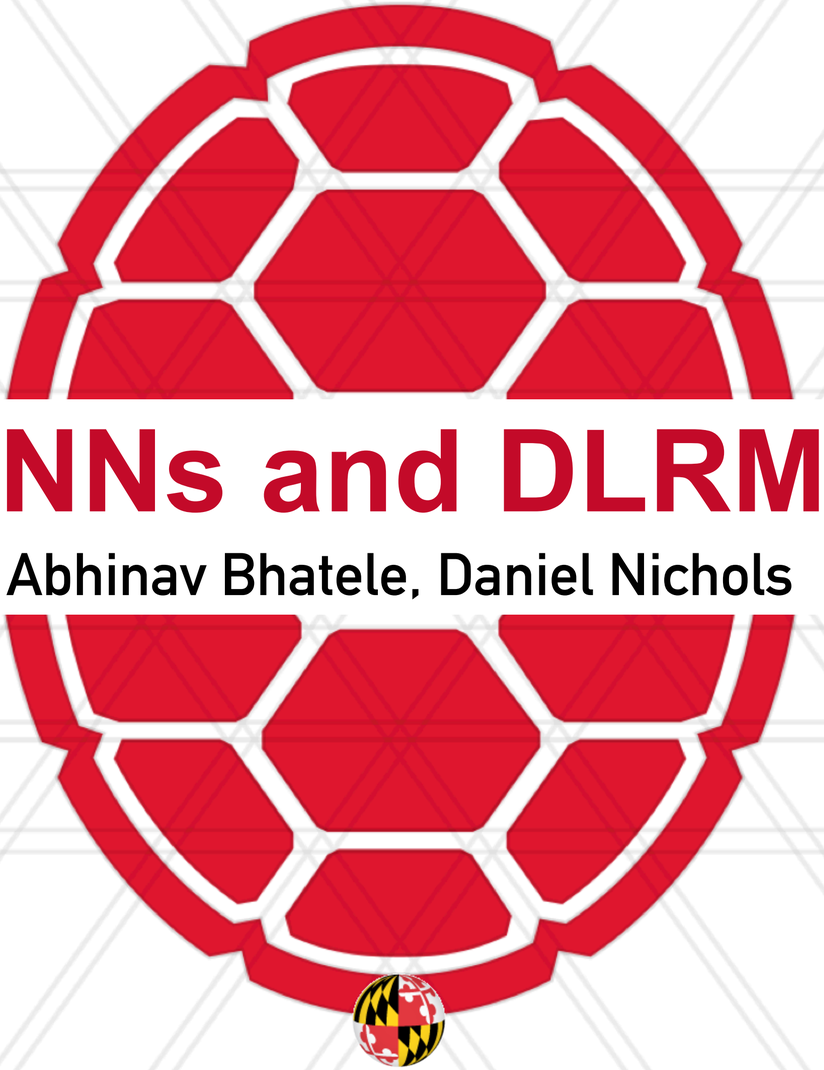
Systems for Machine Learning (CMSC828G)



