Graph Neural Networks for Predicting Wastewater Service Type At A Land Parcel Level

Sylvia Imanirakiza[1] **Nana Akua Agyemang Sereboo**[1]
Nelson da Luz[2] **Emily Kumpel**[2] **Jay Taneja**[1]
[1]Manning College of Information & Computer Sciences, UMass Amherst
[2]Department of Civil & Environmental Engineering, UMass Amherst
{simanirakiza,nagyemangser,ndaluz-dw,ekumpel,jtaneja}@umass.edu

Abstract

America lacks a unified, parcel-scale map of wastewater service-type infrastructure; the last nationally comprehensive collection was in 1990, and while newer sources exist, they are decentralized and often non-spatial, complicating integration. In this paper, we present an empirical study evaluating Graph Neural Networks (GNNs) for node classification of buried sanitation infrastructure mapping. Using one million parcel-level sanitation records from Florida, we construct sampled graphs of 1k, 5k, and 10k nodes, where each land parcel is modeled as a graph node with spatial and functional edges.We evaluate three classical GNNs in both transductive and inductive settings. The findings demonstrate that graph structure captures meaningful spatial dependencies in wastewater infrastructure.

1 Introduction

Most wastewater systems are either centralized sewers that convey waste to a treatment plant or decentralized onsite systems (septic tanks) that treat waste locally [1]. Despite their substantial health [2, 3], environmental [4–6] and economic [7] benefits, the United States (U.S.) lacks up-to-date comprehensive parcel-level data on who is served by which type of wastewater system, limiting risk management, investment planning, and sustainability efforts. Since the last comprehensive national collection of wastewater infrastructure data for the U.S. in 1990, estimates of onsite system use vary widely, and state datasets are sparse and inconsistent, making the lack of reliable parcel-level data a critical challenge to prioritizing risks and buried infrastructure management.

Advances in machine learning have improved water and wastewater applications (condition assessment [8, 9], defect detection [10], computer vision mapping [11–13] amongst others, but *locating* buried sanitation systems remains an open problem. Most recent work uses a two-stage Random Forest to predict parcel service type, achieving 91.8% in Florida and 81.9% out-of-sample in Virginia, while considering parcels independently and without exploiting the networked layout of sanitation systems (roads, treatment plant proximity, neighbourhood design) [14].

In this work, we address this problem by formulating the task as node classification (sewer or septic) on graphs, and evaluating three standard GNNs: Graph Convolutional Networks (GCN) [15], GraphSAGE (SAmple and aggreGatE)[16], and Graph Attention Networks (GAT)[17], as shown in Figure 1. The base observational unit in our study is the land parcel, defined as a distinct piece of land designated for tax purposes. Our analysis focuses on parcels located within the state of Florida. **Our contributions are:** (1) a compact, scalable graph construction that uses feature similarity and spatial context across multiple sample sizes and (2) a comparative study based on two training frameworks for graph neural networks, inductive and transductive setting.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: UrbanAI.

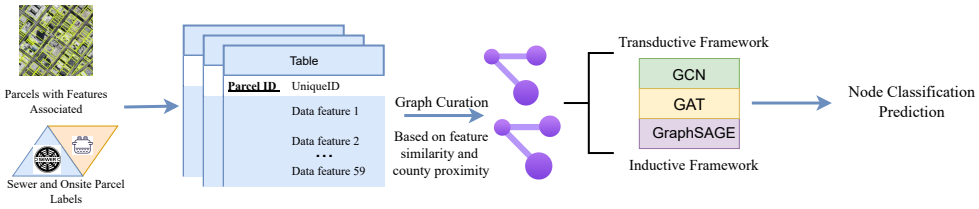


Figure 1: We infer graph structures across multiple samples and sizes where each parcel of land represents a node, and edges are defined based on spatial and functional relationships. We comparatively evaluate the performance of graph neural networks under transductive and inductive frameworks.

