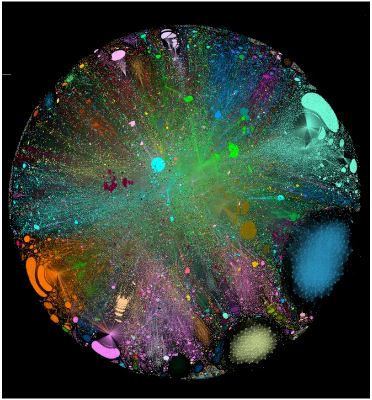
Taxonomy of reduction matrices for Graph Coarsening

Antonin Joly, Nicolas Keriven, Aline Roumy

Motivation

- Graph such as recommender systems (Reddit) too big to enter GPU
- Graph Coarsening is a solution, along with graph condensation and node sampling strategies



Background

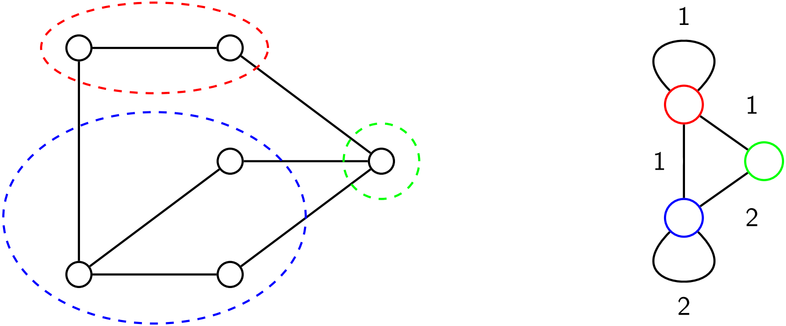


Figure: Graph Coarsening with coarsening ratio of 4/7

Background

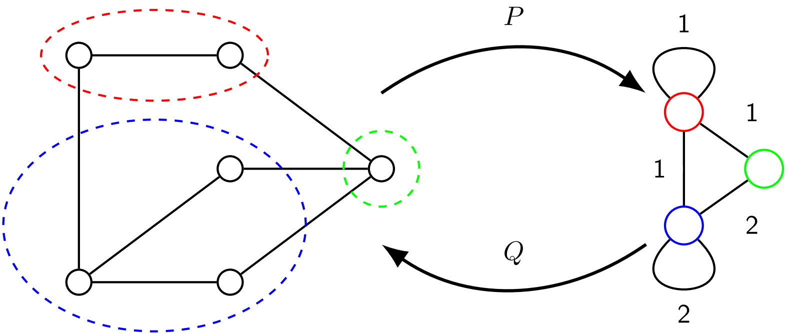


Figure: Graph Coarsening with coarsening ratio of 4/7

Background

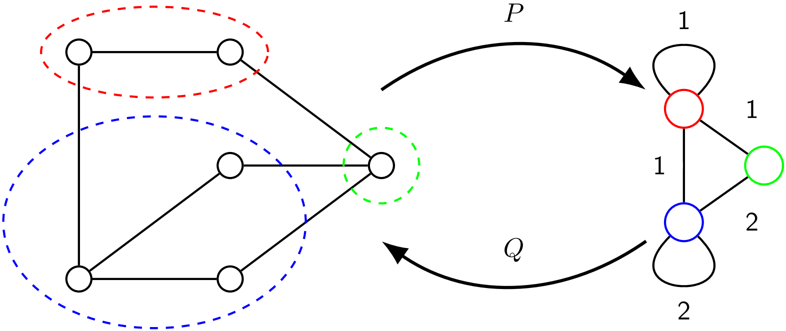


Figure: Graph Coarsening with coarsening ratio of 4/7

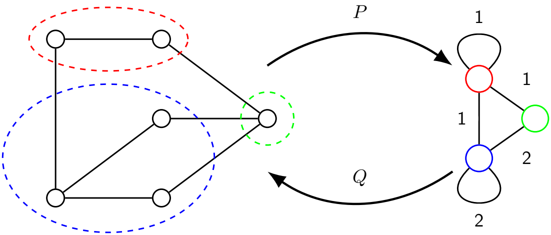
Do P and Q play similar roles?

