

NGSA: NETWORK SCIENCE ANALYTICS
CENTRALESUPELEC and ESSEC BUSINESS SCHOOL

Assignment 2

Instructor: Fragkiskos Malliaros
TA: Abdulkadir Çelikkanat

Due: **April 1, 2018 at 23:00**

Description

The goal of this assignment is to predict missing links in a citation network of research articles. A citation network is represented as a graph $G = (V, E)$, where the nodes correspond to scientific articles and the existence of a directed edge between nodes u and v , indicates that paper u cites paper v . Each node (i.e., article) is also associated with information such as the title of the paper, publication year, author names and a short abstract. A number of edges have been randomly removed from the original citation network.

Your goal is to accurately reconstruct the initial network using graph-theoretical and textual features, and possibly other information. Your solution can be based on supervised or unsupervised techniques for link prediction or on a combination of both. You should aim for the maximum $F1$ score.

Evaluation and Rules of the Competition

The assignment is a Kaggle competition. Detailed information about the competition, the rules, the evaluation and the data can be found in the Kaggle website:

<https://www.kaggle.com/t/e0958175c55d4dd9b43161033567c74b>

Please read in detail the rules of the competition and the evaluation (grading) process.

How to Submit

Please complete the second assignment in groups of **3-4** students (preferably, the same team as in the project of the course). No late assignments will be accepted.

1. **Kaggle submission:** submission of your solution in the kaggle platform (team submission – pick also a name for your team).
2. **Report:** *typeset* your report (**PDF** file only). The submission of the report (max **3 pages**) should be made on **gradescope** (Assignment 2; Entry Code: 974JYB).
3. **Code:** prepare a .zip file (`code_name_of_your_team.zip`) containing the code that is needed to reproduce your submission. The file should be sent by email to: *fragkiskos.me@gmail.com*, subject: “NGSA - Assignment 2 Code - Name of Your Team”.