Putting Preferential Attachment to the Test:

A search for alternative attachment policies for scale-free network formation

Thomas Young & Christoph Uhl
2019
https://github.com/thyo9470/Preferential-Attachment

Introduction

Null models, such as the Erdős–Rényi and Watts-Strogatz, have been important in helping determine specific characteristics in observed networks. For years, these models were used in misleading ways when compared to scale-free networks. Scale-free graphs are defined by their power-law degree distribution, which causes most of the edges to be connected to a small set of nodes. Examples include the World Wide Web (WWW), protein-protein interactions, citation networks, and wealth distribution. The problem with using Erdős–Rényi and Watts-Strogatz graphs as null models is that they do not create graphs with a power-law degree distribution, thus inaccurately representing the real-world network being studied.

This problem was addressed in 1999 in the paper “Emergence of Scaling in Random Networks” by Albert-László Barabási and Réka Albert. The solution was to define a new null model based on two principles. First, that “networks expand continuously by the addition of new vertices” and second, “new vertices attach preferentially to sites that are already well connected” (Barabási & Albert, 1999). The solution proposed was well received and adopted by the network science community. This paper will focus on the second statement made; saying nodes are more likely to attach to high degree nodes, also known as preferential attachment.

Although preferential attachment has been well received, the paper seems to skim over parts, leading to the accuracy of the attachment policy being overestimated. It talks about why the Erdős–Rényi and Watts-Strogatz models are not sufficient for creating scale-free models, but it falls short when it comes to exploring alternative attachment policies that could create scale-free networks. The paper also states that the “development of the power-law scaling in the model indicates that growth and preferential attachment play an important role in network development.” Since they are able to create scale-free networks with preferential attachment, It is assumed this is sufficient to prove the policy. The following question was created to address these holes in the paper: are there other attachment policies that create networks similar to preferential attachment, thus demonstrating other possible ways for scale-free networks to emerge? If a new attachment policy is able to imitate preferential attachment and generate a scale-free network, this means that there could be ambiguity in the main graph attribute that leads to the development of a scale-free network.

There have already been several papers that try and cover this topic, however, they tend to focus on a specific attachment policy or method that might substitute or supplement preferential attachment. In the paper “Growing network with local rules: preferential attachment, clustering hierarchy, and degree correlations” by Alexei Vázquez, various local attachment policies are explored to explain in more detail the source of structure (Clustering hierarchy and degree correlation) in various networks. The policies explored include random walks to describe the WWW and citations networks, connecting nearest neighbors for social networks, and duplication divergence in gene networks (2003). The Vázquez study begins to shed light on alternative attachment policies, however, more focus is put on specific instances of scale-free networks instead of more general policies, which our paper aims to address. There are other examples of exploring alternatives to preferential attachment, though not very many (Berger, Borgs, Chayes, Dsouza, & Kleinberg, 2005; Barrat, Barthélemy, & Vespignani, 2004).

Method

To explore various attachment policies that might mimic preferential attachment, the following was done. First, a set of attachment policies that could potentially lead to the formation of a scale-free network were selected. For each policy, a seed graph (most of the time being two connected nodes) was created. The graphs were grown by adding a single node at a time. Each node was then connected to one or more nodes based on the attachment policy being tested. Each graph was grown to a size of 1,000 nodes, and 10,000 trials were run. The data collected was used to satisfy two phenomena. The first is that the data imitates preferential attachment's method of creating graphs where attachment is proportional to degree, indicating other attachment policies can disguise themselves as preferential attachment. This was measured by comparing node degree against the probability of a node with degree k having a new node attached to it. Preferential attachment was indicated by a positive linear correlation. The second phenomenon being the graph that is constructed forms a scale-free network. This was measured through the degree distribution and degree complementary cumulative distribution (CCDF). The degree distribution followed a power-law trend where $Pr(k) \sim k \sim\alpha$, where α was most often between 2 and 3 with some variation (Newman, 2018). To get the value of alpha for each graph, we used the curve_fit function from the scipy python package (Pauli, 2019). The rest of this section will outline the attachment policies we used and their significance.

Preferential Attachment

For preferential attachment, we compared our policies to two different models, Price Model (Newman, 2018) and Barabási–Albert model (Barabási & Albert, 1999).

