



OPEN Quantifying the non-isomorphism of global urban road networks using GNNs and graph kernels

Linfang Tian^{1✉}, Weixiong Rao¹, Kai Zhao² & Huy T. Vo³

A novel concept of quantifying graph non-isomorphism is introduced to measure structural differences between graphs, and thus overcoming the strict limitations of traditional graph isomorphism tests. This paper trains Graph Neural Networks (GNNs) and graph kernels to classify urban road networks and proposes using graph classification accuracy as a metric to quantify graph non-isomorphism. Experimental results demonstrate that Edge Convolutional Neural Network (EdgeCNN) not only leverages node attributes effectively but also fully utilizes edge features, achieving an 85% classification accuracy, which surpasses that of the Weisfeiler-Lehman (WL) kernel algorithm (80%). This finding challenges the claim that “GNNs are at most as powerful as the WL test in distinguishing graph structures.” Furthermore, the paper explores the non-isomorphism of 10,361 road networks from 30 cities worldwide, providing valuable insights for future urban development.

As urbanization accelerates, traffic problems have become increasingly severe. In this context, a thorough analysis of the structural characteristics of urban road networks (URNs) can provide a scientific basis for evaluating and planning these networks. In recent years, significant progress has been made in global URN research, exploring various characteristics such as road length distribution, relative angle¹, degree distribution, and betweenness centrality², clustering coefficient^{3,4}, and block shape⁵. However, these metrics often focus on the immediate neighbor relationships of nodes (e.g., node degree, clustering coefficient) and fail to fully consider the overall structural characteristics and node attributes of the graph. Recently, many scholars have proposed using Graph Neural Networks (GNNs) to study road networks, aiming to capture their complex features^{6–11}. GNNs take into account multi-hop neighbor information and node attributes. For instance,¹² trains GNNs to perform link prediction in road networks to quantify spatial homogeneity, although their average recall and precision rates did not exceed 50%. This paper trains GNNs and graph kernels to perform graph classification on road networks, and proposes using graph classification accuracy as a metric to quantify the non-isomorphism of road networks. This approach achieves a high classification accuracy of 85%.

To quantify non-isomorphism, it is necessary to first understand the concept of isomorphism¹³. In the fields of biochemistry¹⁴ and gene networks^{15–17}, isomorphism issues frequently arise. For instance, molecules with similar structures often exhibit analogous functional properties^{18,19}. Graph isomorphism represents a strict relationship, where isomorphic graphs are structurally identical, sharing the same number of nodes, edges, spectral properties, and degree sequences. To overcome the limitations imposed by graph isomorphism, we propose a novel concept: the quantification of non-isomorphism. This concept establishes a more flexible relationship, enabling the measurement of the difference between two graphs using appropriate metrics. This definition holds particularly valuable in broader contexts, especially in practical scenarios where strict graph isomorphism is unnecessary.

This paper proposes using graph classification methods to quantify graph non-isomorphism. Graph classification tasks have found widespread application in various fields, including traffic network optimization and biochemistry. For example, these methods have been used to classify chemical molecular graphs to determine their activity^{20–22}, solubility^{23,24}, and toxicity²⁵. Over the past 30 years, the most prominent graph classification methods have primarily fallen into two categories: Graph Neural Networks (GNNs) and graph kernel methods^{26,27}. GNNs are deep neural network models specifically designed for processing graph-structured data. They have demonstrated exceptional performance across various applications, including social networks, citation networks, bioinformatics, and recommendation systems^{28–39}. On the other hand, graph kernel methods address the problem by transforming complex graph classification tasks into linearly separable problems in high-dimensional spaces via kernel functions²⁷. The Weisfeiler–Lehman (WL) test is an iterative method

¹School of Software Engineering, Tongji University, Shanghai 201804, China. ²J. Mack Robinson College of Business, Georgia State University, Atlanta 30301, USA. ³The City College of New York, City University of New York and New York University, New York 10031, USA. ✉email: tianlinfang@tongji.edu.cn; 287469917@qq.com

for detecting graph isomorphism^{40–42}, providing an efficient mechanism for updating and propagating node labels within graph kernel methods. Graph classification accuracy, ranging from 0 to 1, is a widely employed metric in machine learning tasks. In our work, we harness this metric to quantify graph non-isomorphism. The underlying rationale is that isomorphic graphs are indistinguishable from one another; hence, graphs with high classification accuracy demonstrate a pronounced non-isomorphism. Simply put, the higher the accuracy, the larger the non-isomorphism. This paper specifically utilizes the Weisfeiler–Lehman (WL) kernel and Graph Isomorphism Network (GIN), which are designed to address the challenges associated with isomorphic graphs. The paper “How Powerful are Graph Neural Networks” posits that “GNNs are at most as powerful as the WL test in distinguishing graph structures.”

