Popularity Tutorial

Maya Dalton

September 24, 2024

What is popularity?

Popularity, also called *preferential attachment*, means that the more connected a node is, the more likely it is to receive new links. Nodes with higher degree have stronger ability to attract links added to the network. Barabasi and Albert (1999) introduced this concept based on the fact that vertices in many large networks follow a *scale-free power-law distribution*. This means: (i) networks expand continuously by the addition of new vertices, and (ii) new vertices attach to nodes that are already well connected.

An example of this would be modeling the growth of paper citations. Wu et. al. (2013) models this as new papers attach themselves to the citation network based only on the popularity of the currently existing papers. Simply, the "rich-get-richer" in terms of paper citations.

The Data for Today

Colby, Darren, 2021, "Chaos from Order: A Network Analysis of In-fighting Before and After El Chapo's Arrest". https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/L4AXQT

The author examines the relationship between leadership decapitation and in-fighting in drug cartels in Mexico. They hypothesize that leadership decapitation will weaken alliances between armed actors, lead to greater *preferential attachment* in networks of cartels and militias, and lead to decreased clustering as cartels seek to expand their power. Includes a network dataset of episodes of in-fighting among cartels and the militias five years before and after "El Chapo" (the former leader of the Sinaloa Cartel), was arrested in 2016.

A quick note: using `igraph`

For the first part of the tutorial, I'll use `igraph` for visualization, which is a bit different than the `network` package. If you run into errors in the later code, run:

```
detach("package:igraph", unload = TRUE)
```

Figure 1a. Cartel Network Pre−Arrest

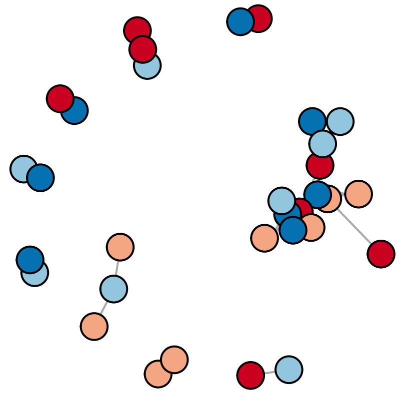
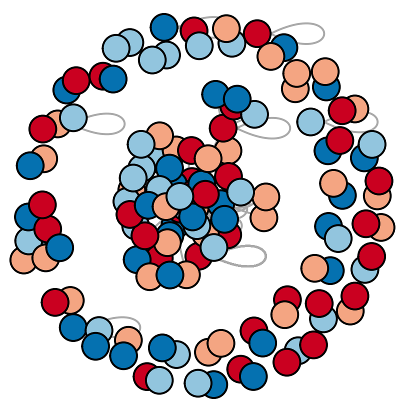


Figure 1b. Cartel Network Post−Arrest



Based on Figures 1a-b, there are a lot more nodes after El Chapo's arrest as groups engaged in in-fighting, disbanded, created off-shoot groups, and connected to others. We want to examine the attachment of new nodes to existing nodes using the concept of preferential attachment. First, let's look at degree centrality. In this case, a few nodes should have higher degrees, with many having smaller degrees. This signals that the higher degree nodes will attract the new nodes.

As shown in Figures 2a-b, before the arrest, the nodes were distributed across centrality scores. After the arrest, most nodes had low degrees with a handful having high degrees.

Figure 2a. Degree Centrality Distribution (Pre−Arrest)

