

**FIRST SEMESTER 2021-2022**

# Course Handout Part II

Date: 20-08-2021

In addition to part-I (General Handout for all courses appended to the time table) this portion gives further specific details regarding the course.

*Course No.* : ***CS F426***

## Course Title : **Graph Mining**

## Instructor-in-Charge : **Dr. Apurba Das**

**Scope and Objective of the Course:**

This course studies managing and mining graphs which are massive and cannot held in main memory as the size of the applications are often very large. Classic examples of graphs are web, social networks, computational biology, communication networking etc. In some cases, the entire graph is not available in the form of continuous stream. Edges are received continuously with time. The course includes the basics of the graphs, static and dynamic graphs, PageRank & random walks, graph or graph node classification, graph clustering, community detection, anomaly detection, frequent sub-graph mining. The course is designed to provide students with an understanding parallel and streaming graph mining to deal with massive graphs and evolving graphs. The course also aims at providing a holistic view of dealing and mining with graphs.

**Textbooks:**

1. T1: Jure Leskovec, Anand Rajaraman, Jeff Ullman. Mining of Massive Datasets. Book 3rd edition. 2020, Cambridge University Press. [<http://www.mmds.org/> ]

**Reference books**

1. **R1:** Agarwal Charu C. and Wang Haixun, Managing and Mining Graph Data, Springer
2. **R2:** Social Media Mining - An introduction, Reza Zafarani, Mohammad Ali Abbasi, Huan Liu , Cambridge University Press , 2014 [<http://dmml.asu.edu/smm/SMM.pdf>]
3. **R3:** Chakrabarti, D. and Faloutsos, C., 2012. Graph mining: laws, tools, and case studies. Synthesis Lectures on Data Mining and Knowledge Discovery, 7(1), pp.1-207.
4. **R4:**Mohammad Hossein Bateni Soheil Behnezhad Mahsa Derakhshan MohammadTaghi Hajiaghayi Raimondas Kiveris Silvio Lattanzi Vahab Mirrokni, Affinity Clustering: Hierarchical Clustering at Scale, NIPS 2017, pp. 6867-6877 [<https://research.google/pubs/pub46700/> ]
5. **R5:** Fenxiao Chen, Yun-Cheng Wang, Bin Wang and C.-C. Jay Kuo, Graph representation learning: a survey DOI: https://doi.org/10.1017/ATSIP.2020.13, Cambridge University Press: 28 May 2020 [<https://arxiv.org/pdf/1909.00958.pdf> ]
6. **R6:** William L. Hamilton, Rex Ying, Jure Leskovec, Representation Learning on Graphs: Methods and Applications, [<https://www-cs.stanford.edu/people/jure/pubs/graphrepresentation-ieee17.pdf> ]
7. **R7:** Easley, D. and Kleinberg, J., 2010. Networks, crowds, and markets: Reasoning about a highly connected world. Cambridge University Press.
8. **R8**: R Angles, et al. Foundations of Modern Query Languages for Graph Databases, 2017, ACM Computing Surveys

# **R9**: D Koutra and Faloutsos, C., Individual and Collective Graph Mining: Principles, Algorithms, and Applications, 2017, <https://www.morganclaypool.com/doi/10.2200/S00796ED1V01Y201708DMK014>

1. **R10**: L. Lu and T Zhou, Link Prediction in Complex Networks: A Survey

# **R11**: M Needham and A. E. Hodler, Graph Algorithms: Practical Examples in Apache Spark and Neo4j, [https://neo4j.com/graph-algorithms-book/]

**Course Plan:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Lecture No.** | **Learning objectives** | **Topics to be covered** | **Chapter in the Text Book** |
| 1 | **M1:**  **Graph Basics**  **Static and dynamic graphs**  To understand graphs and their applications  To understand how to traverse graphs and perform random walks to rank nodes for different applications  To learn how to do computations in parallel for massive graphs | Course Overview  Graph Terminology and Applications | R1 Ch 1,  R2 Ch 1 |
| 2-4 | Static Graphs and basic algorithms | R2 Ch 2  R1 Ch 2 |
| 5-6 | Dynamic Graphs | Class Notes |
| 7-10 | Link Analysis  PageRank, Connectivity, triangle computation, Personalized Random walk, parallel computation | T1 Ch 5 |
| 11-12 | **M2**: **Managing Graph Data**  To study how to manage, index and query graph data | Representing and Using Graph Data | R1: Ch 2.1, 2.2 |
| 13-14 | Indexing & Querying Graph data  Feature Based Graph Index  Indexing for Similarity Search | R1: Ch 5 |
| 14-15 | Graph Querying Algorithms  Patterns and querying for Patterns  Navigational Querries | R8 |
| 16-19 | **M3:**  **Graph Mining**  To study how to search graphs based on key words  To understand how to classify graph nodes/graphs with labeled data or without labeled data  To learn how to detect communities from a graph and how to compare two graphs/graph nodes  To learn how to detect anomalies and mining frequent subgraphs  To understand how graph mining is required and applied to real problems | Node Classification  Link Based methods  Belief Propagation  Semi-supervised learning  Bipartite Graphs  SimRank  Link Prediction  Graph classification | R1 Ch 11 R9  R10 |
| 20-24 | Graph Clustering  Clustering in Bipartite graphs  Hierarchical Clustering  Community Detection  Graph Partitioning  Spectral methods  Graph Similarity and Alignment  Iterative similarity methods  Edge similarity scores  Network Similarity Methods | T2: 10.1-10.6  R4  R1 Ch 9 |
| 25-29 | Sub-graph mining  Sub-graph Enumeration  Frequent Subgraph Mining  Apriori-based Approach  Pattern-growth Approach  Mining Significant Graph Patterns  Branch-and-Bound Approach | T1 10.7 |
| 30 | Case Study |  |
| 31-36 | **M4:**  **Advanced Topics** | Deep learning for graphs  Graph Convolution Networks (GCN)  Link Prediction  Graph Representation Learning | R5  R11 |
| 37-38 | Distributed Graph Algorithms | Research Papers |
| 39-42 | Streaming Graphs  Stream Model for massive graphs  Graph streams | T2 Ch 4  T1 Ch13 |

**Evaluation Scheme:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Component** | **Duration** | **Weightage (%)** | **Date & Time** | **Nature of Component** |
| Mid-Semester Test | 90 Mins. | 30 | 19/10/2021 1.30 - 3.00PM | Open Book |
| Lab based evaluation (3 Nos.) |  | 25 | TBA | Open Book |
| Project |  | 10 | TBA | Open Book |
| Comprehensive Examination | 120 Mins. | 35 | 15/12 AN | Open Book |

Note: ***40% of the evaluation to be completed by midsem grading.***

**Consultation Hour:**

To be announced in the class

**Notices:**

**Make-up Policy:**

Prior Permission is must and Make-up shall be granted only in genuine cases based on individual’s need and circumstances.

**Academic Honesty and Integrity Policy:**

Academic honesty and integrity are to be maintained by all the students throughout the semester and no type of academic dishonesty is acceptable.

**INSTRUCTOR-IN-CHARGE**

**Apurba Das**