University of Witwatersrand
School of Computer Science and Applied Mathematics

Persistent Neural Guidance

for the Quadratic Assignment Problem

A learned initialization, fed through a nuclear-norm ascent that pushes a relaxed assignment toward the vertices of the Birkhoff polytope.

Anesu Pisira, MSc Candidate

Mathematical Optimization and Control for Industrial and
Societal Challenges in Sub-Saharan Africa · ICMS · Pretoria, Sept 2026

The Quadratic Assignment Problem

Anesu Pisira
MSc Candidate

- Assign n facilities to n locations, given a flow matrix F and a distance matrix D
- Objective: minimize total flow-weighted distance
- NP-hard — and the flow term makes the search landscape genuinely rugged, not just large

$$\min_X \ \mathrm{tr}\left(FXD^TX^T\right) \qquad X \text{ a permutation matrix}$$

Today's talk is about the solver how a learned signal and a classical proximal method combine to search this landscape.

Anesu Pisira
MSc Candidate

Five stages, one trajectory

1

GNN + message passing

F, D become graphs. Embeddings Z_F, Z_D emerge.

→

2

Combine

Inner products + Sinkhorn → X_init, doubly-stochastic.

→

3

Nuclear-norm ascent

Proximal iterations pull X toward a vertex of B_n.

→

4

Round

Hungarian algorithm snaps the fractional X to 0/1.

→

5

Refine

Local perturbations polish the rounded solution.

Stages 1–2 run once. Stage 3 iterates. Stages 4–5 run once at the end.

Anesu Pisira
MSc Candidate

F and D become graphs

- F and D are each written as a weighted graph — nodes are facilities (for F) or locations (for D), edges carry the flow or distance weight
- A message-passing GNN runs several rounds: each node aggregates a weighted signal from its neighbors and updates its own embedding
- **Output: Z_F and Z_D — learned vector embeddings, one per facility and one per location**

This is standard GNN machinery — the departure from convention starts at the next stage.

GNN + msg pass	Combine	Nuclear-norm ascent	Round	Refine

Anesu Pisira
MSc Candidate

From embeddings to a starting matrix

$$\text{score}(i, j) = Z_F[i] \cdot Z_D[j]$$

an inner product per facility–location pair — no learning happens in this step, just combination

Sinkhorn: alternate row- and column-normalization

$$X \leftarrow \text{row-normalize}(X) \quad \rightarrow \quad X \leftarrow \text{col-normalize}(X) \quad \rightarrow \quad \text{repeat}$$

$X_{init} \in B_n$ **(the Birkhoff polytope: doubly-stochastic matrices)**

A soft, learned starting point — not yet a permutation, but already informed by both F and D.

GNN + msg pass	Combine	Nuclear-norm ascent	Round	Refine

Anesu Pisira
MSc Candidate

A relaxation, plus one unconventional term

$$\min \quad \operatorname{tr}(F\, X\, D^{T}\, X^{T}) \quad - \quad \|X\|_{*} \qquad\qquad \text{s.t.} \quad X \in B_{n}$$

$\operatorname{tr}(F\, X\, D^{T}\, X^{T})$

The classical Koopmans–Beckmann QAP relaxation ; total flow-weighted distance, now over doubly-stochastic X rather than a hard permutation.

$- \|X\|_{*}$

The unconventional term. Maximizing the nuclear norm pushes the relaxed solution toward the discrete vertices of B_n — next slide.

Anesu Pisira
MSc Candidate

Nuclear norm, used backwards

Usual role: minimize ‖X‖_* to encourage low rank (matrix completion, recovery)

Here: maximize ‖X‖_* over X ∈ B_n — to push toward full rank, structured extremes

1. Birkhoff–von Neumann: every doubly-stochastic matrix is a convex combination of permutation matrices — the vertices of B_n.

2. The nuclear norm is convex. By Jensen's inequality, a convex function over a convex combination is maximized at the extreme points — the vertices themselves.

3. **A permutation matrix is orthogonal, so ‖P‖_* = n exactly — the maximum attainable value on B_n.**

So: maximizing ‖X‖_* is a principled pull toward the discrete vertices — not a heuristic penalty bolted on.

Anesu Pisira
MSc Candidate

The learned start

`-‖X‖_* is concave, so it's linearized at each iterate via its subgradient:`

$$G_t = U_t V_t^T \quad \text{from the economy SVD of } X_t$$

Recomputed fresh at every iteration — G_t is not just G_0 held constant. So where does "persistent" come from?

