Network Analysis (Project Proposal)

Sara Vellone, Digital Humanities and Digital Knowledge, 1056935

Context

General Context:

- Scientific discoveries
- Statistics and Computer Science for predictions

Specific Applications:

- Science in Social Networks
- News spreading through Social Networks (Twitter)
- Popularity prediction

Problem and Motivation

The starting point is a paper, *Predicting the impact of online news articles,* which is based on a dataset in which are contained 40+ million tweets related to COVID-2019. The main idea is to reuse the measures applied for explaining if it is possible to predict the popularity of a scientific article on a dataset that is related to the spreading of the Higgs boson-like particles discover on Twitter, trying to find some similarities.

Datasets

Data were taken from two public sources: Stanford Large Network Dataset Collection for the paper *The Anatomy of a Scientific Rumor* and Awesome Public Dataset for *Predicting the impact online of new articles*.
Following, the list of the tools for managing the data:

- NetworkX – the python library for handling dataset and computing measures
- Matplotlib – the python library for visualizing graphs
- File Viewer Plus – a tool to extract files with .gz format

Measures

Starting from some measures proposed on the paper *Predicting the impact online of new articles,* on the "Higgs Twitter Dataset" will be applied:

- Centrality measures
    - Degree centrality, Closeness centrality, Betweenness centrality, HITS, PageRanks, Eigenvector centrality, Communicability centrality, Load centrality
- Assortativity measures
    - Average neighbour degree, Average degree connectivity