UPC - Complex and Social Networks

Lab 2: Analysis of the degree distribution

Michele Gentili and Simon Van den Eynde

October 13, 2016

1 Introduction

In this lab we will try to select a theoretic model, fitting the degree distribution of syntactic dependency networks in different languages. We will focus on in-degrees.

We considered 5 theoretical models and used the Akaike information criterions (AIC) to decide which model was most favourable. The models we considered were

1. a zeta distribution
2. a zeta distribution with fixed exponent 2
3. a right-truncated zeta distribution
4. a geometric distribution
5. a geometric displaced distribution
6. a displaced poisson distribution

After choosing the statistically best models, we plotted our models to our data to do a visual check of correctness.

And finally recompute the Akaike scores, introducing a better model: Altmann distribution.

Table 1: Difference in AIC for the 5 different models in different languages

	Models				
Language	1	2	3	4	5
Arabic	2.91	175.28	0.00	24188.95	240321.25
Basque	0.24	845.41	0.00	8364.11	50164.59
Catalan	13.96	225.90	0.00	61881.73	913880.09
Chinese	14.69	503.99	0.00	48678.82	618348.40
Czech	8.85	147.28	0.00	91099.39	940677.71
English	45.05	1469.66	0.00	45742.18	739581.47
Greek	3.43	164.11	0.00	17133.83	157137.24
Hungarian	0.00	2220.28	1.71	53469.62	468252.20
Italian	0.49	118.37	0.00	22877.90	245803.51
Turkish	0.00	2740.47	1.98	24615.35	193345.03

2 Results

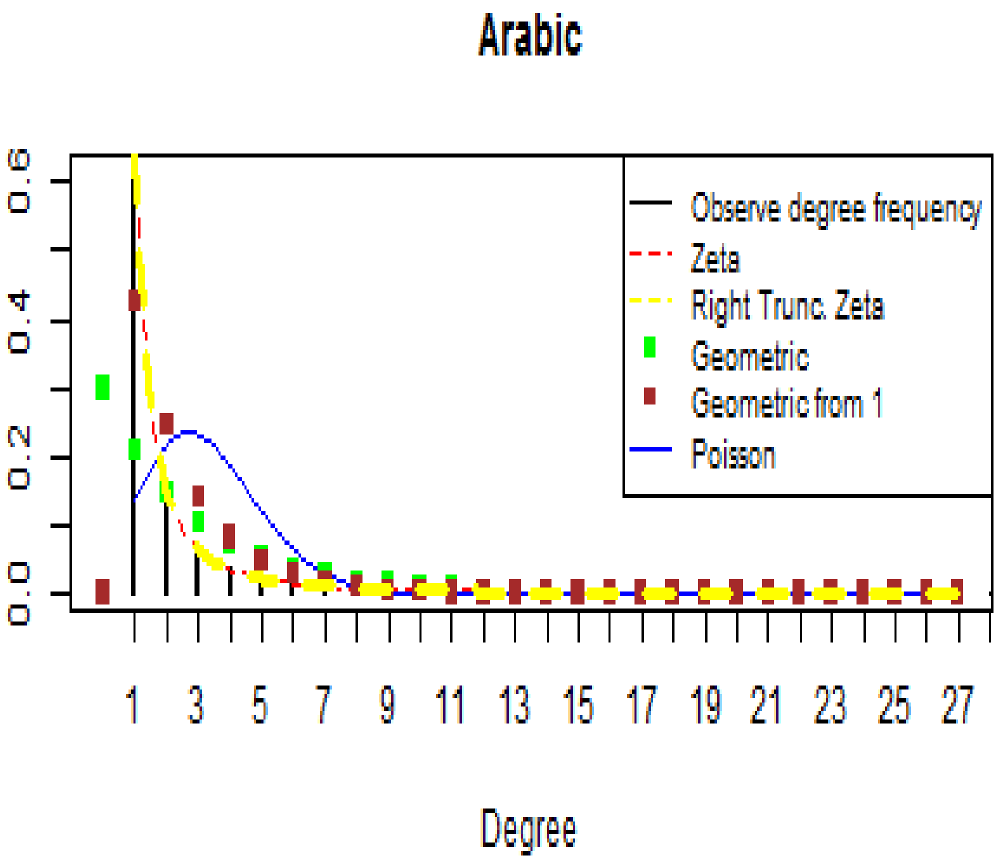
In table 1 we find our main result, the difference in AIC from the best (of the 5) model. We notice that except for Hungarian and Turkish the right-truncated zeta always ends up best.

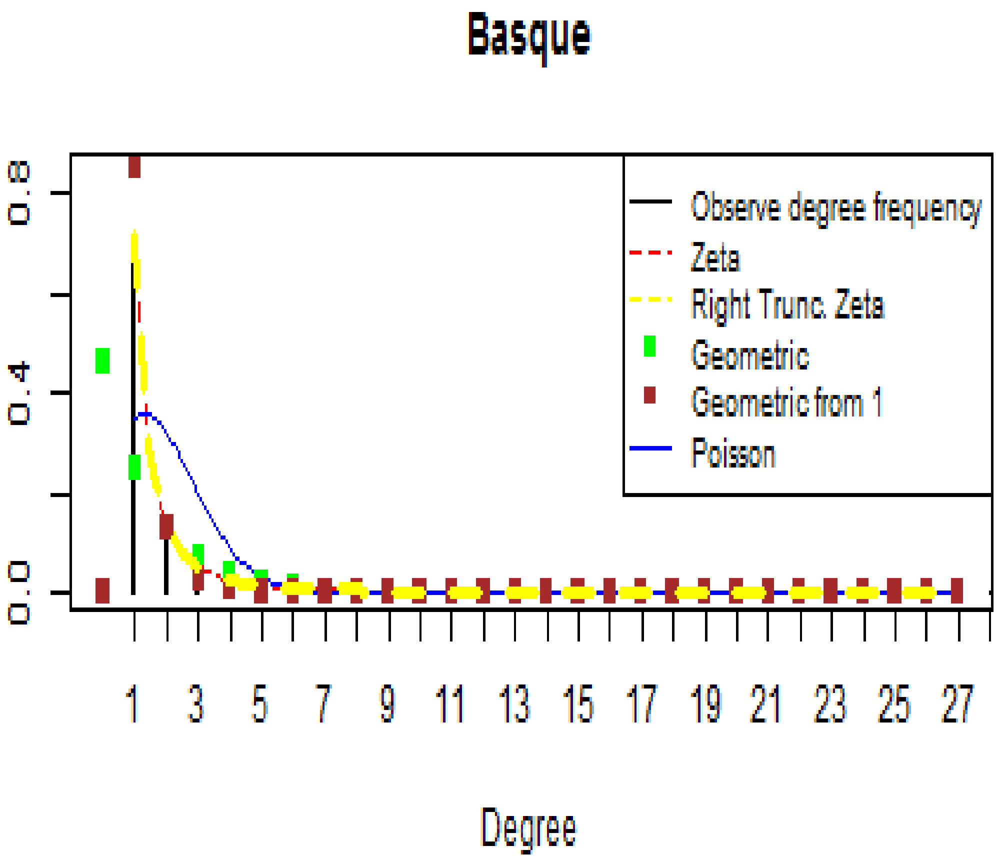
In table 2 we find the RSS-values for the zeta and right-truncated zeta distribution. We notice that in both table 1 and 2 most values for the zeta and right-truncated zeta are very alike.

