

Urban Region Representation Learning: A Positional and Structural Graph Approach

Germano B. Santos , Fabrício A. Silva 

¹Universidade Federal de Viçosa (UFV)

gegenbarcelos@gmail.com, fabricio.asilva@ufv.br

Abstract. *Urban environments are composed of intricate spatial structures and movement patterns, impacting various urban elements like transportation, land use, and economic activity. Firstly, in this dissertation, we review the geospatial artificial intelligence literature on urban region representation models, categorizing them based on learning paradigms, spatial data modalities, and architectural design. Furthermore, we highlight key challenges on region representation models such as the data quality issue, which some datasets are incomplete, presenting missing labeled data. Additionally, urban region models overlook spatial heterophily, where adjacent regions can have varied functions, being an intrinsic characteristic of geographic data. To resolve this, we present HAVANA (Hybrid Attentional Graph Convolutional Network for Semantic Venue Annotation), a model using Graph Neural Networks and human mobility data to advance POI classification improving geographic data quality. Building upon this improved data representation, we introduce a novel spatial heterophily-aware Graph Transformer named FisherGT which is incorporated on HAMURE (Heterophily-Aware Urban Multi-View-Based Region Embedding), a self-supervised multi-stage model. Hence, HAMURE improves representation quality, yielding better outcomes for land use clustering, crime prediction, population density estimation.*

1. Introduction

Cities are complex systems composed of interconnected environments, including neighborhoods, road networks, and amenities such as shopping centers and pharmacies. These elements create a dynamic and livable space for human activities [Dong et al. 2024]. Delineating a city is a challenging task due to its intricate characteristics.

A city can be defined by its physical structure and inherent activities. Form and function are fundamental characteristics that shape the spatial structure and functionality of a region, which can be represented through geographic entities and human mobility patterns [Reia et al. 2025]. While geographic entities, including points of interest (POIs), building footprints, and railway lines, describe the physical structures within a region; human mobility captures movement patterns between the origin and destination places. These aspects reveal latent urban characteristics, helping to describe urban dynamics [Yang et al. 2023], and comprehend human lifestyles [Ma et al. 2024].

With increasing computational capabilities, multimodal models have emerged as a solution to develop general-purpose urban region representation models that can be widely applied to various downstream tasks such as crime prediction [Zhang et al. 2023a, Zhang et al. 2023b], and population density estimation [Li et al. 2023]. Consequently, these models provide valuable insights for urban analytics and planning

[Klemmer et al. 2023, Guo et al. 2024, Li et al. 2024]. Given the evolution of urban region representation models, it has become crucial to comprehensively review, categorize, and analyze these approaches to identify current limitations and future research directions.

Therefore, through a comprehensive review on urban region representation models, we identified two primary challenges data quality and spatial heterophily. These two issues can degrade model performance and limit its usability to diverse cities, resulting in suboptimal outcomes for urban applications.

Regarding quality data, a key issue lies in missing labels for points of interests which impact the performance of spatially-explicit models. With the increasing availability of spatial and mobile phone data, semantic venue annotation can leverage from urban signature composed of human mobility patterns [Ye et al. 2011], spatial neighborhood proximity [Yao et al. 2018], and geographic characteristics such as roads and rivers [Niu and Silva 2021].

Additionally, urban areas exhibit spatial heterogeneity, where neighboring places can play complementary roles when comparing functional categories [Du et al. 2019]; in other words, close places may show different functions. This is a unique feature of geographic space that is often overlooked by urban region representation models that encode cities as graphs. Vanilla Graph Neural Networks are based on homophily assumption, which posits that nearby nodes should be similar - this conflicts with the Third Law of Geography [Zhu and Turner 2022]. Hence, urban region representation models need to incorporate spatial heterophily in order to capture the variety of spatial configurations that can occur even among adjacent areas.

