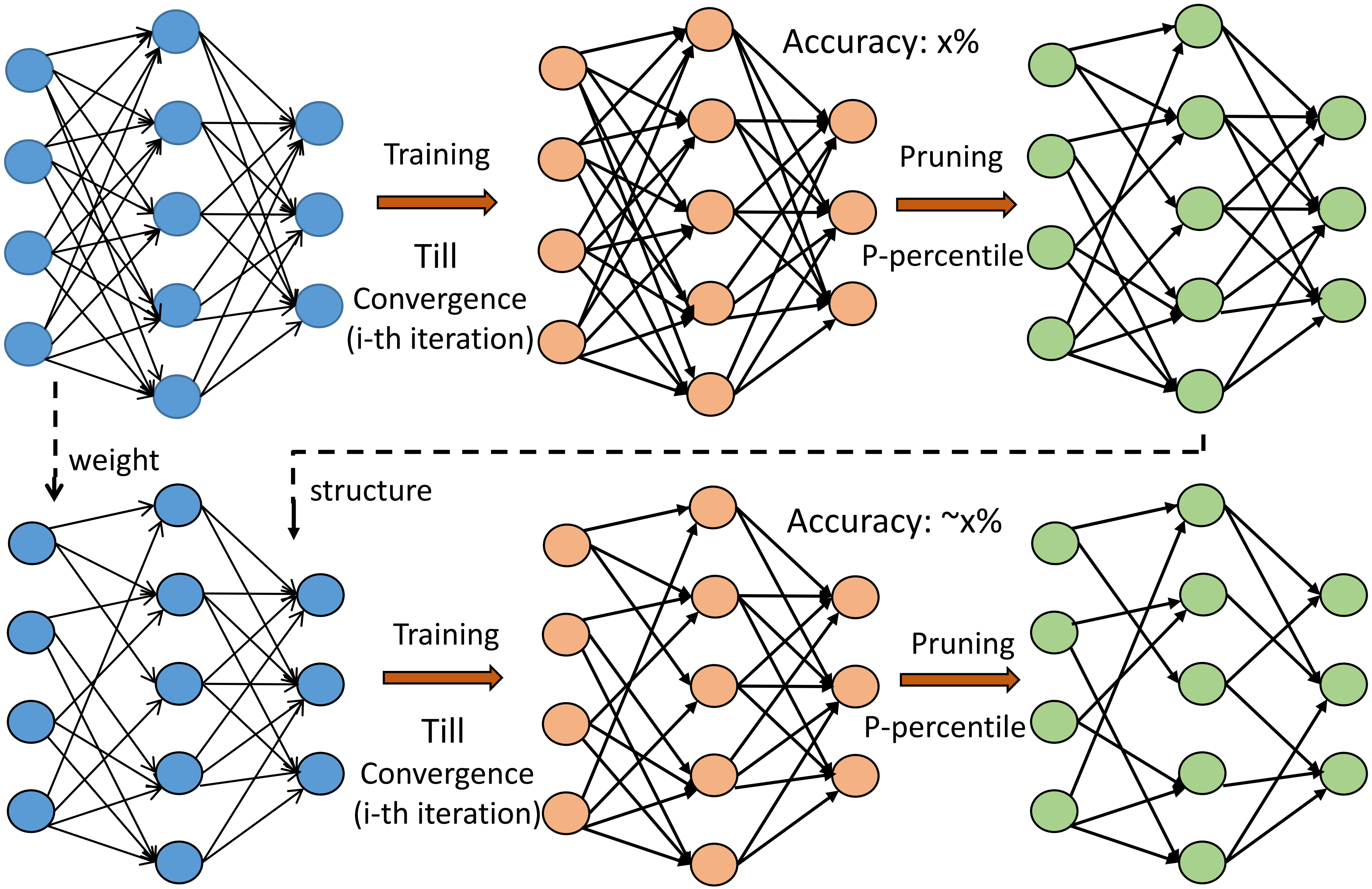
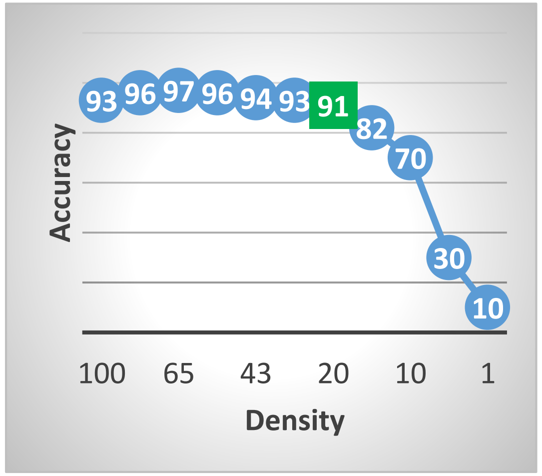
A Study on the Ramanujan Graph Property of Winning Lottery Tickets

