Research project on graph transformer networks

Micahel Fotso - Hassan MKhallati - Abderrahim Mama
X2019

Ecole Polytechnique
January 2022 - March 2022

Outline

I - Introduction

Setup

- PATTERN dataset (2 communities/graph)
- $\frac{q}{n} = 0.25$
- $\frac{p}{n} = 0.5$
- Train set size $= 10000$ graphs
- Test set size $= 2000$ graphs
- Validation set size $= 2000$ graphs
- hidden dimension $\in 40, 80, 120, 160$
- Without Positional Encoding/ with Laplacian Positional Encoding/ with Weisfeiler Lehman Positional encoding
- Graphs size $n \in [60, 160]$

Preliminary tests

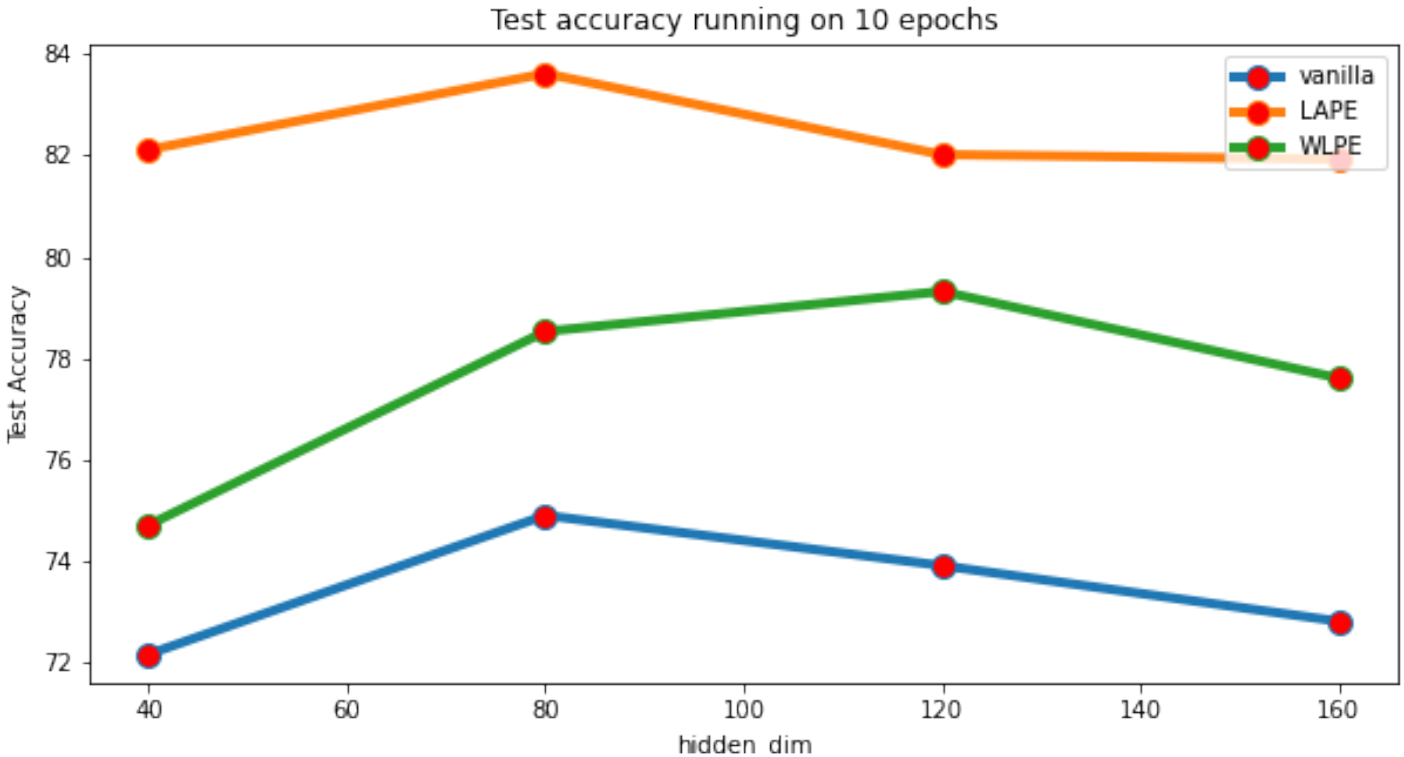

Figure: Test accuracy trained on 10 epochs

II - Different SBM datasets

Setup

- 2 communities/graph
- Graphs size $n \in [80, 120]$
- $\frac{q}{n} = 0.1$
- $\frac{p}{n} \in [0.1, 1]$
- Train set size $= 1000$ graphs
- Test set size $= 200$ graphs
- Validation set size $= 200$ graphs