Only lifting matrix Q contains structural information

- *Coarsened Adjacency*: $A_c = Q^\top A Q$
- *Comb. Laplacian*: $\mathcal{L} = D - A$,

$$Q = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

A well partitioned[1] Q *matrix for* $\mathcal{L}$



[1] Q is said to be well-partitioned if it has exactly one non-zero coefficient per row

Only lifting matrix Q contains structural information

- *Coarsened Adjacency*: $A_c = Q^\top A Q$
- *Comb. Laplacian*: $\mathcal{L} = D - A$,

$$Q = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

A well partitioned[1] *Q matrix for $\mathcal{L}$*

Lemma. (Consistency of Laplacian [1])
Let Q be a well-partitioned matrix. The two following properties are equivalent:

$$Q \text{ is proportional to a binary matrix.}$$
$$\Updownarrow$$
$$\forall A, \quad L(A_c) = Q^\top L Q$$

[1] Andreas Loukas, *Graph Reduction with Spectral and Cut Guarantees*, JMLR 2019.

Only lifting matrix Q contains structural information

Lemma. (Consistency of Laplacian [1])
Let Q be a well-partitioned matrix. The two following properties are equivalent:

$$Q \text{ is proportional to a binary matrix.}$$
$$\Updownarrow$$
$$\forall A, \quad L(A_c) = Q^\top L Q$$

$\Longrightarrow$ **For a fixed lifting matrix Q, what are the admissible reduction matrix P? Can we find a better matrix P than the pseudo inverse?**

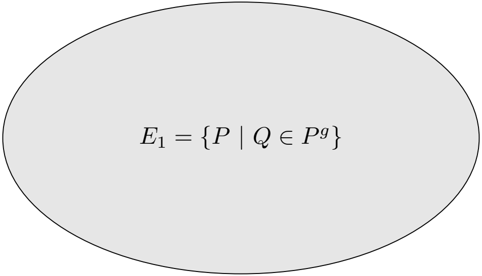
[1] Andreas Loukas, *Graph Reduction with Spectral and Cut Guarantees*, JMLR 2019.

General set of reduction matrices E_1

Lemma. (Generalized Inverse and Π projection)
For a well-partitioned lifting matrix Q, let $\Pi = QP$:

$$\Pi^2 = \Pi \quad \Longleftrightarrow \quad Q \in P^g$$

where P^g is the set of **generalized inverse** of P (more general than Moore-Penrose)



To our knowledge, E_1 does not have a closed-form!

A characterizable subset E_2

Lemma. (Generalized reflexive inverse)
For a well-partitioned lifting matrix Q and a reduction matrix P such that $Q \in P^g$:

$$\operatorname{rank}(P) = n \quad \Longleftrightarrow \quad P \in Q^g$$

Conversely, $P \in Q^g$ implies $Q \in P^g$ and $\operatorname{rank}(P) = n$, such that $E_2 = Q^g \subset E_1$. E_2 is the set of generalized **reflexive** inverse.

Lemma. (Characterization of generalized reflexive inverses of Q)
Let $Q \in \mathbb{R}^{N \times n}$ be a well-partitioned lifting matrix.

$$E_2 = Q^g = \{Q^+ + M\left(I_N - QQ^+\right) \mid M \in \mathbb{R}^{n \times N}\}$$

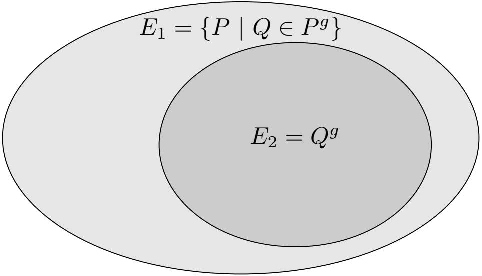
where M can be optimized wrt *anything*, supervised or unsupervised.

A characterizable subset E_2

Lemma. (Characterization of generalized reflexive inverses of Q)
Let $Q \in \mathbb{R}^{N \times n}$ be a well-partitioned lifting matrix.

$$E_2 = Q^g = \{Q^+ + M\left(I_N - QQ^+\right) \mid M \in \mathbb{R}^{n \times N}\}$$

where M can be optimized wrt *anything*, supervised or unsupervised.

