

CSN 05 - Network Dynamics

Laura Cebollero, Pietro Fronte

3rd of December, 2018

1 Introduction

In this delivery we are asked to simulate three network growth models and study them from a statistical point of view.

The first one is the original Barabasi Albert model, and the other two are based on this one with some variations:

1. BA: growth & preferential attachment
2. BA': growth & random attachment
3. BA'': No growth & preferential attachment

We are going to discuss the results and the methods used, as well as cover the implementation done and problems encountered.

2 Implementation

As asked, we have implemented the simulation on these three models with a fixed n_0 , m_0 and t_{max} , their values being:

- $n_0 = 3$
- $m_0 = 2$
- $t_{max} = 10.000$

This means that for the models with growth (B.A. and B.A.''), on $t = 0$, there will exist 3 nodes fully connected. Meaning we are starting with a fully connected graph. This decision is taken in order to avoid ending up with a disconnected graph, since if we start with a disconnected graph, after the first iteration has been performed, the non-selected node will never be selected and thus we will end up with a disconnected node from the overall graph.

Each time t increases on every iteration, a stub is going to be added. Each stub will consist on a node with m_0 edges. In our case, $m_0 = 2$, so each iteration we are going to add two edges to the graph.

This will be performed until reaching t_{max} , when the total number of nodes N will be

$$N = n_0 + t_{max} = 3 + 10.000 = 10.003$$

In the third case, B.A.'' (No growth + preferential attachment), we start with 2.000 nodes and end with the same number of nodes. We have chosen a somewhat big number of nodes as suggested in the lab session statement. More specifically, the suggestion was that $N \geq 1000$, and t_{max} has been preserved as in the other two models to preserve coherency. Having a greater t_{max} would have led to a complete graph which is not what we want to achieve.

In those 2 cases where preferential attachment is what determines where each edge from the stub connects to, the higher the number of edges an existing node has, the higher the probability it has to be the one selected for the new addition.

In the first model and second model we do not have to worry on whether we are creating multiedges or not, for each stub is added in every iteration and it cannot select itself. Because of this, we have not used an

adjacency matrix and only have tracked the degree evolution of 4 selected nodes, as well as the global degree sequence when reaching t_{max} .

However, to avoid having multi-edges on the third model, we have used a sparse adjacency matrix to check whether an edge existed or not before adding it or not. The sparse decision has been taken in order to be more lenient on our computer memories. The usage of such matrix makes the execution of this simulation a little bit slower compared to the other ones.

Below is a table that represents on average the time it has taken us to execute the three simulations:

Table 1: Elapsed times on simulation runs

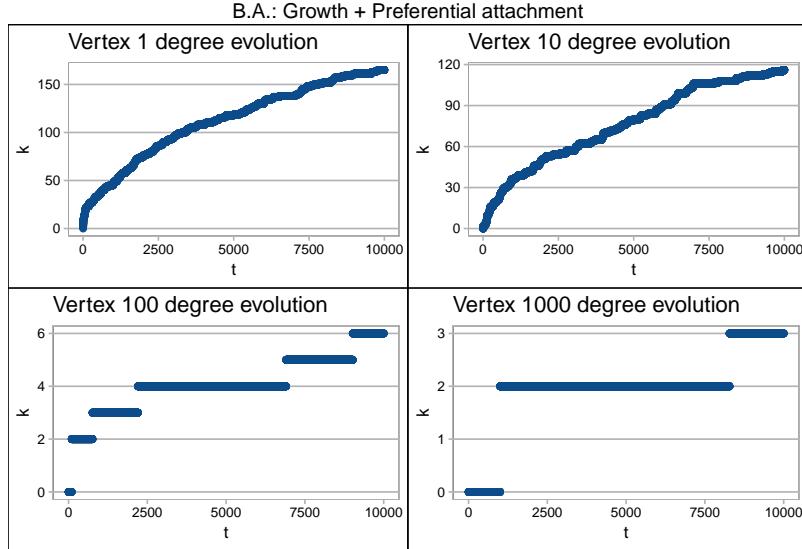
Simulation	$\Delta t(\text{seconds})$
B.A.	5
B.A.'	3
B.A."	35

As mentioned, we can see how the third simulation (No growth + preferential attachment) is the one which has more cost computationally speaking because of the usage of the adjacency matrix.

