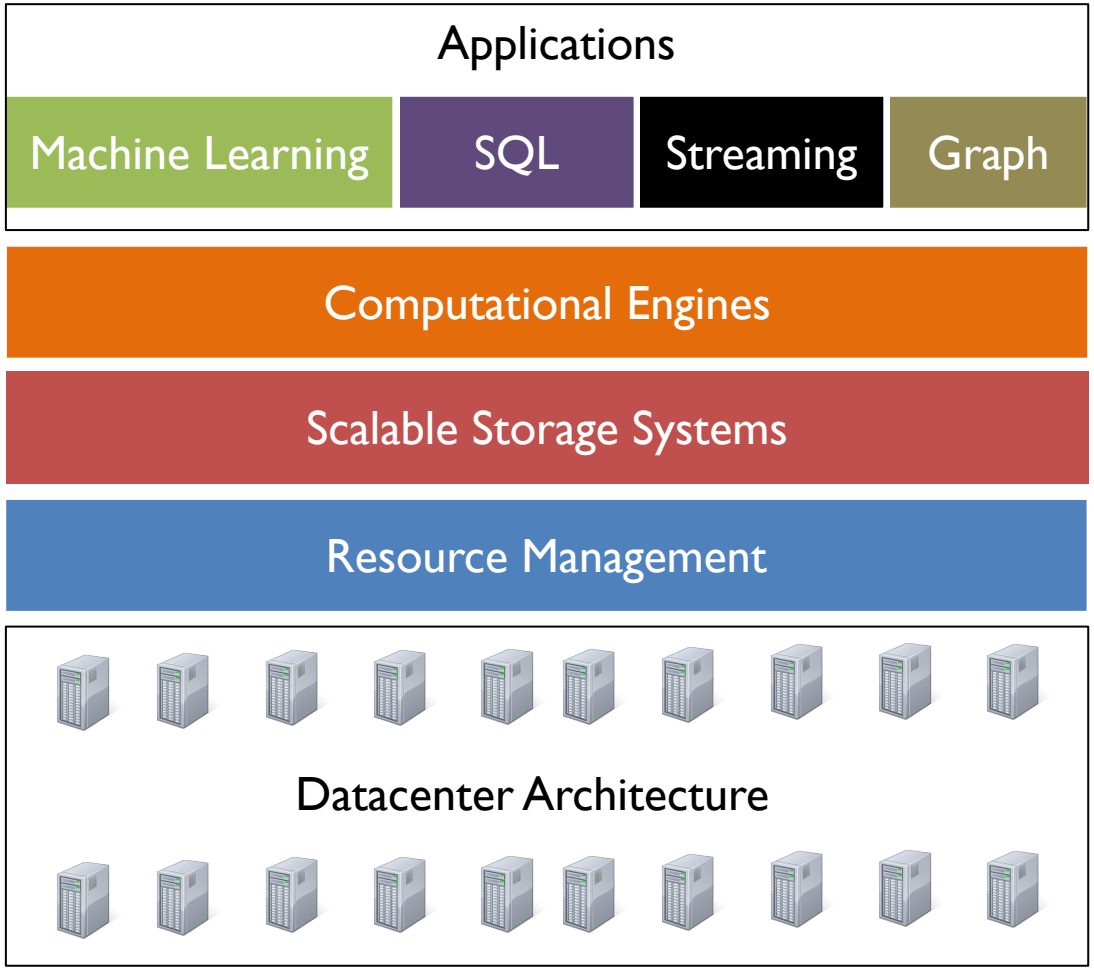
CS 744: PYTORCH-BIGGRAPH

Shivaram Venkataraman

Fall 2020

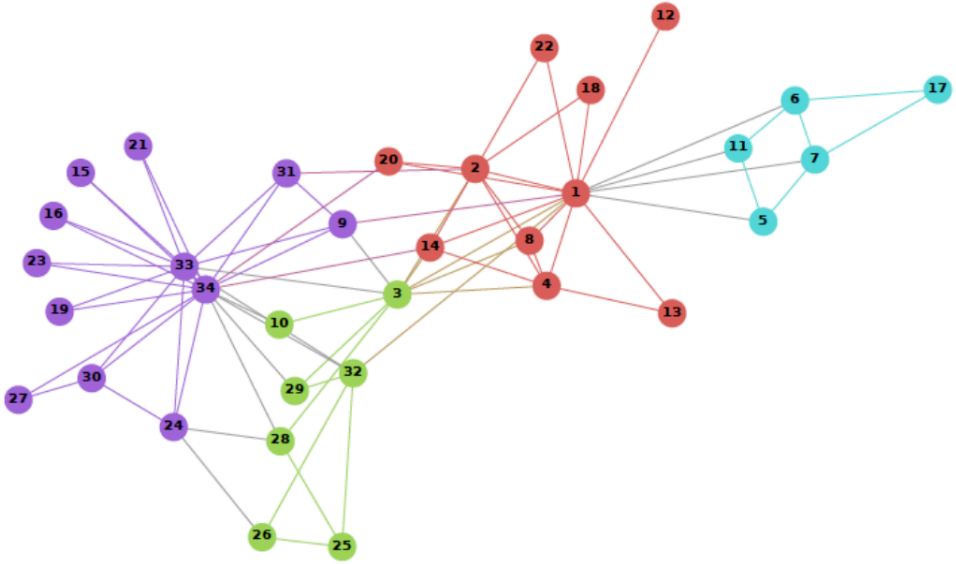
ADMINISTRIVIA

Course Project: Check in by Nov 20th

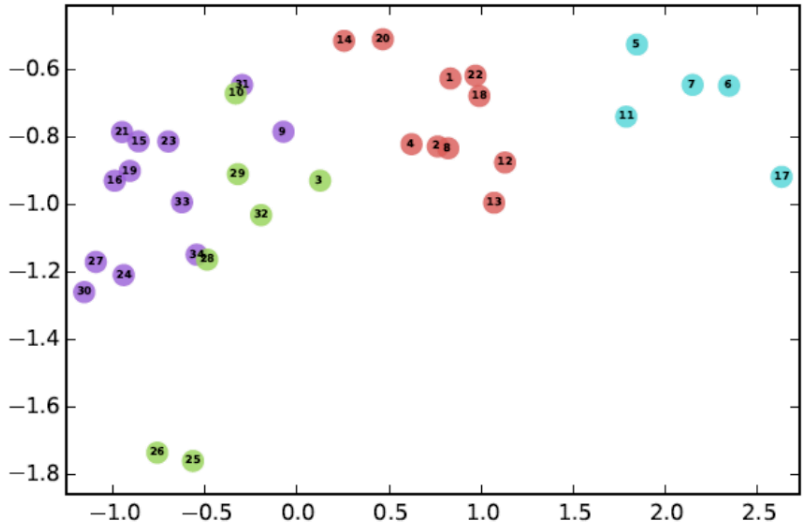

Applications
Machine Learning
SQL
Streaming
Graph
Computational Engines
Scalable Storage Systems
Resource Management
Datacenter Architecture

GRAPH EMBEDDINGS

A



B



TRAINING GRAPH EMBEDDINGS

Vertex	Embedding
12	<0.0, 0.25>
14	<0.33, 0.5>
16	<0.45, 0.6>
…	…

Source	Dest
12	14
12	18
16	14
…	…

$$f(\theta_s, \theta_r, \theta_d) = sim\left(g_{(s)}(\theta_s, \theta_r), g_{(d)}(\theta_d, \theta_r)\right)$$

NEGATIVE SAMPLING

Sample from edges not in the graph!