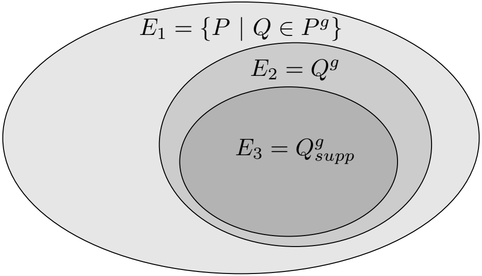


E_2 is easily characterized but the optimized matrix M is dense!

A sparse subset E_3

Lemma. (Generalized reflexive inverse with same support)
Let $Q \in \mathbb{R}^{N \times n}$ be a well-partitioned and **binary** lifting matrix. The set of reflexive generalized inverse of Q with the **same support** as $Q^\top$ is defined as :

$$E_3 = \left\{ P \in \mathbb{R}^{n \times N} \;\middle|\; \begin{cases} supp(P) = supp(Q^\top) \\ \sum_{k=1}^{N} P_{ik} = 1 \quad \forall i \in [1, n] \end{cases} \right\}$$



Same support as Q^T is sparse (N non zero terms vs. $n \times N$)!

An example of score : RSA

Definition. (Restricted Spectral Approximation) (**RSA**)
Consider a subspace $\mathcal{R} \subset \mathbb{R}^N$, a Laplacian L, a lifting matrix Q and a reduction matrix P, and $\|x\|_L = \sqrt{x^\top Lx}$. The *RSA constant* $\epsilon_{L,Q,\mathcal{R}}(P)$ is defined as :

$$\epsilon_{L,Q,\mathcal{R}}(P) = \sup_{x \in \mathcal{R}, \|x\|_L = 1} \|x - QPx\|_L$$

Many classical coarsening algorithms aim to minimize the RSA

Example of reduction matrices

- $P_{MP} = Q^{+} = (Q^{\top} Q)^{-1} Q^{\top}$

Exact solution:

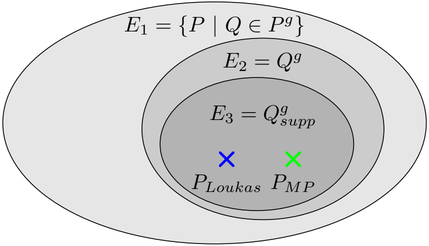
$$\arg\min_{P} \sup_{x \in \mathbb{R}^N, \|x\|_2 = 1} \|x - QPx\|_2$$

RSA :

$$\arg\min_{P} \sup_{x \in \mathcal{R}, \|x\|_L = 1} \|x - QPx\|_L$$

Example of reduction matrices

- $P_{MP} = Q^+ = (Q^\top Q)^{-1} Q^\top$
- $P_{Loukas} = Q_l^+ \dots Q_1^+$



Example of reduction matrices

- $P_{MP} = Q^{+} = (Q^{\top}Q)^{-1}Q^{\top}$
- $P_{Loukas} = Q_l^{+} \dots Q_1^{+}$
- $P_{rao} = L_c^{+} Q^T L$

Inspired from:

$$\arg\min_{P} \sup_{x \in \mathbb{R}^N, \|x\|_L = 1} \|x - QPx\|_L$$

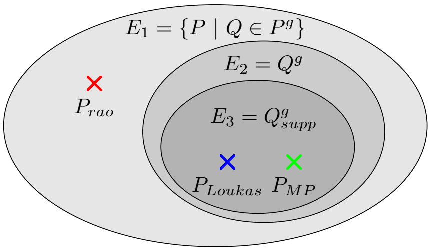
RSA :

$$\arg\min_{P} \sup_{x \in \mathcal{R}, \|x\|_L = 1} \|x - QPx\|_L$$

[2] C Radhakrishna Rao, Sujit Kumar Mitra, et al. *Generalized inverse of a matrix and its applications*, Proceedings of the sixth Berkeley symposium on mathematical statistics and probability, 1972.

