Tetali in [2] proved the relation from electrical perspective of the networks. Our approach is aligned with the spectral view of the networks. Let $v_1, v_2, \ldots, v_n$ be a set of orthonormal set of eigen vectors of the matrix $N = D^{-\frac{1}{2}} A D^{-\frac{1}{2}} = D^{\frac{1}{2}} W D^{-\frac{1}{2}}$ and $\lambda_1, \lambda_2, \ldots, \lambda_n$ be the corresponding eigen vectors. $D$ is the degree matrix, $A$ is the adjacency matrix and $W$ is the random walk matrix with $W = D^{-1}A$. Let $v_{ij}$ denote the $j$th entry of the eigen vector $v_i$. Then according to [1] the hitting time is given by,

$$H_{uv} = 2m \sum_{k=2}^{n} \frac{1}{1-\lambda_k} \left( \frac{v_{kv}^2}{d_v} - \frac{v_{ku} v_{kv}}{\sqrt{d_u d_v}} \right) \tag{1}$$

As $K_{uv} = H_{uv} + H_{vu}$ we obtain,

$$K_{uv} = 2m \sum_{k=2}^{n} \frac{1}{1-\lambda_k} \left( \frac{v_{kv}}{\sqrt{d_v}} - \frac{v_{ku}}{\sqrt{d_u}} \right)^2 \tag{2}$$

Let us examine the quantity

$$\begin{aligned}
\sum_{w \in V} \pi_w K_{wv} &= \sum_{w} \sum_{k=2}^{n} \frac{d_w}{1-\lambda_k} \left( \frac{v_{kv}}{\sqrt{d_v}} - \frac{v_{kw}}{\sqrt{d_w}} \right)^2 \\
&= \sum_{k=2}^{n} \frac{1}{1-\lambda_k} \sum_{w} \left( v_{kv} \sqrt{\frac{d_w}{d_v}} - v_{kw} \right)^2 \\
&= \sum_{k=2}^{n} \frac{1}{1-\lambda_k} \left( \frac{v_{kv}^2}{d_v} \sum_{w} d_w - 2 \sum_{w} \frac{v_{kv} v_{kw}}{\sqrt{d_v}} + \sum_{w} v_{kw}^2 \right) \\
&= \sum_{k=2}^{n} \frac{1}{1-\lambda_k} \left( 2m \frac{v_{kv}^2}{d_v} + 1 \right) \qquad (3)
\end{aligned}$$

We get 3 by using the fact that $v_k$ is of unit length and orthogonal to $v_1$ for $k \geq 2$. We obtain another result:

$$\sum_{w} (K_{wv} - K_{uw}) = 2m \sum_{k=2}^{n} \frac{1}{1-\lambda_k} \left( \frac{v_{kv}^2}{d_v} - \frac{v_{ku}^2}{d_u} \right) \tag{4}$$

Let us rewrite $H_{uv}$ as

$$\begin{aligned}
H_{uv} &= m \sum_{k=2}^{n} \frac{1}{1-\lambda_k} \left[ \left( \frac{v_{kv}^2}{d_v} - \frac{2 v_{ku} v_{kv}}{\sqrt{d_u d_v}} + \frac{v_{ku}^2}{d_u} \right) + \left( \frac{v_{kv}^2}{d_v} - \frac{v_{ku}^2}{d_u} \right) \right] \\
&= \frac{1}{2} \left[ K_{uv} + \sum_{w} (K_{wv} - K_{uw}) \right] \qquad (5)
\end{aligned}$$

This is a coding question. Below is the pseudocode to generate SBM. Let $E$ be the set of edges of the graph $G$.

```
V = random_permutation([1,2, ..., 2n])
v1 = V[1,...,n]
v2 = V[n+1, n+2, ..., 2n]

for (u,v) in combinations(V, 2):
    if u and v both in same_block :
        if flip(p):
            G.add_edge(u,v)
    else if u and v in different_block:
        if flip(q):
            G.add_edge(u,v)
```