Uniform Attachment

This acted similarly to Erdős–Rényiand Watts-Strogatz models and was used as another point of reference for the attachment policies.

Temporal Attachment

New nodes attach to sites proportional to how long they have been apart of the network. It was predicted that new nodes would attach to older vertices for a few reasons, the first being the longer a node is around, the more chances it will have to be attached to. This will lead it to have a higher degree than others, imitating preferential attachment. Additionally, there are specific examples within real networks that would justify temporal attachment. Examples include citation networks where older papers are more likely to be cited because they are oftentimes the papers that lay the groundwork for a field and thus will need to be cited often. The same idea can be applied to web pages. Older web pages have both time to improve, making them more preferable to visit. Further, since they have been around for longer, they have had more chances for people to stumble upon them and reference them later on their own websites.

Mean Geodesic Distance

New nodes attach to cites proportional to their mean geodesic distance. This is the same as saying vertices with a large mean geodesic distance are attached to more often. Since the mean geodesic distance can be directly influenced by node degree, it could be possible to imitate preferential attachment. Alongside this, the inverse mean geodesic distance as explored.

Inverse mean geodesic distance policy says that nodes with a low mean geodesic distance are more likely to have a new node attach to it. The significance of this is that many nodes that have a high degree oftentimes have low mean geodesic distances, possibly indicating that new nodes like attaching to well-connected vertices. Reasons this would show up in a real-world network can be exemplified in a citation network. A small mean geodesic distance would be beneficial within a citation network since it would reduce the degrees of separation between you and possible other researchers you might want to collaborate with in the future.

Triadic Closure

New nodes attach to two nodes when entering a graph. The first node is selected uniformly at random from all nodes. With a given probability, the second node is either selected from the neighbors of the first, or from all nodes again. The network formed from this attachment policy values triangle motifs. This type of formation would be well represented in a social network where it is common for connections to be made between a person and their friend's friend.

Local Triangle Count

New nodes attach to two cites where the probability to attach to a cite is proportional to the number of triangle motifs they are part of. Similar to the mean geodesic distance, the local triangle count can be impacted by the node degree. This is because nodes with many edges have more of a chance to connect to nodes to form a triangle motif. Networks such as protein-protein interactions can both show reasons as to why this would be beneficial. In a protein-protein network, the nodes that participate in many triads can possibly indicate that it is useful in many different reaction chains, showing versatility.

Results

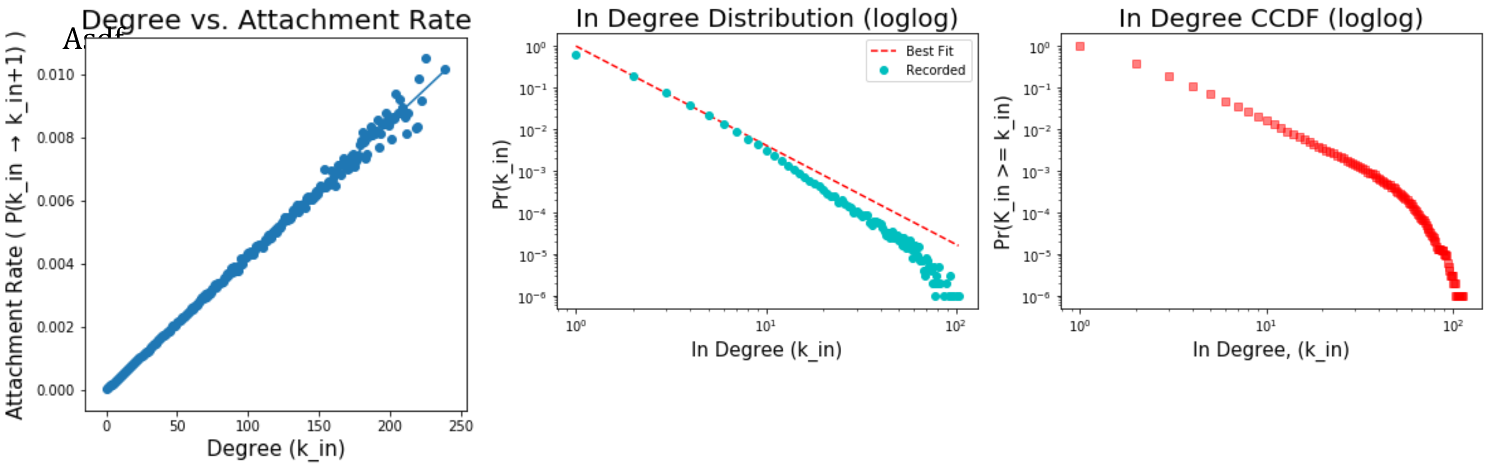
The data collected includes node degree vs. attachment rate (k vs. Pr(k →k+1)), degree distribution, and degree CCDF. and citation networks. Overall most of the degree vs. attachment rate graphs showed some kind of trend indicating some correlation between the two parameters. Also, the degree distribution and degree CCDF graphs were either linear or decayed exponentially when viewed on a loglog scale.

