Self-Supervised Graph Attention Networks for Community-Engaged Lead Contamination Risk Assessment

Supplementary Material

Raphael Anaadumba[1], Nazim A. Belabbaci[1], Yigit Bozkurt[2], Connor Sullivan[2], Pradeep Kurup[2,*], and Mohammad Arif Ul Alam[1,*]

[1]Richard A Miner School of Computer and Information Sciences, University of Massachusetts Lowell, Lowell, USA

[2]Civil and Environmental Engineering, University of Massachusetts Lowell, Lowell, USA

pradeep_kurup@uml.edu mohammadariful_alam@uml.edu

Algorithm

Algorithm 1: Within-City Validation Protocol (LOWO) in Flint

Input: Ward-partitioned dataset $\{\mathcal{D}_1, \ldots, \mathcal{D}_9\}$; seeds $\mathcal{S} = \{1, \ldots, 5\}$; $K = 5$ folds; max epochs $E = 100$; patience $p = 5$
Output: LOWO metrics for each held-out ward w (mean ± std over seeds)
for each held-out ward $w \in \{1, \ldots, 9\}$ **do**
  $\mathcal{D}^{\text{test}} \leftarrow \mathcal{D}_w$; $\mathcal{D}^{\text{train}} \leftarrow \bigcup_{u \neq w} \mathcal{D}_u$
  Construct parcel-based training and test graphs using the same deterministic parcel-induced rule
    $G^{\text{train}} \leftarrow (\mathcal{D}^{\text{train}}, E^{\text{train}})$, $G^{\text{test}} \leftarrow (\mathcal{D}^{\text{test}}, E^{\text{test}})$ {Parcel-induced edges only; held-out test nodes are excluded from training-time message passing.}
  for each seed $s \in \mathcal{S}$ **do**
    Initialize K-fold split of $\mathcal{D}^{\text{train}}$ with shuffle=True, random_state=s
    best_val $\leftarrow +\infty$; $\theta^*_{w,s} \leftarrow \emptyset$
    for each fold $k = 1$ to K **do**
      Split $\mathcal{D}^{\text{train}}$ into $\mathcal{D}^{\text{train}}_k$ and $\mathcal{D}^{\text{val}}_k$
      Create node masks for $\mathcal{D}^{\text{train}}_k$ and $\mathcal{D}^{\text{val}}_k$ on G^{train}; build NeighborSampler loaders using only the training graph
      Initialize GAT model with reconstruction head (SSGAT) and Adam optimizer
        using the fixed hyperparameter configuration from the main manuscript
      for epoch $e = 1$ to E **do**
        Update parameters using $\mathcal{L}_{\text{Total}} = \alpha\mathcal{L}_{\text{NLL}} + \gamma\mathcal{L}_{\text{Recon}}$ on training nodes
        Compute validation loss on $\mathcal{D}^{\text{val}}_k$
        Apply early stopping with patience p
      end for
      if validation loss for fold $k <$ best_val **then**
        best_val $\leftarrow$ validation loss; $\theta^*_{w,s} \leftarrow$ model parameters for this fold
      end if
    end for
    Evaluate $\theta^*_{w,s}$ on the held-out ward nodes $\mathcal{D}^{\text{test}}$ and store metrics (Acc., Macro-Prec., Macro-Rec.)
    Save checkpoint $\theta^*_{w,s}$ as `checkpoint_{w,s}.pth`
  end for
  Aggregate test metrics for ward w as mean ± std across $s \in \mathcal{S}$
end for
Return: Ward-wise LOWO performance summary (mean ± std over 5 seeds)

Leave One Ward Out	Acc.	Prec.	Rec.	Acc.	Prec.	Rec.	Acc.	Prec.	Rec.
	Multi-Class			**Current Action**			**Historic Action**		
Ward 1	0.72 ± 0.04	0.64 ± 0.01	0.68 ± 0.04	0.78 ± 0.06	0.57 ± 0.01	0.62 ± 0.01	0.78 ± 0.09	0.56 ± 0.06	0.65 ± 0.13
Ward 2	0.68 ± 0.11	0.57 ± 0.03	0.60 ± 0.05	0.70 ± 0.11	0.53 ± 0.03	0.57 ± 0.05	0.76 ± 0.06	0.54 ± 0.01	0.64 ± 0.03
Ward 3	$\mathbf{0.76 \pm 0.08}$	$\mathbf{0.60 \pm 0.10}$	$\mathbf{0.64 \pm 0.14}$	$\mathbf{0.84 \pm 0.08}$	$\mathbf{0.60 \pm 0.07}$	$\mathbf{0.70 \pm 0.15}$	$\mathbf{0.89 \pm 0.06}$	$\mathbf{0.62 \pm 0.05}$	$\mathbf{0.78 \pm 0.16}$
Ward 4	0.68 ± 0.07	0.54 ± 0.03	0.55 ± 0.04	0.76 ± 0.08	0.51 ± 0.05	0.50 ± 0.05	0.83 ± 0.07	0.53 ± 0.04	0.55 ± 0.05
Ward 5	0.70 ± 0.04	0.64 ± 0.05	0.69 ± 0.07	0.70 ± 0.07	0.60 ± 0.03	0.66 ± 0.03	$\mathbf{0.84 \pm 0.09}$	0.68 ± 0.12	0.78 ± 0.08
Ward 6	0.65 ± 0.07	0.56 ± 0.02	0.58 ± 0.03	0.65 ± 0.08	0.52 ± 0.03	0.56 ± 0.08	0.77 ± 0.08	0.56 ± 0.04	0.65 ± 0.09
Ward 7	0.61 ± 0.04	0.54 ± 0.02	0.55 ± 0.03	0.78 ± 0.09	0.52 ± 0.04	0.52 ± 0.04	0.75 ± 0.11	0.52 ± 0.02	0.54 ± 0.03
Ward 8	0.66 ± 0.09	0.55 ± 0.03	0.54 ± 0.02	0.74 ± 0.07	0.52 ± 0.04	0.53 ± 0.07	0.79 ± 0.13	0.53 ± 0.02	0.56 ± 0.07
Ward 9	0.68 ± 0.03	0.58 ± 0.03	0.60 ± 0.04	0.74 ± 0.02	0.56 ± 0.03	0.61 ± 0.07	0.75 ± 0.07	0.51 ± 0.02	0.54 ± 0.06