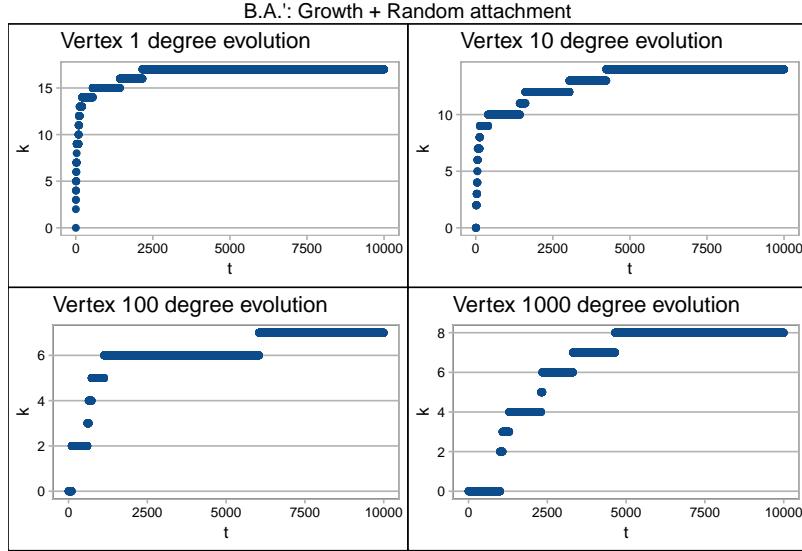
3 Results

Now we are going to see how is the evolution of 4 vertices until reaching t_{max} , as well as what distribution better fits each model.

3.1 Evolution



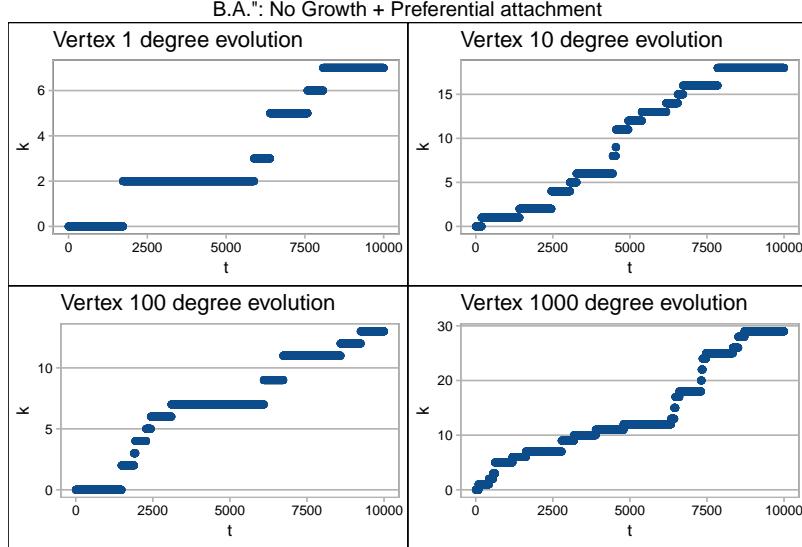
In the plot above we can observe how the first vertex is selected a lot because of the preferential attachment. As the vertex selected chosen is that from a stub added forward in time, clearly its edge degree is going to be lower, thus having a lower preferential attachment for the next iterations. Thus, vertex 100 reaches its highest point at 6 edges, and edge 1000 just 3: the ones created when the node was added as a stub and another iteration where it got selected.



Now that the random attachment is done, we can see how the vertices have more or less the same degree distribution. However, since it is a growth model the first vertices (i.e. 1 and 10) still have more chances to be selected on the first iterations, since the number of existing nodes is lower.

That's why their edge degree reach a higher point than vertices 100 and 1000.

We can also see how on the last 9000 iterations, vertex 1000 got luckier than vertex 100, for it has one more edge than the other one. Thus, we can see how the randomness plays its hand on the attachment of stubs.



