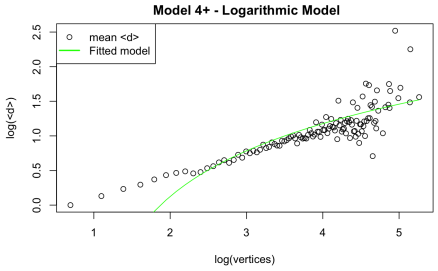

Model 4 - Logarithmic Model
mean <d>
Fitted model
log(<d>)
log(vertices)

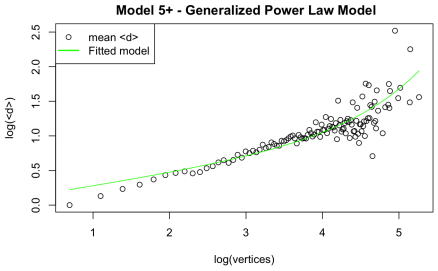

Model 5 - Generalized Power Law Model
mean <d>
Fitted model
log(<d>)
log(vertices)


Model 1+
mean <d>
Fitted model
log(<d>)
log(vertices)

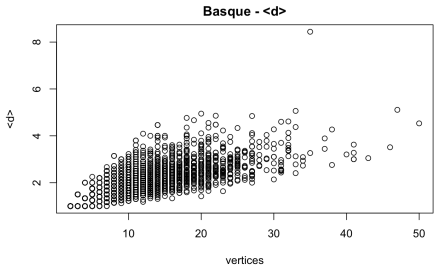

Model 2+ - Power Law Model
mean <d>
Fitted model
log(<d>)
log(vertices)

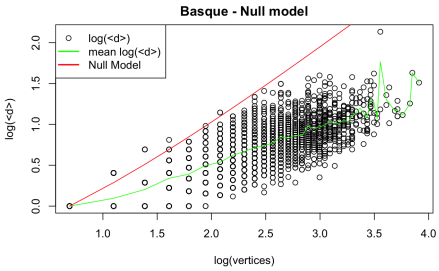

Model 3+ - Exponential Model
mean <d>
Fitted model
log(<d>)
log(vertices)

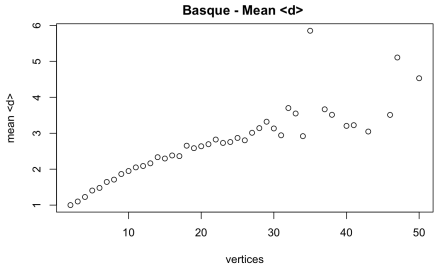

Model 4+ - Logarithmic Model
mean <d>
Fitted model
log(<d>)
log(vertices)

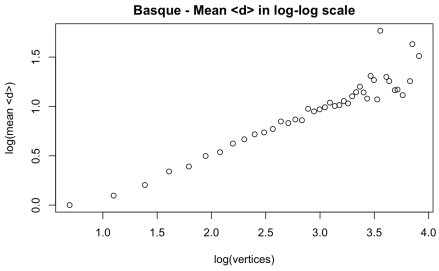

Model 5+ - Generalized Power Law Model
mean <d>
Fitted model
log(<d>)
log(vertices)

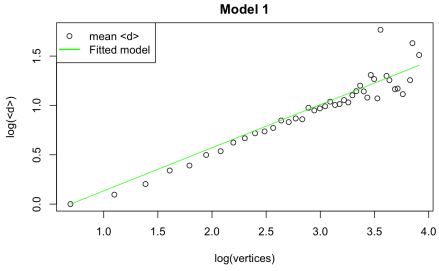
5.2 Basque

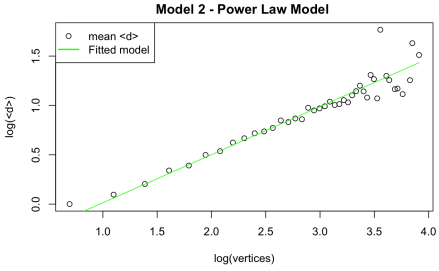

Basque - <d>
<d>
vertices

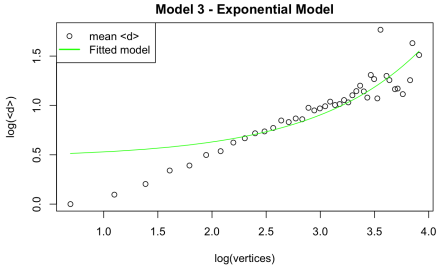

Basque - Null model
log(<d>)
mean log(<d>)
Null Model
log(<d>)
log(vertices)

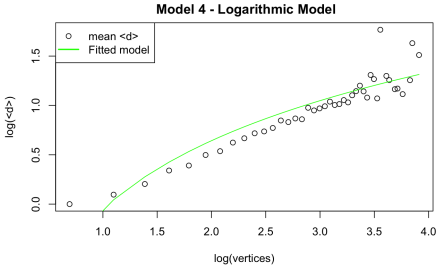

Basque - Mean <d>
mean <d>
vertices

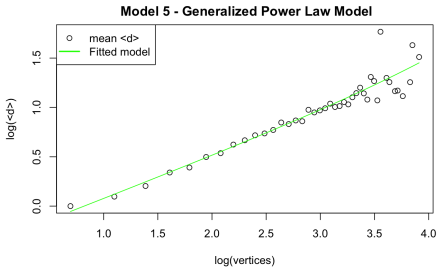

Basque - Mean <d> in log-log scale
log(mean <d>)
log(vertices)

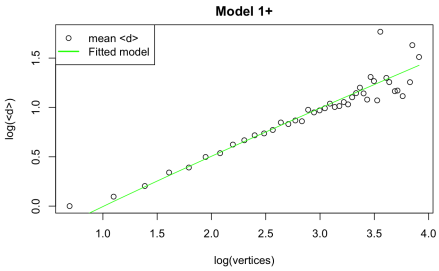

Model 1
mean <d>
Fitted model
log(<d>)
log(vertices)

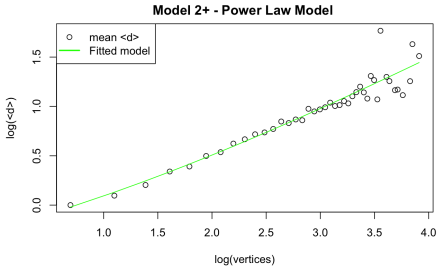

Model 2 - Power Law Model
mean <d>
Fitted model
log(<d>)
log(vertices)

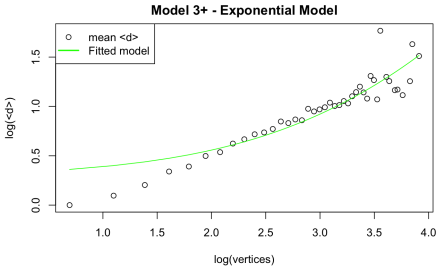

Model 3 - Exponential Model
mean <d>
Fitted model
log(<d>)
log(vertices)

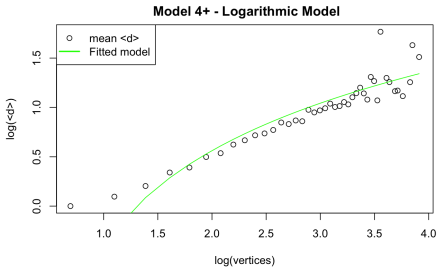

Model 4 - Logarithmic Model
mean <d>
Fitted model
log(<d>)
log(vertices)

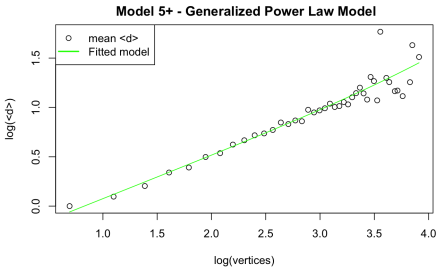

Model 5 - Generalized Power Law Model
mean <d>
Fitted model
log(<d>)
log(vertices)


