CS6208 : Advanced Topics in Artificial Intelligence
Graph Machine Learning

Administrative (Week 6)

Semester 2 2022/23

Xavier Bresson

https://twitter.com/xbresson

Department of Computer Science
National University of Singapore (NUS)

Outline

- Week 7
- Office hours
- Course evaluation
  - Paper review
  - Group project
- Today lecture

Week 7

- Week 7 : No lecture (conference travel)
  - Organizing the workshop “Deep Learning and Discrete Optimization”
  - http://www.ipam.ucla.edu/programs/workshops/artificial-intelligence-and-discrete-optimization
  - Check out online videos if you are interested on DL+NP-hard problems

Artificial Intelligence and Discrete Optimization

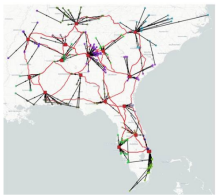
FEBRUARY 27 - MARCH 3, 2023

OVERVIEW | SPEAKER LIST | LODGING | APPLICATION & REGISTRATION

Overview

In recent years, the use of Machine Learning techniques to Operations Research (OR) problems, especially in the Discrete Optimization (DO) a.k.a. Combinatorial Optimization context, opens very interesting scenarios because DO is the “home” of an endless list of decision-making problems that are of fundamental importance in multitude applications.

The workshop will bring together experts in mathematics (optimization, graph theory, sparsity, combinatorics, statistics), operations research (assignment problems, routing, planning, Bayesian search, automation, scheduling), machine learning (deep learning, supervised, self-supervised and reinforcement learning) and artificial intelligence at large (including multi-agent systems, interpretability, fairness, etc.). In addition, the focus will be on



- Algorithmic challenges and potential of the interaction between AI and OR;
- Data requirements in which such an interaction can be profitable; and
- Application areas that are likely to lead to game-changing results (e.g., transportation, supply chain, public policy, energy).

Program Flyer

ORGANIZING COMMITTEE

Xavier Bresson (National University of Singapore)
Bistra Dilkina (University of Southern California (USC))
Andrea Lodi (Cornell University)
Pascal Van Hentenryck (Georgia Institute of Technology)

Speaker List

Matej Balog (DeepMind)
Hamsa Bastani (University of Pennsylvania)
Timo Berthold (Technische Universität Berlin)
Xavier Bresson (National University of Singapore)
Bistra Dilkina (University of Southern California (USC))
Priya Donti (Cornell University)
Tina Eliassi-Rad (Northeastern University)
Paul Grigas (University of California, Berkeley (UC Berkeley))
Tias Guns (KU Leuven)
Stefanie Jegelka (Massachusetts Institute of Technology)
Elias Khalil (University of Toronto)
Thomas Laurent (Loyola Marymount University)
Andrea Lodi (Cornell University)
Rahul Mazumder (Massachusetts Institute of Technology)
Azalia Mirhoseini (Google AI)
Vinod Nair (Google Brain)
Louis-Martin Rousseau (École Polytechnique de Montréal)
Le Song (Georgia Institute of Technology)
Bartolomeo Stellato (Princeton University)
Yuandong Tian (Artificial Intelligence Center)
Pascal Van Hentenryck (Georgia Institute of Technology)
Phebe Vayanos (University of Southern California (USC))
Petar Veličković (DeepMind Technologies)
Ellen Vitercik (Stanford University)
Segev Wasserkug (IBM Research - Israel)

Outline

- Week 7
- Office hours
- Course evaluation
  - Paper review
  - Group project
- Today lecture

Office hours

- From Week 8, I will offer office hours to discuss lectures, etc
  - When/where?
    - Tue 4:30pm-5:30, my office COM2-04-26
  - How?
    - Appointment : Send me an email for priority, xaviercs@nus.edu.sg
    - Walk-in : I will be at my office until 5pm (if nobody at 5pm, office hour stops)
  - 15min max per student (more time if no student is waiting)

Outline

- Week 7
- Office hours
- Course evaluation
  - Paper review
  - Group project
- Today lecture

Evaluation

- This module is 100% CA, i.e., there is no final exam.
- There are 2 components
  - Individual assessment (50%) : 1 paper review
  - Group assessment (50%) : 1 project

Paper review

- Paper review counts for 50% of the final grade
- Deadline is Week 13, Tue April 11th 2023 11:59pm
- Upload the deliveries zipped into the file name "reviewCS6208_your_name.zip" to Canvas – Assignments - Paper review delivery
- Penalty : You will lose 25% of the project grade every late day (except medical certificate).