Preferential Attachment

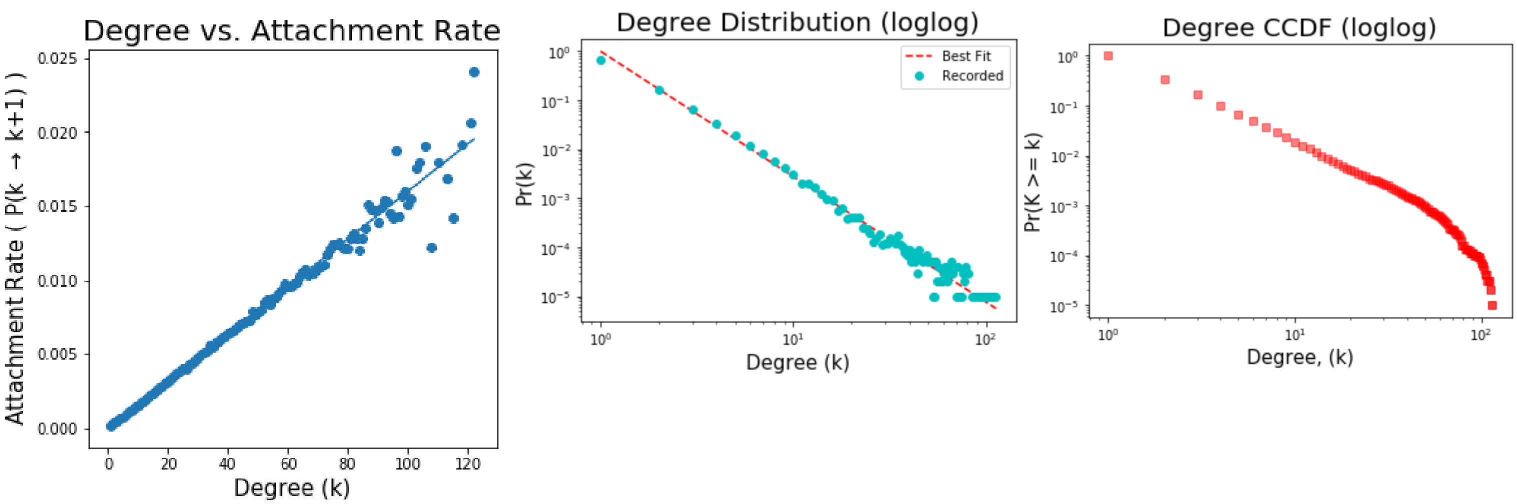
Both of the preferential attachment models exhibit a positive correlation between degree and attachment rate, as expected. This is due to the inherent property of preferential attachment. One characteristic the graph has is that the end begins to decay in accuracy since there are fewer data points to continue the trend. The degree distributions of both models show a power-law with α between 2 and 3 and a sharp dip in the tail of the CCDF.

Price Model

The price model is a directed graph, which means it uses a node's in-degree in for the preferential attachment.