Model 1+
mean <d>
Fitted model
log(<d>)
log(vertices)

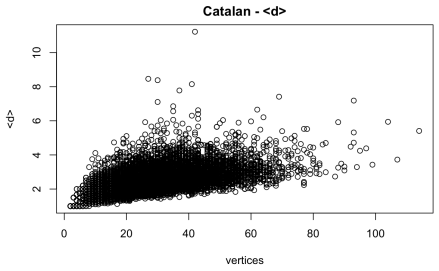

Model 2+ - Power Law Model
mean <d>
Fitted model
log(<d>)
log(vertices)

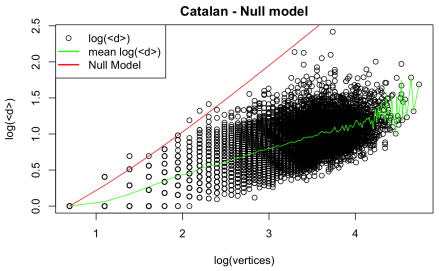

Model 3+ - Exponential Model
mean <d>
Fitted model
log(<d>)
log(vertices)

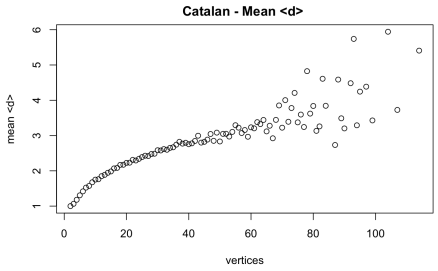

Model 4+ - Logarithmic Model
mean <d>
Fitted model
log(<d>)
log(vertices)

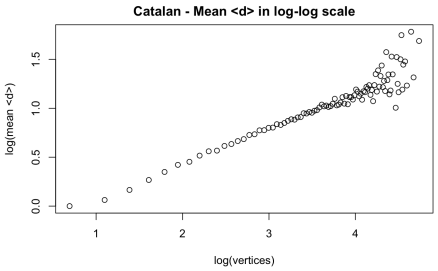

Model 5+ - Generalized Power Law Model
mean <d>
Fitted model
log(<d>)
log(vertices)

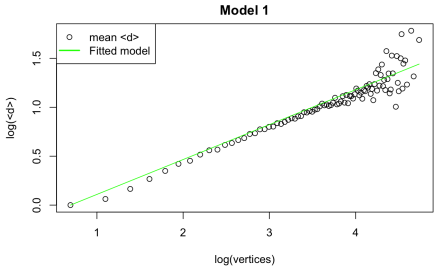
5.3 Catalan

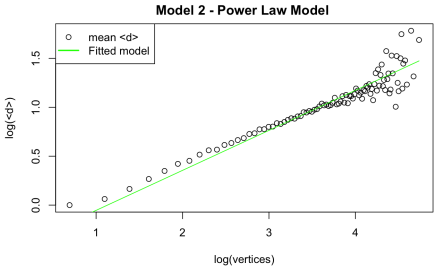

Catalan - <d>
<d>
vertices

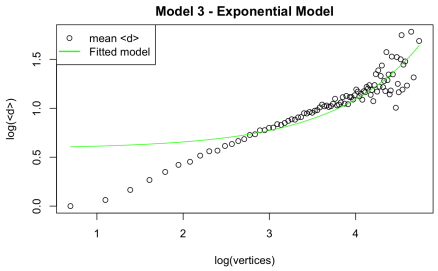

Catalan - Null model
log(<d>)
mean log(<d>)
Null Model
log(<d>)
log(vertices)

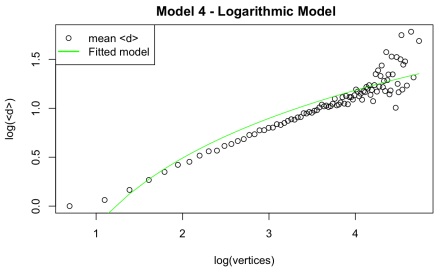

Catalan - Mean <d>
mean <d>
vertices

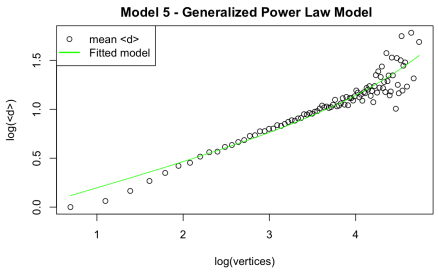

Catalan - Mean <d> in log-log scale
log(mean <d>)
log(vertices)

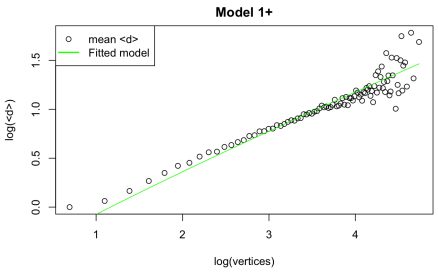

Model 1
mean <d>
Fitted model
log(<d>)
log(vertices)

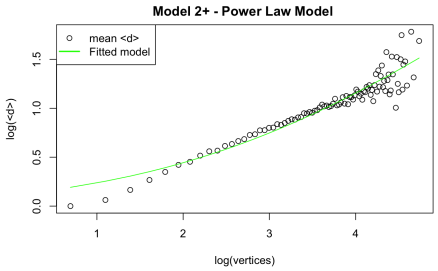

Model 2 - Power Law Model
mean <d>
Fitted model
log(<d>)
log(vertices)

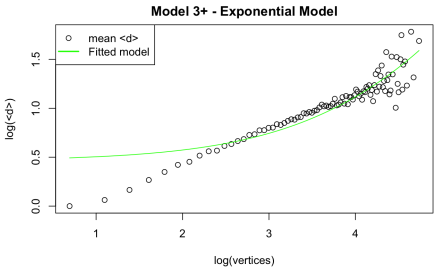

Model 3 - Exponential Model
mean <d>
Fitted model
log(<d>)
log(vertices)

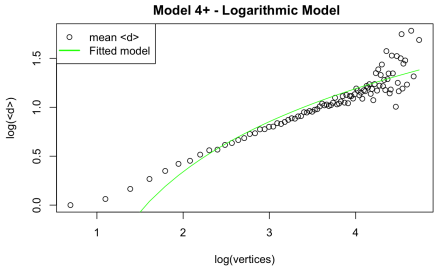

Model 4 - Logarithmic Model
mean <d>
Fitted model
log(<d>)
log(vertices)

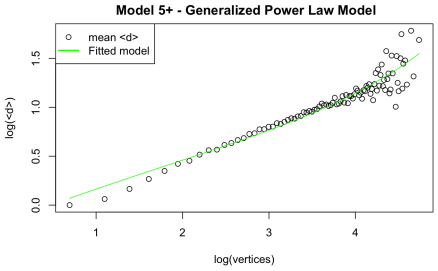

Model 5 - Generalized Power Law Model
mean <d>
Fitted model
log(<d>)
log(vertices)


Model 1+
mean <d>
Fitted model
log(<d>)
log(vertices)

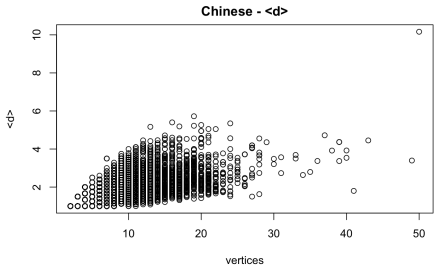

Model 2+ - Power Law Model
mean <d>
Fitted model
log(<d>)
log(vertices)

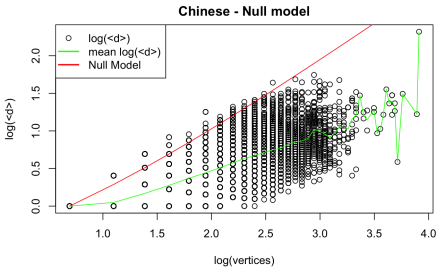

Model 3+ - Exponential Model
mean <d>
Fitted model
log(<d>)
log(vertices)