2. Contributions

The primary objective of this dissertation is to contribute to the fields of semantic venue annotation and urban region representation learning by developing and evaluating advanced graph-based models. The specific objectives are:

- Provide a comprehensive review of urban region representation learning models. This review categorizes existing methodologies on urban region representation learning based on data modalities, architectural design, and learning paradigms, identifying key challenges and future research directions.
- Investigate the use of Graph Neural Networks (GNNs) for semantic venue annotation. This includes evaluating whether human mobility patterns and spatial relationships improve POI categorization accuracy, enhancing data quality.
- Incorporate spatial heterogeneity into urban region representation learning through FisherGT, a novel Graph Transformer aligned with the Third Law of Geography.

In summary, this dissertation seeks to advance the field of urban analytics by developing methodologies that enhance both semantic venue annotation and urban region representation learning, thus enhancing GeoAI.

2.1. Urban Region Representation Learning: A Comprehensive Review

Compared to classical fields like images and speech, the study of spatial characteristics of the environment has gained significant attention in recent times, leading to the emergence of an area known as spatially-explicit geographic artificial intelligence (GeoAI)

[Grekousis 2019]. GeoAI utilizes spatial data to enhance model performance across various issues in urban dynamics, urban area differentiation, and social sensing fields [Liu and Biljecki 2022]. The examination of urban dynamics encompasses the concept of spatial transformation over time, influenced by shifts in human mobility patterns [Yang et al. 2023] and the availability of amenities [Bittencourt and Giannotti 2021].

Comprehending an urban region is essential since it facilitates the association between static attributes and city dynamics. Therefore, urban region representation learning aims to represent an urban area in a vector space considering region intricacies being a challenging computational endeavor due to variety of geographic entities. While urban region representation learning is vital in geographic artificial intelligence, existing reviews fall short in discussing region representation methods.

Hence, we comprehensively review about 40 methods that focus on learning urban region characteristics to apply on downstream tasks such as crime prediction, land use clustering, health index prediction and we noticed that data quality and spatial heterophily are two crucial challenges when modeling spatially-explicit models. Additionally, we categorize these works under a novel taxonomy (Spatial Data, Learning Approaches and Architecture Style). This survey is under review.

The first aspect of taxonomy, i.e., spatial data, encloses Human Mobility, Spatial Entities, Satellite Images and Geo-Tagged Social Media Data. The second aspect, Learning Approches, is divided into Reconstruction, Discriminative, Generative and Hybrid. Lastly, Architecture Style separates the literature over Word-Like, Graph Embedding, Convolutional Neural Network, Transformer and Matrix Decomposition.

2.2. HAVANA: Hybrid Attentional Graph Convolutional Network Semantic Venue Annotation Model

Location-based social networks (LBSNs) capture the intrinsic mobility characteristics of visitation patterns, providing a vast source of data. Nowadays, there are various LBSNs data sources available, such as Yelp, Foursquare, and Gowalla, which frequently have missing features, and incorrect or inconsistent labels [Psyllidis et al. 2022], leading to errors in different domains, such as site selection and urban planning. Therefore, as noticed in survey and [Miliadis and Psyllidis 2021], it is necessary to develop methods that can accurately label the points of interests (POIs), a problem also known as semantic venue annotation, since the place category is essential for gaining in-depth knowledge of user displacements [Ma et al. 2024]. Furthermore, improve data quality, enhance indirectly urban region representation models.

Therefore, we present our contribution, published in BRACIS 2025, a novel model named HAVANA [dos Santos et al. 2024] (**H**ybrid **A**ttentional **G**raph **C**onvolution for **S**emantic **V**enue **A**nnotation) that integrates spatial and spectral graph convolution to enhance the accuracy of semantic venue annotation. Specifically, our model combines the spatial and spectral convolution, represented respectively by Graph Attention Network (GAT) [Veličković et al. 2017] and Auto-Regressive Moving Average (ARMA) [Bianchi et al. 2021], into a hybrid block. The output of this novel block is aggregated using self-attention [Vaswani et al. 2017] and a two-layer feedforward network with residual connections.

