Lab 6 - CSN: Network dynamics

Pierre-Antoine Porte
porte.pierreantoine@gmail.com

Rodrigo Arias Mallo
rodarima@gmail.com

December 16, 2017

1 Introduction

In this session, we are going to generate some data by using 3 different variations of the dynamical principles in Barabasi-Albert models (BA models in the future). Those principles are: vertex growth and preferential attachment. The different variations of models we will implements are:

- **A**: Vertex growth and preferential attachment (original).
- **B**: Vertex growth and uniform random attachment.
- **C**: Suppressed growth and preferential attachment.

Those generator models were simulated and the stored data will let us to analyze mathematical properties. In this report we will show, discuss and explain the results as well as the details of the implementation.

2 Results

For each BA model (A, B and C), two metrics are analyzed, the distribution of the degrees of the nodes, and the evolution of a node as the graph grows over time. We track the evolution of 4 different nodes selected from different points in the simulation. In the figures from 1 to **??** the best models are plotted along with the data. The notation for each model has been slightly changed, to avoid confusion. All models from session 2 where renamed with a T as prefix, and those from the session 3 a D as prefix instead. Those prefixes let us distinguish between say the model $D1$ and $T1$, as one is modeling the degree distribution, and the other the degree over time. The models for the evolution of degree over time can be shown in table 1, the models for the degree distribution are used in the minimum log likelihood form, and the only change is $\gamma = 3$ in model 2 (see table 2 of session 2). The tables 2 to 6 use AIC to measure $\Delta = AIC - AIC_{best}$ of each model.

Model	Function	Parameters
T0	$f(n) = at$	a
T1	$f(t) = (t/2)^b$	b
T2	$f(t) = at^b$	a, b
T3	$f(t) = ae^{ct}$	a, c
T4	$f(t) = a \log t$	a
T5	$f(t) = at^b e^{ct}$	a, b, c
T1+	$f(t) = (t/2)^b + d$	b, d
T2+	$f(t) = at^b + d$	a, b, d
T3+	$f(t) = ae^{ct} + d$	a, c, d
T4+	$f(t) = a \log t + d$	a, d
T5+	$f(t) = at^b e^{ct} + d$	a, b, c, d

Table 1: The list of models to fit the degree over time.

