Complex and Social Networks: Lab session 6

Network dynamics

Sergio Mosquera & Egon Ferri

Contents

1 Introduction

In this session, we are going to simulate different network growth models and analyse their properties from a statistical perspective. This session has three goals: achieving a better understanding of the dynamical principles behind the Barab´asi-Albert model, improving our simulation skills, and applying curve fitting methods (model selection).

This session consists of simulating and analyzing the mathematical properties of the Barabasi-Albert model and two modified versions: one where preferential attachment is replaced by random attachment and another where vertex growth is suppressed [Barabasi et al., 1999]. For every variant we will have to study the growth of vertex degrees over time and their degree distribution.

2 Results

2.1 The Barabasi-Albert model: growth + preferential attachment

The parameters that we choose, that seem to give us good results, are

$m_0 = 1$ $n_0 = 100$

With a lattice as a starting graph.

2.1.1 Vertex growth degree

2.1.1.1 Check visually if the rescaled variant of vertex growth is about the same for every vertex chosen

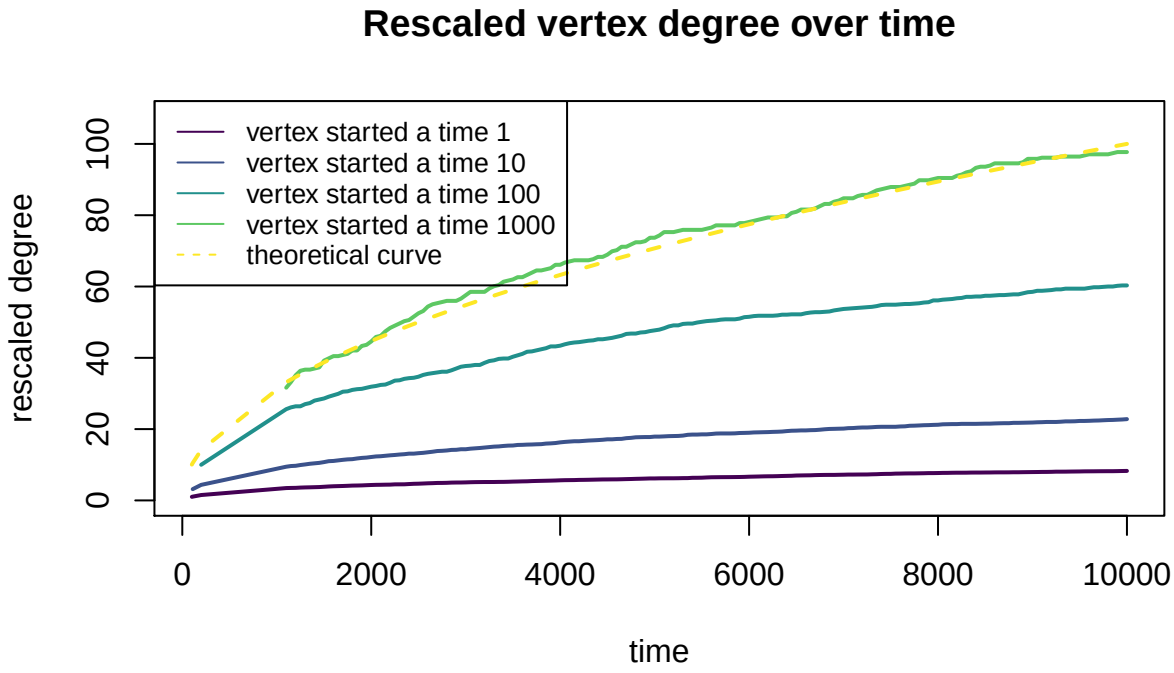
The growth of k_i, the degree of the i-th vertex, as a function of time obeys:

$$k_i(t) \approx m_0 \left(\frac{t}{t_i}\right)^{1/2}$$

This means that:

$$k_i'(t) = t_i^{1/2} k_i(t) \approx m_0 t^{1/2}$$

should be about the same for every vertex, regardless of its arrival time.