G_0 is built from the SVD of the GNN's own X_init — not a random or uniform start. Every subsequent G_t is derived from a trajectory that began there. The GNN's signal doesn't just seed iteration 1; it shapes the ascent direction the whole path follows.

Anesu Pisira
MSc Candidate

What's actually solved at each iteration

$$\min_{Y \in \mathcal{B}_n} \langle \nabla f(X_t), Y - X_t \rangle + \frac{\mu_t}{2} \|Y - X_t\|_F^2 - \lambda \langle G_t, Y \rangle$$

$$\langle \nabla f(X_t), Y - X_t \rangle$$

Linearized QAP objective — the direction that reduces flow-weighted distance.

$$\frac{\mu_t}{2} \|Y - X_t\|_F^2$$

Proximal / trust-region term — keeps each step close to the current iterate.

$$-\lambda \langle G_t, Y \rangle$$

Pull toward the current nuclear-norm ascent direction — the GNN-seeded push toward B_n's vertices.

$$\nabla f(X) = FXD^T + F^T XD$$

GNN + msg pass | **Combine** | **Nuclear-norm ascent** | **Round** | **Refine**

Anesu Pisira
MSc Candidate

From fractional matrix to permutation

Round: Hungarian algorithm

- The proximal loop terminates — fixed iteration budget or near-stationarity — leaving a fractional $X \in B_n$, not yet a hard assignment.
- The Hungarian algorithm finds the optimal 0/1 permutation matrix nearest to this fractional X in linear-assignment cost.

Refine: local perturbation

- Rounding can leave value on the table — the nearest permutation isn't always the best nearby one.
- A small number of local swaps, with perturbation, checks the immediate neighborhood for improvements before accepting the final answer.

This is the one place classical metaheuristic search re-enters the pipeline — everything before it is learned + convex proximal iteration.

GNN + msg pass	Combine	Nuclear-norm ascent	Round	Refine

Instance	n	Optimal	HyGQAP	Time(s)	BMA	SWAT
Chr12a	12	9552	9552	2	0.2	0.4%/8.1s
Chr18a	18	11098	11098	14	0.4	0.4%/8.1s
Esc16a	16	68	68	2	0.0	9.2%/26.8s
Esc16b	16	292	292	2	0.0	9.2%/26.8s
Esc16c	16	160	160	2	0.0	9.2%/26.8s
Had12	12	1652	1652	2	0.2	10.8%/12.11s
Had14	14	2724	2724	2	0.1	10.8%/12.11s
Had16	16	3720	3720	3	0.3	10.8%/12.11s
Src15	15	51140	51140	10	0.3	-
Src20	20	110030	110030	23	0.5	-
Nug15	15	1150	1150	3	0.2	-
Nug20	20	2570	2570	7	0.4	-
Nug25	25	3744	3744	11	0.6	-
Tai12b	12	39464925	39464925	5	0.0	-
Tai15a	15	388214	388214	13	0.2	-
tai15b	15	51765268	51765268	12	0.1	-

Anesu Pisira
MSc Candidate

Any assignment problem with a flow and a distance

Warehouse slotting

Item-to-location assignment, re-solved often as demand shifts.

Hub-and-spoke design

Regional network design — QAP's most direct real-world descendant.

Facility / hospital layout

New-build siting decisions where the assignment is still genuinely open.

Chip / PCB placement

Classical QAP territory — minimizing wire length between connected components.

The solver doesn't know or care what the facilities and locations represent — only F, D, and the polytope. It is therefore most useful where a strict permutation with guaranteed optimality is required.

Anesu Pisira
MSc Candidate

A solver is not an integration

A correct, fast solver answers “what's the best assignment.” It doesn't answer whether that assignment is deployable:

Legacy systems

Existing scheduling / ERP software rarely accepts an external optimizer's output without custom integration work.

Sunk structures

Real facilities have physical, contractual, and capital constraints the model doesn't see.

Validation gap

A domain expert needs to sanity-check every non-obvious swap before it's trusted operationally.

What would help most for implementation:

- A real, bounded pilot problem — small enough to validate end-to-end
- A domain collaborator who can say which swaps are actually infeasible, not just costly

Anesu Pisira
MSc Candidate

One learned signal, one principled pull, one classical finish

- A GNN reads F and D once, producing an initialization that isn't thrown away
- The nuclear norm, maximized rather than minimized, gives a principled reason to push toward the Birkhoff polytope's vertices
- Every proximal step is anchored to a trajectory the learned start actually shaped
- Hungarian rounding and local refinement close the gap from fractional to discrete

Thank You

Any Questions?

Anesu Pisira, MSc Candidate

Persistent Neural Guidance for the Quadratic Assignment Problem