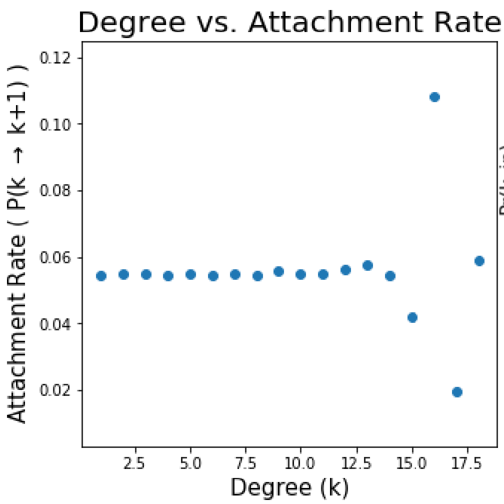


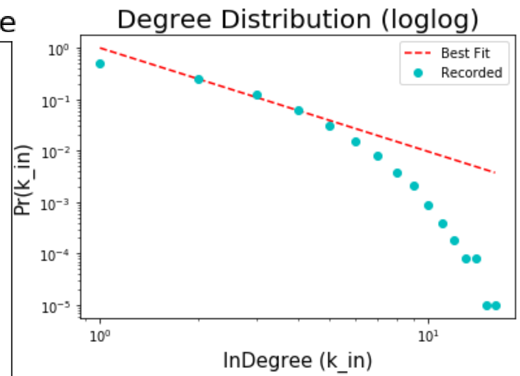
Barabási-Albert

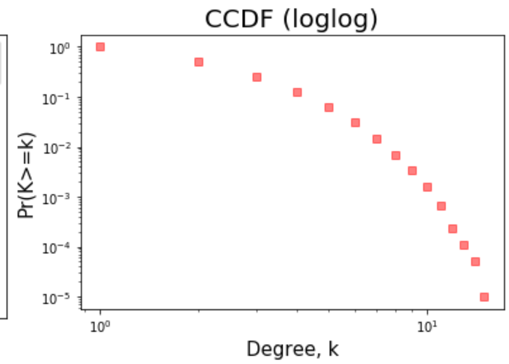


Uniform Attachment

The degree vs. attachment rate of the uniform attachment graph shows that the probability of a node attaching to others does not vary with the degree. Again, the tail is less accurate because it uses a smaller sample size for the larger degrees. Both the degree distribution and CCDF have a non-linear trend on the loglog scale. For this reason, the alpha value is irrelevant.

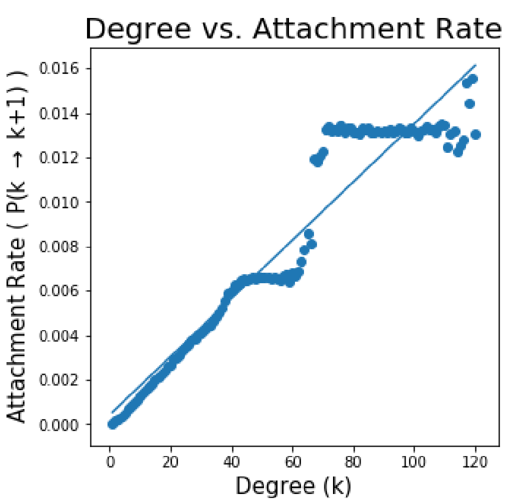


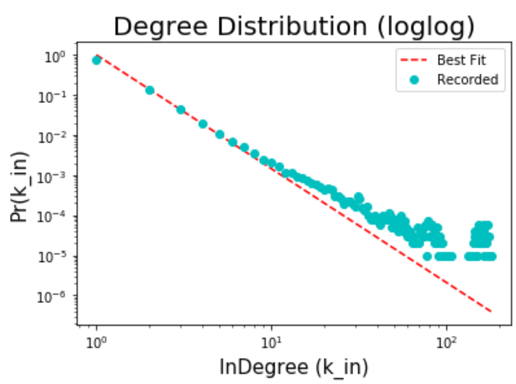


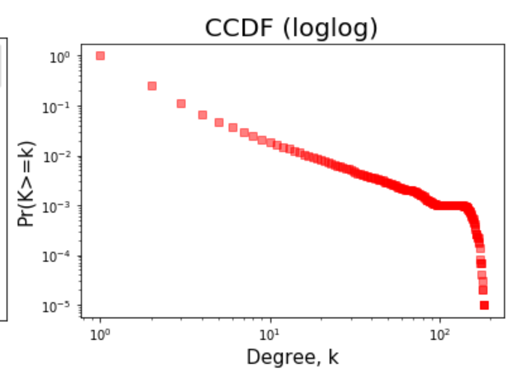


Temporal Attachment

Similar to the preferential attachment models, temporal attachments display a positive linear correlation between degree and attachment rate. The degree distribution is relatively linear with an α value of 2.83. Finally, the loglog CCDF is linear until the end where it drops off around k=150.