Our contributions

First, the paper introduces the concept of quantifying graph non-isomorphism, addressing the strict limitations of traditional graph isomorphism tests. Specifically, this method goes beyond the binary judgment of “isomorphic” or “non-isomorphic” by quantifying non-isomorphism through classification accuracy. It provides a continuous measure of the structural and attribute similarity between two graphs (e.g., road networks), which is more suited for handling complex heterogeneous graphs. Second, we challenge the notion that “GNNs are at most as powerful as the WL test in distinguishing graph structures.” We find this assertion to be constrained to homogeneous graphs, overlooking the diversity and significance of node attributes. In reality, node attributes play a crucial role in the evaluation of graphs. Third, our analysis reveals that, compared to WL test-based graph kernel methods, GNNs offer greater scalability in handling node attributes. GNNs can fully and effectively utilize node attributes, thereby enhancing the accuracy of heterogeneous graph classification. Besides, this study examines fine-grained global URN data and provides a preliminary analysis of the evolution of world cities from the perspective of URNs. The findings demonstrate that the non-isomorphism of inter-city road networks reveals patterns of disparity that are not captured by existing road network statistical indicators. These differences are particularly evident across regions such as Europe, North America, and the Asia-Pacific, highlighting profound global urban interaction patterns. These indicators and research findings hold significant value for practice across various disciplines and stakeholders: (1) They can assist sociologists in evaluating the equity of spatial infrastructure⁴³ and provide a scientific foundation for the planning of URNs, aiding in the development of more rational and evidence-based road network planning strategies⁴⁴. (2) The non-isomorphism of inter-city road networks can serve as a metric for assessing structural differences, facilitating the transfer of policies across cities. This includes initiatives on autonomous vehicles and accident prevention, which are especially beneficial for transferring knowledge from developed countries to developing ones.

Results

We conducted a graph classification analysis on the road networks of 10,361 instances across 30 cities worldwide (Fig. 1a–c), using graph classification accuracy to quantify their non-isomorphism (Fig. 2a–c). The GIN surpassed 75% classification accuracy (Fig. 2d), while the WL kernel achieves a nearly 80% accuracy (Fig. 2d–f). We emphasize that the non-isomorphism rate serves as a comprehensive measure of graph differences, encapsulating various existing network statistical metrics, such as average degree and betweenness². We observed that cities with similar levels of development (whether in developed or developing countries), within the same geographic regions (Europe, North America, or Asia-Pacific), or in comparable geographic locations (coastal or inland), exhibit statistically smaller non-isomorphism. This observation reveals a subtle relationship between road network structures and socio-economic environments. The evolution of URNs mirrors the historical processes of socio-economic development and population concentration. Significant differences in these networks may stem from diverse historical contexts or the influence of differing infrastructure planning policies. This highlights the intricate relationship between infrastructure networks and socio-economic environments. These insights are crucial for understanding the global interactions of URNs in both developed and developing countries.

EdgeCNN performs best

The EdgeCNN^{45,46} achieves a nearly 85% accuracy in graph classification, surpassing both the WL kernel⁴⁷ and GIN^{19,48,49} (Fig. 2d). The WL kernel is renowned for its effectiveness in handling isomorphism issues, while the GIN is a deep neural network specifically designed to address graph isomorphism problems. Two principal factors influence graph classification performance: node feature extraction and edge feature extraction. In some chemical molecular graphs, edge features significantly impact performance due to their close relationship with the structure of specific subgraphs. Conversely, in some social networks, node features have a more significant impact on graph classification performance. The EdgeCNN excels because, unlike other GNNs, it applies convolution operations on edge features between nodes and their neighboring nodes. Compared to the WL kernel, which is limited to processing discrete node attributes, EdgeCNN fully leverages multi-dimensional continuous node attributes, demonstrating a more generalized and robust ability to distinguish graphs. In the following text, the term “non-isomorphism rate” will be consistently used to refer to the graph classification accuracy of EdgeCNN.

Three regions

According to the non-isomorphism rate of road networks, these world’s top 30 cities can be roughly divided into three groups: Europe (average non-isomorphism rate between 0.777 and 0.838), North America (0.868–0.931), and the Asia-Pacific region (0.807–0.887) (Fig. 2g).

