BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI
WORK INTEGRATED LEARNING PROGRAMMES
Digital

Part A: Content Design

Course Title	Graphs - Algorithms and Mining
Course No(s)	DSECL ZG521
Credit Units	**5**
Credit Model	
Content Authors	Poonam Goyal

Course Objectives

No	
CO1	To learn how to mine graphs
CO2	To learn how to apply these techniques to specific applications such as web, social science, computer networks, neuroscience, etc.
CO3	To learn about analyzing large datasets using scalable and practical methods.

Text Book(s)

T1	-
T2	-

Reference Book(s) & other resources

R1	Agarwal Charu C. and Wang Haixun, ***"Managing and Mining Graph Data"***
R2	Jure Leskovec, Anand Rajaraman, Jeff Ullman. "***Mining of Massive Datasets***". Book 2nd edition. Cambridge University Press.
R3	Tan P. N., Steinbach M & Kumar V. "***Introduction to Data Mining"*** Pearson Education, 2006

Content Structure

1. Introduction
   - 1.1. What are Graphs?
   - 1.2. Graph Applications
   - 1.3. Preliminaries
2. Static Graphs
   - 2.1. Degree distribution, Power Laws,
   - 2.2. Diameter
   - 2.3. Connected components
   - 2.4. Eigen-spokes
   - 2.5. Scalability
3. Dynamic graphs
4. PageRank, Connectivity and Triangle Computations
   - 4.1. Personalized Random Walk
   - 4.2. Parallel Computation
5. Key word search on graph data
6. Node Classification
   - 6.1. Belief Propagation
   - 6.2. Semi-Supervised Learning
   - 6.3. Node Similarity
7. Graph Clustering
   - 7.1. Community Detection

7.2. Graph Partitioning
8. Graph Similarity and Alignment
8.1. Network Similarity Methods
8.2. Network Alignment
9. Graph Summarization
10. Anomaly Detection
11. Sub-graph mining
11.1. Sub-graph Enumeration
11.2. Frequent Subgraph Mining
12. Streaming Graphs
13. Deep learning for graphs

Learning Outcomes:

No	Learning Outcomes
LO1	To understand graphs and their applications
LO2	To understand the role of distributed computing in data intensive data mining
LO3	To study how to search graphs based on key words
LO4	To understand how text mining is different from data mining and how to mine it
LO5	To understand what goes into the web search and to study methods of web search and their improvements
LO6	To understand how to mine complex structures other than records while retaining the relations among the entities

Readings

1. U Kang, Charalampos E. Tsourakakis, Ana Paula Appel, Christos Faloutsos, and Jure Leskovec, **Radius Plots for Mining Tera-byte Scale Graphs: Algorithms, Patterns, and Observations**, SIAM International Conference on Data Mining (SDM) 2010.
2. C. E. Tsourakakis, **Fast Counting of Triangles in Large Real Networks: Algorithms and Laws**, ICDM, 2008.
3. J. Leskovec, L. Backstrom, R. Kumar, A. Tomkins. **Microscopic Evolution of Social Networks**. KDD 2008.
4. J. Leskovec, D. Chakrabarti, J. Kleinberg, C. Faloutsos. **Realistic, Mathematically Tractable Graph Generation and Evolution, Using Kronecker Multiplication**. ECML/PKDD, 2005.
5. M. McGlohon, L. Akoglu, C. Faloutsos. **Weighted Graphs and Disconnected Components: Patterns and a Generator**. ACM SIGKDD, 2008.
6. R. Kumar, J. Novak, A. Tomkins. **Structure and evolution of online social network**s. KDD, 2006.
7. Brin, S. and Page, L. (1998) **The Anatomy of a Large-Scale Hypertextual Web Search Engine**. In: Seventh International World-Wide Web Conference (WWW 1998), April 14-18, 1998, Brisbane, Australia
8. Jon M. Kleinberg. **Authoritative Sources in a Hyperlinked Environment**. Journal of the ACM (JACM), 46(5):604-632, 1999.
9. Page, Lawrence and Brin, Sergey and Motwani, Rajeev and Winograd, Terry (1999**). The PageRank Citation Ranking: Bringing Order to the Web**. Technical Report. Stanford InfoLab.
10. Danai Koutra, Tai-You Ke, U Kang, Duen Horng (Polo) Chau, Hsing-Kuo Kenneth Pao, and Christos Faloutsos. **Unifying Guilt-by-Association Approaches: Theorems and Fast Algorithms**. ECML PKDD 2011.

11. Keith Henderson, Brian Gallagher, Tina Eliassi-Rad, Hanghang Tong, Sugato Basu, Leman Akoglu, Danai Koutra, Christos Faloutsos, Lei Li. **RolX: structural role extraction & mining in large graphs**. In Proceedings of 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2012.