Two options

1. According to data distribution

$$\mathcal{L} = \sum_{e \in G} \sum_{e' \in S'_e} \max(f(e) - f(e') + \lambda, 0))$$

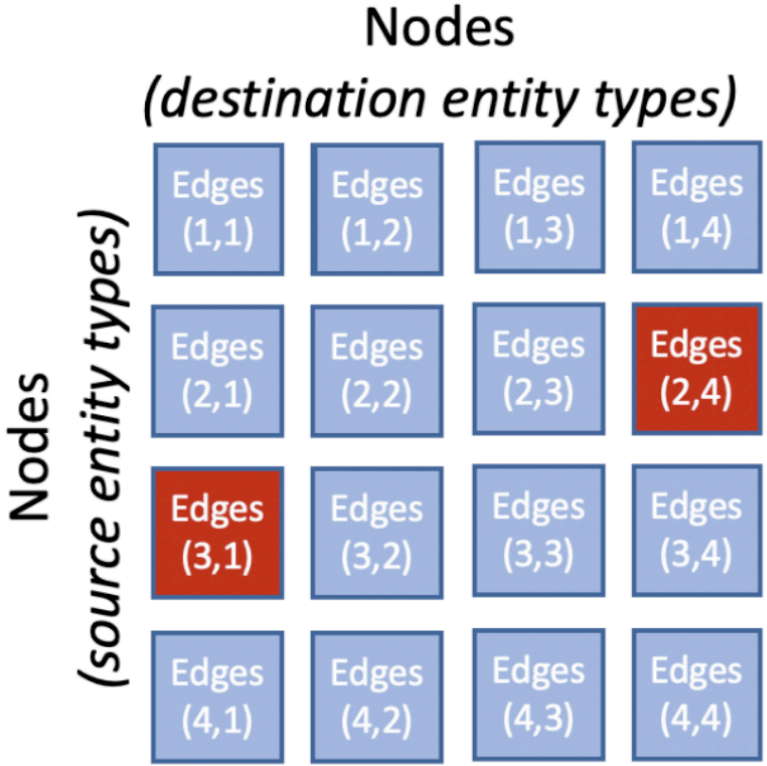
2. Uniformly

SCALING CHALLENGES

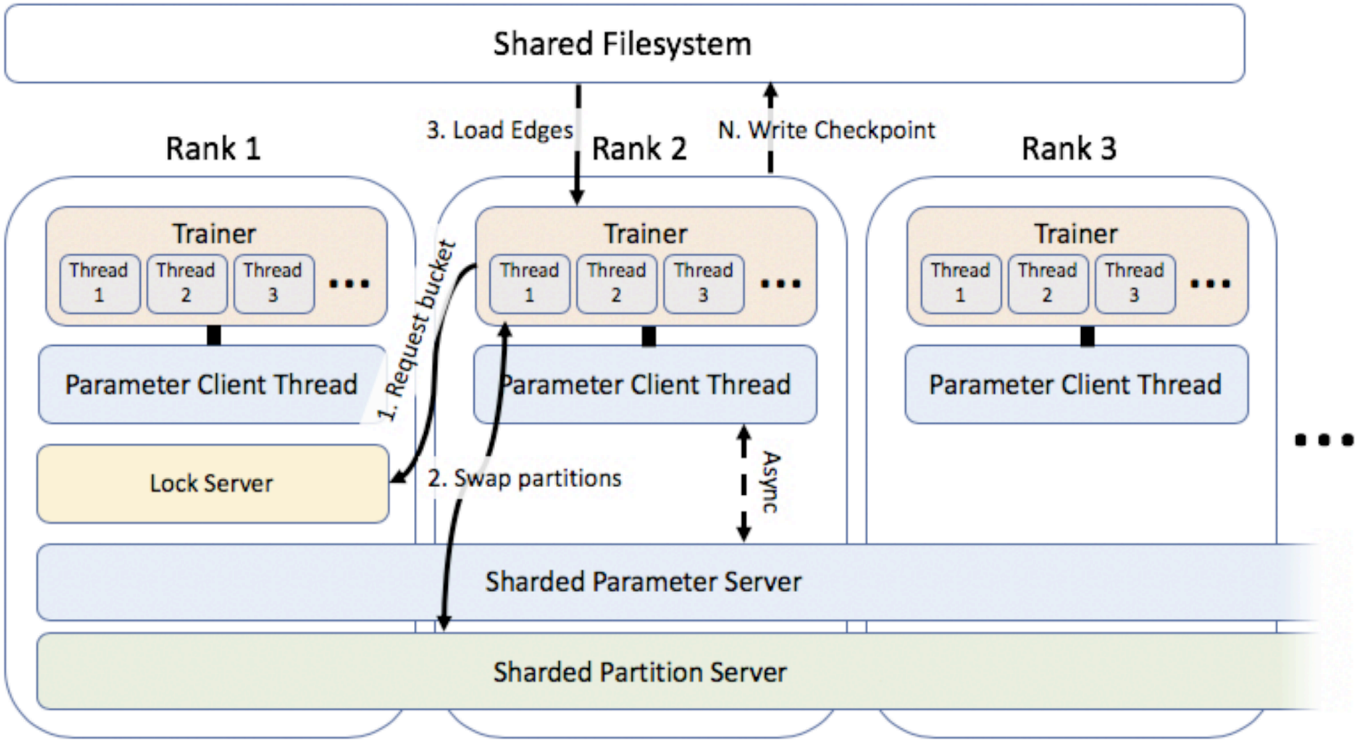
Fast enough to embed graphs with 10^{11}–10^{12} edges in a reasonable time