European cities generally exhibit lower non-isomorphism rates, which can be attributed to their early establishment. The frameworks of most European cities were formed before the 18th century and evolved naturally, lacking unified planning. Consequently, significant updates were difficult, resulting in generally

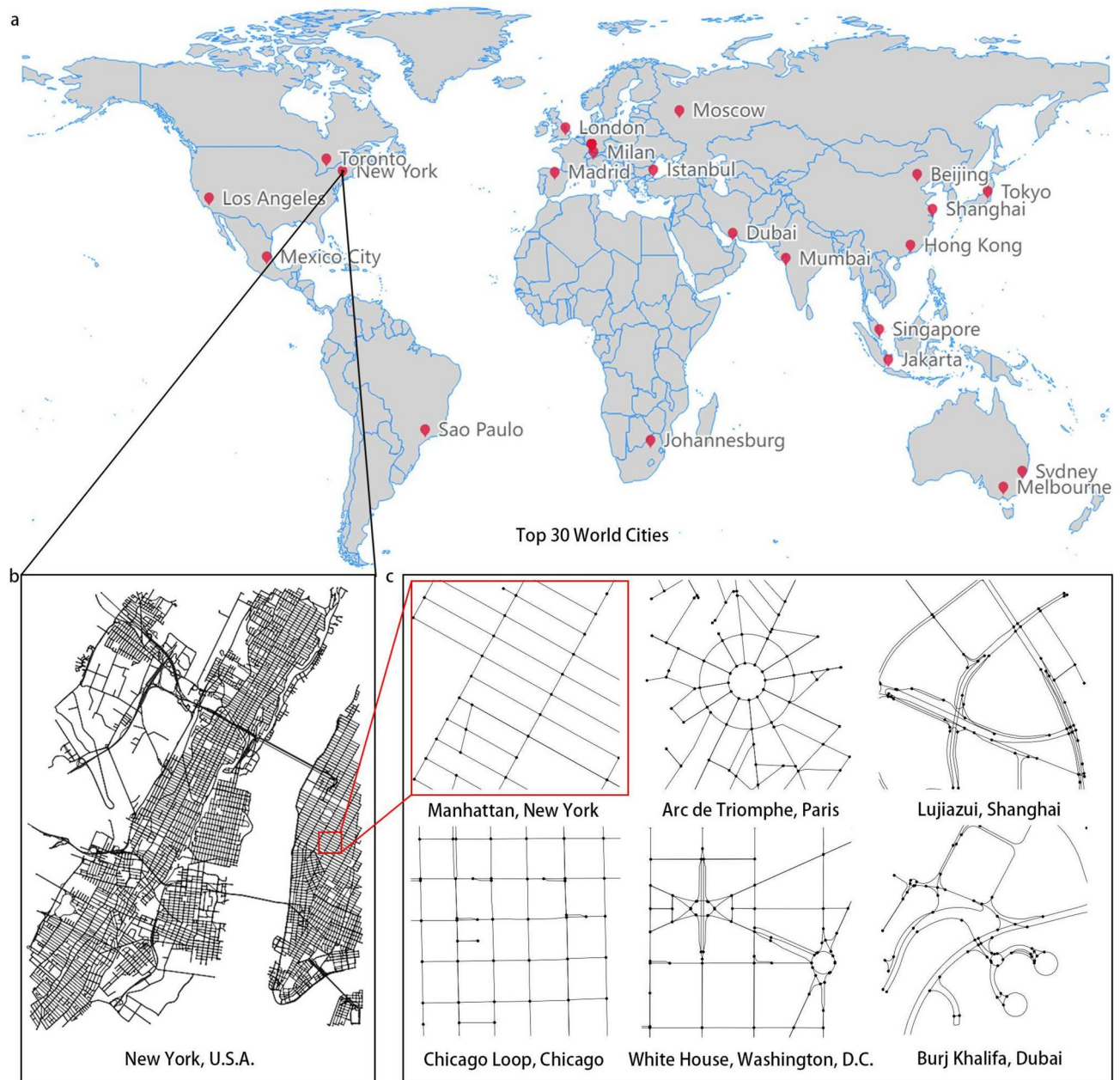


Fig. 1. Global URN Dataset. **(a)** Top 30 Global Cities and Their Typical Road Networks. According to the 2020 ranking by the Globalization and World Cities Research Network (GaWC), the top 30 global cities are predominantly located in Europe, North America, and the Asia-Pacific region. While most of these cities are coastal or semi-coastal (e.g., Toronto), there are inland exceptions such as London, Beijing, and Paris. **(b)** Road Network in Downtown New York City. The road network in downtown New York City is highly organized, with Manhattan featuring a particularly dense and regular grid of roads. **(c)** Typical Road Network Types. Left: Grid Pattern, as exemplified by Manhattan and the Chicago Loop. Middle: Radial Pattern, as exemplified by the Arc de Triomphe and the White House. Right: Freeform Pattern, as exemplified by Lujiazui and the Burj Khalifa. Note: The map data for sections b and c are provided by OpenStreetMap contributors.