12. Sujith Ravi, Qiming Diao. **Large Scale Distributed Semi-Supervised Learning Using Streaming Approximation**. AISTATS 2016.

13. J. Ugander, Lars Backstrom. **Balanced Label Propagation for Partitioning Massive Graphs**. In Proceedings of the sixth ACM international conference on Web search and data mining (WSDM), 2013.

14. Weizhong Zhao, Gang Chen, Xiaowei Xu. AnySCAN: **An Efficient Anytime Framework with Active Learning for Large-scale Network Clustering**. In Proceedings of IEEE 17th International Conference on Data Mining (ICDM'17).

15. Sucheta Soundarajan, Tina Eliassi-Rad, Brian Gallagher. **A Guide to Selecting a Network Similarity Method**. SDM 2014: 1037-1045.

16. Si Zhang, Hanghang Tong**. FINAL: Fast Attributed Network Alignment**. KDD 2016.

17. D. Koutra, U. Kang, J. Vreeken, C. Faloutsos. **Summarizing and Understanding Large Graphs.** Special Issue of Statistical Analysis and Data Mining, "Best of SDM 2014". October 2014.

18. Neil Shah, Danai Koutra, Tianmin Zou, Brian Gallagher, Christos Faloutsos. **TimeCrunch: Interpretable Dynamic Graph Summarization**. Conference on Knowledge Discovery and Data Mining (KDD), August 2015.

19. Saket Navlakha**. Learning the Structural Vocabulary of a Network. Neural Computation** 29, 287–312 (2017).

20. Manish Purohit, B. Aditya Prakash, Chanhyun Kang, Yao Zhang, and V.S. Subrahmanian**. Fast Influence-based Coarsening for Large Networks**. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '14).

21. Dan Feldman, Sedat Ozer, and Daniela Rus. **Coresets for Vector Summarization with Applications to Network Graphs**. In Proceedings of the 34th International Conference on Machine Learning (ICML '17).

22. Aditya Grover and Jure Leskovec. 2016**. node2vec: Scalable Feature Learning for Networks**. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). ACM, New York, NY, USA, 855-864.

23. Leonardo F. R. Ribeiro, Pedro H. P. Saverese, Daniel R. Figueiredo. **Struc2vec: Learning Node Representations from Structural Identity.** In Proceedings of the 23rd ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '17).

24. Anjuli Kannan, Karol Kurach, Sujith Ravi, Tobias Kaufmann, Andrew Tomkins, Balint Miklos, Greg Corrado, Laszlo Lukacs, Marina Ganea, Peter Young, Vivek Ramavajjala. **Smart Reply: Automated Response Suggestion for Email.** In Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '16).

25. Thomas N Kipf and Max Welling**. Semi-Supervised Classification with Graph Convolutional Networks**. International Conference on Learning Representations (ICLR) 2017.

26. Yanfang Ye, Shifu Hou, Yangqiu Song. **HinDroid: An Intelligent Android Malware Detection System Based on Structured Heterogeneous Information Network**. In Proceedings of the 23rd ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '17). Best Data Science Paper

27. Bryan Perozzi, Leman Akoglu, Patricia Iglesias Sánchez, and Emmanuel Müller. 2014. **Focused clustering and outlier detection in large attributed graphs**. KDD 2014.

Part B: Learning Plan

Academic Term	Second Semester 2019-2020
Course Title	Graphs - Algorithms and Mining
Course No	
Lead Instructor	Poonam Goyal

Contact Session 1

Type	Content Ref.	Topic Title	Study/HW Resource Reference
Pre CH		Introduction What are Graphs? Graph Applications Preliminaries Introduction to course	R1: Ch 1 & 2 and Class Slides
During CH			
Post CH			

Contact Session 2

Type	Content Ref.	Topic Title	Study/HW Resource Reference
Pre CH		Static Graphs Degree distribution, Power Laws, Diameter Connected components Eigen-spokes Scalability	R1: Ch 3 and Class Slides
During CH			
Post CH			

Contact Session 3

Type	Content Ref.	Topic Title	Study/HW Resource Reference
Pre CH		Static Graphs Degree distribution, Power Laws, Diameter Connected components Eigen-spokes Scalability	R1: Ch 3 and Class Slides
During CH			
Post CH			

Contact Session 4

Type	Content Ref.	Topic Title	Study/HW Resource Reference
Pre CH		Dynamic graphs	Class Slides
During CH			
Post CH			

Contact Session 5

Type	Content Ref.	Topic Title	Study/HW Resource Reference
Pre CH		PageRank, Connectivity and Triangle Computations Personalized Random Walk Parallel Computation	R2 Ch 5 and Class Slides
During CH			
Post CH			

Contact Session 6

Type	Content Ref.	Topic Title	Study/HW Resource Reference
Pre CH		PageRank, Connectivity and Triangle Computations Personalized Random Walk Parallel Computation	R2 Ch 5 and Class Slides
During CH			
Post CH			

