CSN: lab 6

Sergio H. Martínez Mateu, Martín Dans

28/11/2019

Introduction

The goal of this laboratory work is to study some of the properties of dynamic networks. We will be looking at two main aspects, the evolution of the degrees over time and the final degree distribution. Three different versions dynamic network generating models are considered: the Barabasi-Albert, a variant where preferential attachment is substituted by random attachment and a third variant where there preferential attachment but no vertex growth.

Results

In the following table, we show the parameters that we used in our simulations of network generating models. m0 is the number of edges added at each time, n0 the number of initial vertices and the initial configuration means the type of graph that is used for the initial vertices.

Degree evolution

The first thing that we want to check is that the evolution of the degrees over the time generation steps follow approximately the corresponding theoretical curves. To do this, we rescaled the degree for which the curve should be indenpendent of the arrival time, i.e., similar for each of the vertices. At the same time, we plot the theoretical curve.

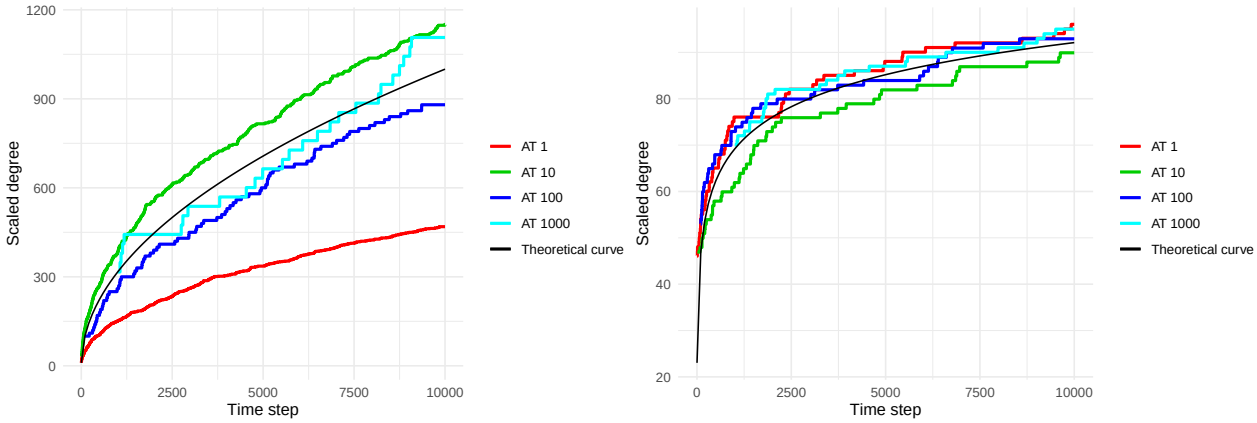
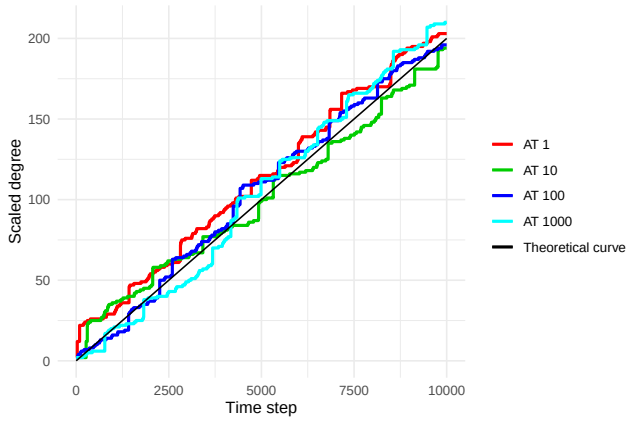


Table 1: Initial settings for the simulation algorithm

Version	m0	n0	Initial_Conf
Barabasi-Albert	10	10	Ring
Random Attachment	10	100	Ring
No Growth	10	1000	Ring



From the previous plots we conclude that our simulations worked well since the empirical curves follow approximately the theoretical curve. After running the simulations several times we observed the relatively large deviations observed in some cases is due to random variability, because all 4 trajectories appear sometimes above and sometimes below the theoretical curve. This could have been shown explicitly by computing the average of several simulations, but we had to prioritize other parts of the work.

Next we show the results of the model selection process based on the AIC of the fitted models. At the beggining we were not sure if we had to fit the evolution model for the chosen vertices individually or all at once. For this reason, we did both versions of the analysis.

