Graph-Based Analysis of Social Networks and Big Data

Rationale

Recent changes in the World Wide Web (WWW) facilitated media information sharing among users, user-based content creation and remote collaboration. Social networks allow users to post public profiles and share them with friends, consequently creating virtual communities and networks. Posting relevant media data (e.g., images, videos, music and text) in social media sites forces an increase of multimedia data storage, communication and consumption. Social media are concentrated on the creation and exchange of user-generated content, allowing users to create, search, share, rate and access multimedia data, thus creating a totally new media experience. This course addresses these scientific and technological issues, that is, the confluence of graph analysis, network theory, digital media, machine learning, and big data analysis.

Course Description

Mathematical foundations of social media analysis, and a comprehensive introduction to the use of graph theory in the study of social and digital media. Study of the confluence of graph analysis, network theory, big data analysis, and signal processing. Algebraic and combinatorial graph theory will be particularly applied for the analysis of graphs, in social media studies. Techniques such as spectral clustering, distributed tensor decomposition, matching, and random walks will be discussed. As social media inherits strong big data issues related to both size and content of the stored multimedia, emphasis will be placed on the analysis of big data. Parallel and distributed
programming techniques to address graph partitioning problems (e.g., clustering) will be discussed.

**Topic List**

Topics may include but won’t be limited to:

1. **Graphs in Social and Digital Media**
   - Dominant social networking/media platforms
   - Collecting data from social media sites
   - Social media graphs (Facebook, Twitter, etc.)
   - Graph storage formats and visualization
   - Big data issues in social land digital media

2. **Mathematical Preliminaries: Graphs and Matrices**
   - Adjacency, and Laplacian matrices
   - Review of relevant Graph Theoretical algorithms (e.g., Depth-first-search, Dijkstra’s shortest path, minimum spanning trees, etc.)
   - Centrality, betweenness and closeness

3. **Algebraic Graph Analysis**
   - Spectral graph theory
   - Similarity matrix and nearest neighbor graph
   - Random graph generation and models
   - Spectral graph generation
   - Graph clustering
   - Graph matching
   - Spectral graph matching
   - Random walks

Programming Assignment (spectral clustering, matching, or random walks)
4. Label Propagation and Information Diffusion in Graphs
   - Graph construction approaches
   - Label inference methods
   - Social networks diffusion models

5. Graph-based Pattern Classification and Dimensionality Reduction

6. Matrix and Tensor Factorization with Recommender System Applications

7. Big Data Analytics of Social Networks

8. Semantic Model Adaptation for Evolving Big Social Data
   (a) Parallel and Distributed Approaches for Big Data Analysis
      - Parallel probabilistic latent semantic analysis
      - Parallel spectral clustering
      - Distributed tensor decomposition
   (b) Programming Assignment (Parallel Probabilistic Latent Semantics, Parallel Spectral Clustering, or Distributed Tensor Decomposition)

Recommended Book
- Graph-Based Social Media Analysis, Edited by Ioannis Pitas, CRC Press (2016) — ISBN 9781498719056

Learning Goals
The student must be able to demonstrate a working knowledge of the theoretical foundations of Graph Analytics by showing proficiency in following topics:

- Theoretical methods to determine social influence in media networks by application of known graph theoretical algorithms.
- Graph partitioning (clustering) by application of spectral, matching, or random walks techniques.
• Learn about social networks diffusion models and dimensionality reduction methods.

• Apply parallel/distributed techniques to handle big data problems. Besides clustering problems, partitioning of data to be distributed on a cluster of nodes to solve other graph theory algorithms will be investigated.

Assessment

• Programming exercises will be assigned to make sure students are capable of applying theoretical knowledge to implement algorithms for partitioning problems as well as implementation of parallel/distributed techniques to handle big data. 50%

• Midterm exam to demonstrate proficiency in but not limited to techniques to measure social influence, and partitioning methods for clustering problems. 25%

• Final exam to demonstrate knowledge in but not limited to graph diffusion models, dimensionality reduction methods, and parallel/distributed partitioning problems of large graphs. 25%

Contact Information
Professor Louis Petingi- email:louis.petingi@csi.cuny.edu