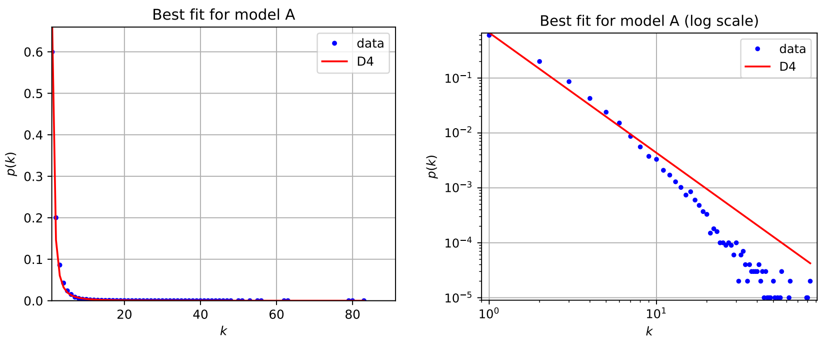

Figure 1: Distribution degree for model A

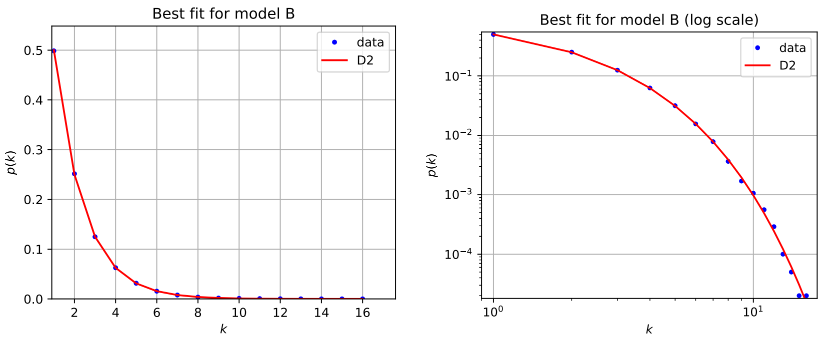

Figure 2: Distribution degree for model B

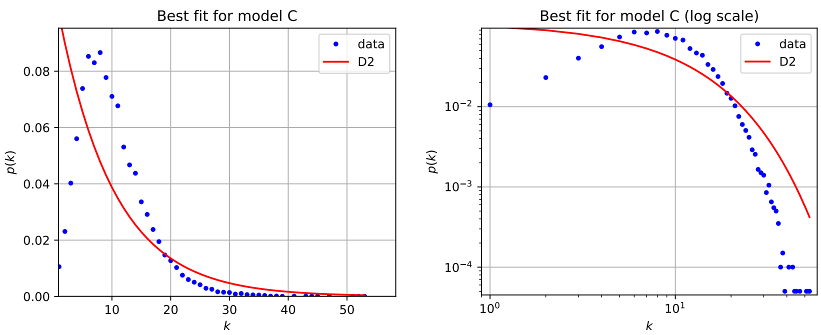

Figure 3: Distribution degree for model C

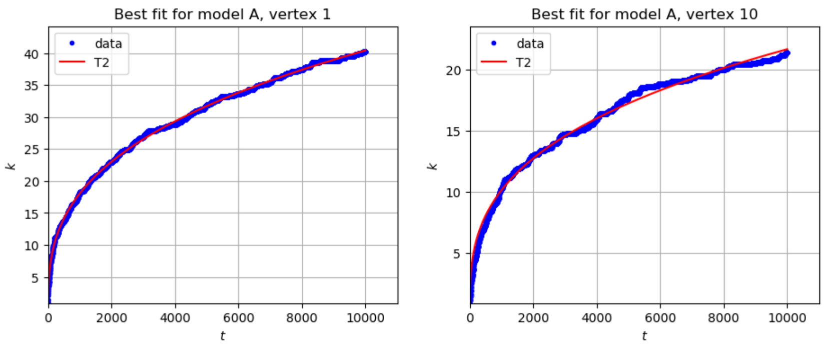

Figure 4: Degree over time for model A with vertex at $t = 1$ and $t = 10$

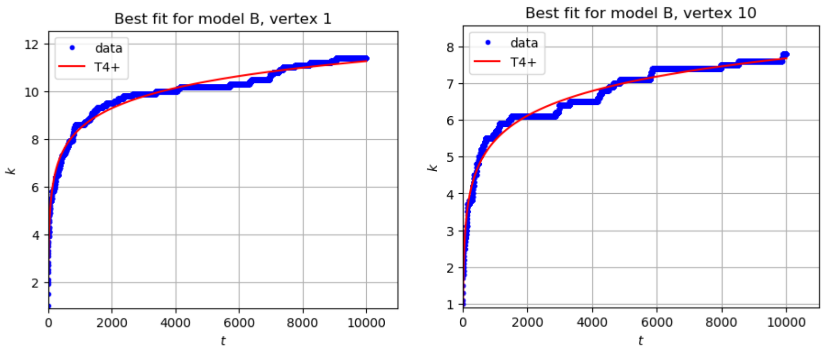

Figure 5: Degree over time for model B with vertex at $t = 1$ and $t = 10$

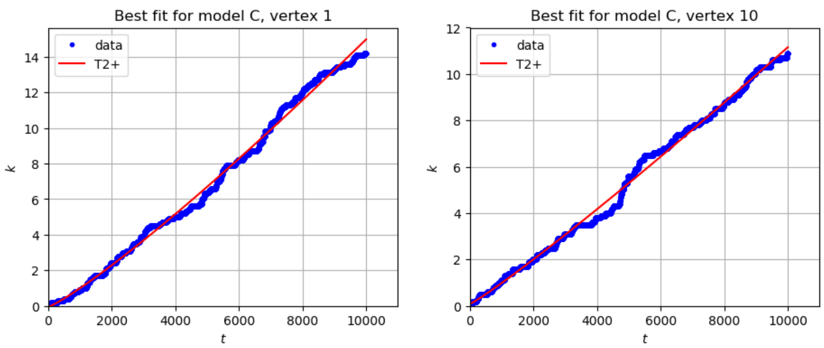

Figure 6: Degree over time for model C with vertex at $t = 1$ and $t = 10$

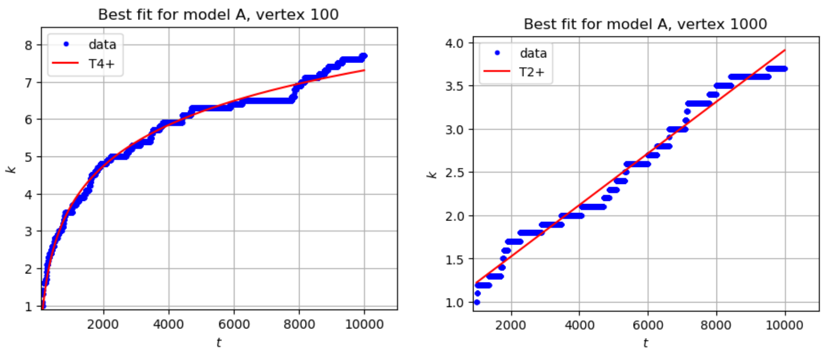

Figure 7: Degree over time for model A with vertex at $t = 100$ and $t = 1000$

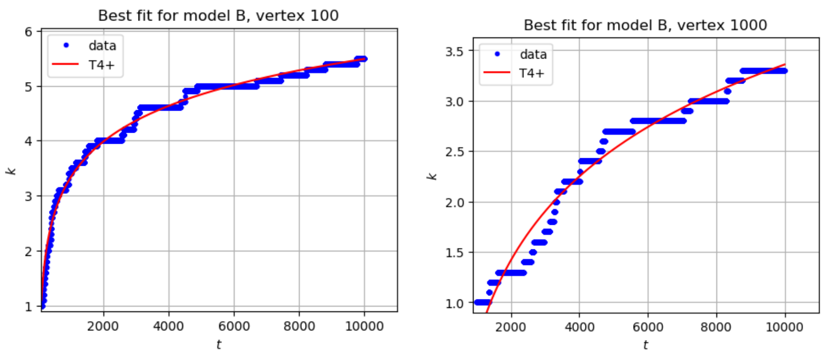

Figure 8: Degree over time for model B with vertex at $t = 100$ and $t = 1000$

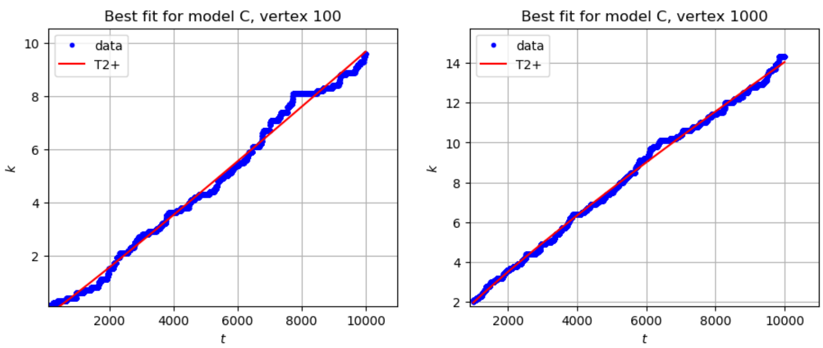

Figure 9: Degree over time for model C with vertex at $t = 100$ and $t = 1000$

3 Discussion

We used constant values for m_0 and n_0 for each model. We never compared the models with more values of m_0 or n_0 for the same model. It could have indicate us if the model behave differently given different graph in input. We could have went further and do this but we preferred to focus on the analysis of our different models with the input specified in the methods section.

3.1 Model A

Vertex degree over time We see in the plots 4 to 7 that the values for k of the selected vertex are very different, depending of the vertex chosen. If we see the value of k at $t = 4000$, for the different graphs with t_i being 1, 10, 100 and 1000, we see the values 30, 15, 6 and 2. It makes sense that latter vertices added to the graph have a slower grow in degree, as there are already some large popular vertex (with high degree), so the preferential attachment algorithm assigns very little probability to the new vertex to be selected for attachment.

We checked if the power-law dependency with 1/2 exponent gives the best fit to all the time series as required by the statement. This power-law dependency is not always the best fit, but is close. This power-law is model T1, which has an AIC higher than model T2 (see tables 3 to 6). In fact the best model seems to be model T2, defined as $f(t) = at^b$, which is the exponential growth. It makes sense because with one free parameter we can better optimize this model to fit the data, by adapting b to somewhat more close to 1/3. As we are using $m_0 = 1$, the equation 2 from session 6 is the same as model T1. If we look at table 8 to 11, we see that b is close to 1/2 only for $t_i = 1000$ (table 11).

