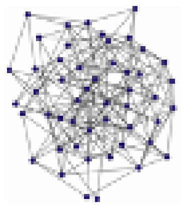
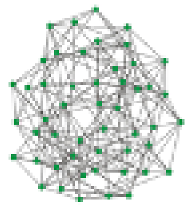
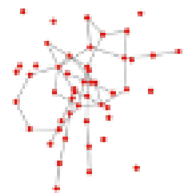


HOMEWORK ASSIGNMENT 1: COMPARISON BETWEEN DISTRIBUTIONS

ENGINEERING 5350
NETWORKS THEORY II FALL 2021
PROFESSOR GABRIEL CWILICH

PROBLEM 1: Binomial distribution versus Poisson distribution

The binomial distribution of probabilities describes a situation when there are N independent identical variables, each of which can take two possible states, with probability p and $1-p$ respectively. Say, a bunch of non-interacting spins that can take two states "up" or "down". It is clear that in such a system, the probability of having exactly k spins pointing "up" (and $N-k$ pointing down) will be just

$$p_k = p^k (1-p)^{N-k} \binom{N}{k}$$

As we saw last semester, this is very useful in the context of the Erdös-Renyi model of a network of n nodes to describe probability of a node having degree k, simply because that node might be connected (or not be connected) to each of the other $n-1$ nodes with probability p (or $1-p$) . So we get

$$p_k = p^k (1-p)^{n-1-k} \binom{n-1}{k}$$

We also saw that this binomial distribution can be approximated in the case when $n \to \infty$ and $p \to 0$, in such a way that $n \times p \to c$ by the Poisson distribution.

$$p_k = \frac{c^k}{k!} \exp^{-c}$$

To be precise c which is the average degree of this Poisson distribution happens to be $c \equiv (n-1)p$ but in the limit of $n \to \infty$ of course $(n-1)p \approx np$.

(1) Compare the binomial and the Poissonian distribution by looking at several values of the corresponding p_k, and displaying them in a table. Remember that for the comparison to be meaningful the average value of the degree of the two distributions has to be the same, so pick the c of the Poissonian, and the n and the p of the binomial accordingly, Choose values of k so that you can compare how good the approximation is when k is smaller, of the order, bigger, much bigger, and much much bigger than average degree.

(2) Repeat this for three different values of the average degree, to see if the approximation remains similarly good for different values of it.

(3) For at least one of the choices of the average degree (probably one where it is not too close to one, plot carefully the p_k for the Poissonian and the binomial on one graph to see how the values overlap.

PROBLEM 2: Rate of decay of different power laws

We will here compare the decay of power laws with different exponents α to see how they behave. We will concentrate in the following cases:

a) a very slow decaying power law (with $\alpha = 2$) which as we will see later leads to a system where the node or nodes of highest degrees are connected to almost any other node in the network (condensation).
b) a slow decaying power law characteristic of many real life networks like social networks, etc. ($\alpha = 2.5$).
c) The marginal case of the Barabasi-Albert ($\alpha = 3$).
d) A fast decaying power law ($\alpha = 4$) . In fact decaying so fast that it will be hard to pick it up from experimental data, since you might confuse it with an exponential decay.

The degrees should be normalized, of course, so that the comparison is meaningful. You could use the approximate normalization which is valid in the tail of the distribution of the exact normalization, as discussed in the lecture. But be consistent: in all cases use the same one. If you have access to some software that gives numerical values of the Zeta function (like Mathematica does), perhaps the exact normalization is the easiest.

$$p_k = \frac{1}{\zeta(\alpha)} k^{-\alpha}$$

This is for the case of a pure power law starting from $k = 1$,

$$p_k = \frac{1}{\zeta(\alpha, k_{min})} k^{-\alpha}$$

while this is for a truncated power law starting from $k = k_{min}$

Remember that the different distributions will not have now the same average degree, since as we saw in class that average degree (for the case of exact normalization) depends on the exponent α, and is:

$$< k >= \frac{\zeta(\alpha-1)}{\zeta(\alpha)} \qquad and \qquad < k >= \frac{\zeta(\alpha-1, k_{min})}{\zeta(\alpha, k_{min})}$$

respectively. Obviously for each value of α you get a different average degree, but those numbers do not change that much. For example $< k >_{\alpha=2.5}= 1.94737$, $< k >_{\alpha=3}= 1.36843$ and $< k >_{\alpha=4}= 1.110363$. The average degree for $\alpha = 2$ is a special case, since it diverges. Why is that? That means that, since we are looking at the values of k so far from the average of the distribution (in the tail), it does not matter that much that the average degrees of the distributions we are comparing are not the same

(1) For the case of pure power laws, compare the values of the p_k for the four cases described above. Perhaps you will see better the difference if you compare by pairs (2 vs 2.5 , then 2.5 vs 3, then 3 vs 4). Look for different values of k until the difference in p_k for different α becomes as big as an order of magnitude at least to get a sense of how they decay.