~100 embedding parameters per node → require 800GB of memory!

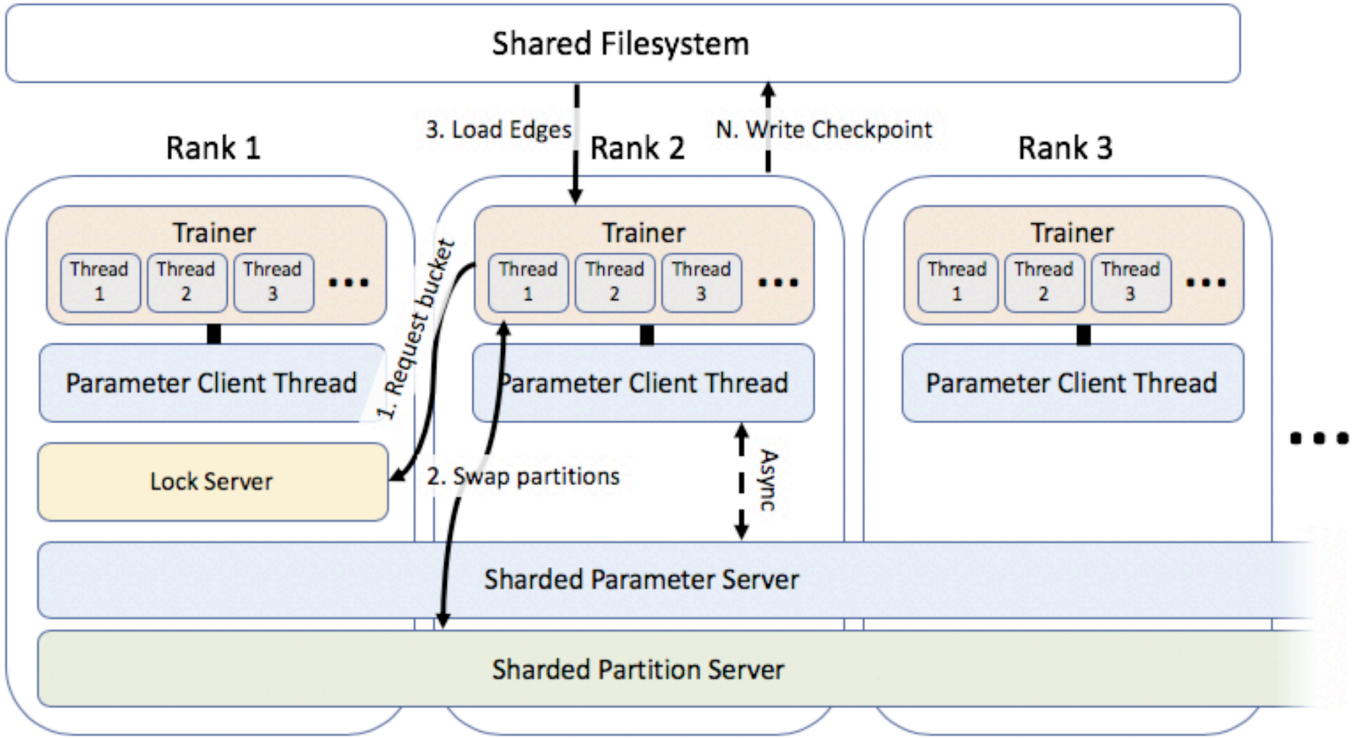
GRAPH PARTITIONING



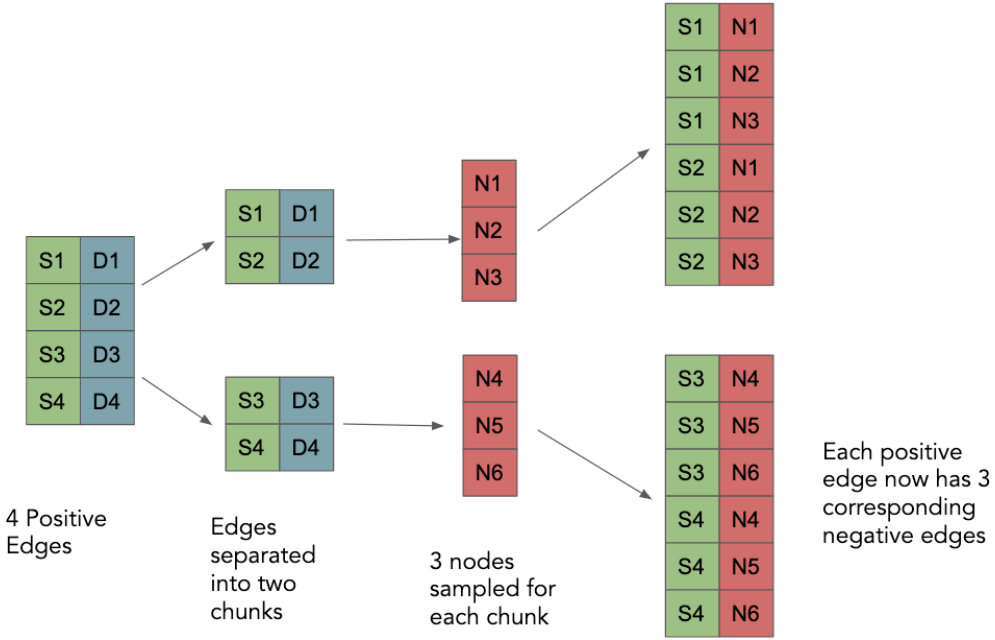
SYSTEM DESIGN



DISTRIBUTED EXECUTION



BATCH NEGATIVE SAMPLING



SUMMARY

Graph Embeddings: Learn embeddings from graph data for ML

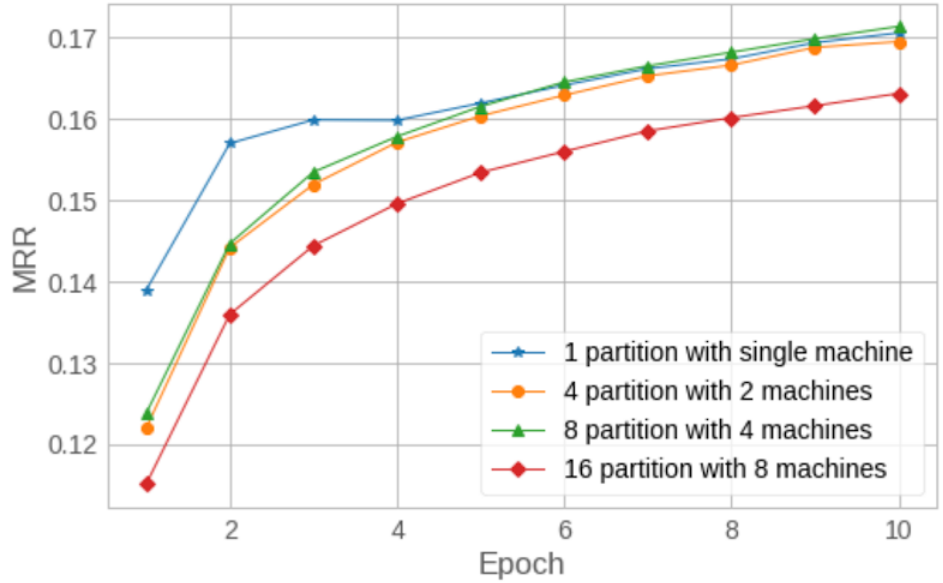
Partition graph into buckets for scalability

