Unsupervised Clustering using Pseudo-semi-supervised Learning

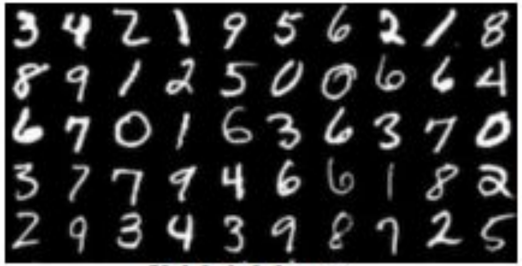
Divam Gupta, *Ramachandran Ramjee, Nipun Kwatra, Muthian Sivathanu*

Carnegie Mellon University

Microsoft Research

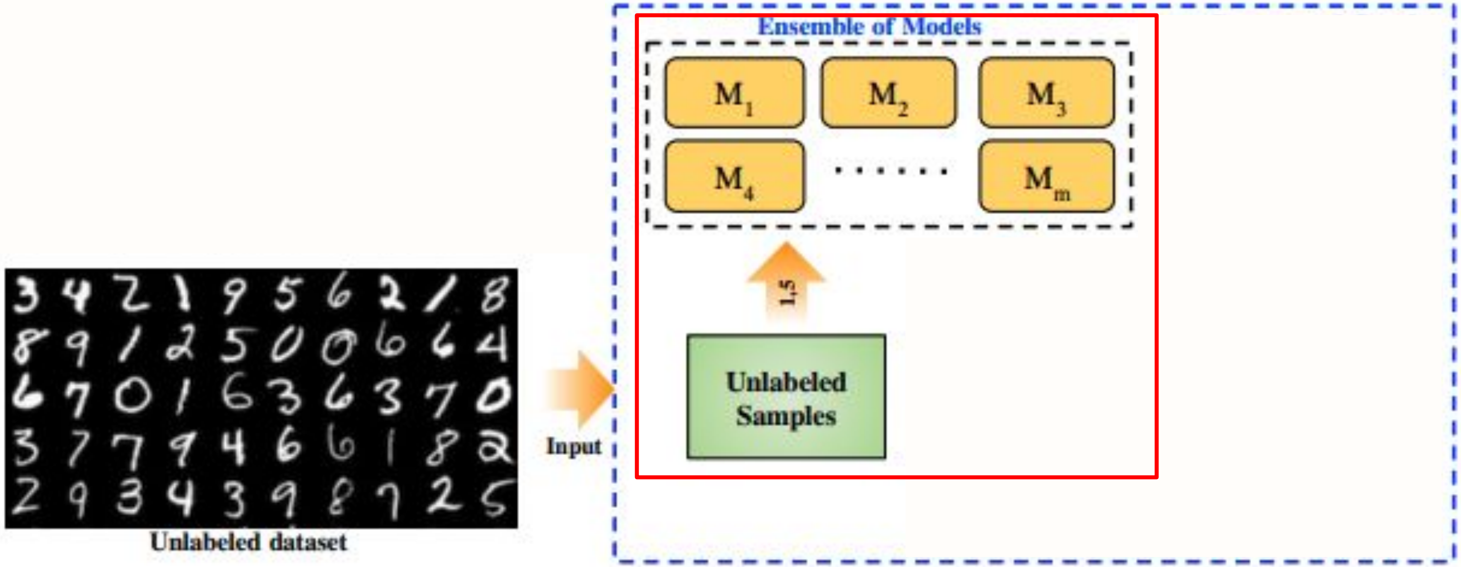
Introduction

- Leverage semi-supervised models to improve unsupervised clustering performance
- In place of labels, generate pseudo labels in an unsupervised manner
- Novel technique leveraging ensembles to generate pseudo labels
- Our method outperforms state of the art clustering results for multiple image and text datasets



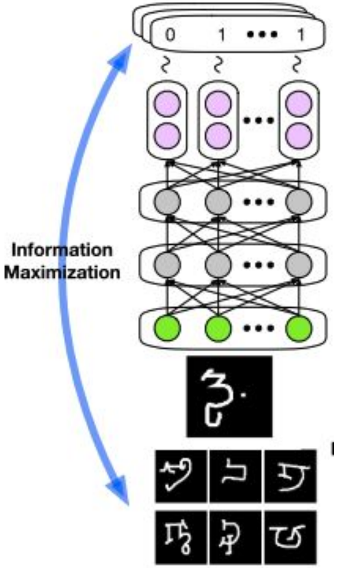
Unlabeled dataset

We start with an unlabeled dataset



Step 1

Train all the individual models on the unlabeled dataset using the unsupervised clustering loss