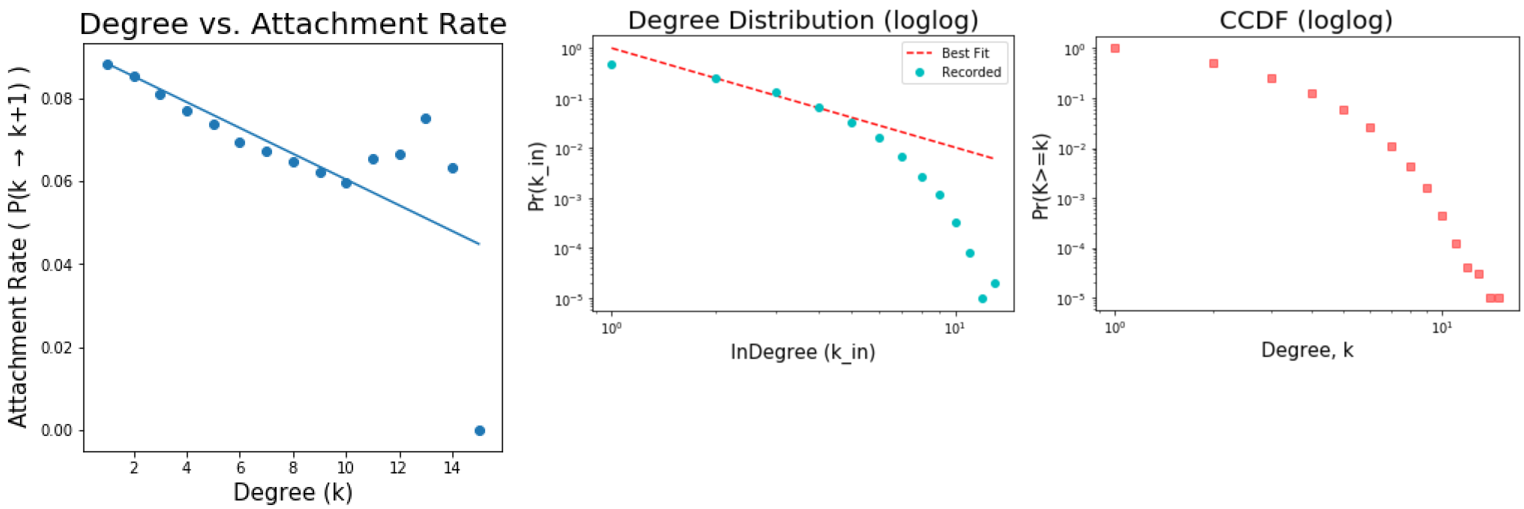






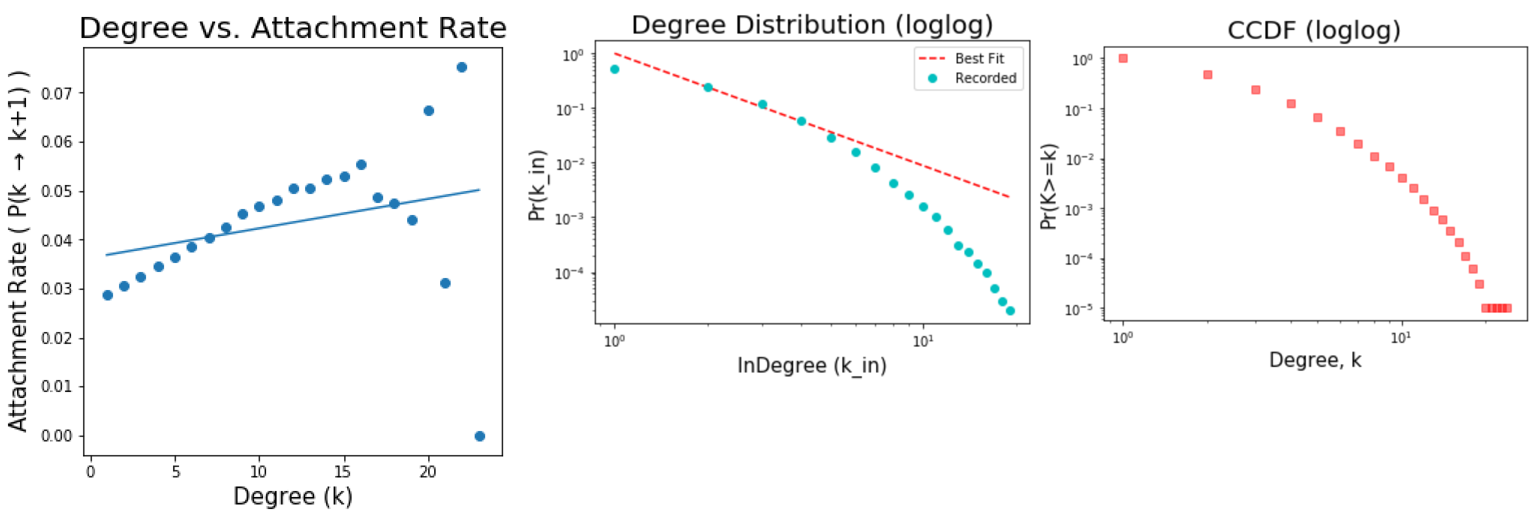
Mean Geodesic Distance

Both of the variations of mean geodesic distance attachment used show an exponential decrease on a loglog scale for degree distribution and degree CCDF. Again, for this reason, the α value is ignored. There is a negative linear correlation between degree