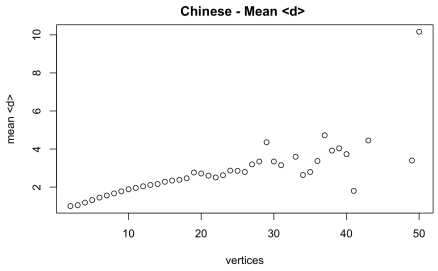

Model 4+ - Logarithmic Model
mean <d>
Fitted model
log(<d>)
log(vertices)

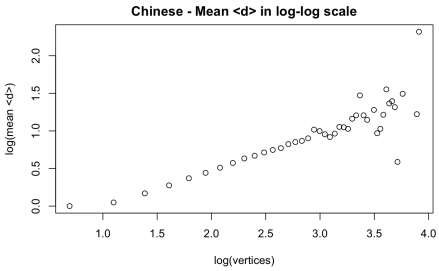

Model 5+ - Generalized Power Law Model
mean <d>
Fitted model
log(<d>)
log(vertices)

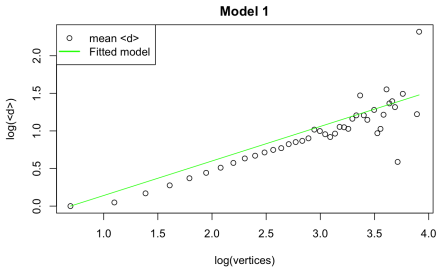
5.4 Chinese

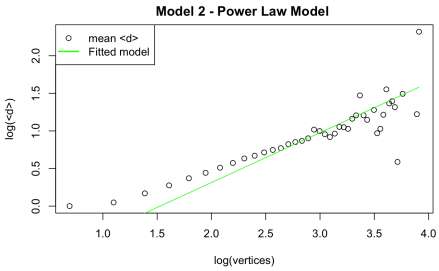

Chinese - <d>
<d>
vertices

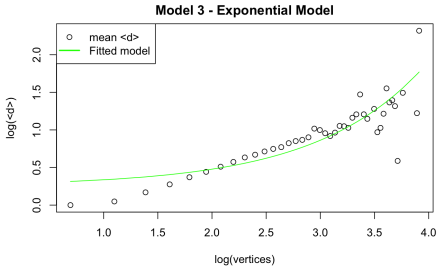

Chinese - Null model
log(<d>)
mean log(<d>)
Null Model
log(<d>)
log(vertices)

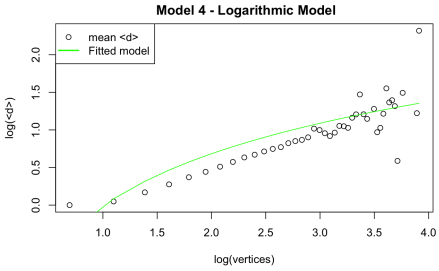

Chinese - Mean <d>
mean <d>
vertices

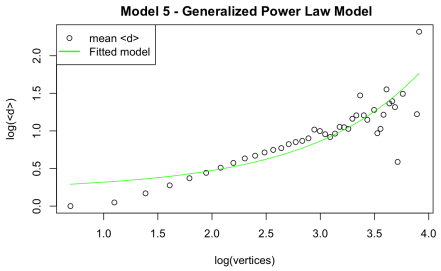

Chinese - Mean <d> in log-log scale
log(mean <d>)
log(vertices)

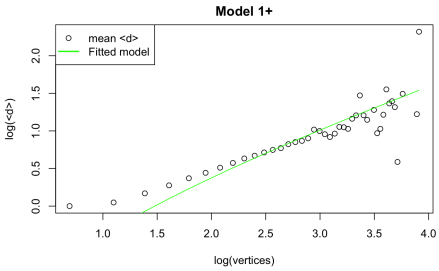

Model 1
mean <d>
Fitted model
log(<d>)
log(vertices)

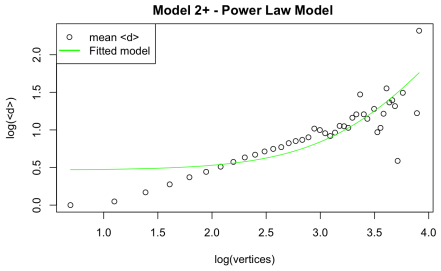

Model 2 - Power Law Model
mean <d>
Fitted model
log(<d>)
log(vertices)

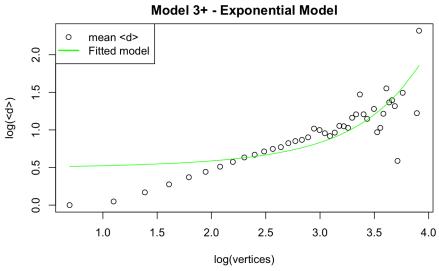

Model 3 - Exponential Model
mean <d>
Fitted model
log(<d>)
log(vertices)

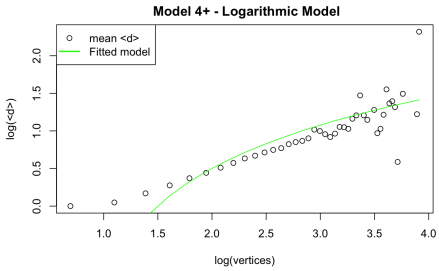

Model 4 - Logarithmic Model
mean <d>
Fitted model
log(<d>)
log(vertices)

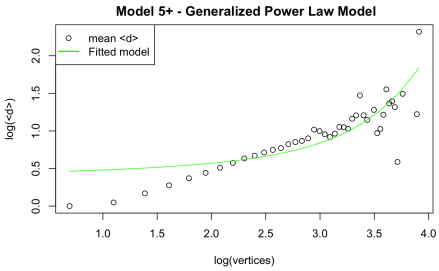

Model 5 - Generalized Power Law Model
mean <d>
Fitted model
log(<d>)
log(vertices)


Model 1+
mean <d>
Fitted model
log(<d>)
log(vertices)

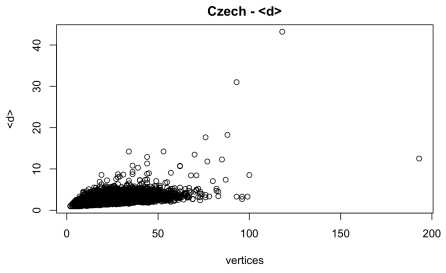

Model 2+ - Power Law Model
mean <d>
Fitted model
log(<d>)
log(vertices)

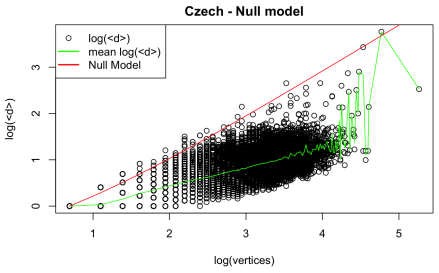

Model 3+ - Exponential Model
mean <d>
Fitted model
log(<d>)
log(vertices)

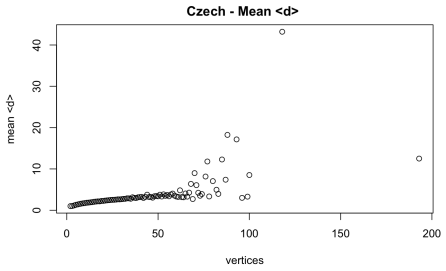

Model 4+ - Logarithmic Model
mean <d>
Fitted model
log(<d>)
log(vertices)

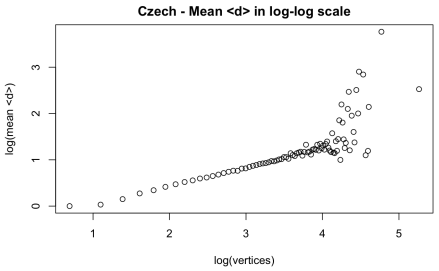

Model 5+ - Generalized Power Law Model
mean <d>
Fitted model
log(<d>)
log(vertices)