Contact Session 7

Type	Content Ref.	Topic Title	Study/HW Resource Reference
Pre CH		Key word search on graph data	R1 Ch 8 and Class Slides
During CH			
Post CH			

Contact Session 8

Type	Content Ref.	Topic Title	Study/HW Resource Reference
Pre CH	See Class Slides	Node Classification Belief Propagation Semi-Supervised Learning Node Similarity	Class Slides
During CH			
Post CH			

Contact Session 9

Type	Content Ref.	Topic Title	Study/HW Resource Reference
Pre CH	See Class Slides	Graph Clustering Community Detection Graph Partitioning Graph Similarity and Alignment Network Similarity Methods Network Alignment	R1 Ch 9
During CH			
Post CH			

Contact Session 10

Type	Content Ref.	Topic Title	Study/HW Resource Reference
Pre CH	See Class Slides	Graph Clustering Community Detection Graph Partitioning Graph Similarity and Alignment Network Similarity Methods Network Alignment	R1 Ch 9
During CH			
Post CH			

Contact Session 11

Type	Content Ref.	Topic Title	Study/HW Resource Reference
Pre CH	See Class Slides	Graph Summarization	Class notes (Research Papers)
During CH			
Post CH			

Contact Session 11

Type	Content Ref.	Topic Title	Study/HW Resource Reference
Pre CH	See Class Slides	Anomaly Detection	See Class Slides
During CH			
Post CH			

Contact Session 12

Type	Content Ref.	Topic Title	Study/HW Resource Reference
Pre CH	See Class Slides	Sub-graph mining Sub-graph Enumeration Frequent Subgraph Mining	See Class Slides
During CH			
Post CH			

Contact Session 13

Type	Content Ref.	Topic Title	Study/HW Resource Reference
Pre CH	See Class Slides	Sub-graph mining Sub-graph Enumeration Frequent Subgraph Mining	See Class Slides
During CH			
Post CH			

Contact Session 14

Type	Content Ref.	Topic Title	Study/HW Resource Reference
Pre CH	See Class Slides	Streaming Graphs	See Class Slides
During CH			
Post CH			

Contact Session 15

Type	Content Ref.	Topic Title	Study/HW Resource Reference
Pre CH	See Class Slides	Streaming Graphs	See Class Slides
During CH			
Post CH			

Contact Session 16

Type	Content Ref.	Topic Title	Study/HW Resource Reference
Pre CH	See Class Slides	Deep learning for graphs	See Class Slides
During CH			
Post CH			

Evaluation Scheme:
Legend: EC = Evaluation Component; AN = After Noon Session; FN = Fore Noon Session

No	Name	Type	Duration	Weight	Day, Date, Session, Time
EC-1	~~Quiz-I~~/ Assignment-I	Online	-	30%	The assignments will be given after every contact session, but submission will be once in a month (dates will be declared along with the assignment) *First Assignment is already given and deadline is also over*
	~~Quiz-II~~/ Assignment-II				
	~~Quiz-III~~/ Assignment-III				
	Assignment- III				
	Assignment-V				
EC-2	Mid-Semester Test	Closed Book	90 minutes	30%	TBA
EC-3	Comprehensive Exam	Open Book	150 minutes	40%	TBA

Syllabus for Mid-Semester Test (Closed Book): Topics in Session Nos. 1 to 16
Syllabus for Comprehensive Exam (Open Book): All topics (Session Nos. 1 to 32)

Important links and information:

Elearn portal: https://elearn.bits-pilani.ac.in
Students are expected to visit the Elearn portal on a regular basis and stay up to date with the latest announcements and deadlines.
Contact sessions: Students should attend the online lectures as per the schedule provided on the Elearn portal.
Evaluation Guidelines:

1. ~~EC-1 consists of either two Assignments or three Quizzes. Students will attempt them through the course pages on the Elearn portal. Announcements will be made on the portal, in a timely manner.~~
2. For Closed Book tests: No books or reference material of any kind will be permitted.
3. For Open Book exams: Use of books and any printed / written reference material (filed or bound) is permitted. However, loose sheets of paper will not be allowed. Use of calculators is permitted in all exams. Laptops/Mobiles of any kind are not allowed. Exchange of any material is not allowed.
4. If a student is unable to appear for the Regular Test/Exam due to genuine exigencies, the student should follow the procedure to apply for the Make-Up Test/Exam which will be made available on the Elearn portal. The Make-Up Test/Exam will be conducted only at selected exam centres on the dates to be announced later.

It shall be the responsibility of the individual student to be regular in maintaining the self study schedule as given in the course handout, attend the online lectures, and take all the prescribed evaluation components such as Assignment/Quiz, Mid-Semester Test and Comprehensive Exam according to the evaluation scheme provided in the handout.