Universidad
Carlos III de Madrid

Homework 2. Link Prediction

Master in Computational Social Science

Esteban Moro
emoro@math.uc3m.es
2022-23

Homework 2

Link prediction Project

1. **Objective**: study the problema of link prediction in a social network
2. **Intro**
   - Choose a network
     - Use the network you analyzed in homework 2
     - If you want, you can change the network to another one. But keep in mind that the network should have enough number of links (typically more tan 5000)
     - Register your choice in Aula Global: https://aulaglobal.uc3m.es/mod/forum/view.php?id=4420611
   - The network can be in any context, but preferably a social network.
   - The number of nodes should be larger tan 500 and it should have around 5000 links.
   - Import the network data into R and create a graph using .

Homework 2

Link prediction Project

3. **Steps**:

- Delete a fraction of real edges in the network and create a table of those links deleted (positive class) and of links non-present (negative class)
- Generate a number of proximity/similarty metrics heuristics for each link in the positive and negative class
- Train a binary classifier to predict the links, i.e., to predict the class (positive/negative) using those heuristics. Use cross-validation.
- Evaluate the precision of the model. Which heuristic is the most important. Why do you think it is the most important?
- Comment on potential ways to improve the link prediction

Homework 2

4. **Evaluation.** This are the main points in the evaluation:
    - The network chosen is not used by any other student (try not to overlap)
    - All questions and steps in 3. are answered correctly and the notebook is clearly presented
5. **Format and deadline**
    - Homework can be done in groups up to two people
    - You can use R or Python
    - You will have to produce a markdown notebook using Rmarkdown or Jupyter
    - You will have to present a report:
        - Include a section in which you address each of the steps in point 3. and answer the questions.
        - You have to show the code used to solve the question and the answer.
        - Hand out the pdf or html version of the notebook
6. **Deadline:** Monday, May 8th 2023

Questions and Answers

- How can I know what networks other students have chosen?
  - As soon as you have decided which one to choose I will publish it in Aula Global
- Can I use a network with 1 million nodes
  - Up to you, but it will take you more time to do the report ☺
- How much weight is this homework in the final mark?
  - 30% as stated at the beginning of the course
- Can we present the homework as a group?
  - Yes, in groups of up to two persons.
- What if the link prediction algorithm is not working?
  - Many thinks might have happened. Most likely your network is “special” and does not have the properties of social networks we saw in class.