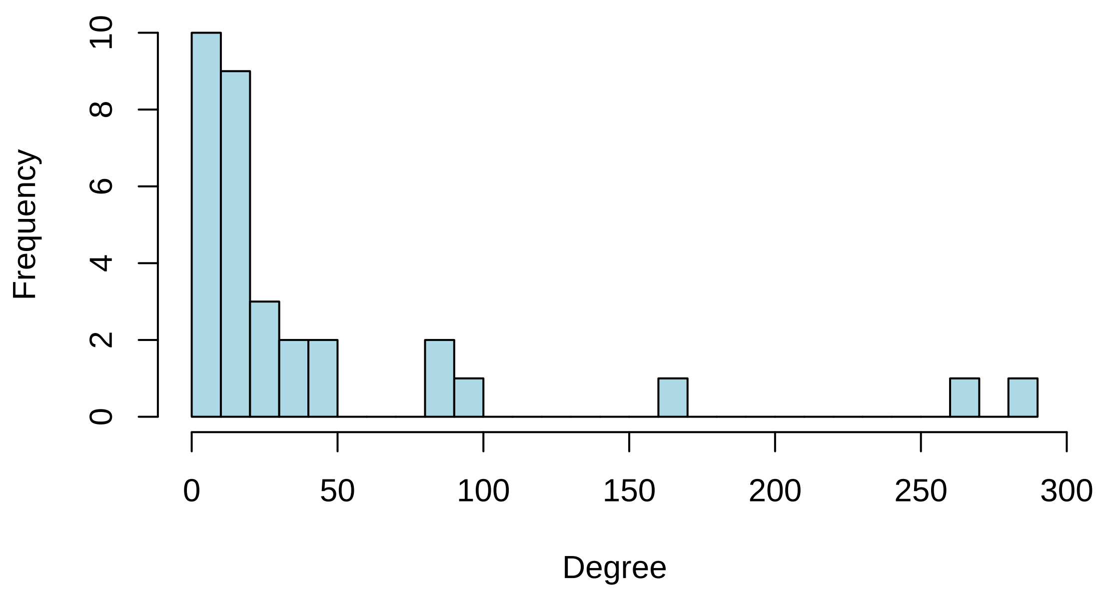
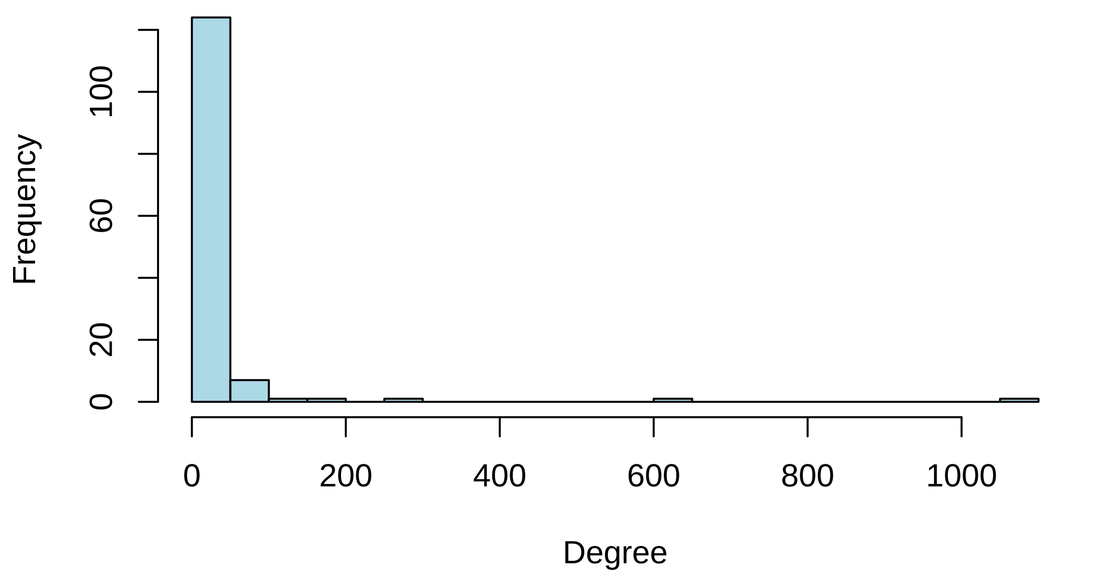


Figure 2b. Degree Centrality Distribution (Post−Arrest)



Now, we can also plot based on degree centrality. We see the same thing as the histograms. Before the arrest, there were only a few cartel groups engaging with one another, all with lower centralities. However, after the arrest, only a handful of cartels have larger degrees meaning they are more central to the network and “take in” disbanded cartels after El Chapo’s death.

