Scale-freeness and Growth Stability of Realistic Network Models

Barabási-Albert Model

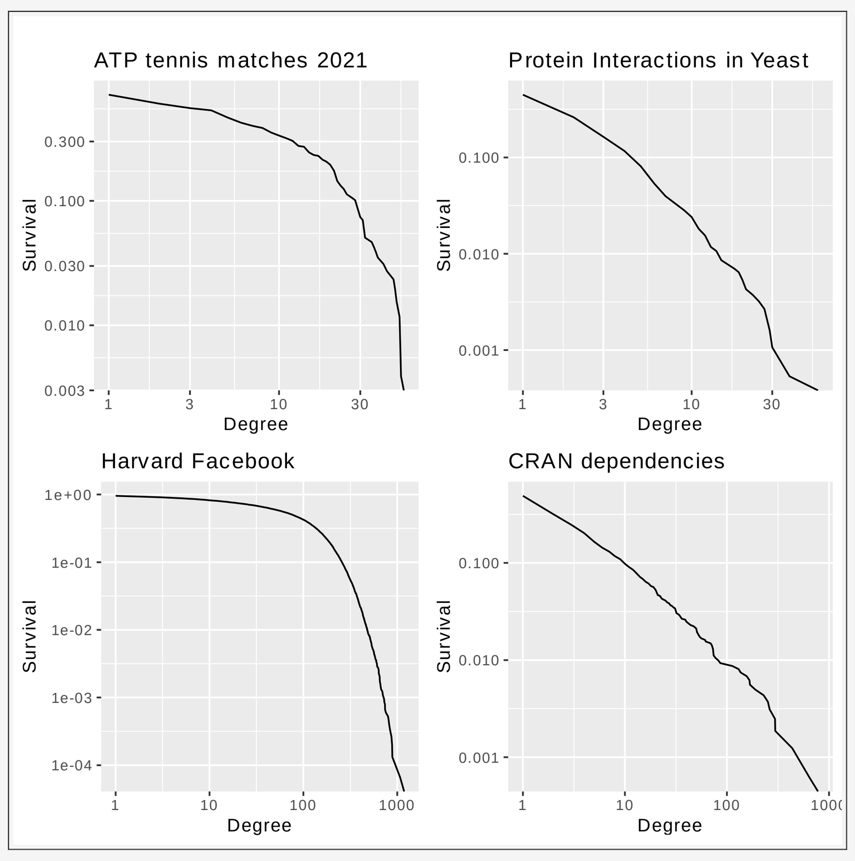
This is a model that generates networks in an interpretable and simple way. This is why it is often used to model real network formation. It works by repeating these steps:

1. **Growth:** At each time step, add a node to the network that will connect to $m \leq m_0$ nodes (already in the network) with m edges.

2. **Preferential Attachment:** The probability that an edge from the new node connects to node i is proportional to its current degree k_i i.e $k_i / \Sigma_j k_j$.

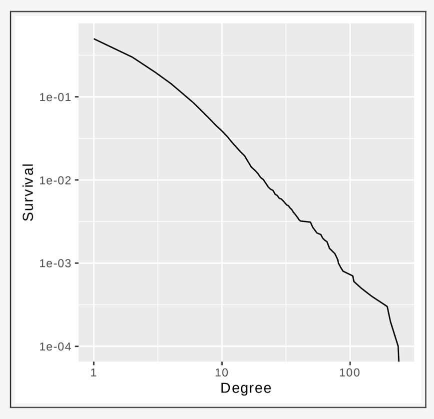
This model produces networks with a degree distribution that roughly follows a power law.

Some examples of the degree survival functions of real networks



Barabási-Albert Not Realistic?

While the Barabási-Albert model is attractive to use due to its simplicity and interpretability, it can only produce networks that have a power law degree distribution. An example of what this mean the degree survival function looks like is below:



Clearly this looks like some of the plots to left, but seems to look almost nothing like others. Our aim to to produce a model that can make networks with more realistic behaviour.

