IITB-Monash Research Academy

Indian Institute of Technology Bombay | MONASH University

An Indian-Australian research partnership

Project Title:	**Representation Learning of Domain-specific Graphs**	
Project Number	**IMURA0761**	
Monash Main Supervisor (Name, Email Id, Phone)	Yuan-Fang Li, yuanfang.li@monash.edu	*Full name, Email*
Monash Co-supervisor(s) (Name, Email Id, Phone)		
Monash Head of Dept/Centre (Name,Email)	Alan Dorin, alan.dorin@monash.edu	*Full name, email*
Monash Department:	Faculty of Information Technology	
Monash ADRT (Name,Email)	Sue, McKemmish, sue.mckemmish@monash.edu	*Full name, email*
IITB Main Supervisor (Name, Email Id, Phone)	Ganesh Ramakrishnan, ganesh@cse.iitb.ac.in	*Full name, Email*
IITB Co-supervisor(s) (Name, Email Id, Phone)		*Full name, Email*
IITB Head of Dept (Name, Email, Phone)	Uday Khedkar, head@cse.iitb.ac.in	*Full name, email*
IITB Department:	Department of Computer Science and Engineering	

Research Clusters:

Research Themes:

Highlight which of the Academy's CLUSTERS this project will address? *(Please nominate JUST **one.** For more information, see www.iitbmonash.org)*		**Highlight which of the Academy's Theme(s) this project will address?** *(Feel free to nominate more than one. For more information, see www.iitbmonash.org)*	
1	*Material Science/Engineering (including Nano, Metallurgy)*	***1***	*Advanced computational engineering, simulation and manufacture*
2	*Energy, Green Chem, Chemistry, Catalysis, Reaction Eng*	2	*Infrastructure Engineering*
3	*Math, CFD, Modelling, Manufacturing*	3	*Clean Energy*
4	***CSE, IT, Optimisation, Data, Sensors, Systems, Signal Processing, Control***	4	*Water*
5	*Earth Sciences and Civil Engineering (Geo, Water, Climate)*	5	*Nanotechnology*
6	*Bio, Stem Cells, Bio Chem, Pharma, Food*	6	*Biotechnology and Stem Cell Research*
7	*Semi-Conductors, Optics, Photonics, Networks, Telecomm, Power Eng*	7	*Humanities and social sciences*
8	*HSS, Design, Management*	8	Design

The research problem

Define the problem

Representation learning of graphs is the problem of learning a mapping that embeds nodes and edges of a graph as points in a low-dimensional space. This problem has recently garnered intense research interest in machine learning, as the vector-based representation of graphs has been shown to be useful in many important learning tasks.
Graphs are prevalent in many domains. Graph representation learning that takes into account domain-specific knowledge/information is a novel and challenging problem. This is extremely useful in tasks such as link prediction, node classification, community discovery, etc.

Additionally, the continuous rise in inter-connected data creates the need for summarizing large graphs to extract relevant information [1][2]. It is useful in various areas such as social network analysis, etc. Currently, visualisation of large graphs is almost impossible. The project will also aim to develop a framework for summarization and visualising large graphs, including analysis of the textual content that is available, for example, in social network graphs. Associated technical questions are : How can we summarize large weighted, directed graphs ? How can we build big data, fast analysis systems to parse and understand large amounts of inter-connected data ?
In addition, graph-based user behavior modeling is an exciting area with huge applications. Capturing contextual data to gain insights into the user behavior. Moreover, relational databases can be considered as large hypergraphs. By modeling graphs incorporating contextual and relational information, it can be applied in areas such as biomedical, education, etc [3].

1. Danai Koutra, U Kang, Jilles Vreeken, and Christos Faloutsos. VoG: Summarizing and Understanding Large Graphs. In Proceedings of the 14th SIAM International Conference on Data Mining (SDM), Philadelphia, PA, pages 91–99, 2014.
2. Danai Koutra, U Kang, Jilles Vreeken, and Christos Faloutsos. Summarizing and Understanding Large Graphs. In Statistical Analysis and Data Mining. John Wiley & Sons, Inc., 2015.
3. Beutel, Alex and Akoglu, Leman and Faloutsos, Christos , Graph-Based User Behavior Modeling: From Prediction to Fraud Detection

Project aims

Define the aims of the project

This problem will address the problem of graph representation learning in a domain-specific setting. This project will investigate a number of domains, including software engineering (source code graph, software engineering knowledge graphs, etc.) and biomedicine (domain taxonomies, publications, etc.). It will investigate novel learning algorithms to

- efficiently learn representations of domain graphs,
- effectively make use of domain knowledge and/or other auxiliary information.

Expected outcomes

Highlight the expected outcomes of the project

If completed successfully, we expect to achieve the following outcomes:

- Novel learning algorithms for representation learning of domain-specific graphs.

How will the project address the Goals of the above Themes?

Describe how the project will address the goals of one or more of the 6 Themes listed above.

Capabilities and Degrees Required

List the ideal set of capabilities that a student should have for this project. Feel free to be as specific or as general as you like. These capabilities will be input into the online application form and students who opt for this project will be required to show that they can demonstrate these capabilities.

- Background and experiences in machine learning.
- Strong mathematical knowledge and analytical skills.
- Strong programming skills

Potential Collaborators

Please visit the IITB website www.iitb.ac.in OR Monash Website www.monash.edu to highlight some potential collaborators that would be best suited for the area of research you are intending to float.

Soumen Chakrabarti (IITB)
Lan Du (Monash)

Select up to **(4)** keywords from the Academy's approved keyword list **(available at http://www.iitbmonash.org/becoming-a-research-supervisor/)** relating to this project to make it easier for the students to apply.

Data Science,optimisation,algorithms
Natural Language Processing
Modelling and Simulation