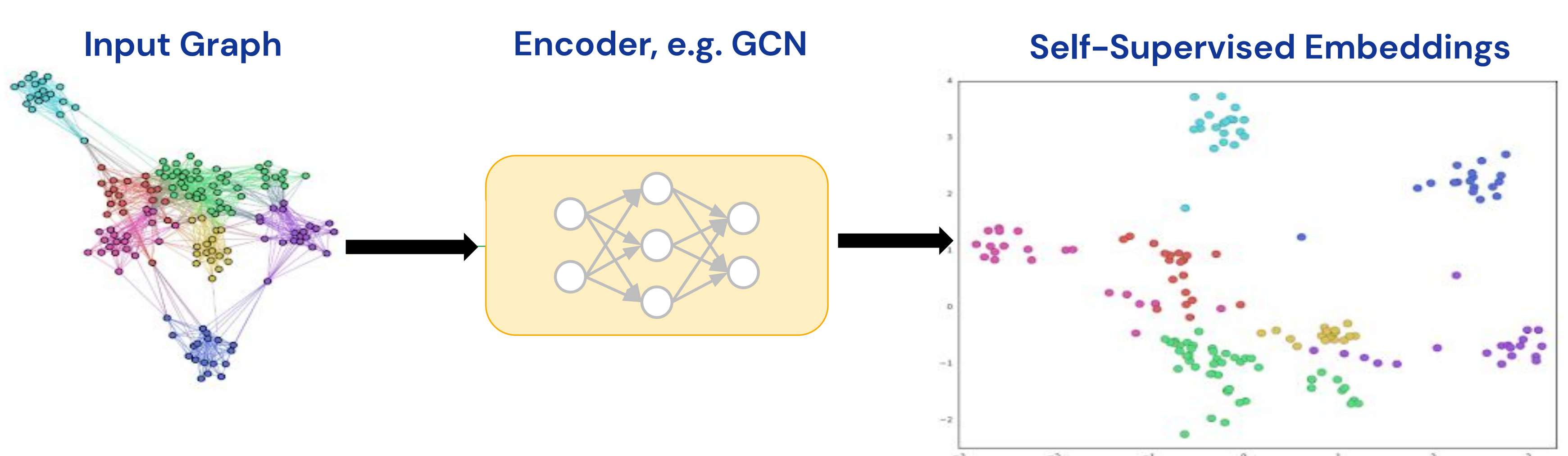
DeepMind

Large-Scale Representation Learning on Graphs via Bootstrapping

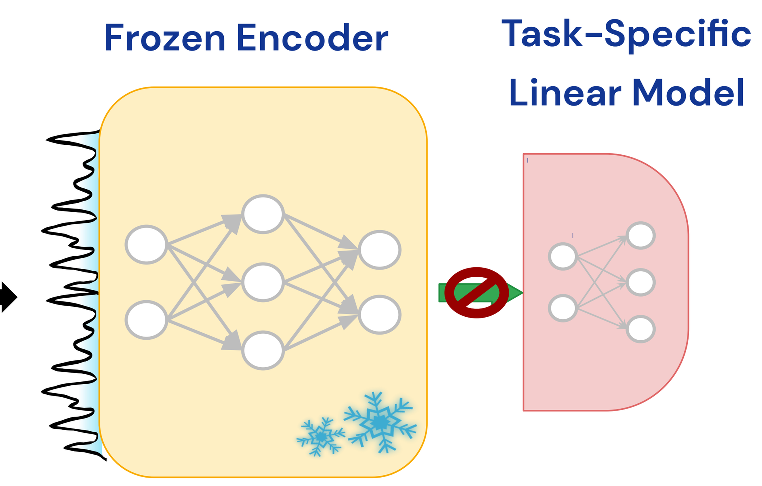
Shantanu Thakoor, Corentin Tallec, Mohammad Gheshlaghi Azar, Mehdi Azabou,
Eva L Dyer, Remi Munos, Petar Veličković, Michal Valko

Self-Supervised Node Representation Learning



Unlabeled data widely available in graphs domain, procuring labels is costly
⇒ Self-supervised learning trains a representation *without labels*