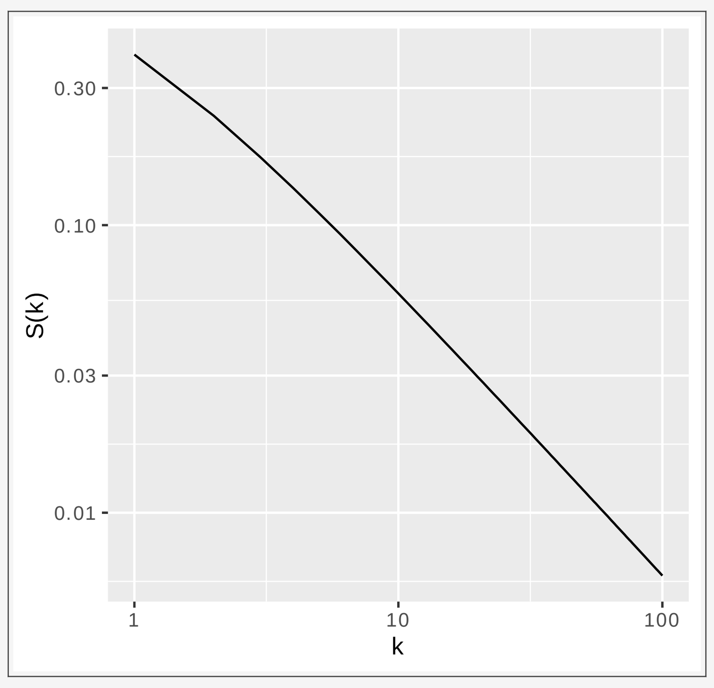
The Power Law

When we talk about a degree distribution following a power law, we mean that it is following the Zeta distribution which has probability mass function:

$$f(k) = \zeta(\alpha+1)^{-1} k^{-(\alpha+1)}, \quad k = 1,2,\ldots$$

For some $\alpha \in \mathbb{R}^{+}$.

This is a very simple model that is only characterised by only one parameter α. An example of what the survival function of this model looks like is below:



The value of α used above is $\alpha = 1$, on the log-log scale it is approximately a straight line and indeed it will be for any value of α.

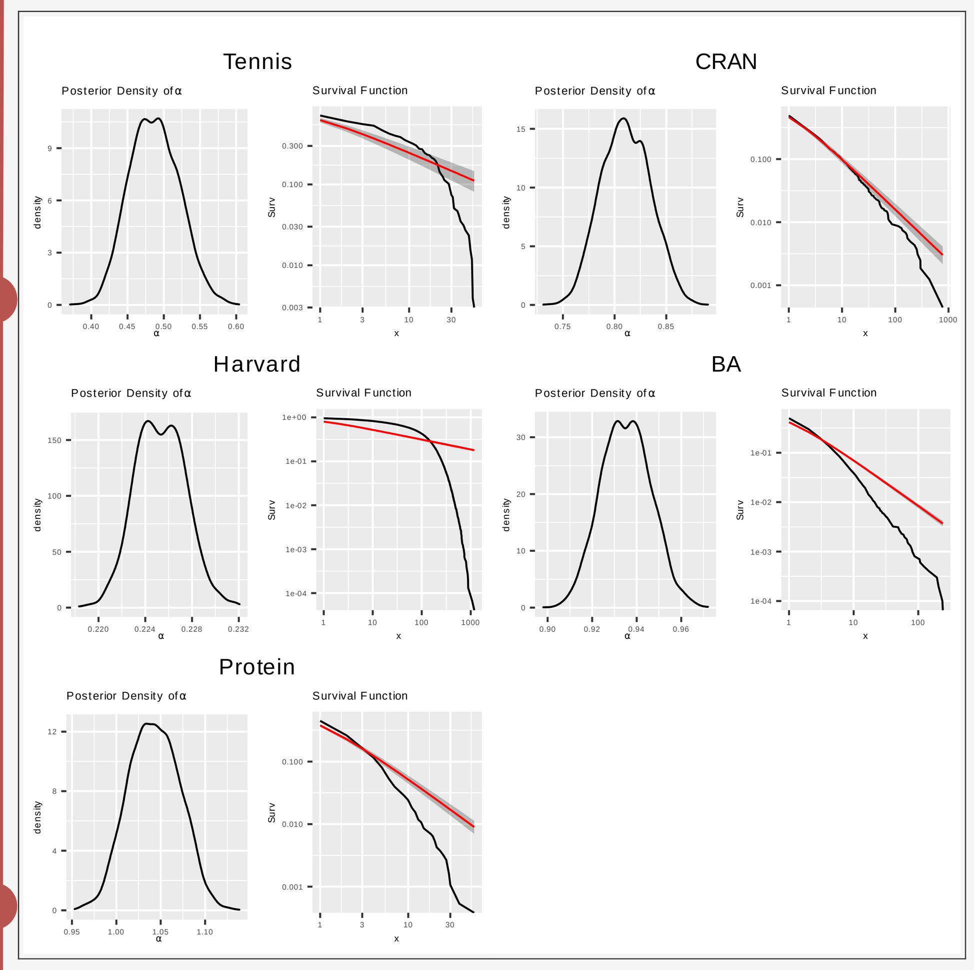
What are we doing?