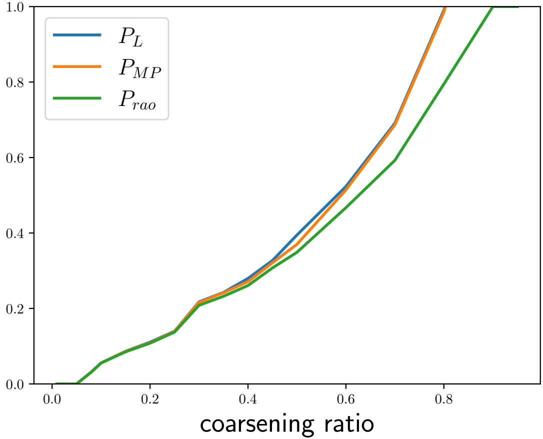
Example of reduction matrices

- $P_{MP} = Q^+ = (Q^\top Q)^{-1} Q^\top$
- $P_{Loukas} = Q_l^+ \dots Q_1^+$
- $P_{rao} = L_c^+ Q^T L$



[2] C Radhakrishna Rao, Sujit Kumar Mitra, et al. *Generalized inverse of a matrix and its applications*, Proceedings of the sixth Berkeley symposium on mathematical statistics and probability, 1972.

Results for the RSA



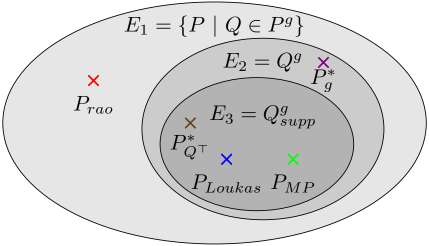
Conclusion

Key messages:

- Q contains all the structural information of the coarsened graph
- P has a degree of freedom and we propose a taxonomy of the admissible reduction matrices

Outlooks :

- Optimizing the RSA in E_2 and E_3



Conclusion

Learn more about optimizing the RSA and its application to GNN in our new preprint available : `https://arxiv.org/abs/2506.11743`

Appendices

References

[1] Andreas Loukas, *Graph Reduction with Spectral and Cut Guarantees*, JMLR 2019.

[2] C Radhakrishna Rao, Sujit Kumar Mitra, et al. *Generalized inverse of a matrix and its applications*, Proceedings of the sixth Berkeley symposium on mathematical statistics and probability, volume 1, pages 601–620. University of California Press Oakland, CA, USA, 1972.

[3] Roger Penrose. *A generalized inverse for matrices*. In Mathematical proceedings of the Cambridge philosophical society, volume 51, pages 406–413. Cambridge University Press, 1955.

Adaption of Loukas coarsening algorithm

Algorithm Loukas algorithm Adapted

Require: Adjacency matrix A, Laplacian $L = f_L(A)$, propagation matrix S, a coarsening ratio r , preserved space $\mathcal{R}$, maximum number of nodes merged at one coarsening step : n_e

```
n_obj ← int(N − N × r) the number of nodes wanted at the end of the algorithm.
compute cost matrix B_0 ← V V^T L^{-1/2} with V an orthonormal basis of R
Q ← I_N
while n ≥ n_obj do
    Make one coarsening STEP l
    Create candidate contraction sets.
    For each contraction C, compute cost(C, B_{l-1}, L_{l-1}) = ||Π_C B_{l-1} (B_{l-1}^T L_{l-1} B_{l-1})^{-1/2}||_{L_C} / (|C| − 1)
    Sort the list of contraction set by the lowest score
    Select the lowest scores non overlapping contraction set while the number of nodes merged is inferior to min(n − n_obj, n_e)
    Compute Q_l, Q_l^+, uniform intermediary coarsening with contraction sets selected
    B_l ← Q_l B_{l-1}
    Q ← Q_l Q
    A_l ← (Q_l^+)^⊤ A_{l-1} Q_l^+ − diag((Q_l^+)^⊤ A_{l-1} Q_l^+) 1_n)
    L_{l-1} = f_L(A_{l-1})
    n ← min(n − n_obj, n_e)
end while
Compute S_c^{MP} = PSQ
return P, Q, S_c^{MP}
```