We can see that our theoretical result is not really confirmed graphically . The theoretical curve (exponent 1/2) seems a good approximation only to the scaled timeseries that starts at node 1000, and for all the others overfits. We tried to see if once we reach an high starting time the result was stable, but for t_i with $i > 1000$ it underfits. We tried to explore a little bit to understand what's going on, but with scarse results. We discovered that the starting number of the lattice influences the timeseries a bit, but that does not seems the main problem.

2.1.1.2 Model fit

Table 1: Akaike information criterion of each model

	0	1	2	3	4	0+	1+	2+	3+	4+
timeserie1	603.7190	16.13203	-377.0671	-11.79627	344.34072	-204.64018	-453.3365	-481.8214	-414.9303	-437.4308
timeserie2	695.2858	155.70477	-339.1372	199.58954	371.95926	87.40495	-210.2310	-735.7529	-483.4137	-730.1427
timeserie3	510.4377	-45.98340	-633.6469	-69.20611	215.41434	-232.79351	-648.8304	-658.8469	-677.8678	-678.2926
timeserie4	248.4257	-450.40078	-531.4087	-129.18927	77.68543	-250.92784	-504.7668	-683.7835	-641.6303	-690.2140

Table 2: AIC differences with respect to the best AIC in our ensemble of models

	0	1	2	3	4	0+	1+	2+	3+	4+
timeserie1	1085.5404	497.9535	104.75431	470.0252	826.1621	277.1812	28.48497	0.000000	66.8910853	44.390618
timeserie2	1431.0387	891.4576	396.61568	935.3424	1107.7121	823.1578	525.52190	0.000000	252.3392044	5.610159
timeserie3	1188.7303	632.3092	44.64568	609.0865	893.7070	445.4991	29.46225	19.445740	0.4248019	0.000000
timeserie4	938.6397	239.8132	158.80529	561.0247	767.8994	439.2861	185.44714	6.430514	48.5837168	0.000000

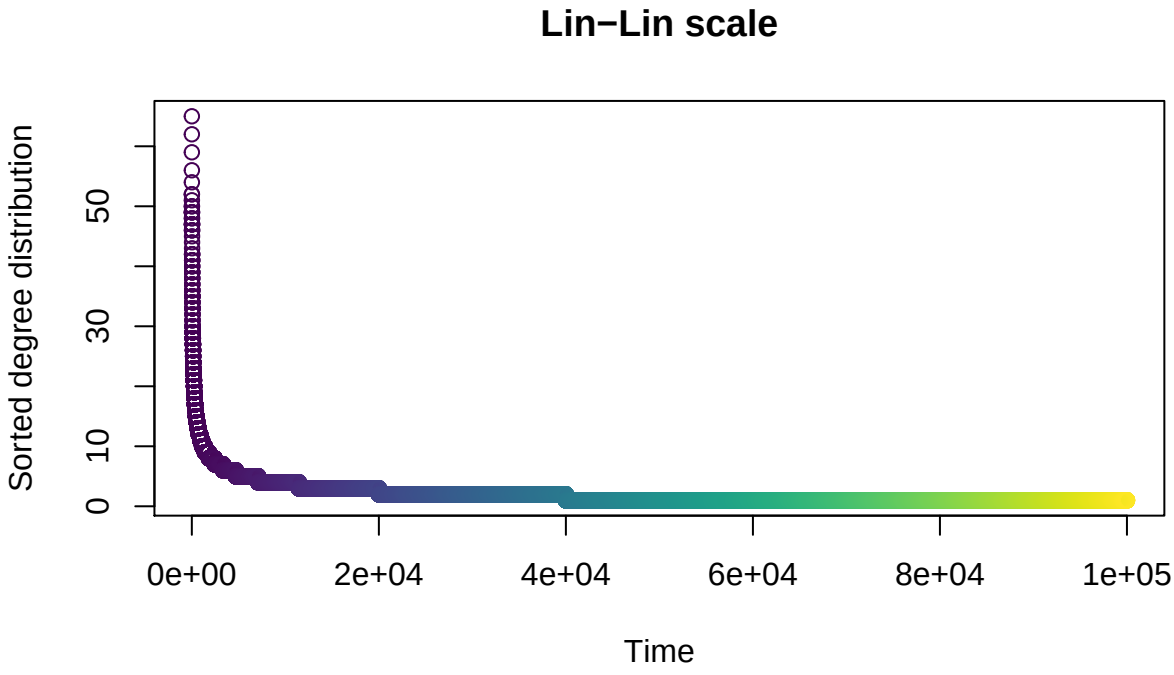
Table 3: best fittings for model parameters