Paper review

- Paper review
  - You can choose any paper to review related to graph machine learning
  - You may select one paper from my selection in Canvas-Files-Papers selection (50 papers)
- You are free to review the paper in any way you like, but it must be insightful.
- Example of review
  - What is the task?
  - What are the literature, state of the art, limitations?
  - What is the new proposed idea/contribution?
  - Is the new technique mathematically sound?
  - What are the datasets? Benchmark datasets?
  - What do experiments demonstrate?
  - What conclusion?
  - What can be improved next?

Paper review outcome

- Deliveries
  - Report : 1 latex-style pdf
  - Code with Pytorch, TensorFlow, DGL, PyG, etc
  - GitHub repository
  - Video presentation : 2min

- Zip everything and upload to Canvas – Assignments - Paper review delivery with file name "reviewCS6208_your_name.zip"
  - Deadline is Week 13, Tue April 11th 2023 11:59pm
  - Penalty : You will lose 25% of the review grade every late day (except medical certificate).

Report

- The report must be clear and concise.
- The length of the report is 1 latex-style page maximum.
  - The length is strict, no more than 1 page.
  - You will lose 50% of the grade if more than 1 page.
  - An additional page containing only cited references are allowed (i.e. references do not count in the one-page limit).
- Recommendation : You may use overleaf to write the report.

Code

- A code prototype must reproduce the proposed technique.
- The code may use GNN libraries like DGL or PyG.
- It does not have to be long.
- The code can use the same or different dataset(s) from the paper.
- However, the code cannot be the same code as an existing online code e.g. GitHub (plagiarism).
- Besides, the code cannot use a pre-defined function of the paper technique implemented by DGL, PyG, etc.
- It is important to reproduce the proposed technique as much as possible from scratch in order to demonstrate a good understanding of the paper.

Video presentation

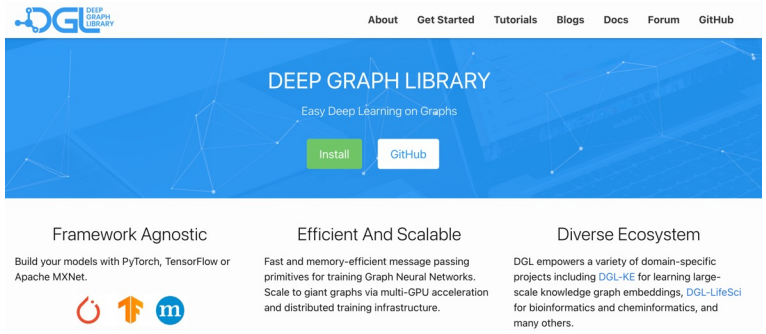
- The review presentation must be clear and concise.
- Use slides
- The length of the presentation is 2min maximum.
  - The time is strict, no more than 2min.
  - You will lose 50% of the grade if more than 2min.
- Recommendation : You may use zoom to record the video.

Marking scheme

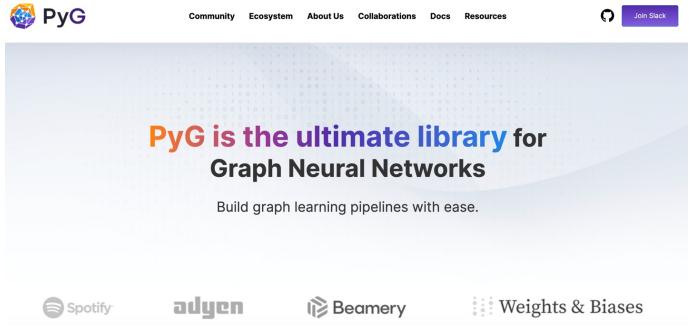
- Deliveries of report, code on GitHub, video presentation will receive 2/3 of the grade if well executed.
- Anything that demonstrates innovations, initiatives will count for the remaining 1/3 of the grade.

GNN libraries

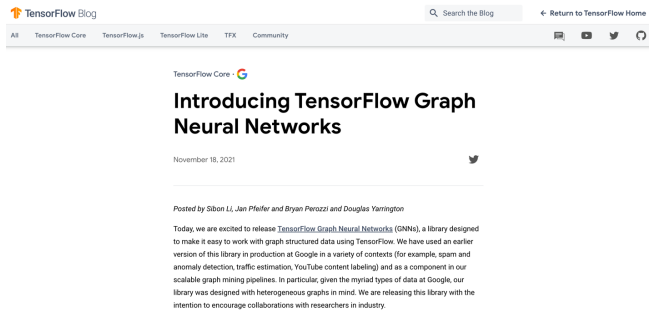
- Amazon DGL (Deep Graph Library)
  - First released on Dec 2018 (11k+ stars)
  - PyTorch / TensorFlow / MxNet
  - https://www.dgl.ai



- Kumo PyG (PyTorch Geometric)
  - First released on Mar 2019 (16k+ stars)
  - PyTorch
  - https://www.pyg.org



- TensorFlow GNN
  - First released on Nov 2021 (1k+ stars)
  - https://github.com/tensorflow/gnn



GPU

- The project is not about running long experiments with the best possible GPUs.
- Google Colab (free GPU with limitations), Google Cloud (600hr of free GPU), and your computer/laptop are enough.
  - Read lecture "google_gpu.pdf" for Google Cloud GPU setup.