Table 1: **Within-City Validation (Self-Supervised).** Ward-wise performance summary across lead contamination risk classification settings: Multi-Class, EPA Current Action, and Historic Action. Results are reported as mean ± standard deviation across five independent seeds.

Leave-One-Ward-Out	Acc.	Prec.	Rec.	Acc.	Prec.	Rec.	Acc.	Prec.	Rec.
	Multi-Class			**Current Action**			**Historic Action**		
Ward 1	0.64 ± 0.05	0.54 ± 0.03	0.55 ± 0.04	0.72 ± 0.08	0.54 ± 0.04	0.57 ± 0.07	0.71 ± 0.11	0.54 ± 0.03	0.61 ± 0.04
Ward 2	0.70 ± 0.09	0.54 ± 0.05	0.56 ± 0.06	0.76 ± 0.10	0.54 ± 0.03	0.60 ± 0.07	0.82 ± 0.08	0.52 ± 0.02	0.54 ± 0.03
Ward 3	0.78 ± 0.08	0.62 ± 0.09	0.67 ± 0.12	0.80 ± 0.09	0.56 ± 0.04	0.63 ± 0.09	0.89 ± 0.06	0.66 ± 0.11	0.84 ± 0.14
Ward 4	0.64 ± 0.07	0.54 ± 0.03	0.56 ± 0.05	0.80 ± 0.08	0.52 ± 0.04	0.54 ± 0.04	0.86 ± 0.03	0.54 ± 0.04	0.56 ± 0.08
Ward 5	0.69 ± 0.05	0.67 ± 0.04	0.71 ± 0.03	0.72 ± 0.10	0.63 ± 0.05	0.75 ± 0.09	0.84 ± 0.09	0.66 ± 0.07	0.79 ± 0.01
Ward 6	0.64 ± 0.10	0.57 ± 0.06	0.58 ± 0.04	0.72 ± 0.05	0.55 ± 0.03	0.60 ± 0.05	0.71 ± 0.07	0.55 ± 0.02	0.60 ± 0.03
Ward 7	0.63 ± 0.09	0.53 ± 0.02	0.54 ± 0.03	0.82 ± 0.05	0.54 ± 0.04	0.53 ± 0.05	0.80 ± 0.15	0.51 ± 0.03	0.53 ± 0.06
Ward 8	0.67 ± 0.11	0.54 ± 0.02	0.54 ± 0.02	0.69 ± 0.07	0.50 ± 0.01	0.50 ± 0.01	0.78 ± 0.07	0.53 ± 0.03	0.57 ± 0.08
Ward 9	0.65 ± 0.05	0.54 ± 0.04	0.56 ± 0.05	0.78 ± 0.05	0.55 ± 0.04	0.56 ± 0.04	0.72 ± 0.08	0.54 ± 0.01	0.61 ± 0.04

Table 2: **Within-City Validation (Supervised).** Ward-wise supervised GAT performance under the Leave-One-Ward-Out (LOWO) protocol. Results are reported as mean ± standard deviation across five independent seeds.

