NETWORK SCIENCE FOR DATA ANALYTICS

2019781-1/2024-I

MINI PROJECT – FINAL REPORT

1. GOAL

To present the achievements of your mini project in which you apply some of the concepts and methods studied in class by trying to reproduce an application of network science reported in a research paper.

2. DESCRIPTION

The goal of this assignment is to present the results of your team´s mini project. In this report, you will need to describe in detail the network science methods implemented and used to carry out your network analysis to solve the stated problem in your mini project proposal, as well as describing the final results.

3. DELIVERABLES

This homework includes three main components:

A. ***WRITTEN REPORT.*** A written report, which must be presented in the template of the IEEE Transactions on Network Science and Engineering Journal available at

https://template-selector.ieee.org/secure/templateSelector/format?publicationTypeId=1&titleId=152&articleId=1.

This document will need to include at least the following sections:

PAPER CITATION:

You need to include the citation to the selected paper that reports the work that you will attempt to replicate.

PROBLEM DESCRIPTION:

Briefly summarize the problem that was solved.

PROJECT GOAL AND SCOPE: The goal of the project consists of a few sentences with an executive statement of the objectives of the project and what you expect to accomplish.

CASE STUDY: Summarize the case study that your mini project attempted to reproduce.

NETWORK DATA SET: You will need to provide a general description of the data set that was analyzed.

IMPLEMENTED NETWORK SCIENCE APPROACH: Describe in general terms the implemented network science approach. The types of network models and methods that you used need to be mentioned.

Important note: as a guide to perform your network analysis, a briefing on graph analytics is presented below. It is very important to notice that to perform your analysis, you will not need to implement the graph algorithms, but rather you are advised to select the proper software tools and just apply them to analyze your particular network.

Taken from: https://livebook.manning.com/book/graphs-and-network-science-an-introduction/chapter-1/

Graph analytics: The graph analytics focuses on connections and dependencies between entities in a network.

1. Graph-based pattern queries

Graph-based pattern queries are useful when you want to find explicit local patterns within the network. As you need to describe the pattern with the graph query language, you need to know beforehand what you are looking for. The graph patterns could be as simple as to find all the neighbors of a specific node to more complicated such as analyzing the customer path through the market funnel. You can also use graph queries to calculate graph statistics such as the number of nodes and relationships, the node degree distribution, and more.

2. Graph algorithms

The term graph algorithms refer to more global and iterative analysis, where we consider the whole network by the graph algorithm. For example, you want to learn the overall community structure of a network or find the most central nodes. The graph algorithms can be grouped into several categories.

2.1. Pathfinding algorithms

Pathfinding algorithms are usually used to find the shortest path between a pair of nodes in a network. Other pathfinding algorithms try to connect a given network with the least amount of relationships or the sum of their weights. They are used in transportation, communication, and logistics network to find the optimal routes.

2.2 Centrality algorithms

Centrality algorithms can find the most central nodes and identify their role in a network based on the graph topology. These algorithms are used in scenarios when you want to identify the influencers in a social network, find bottlenecks in a transportation network, or bridges between groups of nodes.

2.3. Community detection algorithms

The function of community detection algorithms is to find clusters or communities of nodes within a network. You can also use them to evaluate how tightly-knit the network as a whole is. They are used to find isolated groups, reveal tight clusters of nodes, and discover the overall community structure. This information helps predict similar behavior or preferences and estimate the resilience and robustness of a group.

2.4. Similarity algorithms

The role of similarity algorithms is to detect similar nodes in a network based on the graph topology or their properties. The output of those algorithms is usually a new similarity network between the nodes.

2.5. Node embedding

Node embedding algorithms aim to encode nodes in the embedding space so that the similarity in the embedding space approximates the similarity in the network. Graph neural networks also fall into this category.

The output of these algorithms is a vector for each node that represents its position in a network. These vectors can be used in a downstream machine learning workflow, or they can be used to infer a similarity network using the kNN algorithm. Some of the node embedding algorithms are: Node2Vec, GraphSAGE and TransEk.