Optimizing RSA in E_2 and E_3

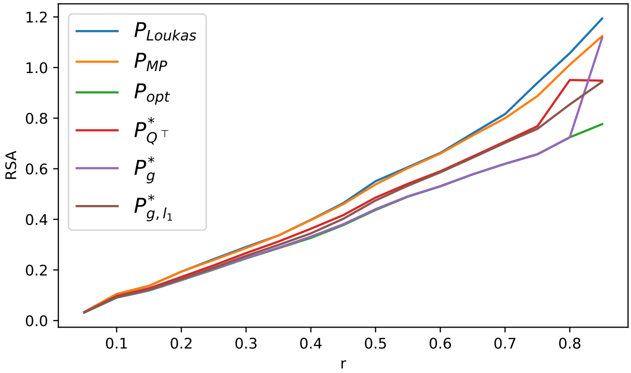


Figure: RSA optimization for Cora graph, combinatorial Laplacian $\mathcal{L}$

Training on coarsened graph procedure

Algorithm Training Procedure

Require: Adjacency A, node features X, desired propagation matrix S, preserved space $\mathcal{R}$, Laplacian L, a coarsening ratio r

P, Q, $S_c^{MP} \leftarrow$ Coarsening-algorithm$(A, L, S, r, \mathcal{R})$
$X_c \leftarrow PX$
Initialize model (SGC or GCNconv)
for N_{epochs} iterations **do**
  compute coarsened prediction $\Phi_\theta(S_c^{MP}, X_c)$
  uplift the predictions : $Q\Phi_\theta(S_c^{MP}, X_c)$
  compute the cross entropy loss $J(Q\Phi_\theta(S_c^{MP}, X_c))$
  Backpropagate the gradient
  Update θ
end for

Training GNN on G_c (experiments)

Table: Accuracy in % for node classification with SGC and GCNconv on different coarsening ratio

SGC	Cora			Citeseer		
r	0.3	0.5	0.7	0.3	0.5	0.7
P_{Loukas}	80.5 ± 0.0	79.7 ± 0.0	76.8 ± 0.0	72.6 ± 0.3	71.7 ± 0.1	**69.7** ± 0.7
P_{MP}	80.5 ± 0.0	**80.1** ± 0.0	77.7 ± 0.0	72.8 ± 0.5	72.7 ± 0.0	69.5 ± 0.3
P_{opt}	77.1 ± 0.6	75.9 ± 0.1	73.8 ± 0.3	70.9 ± 0.2	70.2 ± 0.1	67.3 ± 0.4
$P^*_{Q^\top}$	80.3 ± 0.0	80.0 ± 0.1	77.2 ± 0.0	72.7 ± 0.3	72.6 ± 0.5	67.6 ± 0.2
P^*_g	**80.7** ± 0.0	80.0 ± 0.0	77.6 ± 0.0	72.6 ± 0.2	**72.7** ± 0.0	68.6 ± 0.4
P^*_{g,l_1}	80.4 ± 0.0	79.2 ± 0.0	**78.3** ± 0.0	**73.0** ± 0.0	71.2 ± 0.1	69.2 ± 0.4
Full Graph		81.0 ± 0.1			71.6 ± 0.1	
GCN	**Cora**			**Citeseer**		
r	0.3	0.5	0.7	0.3	0.5	0.7
P_{Loukas}	**80.6** ± 0.8	80.5 ± 1.0	78.1 ± 1.4	71.0 ± 1.6	72.2 ± 0.6	70.4 ± 0.8
P_{MP}	80.4 ± 1.0	80.7 ± 0.9	78.6 ± 0.9	70.8 ± 1.9	72.1 ± 1.0	**71.0** ± 1.0
P_{opt}	73.7 ± 1.5	63.3 ± 1.4	55.11 ± 2.4	64.6 ± 0.7	50.4 ± 1.6	42.6 ± 4.0
$P^*_{Q^\top}$	80.5 ± 0.9	80.9 ± 0.6	78.0 ± 0.9	**71.1** ± 1.5	**72.3** ± 0.7	70.0 ± 0.9
P^*_g	**80.6** ± 1.1	**81.3** ± 0.6	**78.7** ± 0.9	**71.1** ± 1.7	72.1 ± 1.2	69.6 ± 1.0
P^*_{g,l_1}	80.4 ± 0.9	80.0 ± 0.9	78.2 ± 0.7	70.2 ± 1.8	66.8 ± 1.1	66.7 ± 1.2
Full Graph		81.3 ± 0.8			70.9 ± 1.4	

Dataset Presentation

Dataset	# Nodes	# Edges	# Features	#classes
Cora PCC	2,485	10,138	1,433	7
Cora70	746	3,716	1,433	7
Citeseer PCC	2,120	7,358	3,703	6
Citeseer70	636	2,122	3,703	6