irregular road networks. Amsterdam's road network has a non-isomorphism rate of 0.777, the lowest in the study. This can be explained by the city's layout, which includes over 160 intersecting rivers, more than 1000 bridges of various styles, narrow roads, and a compact structure. Frankfurt follows closely with a road network non-isomorphism rate of 0.777. Notably, five European cities—Frankfurt, Warsaw, Zurich, Amsterdam, and Brussels—have lower non-isomorphism rates. We speculate that these cities, all part of the European Union, frequently engage in political, economic, and cultural exchanges, thus exhibiting similar and diverse road network characteristics.

In contrast, there is a significant difference between the orderly uniformity of modern road layouts and the complex diversity of traditional road structures. North American cities show higher non-isomorphism rates

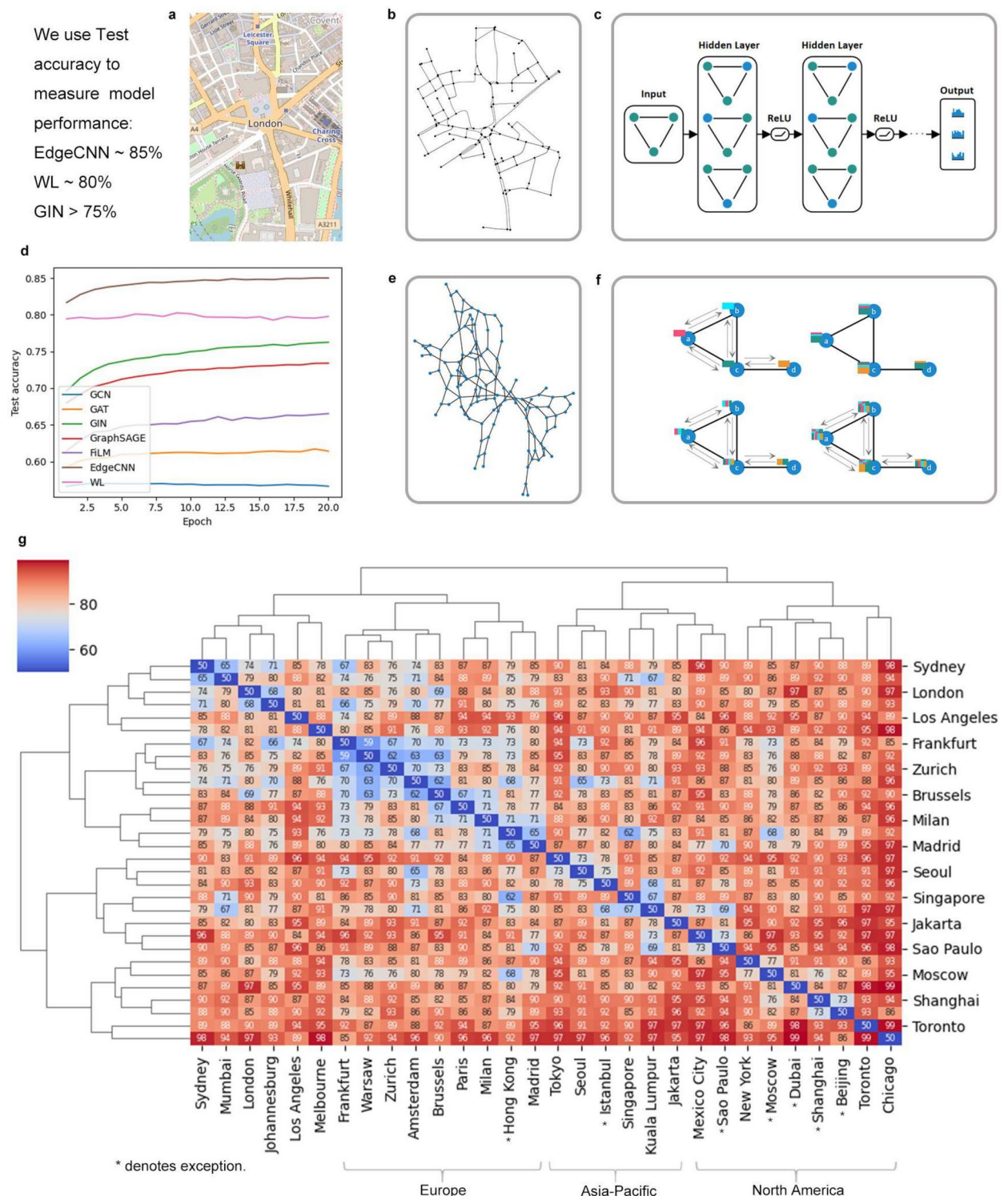


Fig. 2. Graph Classification Models and Their Accuracy. **(a)** Street map of central London. **(b)** The road network, utilized for GNNs. The node attributes are latitude and longitude coordinates normalized to the $[0, 1]$ interval. **(c)** GNN. The embedding representations of all nodes are processed by *torch.mean()*. **(d)** Test accuracy of graph classification. The EdgeCNN achieves an accuracy close to 85%. The WL kernel reaches nearly 80%, while the GIN exceeded 75%. **(e)** The topology of the road network, used for the WL kernel. The node labels are assumed identical. **(f)** WL kernel. The node labels (colors) are iteratively refined through color refinement. **(g)** Graph classification accuracy for 30 cities obtained through the EdgeCNN. Maps a, b, and e are copyrighted by OpenStreetMap contributors.

in their road networks. This is due to the relatively short history of large-scale population concentration in North American cities. Their URN planning, influenced by European cities, follows standardized land ordinance policies, which can be traced back to the grid plan of Philadelphia, Pennsylvania, in the early 18th century⁵⁰. Chicago's road network has a non-isomorphism rate of 0.931, significantly higher than other cities. Chicago boasts one of the most consistent and regular grid road networks in the world. When landing at O'Hare International Airport at night, passengers can see the city resembling a vast circuit board, with streets running uniformly north-south or east-west. In Mr. Boeing's study, all 16 cities with the world's most regular road networks are located in North America⁵¹.

The non-isomorphism rates of road networks in the Asia-Pacific region fall between those of Europe and North America. This is due to the relatively late urbanization process in the Asia-Pacific region. The road networks of old city areas primarily evolved naturally and are less regular; however, the road networks of new city areas, developed mostly after North American cities, adopted more systematic planning and construction, making their road networks more regular than those in developed countries.

