Homework 1: Comparison Between Distributions

Ben Zuckier

Network Science II

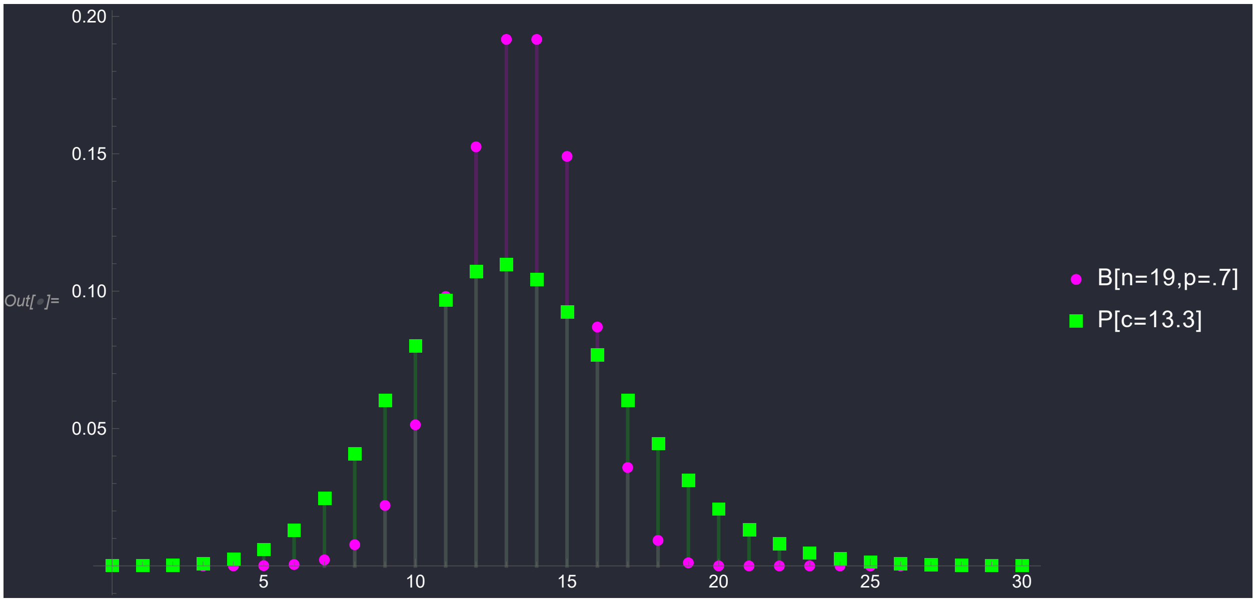
1 Binomial vs Poisson

First, we need to make sure that the average value of the degree of each distribution is the same. This can be computed by the expected value. The expected value for the Binomial Distribution is np and for the Poisson Distribution it's c. We just need to set np equal to c.

Now we have our chart comparing the distributions at different values of n, p, c, and k, where B[n,p] is the Binomial distribution with the n trials and probability of success p (evaluated at different values $x = k$ for each) and P[c] represents a Poisson distribution with mean c (evaluated at different values $x = k$).

	B[19,0.1]	P[1.9]	B[19,0.3]	P[5.7]	B[19,0.5]	P[9.5]	B[19,0.7]	P[13.3]
k=0	0.13	0.14	0.00	0.00	1.9e	0.00	1.2e-10	1.77e-6
k=1	0.28	0.28	0.00	0.01	0.00	0.00	5.2e-9	0.00
k=2	0.28	0.27	0.03	0.05	0.00	0.00	1.1e-7	0.00
k=3	0.17	0.17	0.08	0.10	0.00	0.01	1.4e-6	0.00
k=4	0.07	0.08	0.14	0.14	0.00	0.02	0.00	0.002
k=5	0.02	0.03	0.19	0.16	0.02	0.04	0.00	0.01
k=6	0.00	0.00	0.19	0.15	0.05	0.07	0.00	0.01
k=7	0.00	0.00	0.15	0.12	0.09	0.10	0.00	0.02
k=8	0.00	0.00	0.09	0.09	0.14	0.12	0.00	0.04
k=9	0.00	0.00	0.05	0.05	0.17	0.13	0.02	0.06
k=10	3.6e-6	0.00	0.02	0.03	0.18	0.12	0.05	0.8
k=11	3.3e-7	4.4e-6	0.01	0.02	0.14	0.11	0.10	0.10
k=12	2.4e-8	6.9e-7	0.00	0.01	0.10	0.08	0.15	0.11
k=13	1.4e-9	1.0e-7	0.00	0.00	0.05	0.06	0.19	0.11
k=14	6.9e-11	1.4e-8	0.00	0.00	0.02	0.04	0.19	0.10
k=15	2.5e-12	1.7e-9	0.00	0.00	0.01	0.03	0.15	0.09
k=16	7.1e-14	2.1e-10	1.4e-6	0.00	0.00	0.02	0.09	0.08
k=17	1.4e-15	2.3e-11	1.1e-7	0.00	0.00	0.01	0.04	0.06
k=18	1.7e-17	2.4e-12	5.2e-9	0.00	0.00	0.01	0.01	0.04
k=19	1.0e-19	2.4e-13	1.2e-10	6.3e-6	1.9e-6	0.00	0.00	0.03