Figure 3a. Pre−Arrest Network Based on Degree Centrality

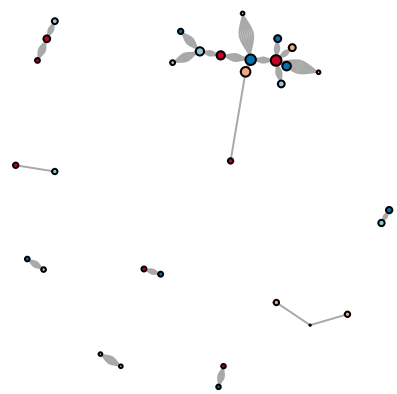
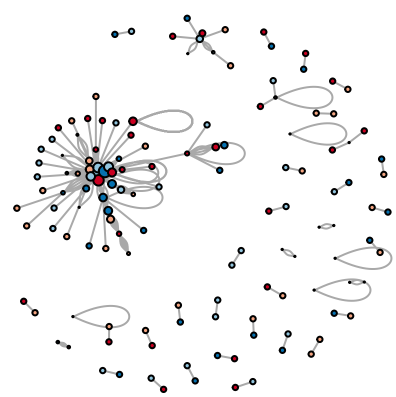


Figure 3b. Post−Arrest Network Based on Degree Centrality



Preferential Attachment - A Simulation

The **PAFit** package provides a framework to model attachment mechanisms. To show this package, I'll do some simulations first. I simulate a network starting with two nodes and one edge, then one new node with 5 new edges is added at each time-step until the number of nodes is N = 1000.

You can see in Figures 4a-b, the network at t=10 (4a) is smaller than at t=100 (4b), displaying the introduction of new nodes over time.

Figure 4a. Network at t=10

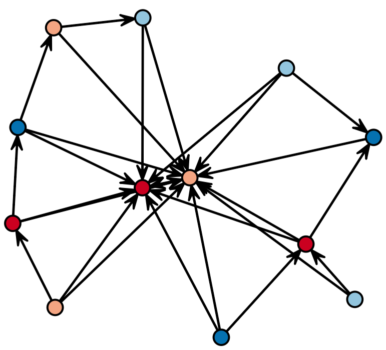
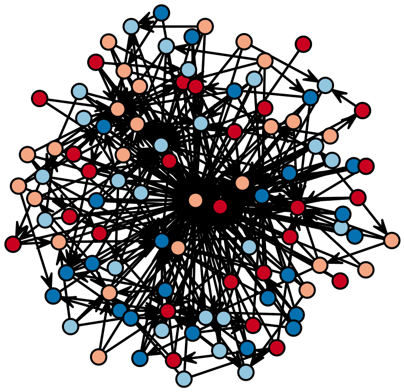


Figure 4b. Network at t=100



Figures 5a-b show the distribution of degree centrality at t=10 and t=100. Again, these distributions show that there are only a few nodes with higher centrality, with more nodes at lower centrality.