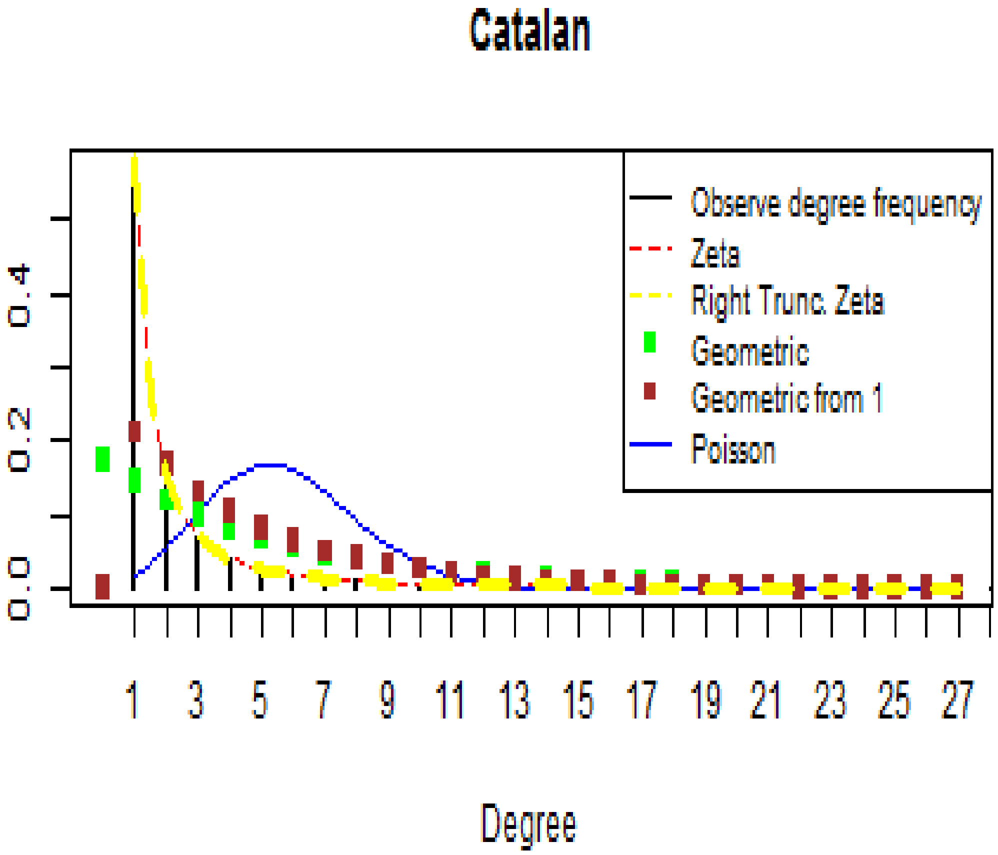
Table 2: RSS values for zeta and right-truncated zeta distribution with optimised parameters

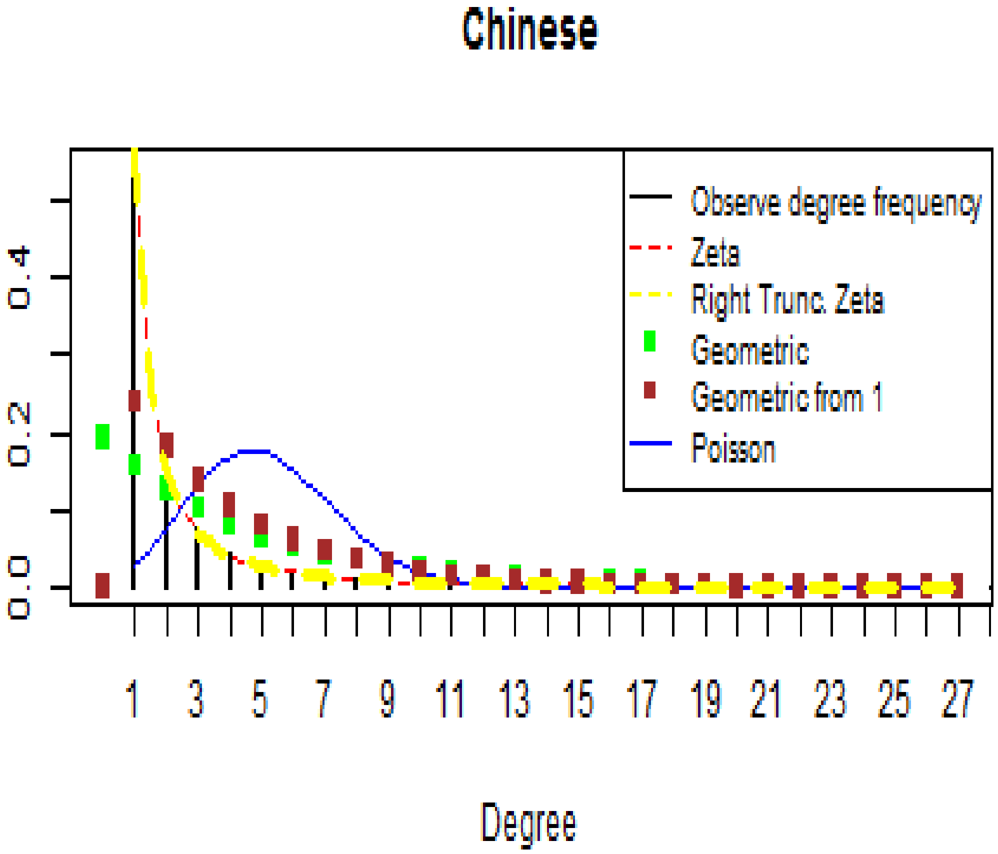
	zeta	RT_zeta
Arabic	9.54e-04	9.33e-04
Basque	1.68e-05	1.78e-05
Catalan	1.12e-04	1.03e-04
Chinese	5.21e-04	4.94e-04
Czech	4.12e-05	3.82e-05
English	3.46e-04	2.90e-04
Greek	4.38e-04	4.15e-04
Hungarian	3.93e-04	3.94e-04
Italian	1.82e-04	1.75e-04
Turkish	6.92e-05	6.92e-05

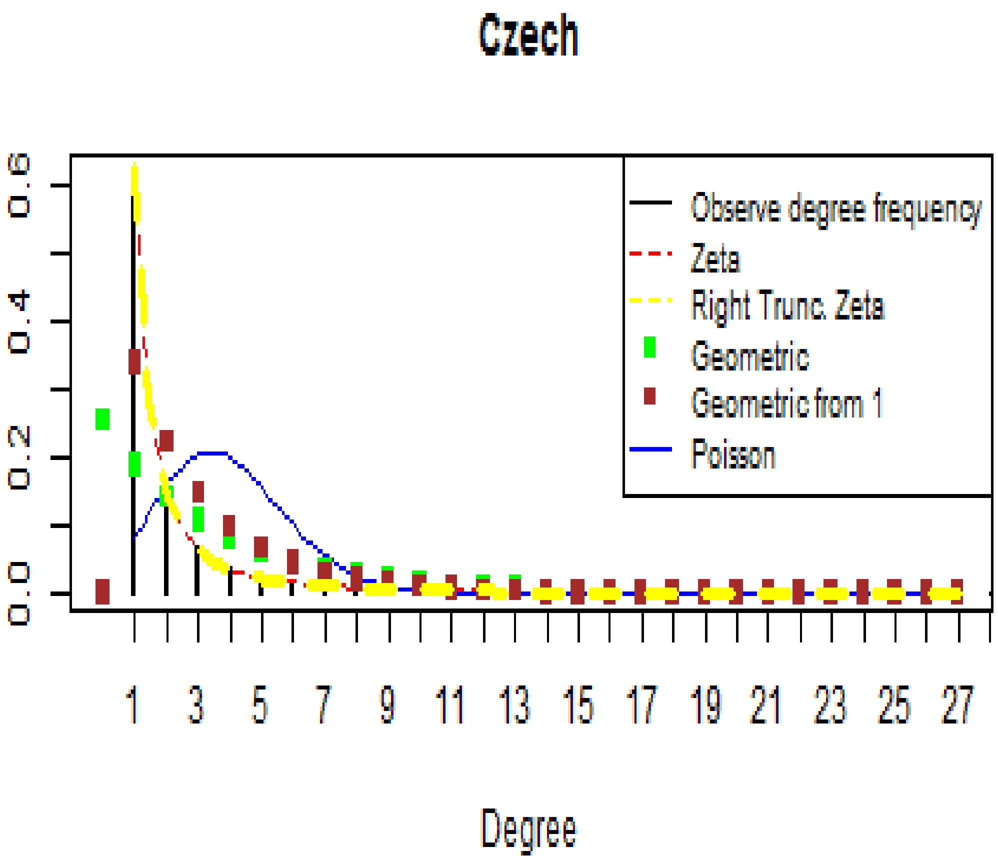
In table 3, as we expected Altmann function has the best score, so all the other values are shifted of the difference between the previous best model score and the Altmann score.

