LINFO1115: Project – Power Grid Analysis

Peter VAN ROY Lucile DIERCKX

Academic Year: 2023 – 2024

Context

The safe and efficient working of a power grid relies heavily on its resilience. Understanding its structure and identifying the importance of each node is crucial for preventing blackouts and maintaining a reliable electricity supply. This project focuses on analyzing an electrical power grid, its structure, and identifying the most important elements in the network.

1 Format

Every entry of the dataset represents the connection between two nodes in the electrical grid. The dataset has two columns: `Src` and `Dst`. A line of the dataset `(u, v)` means therefore that node `u` is connected to node `v`.
An electrical connection allows the sending of electricity in both directions of the link. Therefore, the graph is considered undirected in all questions below except for task 3.

2 In practice

This project is made of 5 different questions, summing up to a total of 20 points, if you do not wish to answer all the questions, it is allowed. You do not need to have the answer to a previous question to be able to answer the next one.

For this project, you have to use Python language. (Note that if the code does not compile when we run it, you will not succeed in the project). We request you to implement all algorithms for the project by yourself, without using any library that implements it already. You may of course use the available implementation, but for validation purposes only.
This assignment must be completed by **group of two students** (preferably) **or alone**. If you do not manage to find a partner, please start by requesting a partnership via the Moodle forum. Then, **if and only if** you cannot find anyone else to work with, get in touch with Lucile Dierckx.
Unless specified otherwise, for each of the items below, we request you to comment on your quantitative results. For each algorithm, indicate also the temporal complexity in your report. **This complexity should be expressed using the number of vertices v and/or the number of edges e**.

You will have to submit your code on Inginious (`https://inginious.info.ucl.ac.be/course/LINFO1115`) (a code template is available on Moodle with some details about what should be returned for each question) and you will have to write a report to analyze the obtained results.

2.1 Task 1: (6 points/20)(Easy)

For this exercise, consider the graph as an undirected graph.

In this first exercise, we are interested in knowing more about the general information of the graph. Therefore, we ask you to count:

1. The average degree of the nodes (average number of connections per node) and a histogram of the degree distribution of the graph (number vs degree).
2. The number of bridges in the graph.
3. The number of local bridges in the graph.

How can you interpret those numbers in the context electrical power grid?

2.2 Task 2: (4 points/20)(Intermediate)

For this exercise, you have to consider the graph as an undirected graph.

As redundancy is important to ensure that the apparition of a failure is not critical for the system, we wish to analyze whether pairs of connected nodes have a lot of common neighbors. This information can be represented with the similarity measure between two connected nodes A and B:

$$s(A,B) = \frac{\#\text{ neighbors common to A and B}}{\text{total amount of distinct nodes that are neighbors of A and/or B}}$$

Compute the similarity score of each pair of nodes connected with an edge and report the average similarity score for the network. Additionally, present the cumulative graph of the percentage of ~~nodes~~ **edges** vs the similarity score. Interpret the obtained results.

2.3 Task 3: (4 points/20)(Intermediate)

For this exercise, you have to consider the graph as a directed graph.

Let us now imagine that the edges between two nodes are not bidirectional anymore. Instead, the electrical grid is a directed graph with edges going from the `Src` nodes to the `Dst` ones. We want to measure which nodes are well-served and easy to reach. Therefore, we want to compute the PageRank score of each node in the network.

The common PageRank (PR) score of a node p is usually computed recursively as

$$PR(p) = \frac{(1-d)}{N} + d \sum_{n \in B(p)} \frac{PR(n)}{Nout_n}$$

where N is the total number of nodes in the graph, $B(p)$ is the set of nodes pointing to p, $Nout_n$ is the number of outgoing links of node n and d is the damping factor, having a value of 0.85. After convergence, the PageRank scores need to be normalized so that the total of all PageRank scores sums to one. Note that the system can be considered as "converged" when the total sum of value updates is $\leq 10^{-6}$.

We ask you to compute the PageRank score of each node. Give the id number of the node having the highest weighted PageRank score and indicate its value.

2.4 Task 4: (3 points /20)(Advanced)

For this exercise, you have to consider the graph as an undirected graph.

The small-world phenomenon describes the observation that most pairs of people in the world are connected by a short chain of acquaintances (usually with an average length of 5 or 6 for social relations).
We would like to check whether the small-world phenomenon is also observable in this network. Therefore you have to measure the distance of the shortest path between each pair of nodes. Then generate a graph of the number of paths having a given distance value (number vs length), you can find an example of such a graph in the slides of the first course. What is the diameter of the electrical network (the longest shortest path between two nodes)? Interpret the obtained measurements.

2.5 Task 5: (3 points /20)(Advanced)

For this exercise, you must consider the graph as undirected.

A last commonly used measure to obtain an indication about the importance of each node is called the betweenness centrality. This is calculated for each node by computing the shortest path between each pair of nodes and counting the number of times each node of the network is part of such shortest path. We ask you to give the id number of the node with the highest betweenness centrality and the computed value.

3 Deliverables

Your rapport should be **maximum 4 pages** long, should have the following structure, and should contain the following information:

Name and NOMA of the two students

1. Task 1
    1.1. Results: Indicate the average degree of the nodes, the number of bridges, and the number of local bridges.
    1.2. Graph: Show the histogram of the degree distribution.
    1.3. Interpretation: Interpret the obtained values in the context of electrical grids.
    1.4. Complexity: Indicate the time complexity of your algorithms.

2. Task 2

   2.1. Result: Give the average similarity score for the network.

   2.2. Graph: Show the graph of the cumulative graph of similarity score for the ~~nodes~~ **edges**. Interpret.

   2.3. Complexity: Indicate the time complexity of your algorithm.

3. Task 3

   3.1. Result: Indicate the highest PageRank score you found and the id number of the related node.

   3.2. Complexity: Indicate the time complexity of your algorithm.

4. Task 4

   4.1. Result: Indicate the diameter of the graph.

   4.2. Graph: Show the graph of the number of paths having a given distance. Interpret.

   4.3. Complexity: Indicate the time complexity of your algorithm.

5. Task 5

   5.1. Result: Give the id number of the node with the highest betweenness centrality and the obtained score.

   5.2. Interpretation: Interpret the obtained values in the context of electrical grids.

   5.3. Complexity: Indicates the time complexity of your algorithm.

6. Appendix: Give your code in appendix

4 Deadline

The assignment is due by **Friday, April 19th, 2024 at 23:59**. The code should be submitted on Inginious (`https://inginious.info.ucl.ac.be/course/LINFO1115`) by the same deadline. The report must be handed in on Moodle as a **zip file containing both your report and your complete source code**. **Only one student per group of two should submit on Inginious and Moodle (preferably the same student for both)**.