2.3. Fisher Graph Transformers: Geometry-Aware Attention for Spatial Representation Learning

Following the comprehensive review, spatial heterophily is overlooked on urban region representation models, although it is an important characteristic already noted by some works [Xiao et al. 2023, Du et al. 2019, Zhu and Turner 2022]. Therefore, we propose a novel Graph Transformer named FisherGT based on Fisher Embedding Information Matrix approximation using Jacobian Vector Products to improve the attention weights. This novel method FisherGT replace the Euclidean Attention incorporated into scaled-dot-product to Fisher Elliptical Attention inducing an anisotropic interaction geometry between embeddings. This novel method is under review.

In order to tackle spatial heterophily within geographic environment, we incorporated it on a multi-stage urban region representation learning model named HAMURE that comprises a fine-grained stage to capture local correlations from points of interests and building footprints within a region and a coarse-grained stage to capture global information from region-level embeddings.

HAMURE leverages a Fisher Graph Transformer, capturing spatial long-range dependencies [Di Giovanni et al. 2023] while attending to diverse spatial configuration even on close neighborhoods. Moreover, the fine-grained stage focuses on structural local properties of a graph by applying random-walk encoding [Zhang et al. 2018], while Laplacian positional encoding [Kreuzer et al. 2021] is used in the coarse-grained stage to capture a broader spatial context of each region. Additionally, we introduce a self-supervised method that combines three distinct region graph views while preserving intra-view similarity with a joint learning module. Furthermore, we utilize data from OpenStreetMap (OSM) to address challenges related to data availability.

3. Indirect Contributions

Besides the direct contributions presented in dissertation, a model for semantic venue annotation, a comprehensive review and an urban region representation model for handling spatial heterophily with Graph Transformers, there were also indirect contributions which were proposed during master's.

Firstly, we propose a parallel human mobility mining framework [Santos et al. 2025], which served as a tool to extract semantic motifs to understand the shifting behavior of mobility during the easing restrictions of COVID-19. Secondly, we evaluate the site planning of electric charging stations to heavy-duty vehicles using a proposed simulation of queues and capacity of hexagonal network and discrete location models, such as Hexagonal P-Median and Hexagonal Capacitated Location Set Covering Model [dos Santos et al. 2025].

Although these works were not explicitly included in the dissertation, they were projected and published during the master's program. They intersect with urban analytics and contribute to the understanding of urban characteristics, being valuable to compose the scope of this dissertation.

4. Results

In this Section, we present the performance of HAVANA considering the average F1-Score on Texas, California and Florida, and HAMURE on three downstream tasks in

New York City, Chicago, and Singapore compared to state-of-the-art approaches. We also include the results obtained when HAMURE is used with FisherGT compared to Euclidean Transformers (ExpFormer [Shirzad et al. 2023], SGFormer [Wu et al. 2023], Transformer [Dwivedi and Bresson 2020]), and Hyperbolic Transformer (HypFormer [Yang et al. 2024]).

4.1. HAVANA

The results for each category shown by Table 1 demonstrate that HAVANA surpasses the other methods in almost all scenarios, suggesting that the mechanisms adopted, i.e., attention weights and skip-connections, yielded notable performance on the semantic venue annotation task.

Furthermore, the performance on different states suggest that the method proposed exhibits strong generalizability to different mobility patterns and city organizations, as evidenced by improvements in the analyzed metrics. The consistent performance indicates that HAVANA possesses high adaptability to diverse spatio-temporal city characteristics, including the imbalanced category place distribution, as well as the variability of check-ins and users quantities.