URN comparison

Subsequently, we focused our analysis more narrowly. Using EdgeCNN, we classified nodes within the intercity road networks to delve deeper into the intercity differences and explore the underlying causes (Fig. 3). We selected two European cities: Frankfurt (average non-isomorphism rate of 0.777) and Warsaw (0.798); two North American cities: Los Angeles (0.873) and Mexico City (0.881); and two Asia-Pacific cities: Seoul (0.830) and Tokyo (0.887). This study reveals two common characteristics of these URNs. First, city centers, which are typically political and cultural hubs adorned with numerous historical and cultural landmarks, exhibit less regular road networks due to the difficulty of updating them. Second, coastal or riverside areas, influenced by the natural irregularities of water bodies, also display lower network regularity. However, areas adjacent to these coastal or riverside regions tend to have open vistas and high commercial value. These areas, when meticulously planned and developed, exhibit more regular patterns and higher degrees of organization, as seen in the southern region of the Han River in Seoul. Comparison between Frankfurt and Warsaw (non-isomorphism rate 0.592, Fig. 3a). The overall lighter color indicates smaller differences, with the road networks on the north and south sides of the Main River in Frankfurt being less regular. Comparison between Los Angeles and Frankfurt (0.745, Fig. 3b). The slightly darker color compared to panel a suggests small differences in the overall road network. Comparison between Tokyo and Seoul (0.735, Fig. 3c). The lighter color in the areas north of the Han River in Seoul and near Tokyo Bay indicates smaller differences. The southern region of the Han River, due to rapid development in recent years, shows more regular road networks and hence darker colors. In URN relationship network (Fig. 3d), European cities are centrally located and have closer relationships with each other, while North American and Asia-Pacific cities are mostly on the periphery, with more distant relationships. Comparison between Frankfurt and Mexico City (0.956, Fig. 3e). Compared to panel a, the city center areas remain lighter in color, but the surrounding areas are noticeably darker. This indicates that Mexico City's road network is generally more regular, though the city center, with its many historical and cultural landmarks (such as Constitution Square), has lower regularity due to the challenges of major renovations. Comparisons between Los Angeles and Tokyo (0.961, Fig. 3f,g). The overall darker colors indicate a larger difference between them.