2 Data

We use the pre-processed parcel-level indicator tabular dataset from da Luz et al [14]. The target labels are the type of parcel service from the Florida Water Management Inventory [18]. Each land parcel node has 59 features as described in [14], including road-network metrics from U.S. Census TIGER/Line Roads (2023) [19], distance to the nearest wastewater treatment plant from EPA FRS/HIFLD [20], building counts and areas from Microsoft building footprints [21]), housing characteristics from the 2021 ACS [22], incorporation status from TIGER/Line Places (2023) [19], and parcel value from Regrid [23]. Each row corresponds to one parcel with its feature vector and service-type label. We create a graph structure with nodes representing parcels of land and edges capturing spatial and functional relationships.

3 Methodology

3.1 Graph Structure Creation

We generate graph structures from the tabular data focusing on Florida with over a million records representing various attributes of land parcels. We created graph representations for different subsets of parcels (1,000, 5,000, 10,000 and 50,000), sampling across five distinct folds. This sampling was performed to balance computational requirements with representational accuracy for the model. Land parcels are modelled as nodes, labels indicate sewer versus onsite service, and graph edges were constructed using two main strategies:

Feature similarity. We build a k-Nearest Neighbors (k-NN) graph using cosine similarity, connecting parcels whose cosine distance is ≤ 0.2. This links properties with similar wastewater-related attributes and yields a meaningful neighbourhood structure. The threshold must be tuned: if too strict, useful medium-strength links are missed; if too loose, weak noisy links are added [24].

County-Based Proximity. We also connect parcels in the same county to encode local spatial context, potential county-level policy effects and to complement feature-based links for predictive modeling.

All graphs are connected, except for a single 1,000-node fold that contains one isolated node we retain since it does not affect results.

3.2 Model Implementation

In this work, we employ three classical GNN architectures: GCN, GraphSAGE, and GAT. We conduct hyperparameter tuning using 5-fold cross-validation for each model. The Adam optimizer is used with an initial learning rate of 0.001 and an epoch limit of 100. We tune the hidden dimensions from 16, 32, 64, and 128, the learning rates from 0.001, 0.005, and 0.01, and the number of attention heads for GAT from 1, 2, and 4. An exhaustive search is conducted across all datasets, and the best hyperparameters are selected based on the highest average validation accuracy across different samples and node configurations. For GCN, we use 64 hidden dimensions, learning rate of 0.01 and weight decay of 5e-4 with no dropout; for GAT,we use 16 hidden dimensions, learning rate of 0.01, weight decay of 5e-4 and a singular attention head, and for GraphSAGE with a mean aggregator,we use 64 hidden dimensions, learning rate of 0.01 and weight decay of 5e-4.

Experimental Setup We design experiments using two primary frameworks

In the first, the transductive framework, the entire graph is used as input, and masking techniques are applied to split the data into training, validation, and test sets with ratios of 0.8, 0.1, and 0.1, respectively. Features are normalized to prevent data leakage. This setup assumes that some node labels are known during training, while the masked nodes represent unknown labels.

In the second case, the inductive framework, the model is trained on one graph without masked nodes, validated on a distinct set of nodes, and tested on an entirely different graph. This approach mimics real-world situations where models need to generalize across multiple graphs from various regions.

Evaluation Model performance is evaluated using classification accuracy. We compare their performance against the baseline Multi-Layer Perceptron (MLP) and Random Forest from [14] which do not take the graph structure into account. All experiments are conducted using 2 NVIDIA A40 GPUs (132 GB) and an 8-core CPU, for dataset curation and model training. The data curation compute time varies with the number of nodes, the average duration for graphs of size <10k was approximately 30 minutes to 2 hours, and scaled exponentially with increasing number of nodes. For model training, we trained each model in parallel for 100 epochs, this had an average compute <30 minutes per model.

4 Results

We compare GNN performance on graph structures derived from urban sanitation data in terms of classification performance. The results are reported in Table 1 and 2 for the transductive and inductive settings respectively.

