7. Graph Properties and Metrics - II

Bookmark this page

Exercises due Oct 20, 2021 17:29 IST Completed

Small Graph

6/6 points (graded)

Download the file directed_graph.txt. You may use any computational tool for this problem.

The file is in the **edge list** format with each line "i j" indicating a directed edge from vertex i to vertex j.

For example, the edge list of the graph $G = (\{0, 1, 2\}, \{\{0, 1\}, \{0, 2\}\})$ is

0 1

0 2

The example above assumes an undirected graph, but the same edge list can represent the directed graph $G' = (\{0, 1, 2\}, \{(0, 1), (0, 2)\})$.

Answer the following questions.

1. How many nodes does the graph have?

100 ✔ **Answer:** 100

2. How many edges does the graph possess?

1030 ✔ **Answer:** 1030

3. Does the graph contain self loops?

- (•) Yes
- ( ) No

✔

4. Does the graph have directed cycles not involving self loops? *Hint:* Think along the lines of the presence or absence of nodes with no incoming edges other than self loops.

- (•) Yes
- ( ) No

✔

5. This graph was generated using the following model: We fixed the number of nodes and every possible directed edge was selected with a probability p independently of all other edges. Note that in the graph we also consider possible edges that construct self loops. What is the maximum likelihood estimate of p? Provide an answer accurate to at least **five significant figures** (graded to 0.01% tolerance), you can also enter your answer as an exact fraction.

0.103 ✔ **Answer:** 1030/10000

6. Given the above generative model, what is the p-value of observing this graph for the null hypothesis that $p = 0.1$? Assume that the test statistic is $\left|\frac{S_n - \mu_n}{\sigma_n}\right|$, where S_n is the sample mean and μ_n and σ_n are the mean and standard deviation of the sample mean under the null hypothesis. The sample mean is the fraction of number of edges observed to the total number of possible edges. Use the central limit theorem approximation to compute the required probability. Provide an answer accurate to at least three **significant figures** (graded to 1% tolerance).

0.3173105078629159 ✔ **Answer:** 0.3173

Solution:

1. **100.**
2. **1030.**
3. **Yes.** This is evident from examining the edge list for any edge "$i\ i$". For instance, vertex 11 has a self loop.
4. **Yes.** To see this, first disregard the self loops in the graph. Then, since every node has an incoming edge it implies that there exists a directed cycle in the graph. Alternatively, you can also write a program to find a directed cycle in the given graph.
5. **1030/10000.** The maximum likelihood estimate in this case is the number of edges observed divided by the total number of possible edges.
6. First, note that $S_n = 1030/10000$. Under the null hypothesis, $\mu_n = 0.1$ and $\sigma_n = \sqrt{\frac{0.1 \cdot 0.9}{10000}}$. These values come from the central limit theorem for finding the mean of 10000 random numbers distributed according to Binomial(1, 0.1). Hence, the p-value is

$$\begin{aligned} p &= 2 \times P\left(|Z| > \left|\frac{S_n - \mu_n}{\sigma_n}\right|\right) \\ &\approx 2 \times P\left(|Z| > 1\right) \\ &\approx 0.3173 \end{aligned}$$

where Z is a standard normal random variable.

Submit You have used 3 of 6 attempts

Answers are displayed within the problem

Power Law Distribution

The **power law** distribution is defined by the following $\log$-$\log$ relationship between k and p_k:

