A Simulation Study of a Preferential Attachment with Delay

Charles Cooper, Dr. Partha Dey
University of Illinois at Urbana–Champaign

Abstract

Preferential attachment models explain many real-world networks, such as the follower network of social media platforms and the citation network of academic papers. Often in real life, later additions to the network will not be considered by new additions to the network. In this paper, we will account for this phenomenon by introducing a random delay variable to a preferential attachment model, i.e., any new connection considers nodes introduced before the delay. Through simulation, we will compare our model's degree distribution and maximum delay to a model without delay. Then, we will make inferences regarding how induced delay affects the limiting degree distribution of our model. Finally, we will discuss future research which can be conducted relating to dynamic graphs with delay.

Definitions

- **Preferential Attachment** is a growth model where new connections are formed between two nodes proportional to the degree of the pre-existing node.
- A **Supernode** is a node in a graph with a degree far greater than the vast majority of other nodes.
- **Error** as a term will mean the difference between an observed value and the expected value.
- The **tail exponent** of a graph is the exponent for the power-law of the asymptotic degree sequence, assuming it follows a power-law.

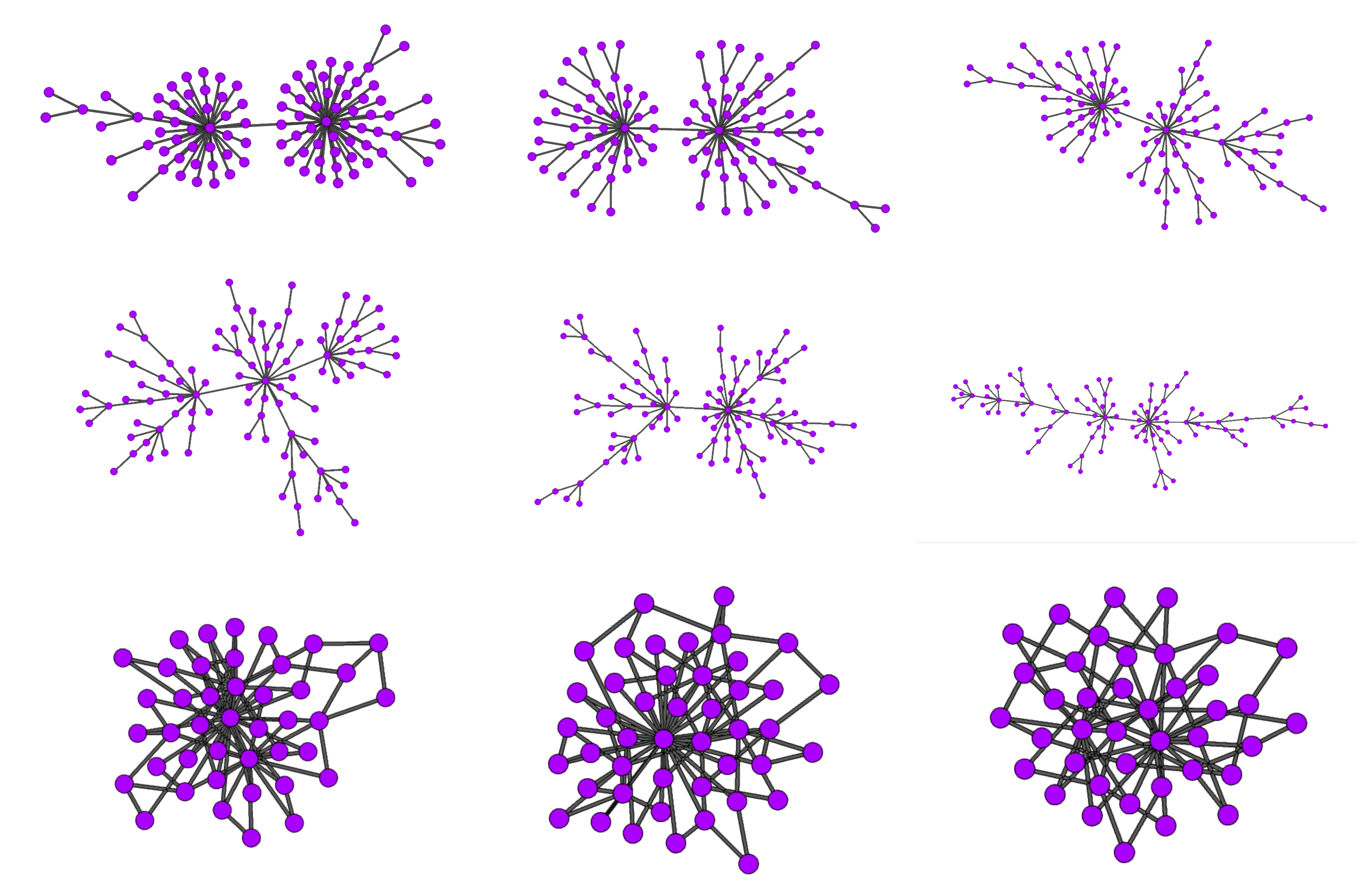
Our Model

- The **original model** we are building from represents a graph sequence dependent on t time where a new node v_t connects to a node, $v_{0 \geq i < t}$ in the existing graph G_{t-1} with probability proportional to the $deg(v_{0 \leq i < t})$. Each node makes m connections. More precisely, v_t chooses a subset of size m from all the nodes in G_{t-1} to connect to.
- **Our model** expands on the above model by adding a delay factor. In our case, a new node v_t will connect to nodes from $G_{t'}$ where $t' = t - \xi_t$ where ξ_t is some discrete random variable taking values from $1, 2, \ldots, t$.