Networks often arise when representing complex systems and the relationships between the components within them. They appear in a huge variety of fields of research from sports modelling to biology. This makes them a very important source of data. One area of interest concerns how these networks formed. We are interested in improving on the models currently used for this.

How good is the power law?

Since the survival function of a power law, it may seem good to use for some of our real networks. So that, in turn, the Barabási-Albert model can be used to model their growth.

However if we fit the power law to our real networks and the simulated one we get the below plots:



We can see that fitting the power law to the entirety of any of our data sets seems to be inadequate, even the simulated network that should have approximately a power law degree distribution.

The Power Law – Integer Generalised Pareto Mixture

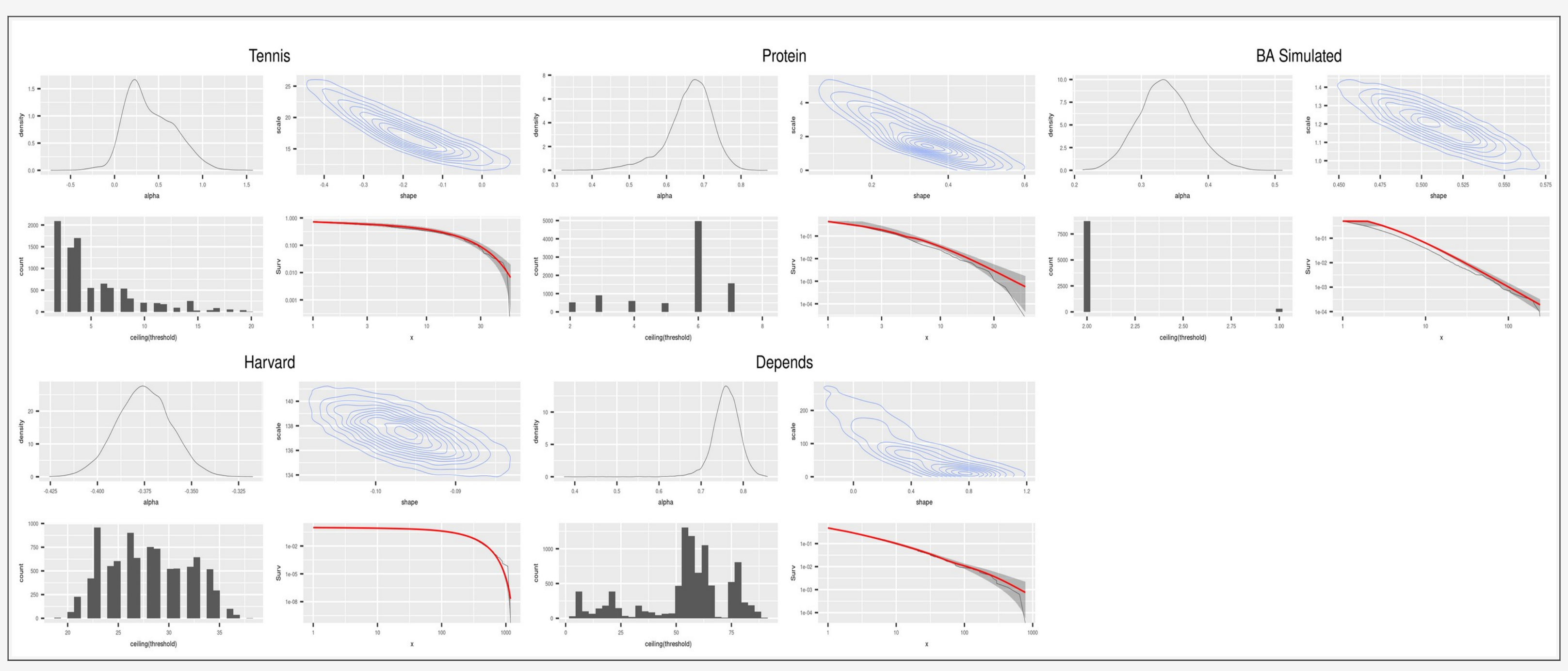
The survival curves of our datasets and indeed a lot of real data, seems to be a straight line in the bulk of the data but then falls off in the right tail. So, perhaps it would be a good idea to model the bulk of the data using a power law and the upper tail using methods from extreme value theory.