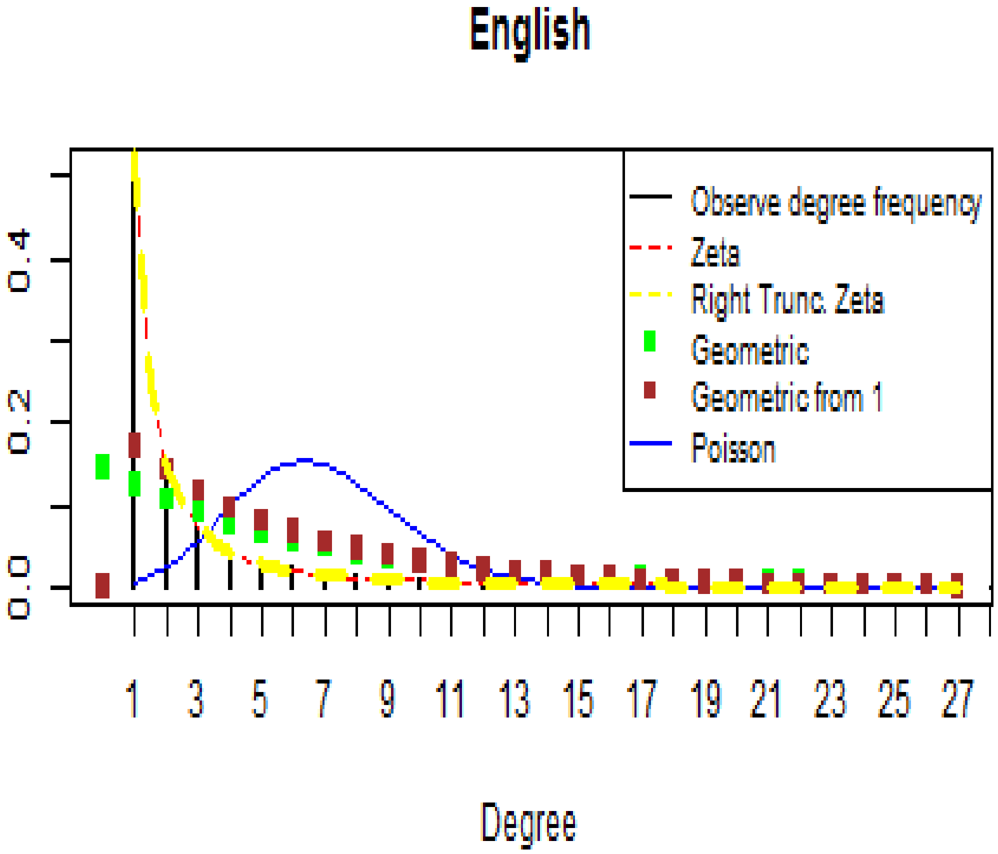

Arabic
Observe degree frequency
Zeta
Right Trunc. Zeta
Geometric
Geometric from 1
Poisson
0.0
0.2
0.4
0.6
1
3
5
7
9
11
13
15
17
19
21
23
25
27
Degree

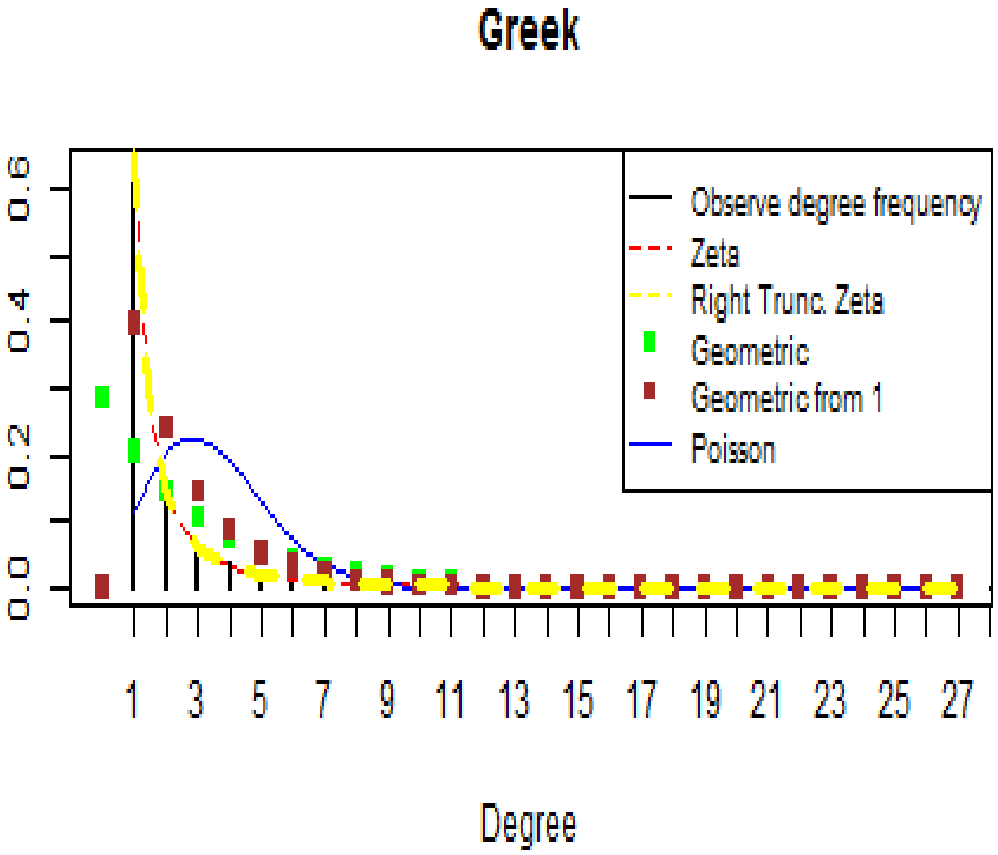

Basque
Observe degree frequency
Zeta
Right Trunc. Zeta
Geometric
Geometric from 1
Poisson
0.0
0.4
0.8
1
3
5
7
9
11
13
15
17
19
21
23
25
27
Degree