Table 2: AIC difference results for the time series data

	Version	ArrivalTime	0	1	2	3	4
1	BA	1	59381.8	871.2	122.4	345796.8	61348.4
4	BA	10	58655.7	10525.6	29.9	347181.4	58308.2
7	BA	100	40242.2	8956.1	637.6	262028.3	46853.2
10	BA	1000	25462.1	15318.1	12291.7	296717.0	31382.1
2	BA_RA	1	69373.1	57998.6	1902.1	316285.1	382.8
5	BA_RA	10	65025.2	52897.7	2757.8	341991.1	6846.0
8	BA_RA	100	63308.3	52160.3	562.6	272794.9	1519.4
11	BA_RA	1000	62705.4	50964.6	1966.4	269834.9	1861.5
3	BA_NG	1	14494.4	32064.1	4877.4	355846.9	49293.1
6	BA_NG	10	10051.6	26414.1	4196.1	248045.4	43160.6
9	BA_NG	100	5119.9	35732.7	3674.5	317283.2	50547.5
12	BA_NG	1000	5472.1	33817.9	462.4	330718.8	46066.7

The results are more or less consistent stratifying by vertex, as expected by the preliminar plots. Therefore, we can focus on the results for all 4 vertices at once. Basically we obtain that for both the Barabasi-Albert and the preferential attachment without vertex growth model 2+ is the best fit. For the random attachment with vertex growth we have that the best evolution model is 4+. In order to see if the fitted models make sense, we show again the evolution plots but with the best fitting curve.

Table 3: AIC difference results for the time series data

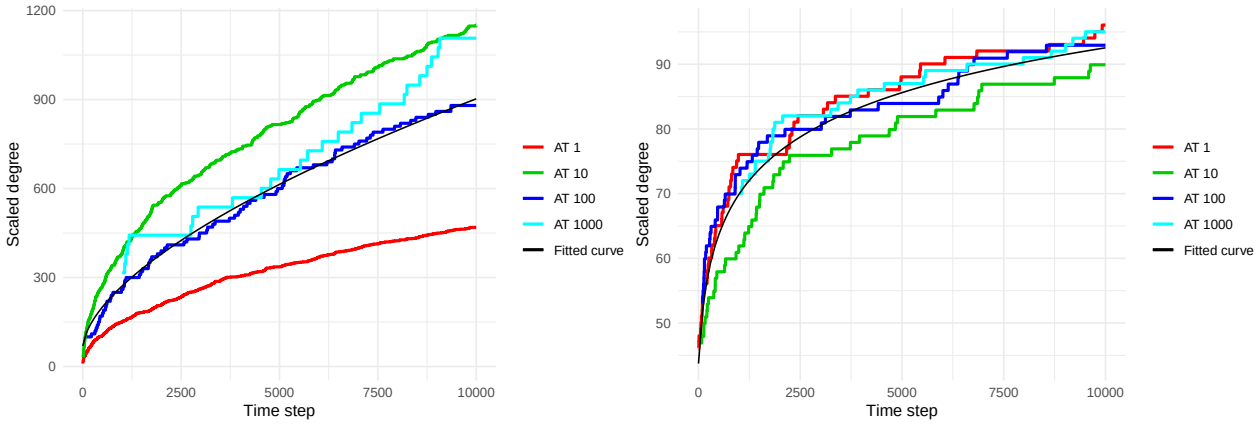
	Version	ArrivalTime	0+	1+	2+	3+	4+
1	BA	1	37949.4	387.0	0.0	339284.3	5563.9
4	BA	10	35767.5	2026.9	0.0	347152.4	12029.4
7	BA	100	17476.4	3117.6	0.0	311614.4	2069.1
10	BA	1000	4656.3	13900.9	0.0	299971.2	5885.6
2	BA_RA	1	26666.5	17399.0	1762.3	316317.7	0.0
5	BA_RA	10	23739.1	13129.4	2517.2	344178.8	0.0
8	BA_RA	100	12197.7	3974.3	0.0	307792.4	1510.8
11	BA_RA	1000	11630.8	6342.5	1536.8	277626.8	0.0
3	BA_NG	1	26.4	22511.0	0.0	346669.7	279.9
6	BA_NG	10	254.6	17585.5	0.0	270899.7	1620.8
9	BA_NG	100	4964.0	20338.9	1841.1	309157.8	0.0
12	BA_NG	1000	666.8	21142.8	0.0	259283.0	1616.3

