



MSc Program - Data and Web Science - ARISTOTLE UNIVERSITY

SOCIAL NETWORK ANALYSIS PROJECT ANNOUNCEMENT – 2019/20

The goal of this project is to get you ready to apply state-of-the-art network analysis algorithms and use common network analysis tools to an application.

Solving everyday problems using network analysis is a challenge and after this project you will be in a position to identify if a problem can be solved using graph mining and proceed accordingly.

The students have to work in teams of 3 people. Only if your project idea is really needy and innovative, we can discuss the possibility of working in a team of 4 people. Each project will be divided into two parts and there will be three types of course projects:

- **Evaluation of different algorithms and models on an interesting dataset.** Πρέπει να βρούμε δικό μας dataset
- **Presentation of a new, theoretical algorithm/model/network metric that produces comparable and better results than already existing ones.**
- **A scalable implementation of already existing (or new) algorithms for processing very large graphs.**

It would be ideal, if your project contains a mixture of all these types. The project should contain at least some amount of mathematical analysis, and some experimentation on real or synthetic data. A very good final result can even lead to the publication of the work in a scientific conference or journal.

The topics that you will focus on are the following:

Topics:

1. Community detection in Social networks (graph partitioning & clustering)

2. Information diffusion in online social networks : (a) Independent Cascade model - b) Linear threshold model - c) new models // choose one of (a), (b), (c)

3. Influence maximization on Social Graphs

Phase I - Deadline: 2/12/19

The first part of the project is a literature review of the latest and more important papers regarding the topic that you have already picked. You will have to choose 7-10 papers that are clearly related to the topic of your choice and carefully read them. You will have **to write a short (5-7 pages) report** (in the form of a paper) where you should address the following:

- **Summary**
 - What is the main technical content of the papers?
 - How do papers relate to the topic?
 - What is the connection between the papers you are discussing?
- **Critique**
 - What are strengths and weaknesses of the papers?
 - What did the authors miss?
 - Present anything that you consider unrealistic!
- **Brainstorming**
 - What are promising future research questions in the direction of the papers?
 - How can these be pursued?
 - A better idea/model/algorithm? A better approach?

You may proceed with the implementation of the second part, Based on the outcome of this first part. You will also present your work to all your colleagues, so make sure you **prepare a proper -short- presentation** where you will present the key points of your report.



MSc Program - Data and Web Science - ARISTOTLE UNIVERSITY

Phase II - Deadline: 20/1/19

The second part is the more practical one. You are asked to collect your own data, or use a publicly available network dataset. It is not obligatory to use a social network, for example you can use a network of proteins. We advise you to choose a set of data that you find appealing or is related to your work or everyday life. You will have to apply the methods – models and algorithms that you have been introduced to during the course and we strongly support to use even stuff that you haven't been taught yet. Another, even better approach would be to build on the results of the first part report and focus on how you can use the findings or algorithms that you have evaluated on your own dataset.

The final project report should be a 6-8 page paper, following the structure of common scientific papers [introduction, related work, approach, results and conclusion]. At the end of the paper, you should also highlight the contributions of individual team members to the project. The project report should contain at least some amount of mathematical analysis, and some experimentation on real or synthetic data.

Regarding the tools you can use in order to complete your report, we recommend python (NetworkX) and Gephi. Gephi is a network visualization tool that you will get the chance to know better in one of the upcoming lab activities. The grading of the final report will be as follows:

- **Introduction/Motivation/Problem Definition**
 - What is it that you are trying to solve/achieve and why does it matter.
- **Related Work**
 - How does your project relate to previous work. Please give a short summary on each paper you cite and include how it is relevant.
- **Model/Algorithm/Method**
 - This is where you give a detailed description of your primary contribution. It is especially important that this part be clear and well written so that we can fully understand what you did.
- **Results and findings**
 - How do you evaluate your solution to whatever empirical, algorithmic or theoretical question you have addressed and what do these evaluation methods tell you about your solution. It is not so important how well your method performs but rather how interesting and clever your experiments and analysis are.
 - We are interested in seeing a clear and conclusive set of experiments which successfully evaluate the problem you set out to solve. Make sure to interpret the results and talk about what can we conclude and learn from your evaluations. Even if you have a theoretical project you should have something here to demonstrate the validity or value of your project (for example, proofs or runtime analysis).
- **Style and writing**
 - Overall writing, grammar, organization and neatness.

Be careful! Plagiarism and Copying other people's work is not accepted!

Make sure you cite everything you use and has not been developed by you or your team!

Try to work together!

Work done by just one or two members of the team will not guarantee a proper grade for the others as well.



MSc Program - Data and Web Science - ARISTOTLE UNIVERSITY

Some Indicative Examples:

- <https://arxiv.org/abs/1111.3919>
- http://www.cs.columbia.edu/~apoorv/Homepage/Publications_files/naacl2012.pdf
- <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0201879>

Interesting articles:

- <http://evelinag.com/blog/2016/01-25-social-network-force-awakens/index.html>
- <https://towardsdatascience.com/network-analysis-to-quickly-get-insight-into-an-academic-field-with-python-cd891717d547>
- <https://www.fastcompany.com/3068474/the-real-difference-between-google-and-apple>
- <https://towardsdatascience.com/who-is-the-most-important-person-in-the-film-industry-61d4fd6980be>

More Demanding examples:

- <http://snap.stanford.edu/class/cs224w-2018/projects.html>
- <http://snap.stanford.edu/class/cs224w-2017/projects.html>
- <http://snap.stanford.edu/class/cs224w-2016/projects.html>

List of resources:

- <https://github.com/briatte/awesome-network-analysis>