In this case we can see how the no-growth and preferential attachment have their effects on the simulation model.

On the first iteration all vertices (2000) have the same chance to get selected. Afterwards, those that got lucky and got selected on the first iterations have a higher probability to get selected again.

In this case, it seems that vertex 1000 got lucky and got selected a lot more than its 3 other counterparts.

3.2 Scaling of vertex degree over time

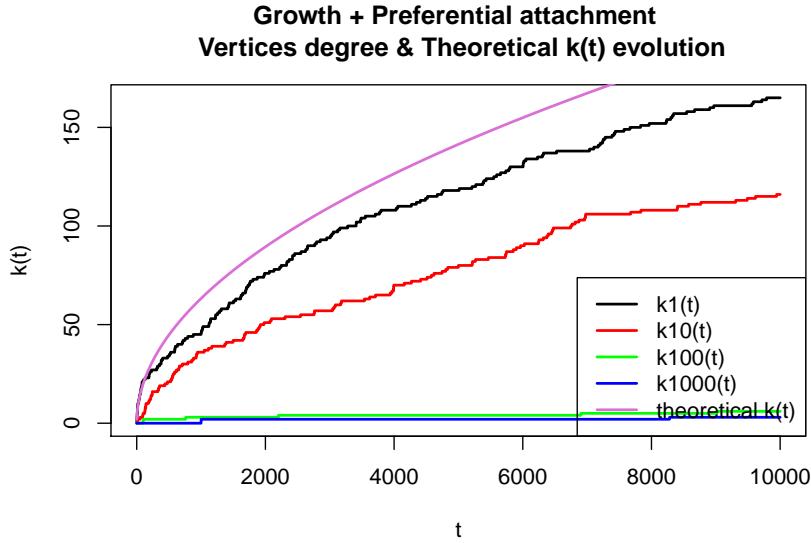
Now let's take a look on how these vertices' evolution grow comparing them with the theoretical k_i . Each simulation has its own theoretical average growth.

3.2.1 Growth and preferential attachment

On the first simulation model, the formula to follow is

$$k''_i(t) = m_0^{t_0} = m_0 \sqrt{t_0}$$

This theoretical evolution has been painted in pink in the plot below.



We can see how vertex 1 follows somewhat closely that theoretical evolution. Of course, the vertices added later in time do not follow this evolution, for they have a lower chance to get chosen on the preferential attachment.

	Model	RSS	AIC	Param1	Param2	Param3
ks.1	log+i	0	-365033.5	82.1391883	1887.4933536	-604.8393
ks.2	plaw 0.5+i	0	-387929.7	1.2637382	-7.7206726	NA
ks.3	linear regr+i	0	-585808.8	0.0003384	2.3995055	NA
ks.4	plaw 0.5+i	0	-338910.5	0.0253580	0.2812562	NA

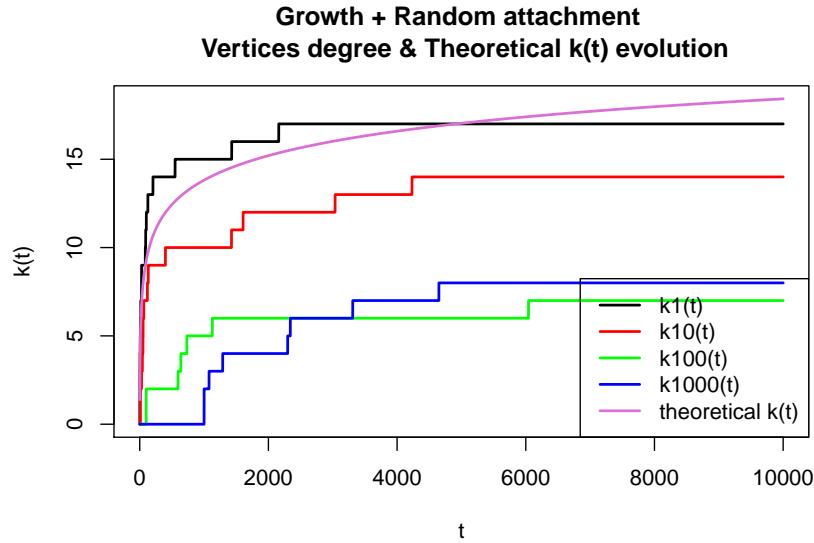
In this table.... TODO: Piero, can you comment on this?