SBM generation

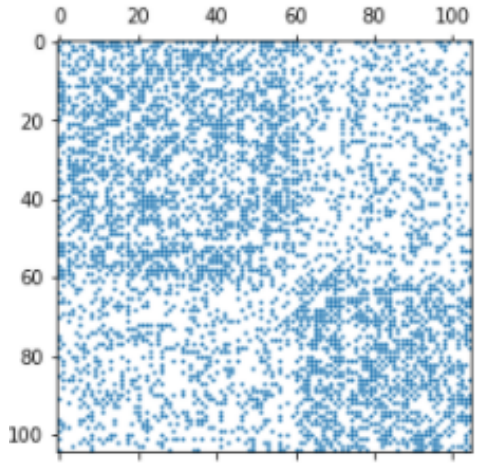

Figure: Adjacency matrix for $\frac{p}{n} = 0.5$

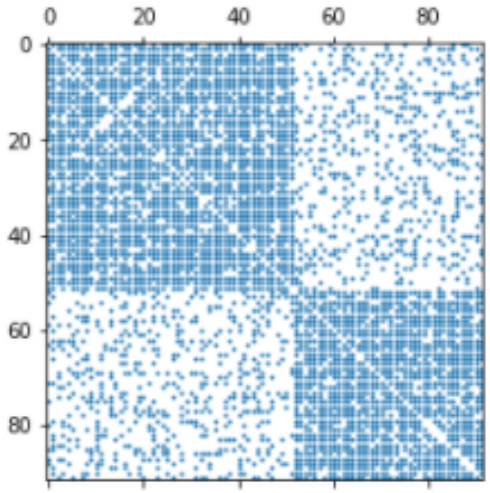

Figure: Adjacency matrix for $\frac{p}{n} = 0.8$

Signal noise ratio

$SNR(k) = \frac{(p-q)^2}{k(p+(k-1)q)}$ for a SSBM$(n, k, \frac{p}{n}, \frac{q}{n})$