Degree distribution The best model for degree distribution is in the plot 1. We can see that the model D4 gives a pretty good approximation of the data. We can compare the model D4 with the others looking at table 2 and the figure 10. Model D3 is the same zeta function with parameter γ fixed to 2. If we look at the table **??** we can see that $\gamma = 2.189$ for model D4. It explains why D3 and D4 are not bad methods either. The model D5 does not converge, maybe because we are using an integer $k_{\max}$ parameter as real.

3.2 Model B

Vertex degree over time We can see again that the growth of the degree depends on the time of arrival For table 3 to 6, best model is as suggested by the statement, the logarithmic model T4+. Indeed, we have 0 as Δ for the vertex degree over time for this model and for every t_i chosen to look at the vertex. We can see that the model T4+ approximate really well the data looking at figure **??**, the model T4 is also very good.

Degree distribution The best model for the degree distribution for the Barabasi Albert model without preferential attachment is the geometric one (model D2). Actually we can see in table 2.

By looking at the figure 11, we can clearly see that the geometric model (D2) is a really good approximation of our data. It even goes through a lot of points of the actual data point we have until $degree \approx 10$.

3.3 Model C

Following the session statement, the value of n_0 needs to meet $1000 \leq n_0 \leq t$, so we used $n_0 = t_{\max}/5 = 2000$.

Vertex degree over time As expected by statement, the degree over time should fit a linear scale. By computing the AIC and making the Δ (see tables 3 to 6) we have seen that the linear model was good. Also, we are really confident by saying it's linear when we look at the plot generated for model C, for every t_is. However we find that model T0, T0+, T2 and T2+ are the best. Model T2 is represented as at^b, as b is close to 1 as we can see in tables 3 to 6, which explain this good fit for model T2 and T2+.