From Table 1 for the transductive setting,we observe the following; GNN models (GCN, GraphSAGE, GAT) consistently outperform the baseline MLP across all graph sizes. For 50k nodes, the GraphSAGE model achieves an accuracy comparable to the random forest proposed by [14]. GraphSAGE exhibits the best performance across all node sizes except for the smallest graph (1k nodes), where GCN marginally surpasses it. GraphSAGE demonstrates higher stability (lower standard deviation) and scalability, achieving an accuracy of 96.42%. GCN performs well on smaller graphs (1k and 5k nodes), achieving the highest accuracy for the 1k-node dataset (83.80%). However, its performance saturates slightly on larger graphs, suggesting potential limitations in scalability. GAT lags behind GraphSAGE and GCN for smaller graphs but performs fairly well for mid-sized graphs (10k nodes) with and accuracy of 93.60%. This model exhibits higher performance variability (higher standard deviation) indicating that attention mechanisms might introduce sensitivity to graph size or training conditions.

Table 1: Performance of GNN models and the baseline MLP in the **transductive setting**. The values indicate the mean and standard deviation accuracy across 5 distinct samples of a given number of parcels(nodes), excluding 50k nodes, where we used one sample. Boldened values indicate best performance for each sample.

Model	1k Nodes	5k Nodes	10k Nodes	50k Nodes
GCN	**83.80% ± 6.83**	93.12% ± 1.04	94.62% ± 0.61	91.72%
GAT	81.20% ± 1.92	88.68% ± 2.15	93.60% ± 2.83%	-
GraphSAGE	83.20 % ± 2.59	**93.60% ± 0.58**	**94.72% ± 0.53**	**96.42**
Baseline MLP	83.20% ± 7.05	83.84% ± 1.66	83.64% ± 0.53	84.52% ± 0.76
RF [14]	-	-	-	93.5%

Table 2: Performance of GNN models and the baseline MLP in the **inductive setting**. The values represent the accuracy of each model when trained on a graph structure with n nodes from Sample Fold 1, validated on a graph structure from Sample Fold 2, and tested on a graph structure from Sample Fold 3. Bold values indicate the best performance.

Model	1k Nodes	5k Nodes	10k Nodes
GCN	64.50	78.54%	76.56%
GAT	69.30%	75.22%	74.47%
GraphSAGE	**81.60** %	**89.52%**	**91.49%**

From Table 2 for the inductive setting,we observe the following; GraphSAGE consistently outperforms both GCN and GAT across all graph sizes. For 1k nodes, it achieves 81.60% accuracy, a significant margin over the other GNN architectures, GAT (69.30%) and GCN (64.50%). For 5k and 10k nodes, its performance increases to 89.52% and 91.49%, respectively. This emphasizes the suitability of the GraphSAGE for the inductive setting, likely due to its ability to learn representations that generalize well to unseen graph structures. GCN exhibits the poorest performance across all graph sizes: Its accuracy is 64.50% for 1k nodes, improving to 78.54% for 5k nodes but declining slightly to 76.56% for 10k nodes. This highlights the structural dependency of the GCN, which makes it less effective in inductive settings. GAT performs better than GCN for smaller graphs (69.30% vs. 64.50% for 1k nodes). Its accuracy for 5k nodes is 75.22%, and for 10k nodes, it slightly declines to 74.47%. This suggests that while attention mechanisms offer advantages for capturing localized structural information, GAT may struggle to generalize across varying graph structures in inductive scenarios.

5 Discussion

Scalability As the number of nodes increases, the number of edges grows quadratically, leading to increasingly complex structures. For instance, the largest graph (50k nodes) has over 41 million edges compared to 17k edges in the smallest graph (1k nodes). The increasing density and connectivity of larger graphs require scalable GNN architectures. Models need to handle high computational demands for node representation learning as the graph size grows. The scalability of structure-based architectures such as GNNs remains an open problem in general since they are yet to show the benefits of scale mainly due to the lower efficiency of sparse operations, large data requirements [25] [26].