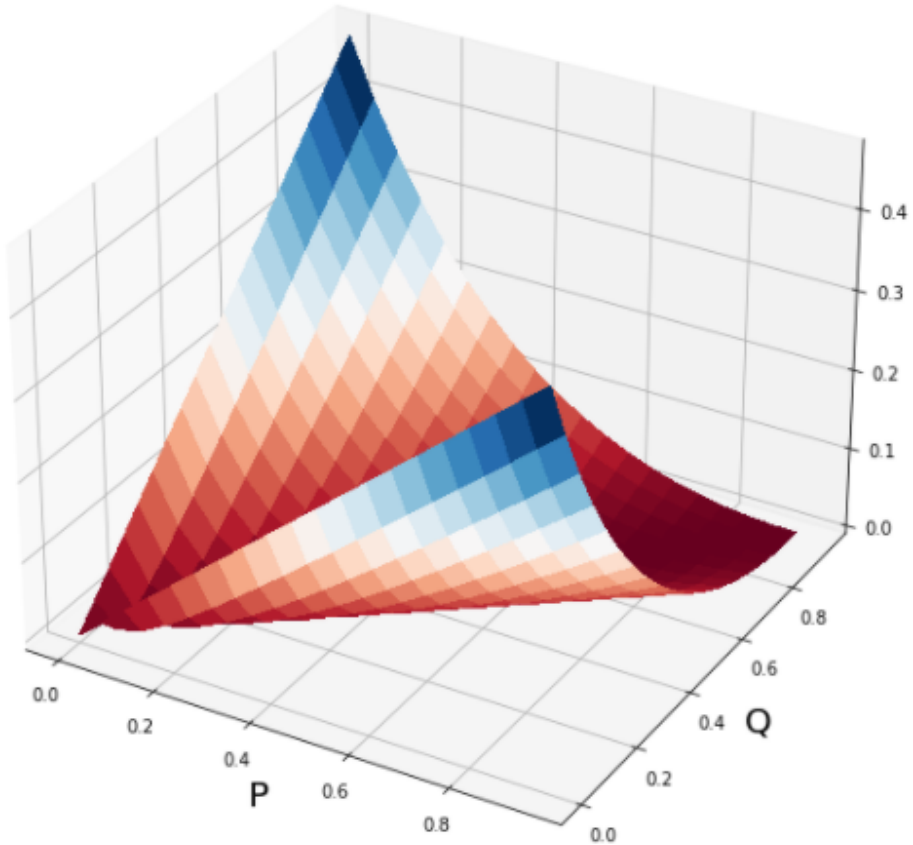

Figure: Signal noise ratio $SNR(2)$

Accuracy for different SNR

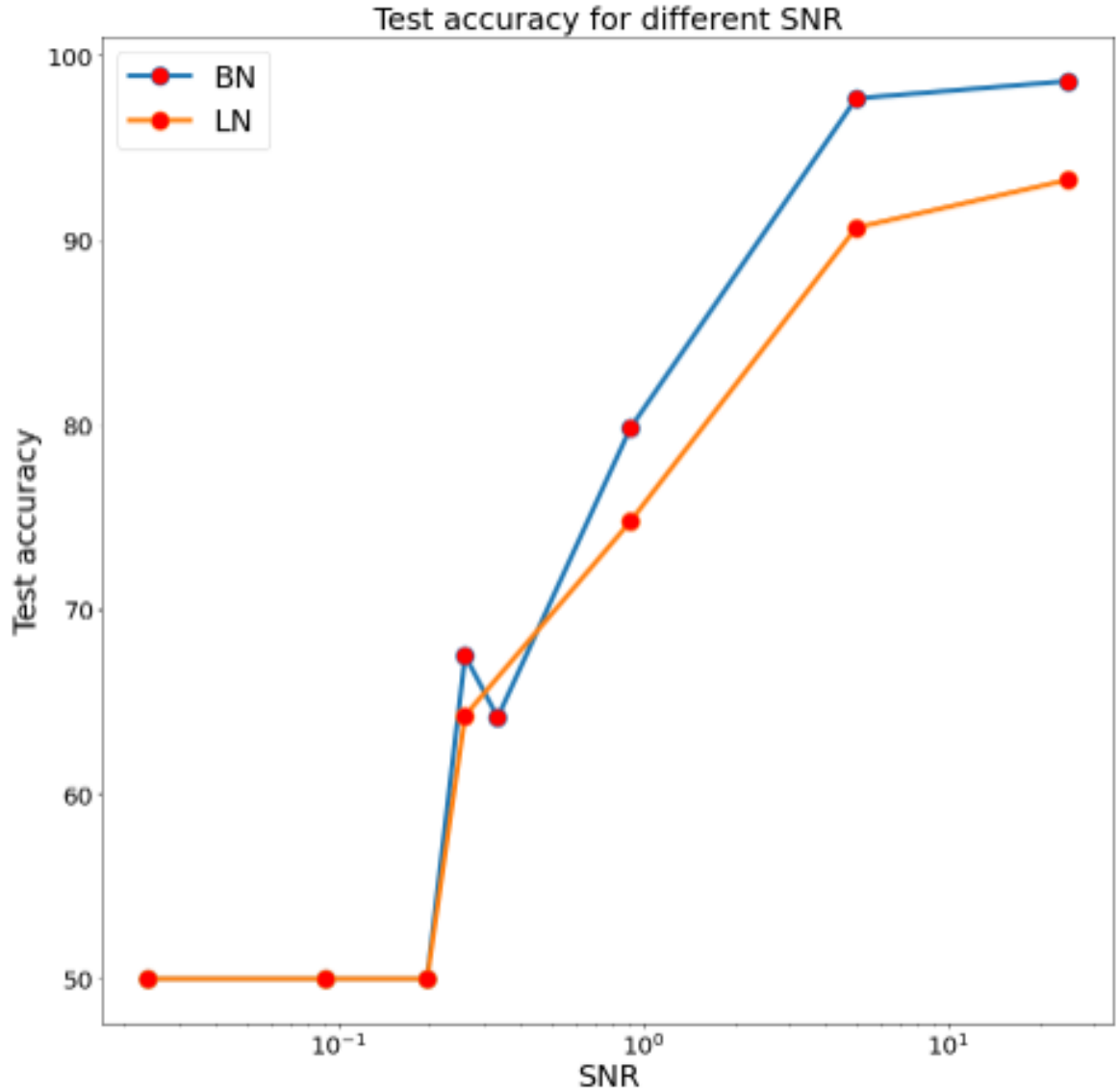

Figure: Accuracy as a function of SNR (semi-log scale)

Accuracy for different SNR

- $o(n)$ vertices are misclassified = **almost exact partition**
- strictly less than half of the vertices are misclassified = **positively correlated partition**

if $SNR \leq 1$:
almost exact partition : not reached.
positively correlated partition : reached.