3.2.2 Growth and random attachment

Now the scaling of vertex degree over time is as follows:

$$k''_i(t) = m_0 \log(m_0 + t - 1)$$

We can see how the first vertex follows it closely and the other 3 also tend to stagnate (later, of course, since they are added later on).



	Model	RSS	AIC	Param1	Param2	Param3
ks.1	linear regr+i	0	-405726.0	0.0002483	15.273684	NA
ks.2	plaw 0.5+i	0	-343653.3	0.0656514	8.502505	NA
ks.3	plaw 0.5+i	0	-328299.6	0.0448315	3.077355	NA
ks.4	linear regr+i	0	-569015.9	0.0007288	2.658978	NA

TODO: Piero, can you comment on this?

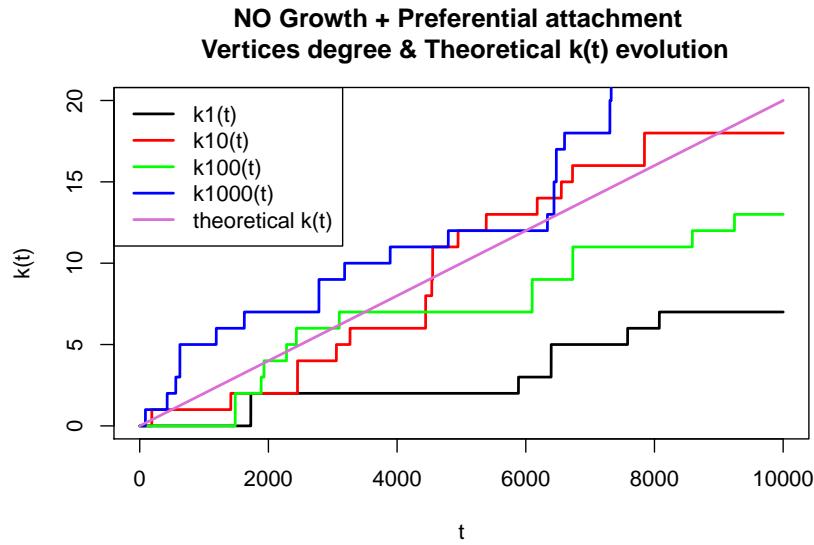
3.2.3 No Growth and preferential attachment

Now

$$k_i(t) = \frac{2m_0}{n_0}t$$

for every vertex.

We can see how, more or less, they all follow this vertex degree scaling. They vary a little because of the preferential attachment.



	Model	RSS	AIC	Param1	Param2	Param3
ks.1	plaw 0.5+i	0	-338937.8	0.0930116	-2.979471	NA
ks.2	plaw 0.5+i	0	-564432.7	0.2660471	-7.821277	NA
ks.3	plaw 0.5+i	0	-304603.6	0.1662186	-3.891142	NA
ks.4	plaw 0.5+i	0	-257043.3	0.3366766	-8.133940	NA

TODO: Piero, can you comment on this??

3.3 Degree distribution

Now let's check which final degree distribution fits better each simulation model.

To select the best distribution fitting our final degree sequence of each model, we are going to use the functions used on lab. session 2.

Afterwards, we are going to select the model that has the lowest AIC value.

3.3.1 Growth and preferential attachment

First we compute the log likelihood of each distribution fitting our degree sequence.

The table below summarizes the AIC and best parameter obtained from each distribution.