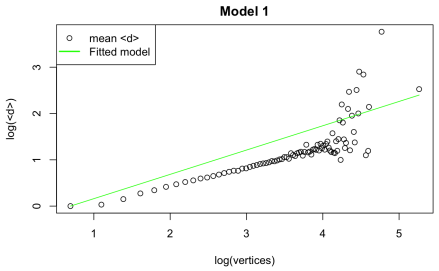
5.5 Czech

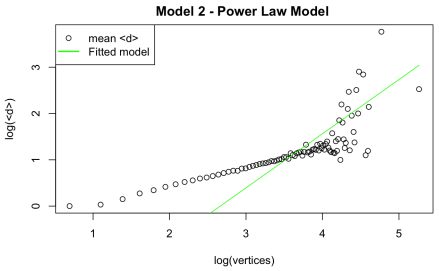

Czech - <d>
<d>
vertices

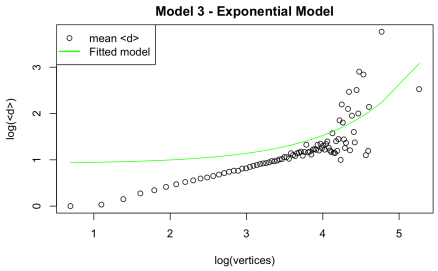

Czech - Null model
log(<d>)
mean log(<d>)
Null Model
log(<d>)
log(vertices)

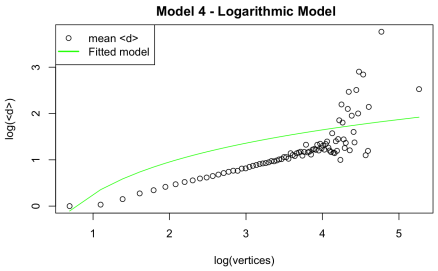

Czech - Mean <d>
mean <d>
vertices

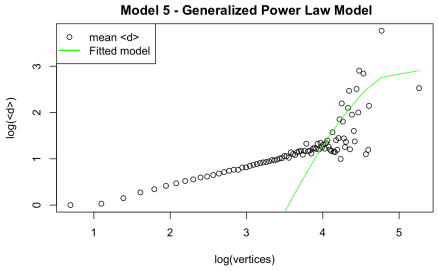

Czech - Mean <d> in log-log scale
log(mean <d>)
log(vertices)

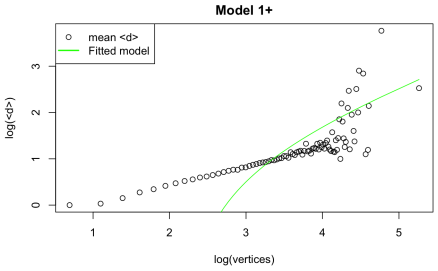

Model 1
mean <d>
Fitted model
log(<d>)
log(vertices)

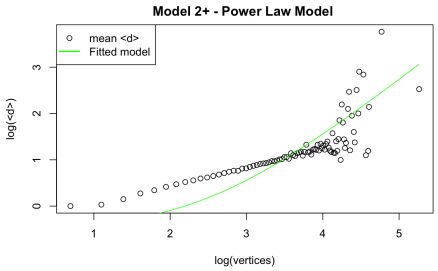

Model 2 - Power Law Model
mean <d>
Fitted model
log(<d>)
log(vertices)

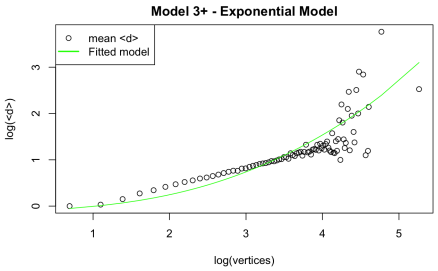

Model 3 - Exponential Model
mean <d>
Fitted model
log(<d>)
log(vertices)

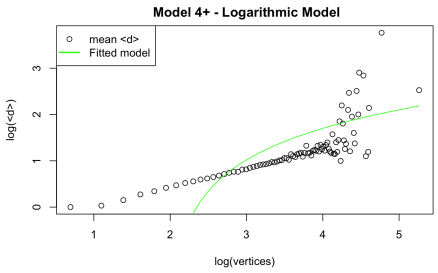

Model 4 - Logarithmic Model
mean <d>
Fitted model
log(<d>)
log(vertices)

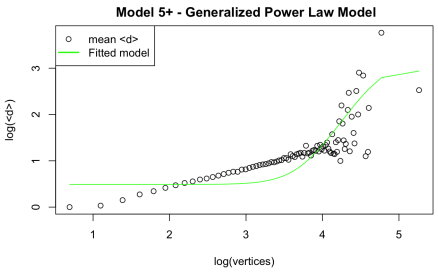

Model 5 - Generalized Power Law Model
mean <d>
Fitted model
log(<d>)
log(vertices)


Model 1+
mean <d>
Fitted model
log(<d>)
log(vertices)

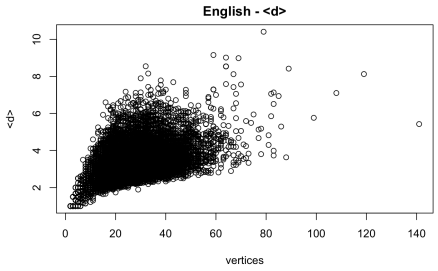

Model 2+ - Power Law Model
mean <d>
Fitted model
log(<d>)
log(vertices)

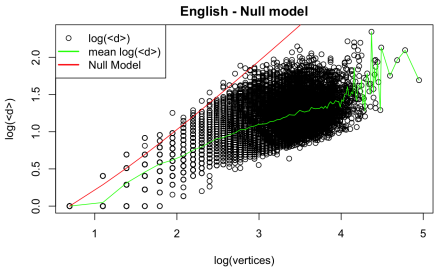

Model 3+ - Exponential Model
mean <d>
Fitted model
log(<d>)
log(vertices)

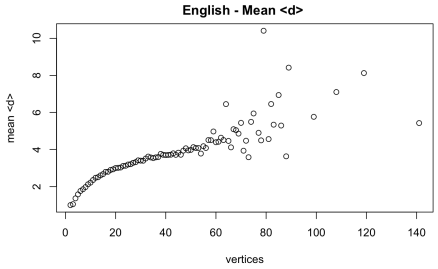

Model 4+ - Logarithmic Model
mean <d>
Fitted model
log(<d>)
log(vertices)

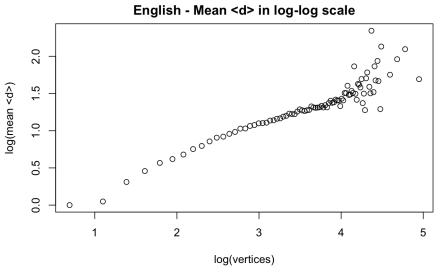

Model 5+ - Generalized Power Law Model
mean <d>
Fitted model
log(<d>)
log(vertices)

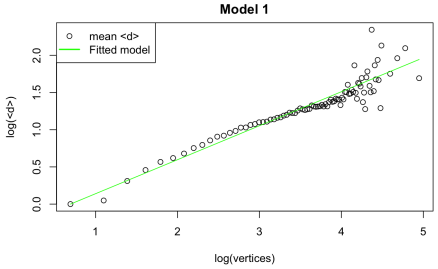
5.6 English

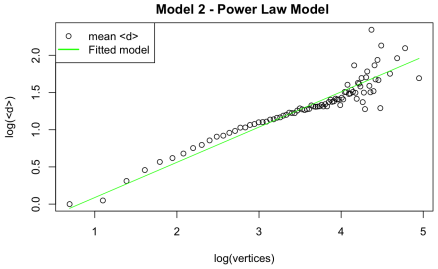

English - <d>
<d>
vertices

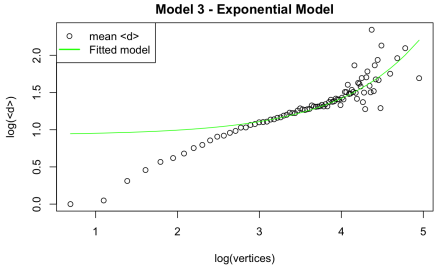

English - Null model
log(<d>)
mean log(<d>)
Null Model
log(<d>)
log(vertices)

