**Exercise 0.1.** Show that $\ln n = \Theta(\lg n)$, and $\lg n = \Theta(\log n)$.

$\ln n$, $\lg n$ and $\log n$ are the usual notations for $\log$ to base $e$, $2$ and $10$ respectively. This is to show that $\log$ functions to different constant bases differ only by constant factors.

**Answer.** From the change of base formula $\log_e n = \frac{\log_2 n}{\log_2 e}$, therefore $\ln n = \frac{\lg n}{\log_e 2}$. Similarly, $\lg n = \frac{\log n}{\log_e 2}$. Since both $\frac{1}{\log_e 2}$ and $\frac{1}{\log_e 2}$ are constants, it follows that $\ln n = \Theta(\lg n)$ and $\lg n = \Theta(\log n)$.

**Exercise 0.2.** Set up the ipython notebook on a system of your choice with networkx. Try it out.

**Exercise 0.3.** Write code to create plots showing the threshold phenomenon for existence of isolated vertices.

**Exercise 0.4.** **Coupon collector problem.** Suppose they are giving out one coupon in each cereal boxes. There are $n$ different types of coupons. You have to collect all $n$ types to win a prize. Show that in expectation you need to buy $n \ln n$ boxes to to win the prize.

**Answer.** Let $X_i$ be the random variable denoting the number of cereal boxes to buy to get $i^{th}$ type coupon after collecting $(i - 1)$ types of coupons. Thus, to collect all $n$ types of coupons we need to buy $X = X_1 + X_2 + \ldots + X_n$ cereal boxes.

Trivially, $X_1 = 1$. As all the coupons are equiprobable, after collecting $i$ types of coupon, the probability to get a new type of coupon in the next cereal box is $\frac{n - i}{n}$. Therefore, expected number of draws to get $i^{th}$ type of coupon after collecting $(i - 1)$ types of coupons is $E(X_i) = \frac{n}{n - i}$. Therefore,

$$
E(X) = E(X_1 + X_2 + \ldots + X_n)
= E(X_1) + E(X_2) + \ldots + E(X_n)
= 1 + \frac{n}{n - 1} + \ldots + \frac{n}{n - i} + \ldots + n
= n \sum_{i=1}^{n} \frac{1}{i}
= \Theta(n \ln n)
$$

The question asked for exactly $n \ln n$, which is not quite correct. However, the actual value closer to $n \ln n$ than the $\Theta$ suggests. (see wikipedia entry for harmonic series: $\sum \frac{1}{i}$.)
Exercise 0.5. Show that for a suitable constant $c$, buying $cn \ln n$ boxes suffices to guarantee that you get at least one coupon of each type with high probability.

Let $A_i$ be the event that $i^{th}$ type of coupon is not collected after buying $k$ boxes. Thus, $P(A_i) = (1 - \frac{1}{n})^k$. Probability of getting at least one of all types of coupons is $(1 - P(\bigcup_{i=1}^{n} A_i))$.

From union bound rule we know that the probability of not getting some types of coupons is:

$$P(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i) = \sum_{i=1}^{n} (1 - \frac{1}{n})^k = \sum_{i=1}^{n} (1 - \frac{1}{n})^{cn \ln n} \leq \sum_{i=1}^{n} e^{-c \ln n} = \sum_{i=1}^{n} n^{-c} = n^{1-c}$$

Therefore, with high probability, we get at least one coupon of each type.

* Exercise 0.6. Random graphs are connected. In class, we showed that above the threshold $p = \ln n/(n - 1)$, isolated vertices are unlikely. However, this does not say that the graph overall is connected. It is possible that the graph itself stays in to two (or more) different connected components with no edge bridging them.

In this exercise, show that this is also unlikely. That is, above the threshold, with high probability, the graph has only a single connected component. [Hint: Formulate the problem to show that for a partition of the graph into multiple subsets, it is unlikely that there will be no edge out of any of them.]

Answer. This was one of the results by Erdos and Renyi. The proof is a bit long, see for example a description in https://www.math.cmu.edu/~af1p/MAA2005/L2.pdf

Exercise 0.7. Show that a connected graph has at least $\Omega(n)$ triads.

Answer. As per the definition of $\Omega(n)$, we need to show that for some constant $c > 0$, number of triad in a connected graph $T > c.n$ for $n > n_0$. 

2
We would prove this by induction. In a minimally connected graph, a tree with three vertices, the number of triad is one. This constitutes the base case with $c = 1/3$.

For induction, we are assuming there is a graph $G$ with $n$ nodes and $T$ triads. In $G$, $T \geq \frac{n}{3}$. Now, we’ll show that after adding nodes and edges to $G$, it still holds this property.

- Adding an edge to $G$: Adding an edge to $G$ only increases the number of triads. Thus, the number of triads in the new graph $T' \geq T \geq \frac{n}{3}$.

- Adding an vertex to $G$: As the new graph is to be connected, there should be at least an edge connecting the newly added vertex ($i$) to one of the vertices ($j$) in $G$. As, $G$ was connected there was at least an edge $jk$. Thus, $ijk$ is a triad in the new graph. Thus, $T' \geq T + 1 \geq \frac{n}{3} + 1 > \frac{1}{3}(n + 1)$.

Therefore, for $n \geq 3$, $T \geq c.n$ where $c = \frac{1}{3}$.