GNNs and DLRMs

Abhinav Bhatele, Daniel Nichols

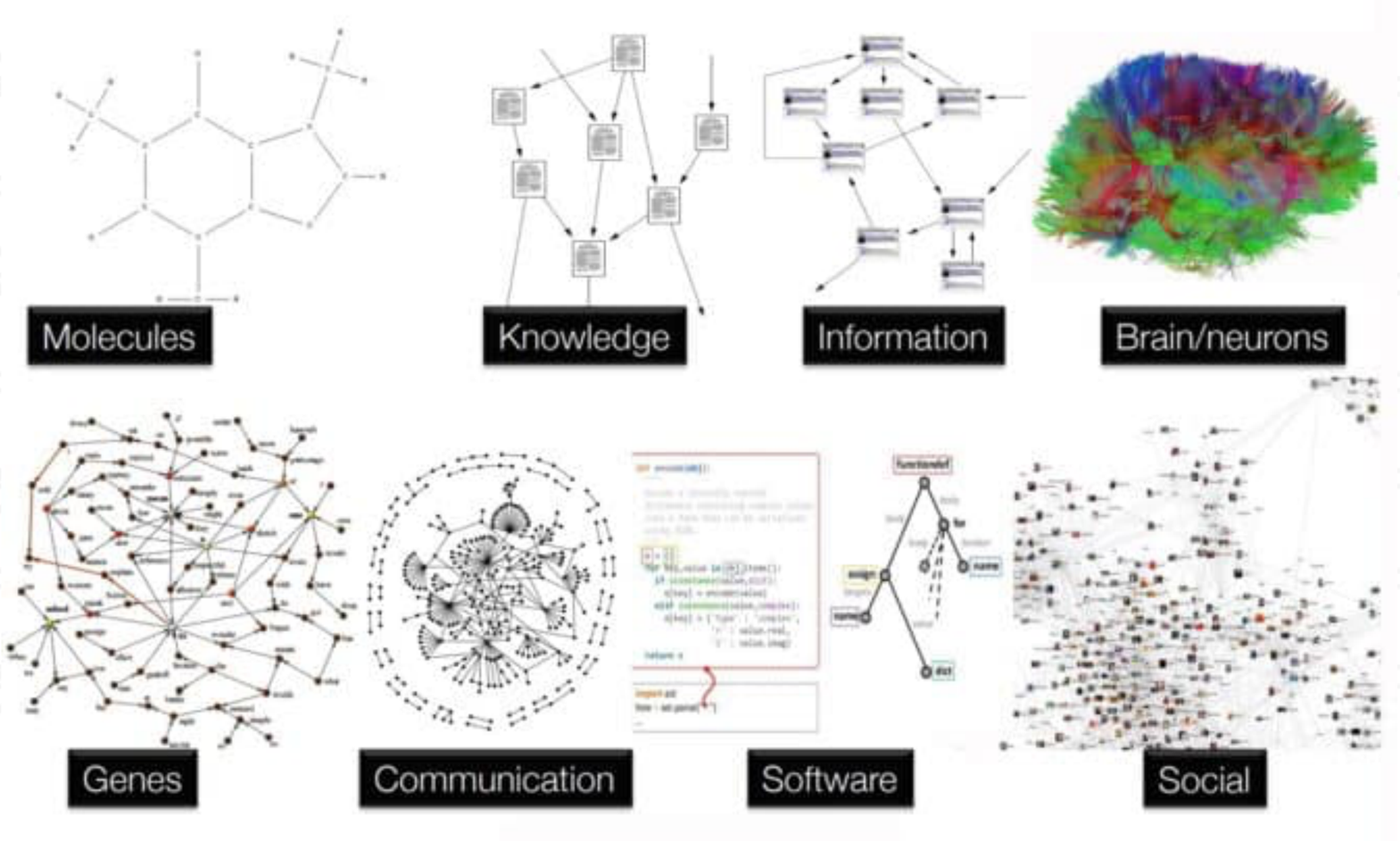
UNIVERSITY OF MARYLAND

Annoucements

- Students who haven't presented in class: submit slides to the same grade scope assignment with a YouTube or Google Drive link to the video
  - Due date: May 1
- Extra credit is due on May 7
  - No extensions, you cannot use the late penalty for this assignment
  - It is optional

Graph Neural Networks

- Drug discovery
- Fraud detection
- Route optimization
- Recommendation systems



https://blogs.nvidia.com/blog/what-are-graph-neural-networks/

Tasks on graphs

- Node-level: predicting the class of individual nodes
- Edge-level: determine if an edge exist or predict the type of edge
- Graph-level: even categorize entire graphs

Graph Convolution Network (GCN) layer

- Aggregation

$$\mathbf{H}^{Li} = SpMM(\mathbf{A}, \mathbf{F}^{Li})$$

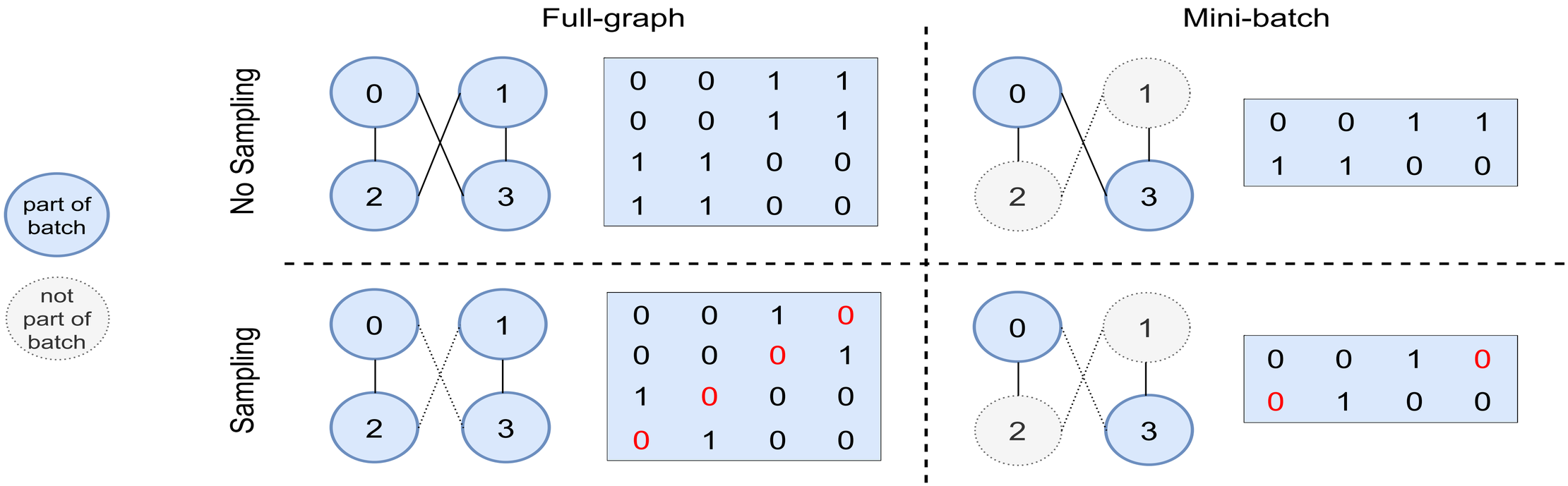
- Combination

$$\mathbf{Q}^{Li} = SGEMM(\mathbf{H}^{Li}, \mathbf{W}^{Li})$$

- Activation

$$\mathbf{F}^{Li+1} = \sigma(\mathbf{Q}^{Li})$$

Different approximations



Challenges with Parallel GNN training

- Graphs are large and irregular
- Significant load imbalance
- Sparse kernels are slow on GPUs
- High communication overhead

UNIVERSITY OF
MARYLAND