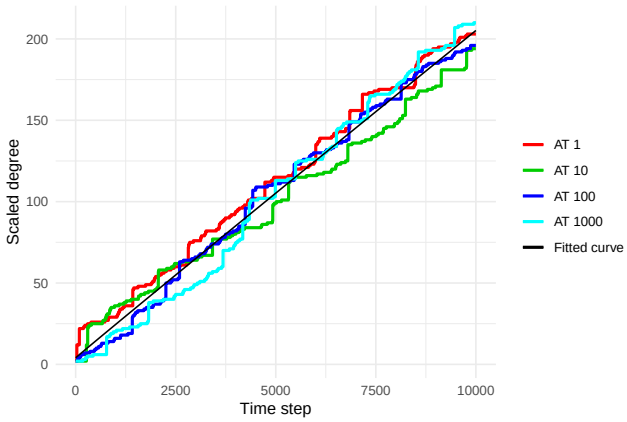
Table 4: AIC difference results for the time series data

	0	1	2	3	4
BA	12429.030	147.6561	56.71293	1045113	1.848458e+04
BA_RA	188762.870	143104.4694	445.36802	1135618	6.418389e+00
BA_NG	2982.011	82312.1740	549.61815	1206563	1.395782e+05

Table 5: AIC difference results for the time series data

	0+	1+	2+	3+	4+
BA	845.83027	102.2522	0.0000	1064885	183.0176
BA_RA	31811.70467	13560.4877	397.6146	1203702	0.0000
BA_NG	15.75837	42572.0179	0.0000	1075374	290.2307





Indeed, the fitted curve describes the mean behaviour of the data. On the other hand, we observe that the models that were selected are of the same family than the theoretical curves, matching also with the expectations. Even though it is not shown here, we found that the fitted parameters are also close to the theoretical ones, but not the same. For example for the Barabasi-Albert model our exponent is 0.61, instead of 0.5. Intercepts were chosen by the selected models, and they are compatible with the theoretical formula, which is actually an approximation.

Degree distribution

Here we show the AIC differences for the fitted degree distribution models.

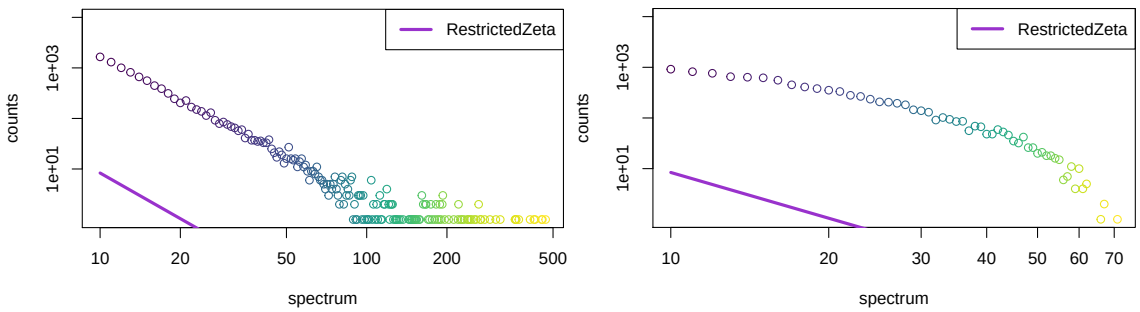
Table 6: AIC difference results for the degree distribution data

	DisplacedPoisson	DisplacedGeometric	RestrictedZeta	Zeta
Barabasi-Albert	133591.60	47326.52	0.00	67860.45
Random Attachment	58073.06	47874.14	0.00	71313.40
No Growth	0.00	730.17	20260.47	4246.86

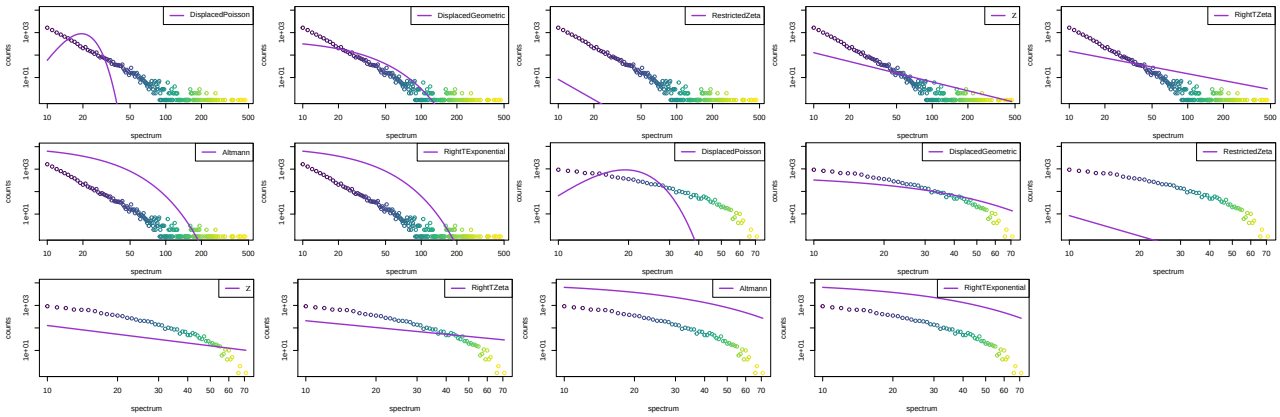
Table 7: AIC difference results for the degree distribution data