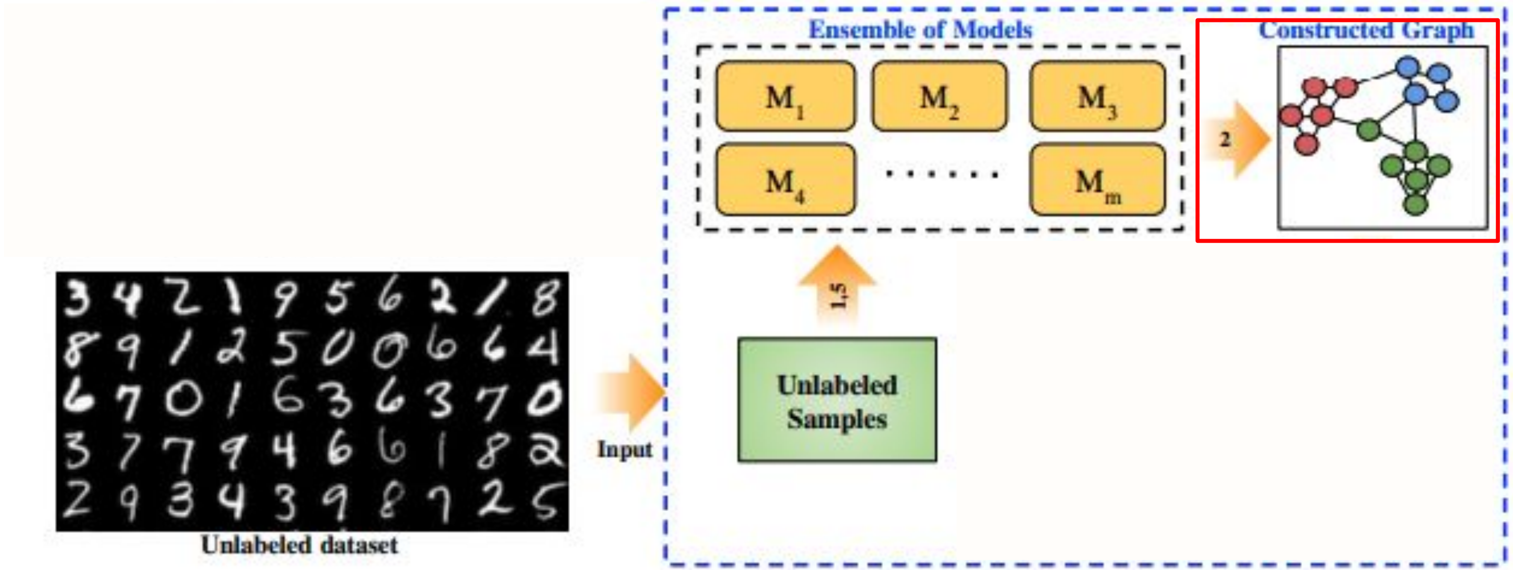
1) Information maximization loss *Hu et al. (2017)*

$$I(X;Y) = H(Y) - H(Y|X)$$

2) Dot product loss *Chang et al. (2017)*

$$D(X_i, X_j) = Y_i^T Y_j, \text{ if } i \neq j$$

All models of the ensemble have same architecture but random initialization with different seeds