Compared with traditional methods

Here, we compared the proposed quantification of non-isomorphism with traditional analysis methods, mainly involving cosine similarity and Spearman correlation analysis. Notably, we examined URNs using the multi-dimensional representation of existing network metrics, rather than relying on GNNs or graph kernels. This decision stems from the fact that these representations offer a more thorough and comprehensive understanding of network topologies. Using *osmnx.basic_stats()*, we obtained nine existing network metrics for the top 30 cities globally (*k_avg*, *edge_length_avg*, *streets_per_node_avg*, *street_length_avg*, *circuitry_avg*, *degree ≤ 2*, *degree = 3*, *degree = 4*, *degree ≥ 5*) (as shown in Fig. 4a), and encoded each URN as a nine-dimensional vector, resulting in 30 nine-dimensional vectors. The bearing distributions of six specific URNs are illustrated in Fig. 4b. Using these 30 nine-dimensional vectors, we conducted traditional cosine similarity and Spearman correlation clustering analyses (as shown in Fig. 4c,d). The results were compared with the graph non-isomorphism achieved through the EdgeCNN (as shown in Fig. 2g). Same point: The EdgeCNN also identifies city groups with high similarity and low diversity, such as Hong Kong and Singapore, with a non-isomorphism rate of 0.625. Both cities are port cities with world-class ports and are significant financial centers in Asia. Additionally, they lack natural resources, were former British colonies, and are significantly influenced by Chinese culture. Furthermore, both cities have developed relatively unrestricted coastal road networks. Different point: Cities that show high similarity with traditional methods but significant differences with the EdgeCNN, such as New York City and Chicago (Fig. 1b,c), have a graph non-isomorphism rate of 0.934. This phenomenon primarily attributes to the denser and more regular road networks, which often have more intersections. Consequently, the message-passing mechanism of the EdgeCNN receives more neighborhood information, enhancing classification accuracy.

Based on the comprehensive analysis, the results of employing EdgeCNN to analyze URNs align closely with previous studies. This alignment substantiates the pivotal role of graph isomorphism in graph classification tasks. Therefore, the higher the graph classification accuracy, the larger the non-isomorphism. Hence, it is both reasonable and feasible to use graph classification accuracy to quantify non-isomorphism. Quantifying non-isomorphism introduces a more flexible approach, overcoming the rigid constraints of graph isomorphism and permitting the measurement of the differences between two graphs through graph classification accuracy. This method finds broader applications in practical scenarios where strict graph isomorphism is not required. The paper "How Powerful are Graph Neural Networks" asserts that "GNNs are at most as powerful as the WL test