We can clearly see that whenever average degree k is near $np = c$ (for the binomial and poissonian respectively) we get a good approximation and the value of the binomial and poissonian are similar. However, if k is smaller or larger, that quickly falls apart. We can see this also in the following graph of the Binomial with $n = 20$, $p = .7$ and the Poissonian with $c = 13.3$ and with k from 0 to 30.

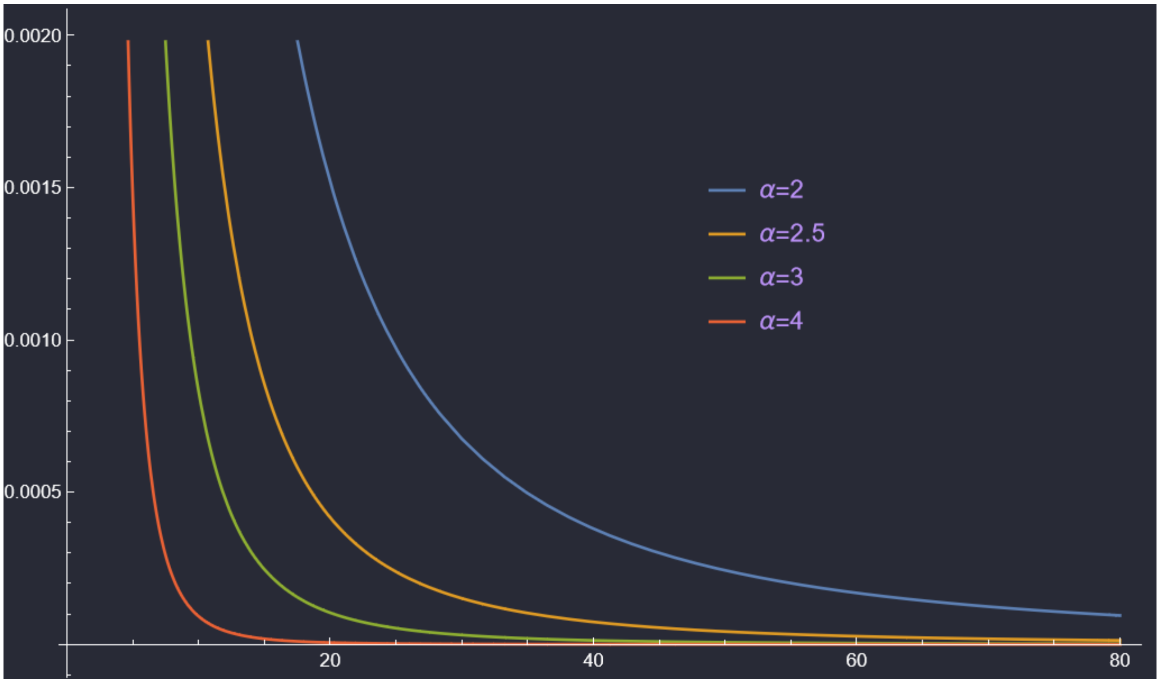

Out[ ]=
0.20
0.15
0.10
0.05
5
10
15
20
25
30
B[n=19,p=.7]
P[c=13.3]

2 Rate of Decay in Power Laws

1. Here we are comparing the decay of power laws with different exponents α, namely $\alpha \in \{2, 2.5, 3, 4\}$ where the power law is given by $\dfrac{k^{-\alpha}}{\zeta(\alpha)}$

   Note that there is no average degree for $\alpha = 2$ because $<k> = \dfrac{\zeta(\alpha-1)}{\zeta(\alpha)}$ and $\zeta(1) = \infty$. This can be explained as we discussed in class that the average degree scales with the size of the network and it's "winner takes all".