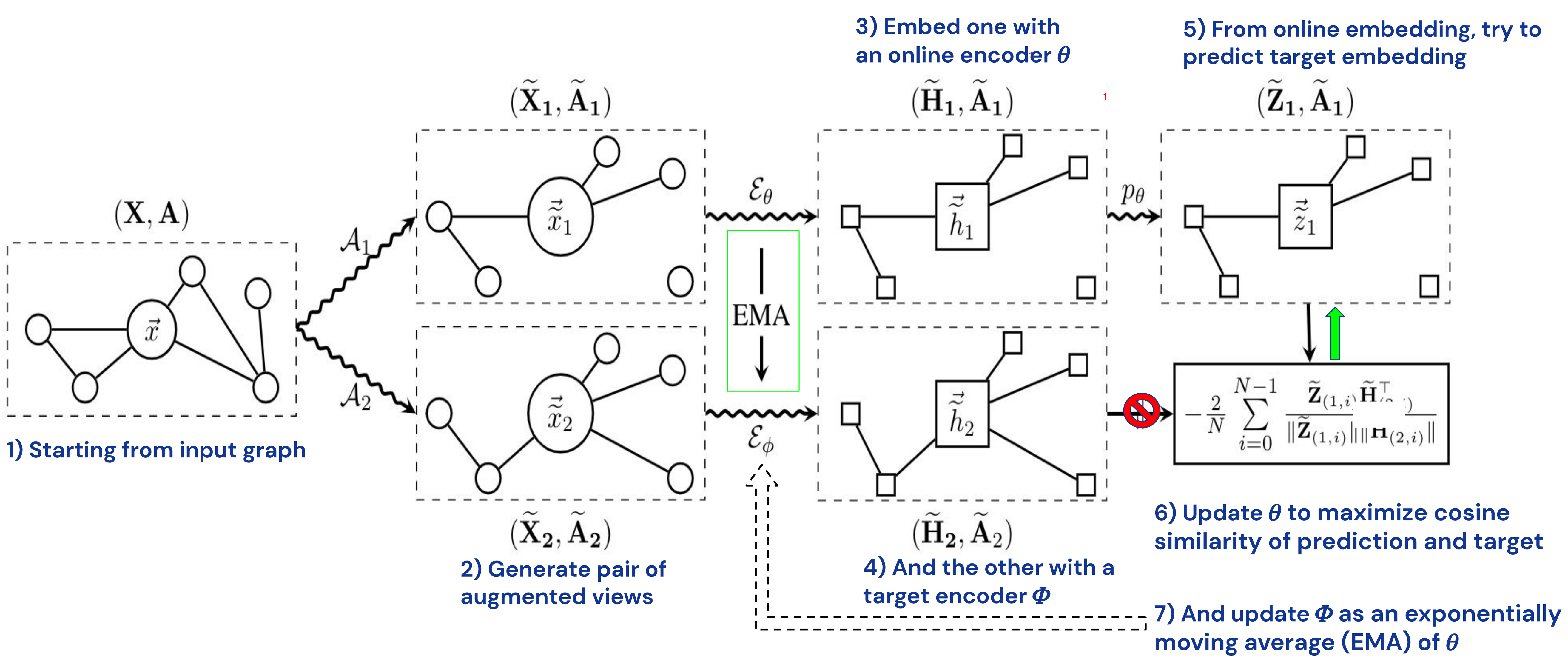
Evaluation of Frozen Embeddings



1) Pretrain representation with unlabeled data

2) On top of frozen features, train simpler model making use of small amount of labelled data

Bootstrapped Graph Latents (BGRL)



Key Advantages:

- No need to define negative examples – particularly hard in graphs domain!
- Computation scales *linearly* – as opposed to *quadratic* all-vs-all contrastive methods
  ⇒ Easily applicable to very large graphs that do not fit in memory!

Experimental Results

Compare under frozen linear evaluation protocol on standard benchmarks against

- DGI (requires defining negative examples)
- GRACE (quadratic all-vs-all contrastive)

★ BGRL matches/exceeds state of the art without negative examples with 5-10x memory savings