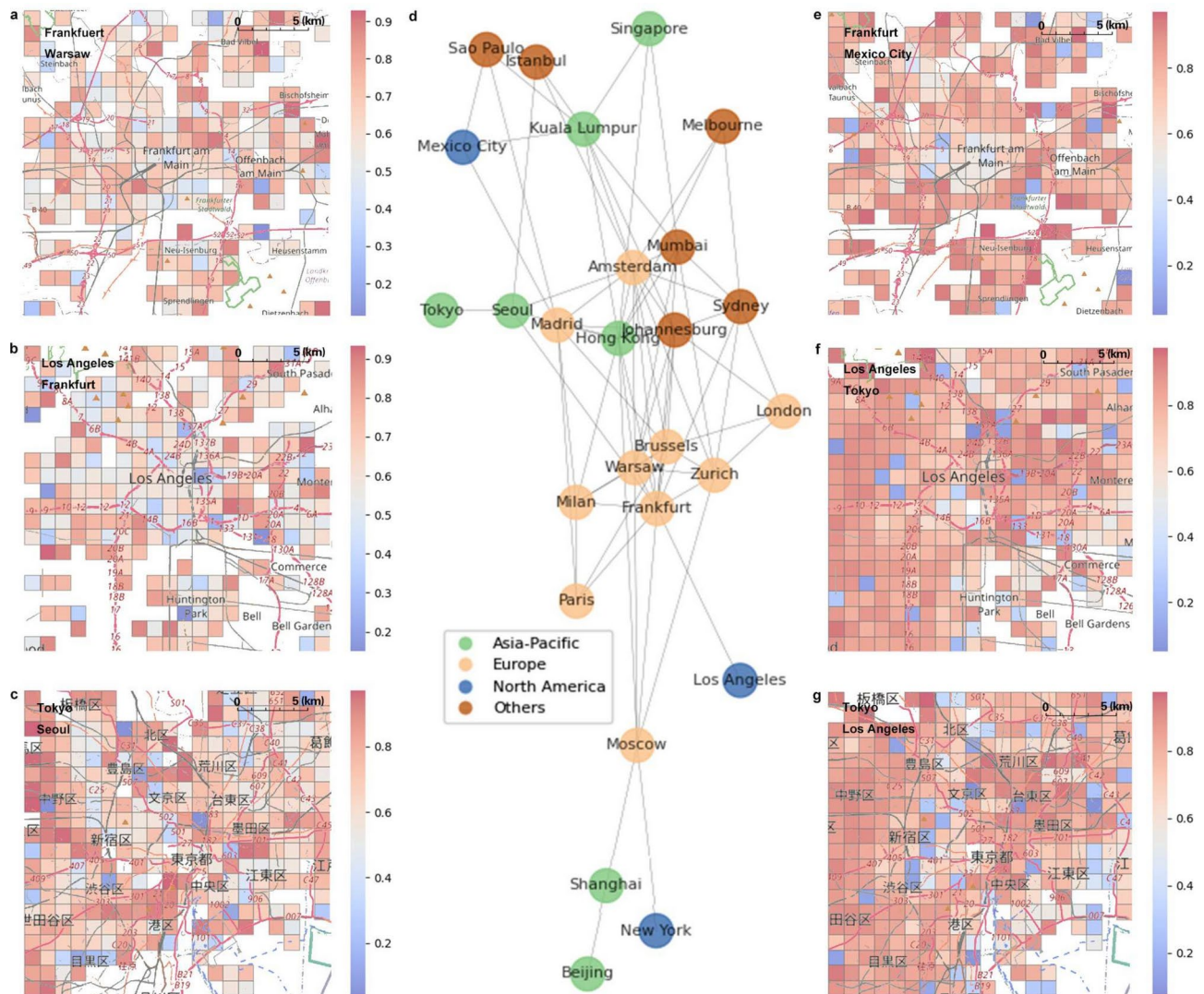


Fig. 3. URN Comparison. We utilized EdgeCNN to classify nodes within intercity road networks. In these visualizations, blue represents a node classification accuracy of 0.0, while red indicates an accuracy of 1.0. The gradient of color reflects the varying levels of node classification accuracy. Our analysis included two European cities: Frankfurt (average non-isomorphism rate of 0.777) and Warsaw (0.798); two North American cities: Los Angeles (0.873) and Mexico City (0.881); and two Asia-Pacific cities: Seoul (0.830) and Tokyo (0.887). (a) Comparison between Frankfurt and Warsaw (non-isomorphism rate 0.592). The overall lighter color indicates smaller differences, with only the road networks on the north and south sides of the Main River in Frankfurt being less regular. (b) Comparison between Los Angeles and Frankfurt (0.745). The slightly darker color compared to panel a suggests minor differences in the overall road network. (c) Comparison between Tokyo and Seoul (0.735). The lighter color in the areas north of the Han River in Seoul and near Tokyo Bay indicates smaller differences. (d) URN Relationship Network. We selected intercity relationships with a non-isomorphism rate ≤ 0.78 . Chicago, Dubai, Toronto, and Jakarta do not appear in the panel because their non-isomorphism rate with other cities exceeded 0.78, partly due to the higher regularity of their road networks. (e) Comparison between Frankfurt and Mexico City (0.956). Compared to panel (a), the city center areas remain lighter in color, but the surrounding areas are noticeably darker. This suggests that while Mexico City's road network is generally more regular, its historical city center, with numerous cultural landmarks such as Constitution Square, exhibits lower regularity due to the challenges of significant renovations. (f, g) Comparisons between Los Angeles and Tokyo (0.961). The overall darker colors indicate a larger difference between these two cities. Upon closer inspection, slight differences can be observed between panels (f, g), which can be attributed to inherent errors in the algorithm. Maps a, b, c, e, f, and g are copyrighted by OpenStreetMap contributors.