	timeserie1	timeserie2	timeserie3	timeserie4
0: a	0.0009226	0.0011292	0.0006858	0.0003709
1: a	0.0777599	0.0954136	0.0578922	0.0311543
2: a	0.1734452	0.2524949	0.1470972	0.0441538
2: b	0.4081249	0.3885280	0.3931883	0.4600865
3: a	3.5769748	4.5385655	2.7223626	1.3642521
3: c	0.0000795	0.0000744	0.0000762	0.0000866
4: a	0.7085048	0.8711034	0.5279360	0.2832841
4: d	-1127.0045532	-1130.8676366	-1125.0793278	-1138.0296248
0+: a	0.0004639	0.0005426	0.0003337	0.0002061
0+: d	3.0998899	3.9643868	2.3797177	1.1138398
1+: a	0.0650544	0.0771826	0.0470089	0.0293025
1+: d	0.9788880	1.4045942	0.8385004	0.1426766
2+: a	0.0267659	3.6973666	0.0735179	1.0757307
2+: b	0.5863867	0.1685381	0.4574619	0.1875749
2+: d	1.6000929	-8.5410508	0.5542739	-3.0536258
3+: a	-8.7205147	-7.6561223	-5.2529956	-2.9099278
3+: c	-0.0000836	-0.0001931	-0.0001214	-0.0001908
3+: d	11.2848923	9.9320271	7.0415381	3.3927460
4+: a	5.3515936	3.0266033	2.8594371	1.1781810
4+: d	6310.3096306	728.4422568	3458.5938903	846.9467777
4+: d2	-44.3973522	-19.1878820	-21.6804618	-7.9543842

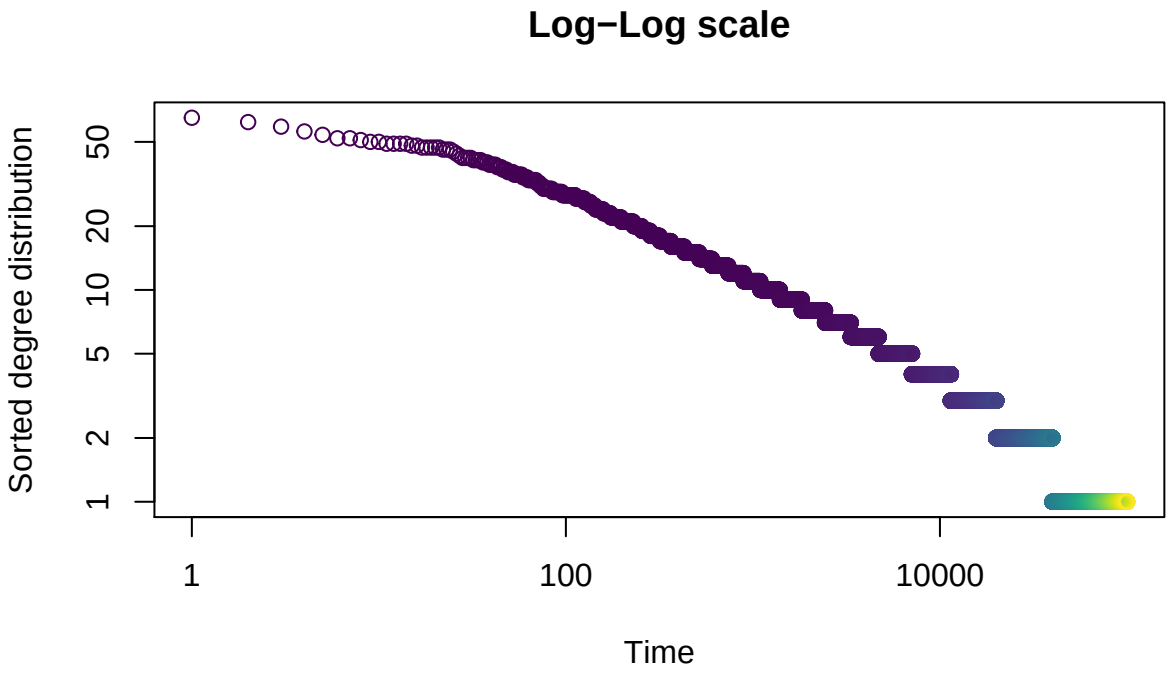
As discussed, model 1 should be a good option to represent the time series. Despite this, the AIC results show a different situation, instead it shows that the models 2+ outperforms for timeseries 1 and 2, while 4+ does the same for timeseries 3 and 4. So in this case for timeseries starting at further moments in time the model with the logarithmic approach performs better than a potential one, while with smaller values is the other way round. At a first glance, we may assume that this behavior will be kept if the simulation was larger and we took more measures at further points of time, but in that case for the values that now are fitted better by model 4+, may be represented by model 2+ instead. But this assumption has to take into account the growth of the degree distribution in the Barabasi Albert graph, which, as can be seen in above plots, gets slower in time.