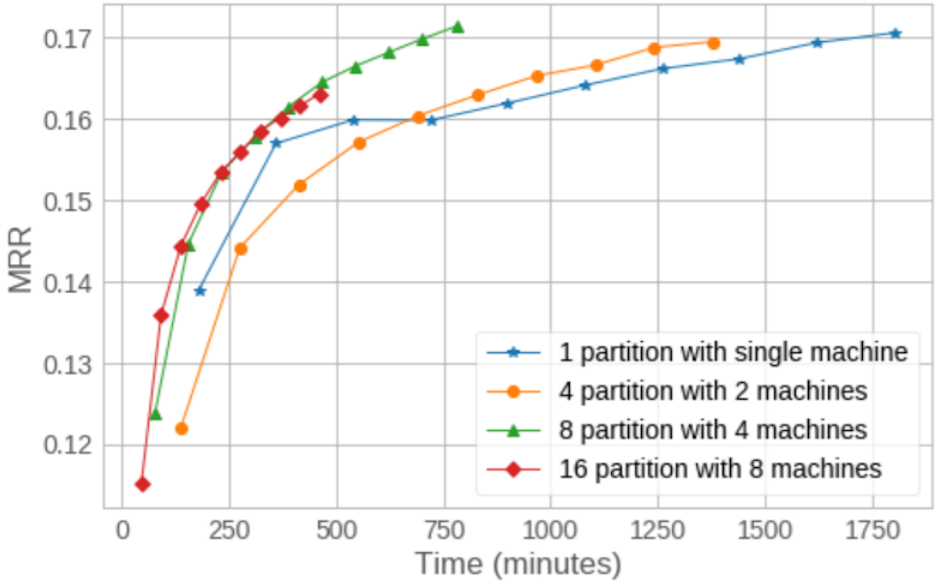
Distributed execution with shared partition server

Batched negative sampling

DISCUSSION

https://forms.gle/YEtECXCUtksoL6Sr8


MRR
0.17
0.16
0.15
0.14
0.13
0.12
2
4
6
8
10
Epoch
1 partition with single machine
4 partition with 2 machines
8 partition with 4 machines
16 partition with 8 machines


MRR
0.17
0.16
0.15
0.14
0.13
0.12
0
250
500
750
1000
1250
1500
1750
Time (minutes)
1 partition with single machine
4 partition with 2 machines
8 partition with 4 machines
16 partition with 8 machines

How does the partitioning scheme used in this paper differ from partitioning used in PowerGraph and why? (from review)

NEXT STEPS

Next class: New module!

Project check-ins by Nov 20th