Degree distribution As we have removed the growth of the graph, the degree distribution (see figure 12) no longer follows power-law distribution. As stated by the statement, the degree distribution for this model should be closer to a binomial distribution. Indeed it looks like it, but we found out it looked even more of a displaced Poisson with on a different scale. If we took $\lambda = 2$, then we would have a faster increasing and decreasing Poisson which is what we want. However, this Poisson has a mean of $\lambda = 2$. Therefore, it does not fit our data which has a mean of approximately 10.

So we displaced the Poisson and adjusted the scale. The lasting problem was that due do this scale and displacement, our model never produces data ≈ 0 whereas the generated model data had a lot of values ≈ 0.

We also made sure that the distribution giving the best fit in not a power-law, it's looking more of a Gaussian one. However since it's not symmetric it was hard to model the data using a normal law, for example.

In the end we chose to use a Poisson distribution as we did in session 2. It's not a really good fit, even the geometric one (D2) seems better. However, the Poisson (D1) has a δ AIC of only 726 (table 2). The plot of the geometric model can be seen in figure 12 in log-scale.

4 Methods

In order to reduce noise, all the models A, B and C were run $R_{\max} = 10$ times. The mean values were computed for the number of vertices of each degree, in the case of the degree distribution, and the mean degree for each time step. Finally this mean values were used for fitting by models.

For the model A and B the initial graph was an empty graph with only one vertex. For C we used an unconnected graph with $t_{\max}$ vertices. Because we have no vertex growth, the number of vertices is constant until the end at $t = t_{\max}$. For the three models we used $m_0 = 1$ as the initial number of edges added at each step. Changing this parameter can affect the final results, but was not tested.

We measured the growth of the vertex degree over time and the degree distribution for each model. The vertex degree was measured over the time for $t_i \in 1, 10, 100, 1000$ successively.

We used python for generating the models, to store the results and to analyze and plot the data. For each BA model with the letter M a folder `data/model`M`/` contains all the results produced by this model. Inside, for each run $0 \leq R < R_{\max}$, the degree sequence is stored in the file `dseq_r`R`.txt`, the degree distribution in `dd_r`R`.txt` and for each t in the arrival time, we produced `dt`t`_r`R`.txt` tracing the degree of the vertex arriving at time t.

4.1 Generating model C

While trying to define how to make the preferential attachment for model C we faced a problem. We were choosing from the edges to be linked to our vertex in an array of probability p (we didn't used the stubs method), with the degree of the node over the sum of the degrees of all nodes as:

```
p[i] = vertex.degree() / sum(graph.all_degrees())
```

However with $m_0 = 5$ and only 4 vertices connected, we cannot choose 5 vertices, and it would not work. To mimic the stub solution, we added one virtual degree to each unconnected node. Therefore our array of probability p was computed as:

```
if vertex.degree == 0:
    p[i] = 1 / sum(graph.all_degrees()) + sum(number_of_nodes_with_degree_0)
else:
    p[i] = vertex.degree() / (sum(graph.all_degrees()) +
        sum(number_of_nodes_with_degree_0))
```

In the end the probability vector was computed from the number of stubs.

4.2 Execution

In order to run the generator of models, the file `sim.py` populates the data from a fixed seed, so it should be reproducible. Then the two programs `fit_dd.py` and `fit_dt.py` perform the fit for the degree distribution and degree over time models, as well as the tables and figures needed in the report.

A Makefile takes care of the building process, so a simple `make` should be enough to build the data and run the models, as well as updating the report.

A Tables

	A	B	C
D1	6.865×10^3	1.242×10^3	7.259×10^2
D2	6.848×10^2	0.000	0.000
D3	2.418×10^2	2.574×10^3	6.135×10^3
D4	0.000	2.526×10^3	6.133×10^3
D5	2.819×10^{-1}	2.332×10^3	6.084×10^3

Table 2: Δ for the degree distribution.

	A	B	C
T0	6.953×10^4	5.997×10^4	1.079×10^4
T1	4.519×10^4	4.838×10^4	3.820×10^4
T2	0.000	4.089×10^3	3.387×10^1
T3	5.749×10^4	2.369×10^4	6.563×10^4
T4	5.713×10^4	1.027×10^1	5.000×10^4
T0+	4.577×10^4	2.284×10^4	4.020×10^3
T1+	2.773×10^4	2.567×10^4	2.614×10^4
T2+	9.833×10^3	1.066×10^4	0.000
T3+	∞	∞	∞
T4+	5.190×10^4	0.000	4.522×10^4

Table 3: Δ for the vertex degree over time for $t_i = 1$.

	A	B	C
T0	5.058×10^4	5.696×10^4	3.295×10^3
T1	2.832×10^4	4.428×10^4	3.568×10^4
T2	0.000	4.270×10^3	7.746×10^1
T3	7.517×10^4	7.561×10^4	6.547×10^4
T4	3.624×10^4	5.879×10^3	4.856×10^4
T0+	2.778×10^4	2.242×10^4	1.463×10^3
T1+	1.228×10^4	1.505×10^4	2.389×10^4
T2+	1.779×10^3	9.452×10^3	0.000
T3+	∞	∞	∞
T4+	2.433×10^4	0.000	4.166×10^4

Table 4: Δ for the vertex degree over time for $t_i = 10$.