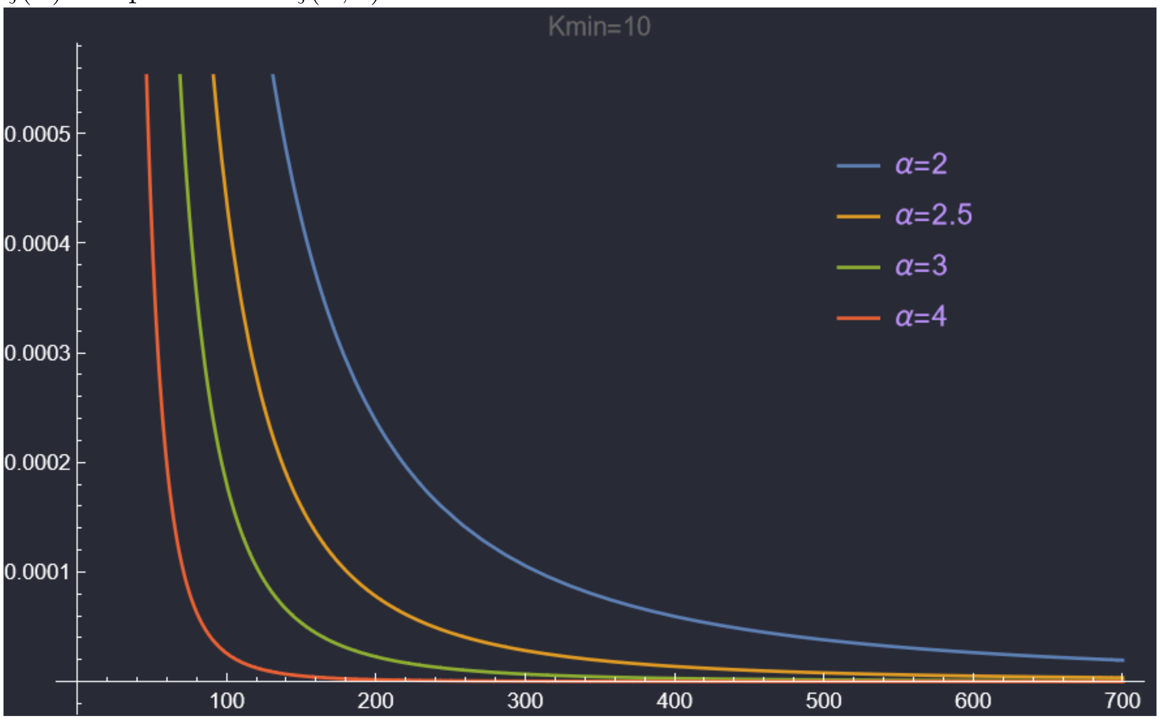
   Here we have a graph of these power laws from 1 to 80.



   Note that this graph is "zoomed" way in on the y axis.
   Already at $k = 36$ the values for each of the power laws differs by an order of magnitude as shown in this chart.

k =	**1**	**4**	**12**	**20**	**28**	**36**	**44**	**52**	**60**	**68**	**76**
$\boldsymbol{\alpha = 2.0}$	6e-1	4e-2	4e-3	2e-3	8e-4	5e-4	3e-4	2e-4	2e-4	1e-4	1e-4
$\boldsymbol{\alpha = 2.5}$	8e-1	2e-2	2e-3	4e-4	2e-4	1e-4	6e-5	4e-5	3e-5	2e-5	2e-5
$\boldsymbol{\alpha = 3.0}$	8e-1	1e-2	5e-4	1e-4	4e-5	2e-5	1e-5	6e-6	4e-6	3e-6	2e-6
$\boldsymbol{\alpha = 4.0}$	9e-1	4e-3	5e-5	6e-6	2e-6	6e-7	3e-7	1e-7	7e-8	4e-8	3e-8