Bithika Pal[1] Arindam Biswas[2] Sudeshna Kolay[1] Pabitra Mitra[1] Biswajit Basu[3]

[1]*Indian Institute of Technology Kharagpur,* [2]*University of Copenhagen,* [3]*Trinity College Dublin*

Lottery Ticket Hypothesis





Frankle, J. and Carbin, M. The lottery ticket hypothesis: Training pruned neural networks. In International Conference on Learning Representations (ICLR), 2019.

Except Density,
what properties of the Winning Tickets
should be analyzed?

Connectivity of the network – free flow of information from input to output

Expanders and Ramanujan Graphs

- Expanders are sparse, yet highly connected graphs
- Connectivity is measured by – **Cheeger Constant** h(G) or rate of Expansion

$$h(G) := \min \left\{ \frac{|\partial A|}{|A|} \ : \ A \subseteq V(G), 0 < |A| \leq \tfrac{1}{2}|V(G)| \right\} \qquad \partial A := \{\{x,y\} \in E(G) \ : \ x \in A, y \in V(G) \setminus A\}$$

- *h(G)* is related with the **spectral gap** of the graph; A r-regular graph with n vertices having adjacency eigenvalues $-r \leq t_n \leq \ldots \leq t_2 \leq t_1 = r$ satisfies the inequality $\frac{r-t_2}{2} \leq h(G) \leq 2\sqrt{r(r-t_2)}$
- We say that a sequence of regular graphs is an **expander family** if

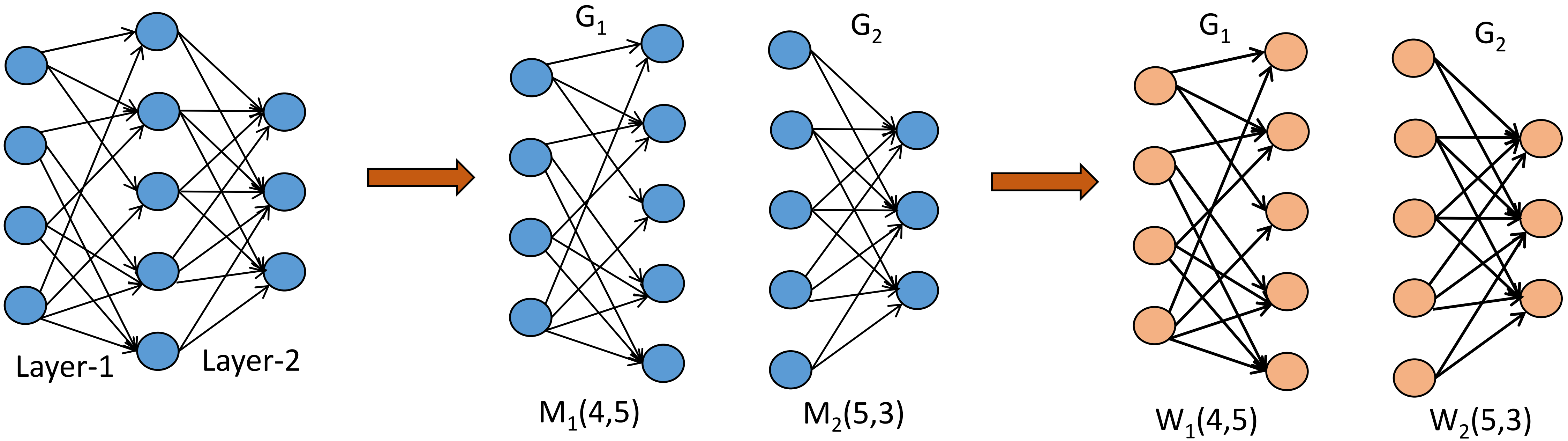
1. All the graphs have the same degree
2. The number of vertices goes to infinity
3. There exists a positive lower bound ϵ such that the expansion constant is always at least ϵ

