CSE 291-F: Graph Mining and Network Analysis

Tentative Syllabus

Department of Computer Science and Engineering, UC San Diego, Spring 2017

Instructor: Fragkiskos Malliaros
Email: fmalliaros@ucsd.edu
Office: Atkinson Hall, 4111
Web: http://fragkiskos.me

Course Overview

Networks (or graphs) have become ubiquitous as data from diverse disciplines can naturally be mapped to graph structures. *Social networks*, such as academic collaboration networks and interaction networks over online social networking applications are used to represent and model the social ties among individuals. *Information networks*, including the hyperlink structure of the Web and blog networks, have become crucial mediums for information dissemination, offering an effective way to represent content and navigate through it. A plethora of *technological networks*, including the Internet, power grids, telephone networks and road networks are an important part of everyday life. The problem of extracting meaningful information from large scale graph data in an efficient and effective way has become crucial and challenging with several important applications and towards this end, graph mining and analysis methods constitute prominent tools.
The goal of this course is to present recent and state-of-the-art methods and algorithms for exploring, analyzing and mining large-scale networks, as well as their practical applications in various domains (e.g., social science, the web, biology).

Learning Objectives

This course aims to introduce students to the field of mining social and information networks by (i) covering a wide range of topics, methodologies and related applications; (ii) giving the students the opportunity to obtain hands-on experience on dealing with graph data and graph mining tasks through exercises and projects.
We expect that by the end of the course, the students will have a thorough understanding of various graph mining and learning tasks, will be able to analyze large-scale graph data as well as to formulate and solve problems that involve graph structures.

Prerequisites

There is no official prerequisite for this course. However, the students are expected to: (i) have basic knowledge of graph theory and linear algebra; (ii) be familiar with fundamental data mining and machine learning tasks; (iii) be familiar with at least one programming language (e.g., Python or any language of their preference).

Course Schedule (Tentative)

Week	Topics
1	• Introduction and overview of graph mining • Graph theory and linear algebra recap
2	• Real-world network properties (node degree distribution, shortest paths, clustering coefficient, small world phenomenon, properties of time evolving graphs) • Network generative models (random graphs, preferential attachment model, Kronecker graphs, stochastic blockmodels)
3	• Random walks • Graph centrality • Link analysis algorithms (PageRank and HITS)
4	• Graph clustering and community detection (spectral clustering, graph partitioning, modularity-based algorithms, community structure of real-world graphs, overlapping communities)
5	• Node similarity • Link prediction
6	• Graph kernels and graph similarity • Graph classification
7	• Learning embeddings in graphs • Applications in node classification and link prediction
8	• Influential spreaders • Influence maximization
9	• Topics in graph mining and applications • Dense subgraph detection • Rich network structures (signed and multilayer networks) and applications • Anomaly detection • Geo-social and location based networks
10	• Project presentations

Evaluation (Tentative)

The evaluation of the course will be based on the following:

1. **Hands-on practical assignments:** the hands-on assignments (two assignments) will familiarize the students with basic graph mining and analysis tasks (e.g., analysis of a social or information network such as the DBLP co-authorship network; crawl the data and construct the graph; examine the structural properties of the graph; find the underlying communities).
2. **Research project:** this will be the main component for the evaluation of the course. The students are expected to form groups of 3-4 people, propose a topic for their project, and finally, present the project at the end of the class (either in class or in poster session that we will organize).

Reading Material

Most of the material of the course is based on research articles. Some of the topics are also covered by the following books (all of them are publicly available):

- David Easley and Jon Kleinberg. Networks, Crowds, and Markets. Cambridge University Press, 2010.
- Deepayan Chakrabarti and Christos Faloutsos. Graph Mining: Laws, Tools, and Case Studies. Synthesis Lectures on Data Mining and Knowledge Discovery, Morgan & Claypool Publishers, 2012.
- Albert-László Barabási. Network Science. Cambridge University Press, 2016.