Test generalization

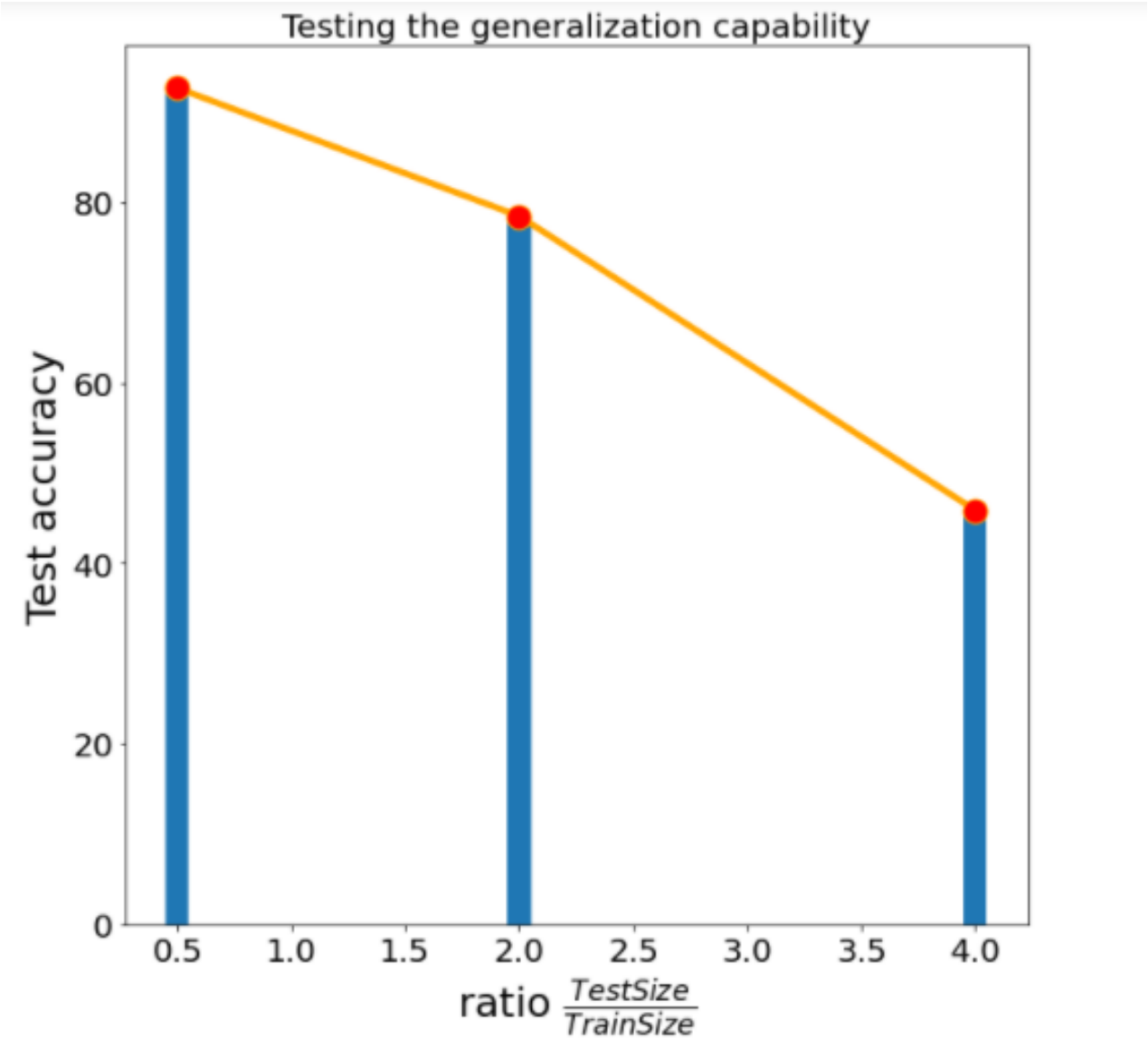

Figure: Test accuracy for fixed training size and different testing sizes

III - Some experiences on small generated dataset

Replication of experiments from the original research paper on the generated dataset