		HAVANA	PGC-NN	STPA	k-FN
Florida	Community	30.20 ± 2.0	11.92 ± 7.6	42.94 ± 13.1	13.95 ± 0.4
	Entertainment	81.93 ± 0.9	71.69 ± 1.9	30.50 ± 8.6	9.45 ± 0.2
	Food	70.47 ± 0.6	62.99 ± 0.8	41.92 ± 1.1	28.94 ± 0.3
	Nightlife	60.71 ± 2.0	38.20 ± 4.0	23.54 ± 7.7	5.71 ± 0.4
	Outdoors	61.18 ± 1.6	52.58 ± 1.6	34.45 ± 8.3	6.27 ± 0.3
	Shopping	58.30 ± 1.0	44.47 ± 5.2	53.83 ± 3.6	23.38 ± 0.2
	Travel	77.47 ± 0.9	70.53 ± 1.0	33.64 ± 7.2	11.27 ± 0.4
California	Community	23.63 ± 1.7	19.08 ± 1.4	43.32 ± 9.5	14.77 ± 0.1
	Entertainment	57.55 ± 1.0	50.44 ± 8.1	26.92 ± 8.4	6.98 ± 0.1
	Food	62.23 ± 0.1	59.99 ± 0.6	44.60 ± 4.1	34.13 ± 0.1
	Nightlife	48.89 ± 0.6	18.28 ± 4.3	19.24 ± 5.6	5.28 ± 0.3
	Outdoors	39.32 ± 1.5	32.08 ± 7.9	36.39 ± 9.7	7.72 ± 0.1
	Shopping	32.98 ± 0.8	17.68 ± 3.3	51.07 ± 7.1	21.69 ± 0.2
	Travel	63.93 ± 0.8	60.44 ± 1.9	25.53 ± 10.9	9.85 ± 0.2
Texas	Community	29.46 ± 1.7	15.00 ± 4.3	49.58 ± 6.7	17.39 ± 0.2
	Entertainment	67.02 ± 0.5	55.79 ± 8.5	30.42 ± 7.7	8.36 ± 0.1
	Food	72.56 ± 0.4	63.78 ± 2.2	43.58 ± 4.8	32.60 ± 0.1
	Nightlife	77.95 ± 0.7	63.07 ± 2.9	24.38 ± 5.0	7.40 ± 0.2
	Outdoors	54.08 ± 0.8	39.94 ± 5.9	31.94 ± 9.0	5.42 ± 0.2
	Shopping	50.77 ± 0.9	24.78 ± 12.6	50.51 ± 3.8	22.73 ± 0.1
	Travel	66.49 ± 1.3	61.23 ± 1.1	24.93 ± 10.0	5.71 ± 0.2

Table 1. Average F1-Score (%) by model, category, and state.

It is still noteworthy to highlight that STPA, PGC-NN, and k-FN do not present satisfactory results, since they exhibit a high variance on F1-Score within different states. Considering that, this lack of consistency in results suggests a higher sensibility to distinct states and spatio-temporal characteristics. On the other hand, as the proposed model comprises a hybrid convolution and weighting attention mechanism, it demonstrates a higher consistency in different contexts. Then, these performance gains compared to the state-of-the-art model reinforce the efficiency of the novel method.

4.2. FisherGT

We conduct experiments on New York City, Chicago and Singapore using embeddings from HAMURE and the SOTA region representation models to contiguously cluster re-

Models	New York City	
	ARI \uparrow	NMI \uparrow
POI-Encoder	0.278	0.593
DGI	0.289	0.601
GAE	0.089	0.345
HGI	0.349	0.641
HAFusion	0.304	0.602
RegionDCL	0.301	0.631
CityFM	0.304	0.626
HAMURE-FisherGT	0.397	0.679
HAMURE-Transformer	0.346	0.631
HAMURE-ExpFormer	0.313	0.622
HAMURE-SGFormer	0.291	0.617
HAMURE-HypFormer	0.350	0.651

Table 2. Land Usage Clustering Performance. Highlighted are the top first, second, third

gions (land use clustering), and for downstream tasks such as crime prediction and population density estimation. To quantitatively evaluate the quality of the resulting clusters, we compare them against known community boundaries using Normalized Mutual Information (NMI) and Adjusted Rand Index (ARI). Higher scores indicate stronger alignment between the clustered spatial communities and real-world administrative. For crime prediction and population density estimation we analyze Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 .