Questions

- If you have any additional questions about the paper review
  - General question : Send me an email and I will answer to everyone at the next lecture
  - Specific question : Use office hour on Tuesdays

Evaluation

- This module is 100% CA, i.e., there is no final exam.
- There are 2 components
  - Individual assessment (50%) : 1 paper review
  - Group assessment (50%) : 1 project

Project

- Project group counts for 50% of the final grade
- Deadline is Week 14, Tue April 18th 2023 11:59pm
- Upload the deliveries zipped into the file name "projectCS6208_name1_name2.zip" to Canvas – Assignments – Group project delivery
- Penalty : You will lose 25% of the project grade every late day (except medical certificate).

Group selection

- Group can be composed of 1 or 2 or 3 members, preferably 2-3 but not more than 3.
  - Use Canvas – Discussions – Project group if you need to find groupmate(s).
- Choose your group wisely
  - Each teammate must contribute *equally* to the project. Clearly, each member has different skills and it is fine to be weaker in maths, coding, etc. However, it is not fine to let the other teammates to do most of the work.
- Minimize conflict and make a short written "contract" at the beginning of the project regarding the commitment of each team member to the project.
  - Upload the "contract" to Canvas – Assignments – Group project contract with the file name "projectCS6208_contract_name1_name2.pdf"
  - Deadline is Week 8, Tue March 7th 2023 11:59pm.
  - Penalty : You will lose 25% of the project grade every late day (except medical certificate).
  - It is fine that commitments change during the project development but the contributions must remain equally distributed.

General project goals

- Project focuses on :
  - Theoretical knowledge received in this course
  - Practical skills with data acquisition, exploration, exploitation, analysis
  - Teamwork with task management
  - Concise and clear communication with report and oral presentation

Specific project goals

- GML task : Identify a task for graph-structured data
- Dataset(s)
  - You may use existing (benchmark) graph dataset(s)
- Baseline(s)
  - Implement baseline GNN(s) on this dataset(s). You may use DGL, PyG, etc
  - Propose one improvement
    - Motivation, description, equation, implementation, result, discussion.
- Demonstrate innovations, initiatives
  - Propose a new dataset, develop a scrapper, dynamic visualization, design new GNN layer, discover unexpected data insights, etc.

Project Philosophy

- This project focuses on
  - The understanding of the fundamental concepts of GML techniques,
  - The practical skills required to develop a data analysis project.
- It is not about using existing GitHub codes.
- It is not about winning a Kaggle competition.
- It is not about ten lines of Keras' code to run deep learning techniques.
- It is not about running long experiments with the best possible GPUs.
  - Google Colab, Google Cloud, and your computer/laptop are enough.
- It is not about getting 90% of accuracy.
- It is about how to design from scratch, debug, understand and train graph neural networks.
- It is about to understand why it works and why it does not.

Project outcome

- Deliveries
  - Report : 1-2 latex-style pdf
  - Code with Pytorch, TensorFlow, DGL, PyG, etc
  - GitHub repository
  - Video presentation : 5min

- Zip everything and upload to Canvas – Assignments – Group project delivery with the file name "projectCS6208_name1_name2.zip"
  - Deadline is Week 14, Tue April 18th 2023 11:59pm
  - Penalty : You will lose 25% of the project grade every late day (except medical certificate).

Report

- The report must be clear and concise.
  - Project motivation and description, data acquisition, data exploration, pre-processing, proposed GML solutions, analysis of results, future development, etc
- Make a section regarding the contribution of each team member.
- The length of the report is 2 latex-style pages maximum.
  - The length is strict, no more than 2 pages.
  - You will lose 50% of the grade if more than 2 pages.
  - An additional page containing only cited references are allowed (i.e. references do not count in the one-page limit).
- Recommendation : You may use overleaf to write the report.