Also we obtain the optimum parameters for each function from the proposed ensemble.

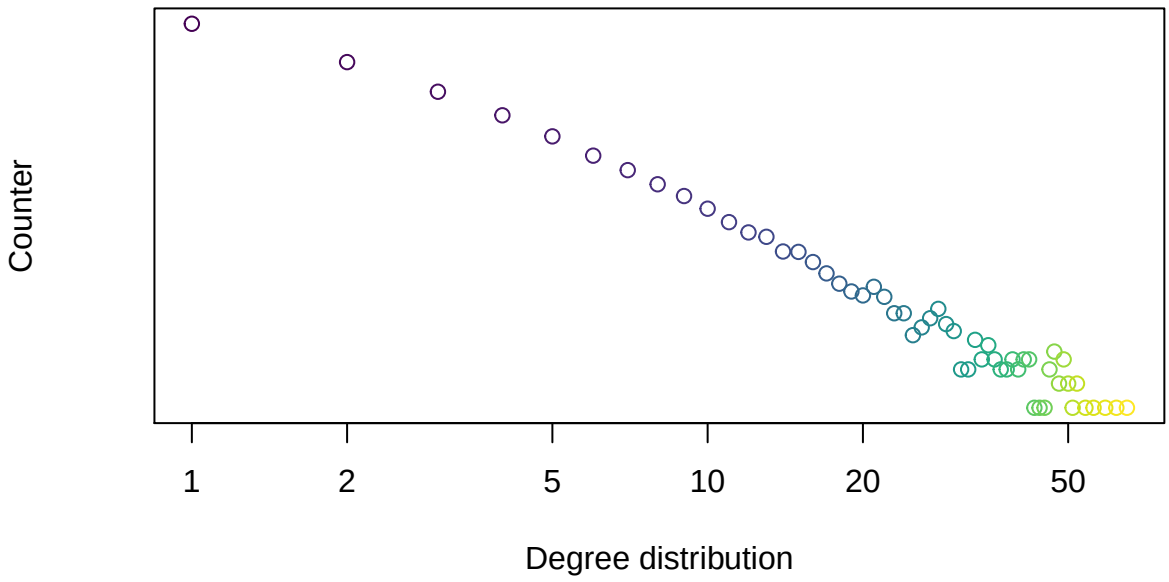
2.1.2 Degree Distribution



The graphical representation with the above plot let us check that there are insights of a Power Law in the degree distribution. This can be confirmed by the following plot using the log-log scale.



For this plot we are taking the logarithm of the data in order to show whether we have a Power Law or not in our data. As we can see in the graphical representation, the linearity of the data suggests the existence of this law. Next steps consist in checking which is the model that better fits the funcion giving the representation of this phenomenon.