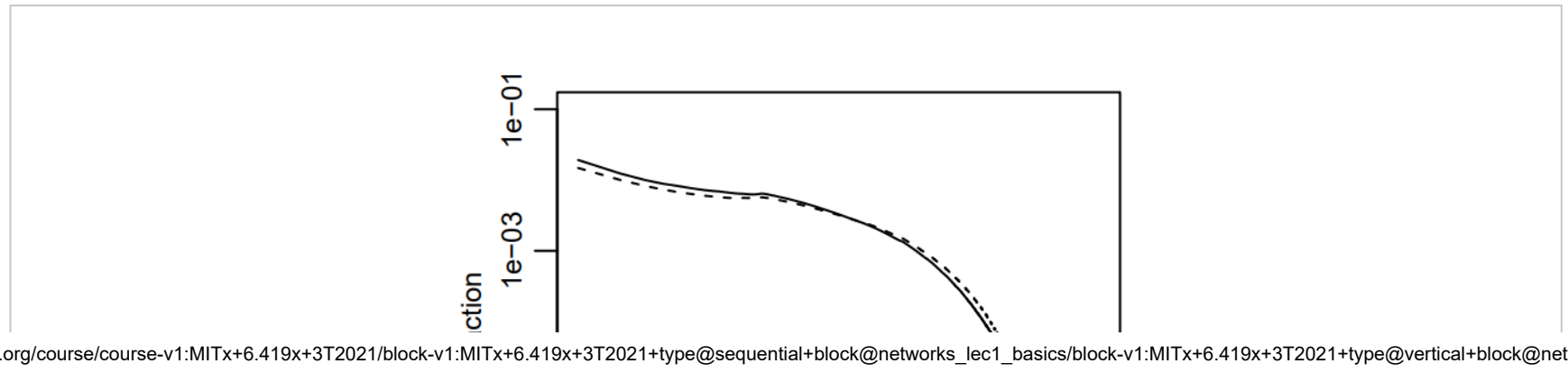
$$\log p_k = -\alpha \log k + c, \quad \alpha, c > 0.$$

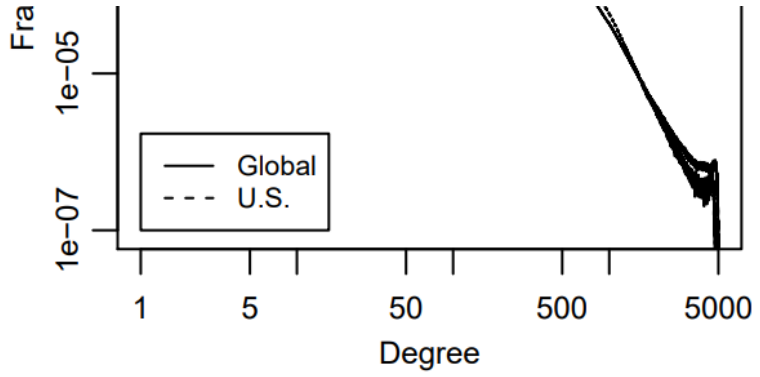
The distribution models scenarios that require the tail of the distribution to decay polynomially rather than exponentially. With a simple transformation, one can see that p_k decays as $k^{-\alpha}$.

The Facebook Graph

1/2 points (graded)

The following is a figure reproduced from the paper "The Anatomy of the Facebook Social Graph" by Ugander et al., 2011. The figure plots the degree vs. fraction of users. In other words, the figure presents the *degree distribution* of the Facebook graph (as it was in 2011). Notice the log-log nature of the plot. For the U.S. users, the degree measures the number of friends also from the United States.





6: Degree Distribution of the Facebook Graph in 2011.

1. **True** or **False**: The degree distribution seems to follow a power law distribution.

- (●) True
- ( ) False ✔

✖

2. **True** or **False**: More than 0.5% of the users have at most 50 friends.

- (●) True
- ( ) False

✔

Solution:

1. **False.** The log-log plot has a curvature and is not nearly linear.
2. **True.** This can be seen from the fact that 10^{-3} is a lower bound for fraction of users with degree k for every $1, 2, \ldots, 50$. Therefore, at least 50×10^{-3} of users have degree at most 50. As $50 \times 10^{-3} > 0.5\%$, the number of users that have at most 50 friends is greater than 0.5%.

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Edge Density

The **edge density** ρ of a graph with n nodes and m edges is defined as

$$\rho = \frac{m}{\binom{n}{2}}.$$

Here $\binom{n}{k}$ is the binomial coefficient given by

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}.$$

This is a metric that captures the fraction of (all possible) edges present in an undirected graph.

Circle Graph

5/5 points (graded)

Let n be a prime number. Consider the following *undirected simple graph* on n nodes $0, 1, \ldots, n-1$: Let k be some integer in the range $0 < k < n$ that parametrizes this graph. Then, for every $i \in [0, n-1]$ we have that node i is connected to node $(i+k)\,\%n$. Here $\%$ is the modulo operator.