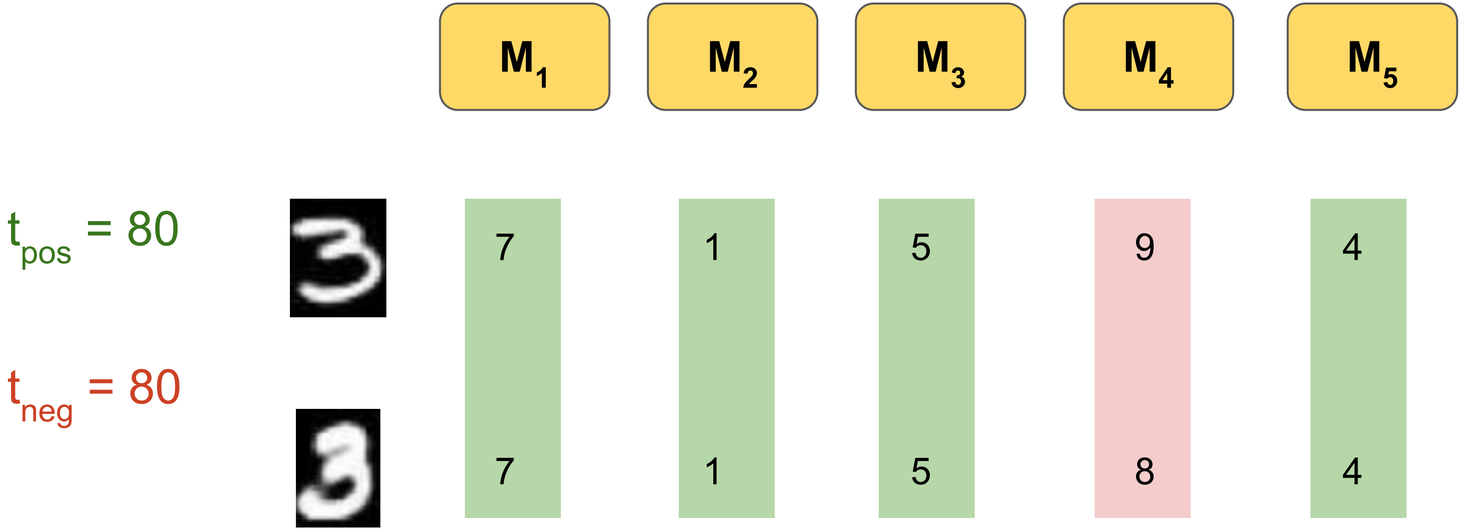


Step 2

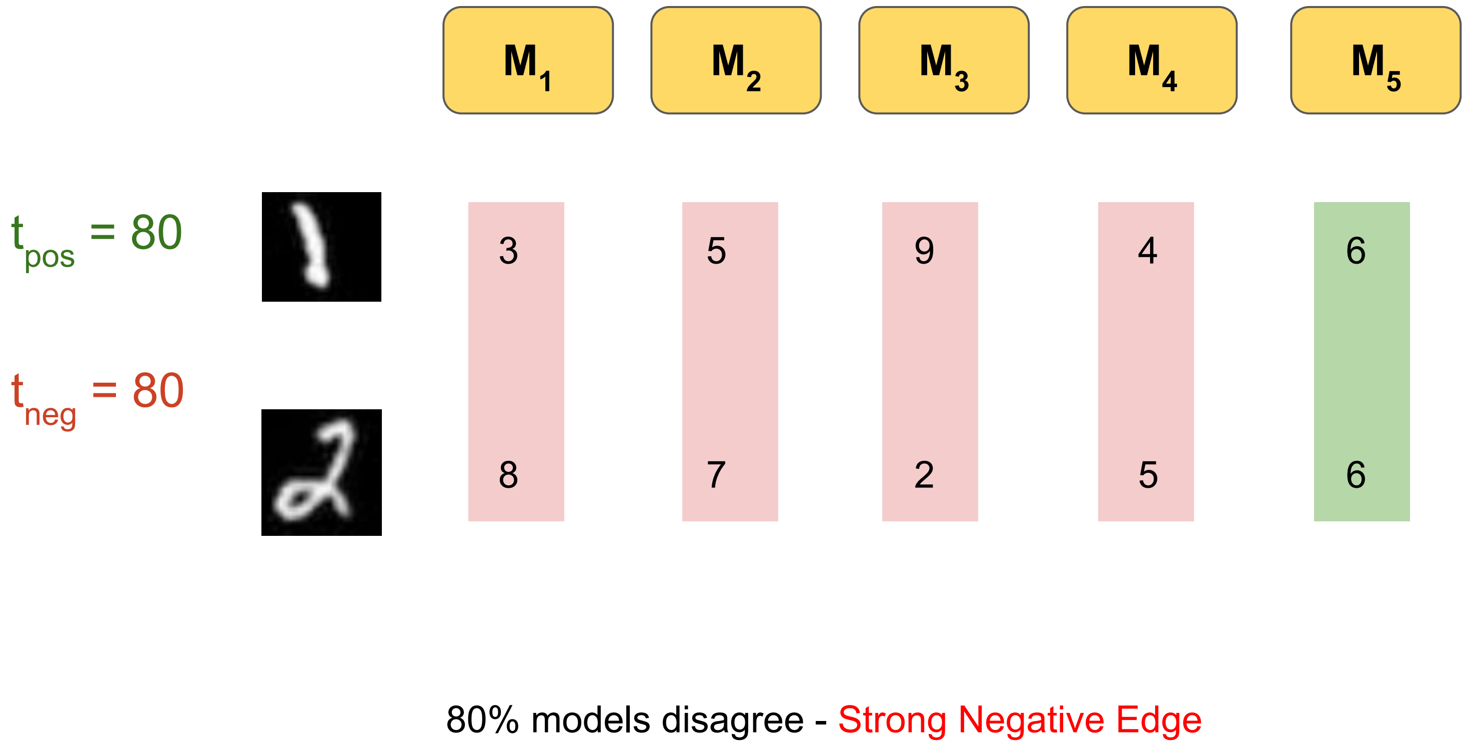
Create pairwise similarity graph

Generating pairwise similarity graph

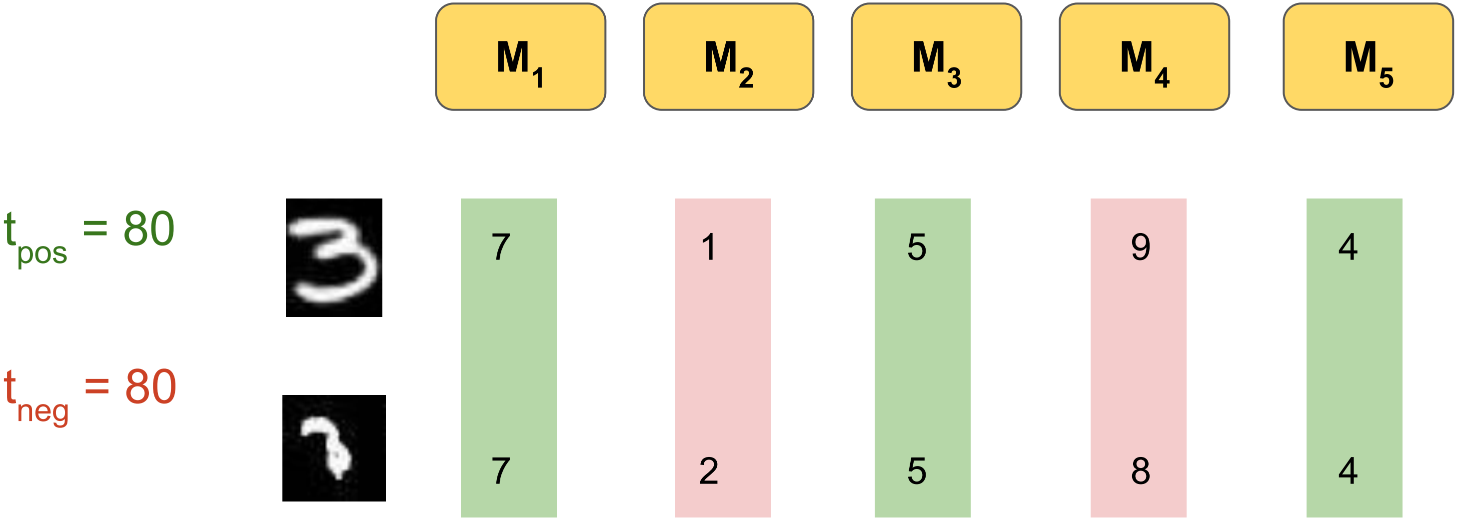
- Every sample in the dataset is a node in the graph
- Add strong positive edge for samples where more than t_{pos} models agree.
- Add strong negative edge for samples where more than t_{neg} models disagree.