2.1.2.1 Model Fitting

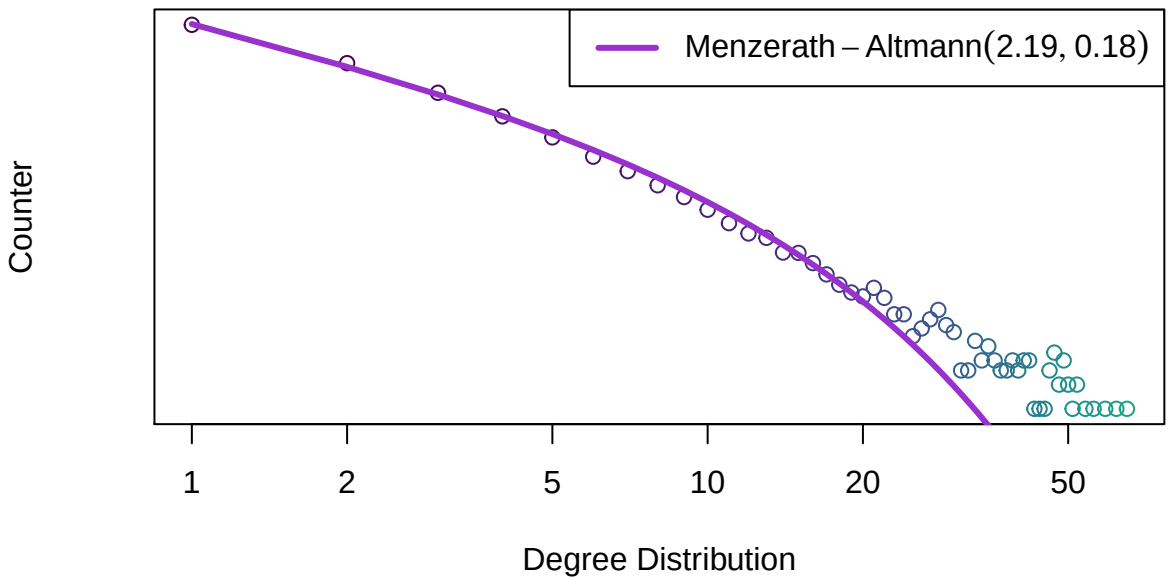
Table 4: AIC differences with respect to the best AIC in our ensemble of models

Displ. Poisson	Displ. Geom	Restricted Zeta	Zeta	R-T Zeta	Menzerath-Altmann
422694.2	277258.1	296404.8	270200.1	269372.8	264460.1

Table 5: best fit for model parametrs

1: λ	2: q	4: γ	5 : γ_2	$kmax$	6: γ_3	δ
1.593598	0.500005	2.187838	2.154483	66	1.510042	0.1804215

From the AIC table we derive that the model providing the best fitting is the Altmann function, in spite of the suggested one which was the Right Truncated Zeta with exponent -3. In addition to that, the optimum parameter for γ chosen by the Zeta distribution is far from the suggested one, instead we get around 2.15.



We have concluded analytically that the Menzerath-Altmann function is the one providing the best fitting for the data, but with the above plot we can also confirm this graphically. The fitting is almost perfect if we omit the dispersion that can be found at the end of the distribution.