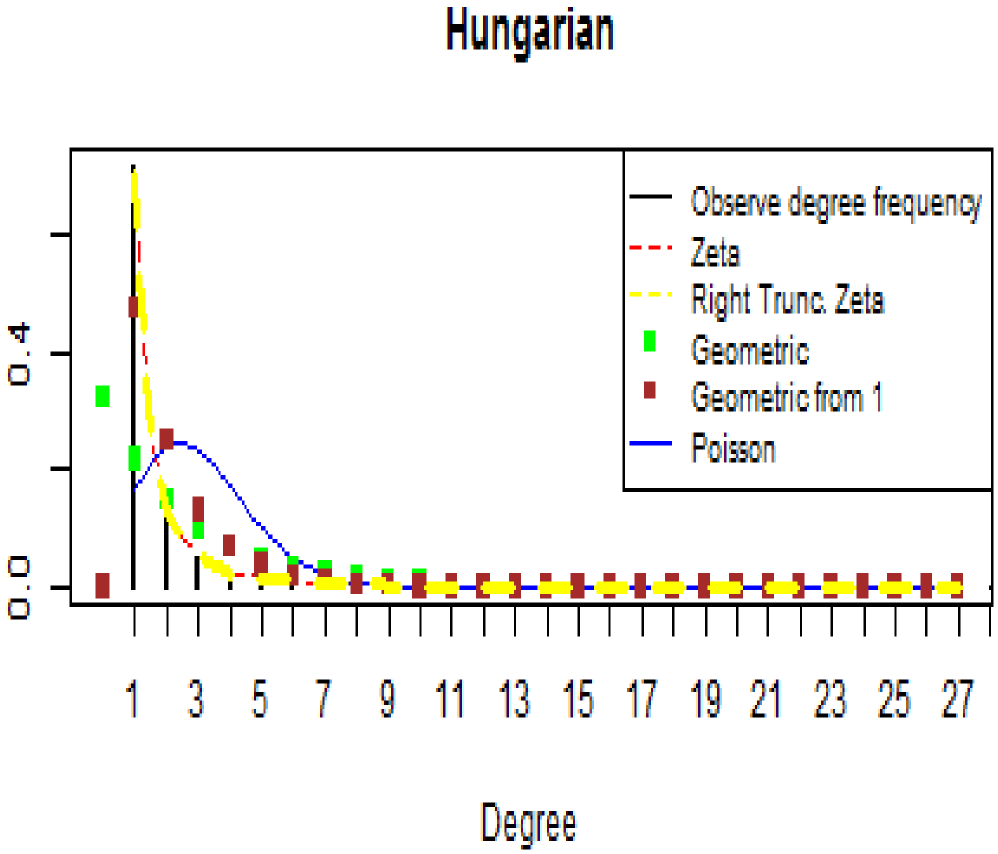

Catalan
Observe degree frequency
Zeta
Right Trunc. Zeta
Geometric
Geometric from 1
Poisson
0.0
0.2
0.4
1
3
5
7
9
11
13
15
17
19
21
23
25
27
Degree

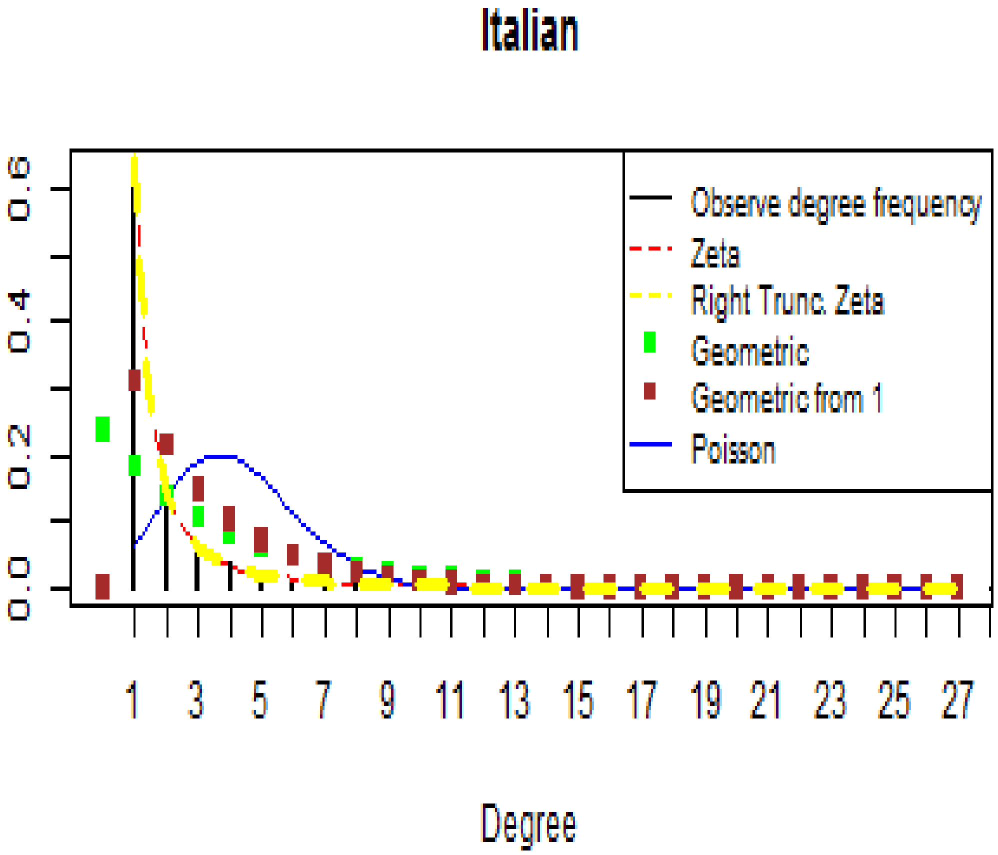

Chinese
Observe degree frequency
Zeta
Right Trunc. Zeta
Geometric
Geometric from 1
Poisson
0.0
0.2
0.4
1
3
5
7
9
11
13
15
17
19
21
23
25
27
Degree

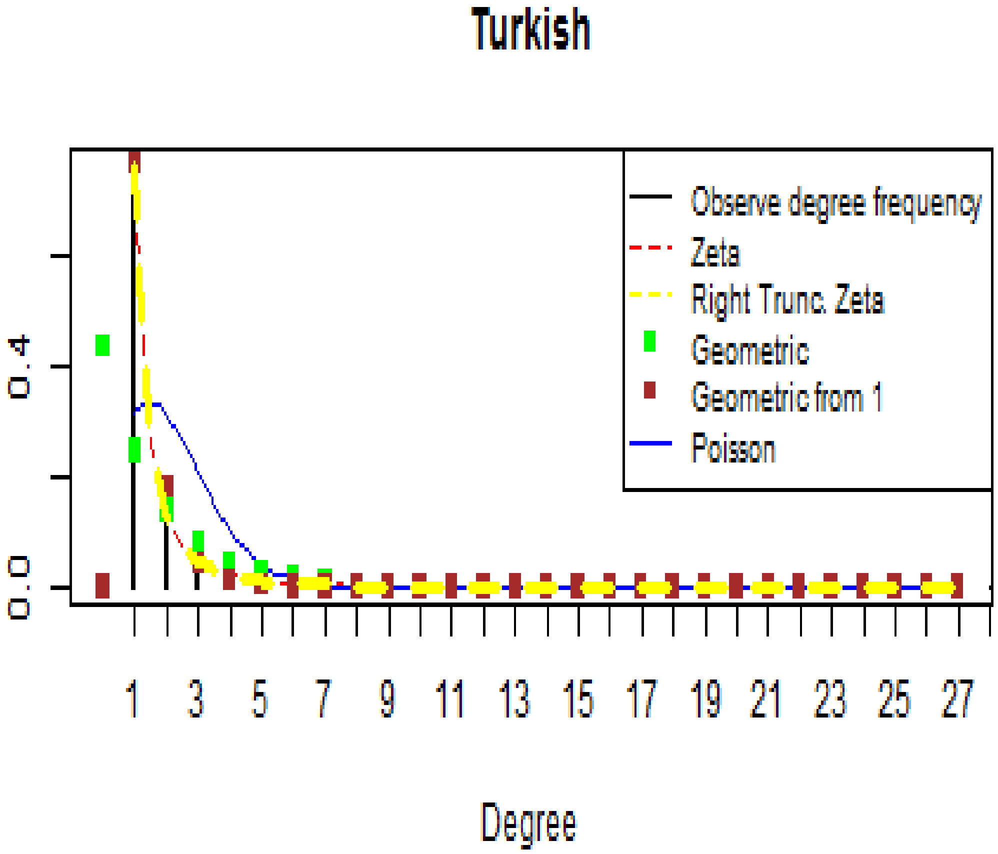

Czech
0.0
0.2
0.4
0.6
Observe degree frequency
Zeta
Right Trunc. Zeta
Geometric
Geometric from 1
Poisson
1
3
5
7
9
11
13
15
17
19
21
23
25
27
Degree