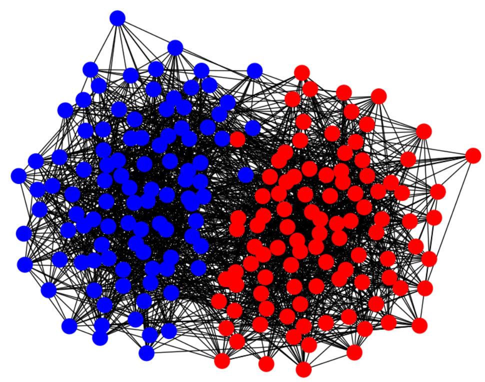
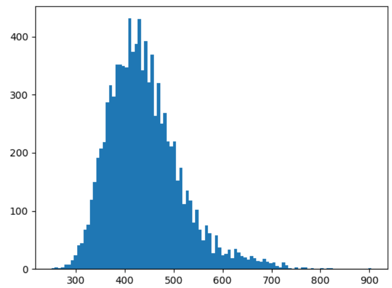
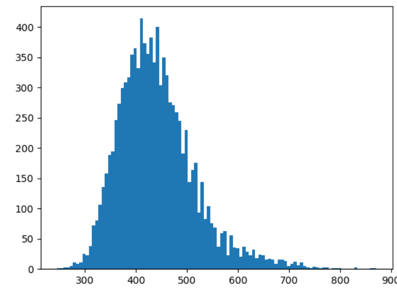


Figure 1: Two blocks of generated sbm graph $G(100, 0.16, 0.04)$

Vertices belonging to two different communities are colored red and blue.




(a) Histogram of commute time within same block



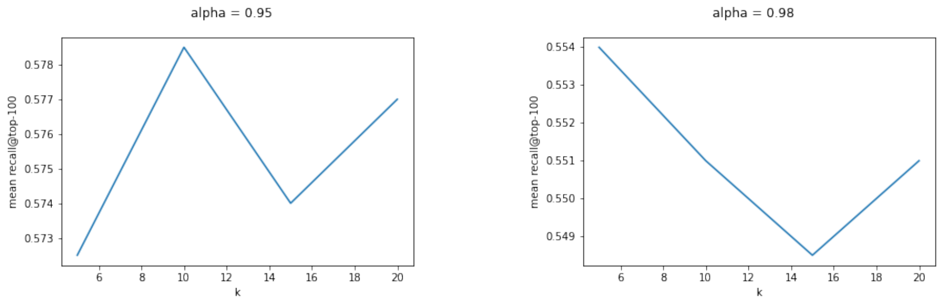

(b) Histogram of commute time. within different blocks

Figure 2: Histogram of commute times

Table 1 below shows the empirical mean and variance of commute time within same and different blocks. Note that mean commute time within different blocks have higher time than within same block. Note that commute time is calculated from Eqn 2.

| | mean | variance |
|---|---|---|
| In | 440.78326 | 5591.5689 |
| Different | 440.9714 | 5680.7047 |

Table 1: Empirical mean and variance of commute time




(a) alpha=0.95

(b) alpha=0.98

Figure 3: recall@top-100 for different parameters

| | k | | | |
|---|---|---|---|---|
| alpha | 5 | 10 | 15 | 20 |
| 0.95 | 0.5725 | 0.5785 | 0.5740 | 0.5770 |
| 0.98 | 0.5540 | 0.5510 | 0.5485 | 0.5509 |

Table 2: Recall@top-100 for different alpha and k values

There are 1000 nodes in the test dataset of 'Cora'.

```
In[2]: from torch_geometric.datasets import Planetoid
In[3]: dataset = Planetoid(root='/tmp/Cora', name='Cora')
In[4]: data = dataset[0]
In[5]: data.test_mask.sum().item()
Out[5]: 1000
```

