Orientation – Graph Mining

Prof. O-Joun Lee

Network Science Lab
Dept. of Artificial Intelligence
The Catholic University of Korea

e-mail: ojlee@catholic.ac.kr

네트워크 과학 연구실
NETWORK SCIENCE LAB

가톨릭대학교
THE CATHOLIC UNIVERSITY OF KOREA

Contents

- About This Course
- Reference Book
- Course Schedule
- Assessment
- Assignments
- Instructor Information
- Teaching Assistant Information

1. About This Course

Goal

- This lecture provides understandings of concepts and methodologies of graph mining.
- Students will learn fundamental concepts and properties of the networked data.
- Based on the concepts, the students will learn metrics and analysis methods for structural properties of graphs.
- The students will obtain capabilities for analyzing unstructured big data with graphs.

Contents

- Fundamental concepts of network science
- Power laws and scale-free networks
- Structure, nodes, and links analysis
- Network communities
- Node classification and link prediction
- Diffusion of information
- Influence propagation

2. Reference Book

- Students do not need to buy textbooks
- Most of the lecture materials will be provided in digital format (e.g., PDF files)

Textbooks

- William L. Hamilton, “Graph Representation Learning Book,” Morgan and Claypool, 2020
- David Easley and Jon Kleinberg, “Networks, Crowds, and Markets: Reasoning About a Highly Connected World,” Cambridge University Press, 2010
- Albert-László Barabási, “Network science,” Cambridge University Press, 2017

3. Course Schedule

- **Week 1: Introduction to Graph Mining**
  - Overview of graph mining
  - Types of graphs and their applications
  - Basic graph concepts (nodes, edges, degree, etc.)

- **Week 2: Graph Representation and Storage**
  - Adjacency matrix and list
  - Sparse matrix representations
  - Graph databases and storage systems

- **Week 3: Centrality Measures**
  - Degree centrality
  - Betweenness centrality
  - Closeness centrality
  - Eigenvector centrality
  - PageRank and its application in web search

3. Course Schedule

- **Week 4: Graph Visualization**
  - Visualization techniques (spring-embedded, circular, etc.)
  - Tools for graph exploration and visualization (Gephi, Cytoscape, etc.)

- **Week 5: Community Detection**
  - Definition of communities and their properties
  - Clustering techniques (k-means, hierarchical, etc.)
  - Modularity and its variants
  - Tools for community detection (Louvain, Infomap, etc.)

- **Week 6: Link Prediction**
  - Types of links (positive, negative, neutral)
  - Commonly used link prediction methods (common neighbors, Jaccard coefficient, etc.)
  - Tools for link prediction

3. Course Schedule

- **Week 7: Subgraph Mining**
  - Frequent subgraph mining (FSM)
  - FSM algorithms (gSpan, FSG, etc.)
  - Tools for frequent subgraph mining (Traceminer, etc.)
- **Week 8: Mid-term Exam**
- **Week 9: Graph Kernels**
  - Overview of graph kernels
  - WL relabeling process
  - Applications of graph kernels
- **Week 10: Node Classification**
  - Overview of node classification
  - Feature extraction methods
  - Classification algorithms (SVM, Random Forest, etc.)
  - Evaluation metrics

3. Course Schedule

- **Week 11: Graph Applications in Social networks**
  - Overview of graph applications in social networks
  - Social network analysis
  - Community detection and link prediction in social networks

- **Week 12: Graph Applications in Bioinformatics**
  - Overview of graph applications in biological networks
  - Graph pattern mining
  - Graph classification task in Bioinformatics

- **Week 13: Graph Applications in web graph**
  - Overview of graph applications in web graph
  - Web graph crawling
  - Web graph analysis

3. Course Schedule

- **Week 14: Additional Topics and Wrap-up**
  - Graph neural networks
  - Graph Transformers
  - Summary of key concepts
  - Future directions in graph

- **Week 16: Final Exam**

4. Assessment Criteria and Grading Scale

Assessment

- Midterm exam: 30%
- Final exam: 30%
- Homework: 30%
- Attendance: 10%
- All test questions are made up within lecture slides and assignments

Tentative Grading Scale

- A: 80~100
- B: 60~80
- C: ~60
- F: Absence from exams

Weekly Programming Practice

- Performing practical programming in one class every week

6. Instructor Information

O-Joun Lee (이오준)

- Email: ojlee@catholic.ac.kr
- Office: T404
- Office phone number: 5516
- Office hour: visit office anytime
- Please feel free to contact me for any class issues

7. Teaching Assistant Information

Van Thuy Hoang

- Email: hoangvanthuy90@gmail.com
- Office: B348

Tien-Bach-Thanh Do

- Email: osfa19730@catholic.ac.kr
- Office: B348

네트워크 과학 연구실
NETWORK SCIENCE LAB

가톨릭대학교
THE CATHOLIC UNIVERSITY OF KOREA