80% models agree - Strong Positive Edge

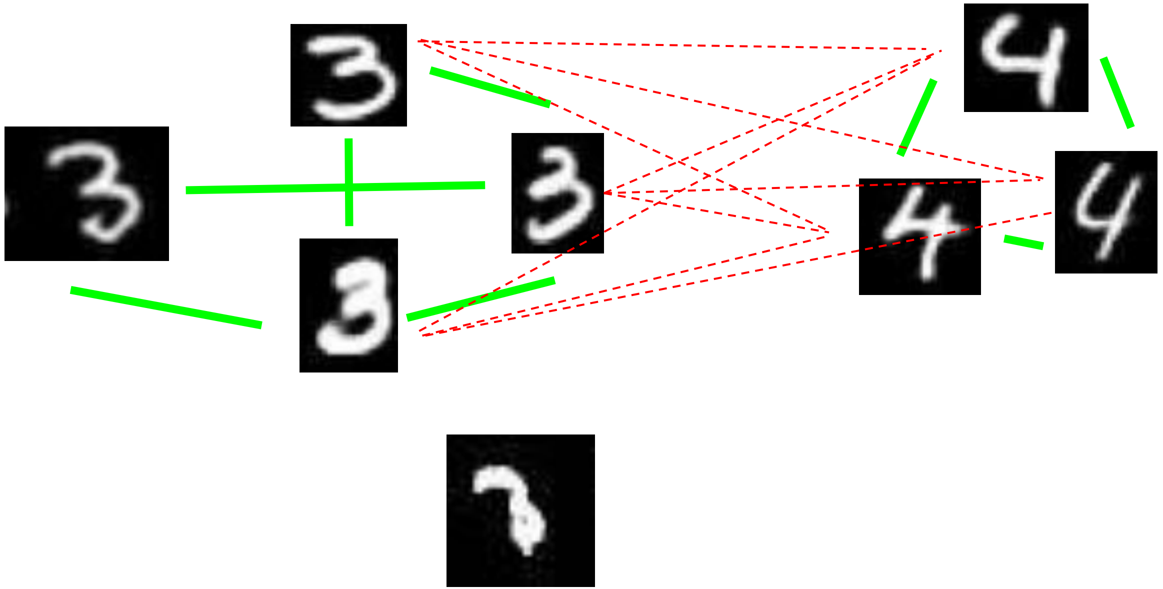


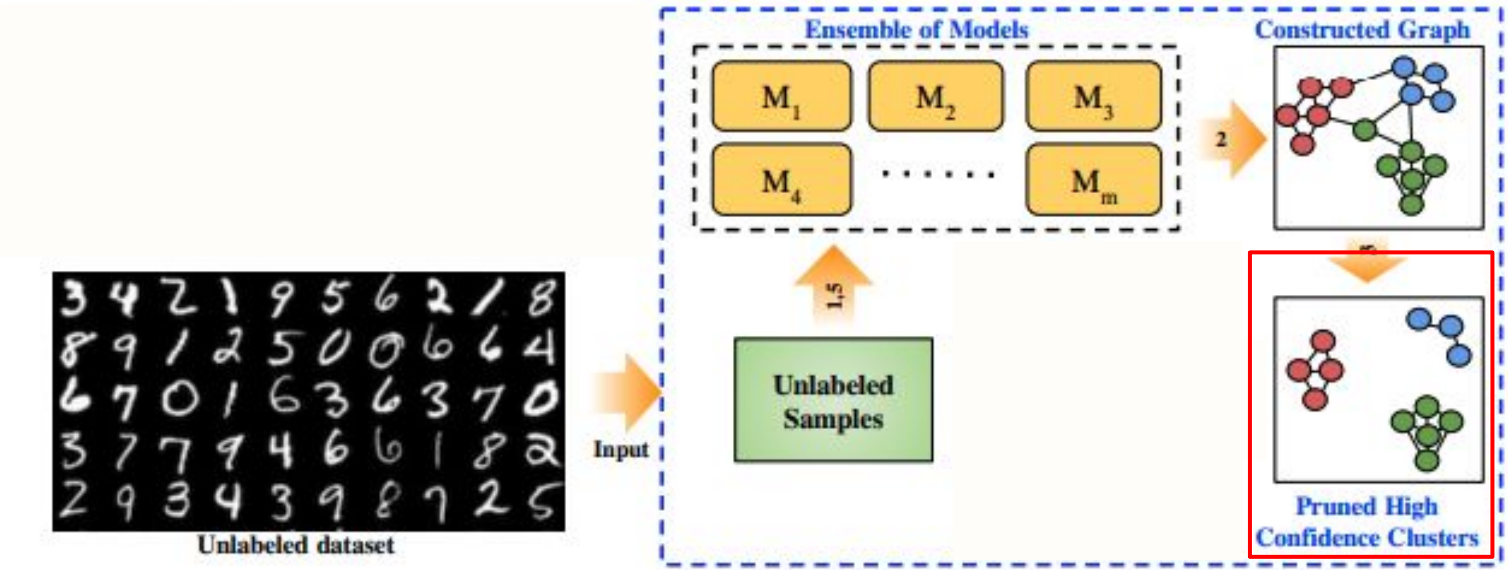
t_{pos} = 80
t_{neg} = 80
M_1
M_2
M_3
M_4
M_5
3
8
5
7
9
2
4
5
6
6
80% models disagree - Strong Negative Edge