Video presentation

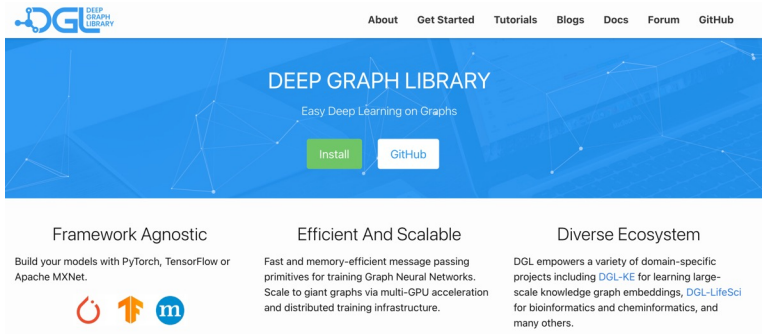
- The project presentation must be clear and concise.
  - Project motivation and description, data acquisition, data exploration, pre-processing, proposed GML solutions, analysis of results, future development.
- Each person must present her/his contribution to the project.
  - You will receive grade zero if you do not present your contribution.
- Use slides.
- The length of the presentation is 5min maximum.
  - The time is strict, no more than 5min.
  - You will lose 50% of the grade if more than 5min.
- Recommendation : You may use zoom to record the video.

Marking scheme

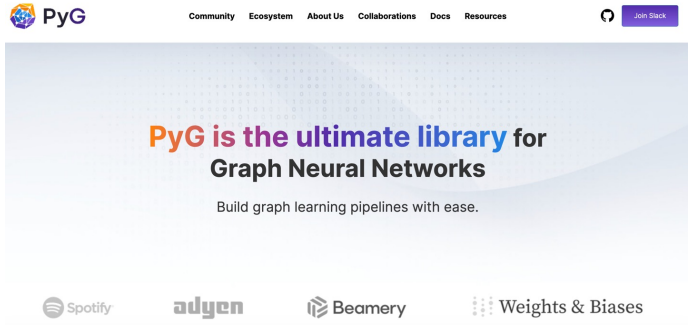
- Deliveries of report, code on GitHub, video presentation will receive 2/3 of the grade if well executed.
- Anything that demonstrates innovations, initiatives will count for the remaining 1/3 of the grade.

GNN libraries

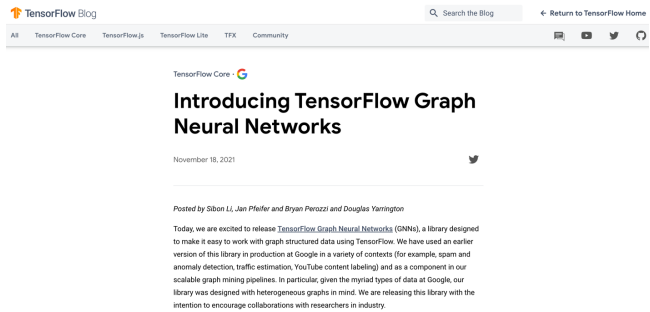
- Amazon DGL (Deep Graph Library)
  - First released on Dec 2018 (11k+ stars)
  - PyTorch / TensorFlow / MxNet
  - https://www.dgl.ai



- Kumo PyG (PyTorch Geometric)
  - First released on Mar 2019 (16k+ stars)
  - PyTorch
  - https://www.pyg.org



- TensorFlow GNN
  - First released on Nov 2021 (1k+ stars)
  - https://github.com/tensorflow/gnn



GPU

- The project is not about running long experiments with the best possible GPUs.
- Google Colab (free GPU with limitations), Google Cloud (600hr of free GPU), and your computer/laptop are enough.
  - Read lecture "google_gpu.pdf" for Google Cloud GPU setup.

Review paper and group project

- The topics of the review paper and the group project must be different.
  - For example, you cannot review a paper on molecular generation and develop a project on molecular generation. But you you can review a paper on graph recommendation and develop a project on molecular generation.
  - There must be significant differences between the review paper and the group project.

Questions

- If you have any additional questions about the project
  - General question : Send me an email and I will answer to everyone at the next lecture
  - Specific question : Use office hour on Tuesdays

Outline

- Week 7
- Office hours
- Course evaluation
  - Paper review
  - Group project
- Today lecture

Outline

- This course focuses on the foundations of graph machine learning (GML).
- The course has three main parts

  Part 1 : GML without feature learning (before 2014)

  Part 2 : GML with shallow feature learning (2014-2016)

  Part 3 : GML with deep feature learning, a.k.a. Graph Neural Networks (after 2016)

Tentative Lectures

- Introduction to Graph Deep Learning
- Part 1: GML without feature learning (before 2014)
  - Introduction to Graph Science
  - Graph Analysis Techniques without Feature Learning
    - Graph clustering
    - Classification
    - Recommendation
    - Dimensionality reduction
- Part 2 : GML with shallow feature learning (2014-2016)
  - Shallow graph feature learning
- Part 3 : GML with deep feature learning, a.k.a. GNNs (after 2016)
  - Graph Convolutional Networks (spectral and spatial)
  - Benchmarking GNNs
  - Graph Positional Encoding
  - Graph ViT/MLP-Mixer
  - Generative GNNs and molecular science
  - Combinatorial optimization
  - GNNs for Recommendation
  - GNNs for knowledge graphs
  - Theory of GNNs

Questions?