4.2.1. Land Usage Clustering

Regarding Normalized Mutual Information, HAMURE outperforms other methods as shown in Table 2, indicating its strong alignment with region clusters. Methods like RegionDCL, CityFM, and HAFusion perform well but fall short compared to HAMURE, highlighting the complexity of regionalization tasks. POI-Encoder and DGI yield lower scores ($NMI < 0.61$, $ARI < 0.29$), showing weaker spatial organization in their embeddings. Note that the Hyperbolic Transformer variant also outperform all state-of-the-art approaches, which can be attributed to the hyperbolic attention integrated into the model. This finding indicates that the Land Use Clustering task can benefit from hierarchical modeling and the use of hyperbolic space, although elliptical attention surpass hyperbolic attention.

4.2.2. Crime Prediction

Similarly, crime prediction demonstrates spatial heterogeneity, as certain types of crimes are more likely to occur in a specific area [Wang et al. 2016]. Models like RegionDCL, which rely heavily on long-range dependencies, and do not account for local spatial correlations explicitly, may overweight distant neighbors with limited influence, reducing their capability to capture these localized patterns effectively.

HAMURE with FisherGT achieves the strongest performance in New York City among all compared methods. This indicates that Fisher Elliptical Attention improves the separability of local spatial clusters in urban environments, a property that is not effectively captured by other geometric spaces such as Euclidean or Hyperbolic. In contrast, the Fisher-based variant does not outperform HGI in Chicago — although it remains competitive — because HGI is especially effective at capturing local spatial correlations through its area-based regularization, particularly in Chicago. The Spearman correlation

Models	Chicago			New York City		
	RMSE ↓	MAE ↓	R2 ↑	RMSE ↓	MAE ↓	R2 ↑
POI-Encoder	6865.58 ± 1422.87	4853.19 ± 547.63	0.151 ± 0.078	118.82 ± 31.20	87.59 ± 13.85	0.311 ± 0.016
DGI	7000.49 ± 1474.94	4935.26 ± 411.46	0.120 ± 0.093	130.89 ± 29.70	94.65 ± 11.59	0.212 ± 0.016
GAE	7241.67 ± 1481.74	5153.21 ± 414.53	0.029 ± 0.046	129.84 ± 33.10	93.73 ± 16.18	0.173 ± 0.012
HGI	5989.65 ± 1102.95	4120.60 ± 437.60	0.341 ± 0.120	116.99 ± 28.16	84.39 ± 13.19	0.365 ± 0.019
RegionDCL	6827.63 ± 1433.99	4767.28 ± 364.36	0.160 ± 0.097	119.92 ± 30.58	86.86 ± 13.56	0.375 ± 0.011
HA Fusion	7549.97 ± 1237.57	5444.31 ± 385.77	0.032 ± 0.030	135.24 ± 36.14	95.60 ± 14.53	0.093 ± 0.027
CityFM	6679.08 ± 1471.02	4629.56 ± 390.03	0.198 ± 0.063	121.03 ± 30.03	88.67 ± 13.67	0.300 ± 0.017
HAMURE-FisherGT	6340.31 ± 1499.70	4508.04 ± 541.80	0.283 ± 0.085	111.12 ± 28.17	78.52 ± 12.82	0.488 ± 0.018
HAMURE-Transformer	6756.74 ± 1457.64	4720.02 ± 396.34	0.184 ± 0.060	116.93 ± 29.74	82.70 ± 9.83	0.346 ± 0.017
HAMURE-Expformer	6362.97 ± 1397.42	4557.36 ± 340.84	0.273 ± 0.092	120.20 ± 33.06	86.30 ± 12.18	0.291 ± 0.021
HAMURE-SGFormer	7279.66 ± 1487.94	5258.26 ± 396.68	0.051 ± 0.039	128.89 ± 33.66	95.77 ± 16.29	0.240 ± 0.007
HAMURE-Hypformer	6755.19 ± 1425.48	4809.35 ± 355.81	0.182 ± 0.061	131.66 ± 36.13	94.09 ± 14.66	0.060 ± 0.077

Table 3. Crime Prediction Performance. Highlighted are the top first, second, third

between area and crime count is higher in Chicago (0.23) than in New York City (0.04), supporting the idea that incorporating area into the model can lead to more information leakage in Chicago.