Generalization in Inductive Settings GraphSAGE's better performance in comparison to other models in inductive settings can be attributed to its use of neighborhood aggregation that utilizes the node attributes effectively, rather than only depending on graph topology. The GCN's dependency on fixed graph structure reduces its effectiveness when transferring to new graph structures, making it less efficient for inductive tasks. GAT's attention mechanism may struggle to generalize across graph structures, suggesting a need for better regularization, training strategies and graph representation learning in inductive settings. The performance gains observed for all models as graph size increases highlights the potential of improving generalization and model performance by training on larger graphs (using the entire dataset of over 1 million parcels).

Implications: The GNN approaches take advantage of the inherent networks involved in sanitation system construction, linked with roads, wastewater treatment facilities and neighborhood design. The RF approach in [14] treats each parcel independently, preventing exploitation of the networked layout for improving predictions and further network analysis. Additionally, the inductive setting is much more appropriate for the out-of-sample predictions that lays groundwork for contributing to a wastewater infrastructure national inventory [14].

6 Limitations and Future Work

While the results show that performance improves as graph size increases, training large graphs with millions of nodes remains a significant challenge. We would like to explore optimizing GNN architectures for scalability, such as leveraging techniques like graph sampling, mini-batching, or distributed training to handle large-scale graphs efficiently.

7 Conclusion

We present preliminary findings for the effectiveness of Graph Neural Networks (GNNs) for node classification in mapping sanitation infrastructure, addressing a gap in wastewater mapping data. Our findings reveal GraphSAGE as a highly effective and scalable model, outperforming GCN and GAT in inductive and transductive settings. While GCN performs well in smaller, fixed graph scenarios and GAT shows potential for diverse structures, their limitations highlight opportunities for future improvement. This research shows the potential of GNNs to extract and leverage complex spatial and functional relationships in urban infrastructure data, offering a significant step toward supporting wastewater mapping systems.

References

[1] World Health Organization (WHO) and United Nations Children's Fund (UNICEF). Sanitation | jmp. URL https://washdata.org/topics/sanitation. Accessed: 2021-09-27.

[2] Kartiki S. Naik and Michael K. Stenstrom. Evidence of the influence of wastewater treatment on improved public health. *Water Science and Technology*, 66(3):644–652, 2012. doi: 10.2166/wst.2012.144. URL https://doi.org/10.2166/wst.2012.144.

[3] Annette Prüss-Ustün, Jennyfer Wolf, Jamie Bartram, Thomas Clasen, Oliver Cumming, Matthew C. Freeman, Bruce Gordon, Paul R. Hunter, Kate Medlicott, and Richard Johnston. Burden of disease from inadequate water, sanitation and hygiene for selected adverse health outcomes: An updated analysis with a focus on low- and middle-income countries. *International Journal of Hygiene and Environmental Health*, 222(5):765–777, 2019. doi: 10.1016/j.ijheh.2019.05.004. URL https://doi.org/10.1016/j.ijheh.2019.05.004.

[4] Natacha Brion, Michel A. Verbanck, Willy Bauwens, Marc Elskens, Margaret Chen, and Pierre Servais. Assessing the impacts of wastewater treatment implementation on the water quality of a small urban river over the past 40 years. *Environmental Science and Pollution Research*, 22(16):12720–12736, 2015. doi: 10.1007/s11356-015-4493-8. URL https://doi.org/10.1007/s11356-015-4493-8.

[5] I. de Guzman, A. Elosegi, D. von Schiller, J. M. González, L. E. Paz, B. Gauzens, U. Brose, A. Antón, N. Olarte, J. M. Montoya, and A. Larrañaga. Treated and highly diluted, but wastewater still impacts diversity and energy fluxes of freshwater food webs. *Journal of Environmental Management*, 345:118510, 2023. doi: 10.1016/j.jenvman.2023.118510. URL https://doi.org/10.1016/j.jenvman.2023.118510.