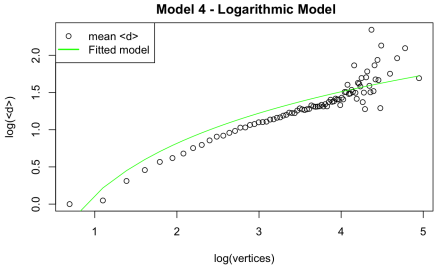

English - Mean <d>
mean <d>
vertices

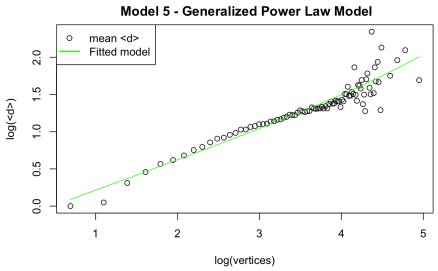

English - Mean <d> in log-log scale
log(mean <d>)
log(vertices)

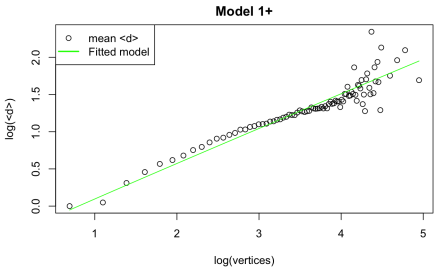

Model 1
mean <d>
Fitted model
log(<d>)
log(vertices)

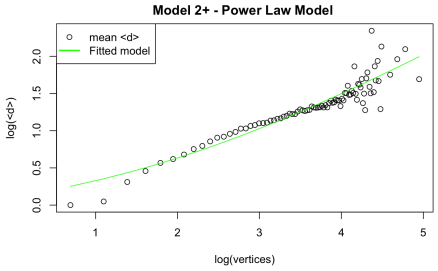

Model 2 - Power Law Model
mean <d>
Fitted model
log(<d>)
log(vertices)

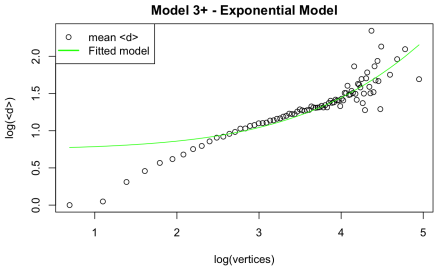

Model 3 - Exponential Model
mean <d>
Fitted model
log(<d>)
log(vertices)

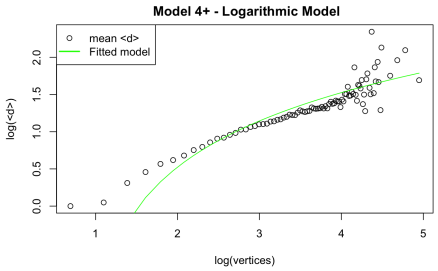

Model 4 - Logarithmic Model
mean <d>
Fitted model
log(<d>)
log(vertices)

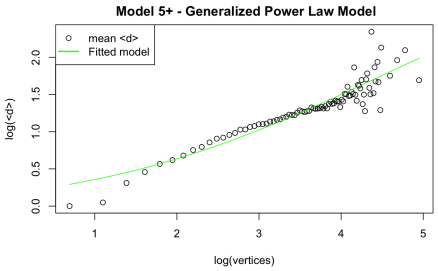

Model 5 - Generalized Power Law Model
mean <d>
Fitted model
log(<d>)
log(vertices)


Model 1+
mean <d>
Fitted model
log(<d>)
log(vertices)

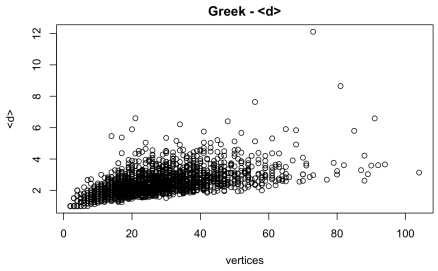

Model 2+ - Power Law Model
mean <d>
Fitted model
log(<d>)
log(vertices)

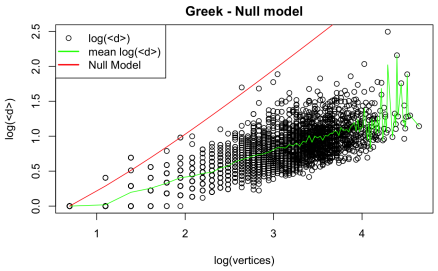

Model 3+ - Exponential Model
mean <d>
Fitted model
log(<d>)
log(vertices)

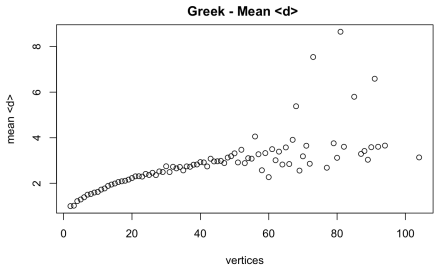

Model 4+ - Logarithmic Model
mean <d>
Fitted model
log(<d>)
log(vertices)

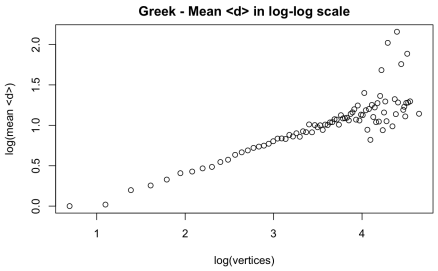

Model 5+ - Generalized Power Law Model
mean <d>
Fitted model
log(<d>)
log(vertices)

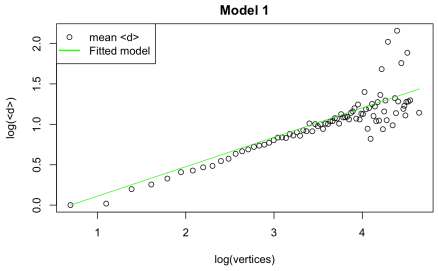
5.7 Greek

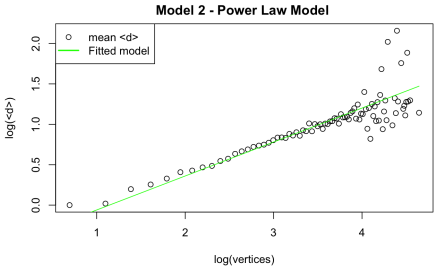

Greek - <d>
<d>
vertices

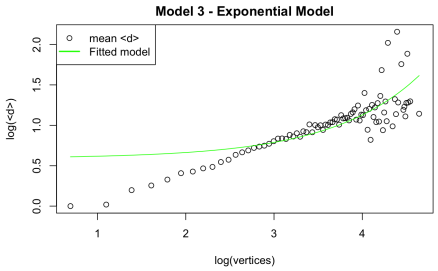

Greek - Null model
log(<d>)
mean log(<d>)
Null Model
log(<d>)
log(vertices)