4.2.3. Population Density Estimation

Regarding population density estimation, it is important to note the performance contrast between HAMURE, HGI and RegionDCL. HGI outperforms other baselines methods which can be attributed to HGI’s regularization of final output with region area, thus leaking the density information for this task.

Models	Singapore			New York City		
	RMSE ↓	MAE ↓	R2 ↑	RMSE ↓	MAE ↓	R2 ↑
POI-Encoder	8626.89 ± 1442.36	6314.69 ± 1023.34	0.290 ± 0.173	8315.72 ± 606.42	6366.65 ± 398.25	0.406 ± 0.006
DGI	8233.45 ± 1341.23	6196.14 ± 997.85	0.356 ± 0.139	8016.65 ± 378.85	6004.78 ± 207.23	0.408 ± 0.063
GAE	7689.51 ± 1235.03	5892.58 ± 731.92	0.432 ± 0.120	9829.22 ± 608.64	7537.82 ± 411.96	0.115 ± 0.064
HGI	7222.71 ± 1553.04	5193.49 ± 1096.29	0.497 ± 0.175	6395.49 ± 462.01	4606.63 ± 273.82	0.625 ± 0.027
RegionDCL	7171.00 ± 1792.72	5381.59 ± 1060.04	0.478 ± 0.137	6964.53 ± 541.51	4933.97 ± 383.626	0.552 ± 0.042
CityFM	8863.76 ± 2235.90	6898.97 ± 1283.98	0.243 ± 0.141	8610.54 ± 681.52	6727.13 ± 456.16	0.327 ± 0.005
HA Fusion	10174.90 ± 1110.59	7757.68 ± 519.19	0.030 ± 0.080	9504.55 ± 877.47	7470.04 ± 616.78	0.172 ± 0.027
HAMURE-FisherGT	7078.00 ± 1196.86	5327.30 ± 817.12	0.518 ± 0.13	6209.19 ± 390.33	4435.84 ± 296.62	0.645 ± 0.041
HAMURE-Transformer	7193.00 ± 1272.76	5423.66 ± 836.84	0.502 ± 0.144	6575.01 ± 422.39	4686.96 ± 295.72	0.602 ± 0.046
HAMURE-Expformer	8071.09 ± 1486.49	5740.24 ± 969.39	0.371 ± 0.184	6836.03 ± 424.57	4905.69 ± 289.94	0.570 ± 0.043
HAMURE-SGFormer	8859.41 ± 1486.49	6832.93 ± 952.40	0.255 ± 0.159	8289.10 ± 356.48	6130.62 ± 213.77	0.369 ± 0.052
HAMURE-Hypformer	8179.90 ± 1334.62	6232.28 ± 873.64	0.362 ± 0.142	8193.18 ± 506.22	6389.43 ± 435.24	0.382 ± 0.077

Table 4. Population Density Performance. Highlighted are the top first, second, third

Table 4 show that HAMURE Fisher-based achieves the best performance regarding population density estimation in Singapore and New York City. Additionally, RegionDCL partitions the space using street networks, with the assumption that this enables finer-granularity that better captures urban diversity. On the other hand, HAMURE relies on predefined spatial boundaries, being much simpler model. Hence, these results demonstrate that HAMURE is indeed capable of capturing spatial diversity incorporating Fisher Elliptical Attention into Graph Transformers even being computationally simpler than RegionDCL.

5. Conclusion and Future Work

Our proposed methods address two central challenges in urban analytics: the presence of missing or incomplete place categories and the impact of heterophily in spatial interaction

graphs. By proposing methods that handle these issues, we enhance the robustness and semantic understanding of urban region representations, enhancing downstream tasks and enabling stakeholders to make more informed decisions.