LapPE	Sparse Graph		Full graph	
	Test Perf	Train Perf	Test Perf	Train Perf
Batch Norm:False; Layer Norm:True				
false	82.28	83.22	53.98	54.80
true	81.27	80.50	50.933	53.10
Batch Norm:True; Layer Norm:False				
false	81.98	86.66	53.87	55.07
true	**91.38**	89.93	86.13	84.12

Dataset	LapPE	L	#Param	Sparse Graph Test Perf.±s.d.	Sparse Graph Train Perf.±s.d.	Sparse Graph #Epoch	Sparse Graph Epoch/Total	Full Graph Test Perf.±s.d.	Full Graph Train Perf.±s.d.	Full Graph #Epoch	Full Graph Epoch/Total
Batch Norm: `False`; Layer Norm: `True`											
ZINC	✗	10	588353	0.278±0.018	0.027±0.004	274.75	26.87s/2.06hr	0.741±0.008	0.431±0.013	196.75	37.64s/2.09hr
	✓	10	588929	0.284±0.012	0.031±0.006	263.00	26.64s/1.98hr	0.735±0.006	0.442±0.031	196.75	31.50s/1.77hr
CLUSTER	✗	10	523146	70.879±0.295	86.174±0.365	128.50	202.68s/7.32hr	19.596±2.071	19.570±2.053	103.00	512.34s/15.15hr
	✓	10	524026	70.649±0.250	86.395±0.528	130.75	200.55s/7.43hr	27.091±3.920	26.916±3.764	139.50	565.13s/22.37hr
PATTERN	✗	10	522742	73.140±13.633	73.070±13.589	184.25	276.66s/13.75hr	50.854±0.111	50.906±0.005	108.00	540.85s/16.77hr
	✓	10	522982	71.005±11.831	71.125±11.977	192.50	294.91s/14.79hr	56.482±3.549	56.565±3.546	124.50	637.55s/22.69hr
Batch Norm: `True`; Layer Norm: `False`											
ZINC	✗	10	588353	0.264±0.008	0.048±0.006	321.50	28.01s/2.52hr	0.724±0.013	0.518±0.013	192.25	50.27s/2.72hr
	✓	10	588929	**0.226±0.014**	0.059±0.011	287.50	27.78s/2.25hr	0.598±0.049	0.339±0.123	273.50	45.26s/3.50hr
CLUSTER	✗	10	523146	72.139±0.405	85.857±0.555	121.75	200.85s/6.88hr	21.092±0.134	21.071±0.037	100.25	595.24s/17.10hr
	✓	10	524026	**73.169±0.622**	86.585±0.905	126.50	201.06s/7.20hr	27.121±8.471	27.192±8.485	133.75	552.06s/20.72hr
PATTERN	✗	10	522742	83.949±0.303	83.864±0.489	236.50	299.54s/19.71hr	50.889±0.069	50.873±0.039	104.50	621.33s/17.53hr
	✓	10	522982	**84.808±0.068**	86.559±0.116	145.25	309.95s/12.67hr	54.941±3.739	54.915±3.769	117.75	683.53s/22.77hr

Modification of the network architecture

Before:
$h_i^0 = \mathbf{h_i^0} + \lambda_i$

After:
$h_i^0 = concat([h_i^0, \lambda_i])W$
where W is a weight of some linear layer with $input_{dim} = 2 * h_{dim}$ and $output_{dim} = h_{dim}$

Comparison of the two architectures

PE	**Sparse Graph**		**Full graph**	
	Test Perf	**Train Perf**	**Test Perf**	**Train Perf**
Batch Norm:False; **Layer Norm**:True				
sum	81.27	80.50	50.93	53.10
concat	**85.11**	**85.09**	**52.63**	**54.83**
Batch Norm:True; Layer Norm:False				
sum	91.38	89.93	86.13	84.12
concat	83.29	84.14	75.92	82.68

Conclusion

Conclusion

- Reproduction of the results on the original dataset and on a generated dataset
- Extension of the results with the study of the signal on noise
- Study of the ability to generalize to bigger grapher
- Modification of the network architecture