60% models agree - No Edge

Example pairwise similarity graph

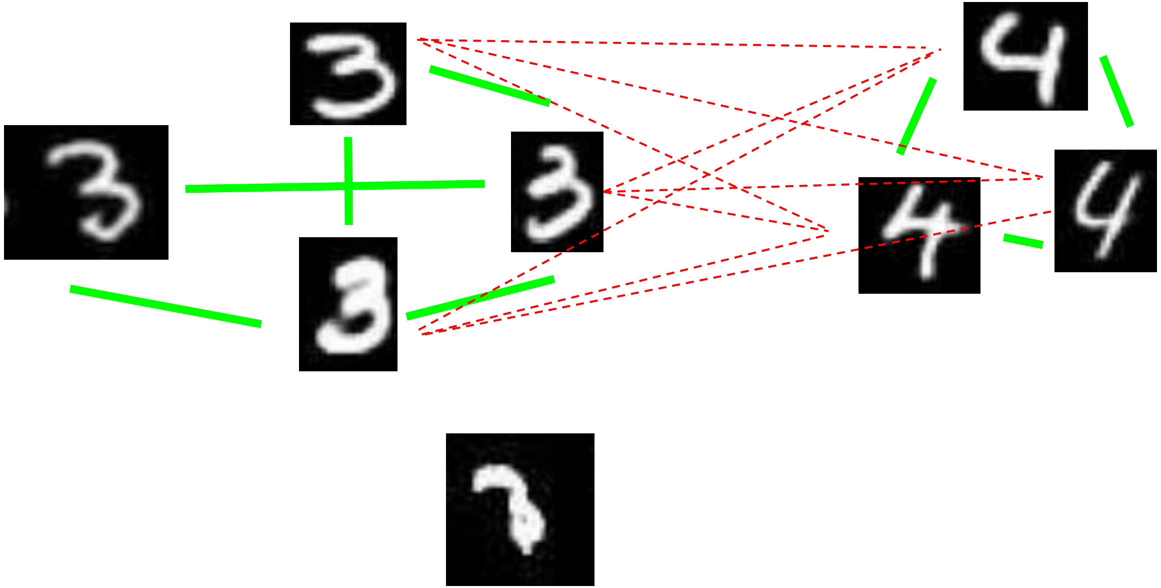




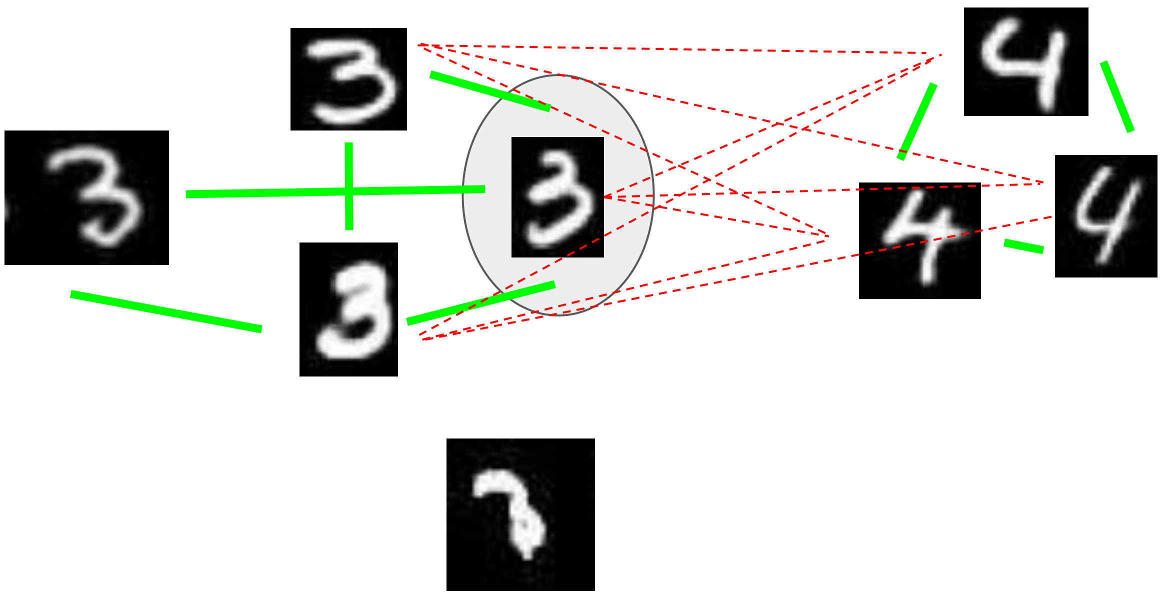
Step 3

Generate pseudo labels by pruning the graph nodes

Generating pseudo labels

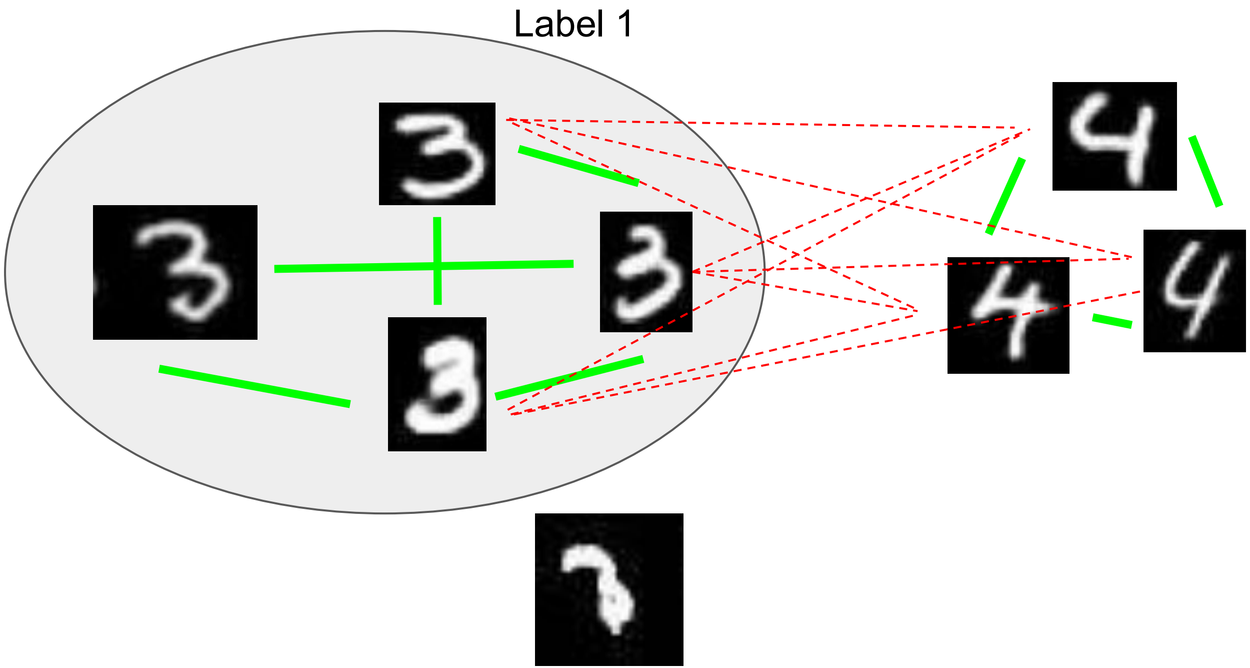