2.2 Growth + random attachment

2.2.1 Vertex growth degree

2.2.1.1 Check visually if the rescaled variant of vertex growth is about the same for every vertex chosen

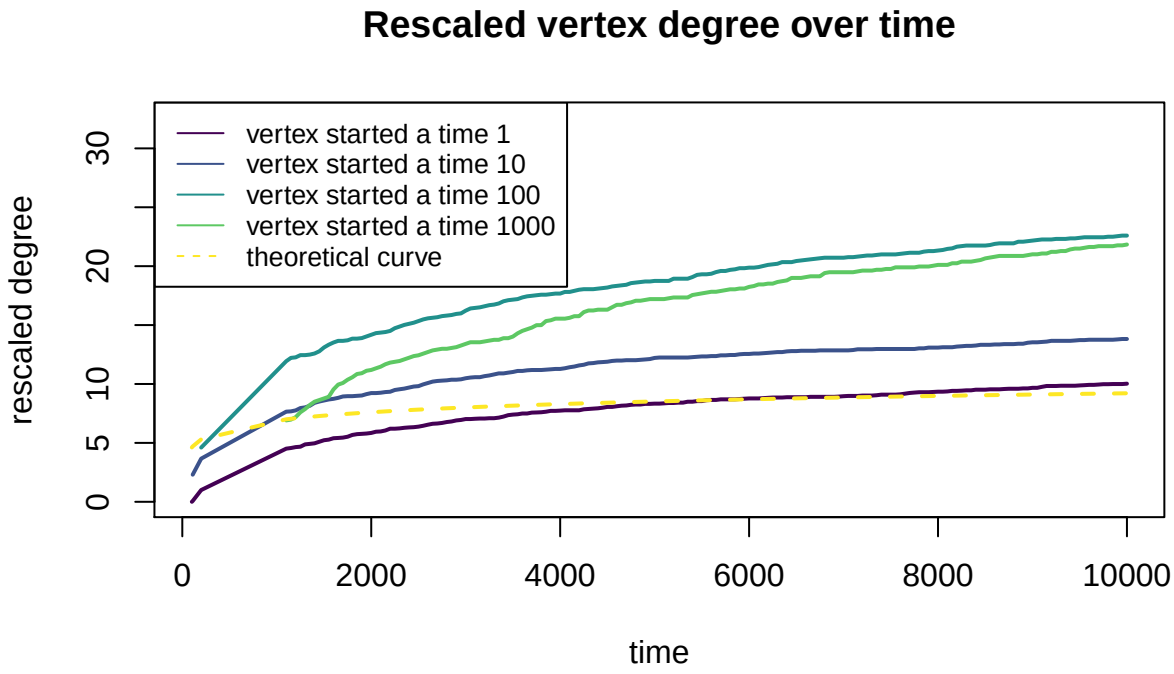
If preferential attachment is replaced by random attachment, the growth of ki (the degree of the i-th vertex) as a function of time obeys [Barab´asi et al., 1999]

$$k_i(t) \approx m_0(log(m_0 + t - 1) - log(n_0 + t_i - 1) + 1)$$

This means that

$$k_i''(t) = k_i(t) + m_0 log(n_0 + t_i - 1) - m_0 \approx m_0 log(m_0 + t - 1).$$

For `barabasi.game` R package we can tune some parameters in order to change the behavior of the preferential attachment, such that it becomes random. The probability of a node i to generate an edge with some random node is calculated as $P(i) \sim k(i)^{power} + a$. Taking into account that we can modify the exponent, if set to 0 we will state that all the vertices have the same probability.



2.2.1.2 Model fit

Table 6: Akaike information criterion of each model

	0	1	2	3	4	0+	1+	2+	3+	4+
timeserie1	679.9732	385.9680	-671.6512	-131.65400	-297.51624	-186.6527	-409.1378	-817.9740	-735.2064	-821.1447
timeserie2	679.4167	368.1332	-577.6539	-86.90894	-199.23008	-144.7309	-359.7614	-727.2493	-533.7887	-724.7971
timeserie3	609.9533	281.4954	-637.5888	-143.52654	-155.39985	-213.4697	-451.5365	-677.6392	-555.4982	-658.2045
timeserie4	316.4559	-260.8930	-359.9584	-34.92398	90.23695	-129.9259	-326.0655	-157.9579	-694.9110	-678.1730

Table 7: AIC differences with respect to the best AIC in our ensemble of models

	0	1	2	3	4	0+	1+	2+	3+	4+
timeserie1	1501.118	1207.1127	149.49352	689.4908	523.6285	634.4920	412.0069	3.170774	85.9383	0.000000
timeserie2	1406.666	1095.3825	149.59539	640.3403	528.0192	582.5183	367.4879	0.000000	193.4606	2.452144
timeserie3	1287.592	959.1346	40.05041	534.1126	522.2393	464.1695	226.1027	0.000000	122.1410	19.434658
timeserie4	1011.367	434.0179	334.95256	659.9870	785.1479	564.9851	368.8454	536.953078	0.0000	16.738004