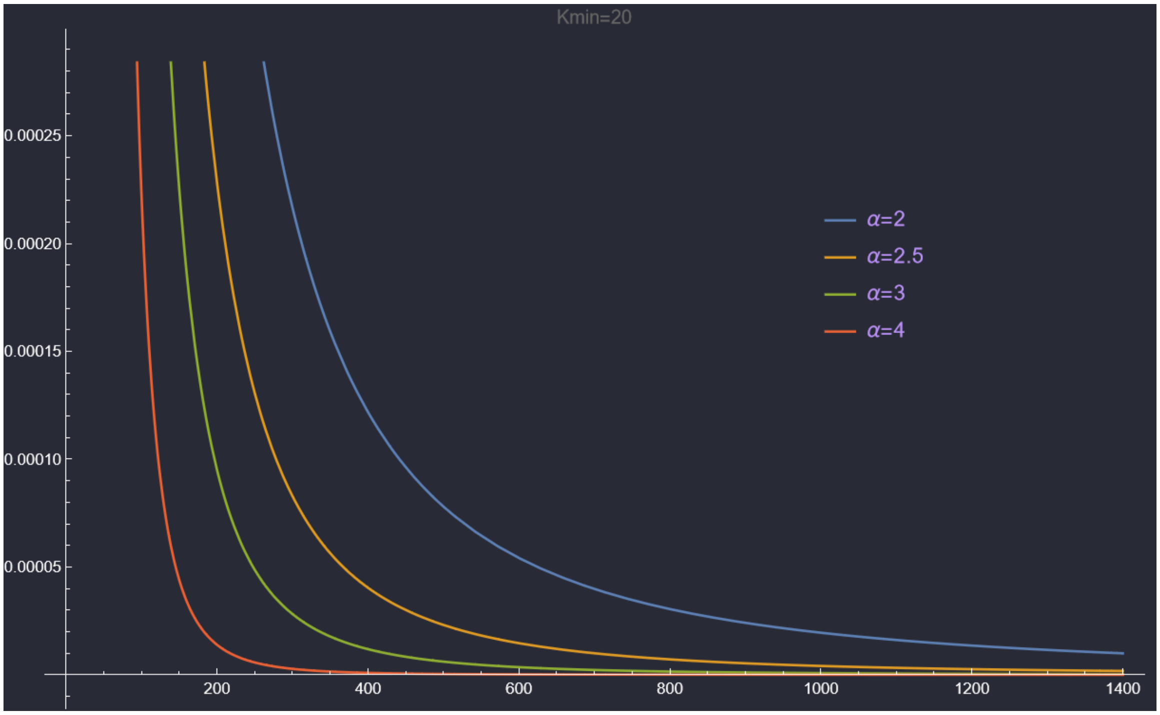
2. Now we will take a look at truncated power laws with $k_{\min} = 10$, again for αs $\in \{2, 2.5, 3, 4\}$ where the truncated power law is given by $\frac{k^{-\alpha}}{\zeta(\alpha, k_{\min})}$. The "usual" $\zeta(\alpha)$ is equivalent to $\zeta(\alpha, 1)$.



Note again that we are zoomed in on the y axis to be able to see what's happening when they diverge, especially here since that only becomes an order of magnitude around $k = 600$ as we can see in this table.

k =	**1**	**50**	**100**	**150**	**200**	**250**	**300**	**400**	**500**	**600**	**700**
$\boldsymbol{\alpha = 2.0}$	1e1	4e-3	1e-3	4e-4	2e-4	2e-4	1e-4	6e-5	4e-5	3e-5	2e-5
$\boldsymbol{\alpha = 2.5}$	4e1	2e-3	4e-4	2e-4	8e-5	4e-5	3e-5	1e-5	8e-6	5e-6	3e-6
$\boldsymbol{\alpha = 3.0}$	2e2	1e-3	2e-4	5e-5	2e-5	1e-5	7e-6	3e-6	1e-6	8e-7	5e-7
$\boldsymbol{\alpha = 4.0}$	3e3	4e-4	3e-5	5e-6	2e-6	7e-7	3e-7	1e-7	4e-8	2e-8	1e-8

Next we have the truncated power laws with $k_{\min} = 20$.