- **Larger** values of *h(G)* signify that the graph is **strongly connected**; However, spectral gap can not be arbitrarily large. For bounded-degree expander family it is **maximal** for **Ramanujan Graphs**
- **Ramanujan Graphs** : $t_2 \leq 2\sqrt{(r-1)} = 2\sqrt{(t_1-1)}$

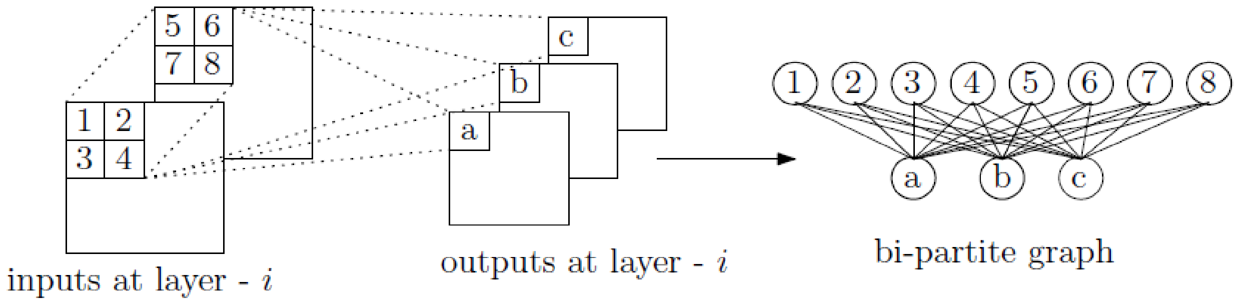
Hoory, S., Linial, N., and Wigderson, A. Expander graphs and their applications. Bull. Amer. Math. Soc. (N.S.), 43(4):439–561, 2006.

Bipartite Graph Structure

2-Layer Feed-Forward NN (pruned)   Unweighted Graphs – Masks   Weighted Graphs – Network weights



Example of a **Convolution** layer with kernel size 2 x 2, 2 input channels and 3 output channels – converted to a bipartite graph

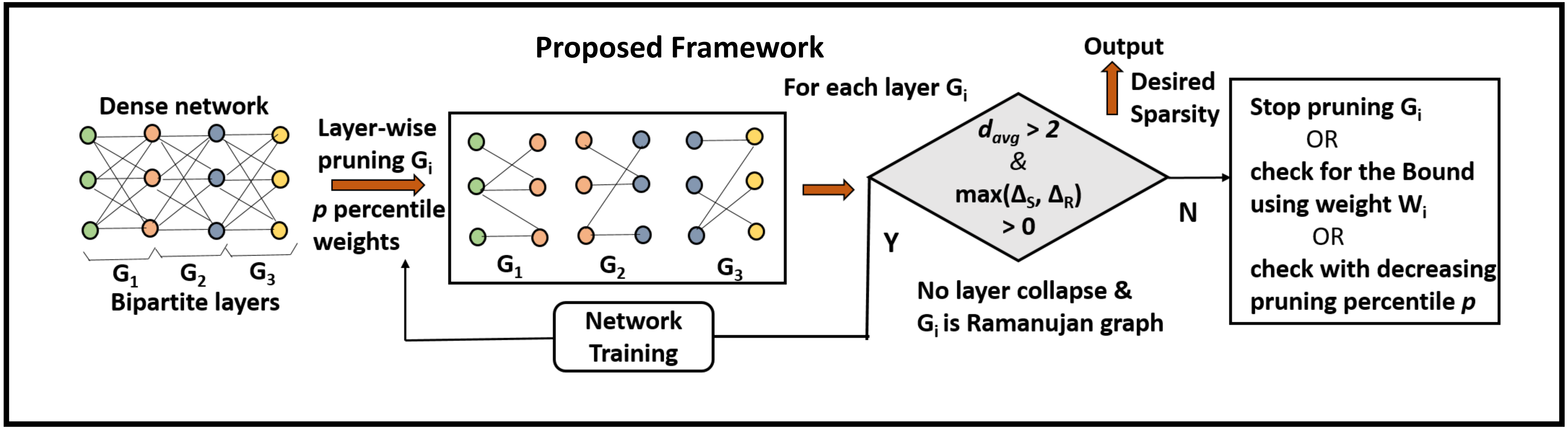


Proposed Framework

Table 1. Different bound criteria on the second largest eigenvalue t_2 of the bi-adjacency matrices