[6] David A. Keiser and Joseph S. Shapiro. Consequences of the clean water act and the demand for water quality. *The Quarterly Journal of Economics*, 134(1):349–396, 2019. doi: 10.1093/qje/qjy019. URL https://doi.org/10.1093/qje/qjy019.

[7] Richard Damania, Sébastien Desbureaux, Aude-Sophie Rodella, Jason Russ, and Esha Zaveri. *Quality Unknown: The Invisible Water Crisis*. World Bank, Washington, DC, 2019. ISBN 978-1-4648-1459-4. doi: 10.1596/978-1-4648-1459-4. URL https://openknowledge.worldbank.org/entities/publication/9880744c-2411-54c2-801f-daa56ab15865.

[8] N. Caradot, M. Riechel, M. Fesneau, N. Hernandez, A. Torres, H. Sonnenberg, E. Eckert, N. Lengemann, J. Waschnewski, and P. Rouault. Practical benchmarking of statistical and machine learning models for predicting the condition of sewer pipes in berlin, germany. *Journal of Hydroinformatics*, 20(5):1131–1147, 2018. doi: 10.2166/hydro.2018.217.

[9] Y. Kleiner, R. Sadiq, and B. Rajani. Modelling the deterioration of buried infrastructure as a fuzzy markov process. *Journal of Water Supply: Research and Technology—Aqua*, 55(2):67–80, 2006. doi: 10.2166/aqua.2006.074.

[10] Xianfei Yin, Yuan Chen, Ahmed Bouferguene, Hamid Zaman, Mohamed Al-Hussein, and Luke Kurach. A deep learning-based framework for an automated defect detection system for sewer pipes. *Automation in Construction*, 109:102967, 2020. doi: 10.1016/j.autcon.2019.102967.

[11] Billie F Spencer Jr, Vedhus Hoskere, and Yasutaka Narazaki. Advances in computer vision-based civil infrastructure inspection and monitoring. *Engineering*, 5(2):199–222, 2019.

[12] Navneet P Singh and Manisha J Nene. Buried object detection and analysis of gpr images: Using neural network and curve fitting. In *2013 Annual International Conference on Emerging Research Areas and 2013 International Conference on Microelectronics, Communications and Renewable Energy*, pages 1–6. IEEE, 2013.

[13] Xisto L Travassos, Sérgio L Avila, and Nathan Ida. Artificial neural networks and machine learning techniques applied to ground penetrating radar: A review. *Applied Computing and Informatics*, 17(2):296–308, 2021.

[14] Nelson da Luz, Jay Taneja, and Emily Kumpel. Look Out Below: Predicting Wastewater Infrastructure Service Type at the Land Parcel Scale. *ACS ES&T Engineering*, October 2025. doi: 10.1021/acsestengg.5c00637. URL https://doi.org/10.1021/acsestengg.5c00637. Publisher: American Chemical Society.

[15] Thomas N. Kipf and Max Welling. Semi-Supervised Classification with Graph Convolutional Networks, February 2017. URL http://arxiv.org/abs/1609.02907. arXiv:1609.02907 version: 4.

[16] William L. Hamilton, Rex Ying, and Jure Leskovec. Inductive Representation Learning on Large Graphs, September 2018. URL http://arxiv.org/abs/1706.02216. arXiv:1706.02216 version: 4.

[17] Shaked Brody, Uri Alon, and Eran Yahav. How Attentive are Graph Attention Networks?, January 2022. URL http://arxiv.org/abs/2105.14491. arXiv:2105.14491.

[18] Florida Department of Health. Florida water management inventory project. Available at:http://www.floridahealth.gov/environmental-health/drinking-water/flwmi/index.html, 2021. Accessed: 2021-10-28.

[19] U.S. Census Bureau. Tiger/line shapefiles, 2024. URL https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html. Accessed: 2025-01-21.

[20] Florida Department of Environmental Protection. Wastewater facility regulation (wafr) - wastewater facilities. Available at:https://www.floridadep.gov/geospatial-open-data, 2018. Accessed: 2021-10-28.