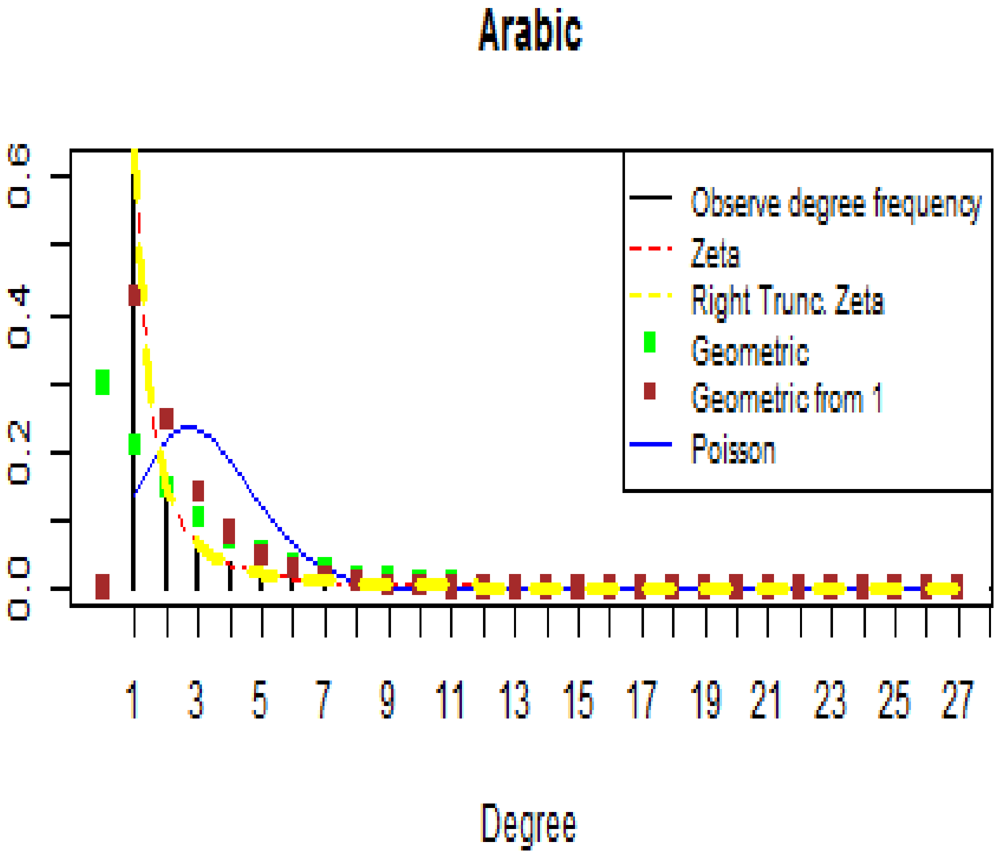

English
0.0
0.2
0.4
Observe degree frequency
Zeta
Right Trunc. Zeta
Geometric
Geometric from 1
Poisson
1
3
5
7
9
11
13
15
17
19
21
23
25
27
Degree

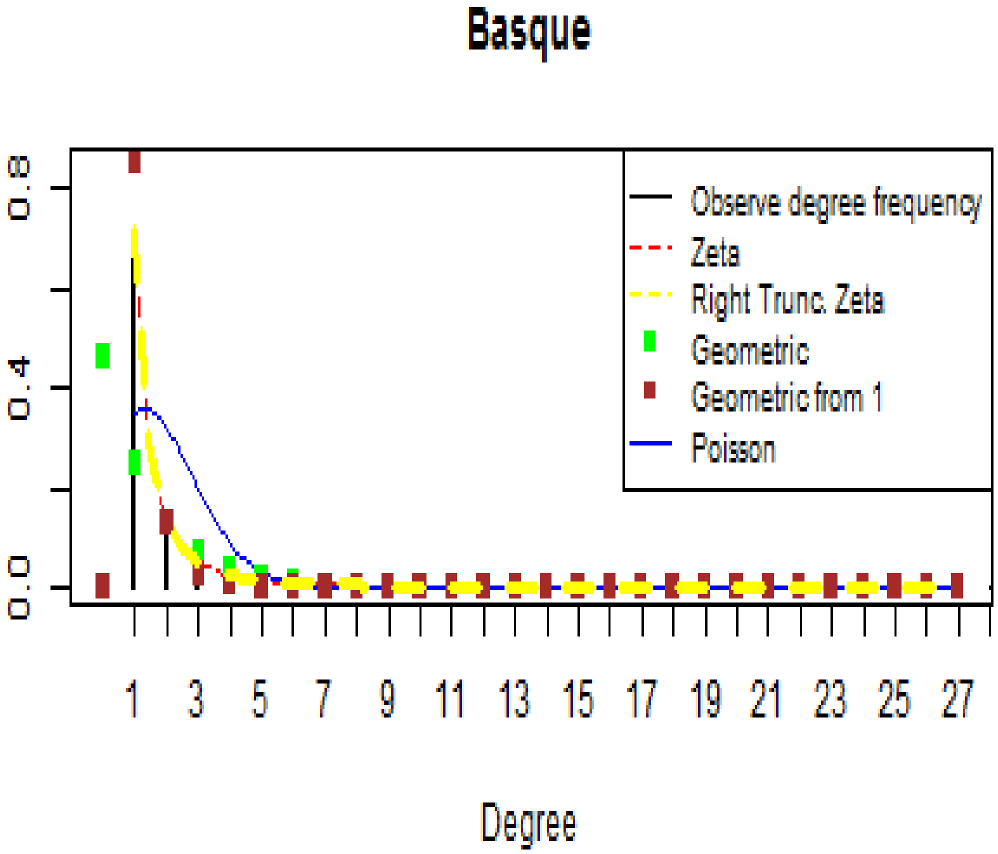

Greek
0.0
0.2
0.4
0.6
Observe degree frequency
Zeta
Right Trunc. Zeta
Geometric
Geometric from 1
Poisson
1
3
5
7
9
11
13
15
17
19
21
23
25
27
Degree

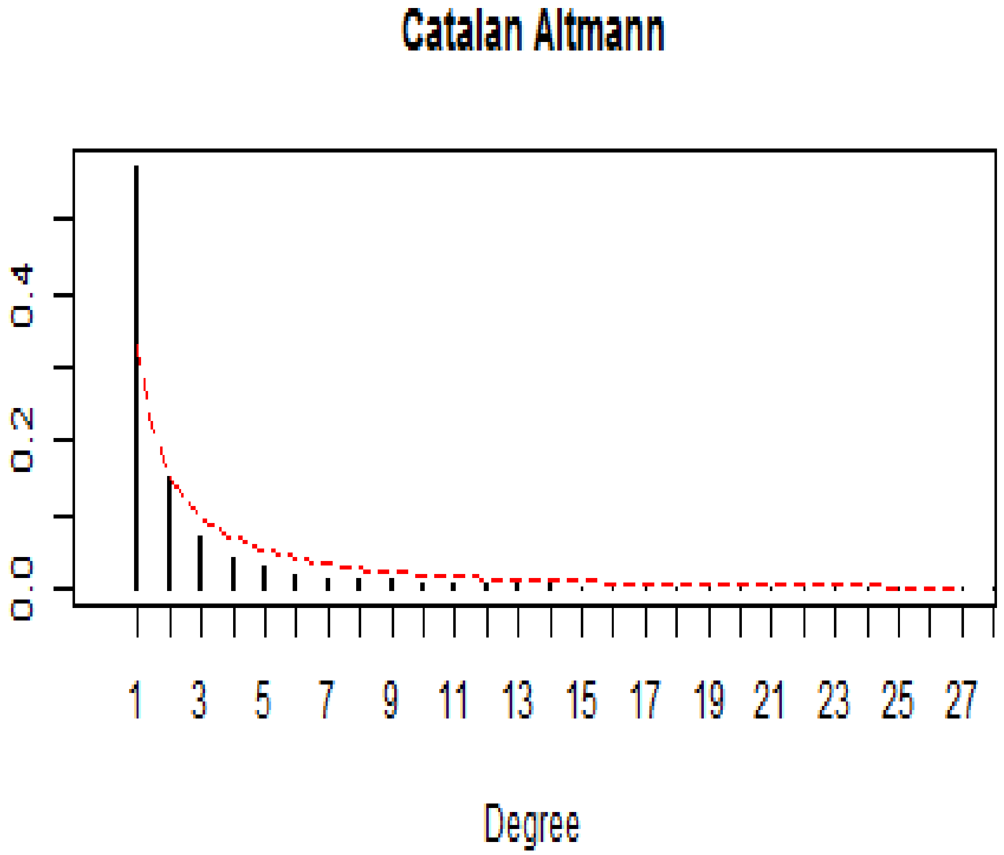

Hungarian
0.0
0.4
Observe degree frequency
Zeta
Right Trunc. Zeta
Geometric
Geometric from 1
Poisson
1
3
5
7
9
11
13
15
17
19
21
23
25
27
Degree