BI-ADJACENCY	EIGENVALUE (eb)	AVERAGE DEGREE (db)
Unweighted ($\mathbf{M}_i$)	$t_2(\mathbf{M}_i) \leq 2\sqrt{t_1(\mathbf{M}_i) - 1}$	$t_2(\mathbf{M}_i) \leq \sqrt{d_{avgL}(\mathbf{M}_i) - 1} + \sqrt{d_{avgR}(\mathbf{M}_i) - 1}$
Weighted ($\mathbf{W}_i$)	$t_2(\mathbf{W}_i) \leq 2\sqrt{t_1(\mathbf{W}_i) - 1}$	
difference on bound	$\Delta_S = (2\sqrt{t_1 - 1} - t_2)/t_2$	$\Delta_R = (\sqrt{d_{avgL} - 1} + \sqrt{d_{avgR} - 1} - t_2)/t_2$

1. Degree is relaxed with average degree using the bounds from universal cover of the graph
2. More sharper estimate is used for Bipartite graph



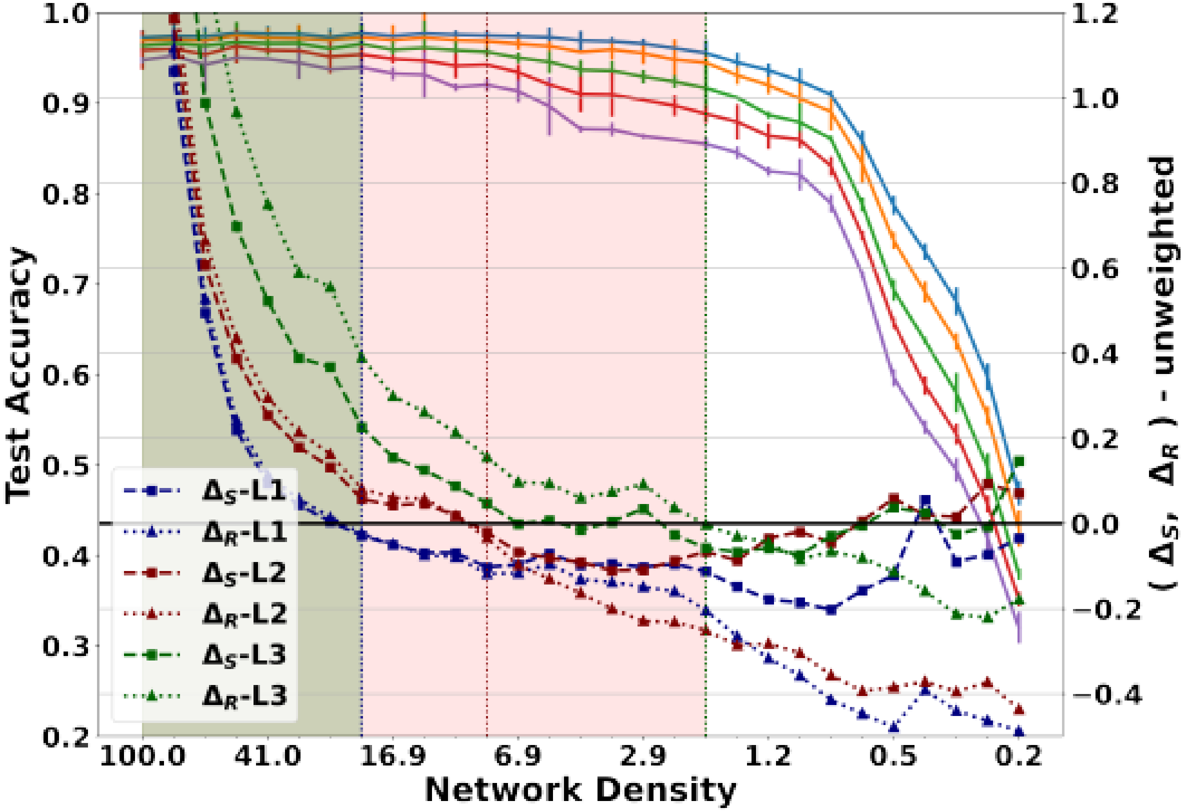
Hoory, S. A lower bound on the spectral radius of the universal cover of a graph. J. Combin. Theory Ser. B, 93(1):33–43, 2005.
Marcus, A.W., Spielman, D. A., and Srivastava, N. Interlacing families I: Bipartite Ramanujan graphs of all degrees. Ann. of Math. (2), 182(1):307–325, 2015.