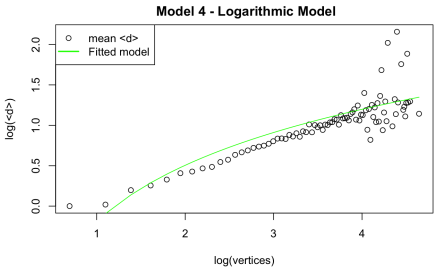

Greek - Mean <d>
mean <d>
vertices

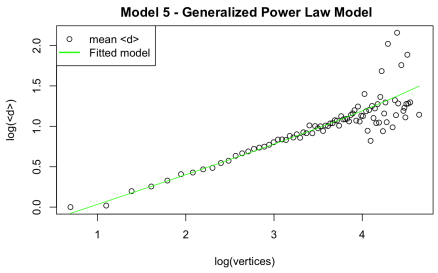

Greek - Mean <d> in log-log scale
log(mean <d>)
log(vertices)

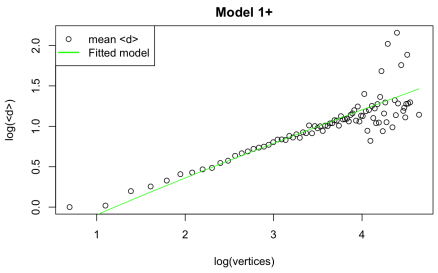

Model 1
mean <d>
Fitted model
log(<d>)
log(vertices)

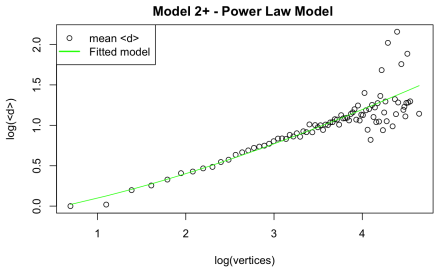

Model 2 - Power Law Model
mean <d>
Fitted model
log(<d>)
log(vertices)

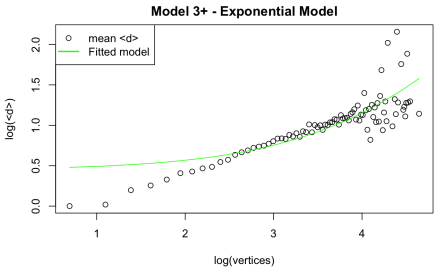

Model 3 - Exponential Model
mean <d>
Fitted model
log(<d>)
log(vertices)

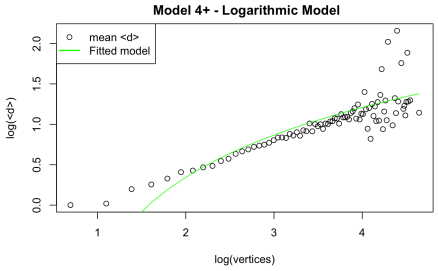

Model 4 - Logarithmic Model
mean <d>
Fitted model
log(<d>)
log(vertices)

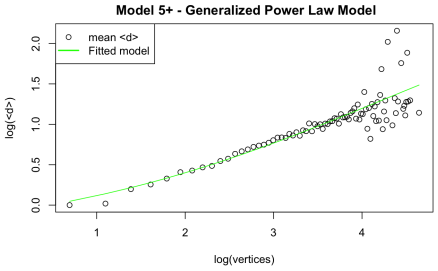

Model 5 - Generalized Power Law Model
mean <d>
Fitted model
log(<d>)
log(vertices)


Model 1+
mean <d>
Fitted model
log(<d>)
log(vertices)

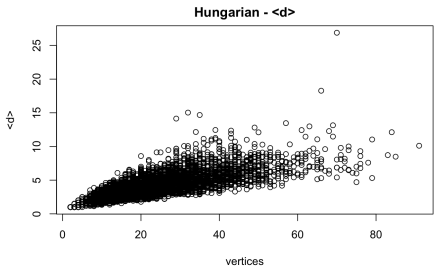


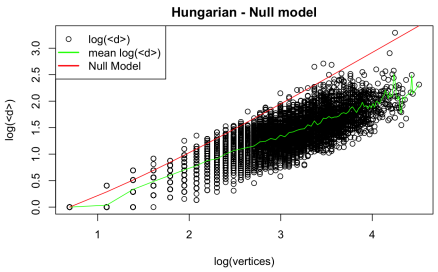


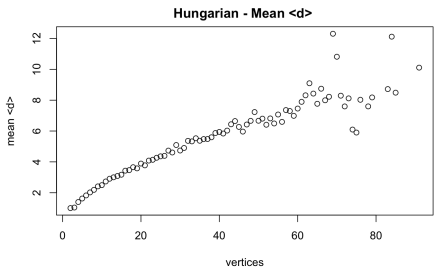


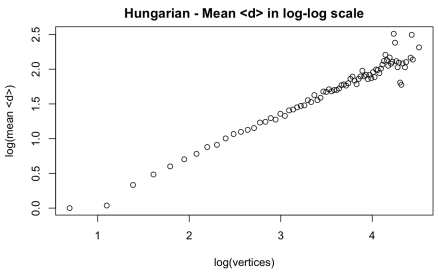


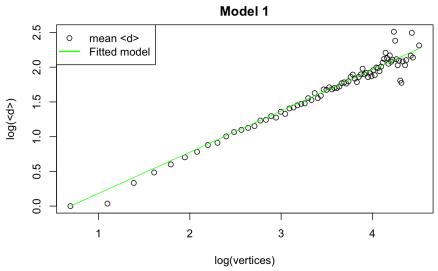
5.8 Hungarian

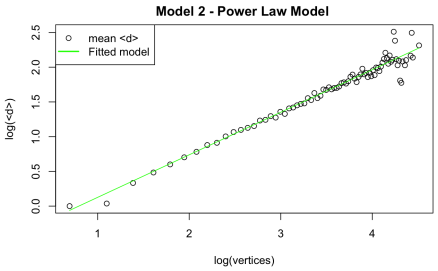


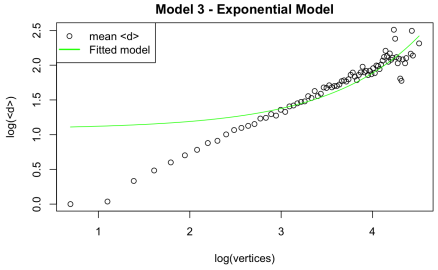


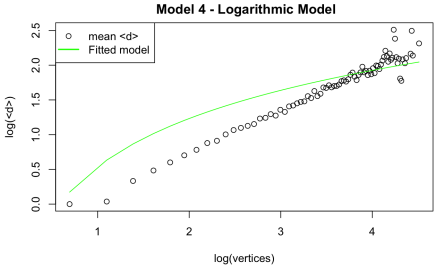


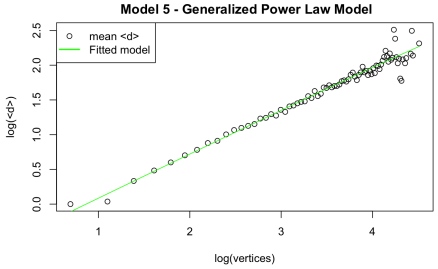


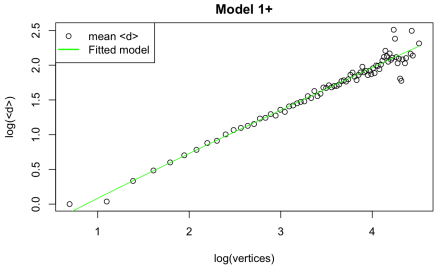

Model 1
mean <d>
Fitted model
log(<d>)
log(vertices)

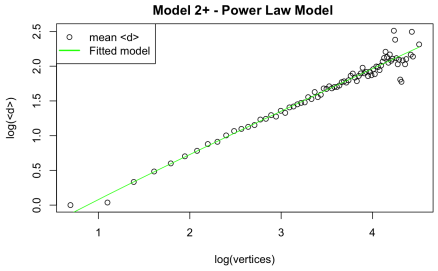

Model 2 - Power Law Model
mean <d>
Fitted model
log(<d>)
log(vertices)

