Social Network Analysis

Instructor: ***Katia Papakonstantinopoulou*, Adjunct Lecturer, AUEB, katia@aueb.gr**
Teaching Assistant: *Panagiotis Liakos*, p.liakos@di.uoa.gr

Overview

This course covers data analysis on social networks, focusing on ways to handle large-scale networks efficiently.
It provides the main theoretical results in social network mining as well as hands-on practice on key issues in the area.

Key Outcomes

By completing the course the students will be able to:

- understand the basic concepts of social networks
- understand the fundamental concepts in analyzing the large-scale data that are derived from social networks
- implement mining algorithms for social networks
- perform mining on large social networks and illustrate the results.

Requirements and Prerequisites

The students should have a basic algorithmic and programming background. Specifically, basic knowledge in the fields of graph theory and artificial intelligence/ database applications will be necessary for this course.

Required Course Materials

There is no required textbook. All course materials will be provided in class and available for downloading.

Students will need to bring their laptops in class in order to try out interactively the material being presented.

Related Books

- David Easley and Jon Kleinberg, Networks, crowds, and markets, Cambridge University Press, 2010.
- Jure Leskovec, Anand Rajaraman and Jeffrey David Ullman, Mining of massive datasets, Cambridge University Press, 2014.

Software/Computing Requirements

The computational aspects of this course will be implemented in Apache Giraph and SNAP, free software environments for graph processing. Students can either set up their programming environment on their own computers or use a VirtualBox image with **pre-installed software provided by the instructor**. We will use the Python and Java programming languages; however, the course **does not assume any prior experience in Python or Java**. We will make use of the SNAP data sets available at the SNAP repository. Other data sets that will be used in class and/or assignments will be available in the e-class page of this course.

Grading

There will be a total of 3 homework assignments (given approximately at the units 2, 3 and 4) that will contribute 60% in the final grade. The remaining 40% will be determined by the in-class final exam. Please note that one needs to get at least 5 (out of 10) in the final exam (independently of the grades in the homework assignments) not to fail the course.

Assignments

There will be a total of 3 homework assignments during the course. Indicative subjects:

- graph data visualization using Gephi
- creation, manipulation, and analysis of large networks using SNAP
- implementation of a Pregel algorithm using Apache Giraph and execution of it for a set of graphs.

Late assignments will either not be accepted or will incur a grade penalty unless due to documented serious illness or family emergency. Exceptions to this policy for reasons of civic obligations will only be made available when the assignment cannot reasonably be completed prior to the due date, you make suitable arrangements, and give notice for late submission in advance.

Attendance Requirements

Class attendance is essential to succeed in this course and is part of your grade. An excused absence can only be granted in cases of serious illness or grave family emergencies and must be documented. Job interviews and incompatible travel plans are considered unexcused absences. Where possible, please notify the instructor in advance of an excused absence.

Students are responsible for keeping up with the course material, including lectures, from the first day of this class, forward. It is the student's obligation to bring oneself up to date on any missed coursework.

Course Syllabus (Tentative)

The course comprises five units of three hours each.

Unit 1: Introduction to Social Network Mining, Graph Models and Node Metrics

- Introduction to social network mining. Illustration of various social network mining tasks with real-world examples. Data characteristics unique to these settings and potential biases due to them.
- Social Networks as Graphs.
- Random graph models/ graph generators (Erdös-Rényi, power law, preferential attachment, small world, stochastic block models, kronecker graphs), degree distributions. Models of evolving networks.
- Node based metrics, ranking algorithms (Pagerank).
- Gephi graph visualization and exploration software – practice.

Unit 2: Social-Network Graph Analysis

- Social network exploration/ processing: graph kernels, graph classification, clustering of social-network graphs, centrality measures, community detection and mining, degeneracy (outlier detection and centrality), partitioning of graphs.
- SNAP system for large networks analysis and manipulation.
- Assignment on SNAP.

Unit 3: Social-Network Graph Analysis and Properties

- Social network exploration/ processing and properties: Finding overlapping communities, similarity between graph nodes, counting triangles in graphs, neighborhood properties of graphs.
- Pregel paradigm and Apache Giraph graph processing system.
- Assignment on Giraph.

Unit 4: Information Diffusion in Social Networks

- Strategic network formation: game theoretic models for network creation/ user behavior in social networks.
- Information diffusion in graphs: Cascading behavior, spreading, epidemics, heterogeneous social network mining, influence maximization, outbreak detection.
- Opinion analysis on social networks: Contagion, opinion formation, coordination and cooperation.
- Assignment on Gephi.

Unit 5: Dynamic Social Networks, Applications and Research Trends

- Dynamic social networks, Link prediction, Social learning on networks.
- Special issues in Information and Biological networks.
- Important applications of social network mining related to the above topics. Research trends.