Experimental Results

- Dataset- MNIST, Architecture – Lenet

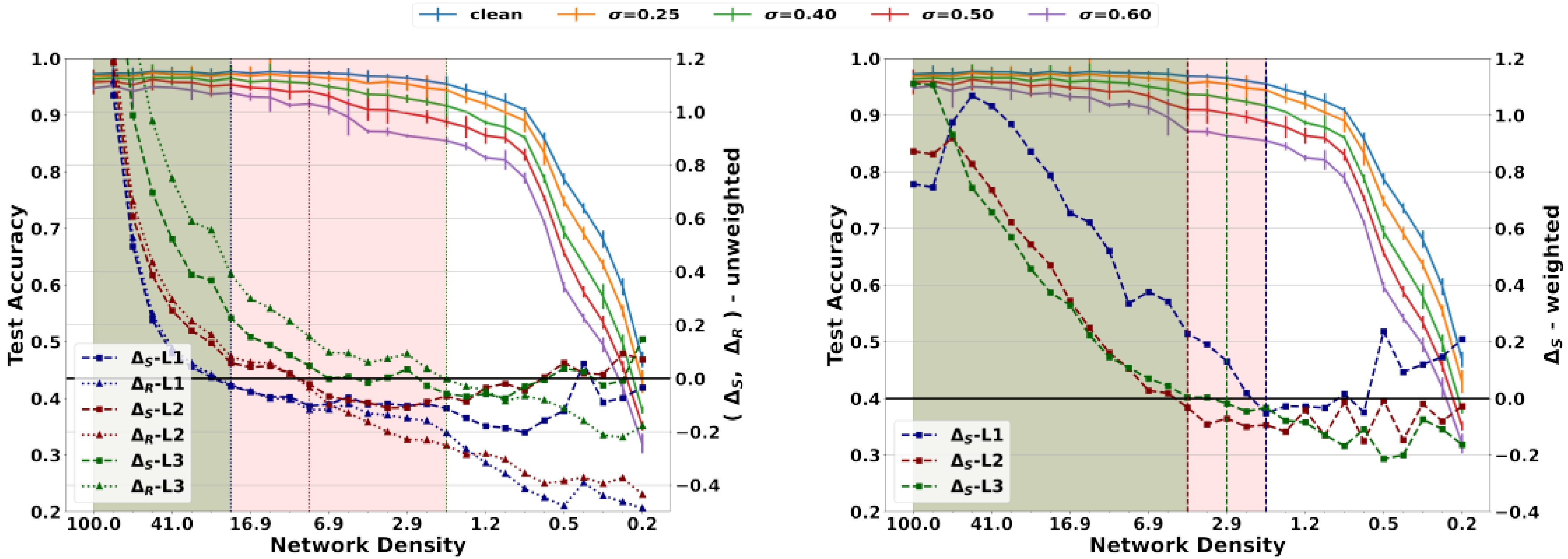
Input(784)	Bipartite graph size
L1 (300)	784 X 300
L2 (100)	300 X 100
L3 (10)	100 X 10

- Ramanujan property holds till Max(Δ_S, Δ_R) > 0
- **Gray** shaded region – all the layers follow Ramanujan graph property
- **Pink** shaded region – some of the layers follow Ramanujan Graph property
- **White** region – None of the layers follows Ramanujan Graph property