	RightTZeta	Altmann	RightTExponential
Barabasi-Albert	61866.80	47328.52	47326.52
Random Attachment	58054.01	47876.14	47874.14
No Growth	2396.35	716.66	714.59

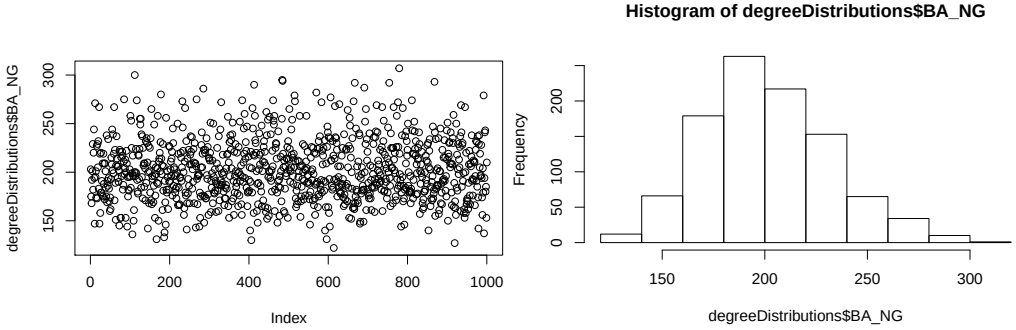
As we see, the selected model for the Barabasi-Albert model is the Zeta distribution with the fixed parameter gamma, the exponent, fixed to 3. Unexpectedly, the same degree distribution model is selected for the random attachment version. For these cases, the plots are:



But obviously this is not correct. We have tried very hard to fix these problems, but were not able to find what we are doing wrong. It is very strange because for other models that are not selected with the AIC, the fitted curve looks better:



The no-growth version, for which a gaussian distribution was expected, is closer to the displaced Poisson with a lambda parameter that matches exactly with the mean degree, which is a very large λ. This is also in agreement with the fact that the Poisson distribution converges to a Normal distribution when $\lambda \to \infty$. However, it is difficult for numerical reasons to plot the fitting curve using the probability predicted by the displaced Poisson function that we defined. Instead, we will just show that the degree distribution in this case indeed looks Gaussian.



Discussion

Regarding the time evolution of the nodes, we were able to check that the scaled degree of the nodes follow approximately the theoretical curves, with some deviations due to random variability. The fitted curves in this case were also in agreement with the theoretical formulas, being selected the models of the corresponding family but with some deviations in the parameters. Since the deviations are not so large, we think that they might be due to the small sample size and the large variability of the trajectories. We expect that with more vertices these should be more precise, and the average of the trajectories match much better to the theoretical result.

Regarding the degree distributions, we have seen that some of the fitted models give reasonable results, but we were not able to find out if our problems were due to the fitted parameters or to the visualization procedure.

Methods

In order to do the simulation we developed in C++ a small program that implemented all the simulations. We defined a class named "representation" that stores the degree of each vertex, the adjancy matrix and a vector of vertices where each vertex appears the same amount of times as it's degree, and with that, we could make fast simulations.

For creating the inital graph we tested with two different types of graph: complete graph and a ring graph achetypes. We realised that using already a medium sized complete graph, the degrees where distorsioned a lot, since the subgraph was absorving almost all the edges. Because that, we tried starting with a ring graph obtaining much better results for the simulations. We used also that graph for initialzing the "No growth" simulation.

The model choice for the time series data was based on the same procedure that we followed on lab 4. We obtained good initial parameters by using linear model versions or approximated linear versions, by taking logarithm or square root, and used the results of the corresponding linear regression as starting paramters. We did some more adjustments to avoid inifinite values. In addition, we changed the nls method from the nls to the nlsLM, which showed less convergence problems.

For the degree distributions, we followed the same procedure as in lab 2, with the additional Altmann and Right Truncated Exponential distributions (we derived the corresponding log likelihood) added to the ensemble of models. Some more adjustments to the starting parameters and the bounds were made, ensuring convergence for all the models.