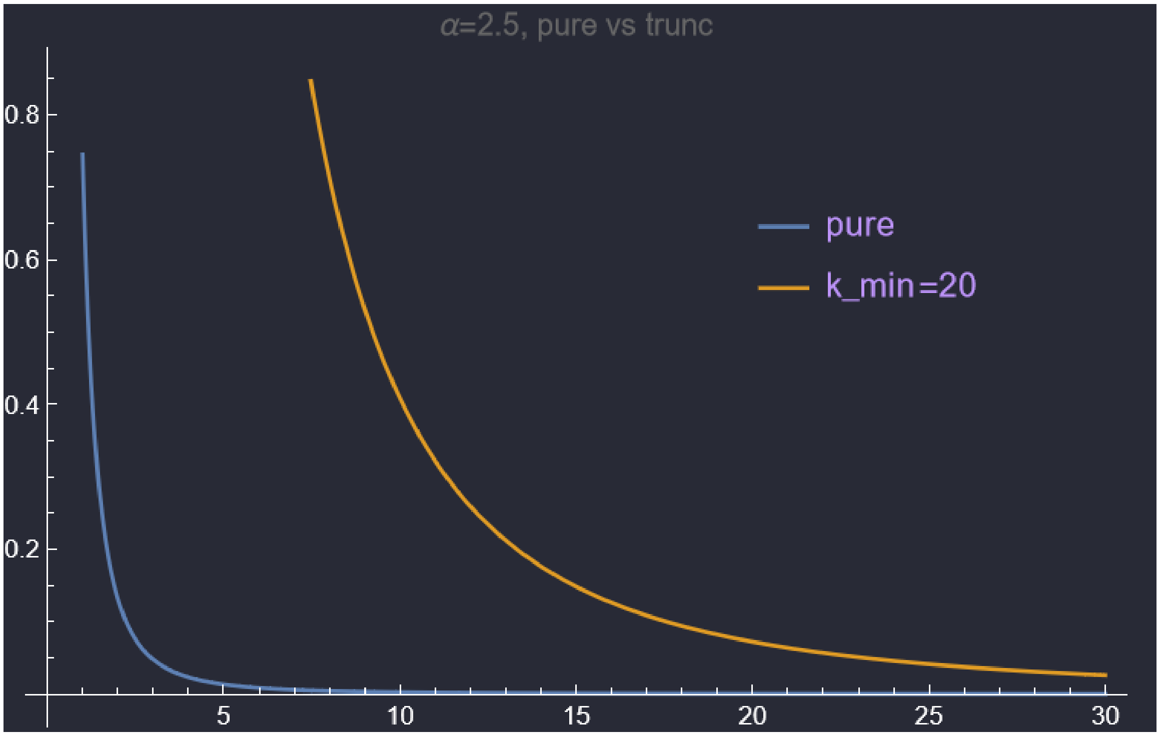


And its associated table:

k =	**1**	**40**	**100**	**200**	**300**	**400**	**500**	**700**	**900**	**1100**	**1300**
$\boldsymbol{\alpha = 2.0}$	2e1	1e-2	2e-3	5e-4	2e-4	1e-4	8e-5	4e-5	2e-5	2e-5	1e-5
$\boldsymbol{\alpha = 2.5}$	1e2	1e-2	1e-3	2e-4	8e-5	4e-5	2e-5	1e-5	5e-6	3e-6	2e-6
$\boldsymbol{\alpha = 3.0}$	8e2	1e-2	8e-4	1e-4	3e-5	1e-5	6e-6	2e-6	1e-6	6e-7	3e-7
$\boldsymbol{\alpha = 4.0}$	2e4	9e-3	2e-4	1e-5	3e-6	9e-7	4e-7	9e-8	3e-8	2e-8	8e-9

We can clearly see in these two cases of the truncated power law that it takes much larger values of k for the differences between the αs to become pronounced. Whereas it took until $k = 50$ for the pure (non-truncated) power law to differ by an order of magnitude, it took the truncated power law starting from $k_{\min} = 10$ until $k = 600$, and starting from $k_{\min} = 20$ until about $k = 1000$. It seems like the higher the $k_{\min}$ for the truncated power law, the slower the different values of α diverge.