Unweighted Graph

Experimental Results



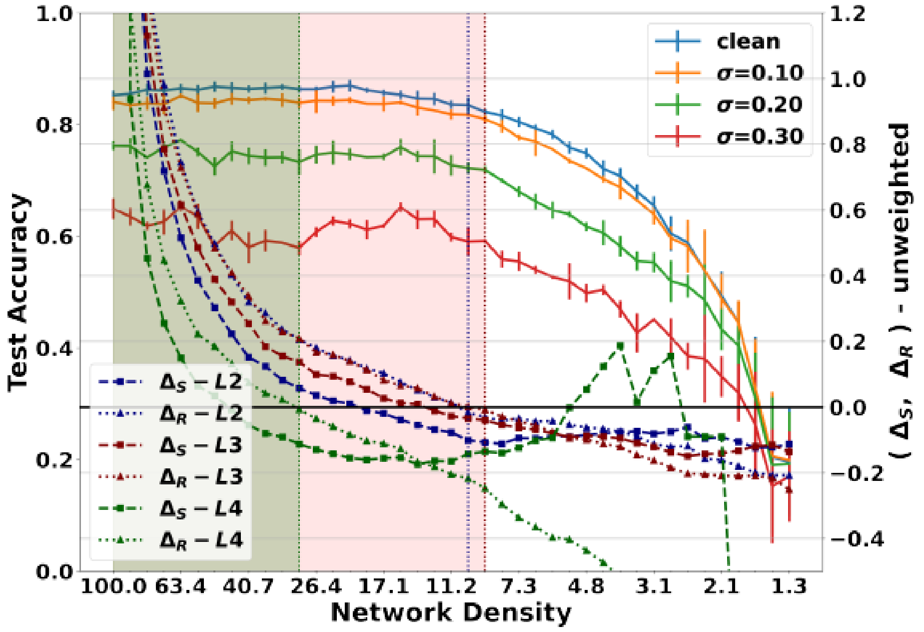
(a) unweighted graph representation

(b) weighted graph representation

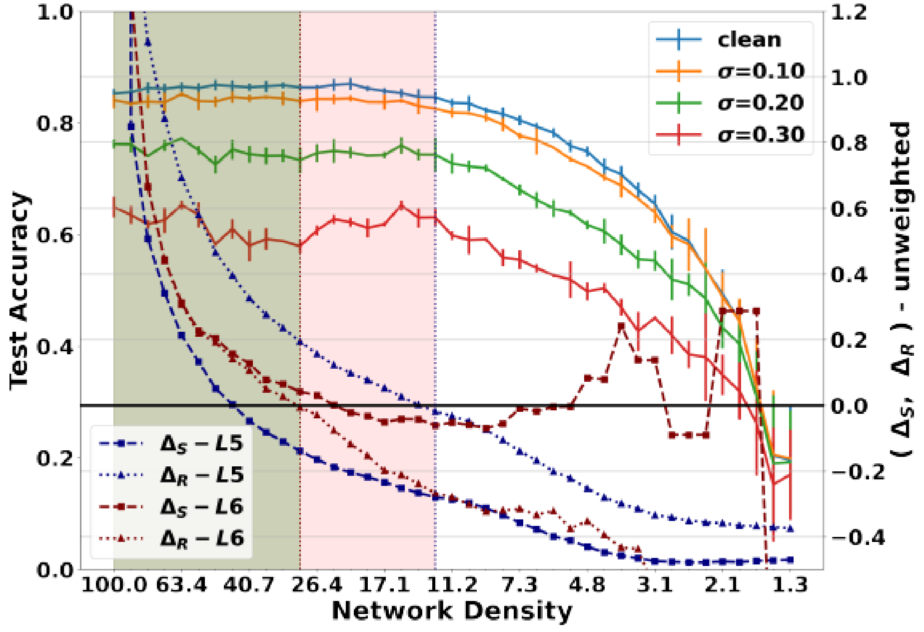
Experimental Results

- Dataset- CIFAR10, Architecture-Conv4 (Subnetwork of VGG19)