Generating pseudo labels



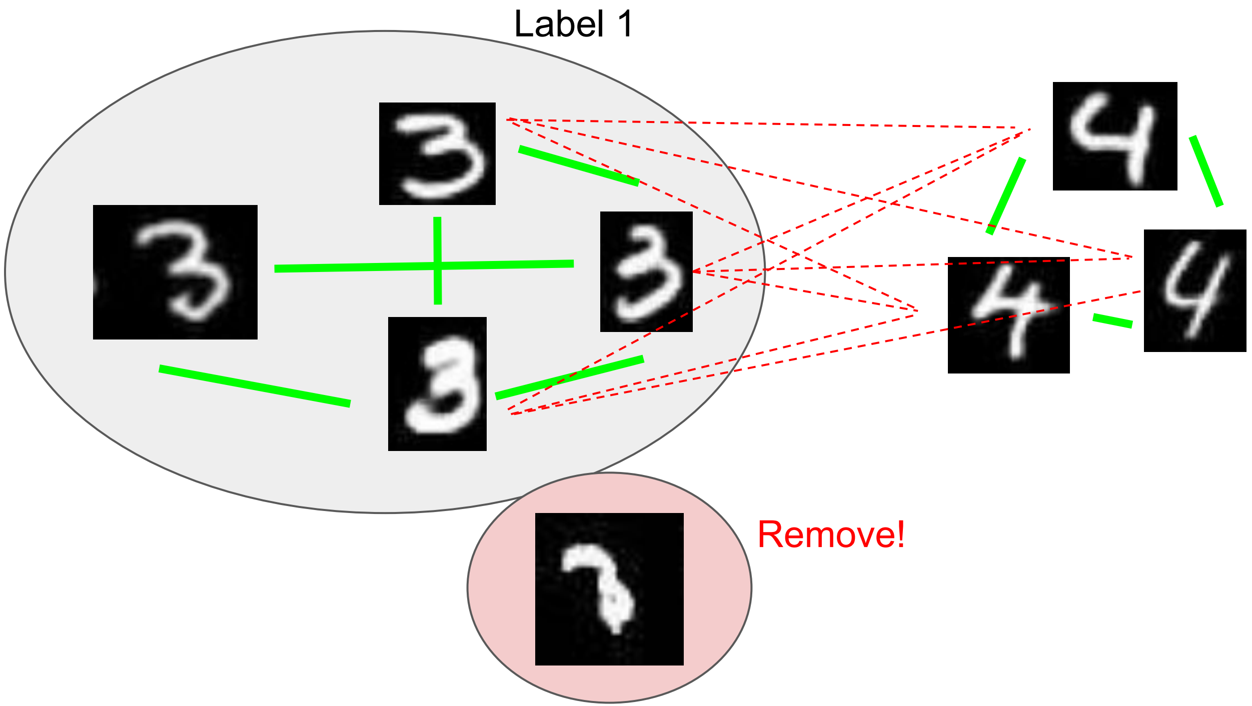
Step 1 Select node with maximum strong positive edges

Generating pseudo labels



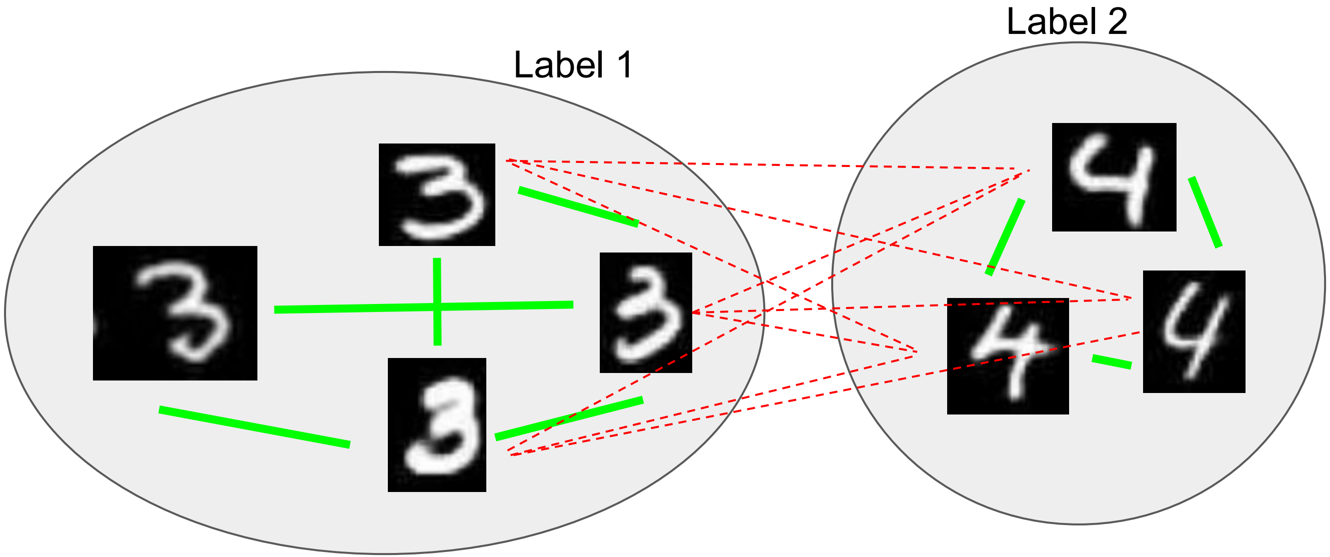
Step 2 Assign a pseudo label to the strong positive neighbors connected to the selected node