	A	B	C
T0	4.537×10^4	5.667×10^4	6.776×10^3
T1	2.513×10^4	4.150×10^4	3.445×10^4
T2	2.465×10^3	1.188×10^4	1.083×10^3
T3	6.899×10^4	7.700×10^4	6.225×10^4
T4	2.847×10^4	2.539×10^4	4.640×10^4
T0+	2.039×10^4	2.846×10^4	3.431×10^2
T1+	9.261×10^3	2.057×10^4	2.008×10^4
T2+	1.800×10^3	1.474×10^4	0.000
T3+	∞	∞	∞
T4+	0.000	0.000	3.870×10^4

Table 5: Δ for the vertex degree over time for $t_i = 100$.

	A	B	C
T0	2.421×10^4	2.393×10^4	1.582×10^4
T1	8.640×10^3	3.786×10^3	3.477×10^4
T2	4.131×10^3	3.099×10^3	4.109×10^1
T3	5.804×10^4	5.415×10^4	6.955×10^4
T4	3.144×10^4	2.650×10^4	4.931×10^4
T0+	2.081×10^{-2}	9.270×10^3	2.438×10^3
T1+	5.672×10^3	2.500×10^3	1.338×10^4
T2+	0.000	2.383×10^2	0.000
T3+	∞	∞	∞
T4+	1.388×10^4	0.000	4.470×10^4

Table 6: Δ for the vertex degree over time for $t_i = 1000$.

Param	A	B	C
$D1$, λ	1.593	1.593	1.001×10^1
$D2$, q	5.000×10^{-1}	5.000×10^{-1}	9.986×10^{-2}
$D4$, γ	2.187	2.075	2.000
$D5$, γ	2.000	2.000	2.000
$D5$, $k_{\max}$	5.100×10^1	5.100×10^1	5.100×10^1

Table 7: Parameters for degree distribution models fitting.

Param	A	B	C
$T0$, a	5.060×10^{-3}	1.527×10^{-3}	1.421×10^{-3}
$T1$, a	4.338×10^{-1}	1.340×10^{-1}	1.100×10^{-1}
$T2$, a	1.616	3.348	3.260×10^{-4}
$T2$, b	3.496×10^{-1}	1.328×10^{-1}	1.166
$T3$, a	1.400×10^{1}	8.637	-1.179×10^{-5}
$T3$, c	1.251×10^{-4}	3.092×10^{-5}	-5.274×10^{-1}
$T4$, a	3.792	1.222	8.369×10^{-1}
$T4$, d_1	-8.487×10^{-1}	1.117	-8.714×10^{-1}
$T0+$, a	3.074×10^{-3}	3.231×10^{-4}	1.531×10^{-3}
$T0+$, d	1.400×10^{1}	8.496	-7.354×10^{-1}
$T1+$, a	3.704×10^{-1}	7.167×10^{-2}	1.784×10^{-1}
$T1+$, d	5.000	5.000	-5.000
$T2+$, a	8.000×10^{-1}	3.936×10^{-1}	3.613×10^{-4}
$T2+$, b	4.157×10^{-1}	3.000×10^{-1}	1.155
$T2+$, d	3.969	5.251	-5.805×10^{-2}
$T3+$, a	2.890×10^{-1}	2.890×10^{-1}	2.890×10^{-1}
$T3+$, c	8.530×10^{-1}	8.530×10^{-1}	8.530×10^{-1}
$T3+$, d	-1.000×10^{1}	-1.000×10^{1}	-1.000×10^{1}
$T4+$, a	4.955	1.230	2.020
$T4+$, d_1	-5.516×10^{-1}	1.413	7.886
$T4+$, d_2	-1.000×10^{1}	-6.005×10^{-2}	-1.000×10^{1}

Table 8: Parameters for the vertex degree over time for $t_i = 1$.