Figure 5a. Degree Distribution at t=10

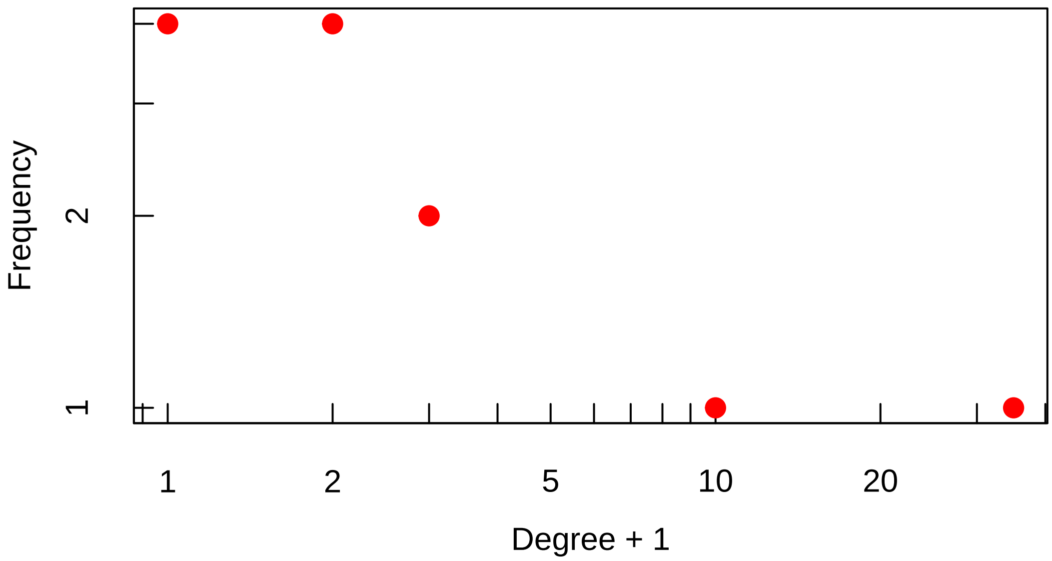
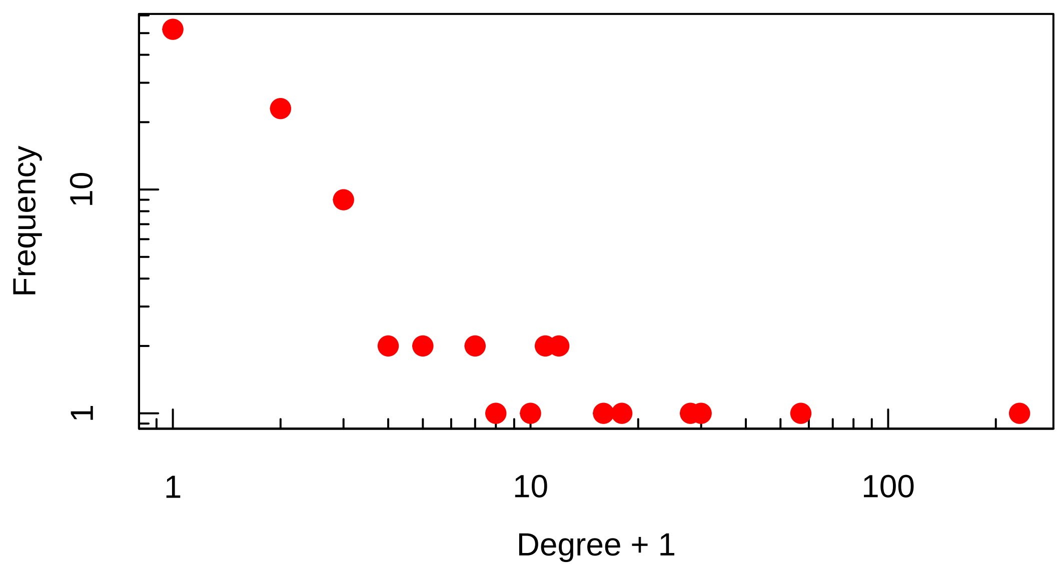


