UNIVERSITÀ
DEGLI STUDI
FIRENZE

Algorithms and Programming for Massive Data

Topics: Eulerian Path and Clustering Coefficient for Graph.

University of Florence

Armand Palla

UNIVERSITÀ DEGLI STUDI FIRENZE

1.Introduction

1. Building the graph of provinces P using NetworkX.

2. Construct the random Graph.

3.Eulerian Path

4. Clustering Coefficients

5. Pandas and Networkx.

6. Conclusions.

7.References.

UNIVERSITÀ
DEGLI STUDI
FIRENZE

1. Building the graph of provinces P using NetworkX.

Data:

a. Datas are of type JSON and referred to provinces and regions .

b. The Regions dataset is made up of 16 variables and many variables are of type int64 and float 64.

c. The Province dataset is made up of 12 variables

d. The columns in common in the two datasets are:

1.data 2.stato 3.codice_regione 4.note 5.totale_casi 6. denominazione_regione

UNIVERSITÀ
DEGLI STUDI
FIRENZE

1. Building the graph of provinces P using NetworkX.

The **main functions** for building the graph are:

1.def plot_graph that it is used for plotting with networkx.

2.def provinces_graph that it is used for building the provinces graph.

3. def set_edges that it is used to set the edges in the graph.

UNIVERSITÀ
DEGLI STUDI
FIRENZE

2.Construct the random Graph.

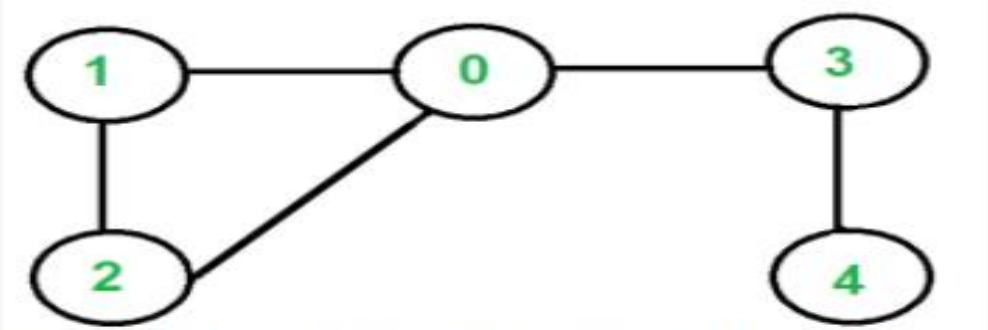
a. def random_graph(nodes_num, x_low, x_high, y_low, y_high):

Generate 2000 pairs of double (x,y) with x in [30,50) and y in [10,20) .

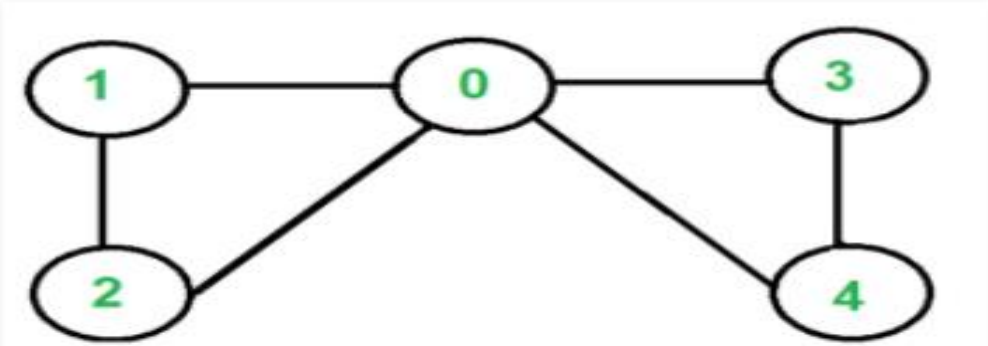
Build a graph R where each pair is a node and two nodes are connected with the rule above.

3.Eulerian Path

Eulerian path is a path in graph that visits every edge exactly once. Eulerian Circuit is an Eulerian Path which starts and ends on the same vertex.



The graph has Eulerian Paths, for example "4 3 0 1 2 0", but no Eulerian Cycle. Note that there are two vertices with odd degree (4 and 0)



The graph has Eulerian Cycles, for example "2 1 0 3 4 0 2"
Note that all vertices have even degree

UNIVERSITÀ DEGLI STUDI FIRENZE

3.Eulerian Path

How to find whether a given graph is Eulerian or not?
The problem is same as following question. “Is it possible to draw a given graph without lifting pencil from the paper and without tracing any of the edges more than once”.

Eulerian Cycle
An undirected graph has Eulerian cycle if following two conditions are true.
a) All vertices with non-zero degree are connected. We don’t care about vertices with zero degree because they don’t belong to Eulerian Cycle or Path (we only consider all edges).
b) All vertices have even degree.

Eulerian Path
An undirected graph has Eulerian Path if following two conditions are true.
a) Same as condition (a) for Eulerian Cycle.
b) If zero or two vertices have odd degree and all other vertices have even degree.

UNIVERSITÀ
DEGLI STUDI
FIRENZE

3.Eulerian Path

How does this work?

In Eulerian path, each time we visit a vertex v, we walk through two unvisited edges with one end point as v. Therefore, all middle vertices in Eulerian Path must have even degree. For Eulerian Cycle, any vertex can be middle vertex, therefore all vertices must have even degree.

The base case of our algorithm is: if the number of vertices with an odd number of edges(odd degree) is greater than 2 then there is no Eulerian path.

UNIVERSITÀ
DEGLI STUDI
FIRENZE

4. Clustering Coefficients

In graph theory, a **clustering coefficient** is a measure of the degree to which nodes in a graph tend to cluster together.