Param	A	B	C
$T0$, a	2.742×10^{-3}	1.055×10^{-3}	1.089×10^{-3}
$T1$, a	2.372×10^{-1}	9.255×10^{-2}	8.740×10^{-2}
$T2$, a	1.036	1.795	5.836×10^{-4}
$T2$, b	3.301×10^{-1}	1.592×10^{-1}	1.070
$T3$, a	-3.305×10^{-1}	-3.309×10^{-1}	-3.518×10^{-1}
$T3$, c	-1.363	-1.363	-5.655×10^{-1}
$T4$, a	2.079	8.200×10^{-1}	6.926×10^{-1}
$T4$, d_1	-9.843	-9.408	-9.875
$T0+$, a	1.237×10^{-3}	2.664×10^{-4}	1.121×10^{-3}
$T0+$, d	1.037×10^{1}	5.424	-2.203×10^{-1}
$T1+$, a	1.718×10^{-1}	3.643×10^{-2}	1.394×10^{-1}
$T1+$, d	4.885	4.303	-3.958
$T2+$, a	8.000×10^{-1}	3.138×10^{-1}	4.948×10^{-4}
$T2+$, b	3.546×10^{-1}	3.000×10^{-1}	1.088
$T2+$, d	7.982×10^{-1}	2.893	7.641×10^{-2}
$T3+$, a	2.890×10^{-1}	2.890×10^{-1}	2.890×10^{-1}
$T3+$, c	8.530×10^{-1}	8.530×10^{-1}	8.530×10^{-1}
$T3+$, d	-1.000×10^{1}	-1.000×10^{1}	-1.000×10^{1}
$T4+$, a	3.238	9.702×10^{-1}	1.874
$T4+$, d_1	-8.242	-1.528	1.197×10^{1}
$T4+$, d_2	-1.000×10^{1}	-1.258	-1.000×10^{1}

Table 9: Parameters for the vertex degree over time for $t_i = 10$.

Param	A	B	C
$T0$, a	9.486×10^{-4}	7.344×10^{-4}	9.372×10^{-4}
$T1$, a	8.295×10^{-2}	6.437×10^{-2}	7.405×10^{-2}
$T2$, a	4.223×10^{-1}	7.346×10^{-1}	3.157×10^{-4}
$T2$, b	3.132×10^{-1}	2.201×10^{-1}	1.122
$T3$, a	-5.000×10^{-1}	-5.000×10^{-1}	-5.000×10^{-1}
$T3$, c	-5.000×10^{-1}	-5.000×10^{-1}	-5.000×10^{-1}
$T4$, a	7.179×10^{-1}	5.630×10^{-1}	5.677×10^{-1}
$T4$, d_1	-1.500×10^{1}	-1.500×10^{1}	-1.500×10^{1}
$T0+$, a	4.272×10^{-4}	2.400×10^{-4}	1.007×10^{-3}
$T0+$, d	3.674	3.410	-4.628×10^{-1}
$T1+$, a	5.852×10^{-2}	3.329×10^{-2}	1.282×10^{-1}
$T1+$, d	1.864	2.366	-4.056
$T2+$, a	4.985×10^{-1}	2.886×10^{-1}	6.449×10^{-4}
$T2+$, b	3.000×10^{-1}	3.000×10^{-1}	1.047
$T2+$, d	-3.278×10^{-1}	1.054	-3.040×10^{-1}
$T3+$, a	2.890×10^{-1}	2.890×10^{-1}	2.890×10^{-1}
$T3+$, c	8.530×10^{-1}	8.530×10^{-1}	8.530×10^{-1}
$T3+$, d	-1.000×10^{1}	-1.000×10^{1}	-1.000×10^{1}
$T4+$, a	1.614	9.310×10^{-1}	1.757
$T4+$, d_1	5.000×10^{1}	-1.000×10^{1}	-1.000×10^{1}
$T4+$, d_2	-7.569	-3.089	-1.000×10^{1}

Table 10: Parameters for the vertex degree over time for $t_i = 100$.

Param	A	B	C
$T0$, a	4.351×10^{-4}	4.135×10^{-4}	1.468×10^{-3}
$T1$, a	3.608×10^{-2}	3.447×10^{-2}	1.201×10^{-1}
$T2$, a	1.554×10^{-2}	2.501×10^{-2}	4.896×10^{-3}
$T2$, b	5.963×10^{-1}	5.367×10^{-1}	8.642×10^{-1}
$T3$, a	-5.000×10^{-1}	-5.000×10^{-1}	-5.000×10^{-1}
$T3$, c	-5.000×10^{-1}	-5.000×10^{-1}	-5.000×10^{-1}
$T4$, a	3.065×10^{-1}	2.998×10^{-1}	1.013
$T4$, d_1	-1.500×10^{1}	-1.500×10^{1}	-1.500×10^{1}
$T0+$, a	2.980×10^{-4}	2.644×10^{-4}	1.329×10^{-3}
$T0+$, d	9.254×10^{-1}	1.003	9.276×10^{-1}
$T1+$, a	4.077×10^{-2}	3.738×10^{-2}	1.833×10^{-1}
$T1+$, d	-3.598×10^{-1}	-2.239×10^{-1}	-4.921
$T2+$, a	2.958×10^{-4}	3.336×10^{-1}	4.192×10^{-3}
$T2+$, b	1.001	3.000×10^{-1}	8.802×10^{-1}
$T2+$, d	9.264×10^{-1}	-1.835	1.325×10^{-1}
$T3+$, a	2.890×10^{-1}	2.890×10^{-1}	2.890×10^{-1}
$T3+$, c	8.530×10^{-1}	8.530×10^{-1}	8.530×10^{-1}
$T3+$, d	-1.000×10^{1}	-1.000×10^{1}	-1.000×10^{1}
$T4+$, a	1.293	1.222	2.171
$T4+$, d_1	5.000×10^{1}	5.000×10^{1}	-1.000×10^{1}
$T4+$, d_2	-8.397	-7.904	-1.000×10^{1}

Table 11: Parameters for the vertex degree over time for $t_i = 1000$.