Figure 5b. Degree Distribution at t=100

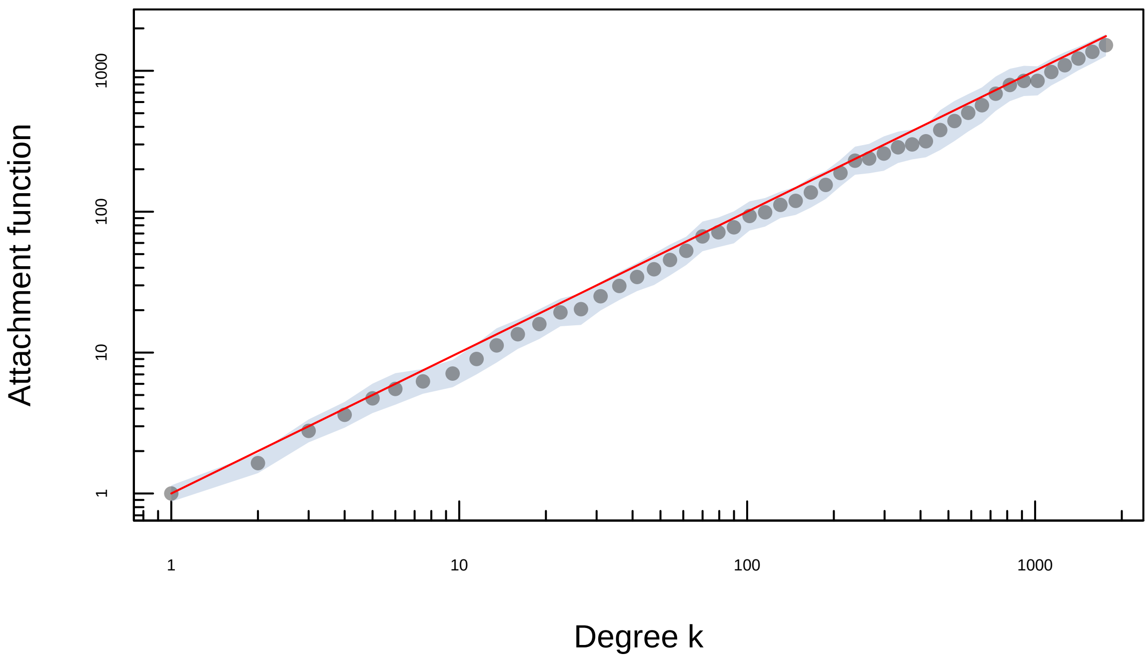


Now, we'll use this package to get the attachment exponents and examine the growth of the simulated network:

```
## Estimation results by the PAFit method.
## Mode: Only the attachment function was estimated.
## Selected r parameter: 0.1
## Estimated attachment exponent: 0.9945336
## Attachment exponent ± 2 s.d.: (0.9858287,1.003238)
## ------------------------------------------
## Additional information:
## Number of bins: 50
## Number of iterations: 29
## Stopping condition: 1e-08
```

Here, we see the estimated attachment exponent (α) is 0.994, meaning the network growth is almost **entirely driven by preferential attachment and new nodes are strictly more likely to connect to highly connected nodes**. In Figure 6 below, the PA coefficient increases across higher degrees in the network as well. This means that the new nodes are going to have very strong attachment to the highest degree nodes over time

Figure 6. Estimated vs. Actual PA Coefficients



Preferential Attachment - Cartel Data

Going back to the cartel data, we can do the same analysis of preferential attachment for the network. We need to first clean up the data and convert it into a PAFit object.

Similar to the simulation, the network at t=1 (which is 2012) includes a lot fewer groups than at t=9 (2020), displaying the introduction of many new nodes after El Chapo's arrest.

Figure 7a. Network at t=1 (2012)

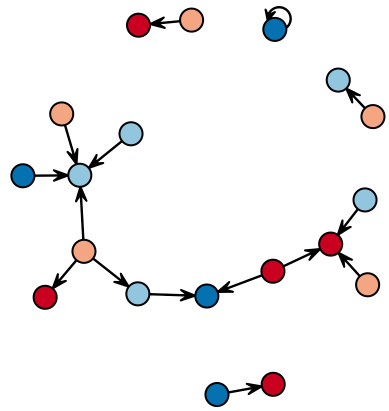
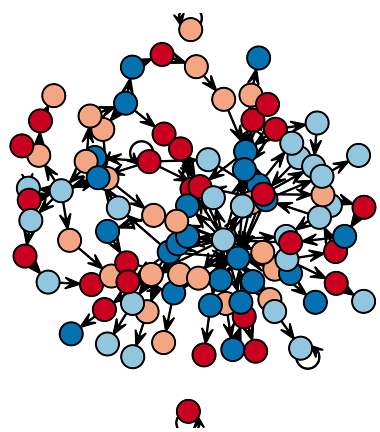


Figure 7b. Network at t=9 (2020)



The distribution plots are also similar to the simulation and Figures 2a and 2b, where t=1 (2012; pre-arrest) has very few nodes in general, as well as fewer nodes with high degrees. However, at t=9 (2020; post-arrest), the distribution shifts to include more nodes in the distribution, with a few nodes having high degrees and more with smaller degrees.

Figure 8a. Degree Distribution at t=1 (2012)