and attachment rate on the mean geodesic distance graphs.



Inverse Mean Geodesic Distance

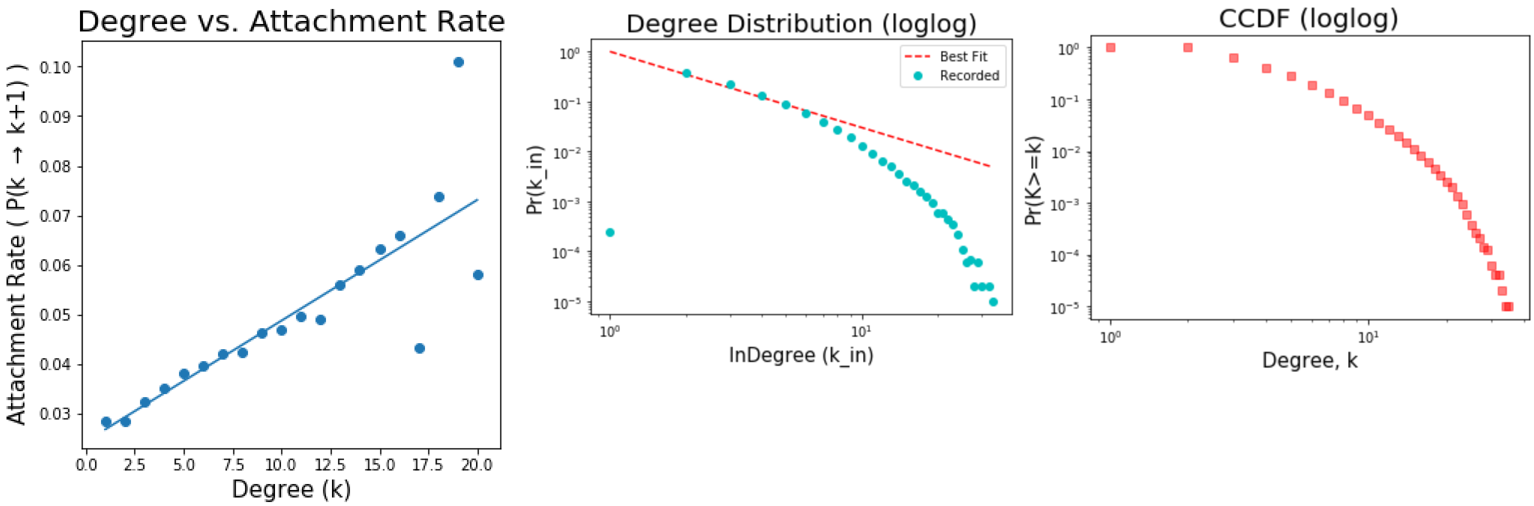
The inverse mean geodesic distance attachment policy produced a positive linear correlation between degree and attachment rate, however, the values are much more spread out and smaller compared to the preferential attachment graph.