1. **True** or **False**: The simple graph is connected.

- (●) True
- ( ) False

✔

2. Let $n \geq 3$. What is the degree of every node?

2 ✔ **Answer:** 2

3. Let $n \geq 3$. How many edges does the simple graph have?

n ✔ **Answer:** n

4. Let $n \geq 3$. What is the edge density ρ of the simple graph?

2/(n-1) ✔ **Answer:** 2/(n-1)

5. Suppose that n is very large (n much larger than 1, or you can consider the limit where $n \to \infty$), is the graph for this choice of n **sparse** or **dense**?

- ( ) Dense
- (●) Sparse

✔

Solution:

1. **True.** Starting with any node i we can form a walk $(i, (i+k)\,\%n, (i+2k)\,\%n, \ldots, (i+(n-1)\,k)\,\%n, i)$. Hence, the graph is connected when n is a prime number. Notice that $k > 0$.
2. **2.** Since n is prime and $n \geq 3$ every node i has two edges: $\{i, (i+k)\,\%n\}$ and $\{(i+(n-1)\,k)\,\%n, i\}$.
3. n**.** We can see this from the walk that we constructed to prove that the graph is connected. Alternatively, we can also see this from the degree sum formula that we showed in an earlier problem: every node has degree 2 and hence the number of edges is equal to $\frac{2n}{2} = n$.
4. $\rho = \frac{n}{\binom{n}{2}} = \frac{2}{n-1}$ as we have shown that the simple graph has n edges.
5. **Sparse.** From the edge density formula we have $\rho \approx \lim_{n\to\infty} \frac{2}{n-1} = 0$, so when n is large the graph tends to be sparse.

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

Sparse or Dense

2/2 points (graded)

Sparse or Dense (choose the most appropriate answer even if you can argue for both, which could happen in some extreme circumstances):

1. Let a graph be defined as follows: The webpages of the internet are the nodes and a directed edge from webpage A to webpage B captures the relationship that B's hyperlink is found in A.

- ○ Dense
- ◉ Sparse ✔

2. Facebook social network friendship graph (excluding “likes").

- ○ Dense
- ◉ Sparse ✔

Solution:

1. **Sparse.** Most webpages that we visit only have a handful of hyperlinks embedded in them.
2. **Sparse.** Every new user that signs up is going to have at most a constant number of friends in their friends list.

Submit You have used 1 of 1 attempt

ℹ Answers are displayed within the problem

Discussion

Topic: Module 3: Network Analysis:Graph Basics / 7. Graph Properties and Metrics - II

Hide Discussion

Add a Post

All Posts

Circle Graph: Explanation not clear to me

question posted 2 months ago by anonymous

Without spoiling, the solution is justified with "Since n is prime and n>3, the number of edges is... ". If the solution was justified by "Since n is non prime and n>1 billion " It would be just as effective in communicating the train of thought, sad to say... I do know what a prime number is and suspect the condition with the module guarantees the second node in the edge is not zero unless the node is prime but still fail to see the point or learn anything from this solution. Any material we can use to cover this at a lower level? At an example level perhaps? Thanks in advance! Edit: Sorry for the rant but if possible, can STAFF please further explain the details behind the answer? Thanks!

This post is visible to everyone.

Add a Response

1 response

Iam_trinh (Community TA)
2 months ago

Hello,

I used inductive reasoning to solve question #2 in the Circle Graph. Once I listed a few prime numbers and begin to work out which nodes map to which nodes, I started seeing patterns and arrived at the answer. I can't think of any way to explain the answer without giving away the answer itself, so I will refrain from explaining... hopefully other TA's or staff can chime in on this.

I guess the idea behind this exercise was to get used to graphs theory, the different definitions, and the different types of graphs, I suppose. Others and I have shared some helpful materials in this post, feel free to check them out: https://courses.edx.org/courses/course-v1:MITx+6.419x+3T2021/discussion/forum/6419x_3T2021_discussion_networks_lec1_basics-tab6_discussion/threads/61581474b9dd920473a133e8

Add a comment

Showing all responses

Add a response:

Preview

Submit

‹ Previous Next ›