Additional note: In this mini project it is not necessary that you perform a statistical evaluation of the implemented network science methods and models, since this is beyond the scope of the class

WEB LINKS TO SOURCE CODE AND EXPLANATORY VIDEO: You will need to provide the access to the source code (Networkx and/or Gephi source code) with the implementations of your analysis. To handle the data, software and documentation of your project, your team is advised to use a Github repository, though, alternative repositories can be used. The information needs to be organized in an intuitive/structured manner. The information to access such repository must be included here.

Also, in this section, you will need to include the information to access the explanatory video of your report (on YouTube or where applicable). Please see some instructions to prepare such video below.

TEAM MEMBERS´ CONTRIBUTIONS. The roles, activities and contributions of each of the team members during the development of this delivery must be concisely described. Please keep in mind that in the different deliveries, you could rotate the roles so that everyone takes a turn and thus gains experience of the different roles. Also, you may want to allocate the crucial roles within the team to the people who would like to take them, are best qualified to carry them out through prior or current experience, or even to people who would like to take on an unfamiliar role in order to gain experience of performing that role. However, each team will need to select a **team coordinator**, who besides supervising and motivating the members of the working group will be the direct communication channel between the professors and the team.

The teams are expected to internally solve any issue or conflict that arises among the members of the team. However, if an issue or conflict could not be solved inside the team, the team coordinator is responsible for reporting it to the professors

The contributions of the team members may be provided in a table as shown below.

Team member	Role	Activities/contributions

USE OF ARTIFICIAL INTELLIGENCE TOOLS (PREPARING THIS DELIVERABLE)

As a reminder, you are allowed to use Artificial Intelligence (AI) tools, such as ChatGPT, to help gather information and generate ideas or source code, but you may not copy and paste information directly from the AI tool and present it as your own without citing it. You are responsible for the

information you submit based on an AI query and for ensuring that it does not contain misinformation, unethical content, or violate intellectual property laws.

You must properly document and cite the use of artificial intelligence tools to ensure academic integrity. When applicable, you are expected to include a statement in your assignment describing which AI tool you used and how you used it. For example, "*ChatGPT was used to draft approximately 50 percent of this document and to provide review assistance. AI-produced content was edited for accuracy and style*".

CONCLUSIONS. Devise the conclusions of your work, address the limitations of your analysis and how your analysis could be enhanced or built on it.

REFERENCES. Each source you cite in the paper must appear in your reference list; likewise, each entry in the reference list must be cited in your document.

***B. SOURCE CODE IMPLEMENTATION*.** You will need to provide the source code (Colab notebooks or others) with the implementations of your analysis. Again, to handle the data, software and documentation of your mini project, your team is advised to use a Github repository, though, alternative repositories can be used. The information needs to be organized in an intuitive/structured manner. The information to access such repository must be included in the written report.

C. EXPLANATORY VIDEO OF YOUR REPORT. You will need to present the mini project final report in a video that will not exceed 5 minutes. Thus, we suggest that you prepare a presentation of no more than 10 slides, which will emphasize the application of network science methods and analysis and your final results. You can use this presentation as a basis to prepare your video. The URL to access this video must be informed in the corresponding section of your report, as explained above.

To help you prepare your video, you are advised to see the tutorial: **Effective Presentation Skills Tutorial** available at https://www.niu.edu/presentations/index.shtml
Please be aware that though some of the tips in this tutorial are intended to in-person presentations, most of them still apply for on-line presentations.

Also, your team will need to participate in the oral discussion of your mini project final results in class.

4. **SUBMISSION**

The file of the mini project final report, in pdf format, need to be submitted via Google Classroom. Only **one submission per team** is required. Make sure ONLY ONE COPY of the deliverable is submitted, thus, you will need to coordinate within your team who will be in charge of submitting the assignment. Students are advised to retain a copy of every assignment submitted for their own record.

- **File to submit:**

Written report: a document with the details of the mini project final report.

Please make sure that in this report you include the web links to:

1. the repository where your source code is available, and
2. the explanatory video of your report, a video presenting your mini project final report.

- **File naming:**

The team must use the following file naming convention for the assignment:

Mini project final report: “NS4DA-MPFR-doc-” appending the assigned team ID. For example, “NS4DA-MPFR-doc-7.pdf” for team ID 7.

5. **DEADLINE (DATE/TIME)**

Written report:

4:00pm, Wednesday, April 3rd, 2024

Oral discussion in class:

Wednesday, April 3rd, 2024.