3. Now we will compare the pure power law against the truncated power law of $k_{\min} = 20$, both with $\alpha = 2.5$.



Aside from the differing starting values here, we see that the graphs are radically different in shape – in the way they "tail off". The pure power law has a much steeper falloff whereas the truncated law is gradual, almost like a catenary. Here is the associated table ($\alpha = 2.5$).

k =	**1**	**10**	**20**	**30**	**40**	**50**	**150**	**300**	**450**	**600**	**750**
pure	7e-1	2e-3	4e-4	2e-4	7e-5	4e-5	3e-6	5e-7	2e-7	8e-8	5e-8
$\boldsymbol{k_m = 2.5}$	1e2	4e-1	7e-2	3e-2	1e-2	7e-3	5e-4	8e-5	3e-5	1e-5	8e-6

It's also interesting to note that from $k = 10$ to around $k = 40$ the difference between the two increases, peaking at $k = 40$, before tapering off until somewhere around $k = 800$ where it stays pretty even (and seems to increase again near $k = 1500$ but I didn't look too closely there).

3 ∂Decay of Power Law, Exponential, and Poissonian

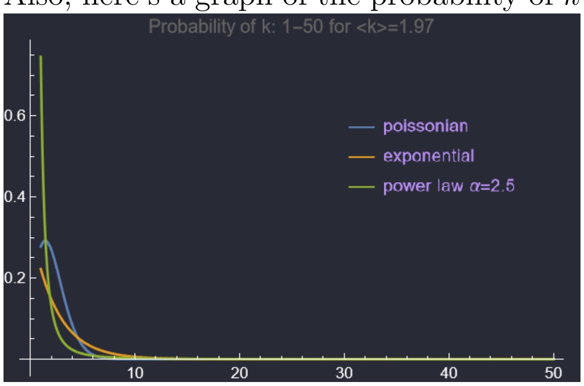
Now we want to compare the rate of decay of the Power Law, Exponential, and Poissonian. To make a comparison relevant, first we have to ensure that our selection of c for our Poissonian and λ for the exponential (and α for the power law) are such that the average degree of all of our distributions is the same. We have the power law, Exponential, and Poissonian (respectively) as

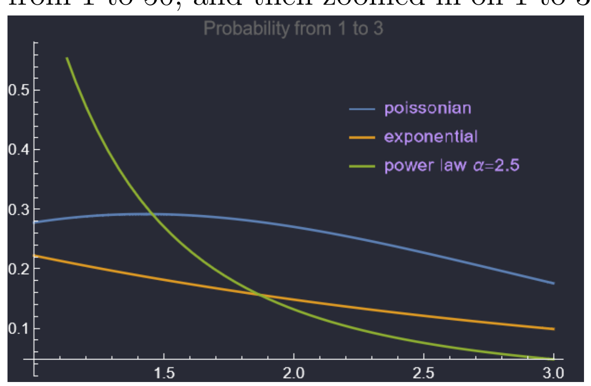
$$\frac{k^{-\alpha}}{\zeta(\alpha)}, \quad \frac{1-\exp(-\lambda)}{\exp(\lambda k)}, \quad \frac{c^k}{k!}\exp(-c)$$

1. We start with the "real world" case of $\alpha = 2.5$. We calculate that the average degree $< k >= \dfrac{\zeta(\alpha-1)}{\zeta(\alpha)} = 1.95$ (to two decimals). Trivially for the Poissonian, we just set $c = 1.95$. To find the value for our Exponential's λ, we can either solve for where the expected value over some values k is equal to our average degree, or using the equation we derived (making sure we sum from 1 to ∞ because we are only considering degrees of 1 and above) $c = \sum_{k=1}^{\infty} \exp\left(-\lambda^k\right) = \dfrac{1}{\exp(\lambda)-1}$. Either way, we find that $\lambda = 0.41$.