B Figures with all models

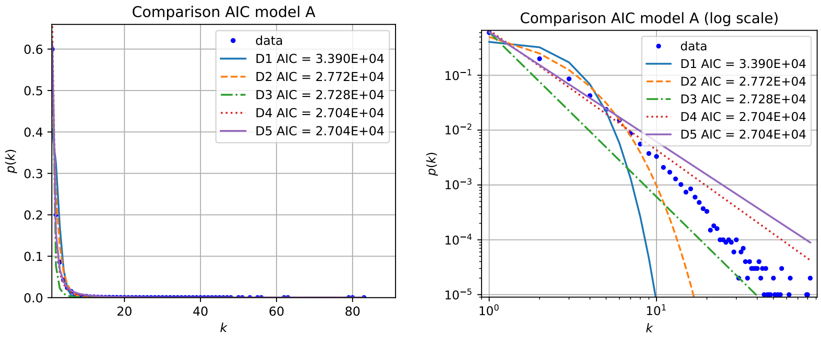

Figure 10: Distribution degree for model A

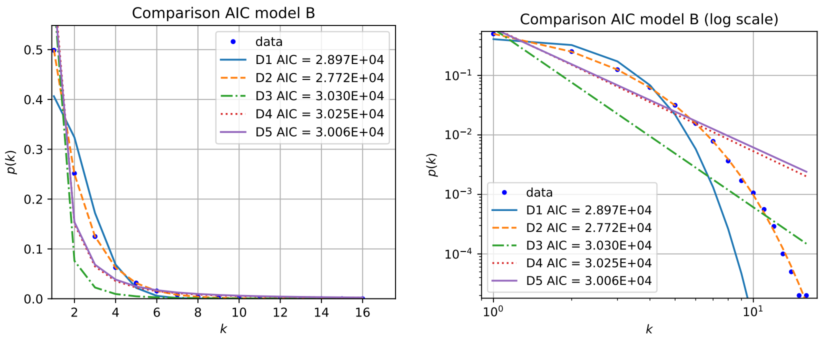

Figure 11: Distribution degree for model B

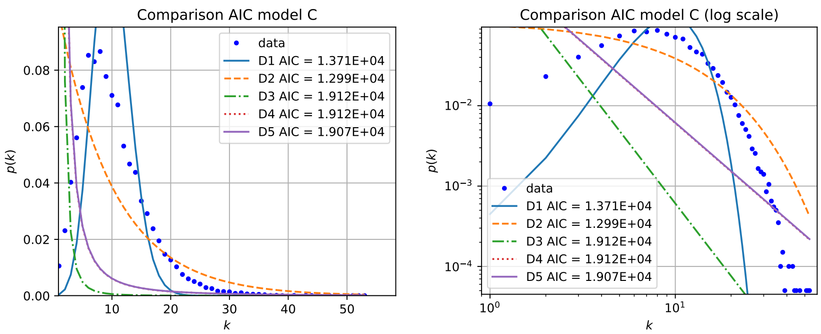

Figure 12: Distribution degree for model C

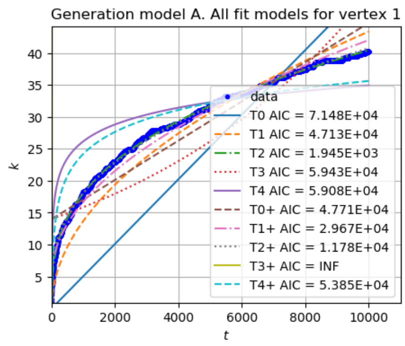

Figure 13: Degree over time for model A with vertex at $t = 1$

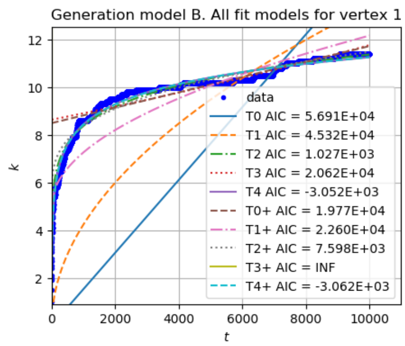

Figure 14: Degree over time for model B with vertex at $t = 1$

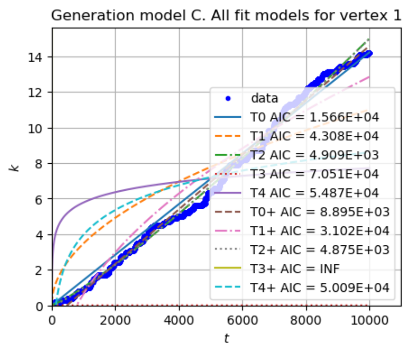

Figure 15: Degree over time for model C with vertex at $t = 1$

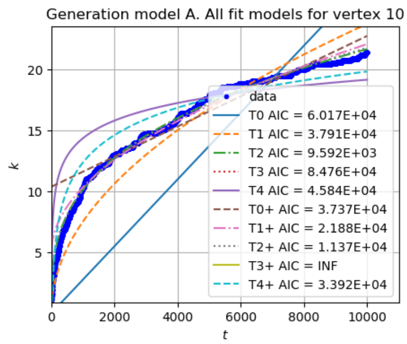

Figure 16: Degree over time for model A with vertex at $t = 10$

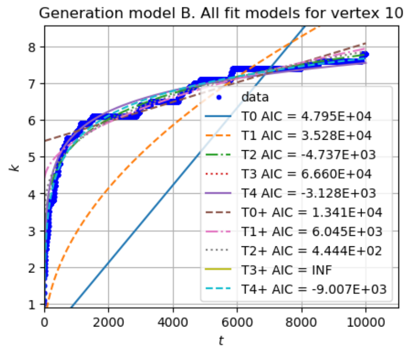

Figure 17: Degree over time for model B with vertex at $t = 10$