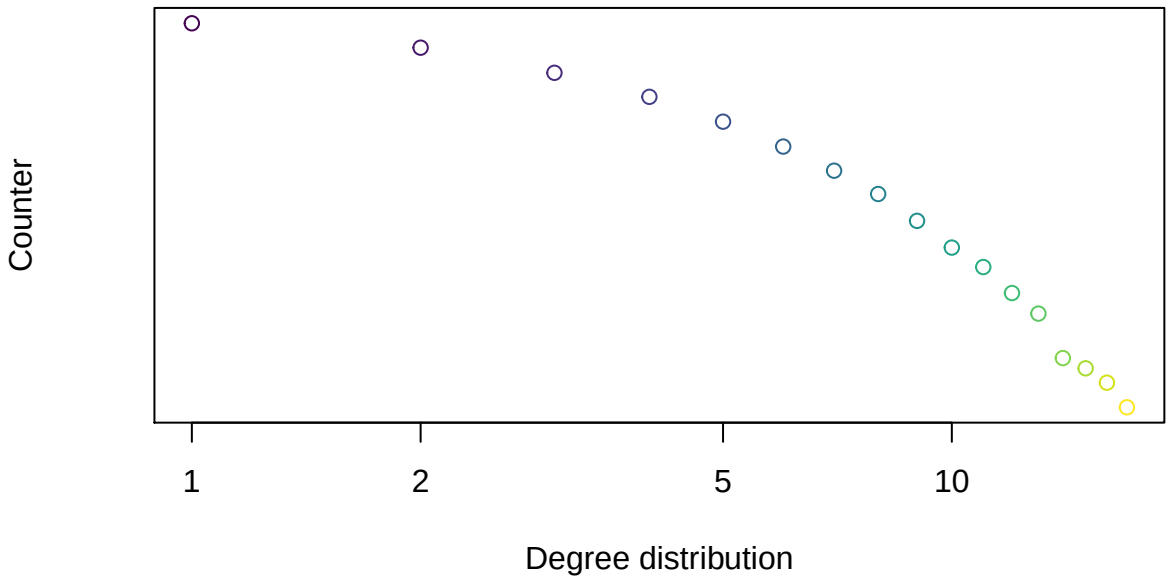
Table 8: best fittings for model parameters

	timeserie1	timeserie2	timeserie3	timeserie4
0: a	0.0007771	0.0008048	0.0006857	0.0004206
1: a	0.0667841	0.0690159	0.0586893	0.0354155
2: a	0.8874648	0.7567211	0.5429409	0.0616687
2: b	0.2030406	0.2251831	0.2447396	0.4365047
3: a	3.9728998	3.9928985	3.3102194	1.6042050
3: c	0.0000405	0.0000448	0.0000487	0.0000815
4: a	0.6109537	0.6330835	0.5384898	0.3227100
4: d	-926.2320529	-1016.9614431	-1050.8562163	-1140.4223246
0+: a	0.0002081	0.0002372	0.0002188	0.0002233
0+: d	3.8460567	3.8360854	3.1560165	1.3334391
1+: a	0.0297486	0.0338830	0.0311390	0.0320218
1+: d	2.8533739	2.7067889	2.1225921	0.2614656
2+: a	25.3442689	45.6211737	2.7077826	91.3367994
2+: b	0.0300819	0.0205900	0.1290717	0.0108042
2+: d	-27.7167598	-49.1799759	-3.7452881	-97.6807027
3+: a	-2.9396249	-3.3407399	-3.0837842	-3.1925519
3+: c	-0.0002324	-0.0002284	-0.0001954	-0.0002544
3+: d	5.9617762	6.2633954	5.5490412	3.5274443
4+: a	1.0131834	1.1161142	1.1429512	1.0404504
4+: d	126.6698830	6.6446176	436.2661628	-50.5741978
4+: d2	-3.6260862	-4.3182873	-5.4411114	-6.2410807

Focusing in the AIC values for the random attachment, we see that model 2+ outperforms in 2 out of 4 cases, while the model 4+ only does it once (timeseries 1). Although model 4+ is not giving the best fit for every case (as the guide suggested), the results are still very good in average for all the timeseries, being for every case either the first or the second (with a very low difference) model giving the best fit. So in absolute terms we would not say that is the model that gives the best predictions among all the ensemble of functions, but in relative terms we can confirm that, in average, is the model that shows the best behavior when using random attachment.

Also if we relate to the parameters for the second part of the exercise, we can check that in fact the estimation of a is around m_0 (initially set to 1), but we fail to match the hypotheses when computing the value of d_2, which was supposed to be 0 but instead we get much higher values for the timeseries where it did not provide the best fit and a very low one (-50) in the timeseries in which it outperforms.