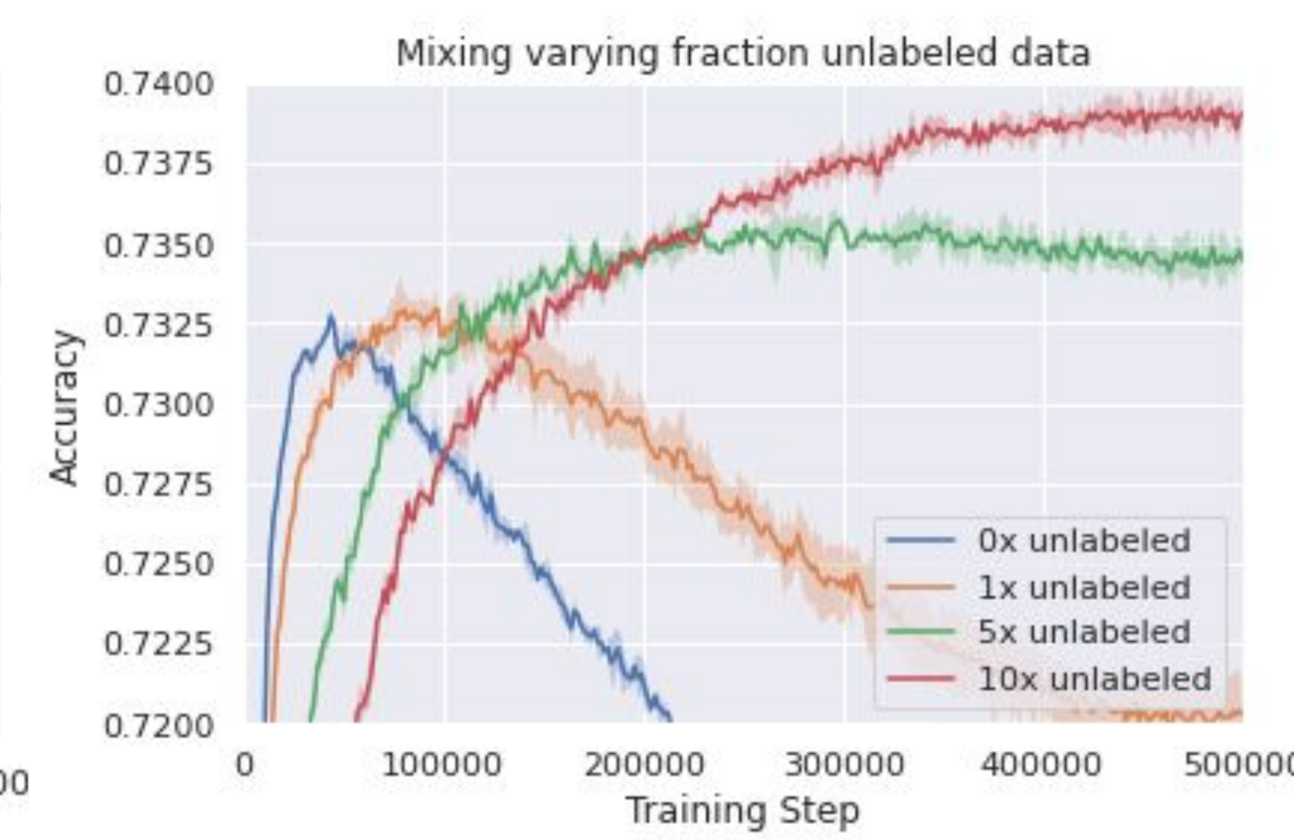
Dataset	Amazon Photos	WikiCS	Amazon Computers	Coauthor CS	Coauthor Phy
#Nodes	7,650	11,701	13,752	18,333	34,493
#Edges	119,081	216,123	245,861	81,894	247,962
DGI accuracy	91.61 ± 0.22	75.35 ± 0.14	83.95 ± 0.47	92.15 ± 0.63	94.51 ± 0.52
GRACE accuracy	92.78 ± 0.45	$\mathbf{80.14 \pm 0.48}$	89.53 ± 0.35	91.12 ± 0.20	OOM
BGRL accuracy	$\mathbf{93.17 \pm 0.30}$	79.98 ± 0.10	$\mathbf{90.34 \pm 0.19}$	$\mathbf{93.31 \pm 0.13}$	$\mathbf{95.73 \pm 0.05}$
GRACE Memory	1.81 GB	3.82 GB	5.14 GB	11.78 GB	OOM
BGRL Memory	**0.47 GB**	**0.63 GB**	**0.58 GB**	**2.86 GB**	**5.50 GB**

Achieved 2nd place on OGB-LSC MAG240M challenge at KDD Cup 2021

- Extremely large-scale (240 million nodes, 1 billion edges)
- Train from subsampled graph neighborhoods, using complex message-passing encoder networks
- Semi-supervised learning setting, using labels to shape representations: 1% of nodes are of interest for classification, other 99% used for self-supervision

★ BGRL provides useful auxiliary signal to shape representations in conjunction with data

★ BGRL leverages vast amounts of unlabeled data to prevent overfitting and achieve state-of-the-art results



Graph Image from https://ai.googleblog.com/2019/06/innovations-in-graph-representation.html