- Our specific **delay variable**, ξ_t, we define as $xi_t = min(t - |G_0| + 1, \lceil u^{\frac{-1}{a}} \rceil)$, where u is a uniform random variable on $(0, 1)$.
- We expect the degree distribution of our model to match that of a no delay preferential model if $\mathbb{E}(d_1) < \infty$. For us, this happens when $a \geq 1$.

Guiding Questions

- **Tail Behavior:** What effect will increased delay have on the tail of the degree distribution? How does this affect scale with delay? Does our model align with the proven results for no-delay models? At what point is the effect of delay noticeable?
- **Maximum Degree:** How does the graph's maximum degree grow with the graph's size? How does delay affect the maximum degree of our graphs? Will the supernodes eventually take up a constant proportion of the total connections?
- **Graph Shape:** When can we see distinct supernodes in our visualizations? Which nodes become supernodes? How many supernodes do we expect there to be given an initial graph?

Graphs



Graphs of size 100 visualized using Graphia. The top six use $m = 1$ with a values 0.1, 0.3, 0.5, 0.7, 1, and 1.2 in that order from top left to bottom right. The bottom three use $m = 3$ and have a values 0.3, 0.6, 0.9 from left to right. The bottom three also have nodes with degree 2 removed for clarity.

Simulation Data and Algorithm

Table 1. Tail Exponents

(a, δ)	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.1
0	4.165	3.555	3.245	3.106	3.067	2.995	2.982	2.898	2.967	2.985	2.955
0.5	4.685	3.969	3.647	3.533	3.459	3.410	3.387	3.311	3.381	3.419	3.409
1	4.916	4.914	3.994	3.767	3.821	3.751	3.750	3.701	3.663	3.687	3.651
-0.5	3.671	3.056	2.771	2.664	2.586	2.554	2.545	2.483	2.518	2.524	2.485

Table 2. Average Max Degrees

(a, δ)	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.1
0	27003	13305	6441	3468	1970	1290	921	721	851	756	685
0.5	24562	11118	4784	2401	1190	746	454	362	275	335	302
1	23131	9884	4230	1761	863	521	364	251	214	220	173
-0.5	32671	19144	11281	7426	5847	4650	3913	3527	3310	4353	3191