Triadic Closure

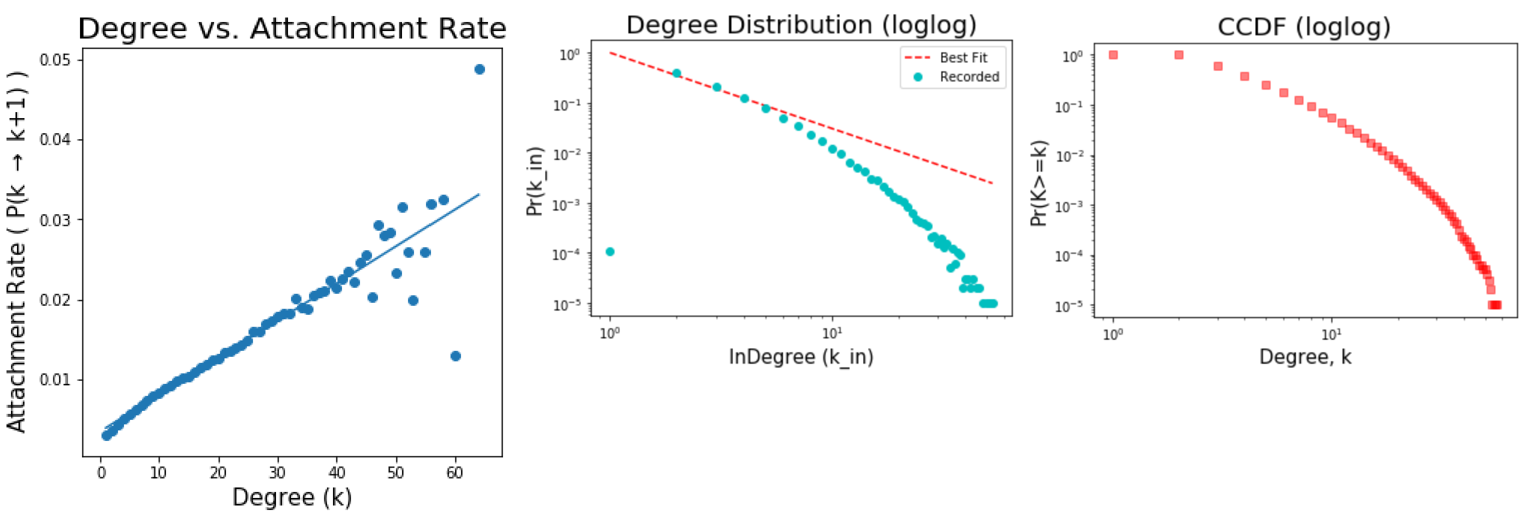
Two variations of triadic closure were tested, one where the second node has a 50% chance to select a neighboring node of the first node picked, and the other with a 90% chance. Both showed a positive linear correlation between degree and attachment rate, however, the degree distribution and CCDF are non-linear and non-representative of scale-free networks. Since it does show the loglog linear trend needed, α is ignored.

50% Probability of Selecting a Neighbor Node.



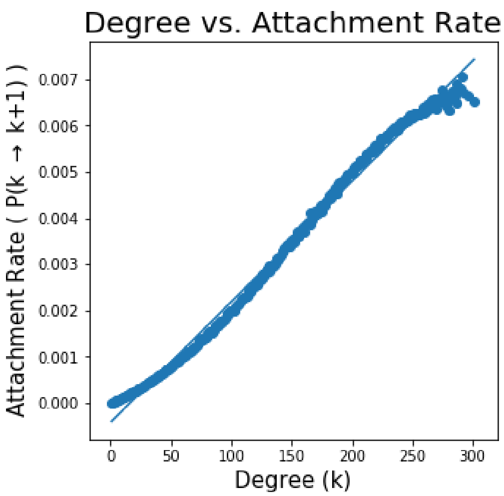
90% Probability of Selecting a Neighbor Node

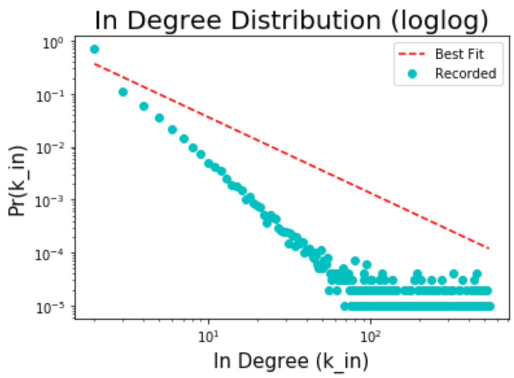
The degree vs. attachment rate graph for 90% was much more dense, which is more similar to preferential attachment than 50%.