Generating pseudo labels

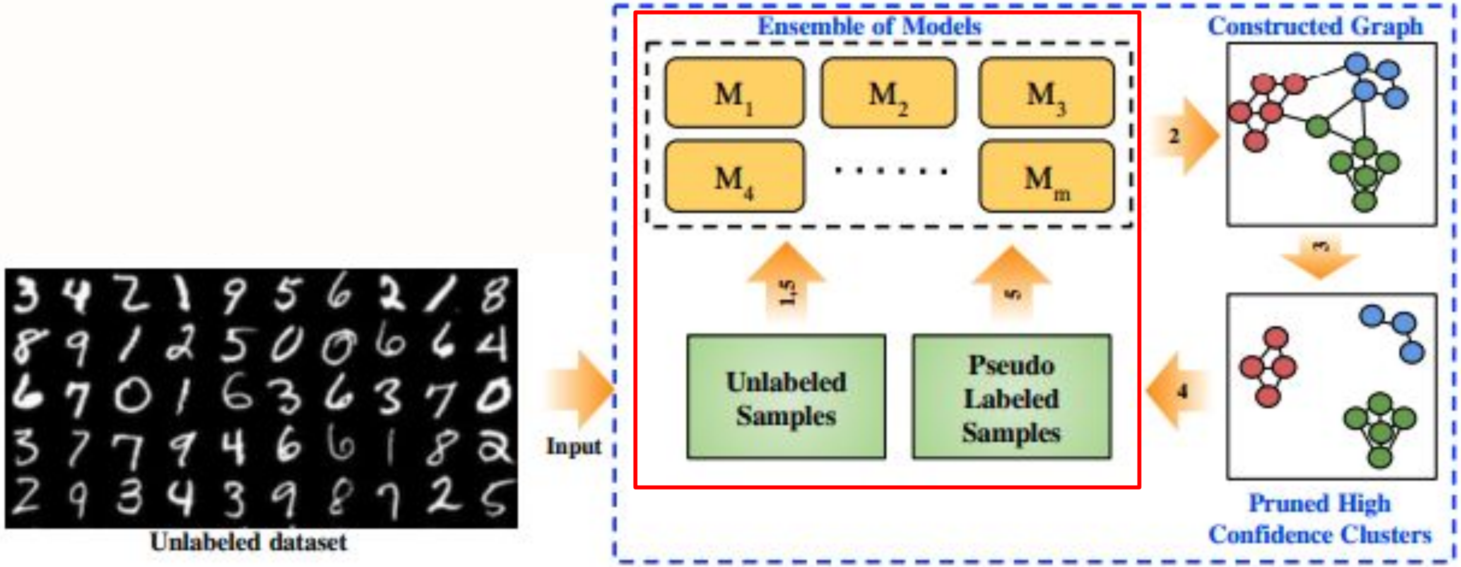


Step 3 Remove nodes not connected by strong negative edges

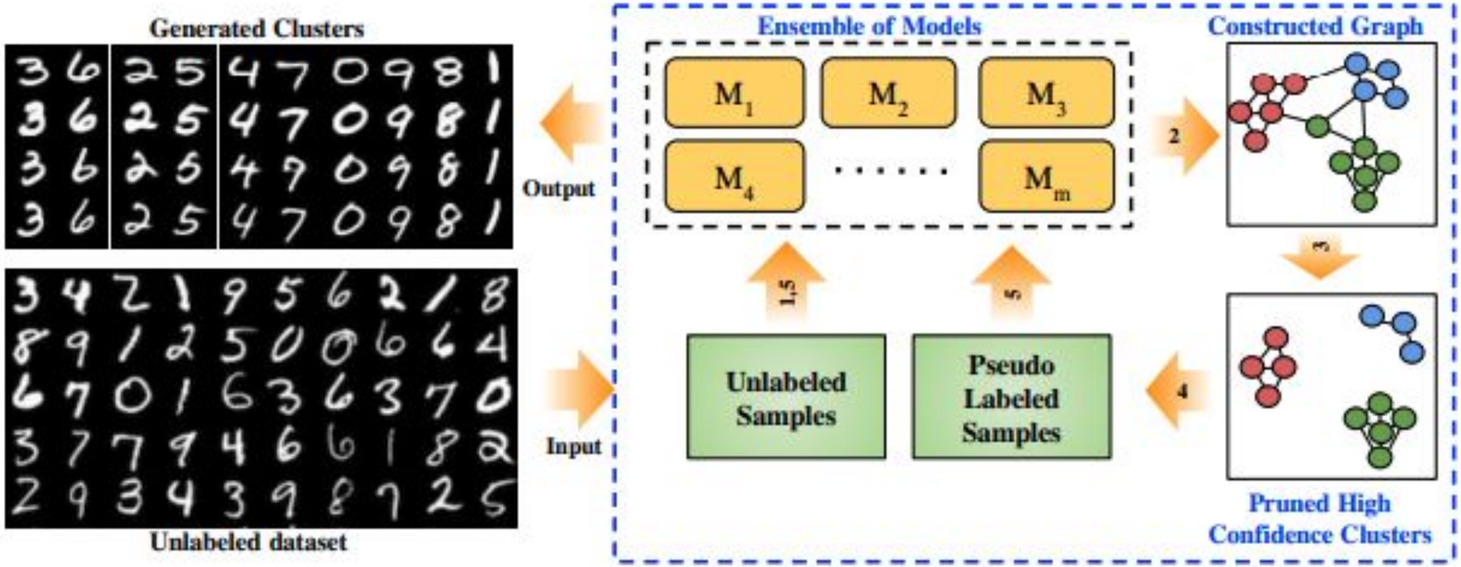
Generating pseudo labels



Repeat the steps K times



Step 4 , 5 Train the models on the Unlabeled samples and pseudo labeled samples

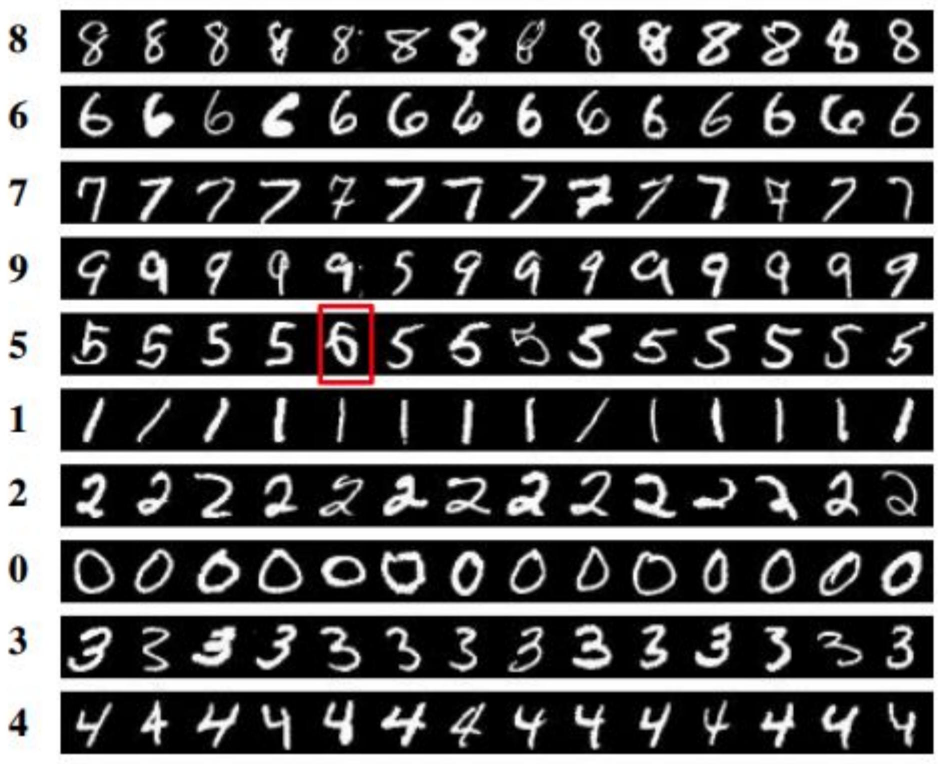


Repeat the steps 2 to step 5

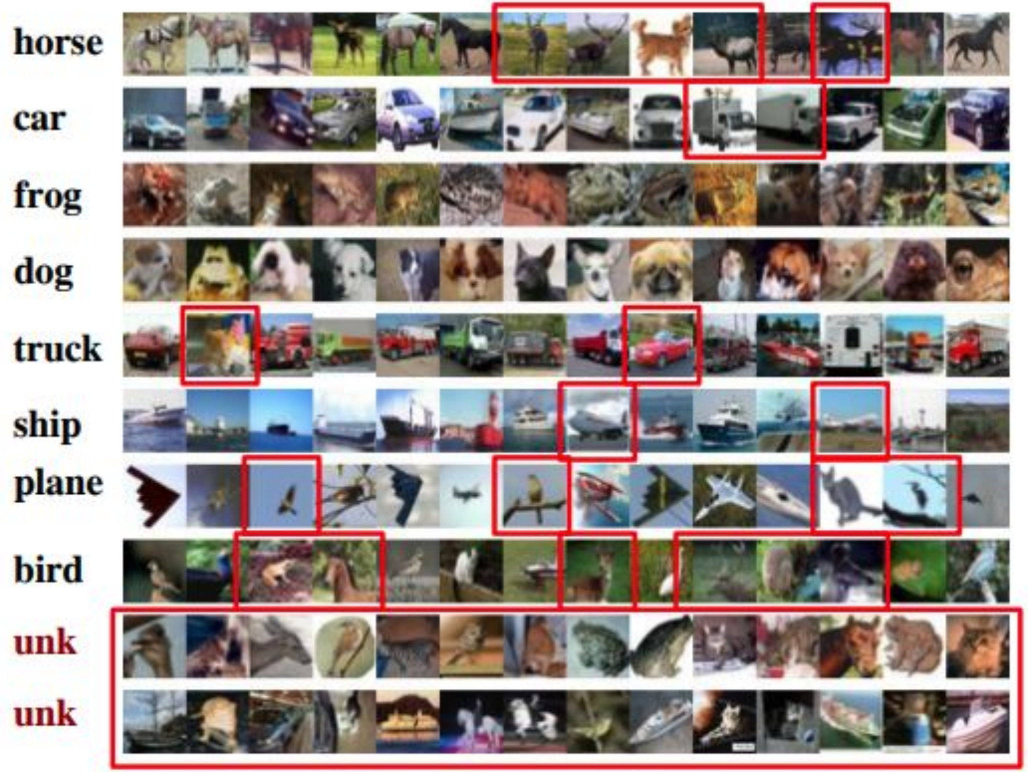
Quantitative evaluation

Method	MNIST	STL	CIFAR10	Reuters	20news
K-means	53.3 (0.1)	85.0 (0.2)	34.4 (0.9)	53.7 (0.4)	14.0 (1.5)
AC	62.1 (0.0)	82.2 (0.0)	42.4 (0.0)	54.9 (0.0)	18.6 (0.0)
dAE+K-means	67.7 (3.0)	20.8 (1.9)	45.2 (2.1)	33.7 (0.2)	7.9 (0.1)
dVAE+K-means	65.2 (3.4)	60.8 (1.9)	44.2 (0.2)	53.7 (1.4)	12.2 (0.2)
DEC Xie et al. (2016)	84.3	78.1 (0.1)	46.9 (0.9)	67.3 (0.2)	30.8 (1.8)