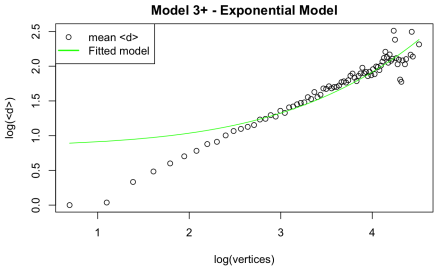

Model 3 - Exponential Model
mean <d>
Fitted model
log(<d>)
log(vertices)

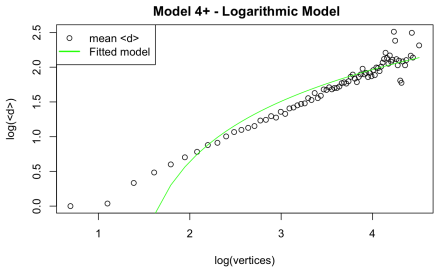

Model 4 - Logarithmic Model
mean <d>
Fitted model
log(<d>)
log(vertices)

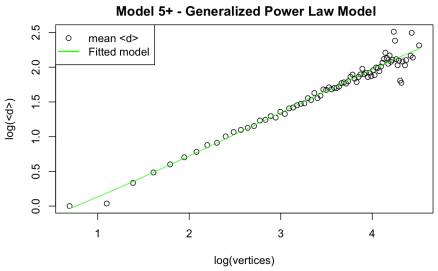

Model 5 - Generalized Power Law Model
mean <d>
Fitted model
log(<d>)
log(vertices)


Model 1+
mean <d>
Fitted model
log(<d>)
log(vertices)

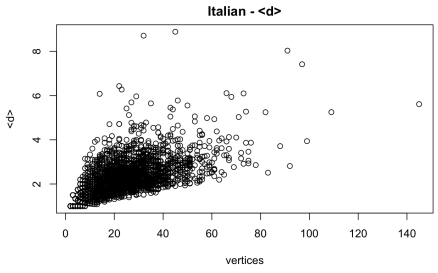

Model 2+ - Power Law Model
mean <d>
Fitted model
log(<d>)
log(vertices)

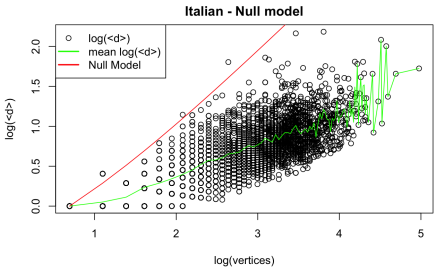

Model 3+ - Exponential Model
mean <d>
Fitted model
log(<d>)
log(vertices)

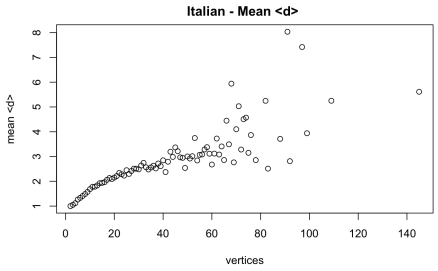

Model 4+ - Logarithmic Model
mean <d>
Fitted model
log(<d>)
log(vertices)

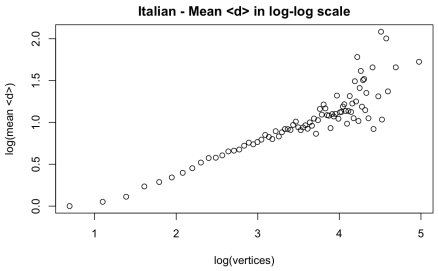

Model 5+ - Generalized Power Law Model
mean <d>
Fitted model
log(<d>)
log(vertices)

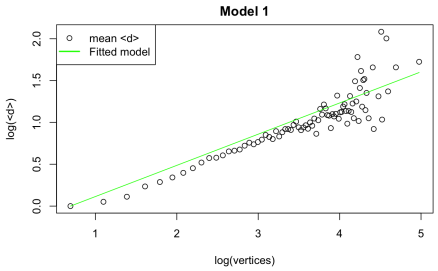
5.9 Italian

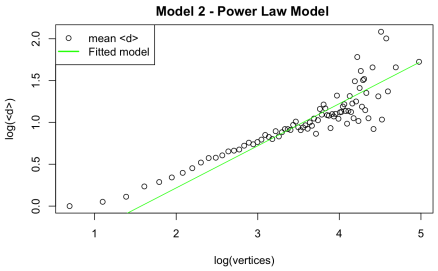

Italian - <d>
<d>
vertices

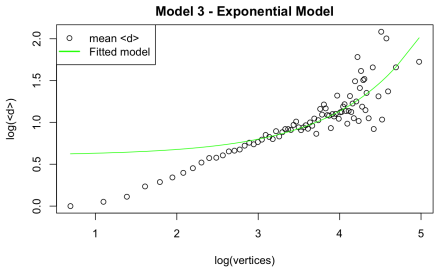

Italian - Null model
log(<d>)
mean log(<d>)
Null Model
log(<d>)
log(vertices)

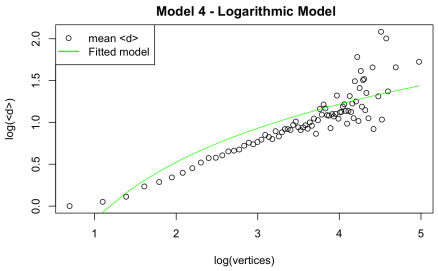

Italian - Mean <d>
mean <d>
vertices

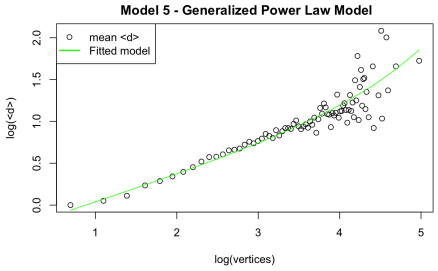

Italian - Mean <d> in log-log scale
log(mean <d>)
log(vertices)

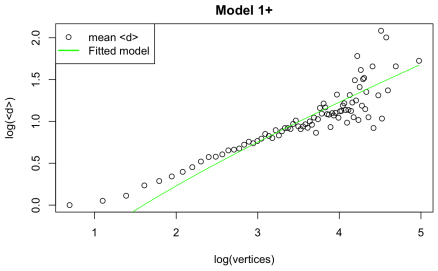

Model 1
mean <d>
Fitted model
log(<d>)
log(vertices)

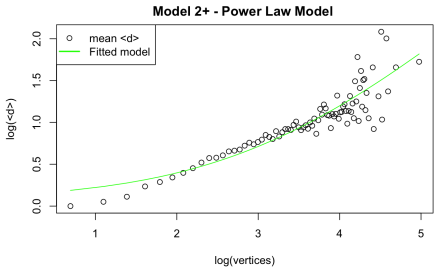

Model 2 - Power Law Model
mean <d>
Fitted model
log(<d>)
log(vertices)

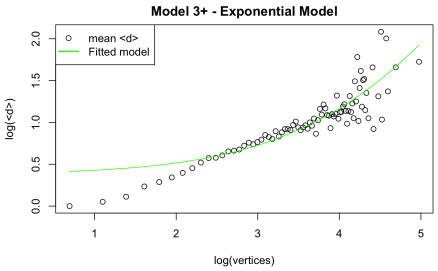

Model 3 - Exponential Model
mean <d>
Fitted model
log(<d>)
log(vertices)

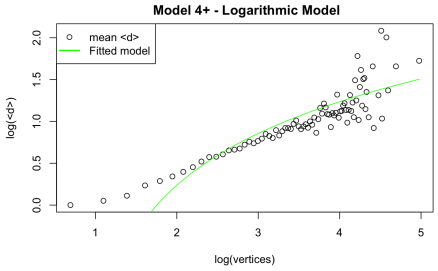

Model 4 - Logarithmic Model
mean <d>
Fitted model
log(<d>)
log(vertices)