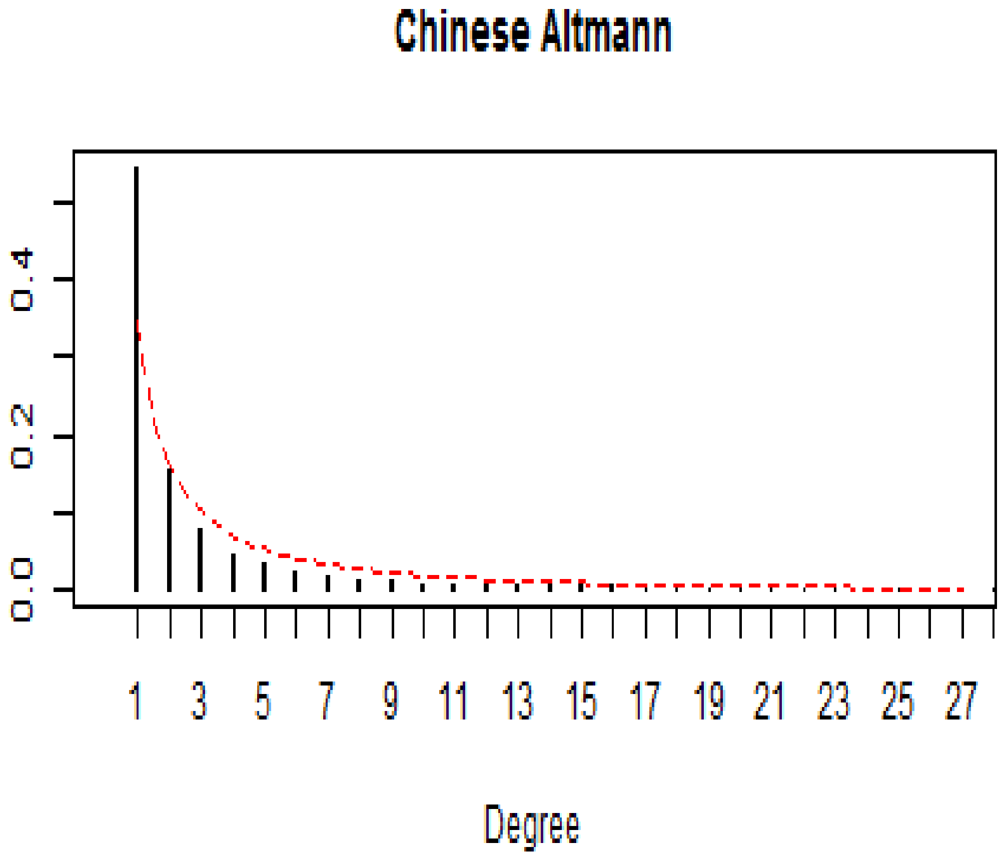

Italian
0.0
0.2
0.4
0.6
1
3
5
7
9
11
13
15
17
19
21
23
25
27
Degree
Observe degree frequency
Zeta
Right Trunc. Zeta
Geometric
Geometric from 1
Poisson

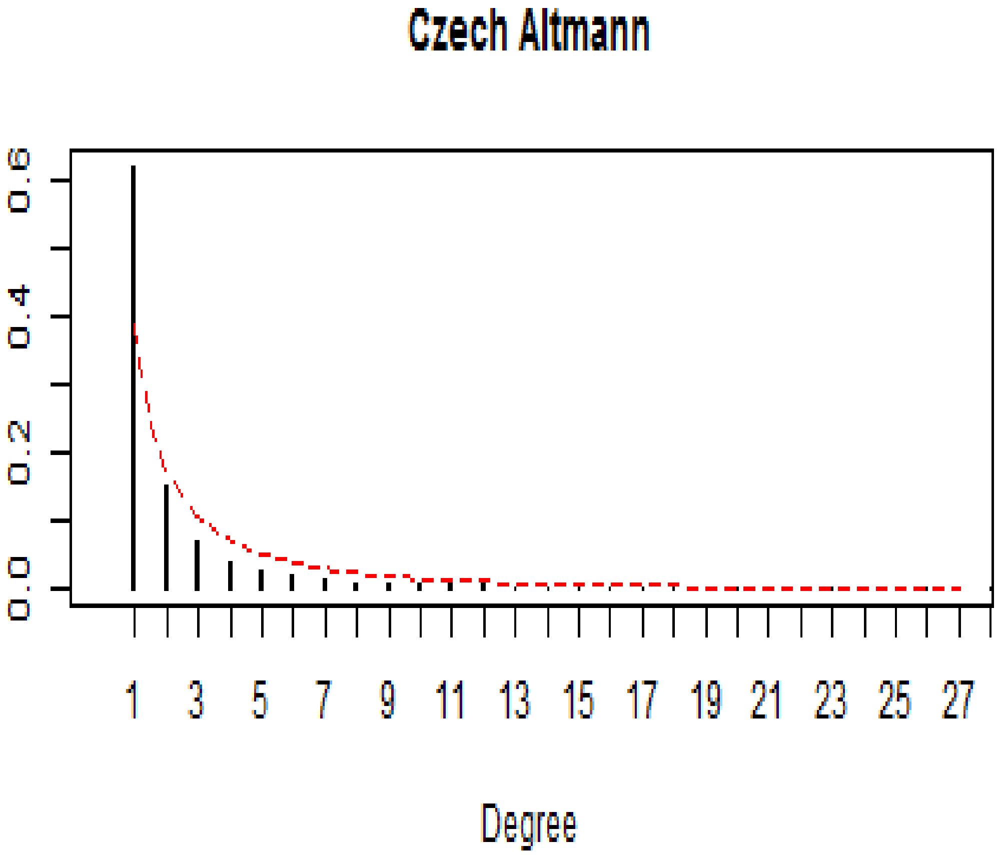

Turkish
0.0
0.4
1
3
5
7
9
11
13
15
17
19
21
23
25
27
Degree
Observe degree frequency
Zeta
Right Trunc. Zeta
Geometric
Geometric from 1
Poisson