Normally a Generalised Pareto distribution (GPD) would be used to model the conditional distribution of the data over a certain threshold. However, our data is not continuous so we require a discretised version of the GPD. This is what we will call the Integer Generalised Pareto Distribution (IGPD).

This model has probability mass function:

$$\mathrm{f}(k) = (1-\phi)\mathbb{I}_{\leq v}(k)\, k^{-(\alpha+1)} \left[\zeta(\alpha+1) - \zeta(\alpha+1, v+1)\right]^{-1} + \phi\, \mathbb{I}_{>v}(k)\, G(k; v, \sigma_0, \xi), \quad x = 1,2,\ldots$$

These are the results of fitting this model to our data sets:

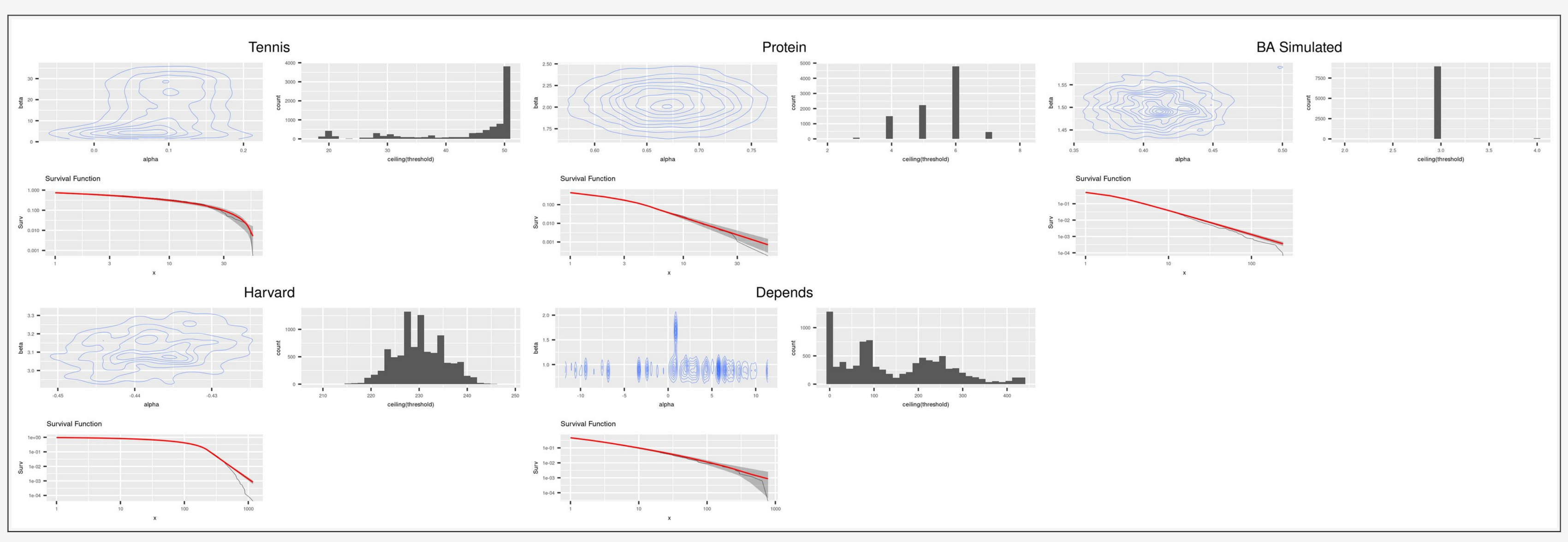


The Power Law Mixture

An alternate route without using extreme value theory, is to just use two power laws with different exponents to model the data. We use a right-truncated Zeta distribution for the bulk of the data, and a left-truncated Zeta distribution for the right tail. The probability mass function of this model is:

$$f(k) = (1-\phi) I_{\leq v}(k)\, k^{-(\alpha+1)} \left[\zeta(\alpha+1) - \zeta(\alpha+1, v+1)\right]^{-1} + \phi\, I_{>v}(k)\, k^{-(\beta+1)} \zeta(\beta+1, v+1), \quad x = 1,2,\ldots$$

The results of fitting this model to the data are below:



Conclusions and The Future

To conclude ...