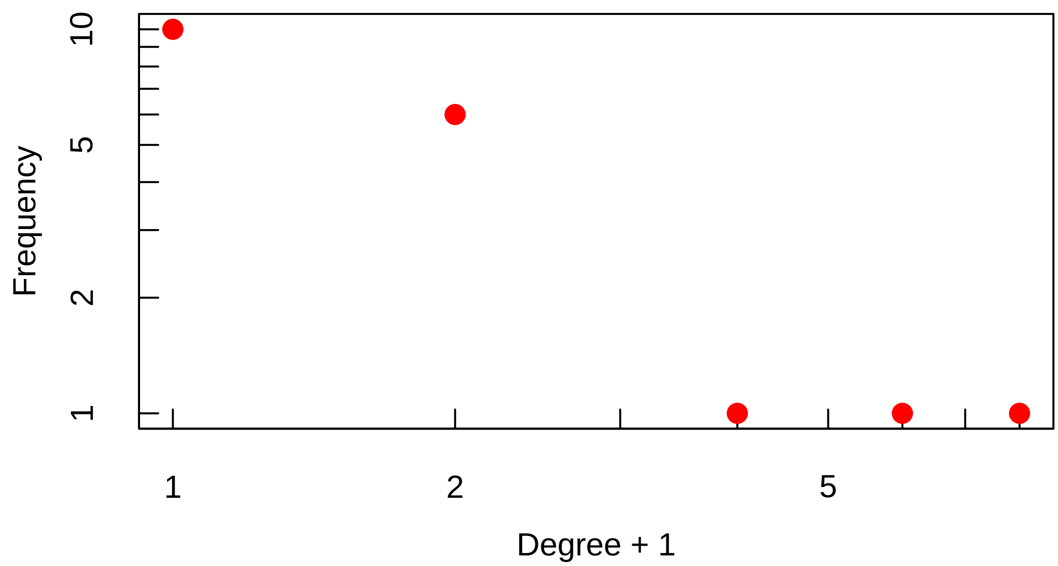
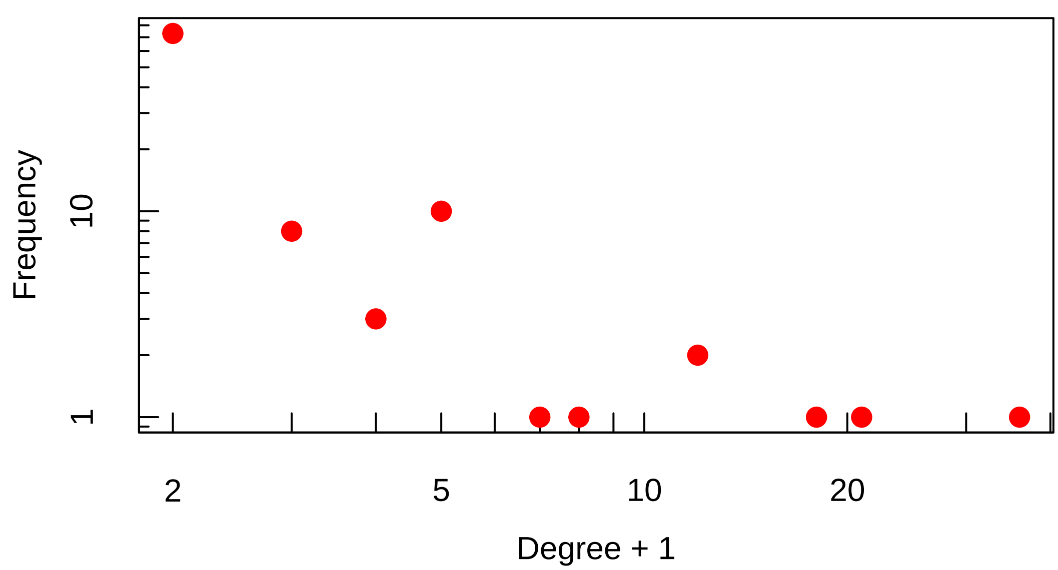


Figure 8b. Degree Distribution at t=9 (2020)




```
## Estimation results by the PAFit method.
## Mode: Only the attachment function was estimated.
## Selected r parameter: 0
## Estimated attachment exponent: 0.965378
## Attachment exponent ± 2 s.d.: (0.6535068,1.277249)
```

```
## ------------------------------------------
## Additional information:
## Number of bins: 36
## Number of iterations: 9
## Stopping condition: 1e-08
```

The estimated attachment exponent (α) is 0.965, meaning the network growth is almost entirely driven by preferential attachment and new nodes are strictly more likely to connect to highly connected nodes. In the context of the data, new off-shoot cartel groups are more likely to connect to established cartel groups.

But what if my data doesn't have a temporal aspect?

You can still use the PAFit package, however, it still expects you to have three columns in your edge list. So you'll need to make a dummy column of time steps which is just a column of 1s for every row.

```
## Containing the result of the PAFit_oneshot method.
## Estimated attachment exponent: 3.706431
## Attachment exponent ± 2 s.d.: (0.8692776,6.543583)
```

The estimated attachment exponent (α) is 3.904 without temporal aspects, meaning the attachment process is *super-linear*. Basically, high-degree nodes are even more likely to attract new connections. The higher the degree of a node, the disproportionately higher the probability that new nodes will attach to it. In the case of the data, the process is highly skewed toward high-degree cartel groups, which act as "hubs" and attract a lot of the new or disbanded groups. If one of these hubs were to suddenly disappear, it would significantly impact the network.

Thank you!!