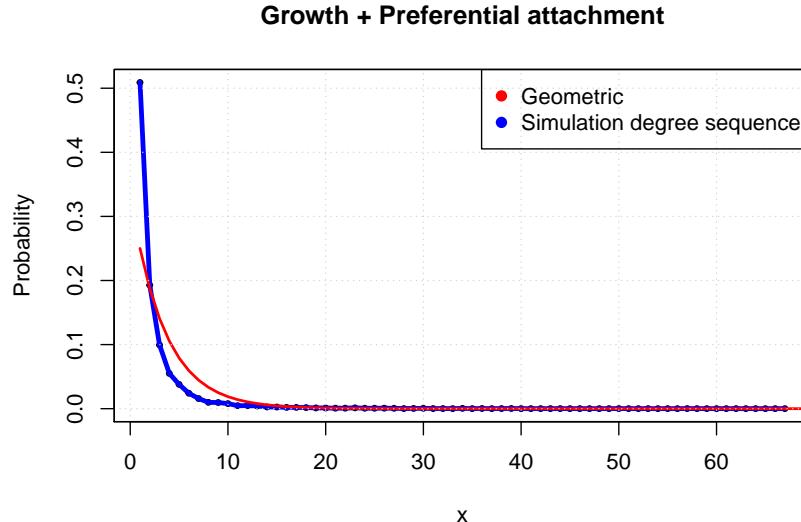
Model	X2LL	AIC	Param1	Param2
Poisson	63152.68	63154.68	3.9197327	NA
Geometric	44995.13	44997.13	0.2500562	NA
Zeta	52262.77	52264.77	1.6275755	NA
RT Zeta	77291.64	77295.64	1.9009266	243

In this case, the distribution selected that has the lowest AIC, thus the model that best fits the final degree sequence is:

```
## Best fitting for B.A. Growth + preferential: Geometric with parameter value 0.2500562
```

In the table above we can see there is a large margin between the Geometric distribution's AIC and the rest.

Let's see how this geometric distribution fits our degree distribution:



As we can see above, the selected geometric distribution fits quite well our simulation degree sequence.

3.3.2 Growth and random attachment

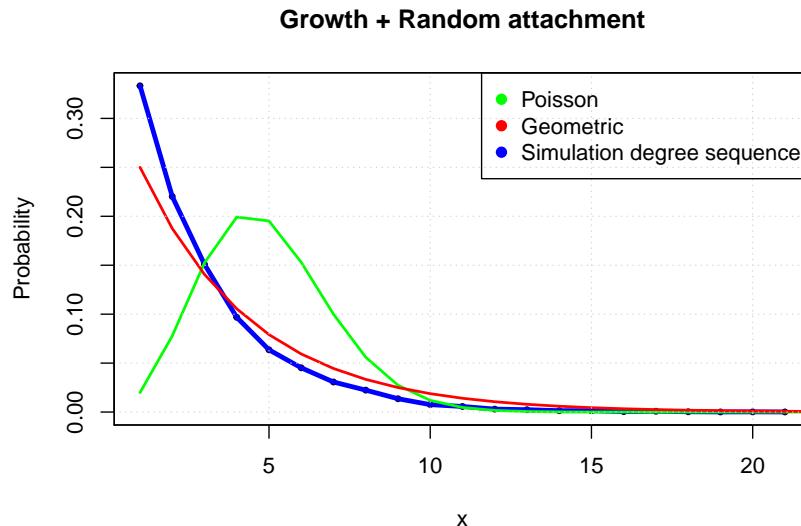
Now we are going to compute the log likelihood of each distribution fitting our degree sequence again for this new degree sequence:

Model	X2LL	AIC	Param1	Param2
Poisson	43145.36	43147.36	3.919413	NA
Geometric	44993.40	44995.40	0.250075	NA
Zeta	56335.76	56337.76	1.578978	NA
RT Zeta	80346.74	80350.74	1.712222	24

Thus, on the second model, the best fitting distribution is:

```
## Best fitting for B.A. Growth + random attachment: Poisson with parameter value 3.919413
```

We can see that the selected Poisson distribution is closely followed by the geometric function, so we are going to plot both of them to see how they fit our true degree sequence distribution:



We can see that although the selected distribution that had the lowest AIC is the Poisson distribution. However, visually speaking, we can see that maybe it is the Geometric one that better fits our data.