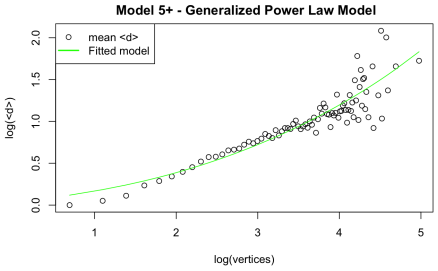

Model 5 - Generalized Power Law Model
mean <d>
Fitted model
log(<d>)
log(vertices)


Model 1+
mean <d>
Fitted model
log(<d>)
log(vertices)

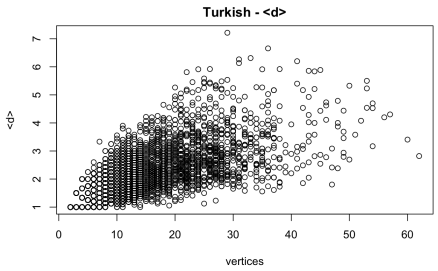

Model 2+ - Power Law Model
mean <d>
Fitted model
log(<d>)
log(vertices)

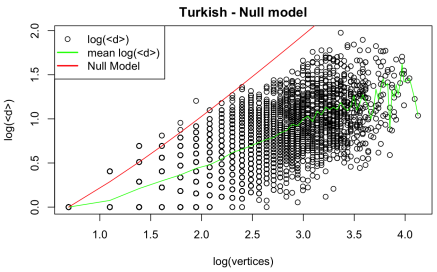

Model 3+ - Exponential Model
mean <d>
Fitted model
log(<d>)
log(vertices)

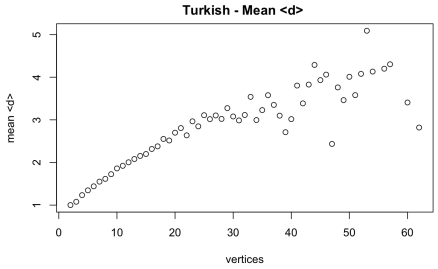

Model 4+ - Logarithmic Model
mean <d>
Fitted model
log(<d>)
log(vertices)

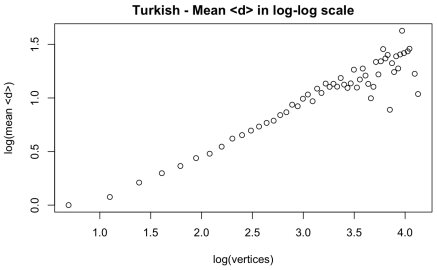

Model 5+ - Generalized Power Law Model
mean <d>
Fitted model
log(<d>)
log(vertices)

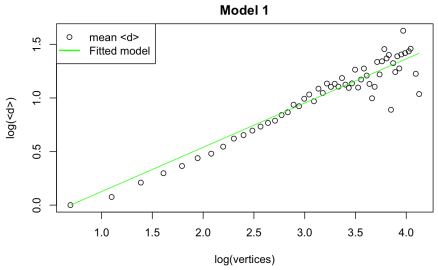
5.10 Turkish

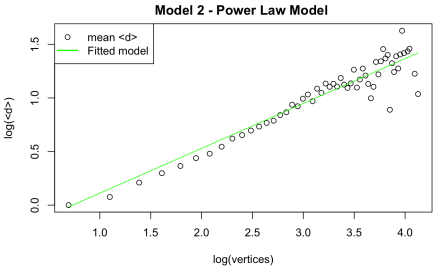

Turkish - <d>
<d>
vertices

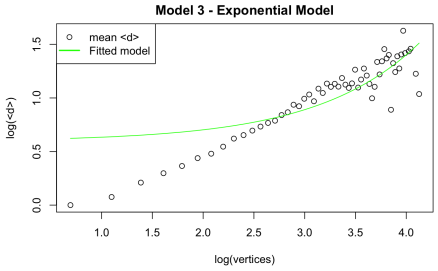

Turkish - Null model
log(<d>)
mean log(<d>)
Null Model
log(<d>)
log(vertices)

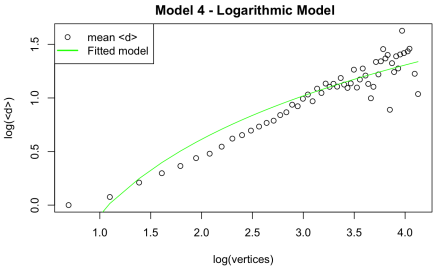

Turkish - Mean <d>
mean <d>
vertices

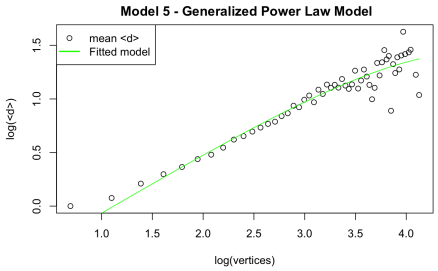

Turkish - Mean <d> in log-log scale
log(mean <d>)
log(vertices)

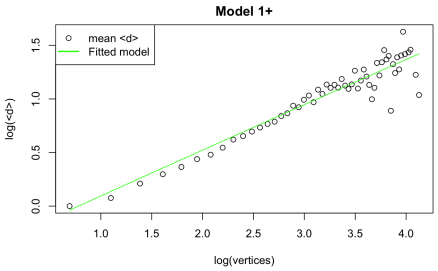

Model 1
mean <d>
Fitted model
log(<d>)
log(vertices)

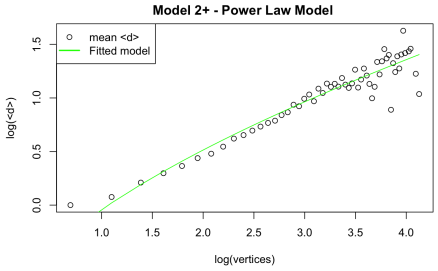

Model 2 - Power Law Model
mean <d>
Fitted model
log(<d>)
log(vertices)

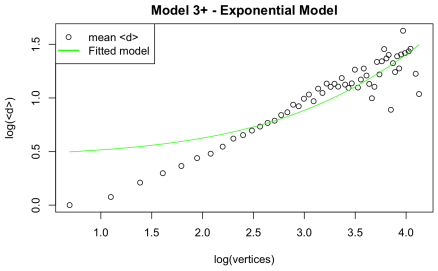

Model 3 - Exponential Model
mean <d>
Fitted model
log(<d>)
log(vertices)

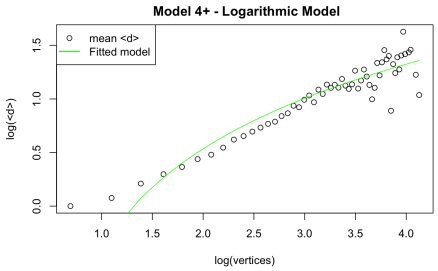

Model 4 - Logarithmic Model
mean <d>
Fitted model
log(<d>)
log(vertices)

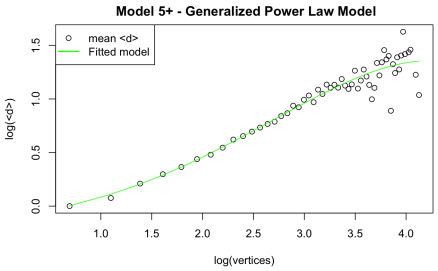

Model 5 - Generalized Power Law Model
mean <d>
Fitted model
log(<d>)
log(vertices)


Model 1+
mean <d>
Fitted model
log(<d>)
log(vertices)


Model 2+ - Power Law Model
mean <d>
Fitted model
log(<d>)
log(vertices)


Model 3+ - Exponential Model
mean <d>
Fitted model
log(<d>)
log(vertices)


Model 4+ - Logarithmic Model
mean <d>
Fitted model
log(<d>)
log(vertices)


Model 5+ - Generalized Power Law Model
mean <d>
Fitted model
log(<d>)
log(vertices)