Nevertheless, promising research directions remain open. First, incorporating multi-modal data sources could broaden the applicability of our methods. Besides that, incorporating topological information and high-order interactions may enhance both expressivity and interpretability of our models.

Acknowledgements

The authors would like to acknowledge the support of the Brazilian Ministry of Science, Technology and Innovations (MCTI), the Manna Team, Fundação Araucária, Softex, CNPq (Grant Number 421548/2022-3), and the Research Support Foundation of the State of Minas Gerais (FAPEMIG), and CAPES.

References

- Bianchi, F. M., Grattarola, D., Livi, L., and Alippi, C. (2021). Graph neural networks with convolutional arma filters. *IEEE transactions on pattern analysis and machine intelligence*, 44(7):3496–3507.
- Bittencourt, T. A. and Giannotti, M. (2021). The unequal impacts of time, cost and transfer accessibility on cities, classes and races. *Cities*, 116:103257.
- Di Giovanni, F., Giusti, L., Barbero, F., Luise, G., Lio, P., and Bronstein, M. M. (2023). On over-squashing in message passing neural networks: The impact of width, depth, and topology. In *International Conference on Machine Learning*, pages 7865–7885. PMLR.
- Dong, L., Duarte, F., Duranton, G., Santi, P., Barthelemy, M., Batty, M., Bettencourt, L., Goodchild, M., Hack, G., Liu, Y., et al. (2024). Defining a city—delineating urban areas using cell-phone data. *Nature Cities*, 1(2):117–125.
- dos Santos, G. B., Melos, G. C., Figueiredo, L. J., Silva, F. A., Silva, T. R., and Loureiro, A. A. (2025). Spatial optimization of charging networks for heavy-duty evs using hexagonal discrete models. *Journal of Internet Services and Applications*, 16(1):268–286.
- dos Santos, G. B., Silva, P. H., Silva, F. A., Silva, T. R. B., and Aylon, L. B. (2024). Havana: Hybrid attentional graph convolutional network semantic venue annotation model. In *Brazilian Conference on Intelligent Systems*, pages 291–305. Springer.
- Du, J., Zhang, Y., Wang, P., Leopold, J., and Fu, Y. (2019). Beyond geo-first law: Learning spatial representations via integrated autocorrelations and complementarity. In *2019 IEEE International Conference on Data Mining (ICDM)*, pages 160–169. IEEE.
- Dwivedi, V. P. and Bresson, X. (2020). A generalization of transformer networks to graphs. *arXiv preprint arXiv:2012.09699*.
- Grekousis, G. (2019). Artificial neural networks and deep learning in urban geography: A systematic review and meta-analysis. *Computers, Environment and Urban Systems*, 74:244–256.