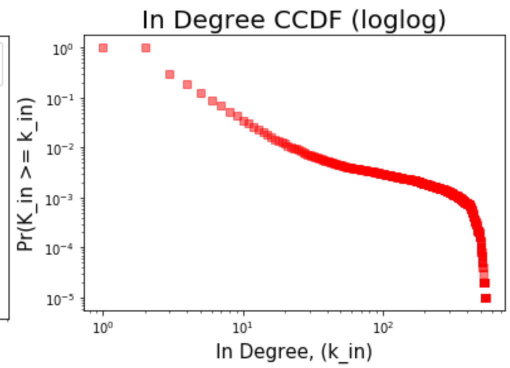


Local Triangle Count

The local triangle count created a similar graph to preferential attachment where there was a positive linear correlation between degree and attachment rate. The degree distribution shows a negative linear trend on the loglog scale, however, the alpha value generated by Scipy's curve_fit was 1.44, which is outside of the expected limits. The CCDF curve is also relatively linear until it drops off at the tail. The major characteristic is that the drop in the degree CCDF is much sharper than the other graphs that imitate preferential attachment.







Discussion

The data collected has shown that there are false positives that imitate the network generation and structure of preferential attachments. The attachment policies found that do so are temporal and local triangle count attachment. This is significant because it shows that there may be other forms of attachment that can explain the creation of scale-free networks.

More specifically, for the two false-positive attachment policies, there are a few key findings. For the local triangle count attachment policy, the finding shows that there might be other network attributes similar to node degree that can contribute to the formation of scale-free networks. Although it is important to note that other attachment policies tested are influenced by node degree, such as mean geodesic distance and triadic closure, This was the only one that created a similar degree vs. attachment rate graph that was also a scale-free network. Additionally, the network formed from local triangle count, although it can be seen as a super-linear scale-free network, the alpha value is outside of the desired 2-3 range. Future work should explore this in more depth to see other attachment policies that are similar to node degree and local triangle count.

The other attachment policy that imitated preferential attachment was temporal attachment. The significance of this is that, although in this process time and node degree grow together, they do not directly impact one another like node degree and local triangle count. Since they can act as independent attributes, it is possible that a different class of parameters might also be responsible for scale-free network formation. Just as Vázquez was exploring attachment policies with respect to specific real-world networks (Vázquez, 2003), the findings from the temporal attachment simulation might indicate that there could be several different underlying reasons for the formation of various real-world networks. Future work should further the findings from this study and Vázquez to determine other attributes and methods that might lead to the formation of scale-free models. As mentioned before, the significance of this study is to question the accuracy of preferential attachment to see if there may be other factors that are being ignored in the formation of scale-free networks. With a better understanding of their formation, network analysis of scale-free models can better represent the root cause of their formation.

References

Barabási, A.-L., & Albert, R. (1999). Emergence of Scaling in Random Networks. Science , 286 (5439), 509–512. doi: 10.1126/science.286.5439.509

Barrat, A., Barthélemy, M., & Vespignani, A. (2004). Weighted Evolving Networks: Coupling Topology and Weight Dynamics. Physical Review Letters , 92 (22). doi: 10.1103/physrevlett.92.228701

Berger, N., Borgs, C., Chayes, J. T., Dsouza, R. M., & Kleinberg, R. D. (2005). Degree Distribution of Competition-Induced Preferential Attachment Graphs. Combinatorics, Probability and Computing , 14 (5-6), 697. doi: 10.1017/s0963548305006930

Newman, M. E. J. (2018). Networks (2nd ed.). Oxford, United Kingdom: Oxford University Press .

Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, CJ Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E.A. Quintero, Charles R Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. (2019) SciPy 1.0–Fundamental Algorithms for Scientific Computing in Python. preprint arXiv:1907.10121

Vázquez, A. (2003). Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. Physical Review E , 67 (5). doi: 10.1103/physreve.67.056104