We have not found the reason for this miss-selection.

TODO: Piero, maybe take a look at this plot above too? :/

3.3.3 No Growth and preferential attachment

Finally, on the third model, let's see what the best distribution fitting our sequence is:

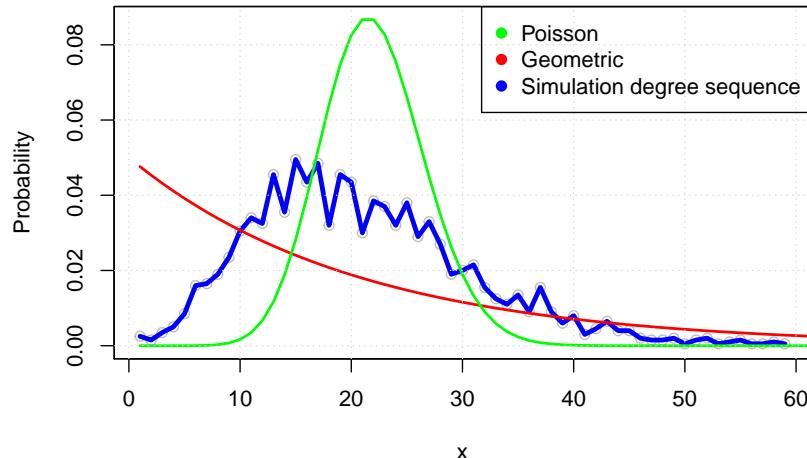
Model	X2LL	AIC	Param1	Param2
Poisson	18592.28	18594.28	21.000000	NA
Geometric	16081.30	16083.30	0.047619	NA
Zeta	20674.17	20676.17	1.289861	NA
RT Zeta	27011.82	27015.83	1.432330	76

We can see that the Geometric distribution is the one with the lowest AIC, followed somewhat closely by the geometric distribution.

```
## Best fitting for B.A. Growth + random attachment: Geometric with parameter value 0.04761905
```

Let's plot the degree sequence against the poisson distribution and geometric one so we can compare them:

NO growth + Preferential Attachment



We can see how both the Poisson and the Geometric distribution may somewhat fit our simulation degree sequence. However, the obtained λ is very large (21), which is what may affect this big peak around $x = 20$, thus making the Geometric distribution a better candidate.

4 Conclusions

To finish with this project, we are going to summarize our conclusions on both the study on vertex scaling in the three models as well as their degree distributions study.

4.1 Vertex scaling

In all the three models we have seen how they follow the theoretical vertex scaling pretty closely.

For the first two cases, since they are growing methods we can see how the evolution of the first vertex follows the theoretical one.

However on nodes that have been added afterwards (i.e. 100) this scaling does not apply on the first model, or the growing stagnates earlier on the second one because of the preferential attachment.

And finally, in the third model, since it is a no growth model, all the vertices seem to follow more or less the theoretical scaling vertex.

In conclusion, as far as the scaling vertex study concerns, the results have been satisfying and meeting our expectations.

4.2 Degree distribution

We have seen how on the first Barabasi Albert model (Growth + preferential attachment), the final degree distribution clearly follows a Geometric distribution.

However, on the other two models it is unclear to us whether the selected distributions are correct or not.

In the second model (Growth + random attachment), Poisson is selected because it has a lower AIC. Although if we take a look at the plot, visually speaking the Geometric distribution seems to fit better our degree sequence.

And in the third model (No growth + Preferential attachment), the Geometric distribution is selected although Poisson could have been selected if the lambda were to be lower. We can see how it follows a somewhat Gaussian distribution with a head and tail on probability 0, which the Geometric does not have (it only has a decreasing tail).

So, to summarize, our conclusions are that there does not exist a clear degree distribution that best fits our degree sequences on the last two models.

4.3 Future work

Some possible future work would be to see whether the best fitting distributions change when

$$n_0 \geq 3$$

and

$$m_0 \geq 2$$

Since we have not had enough time to perform these tests, we have made these two parameters static instead of dynamic. It would be interesting to see the effects on changing them, although it may probably be very time consuming.