2.2.2 Degree Distribution



2.2.2.1 Model Fitting

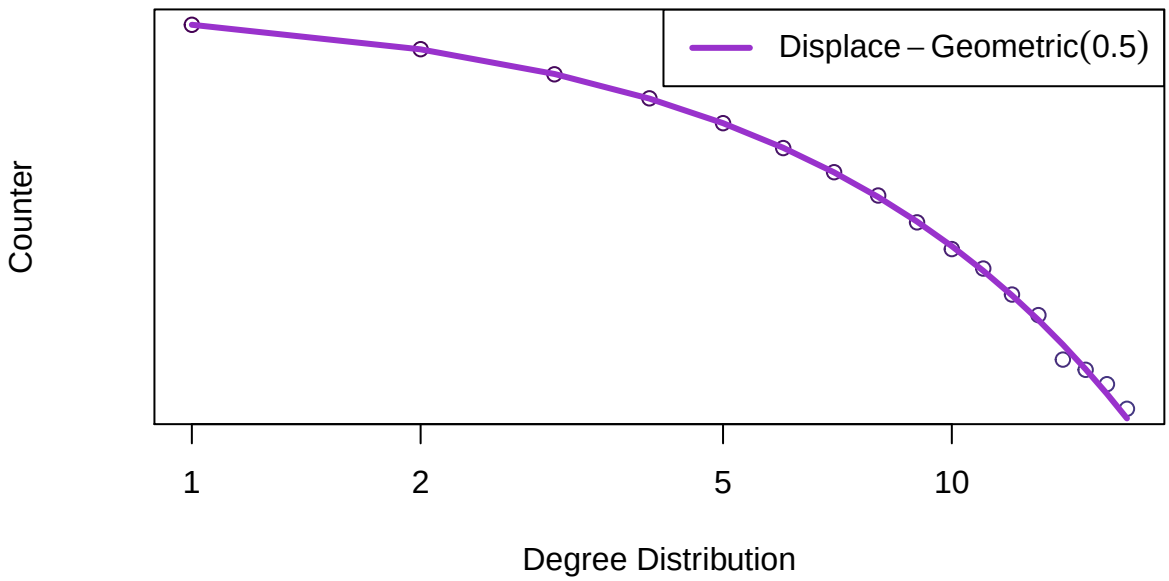
Table 9: AIC differences with respect to the best AIC in our ensemble of models

Displ. Poisson	Displ. Geom	Restricted Zeta	Zeta	R-T Zeta	Menzerath-Altmann
359261.5	277258.1	341541.7	302245.7	295496.1	277260.1

Table 10: best fit for model parametrs

1: λ	2: q	4: γ	5 : γ_2	$kmax$	6: γ_3	δ
1.593598	0.500005	2.076079	1.899167	18	0	0.6931575

Following the results of the AIC, we can see that the model giving the best fit with respect to the models now is no longer a power law, in this case the model giving the best fit is the Displaced Geometric one. It is important to remark that the Altmann function keeps being a good model for the prediction because the AIC results are quite similar to the best model.



In the same way as we have done for the previous section, we want to support our analytical conclusions about the model fitting the best the current behavior by adding a graphical representation. And again, this plot gives us more confidence to confirm the hypothesis that the Geometric distribution is the one that predicst the best the random attachment phenomenon.

3 Conclusions

This lab session allows us to make experiments over the Barabasi-Albert model and how the different approaches on the nodes attachment influences the network growth. In this case we have followed two different paths:

First of all we ran the experiments following a preferential attachment strategy, the usual one in this model. We have taken measures at different points of time and realized that the growth of the vertices in the network was likely to be following a Power Law. Then, using an ensemble of functions we have checked our previous discovery and found out that the Altmann function was providing a very good fit for the evolution in the degree sequence of the graph at the different points of time.

The second part of the experimentation used the random attachment strategy, in which every vertex has the same probability of obtaining a new edge in the next step of the graph generation. This strategy changes the experimentation drastically, instead we needed to show that this approach does not follow a Power Law. To do this we made the same experiments than before and have checked that the function providing the best fit was the Geometric distribution. So, there was no Power Law using the random attachment.