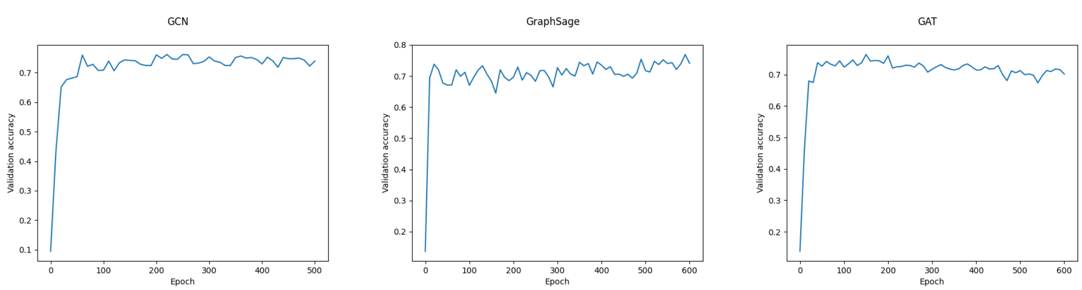
Each model is run with following hyperparameters.

```
python3 train.py --model_type=GCN --dataset=cora --dropout=0.5 --weight_decay=5e-3
--epochs=500
python3 train.py --model_type=GraphSage --hidden_dim=256 --dataset=cora --dropout=0.6
--weight_decay=5e-4 --epochs=600
python3 train.py --model_type=GAT --hidden_dim=64 --dataset=cora --dropout=0.6
--weight_decay=5e-4 --epochs=600
python3 train.py --model_type=APPNP --dataset=cora --dropout=0.5 --weight_decay=5e-3
--epochs=500
python train.py --dataset=enzymes --weight_decay=5e-3 --num_layers=3 --epochs=500
```

Note that due to time constraints careful hyperparameter tuning was not possible. But the results are following. Based on the results in Table 3 'GraphSage' performed best.

| | Accuracy | |
|---|---|---|
| Model Type | Final | Max |
| GCN | 0.739 | 0.761 |
| GraphSage | 0.768 | 0.768 |
| GAT | 0.702 | 0.737 |

Table 3: Validation accuracy for different models




(a) GCN (b) GraphSage (c) GAT

Figure 4: Validation accuracy for different models

For the model APPNP final accuracy is 0.72 and the max accuracy obtained during training is 0.748.

Compared to above still GraphSage performs best.

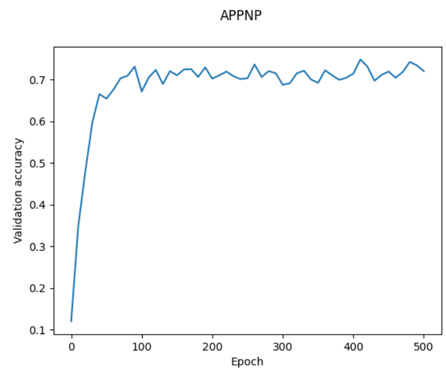



Figure 5: Validation accuracy of APPNP model

## References

[1] László Lovász et al. Random walks on graphs: A survey. *Combinatorics, Paul erdos is eighty*, 2(1):1–46, 1993.

[2] Prasad Tetali. Random walks and the effective resistance of networks. *Journal of Theoretical Probability*, 4(1):101–109, 1991.

# Information sheet
# CS59000-MLG: Computation and Machine Learning on Graphs

**Assignment Submission** Fill in and include this information sheet with each of your assignments. This page should be the last page of your submission. Assignments are due at 11:59pm and are always due on a Thursday. All students please submit their homework via Brightspace. Students can typeset or scan their homework. Make sure that you answer each (sub-)question on a separate page. That is, one answer per page regardless of the answer length. Regarding the programing question, please answer them with your idea/pseudo codes. Students also need to upload their code together. Put all the code for a single question (not for each sub-question) into a single file, compress all the files and this PDF into one file (.zip) and upload the compressed file.

**Late Homework Policy** *Homework are due on Thursdays at 11:59pm PT and one late period expires on the following Monday at 11:59pm PT.* Only one late period may be used for an assignment. Any homework received after 11:59pm PT on the Monday following the homework due date will receive no credit. Once these late periods are exhausted, any assignments turned in late will receive no credit.

**Honor Code** I strongly encourage students to form study groups. Students may discuss and work on homework problems in groups. However, each student must write down their solutions independently, i.e., each student must understand the solution well enough in order to reconstruct it by him/herself. Students should clearly mention the names of all the other students who were part of their discussion group. Using code or solutions obtained from the web (GitHub/Google etc.) is considered an honor code violation. I will randomly check the submissions for plagiarism.

**Your name:** Soham Mukherjee

**Email:** mukher26@purdue.edu **SUID:** 29920150

Discussion Group: ______________________________

I acknowledge and accept the Honor Code.

*(Signed)* SM