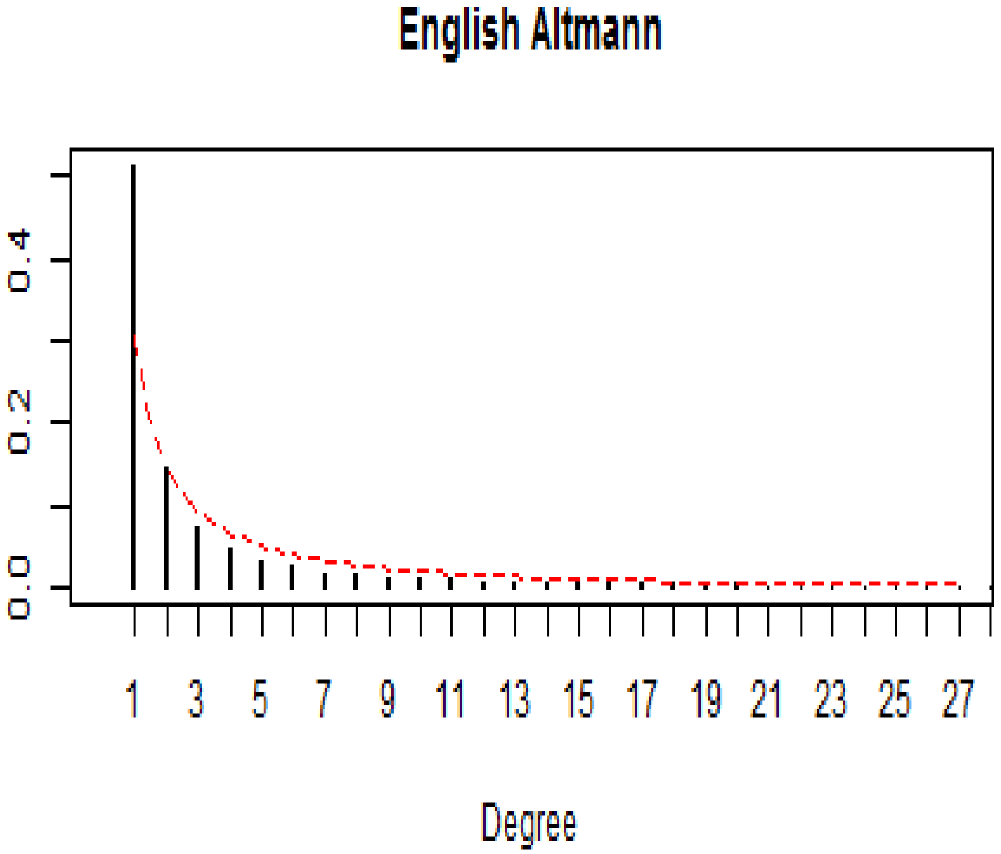

Arabic
0.0
0.2
0.4
0.6
Observe degree frequency
Zeta
Right Trunc. Zeta
Geometric
Geometric from 1
Poisson
1 3 5 7 9 11 13 15 17 19 21 23 25 27
Degree

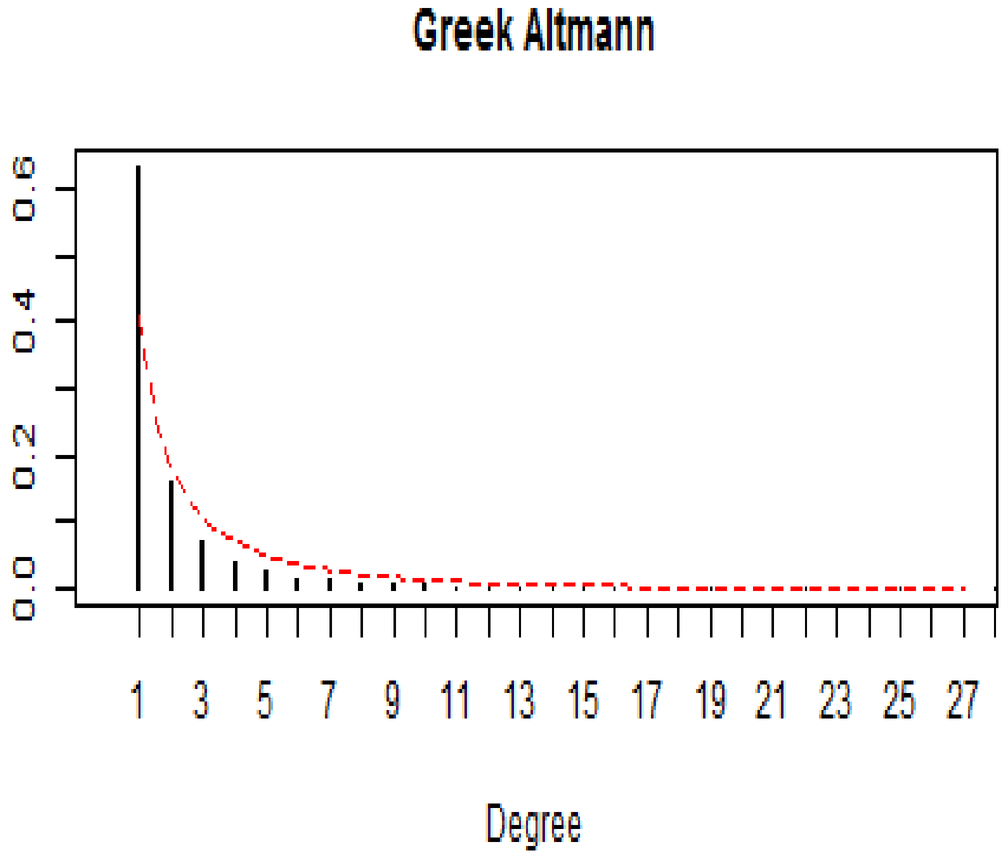

Basque
0.0
0.4
0.8
Observe degree frequency
Zeta
Right Trunc. Zeta
Geometric
Geometric from 1
Poisson
1 3 5 7 9 11 13 15 17 19 21 23 25 27
Degree