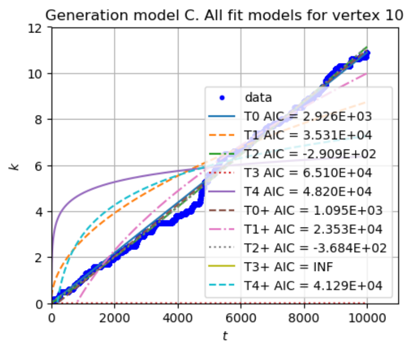

Figure 18: Degree over time for model C with vertex at $t = 10$

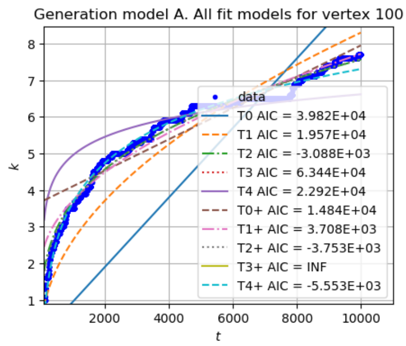

Figure 19: Degree over time for model A with vertex at $t = 100$

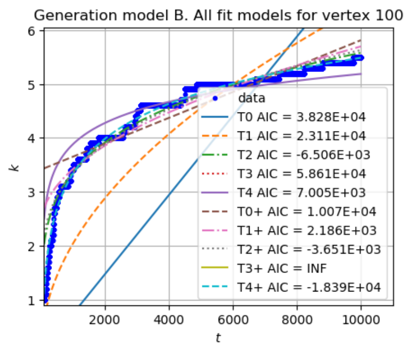

Figure 20: Degree over time for model B with vertex at $t = 100$

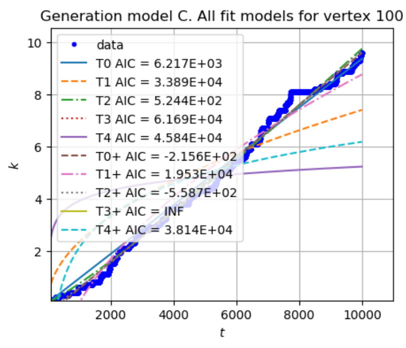

Figure 21: Degree over time for model C with vertex at $t = 100$

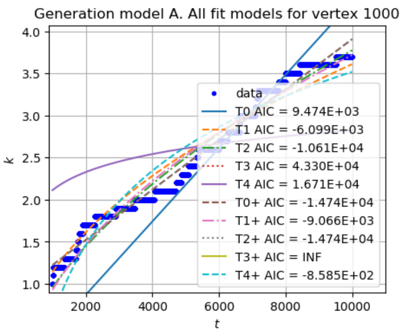

Figure 22: Degree over time for model A with vertex at $t = 10000$

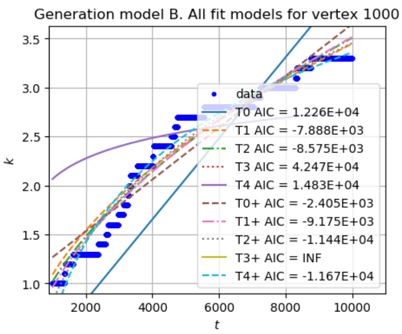

Figure 23: Degree over time for model B with vertex at $t = 1000$

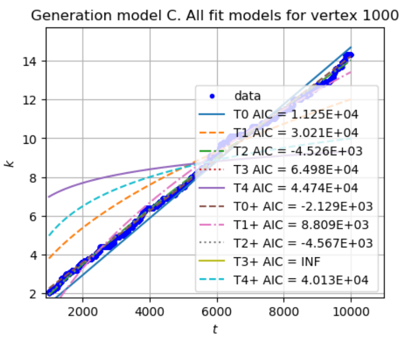

Figure 24: Degree over time for model C with vertex at $t = 1000$