[21] Overture Maps Foundation. Building footprints. Available at:https://overturemaps.org/, 2024. Accessed: 2024-07-17.

[22] US Census Bureau. About the american community survey. Available at: https://www.census.gov/programs-surveys/acs/about.html, 2021. Accessed: 2021-10-28.

[23] Regrid. The regrid parcel schema, 2024. URL https://support.regrid.com/parcel-data/schema. Accessed: 2024-12-13.

[24] Yihang Zhang, Yinfeng Fang, Xixia Yu, Yong Peng, and Zhaojie Ju. Cosine similarity graph attention networks for autism spectrum disorder diagnosis. In *2024 World Rehabilitation Robot Convention (WRRC)*, pages 1–5, 2024. doi: 10.1109/WRRC62201.2024.10696658.

[25] Marco Serafini and Hui Guan. Scalable graph neural network training: The case for sampling. *SIGOPS Oper. Syst. Rev.*, 55(1):68–76, June 2021. ISSN 0163-5980. doi: 10.1145/3469379.3469387. URL https://doi.org/10.1145/3469379.3469387.

[26] Maciej Sypetkowski, Frederik Wenkel, Farimah Poursafaei, Nia Dickson, Karush Suri, Philip Fradkin, and Dominique Beaini. On the scalability of gnns for molecular graphs, 2024. URL https://arxiv.org/abs/2404.11568.

A Technical Appendices and Supplementary Material

A.1 Data Source and Data Aggregation

Defining the geographical observational unit: The base observational unit in this work refers to the land parcel (i.e a distinct piece of land identified for taxation purposes). We perform the calculations for the features(indicator data) for/in relation to individual land parcels. Regrid, a national land parcel data provider, provides the geospatial boundaries for parcels with a standardized data scheme throughout the US.

Attribute Data : For feature engineering for this task, similarly to [14], we utilized domain-informed indicator features that describe the parcel characteristics and would be relevant to infer wastewater infrastructure. In some cases, we combine publicly available datasets at broader scales to generalize the processed features to the observational unit. Adapted from [14], table 3 summarizes the datasets and features used to generate parcel-level variables

Table 3: Summary of datasets and features used for parcel-level analysis.

Attribute	Data Source	Description
Parcel value	[23]	Parcel boundaries and total parcel value; missing values imputed using county or state averages.
Roadways	[19]	Road length and count by type in grids overlaid on parcels; parcel-level values assigned via spatial join.
Treatment plant locations	[18]	Distance from each parcel centroid to the nearest treatment plant.
Building locations and areas	[21]	Parcel-level building counts, total and maximum building area, median building size, and non-building area ratio, calculated via grid overlay and spatial join.
US Census ACS data	[19]	Housing variables (e.g., year built, bedrooms, rooms, household size) aggregated at the Census Block Group level and assigned to parcels based on containing CBG.
US Census places	[19]	Parcels labeled as incorporated, CDP, or undesignated based on intersection (>50% area) with Places polygons.

A.2 Graph Structure Creation

The graph construction was derived by treating each parcel of land as a node and defining relationships between nodes as edges based on specific criteria. Whether a property is connected to a sewer or has an onsite septic system was mapped to labels for classification, with nodes representing individual properties.

The graph edges were constructed using two main strategies:

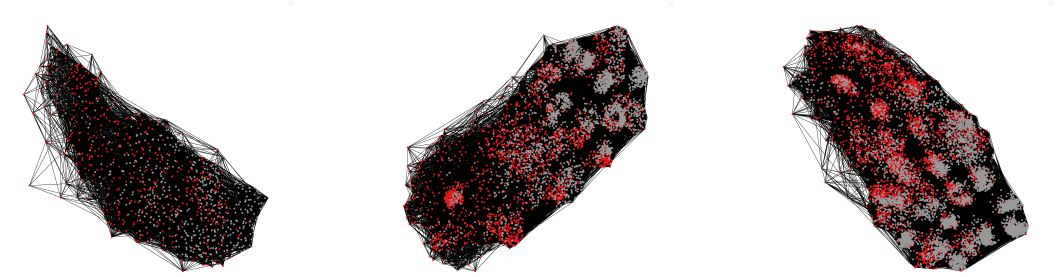
- Feature Similarity: Nodes (properties) were connected based on the similarity of their attributes. A K-Nearest Neighbors (KNN) approach with cosine similarity was employed to connect nodes that had a cosine distance of 0.2 or less. This similarity metric captures properties with similar wastewater characteristics and other relevant features, facilitating the creation of a meaningful neighborhood structure. It is important to choose a cosine distance threshold that is neither too high nor too low [24]. If the threshold is set too high, it captures only the most strongly correlated values, potentially missing meaningful but weaker connections. On the other hand, if set too low, it risks including weakly correlated values, which could introduce noise and reduce the relevance of the identified correlations [24].

- County-Based Proximity: An additional set of edges was established by linking properties within the same county. This spatial proximity criterion ensures that geographically adjacent properties influence each other, capturing local wastewater distribution patterns and potential county-level policy impacts. These two approaches allowed for the formation of edges that capture both attribute-based similarities and spatial relationships, creating a network suitable for predictive modeling.

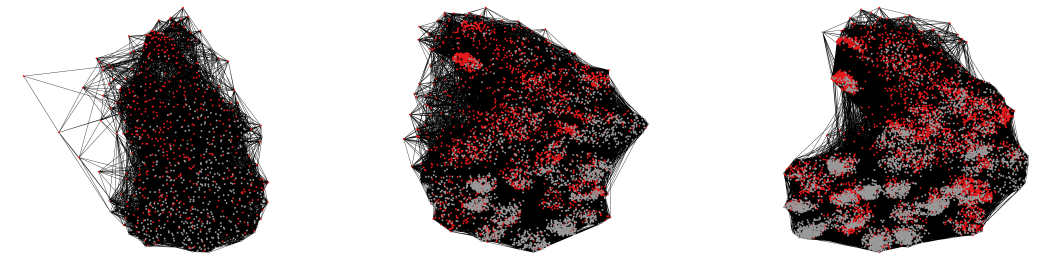
Table 4 summarizes the graphical structures generated using the similarity approach, which incorporates functional (attribute-based) and spatial (county-based) relationships. The datasets includes 59 attributes for each parcel, extracted from the original dataset. Figure 2 shows the visualization of some of the different graph structures.

Table 4: Overview of the graphical structures generated across five sampled folds from the records.

Dataset ID	Number of nodes	Average Num_edges	Stdev Num_edges
Sample(SF) _nrows1k_data	1000	24,767	549
Sample(SF) _nrows5k_data	5000	453,967	8485
Sample(SF) _nrows10k_data	10000	1.737,158	25003
Sample(SF) _nrows50k_data	50000	41,672,307	245914



(a) Sample(1) nrows1k graph (b) Sample(1) nrows5k graph (c) Sample(1) nrows10k graph



(d) Sample(5) nrows1k graph (e) Sample(5) nrows5k graph (f) Sample(5) nrows10k graph

Figure 2: Visualization of the generated similarity network based graph structures with varying configurations.Red: septic system, Grey: Sewer system **a)** The graph structure for dataset sample, Sample(1) nrows1k data with 1000 nodes, 24,163 edges, **b)** The graph structure for dataset sample, Sample(1) nrows5k data with 5000 nodes, 451,513 edges,**c)** The graph structure for dataset sample Sample(1) nrows10k data with 10,000 nodes, 1,727,813 edges, **d)** The graph structure for dataset sample Sample(5) nrows1k data with 1000 nodes, 24,559 edges, **e)**The graph structure for dataset sample Sample(5) nrows5k data with 5000 nodes, 442,797 edges, **f)** The graph structure for dataset sample Sample(5) nrows10k data with 10,000 nodes, 1,720,927 edges.