Catalan Altmann
0.0
0.2
0.4
1
3
5
7
9
11
13
15
17
19
21
23
25
27
Degree

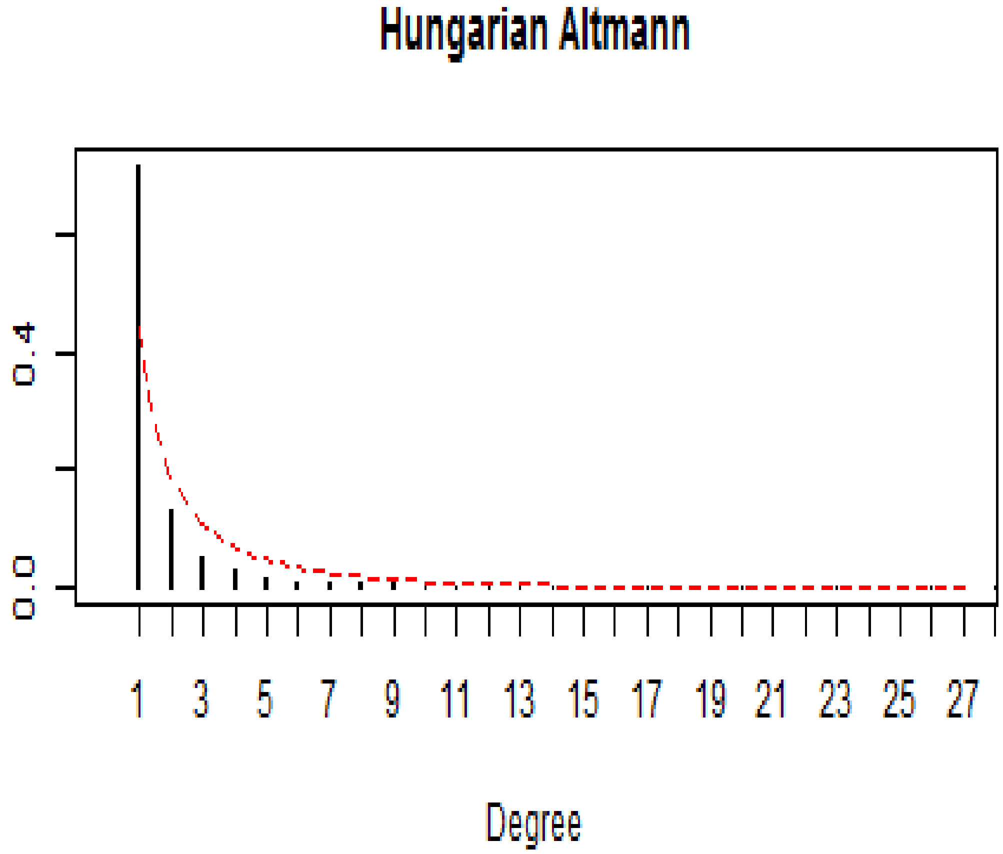

Chinese Altmann
0.0
0.2
0.4
1
3
5
7
9
11
13
15
17
19
21
23
25
27
Degree


Czech Altmann
0.0
0.2
0.4
0.6
1
3
5
7
9
11
13
15
17
19
21
23
25
27
Degree

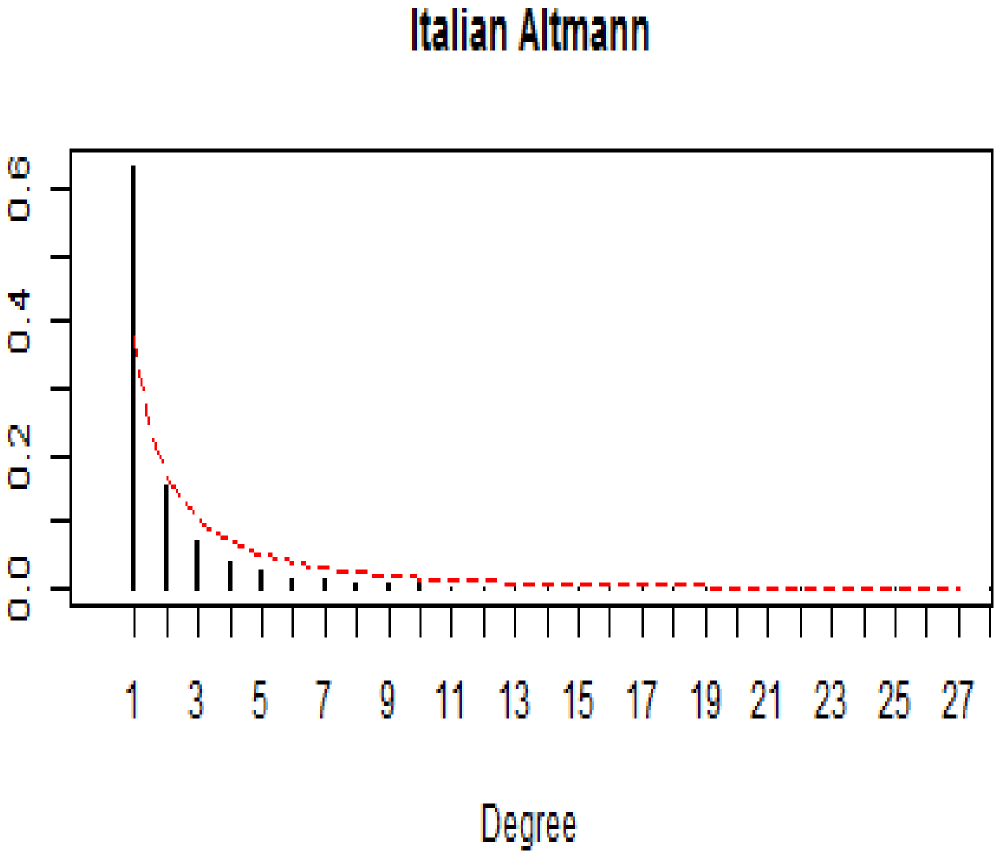

English Altmann
0.0
0.2
0.4
1
3
5
7
9
11
13
15
17
19
21
23
25
27
Degree



Greek Altmann



Hungarian Altmann


Italian Altmann
0.0
0.2
0.4
0.6
1
3
5
7
9
11
13
15
17
19
21
23
25
27
Degree

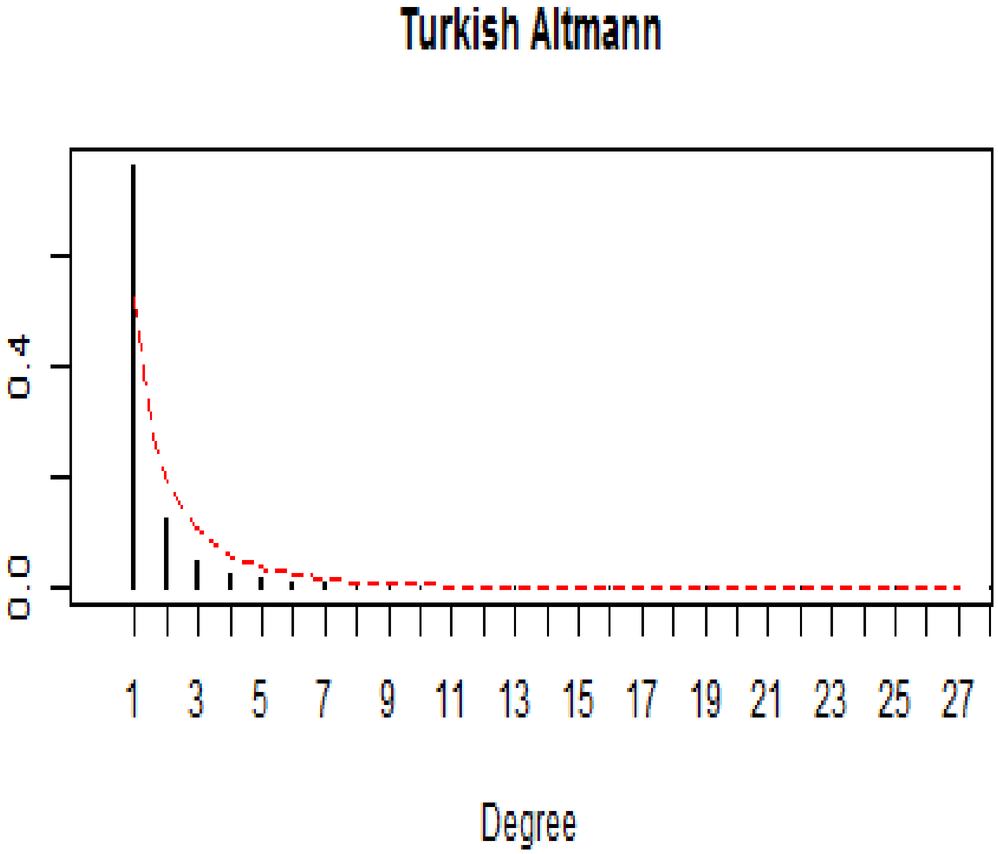

Turkish Altmann
0.0
0.4
1
3
5
7
9
11
13
15
17
19
21
23
25
27
Degree

Table 3: Difference in AIC for the 5 different models in different languages

	zeta	zeta_2	RT_zeta	geom	poisson	geom_corrected	Altmann
Arabic	132092.50	132264.87	132089.59	156278.54	372410.85	137774.84	0.00
Basque	58151.24	58996.42	58151.01	66515.12	108315.60	42796.94	0.00
Catalan	271389.13	271601.08	271375.18	333256.91	1185255.26	318171.69	0.00
Chinese	277583.60	278072.90	277568.91	326247.73	895917.31	309159.19	0.00
Czech	439807.88	439946.31	439799.03	530898.42	1380476.74	485598.28	0.00
English	251394.54	252819.15	251349.49	297091.67	990930.96	287081.79	0.00
Greek	78708.62	78869.30	78705.19	95839.03	235842.43	85490.72	0.00
Hungarian	185034.54	187254.82	185036.24	238504.16	653286.73	204135.21	0.00
Italian	86617.65	86735.54	86617.17	109495.06	332420.68	101040.66	0.00
Turkish	91486.62	94227.09	91488.60	116101.97	284831.65	80961.78	0.00

3 Discussion

When optimising model 3 we noticed that the value of k_{max} stays its starting value and does not change, while manually changing the starting value of k_{max} and thus the value obtained by optimising, significantly improved the AIC-value and the RSS-value. For the english-in-degree-sequence reduces changing k_{max} manually from max degree to 778 the RSS by over 30%.

3.1 Conclusion

Best model it's the Altmann one, the zeta function respects to the previous works better because behave good on the tails, the geometric distribution, even if shifted to 1, it 'loose' too much probability on the tails! Finally as was expected, the AIC has changed as soon as you introduce a better model.

4 Methods

When using the "L-BFGS-B" optimisation method for the right truncated zeta model (model 3), we often encountered errors. We decided to change the optimisation method to the conjugate gradients method "CG". Here we couldn't set any lower bound, but that didn't seem to give any problem for our data.