- Guo, D., Yu, Y., Ge, S., Gao, S., Mai, G., and Chen, H. (2024). Spatialscene2vec: A self-supervised contrastive representation learning method for spatial scene similarity evaluation. *International Journal of Applied Earth Observation and Geoinformation*, 128:103743.
- Klemmer, K., Rolf, E., Robinson, C., Mackey, L., and Rußwurm, M. (2023). Satclip: Global, general-purpose location embeddings with satellite imagery. *arXiv preprint arXiv:2311.17179*.
- Kreuzer, D., Beaini, D., Hamilton, W., Létourneau, V., and Tossou, P. (2021). Rethinking graph transformers with spectral attention. *Advances in Neural Information Processing Systems*, 34:21618–21629.
- Li, Y., Huang, W., Cong, G., Wang, H., and Wang, Z. (2023). Urban region representation learning with openstreetmap building footprints. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1363–1373.
- Li, Z., Xia, L., Tang, J., Xu, Y., Shi, L., Xia, L., Yin, D., and Huang, C. (2024). Urbangpt: Spatio-temporal large language models. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 5351–5362.
- Liu, P. and Biljecki, F. (2022). A review of spatially-explicit geoi applications in urban geography. *International Journal of Applied Earth Observation and Geoinformation*, 112:102936.
- Ma, J., Li, B., and Mostafavi, A. (2024). Characterizing urban lifestyle signatures using motif properties in network of places. *Environment and Planning B: Urban Analytics and City Science*, (4):889–903.
- Milias, V. and Psyllidis, A. (2021). Assessing the influence of point-of-interest features on the classification of place categories. *Computers, Environment and Urban Systems*, 86:101597.
- Niu, H. and Silva, E. A. (2021). Delineating urban functional use from points of interest data with neural network embedding: A case study in greater london. *Computers, Environment and Urban Systems*, 88:101651.
- Psyllidis, A., Gao, S., Hu, Y., Kim, E.-K., McKenzie, G., Purves, R., Yuan, M., and Andris, C. (2022). Points of interest (poi): a commentary on the state of the art, challenges, and prospects for the future. *Computational Urban Science*, 2(1):20.
- Reia, S. M., Anderson, T., de Arruda, H. F., Atwal, K. S., Ruan, S., Kavak, H., and Pfoser, D. (2025). Function and form of us cities. *Computers, Environment and Urban Systems*, 116:102221.
- Santos, G. B., Silva, F. A., Silva, T. R. B., and Aylon, L. B. (2025). Semantic motif-based analysis of post-pandemic mobility adaptation in brazil using mobile phone data. In *2025 IEEE 49th Annual Computers, Software, and Applications Conference (COMPSAC)*, pages 1867–1872. IEEE.
- Shirzad, H., Vellingker, A., Venkatachalam, B., Sutherland, D. J., and Sinop, A. K. (2023). Exphormer: Sparse transformers for graphs. In *International Conference on Machine Learning*, pages 31613–31632. PMLR.

- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*.
- Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., and Bengio, Y. (2017). Graph attention networks. *arXiv preprint arXiv:1710.10903*.
- Wang, H., Kifer, D., Graif, C., and Li, Z. (2016). Crime rate inference with big data. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 635–644.
- Wu, Q., Zhao, W., Yang, C., Zhang, H., Nie, F., Jiang, H., Bian, Y., and Yan, J. (2023). Sgformer: Simplifying and empowering transformers for large-graph representations. *Advances in Neural Information Processing Systems*, 36:64753–64773.
- Xiao, C., Zhou, J., Huang, J., Xu, T., and Xiong, H. (2023). Spatial heterophily aware graph neural networks. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 2752–2763.
- Yang, M., Verma, H., Zhang, D. C., Liu, J., King, I., and Ying, R. (2024). Hypformer: Exploring efficient transformer fully in hyperbolic space. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 3770–3781.
- Yang, Y., Pentland, A., and Moro, E. (2023). Identifying latent activity behaviors and lifestyles using mobility data to describe urban dynamics. *EPJ Data Science*, 12(1):15.
- Yao, Z., Fu, Y., Liu, B., Hu, W., and Xiong, H. (2018). Representing urban functions through zone embedding with human mobility patterns. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18)*.
- Ye, M., Shou, D., Lee, W.-C., Yin, P., and Janowicz, K. (2011). On the semantic annotation of places in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 520–528.
- Zhang, Q., Huang, C., Xia, L., Wang, Z., Li, Z., and Yiu, S. (2023a). Automated spatio-temporal graph contrastive learning. In *Proceedings of the ACM Web Conference 2023*, pages 295–305.
- Zhang, Q., Huang, C., Xia, L., Wang, Z., Yiu, S. M., and Han, R. (2023b). Spatial-temporal graph learning with adversarial contrastive adaptation. In *International Conference on Machine Learning*, pages 41151–41163. PMLR.
- Zhang, Z., Wang, M., Xiang, Y., Huang, Y., and Nehorai, A. (2018). Retgk: Graph kernels based on return probabilities of random walks. *Advances in Neural Information Processing Systems*, 31.
- Zhu, A.-X. and Turner, M. (2022). How is the third law of geography different? *Annals of GIS*, 28(1):57–67.