   If we find p_{50} for each of these distributions, we can find the probability that there will be a node with degree 50. For the poissonian, exponential, and power law (respectively), we find $p_{50} = 1.4 \times 10^{-51}$, $p_{50} = 5.2 \times 10^{-10}$, and $p_{50} = 4.2 \times 10^{-5}$.
   By taking the reciprocal of these numbers, we find the value for the number of nodes for a model in which we are likely to find one node of degree 50.

   That is: 7.2×10^{50} for the poissonian, 1.9×10^{9} for the exponential, and 2.4×10^{4} for the power law at $\alpha = 2.5$ (all normalized to the same average degree of approx 1.95 as explained above).
   Also, here's a graph of the probability of k from 1 to 50, and then zoomed in on 1 to 3



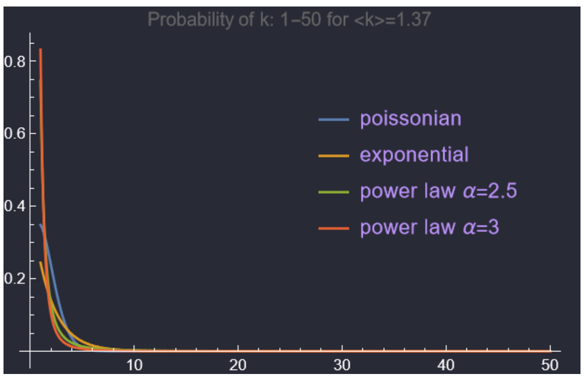


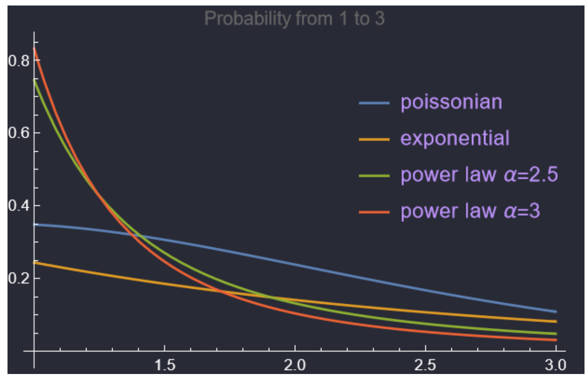
   In addition to the fact that the power law has a much higher probability of having a node of degree 50, it starts much higher as well but then goes under the poissonian at around $k = 1.5$ and the exponential near $k = 1.8$. At around $k = 6.2$ and onward the power law's probability is higher than the poissonian, and from $k = 15$ and onward it's higher than the exponential as well. This shows the much slower "tail off" rate of the power law here, both in the beginning (near 1, for values less than the average degree) and end (including near $k = 50$).

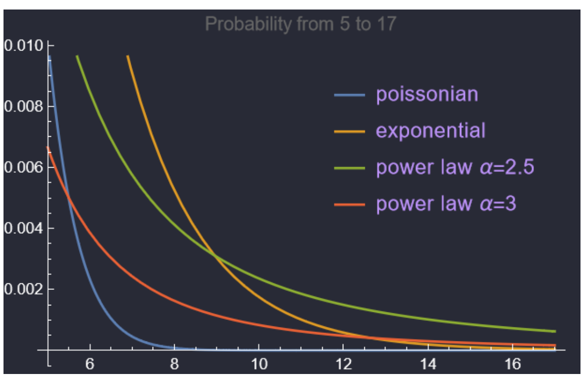
2. Now we move on to to Barabási-Albert model which is a power law of $\alpha = 3$. First we have to recompute the average value $< k >$ which is about 1.37. So we set $c = 1.37$ and calculate that $\lambda = 0.55$. We have a fairly similar conclusion.

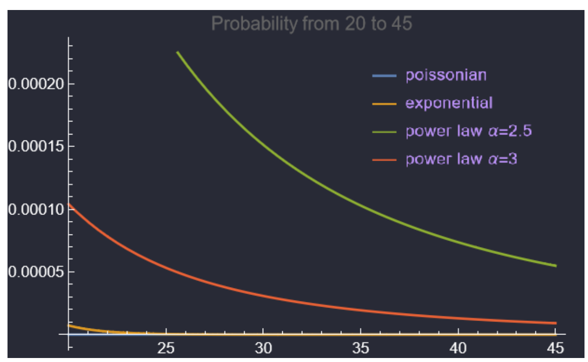
   We need models of size: 1.9×10^{58} for the poissonian, 1.9×10^{12} for the exponential, and 1.5×10^5 for the Barabási-Albert.