Input(3 x 32 x 32)	Bipartite graph size
Conv-L1 (64, 3, 3)	27 x 64
Conv-L2 (64, 3, 3)	576 x 64
Conv-L3 (128, 3, 3)	576 x 128
Conv-L4 (128, 3, 3)	1152 x 128
FC-L5 (256)	8192 x 256
FC-L6 (256)	256 x 256
FC-L7 (10)	256 x 10



(a) Convolutional layers



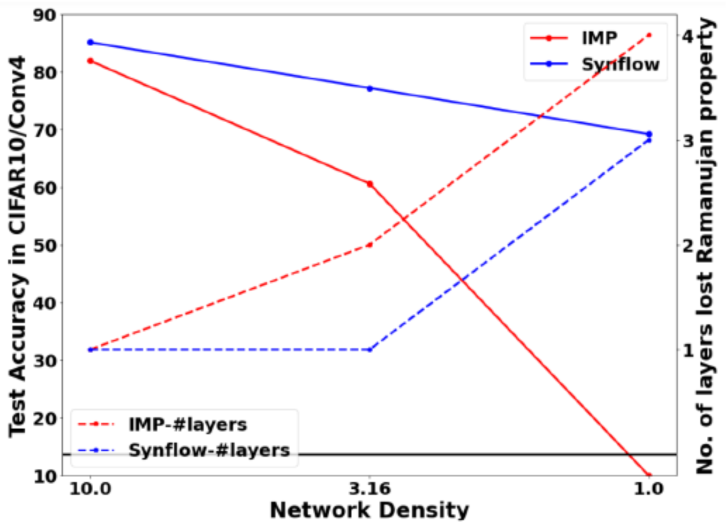
(b) Fully-connected layers

Results on Algorithm Comparison

- Based on different pruning score calculation, three algorithms are chosen
  1. Iterative Magnitude Based Pruning (IMP)
  2. Single shot network pruning based on Connection Sensitivity (SNIP) (Lee et al., 2018)
  3. Synaptic flow based pruning (SynFlow) (Tanaka et al., 2020)

- #algoname-α : Target density ($10^{-\alpha}$ x 100%) pruning
- Significant improvement in IMP in all the datasets

Pruning Algorithm	Lenet/MNIST Density	Lenet/MNIST Accuracy	Conv4/CIFAR10 Density	Conv4/CIFAR10 Accuracy
No Pruning	100.0	97.16	100.0	85.86
IMP-1.0	10.0	97.17	10.0	81.91
IMP-1.5	3.16	93.88	3.16	60.61
IMP-2.0	1.0	45.39	1.0	10
IMP-Bound	**3.6**	**96.74**	**2.54**	**70.58**
IMP*	*3.6*	*95.07*	*2.54*	*28.13*
SNIP-1.0	10.0	97.35	10.0	80.3
SNIP-1.5	3.16	79	3.16	72.95
SNIP-2.0	1.0	49.8	1.0	64.26
SNIP-Bound	**7.64**	**95.41**	**2.24**	**72.06**
SNIP*	*7.64*	*95.18*	*2.24*	*68.7*
SynFlow-1.0	10.0	97.22	10.0	82.5
SynFlow-1.5	3.16	95.92	3.16	77.18
SynFlow-2.0	1.0	49.11	1.0	69.2
SynFlow-Bound	**1.33**	**93.82**	**1.01**	**69.34**
SynFlow*	*1.33*	*67.21*	-	-



Without applying bound SynFlow can preserve Ramanujan Graph property in more number of layers than others

Lee, N., Ajanthan, T., and Torr, P. H. Snip: Single-shot network pruning based on connection sensitivity. arXiv preprint arXiv:1810.02340, 2018.
Tanaka, H., Kunin, D., Yamins, D. L., and Ganguli, S. Pruning neural networks without any data by iteratively conserving synaptic flow. arXiv preprint arXiv:2006.05467, 2020.

Conclusion

- Studied the validity of the lottery ticket hypothesis (LTH) using Ramanujan Graph.
- Three distinct regions demarcated using these bounds - test accuracy varies
- Proposed a layer-wise Ramanujan graph property preserving pruning scheme
- In future work, we will study the spectral approximation to generate more robust winning ticket preserving the Ramanujan graph property.

Thank You