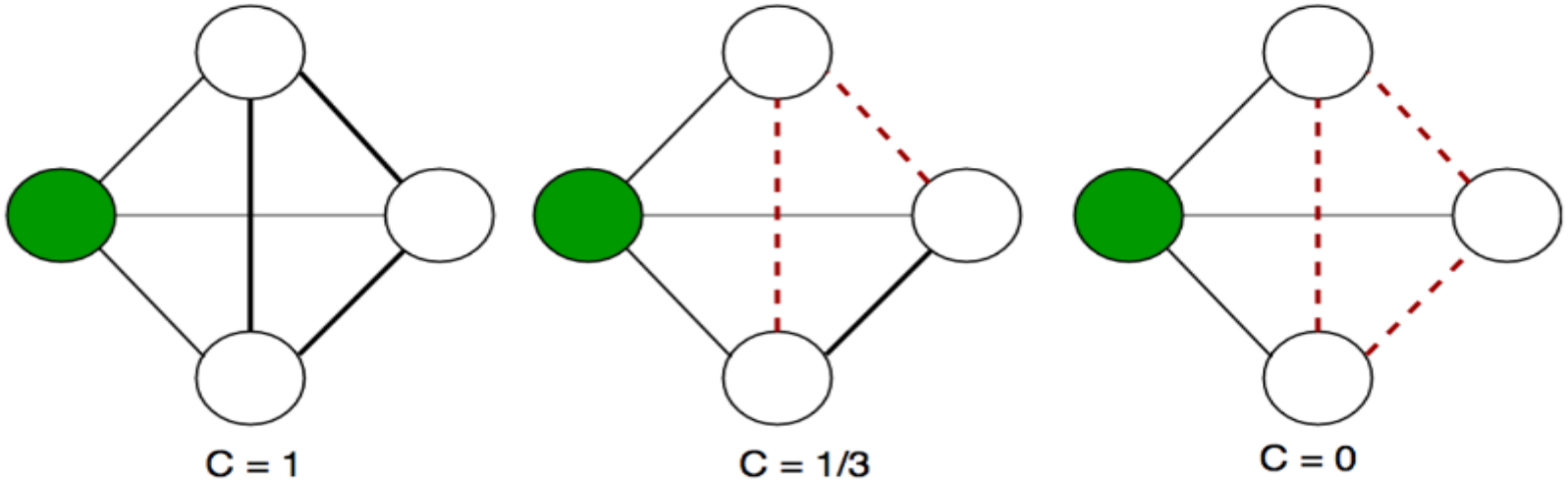
The fields where we can apply this algorithm are real-world networks, and in particular social networks.

The **local clustering coefficient** of a node in a graph quantifies how close its neighbors are to being a clique(complete graph).

Clustering coefficients can be also defined as the measure of the relaive frequency of triangles.

4. Clustering Coefficients

The coefficient of the green node is computed as the proportion of the connections among its neighbors compared with the number of all possible connections.



UNIVERSITÀ DEGLI STUDI FIRENZE

4. Clustering Coefficients

$$C_i = \frac{2|\{e_{jk} : v_j, v_k \in N_i, e_{jk} \in E\}|}{k_i(k_i - 1)}$$

N_i is the set of neighbors of node **v_i.**

e_j_k represent the edge connecting between node **v_j** and node **v_k.**

k_i is the number of neighbors of **v_i.**

0(Star) << Clustering Coeffcients << 1(Clique)

UNIVERSITÀ DEGLI STUDI FIRENZE

4. Clustering Coefficients

	Graph P	Graph R
Network X	0.0156400203704834	0.0312854232788086
List Intersection	0.013122854232788 08	0.0043070316314697266

UNIVERSITÀ
DEGLI STUDI
FIRENZE

4. Clustering Coefficients

Interpretation:

As we can see the time of calculation of the clustering coefficient with **networkx** and with **list intersection** is approximately the same.

We notice a small difference in favour of list intersection ,but to analyze this we need to know how networkx(library)calculates this thing.

UNIVERSITÀ DEGLI STUDI FIRENZE

5. Pandas and Networkx

Pandas:

1. The fundamental high-level building block for doing practical, real world data analysis in Python.
2. A fast and efficient **DataFrame** object for data manipulation with integrated indexing.
3. Flexible **reshaping** and pivoting of data sets.
4. High performance **merging and joining** of data sets.
5. Highly **optimized** for performance.

NetworkX:

1. A Python package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks.
2. Load and store networks in standard and nonstandard data formats.
3. generate many types of random and classic networks.
4. Analyze network structure, build network models, design new network algorithms,

UNIVERSITÀ
DEGLI STUDI
FIRENZE

6.Conclusions

The main intention of this project was to analyze the data about Covid-19 in all the Provinces and Regions in Italy .

Discussing and understanding the Eulerian Path algorithm and Clustering Coefficient methods for an undirected graph.

Using Pandas for data manipulation,visualization and analyzing.
Interpreting different results of the code.

UNIVERSITÀ DEGLI STUDI FIRENZE

7.References.

1. https://en.wikipedia.org/wiki/Clustering_coefficient

2. https://pandas.pydata.org

3. https://networkx.org

4. https://github.com/pcm-dpc/COVID-19

5. https://en.wikipedia.org/wiki/Compartmental_models_in_epidemiology

6. https://medium.com/@1522933668924/using-matplotlib-in-jupyter-notebooks-comparing-methods-and-some-tips-python-c38e85b40ba1

7. https://www.geeksforgeeks.org

UNIVERSITÀ
DEGLI STUDI
FIRENZE

Thank you for your attention.