   Interestingly, the Barabási-Albert ($\alpha = 3$) needs about an order of magnitude more nodes than the power law of $\alpha = 2.5$ to have a high probability of a node of degree 50. In the following graphs we see that both power laws start out about the same but diverge steadily as the Barabási-Albert tails off more quickly than the power law of $\alpha = 2.5$.









4 Facebook (meta?)

1. Using a power law of $\alpha = 2.6$ we find that the probability of having 5000 friends is about 1.85×10^{-10}. But the network is very large, with about 2.85 billion nodes. As a result we expect 0.53 of a person to have 5000 connections, so still likely no one.

   To find the number of people with more than 5000 connections, we calculate $1 - \sum_{k=1}^{5000} p_k$ (where $p_k = (\zeta(\alpha)k^{\alpha})^{-1}$ is the probability of a node having degree k in the power law distribution of $\alpha = 2.6$) and find that the probability is 5.78×10^{-7}. Again, since the network is so large we find that there would be about 1646.39 people with more than 5000 connections. This only makes up about 0.0058% of the network (as we see from the probability), so it's not that significant.

2. If we assume that the Facebook network isn't a pure power law but is instead better modeled by a truncated power law, we have to find the k_{min} for such a network. We know that the average degree of the network is 190 so we solve using the average of a truncated power law as $\frac{\zeta(\alpha - 1, k_{\text{min}})}{\zeta(\alpha, k_{\text{min}})} = 190 \Rightarrow k_{\text{min}} = 71.75$. Rounding to the nearest integer, we get $k_{\text{min}} = 72$.

   The probability that a node will have degree greater than 190 is given by the following: $1 - \sum_{k=191}^{\infty} p_k = .79$ (where $p_k = (\zeta(\alpha, k_{\text{min}})k^{\alpha})^{-1}$ is the probability of a node having degree k in the truncated power law of $\alpha = 2.6$, k_{min}). Note: I used the highest number wolfram would allow instead of infinity, which was 10×10^{84}. This means that 2.26×10^9 nodes will have a degree greater than the average.

   Plugging 5000 directly into $p_k = (\zeta(\alpha, k_{\text{min}})k^{\alpha})^{-1} = 3.58 \times 10^{-7}$ or about 1019.88 nodes would have a degree of 5000 if we didn't impose Facebook's arbitrary friend limit.

   Similarly we take $1 - \sum_{k=0}^{5000} p_k = 1.12 \times 10^{-3}$, meaning about 11 percent of the network, or about 3.19×10^6 nodes, would have a degree larger than 5000. This is obviously significant.

3. Probabilities for $p_k|^{k \in \{10 \times 10^3,\, 20 \times 10^3,\, 50 \times 10^3,\, 100 \times 10^3\}}$

   $$\Rightarrow p_{10000} = 5.9 \times 10^{-8},\ p_{20000} = 9.7 \times 10^{-9},\ p_{50000} = 9 \times 10^{-10},\ p_{100000} = 1.5 \times 10^{-10}$$

   This yields (rounded to the nearest integer) the following expected number of nodes: 168 for $k = 10 \times 10^3$, 28 for $k = 20 \times 10^3$, 3 for $k = 50 \times 10^3$, and 0 for $k = 100 \times 10^3$. Succinctly, we would expect 10k, 20k and 50k, but no one(whole person) with 100k friends. We could set the p to different numbers to see if there would be a half a person with that number, or change our probability to look at a range (between 100k and 200k for example).

Next we can solve for the largest number of nodes k where $p_k N \geq 1$ (at least one whole person) where $N = 3.5 \times 10^9$ the number of nodes in the FB network.

$$1 = p_k N \Rightarrow \texttt{floor}(k) = 71796$$

Therefore the largest number of friends we expect (at least one whole person to have) is 71796.

Note: we can also solve for "5000 friends or more" and we get the alarming max number of 5.8×10^7.