Leave-One-Ward-Out	Acc.	Prec.	Rec.	Acc.	Prec.	Rec.	Acc.	Prec.	Rec.
	Multi-Class			**Current Action**			**Historic Action**		
Ward 1	0.81 ± 0.0013	0.68 ± 0.0131	0.53 ± 0.0028	0.90 ± 0.0005	0.89 ± 0.0320	0.53 ± 0.0003	0.93 ± 0.0000	0.63 ± 0.0000	0.51 ± 0.0000
Ward 2	0.82 ± 0.0023	0.58 ± 0.0165	0.52 ± 0.0047	0.92 ± 0.0005	0.62 ± 0.0107	0.51 ± 0.0003	0.95 ± 0.0006	0.83 ± 0.0231	0.53 ± 0.0061
Ward 3	0.83 ± 0.0014	0.68 ± 0.0115	0.54 ± 0.0024	0.92 ± 0.0000	0.84 ± 0.0000	0.53 ± 0.0000	0.95 ± 0.0008	0.83 ± 0.1370	0.52 ± 0.0004
Ward 4	0.82 ± 0.0025	0.76 ± 0.0262	0.55 ± 0.0032	0.90 ± 0.0007	0.74 ± 0.0187	0.52 ± 0.0029	0.93 ± 0.0004	0.85 ± 0.0298	0.52 ± 0.0002
Ward 5	0.80 ± 0.0015	0.71 ± 0.0135	0.54 ± 0.0031	0.89 ± 0.0014	0.75 ± 0.0345	0.52 ± 0.0043	0.92 ± 0.0009	0.73 ± 0.0278	0.53 ± 0.0062
Ward 6	0.79 ± 0.0014	0.67 ± 0.0113	0.53 ± 0.0022	0.89 ± 0.0006	0.65 ± 0.0144	0.51 ± 0.0003	0.93 ± 0.0005	0.74 ± 0.0196	0.52 ± 0.0003
Ward 7	0.79 ± 0.0018	0.63 ± 0.0122	0.53 ± 0.0021	0.90 ± 0.0005	0.68 ± 0.0149	0.52 ± 0.0016	0.93 ± 0.0004	0.80 ± 0.0245	0.52 ± 0.0002
Ward 8	0.79 ± 0.0016	0.65 ± 0.0160	0.52 ± 0.0020	0.89 ± 0.0000	0.80 ± 0.0000	0.52 ± 0.0000	0.93 ± 0.0000	0.84 ± 0.0000	0.51 ± 0.0000
Ward 9	0.79 ± 0.0020	0.66 ± 0.0154	0.53 ± 0.0020	0.89 ± 0.0009	0.65 ± 0.0139	0.52 ± 0.0022	0.93 ± 0.0004	0.71 ± 0.0125	0.53 ± 0.0002

Table 3: **Within-City Validation (Random Forest).** Ward-wise random forest performance under the Leave-One-Ward-Out protocol. Results are reported as mean ± standard deviation across five independent runs.

Leave-One-House-Out	Acc.	Prec.	Rec.	Acc.	Prec.	Rec.	Acc.	Prec.	Rec.
	Multi-Class			Current Action			Historic Action		
1	0.78 ± 0.04	0.58 ± 0.05	0.63 ± 0.04	0.73 ± 0.05	0.51 ± 0.04	0.56 ± 0.05	0.69 ± 0.09	0.59 ± 0.06	0.64 ± 0.06
2	0.81 ± 0.03	0.63 ± 0.05	0.68 ± 0.05	0.75 ± 0.03	0.52 ± 0.02	0.57 ± 0.03	0.69 ± 0.05	0.60 ± 0.02	0.68 ± 0.04
3	0.78 ± 0.04	0.60 ± 0.04	0.65 ± 0.04	0.73 ± 0.02	0.51 ± 0.02	0.55 ± 0.02	0.71 ± 0.06	0.62 ± 0.05	0.67 ± 0.05
4	0.79 ± 0.01	0.59 ± 0.02	0.65 ± 0.02	0.74 ± 0.02	0.52 ± 0.03	0.55 ± 0.04	0.73 ± 0.05	0.63 ± 0.05	0.69 ± 0.05
5	$\mathbf{0.80 \pm 0.02}$	$\mathbf{0.61 \pm 0.04}$	$\mathbf{0.66 \pm 0.03}$	$\mathbf{0.72 \pm 0.02}$	$\mathbf{0.51 \pm 0.03}$	$\mathbf{0.55 \pm 0.04}$	$\mathbf{0.72 \pm 0.03}$	$\mathbf{0.62 \pm 0.05}$	$\mathbf{0.68 \pm 0.05}$
6	0.79 ± 0.03	0.60 ± 0.04	0.65 ± 0.04	0.72 ± 0.02	0.53 ± 0.03	0.57 ± 0.04	0.68 ± 0.02	0.57 ± 0.03	0.62 ± 0.03
7	0.81 ± 0.02	0.60 ± 0.01	0.65 ± 0.02	0.72 ± 0.02	0.52 ± 0.04	0.57 ± 0.04	0.74 ± 0.05	0.60 ± 0.06	0.66 ± 0.05
8	0.80 ± 0.02	0.62 ± 0.02	0.67 ± 0.02	0.74 ± 0.03	0.52 ± 0.04	0.56 ± 0.03	0.72 ± 0.03	0.63 ± 0.04	0.69 ± 0.03
9	0.78 ± 0.01	0.61 ± 0.02	0.66 ± 0.02	0.73 ± 0.05	0.52 ± 0.06	0.57 ± 0.07	0.71 ± 0.06	0.62 ± 0.03	0.68 ± 0.04

Table 4: **Cross-City Validation (Andover LOHO).** Cross-city validation results under the Leave-One-House-Out (LOHO) protocol in Andover. Results are reported as mean ± standard deviation across five independent seeds.