DeepCluster Caron et al. (2018)	27.9 (1.7)	69.9 (3.2)	37.2 (0.5)	43.1 (4.3)	15.8 (1.2)
Deep RIM Hu et al. (2017)	58.5 (3.5)	92.5 (2.2)	40.3 (3.5)	62.3 (3.9)	25.1 (2.8)
IMSAT (RPT) Hu et al. (2017)	89.6 (5.4)	92.8 (2.5)	45.5 (2.9)	**71.9 (6.5)**	24.4 (4.7)
IMSAT (VAT) Hu et al. (2017)	98.4 (0.4)	94.1 (0.4)	45.6 (0.8)	71.0 (4.9)	31.1 (1.9)
LADDER-IM (ours)	95.0 (2.8)	90.7 (1.8)	49.5 (2.9)	68.2 (2.8)	38.4 (2.5)
LADDER-IM-ensemble (ours)	95.1 (0.4)	91.5 (0.3)	51.5 (0.9)	69.0 (3.4)	40.5 (0.6)
LADDER-DOT (ours)	89.2 (7.2)	76.1 (4.7)	48.0 (1.0)	66.6 (4.8)	25.6 (1.3)
KINGDRA-LADDER-DOT (ours)	98.0 (0.01)	93.5 (1.4)	54.3 (2.5)	**71.9 (3.4)**	28.4 (1.2)
KINGDRA-LADDER-IM (ours)	**98.5 (0.4)**	**95.1 (0.1)**	**54.6 (0.9)**	70.5 (2.0)	**43.9 (1.4)**

Qualitative results MNIST



Qualitative results CIFAR10



Thank you

Project webpage : https://divamgupta.com/pseudo-semi-supervised-clustering