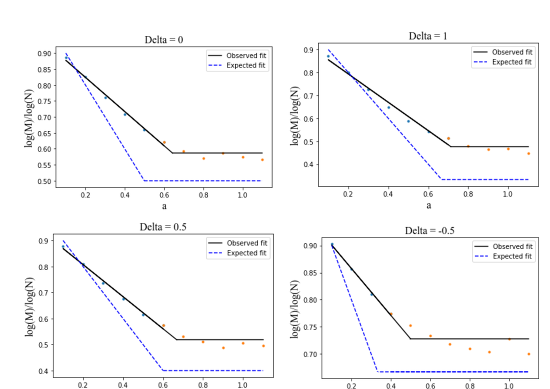
Algorithm 1 Graph Creation

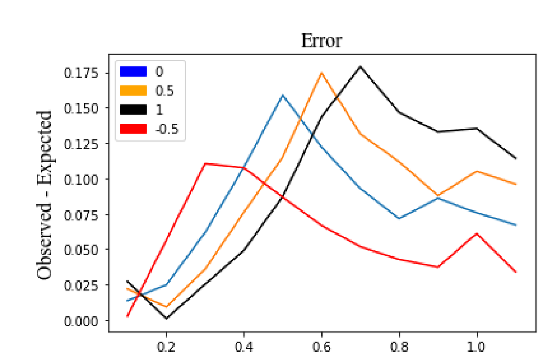
let $repeatedNodes - RN \leftarrow \forall v \in v_0$, add v, $in\deg(v)$ times
let $source = size(v_0)$
let $G = G_0$
let $position = size(RN)$
while $source \leqslant N$ **do**
  let $targets \leftarrow$ random subset of m nodes from RN
  Add a node to G with value $source$ and create edges from it to the nodes in targets
  Add $targets$ to RN.
  Add node v_{source} $m + \delta$ times to repeated nodes
  let p_l be the last element in $position$
  Add $p + 2 * m * \delta$ to $position$
  let d be defined as in **1.4**
  let $d' \leftarrow$ the element at position $size(RN) - d$ in $position$
  $targets \leftarrow$ random subset of nodes from the first d' nodes of RN
  $source += 1$
return G

Five samples of graphs were generated with $m = 1$ and size 100,000. The average respective values were then taken and put into the tables above. Jupyter Notebook with NetworkX package was used to generate the data with the algorithm described on the right.

Plots

Data visualized using the Matplotlib Python package





The plot on the left plots $\frac{\log(M_N)}{\log(N)}$ where M_N is the average max degree from our table above and a is our delay parameter defined in the section **Our Model**. We use the ratio of logarithms instead of the raw M_N because we expect to see a nice linear and then constant relationship with this plot. The plot on the right shows the error in our data versus a.

Transition Point

Examining our graphs, optically there is some transition point where the graphs change shape. We conjecture that each plot can be approximated by a piece-wise function,

$$\frac{\log M_N}{\log N} \approx \begin{cases} 1 - a & \text{if } a \leq 1 - \frac{1}{\theta} \\ \frac{1}{\theta} & \text{if } a \geq 1 - \frac{1}{\theta} \end{cases}$$

We conjecture that as $N \to \infty$, a plot of $\frac{\log M_N}{\log N}$ versus a would fit this function exactly.

Other Conjectures

- We expect given a initial graph G_0, there is some number k dependent on G_0 such that all nodes $v_{i<k}$ in the limiting graph generated from G_0 will always become supernodes. Moreover, no other nodes will. For instance, given a star graph with one starting node, v_1, nodes v_1 and v_2 will become supernodes.
- Past the transition point, we expect our error to decrease linearly. Before that, we did not have enough information to precisely predict what shape the error plot would have, but we can predict that it increases up until the transition point.

Discussion

- **Changing the model:** Regarding preferential attachment models, four elements can be changed in future studies. One may try a broader range of δ, different distributions with different moments for the delay variable, a larger m, or other initial graphs. The interesting factor to change is the delay variable, as we have only considered whether the delay variable has a finite first moment in our study.
- **Future Studies:** As preferential attachment models can be used to model systems where inequality arises, citation networks, social-media following, wealth, etc., one may be interested in reducing the effect of delay in the system to reduce inequality. Moreover, the average entity within the system knows only a tiny proportion of the system, so it is beneficial to consider that a specific entity can influence the entire system with limited knowledge. This can be represented by introducing different mechanisms to our preferential attachment model, such as the ability for nodes to disconnect from each other and reconnect to other nodes, which were added later. Ultimately, there are many open questions regarding dynamic models and induced delay, and we hope this paper assists with future research in this area.

References

References

[1] Barabási Albert-László. *Network science*. Cambridge University Press, 2017.

[2] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[3] Nicolas Curien, Thomas Duquesne, Igor Kortchemski, and Ioan Manolescu. Scaling limits and influence of the seed graph in preferential attachment trees. *J. Éc. polytech. Math.*, 2:1–34, 2015.

[4] Remco Van Der Hofstad. *Random graphs and complex networks*, volume 43. Cambridge university press, 2016.