(2) Repeat the same study but now for a non-pure power law. One starting from $k_{min} = 10$ and one for $k_{min} = 20$, to observe that the differences are not as marked, and you need to get too much bigger values of k to see those differences.

(3) Compare a pure power law, with a non-pure one, but for the same exponent α

PROBLEM 3: Comparing the rate of decay of a power law, an exponential and a Poissonian

Now we want to compare the two fast decaying distributions we have been dealing with so far (the Poissonian associated with the Erdös-Renyi model ($p_k \propto \frac{c^k}{k!}$), and the decaying exponential, ($p_k \propto \exp^{-\lambda k}$) and the slow decaying power laws (which have $p_k \propto k^{-\alpha}$).

Remember that you need to use the normalized expressions for p_k

$$p_k = \frac{c^k}{k!} exp{-c} \qquad p_k = (\exp^{\lambda} - 1) \exp^{-\lambda k} \qquad p_k = \frac{1}{\zeta(\alpha)} k^{-\alpha}.$$

Make sure that you pick values of c and λ so that the average degree of those distributions is the same as the average degree of the power law, so that the comparison is meaningful.

(1) Compare the three distributions for the case in which the power law is a typical real world distribution ($\alpha = 2.5$), which has an average degree of around 2. How big should a network be so that one can expect to find at least one node of degree 50 (twenty five times bigger than the average degree) for each of the three distributions? What does that tell you about the different distributions?

(2) Repeat the same for the case when the power law is the Barabasi-Albert one ($\alpha = 3$). Do you obtain similar conclusions?

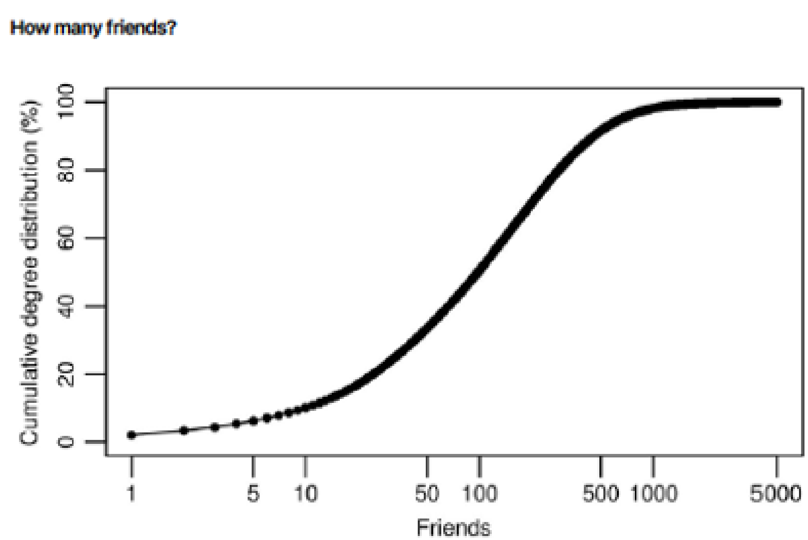
PROBLEM 4: THE FACEBOOK NETWORK

Let us try to see if we can infer something about Facebook. This site gives you some information:

https://www.facebook.com/notes/10158927855913415/

but it does not give you the exponent of the distribution, although perhaps you can find it somewhere else. The network has 2.85 billion nodes. I will assume that the power law distribution that describes the degrees of the nodes (the distribution of the number of friends people have) has an exponent of $\alpha = 2.6$.

(1) If that were the case how many people would you expect to have 5,000 friends? How many people would have <u>more</u> than 5000 friends? Would that be a significant number?



(2) Yet, Facebook puts a cap on the number of friends you can have. You are not allowed to have more than 5,000 friends. One explanation is that it is not a ”pure” power law, starting from $k = 1$, but only starting from a certain value quite bigger than 1. Using the information they give that the average number of friends on the network is 190, calculate what should be the k_{min}. You can do that numerically using the distributions discussed. Then using that value of k_{min}, from the graph above they publish in that site, you can get how many nodes approximately have a degree bigger than that value. Calculate now for the resulting network how many nodes of degree 5000 would you get if you did not apply this arbitrary limit? How many nodes of degree <u>larger</u> than 5000?

(3) In that network network, would you expect to get nodes of degree 10,000, 20,000, 50,000, 100,000? What would be the highest node you would expect?