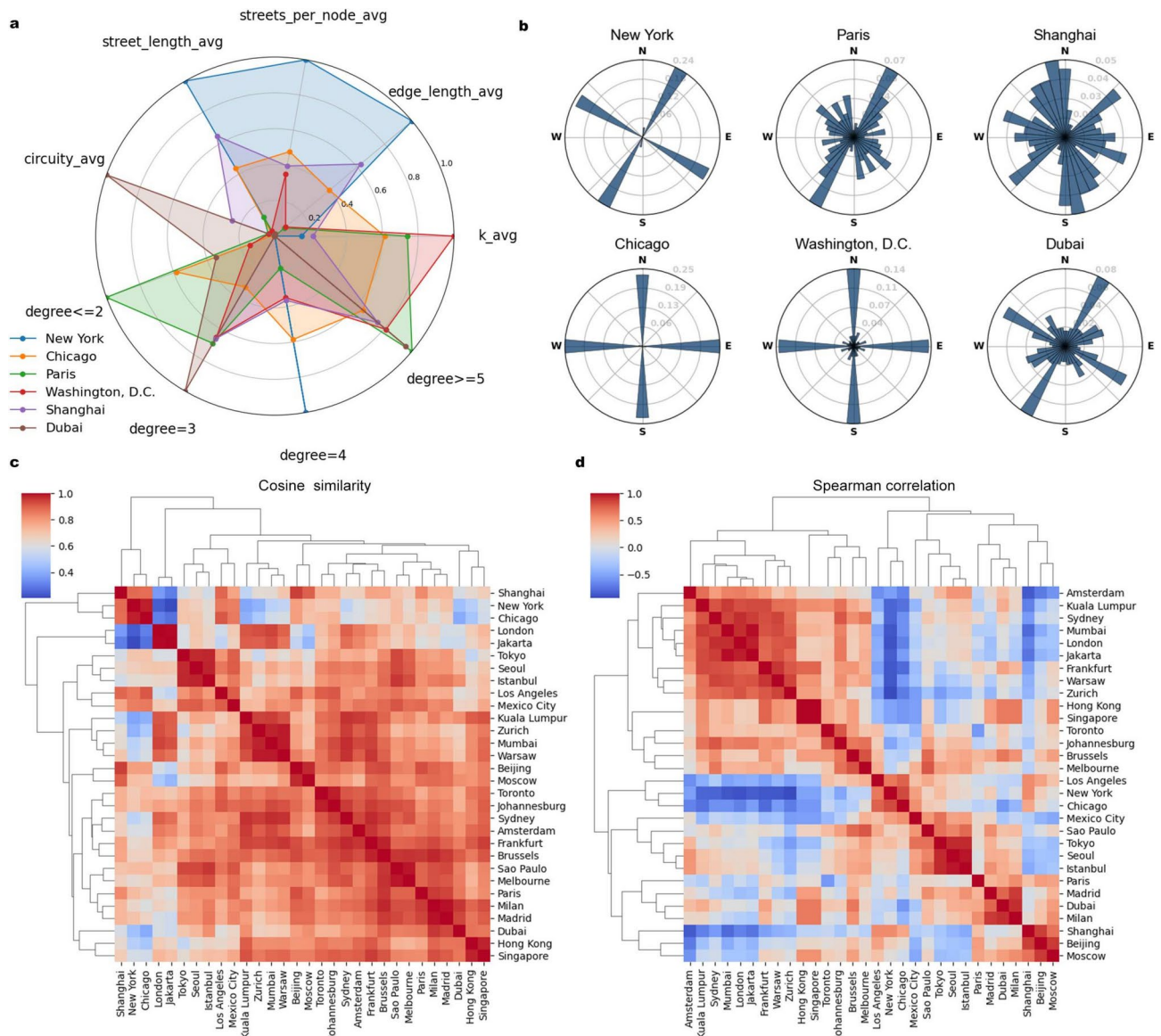


Fig. 4. Traditional URN Analysis Methods. **(a)** Traditional Metrics of Specific URNs. We analyzed 9 traditional metrics and their distributions for 6 specific URNs (as shown in Fig. 1c). In URNs with low regularity, the road network around Burj Khalifa in Dubai exhibits a high proportion of *circuitry_avg*, *degree* = 3, and *degree* = 5, while the road network around the Arc de Triomphe in Paris shows a higher proportion of *degree* ≤ 2 and *degree* = 5. In high-regularity URNs, the road network in Manhattan, New York, has the highest proportion of *degree* = 4, and the road network around the White House in Washington, D.C., has the highest average degree *k_degree*. **(b)** Bearing Distribution of 6 Specific URNs. As shown in Fig. 1c, the bearings of URNs around Manhattan, Chicago Loop, and the White House are highly regular, while the bearings of URNs around the Arc de Triomphe, Lujiazui, and Burj Khalifa show varying degrees of diversity. **(c)** Cosine Similarity Analysis. **(d)** Spearman Correlation Analysis. In sections c and d, we obtained 9 traditional metrics for the top 30 cities worldwide using *osmnx.basic_stats()* (same metrics as in panel a). Each city was encoded as a nine-dimensional vector, forming 30 nine-dimensional vectors. Using this data, we performed cosine similarity and Spearman correlation analyses to compare with the graph non-isomorphism achieved through the EdgeCNN.

in distinguishing graph structures,” which is limited to edge features and does not account for the diversity and importance of node attributes, thereby restricting its relevance to homogeneous graphs. However, in complex heterogeneous graphs, node attributes play a crucial role in graph evaluation. Our research indicates that, in comparison to other GNNs, EdgeCNN performs convolution operations on edge features between nodes and their neighboring nodes. Unlike the WL kernel, which can only process discrete node attributes, EdgeCNN can fully leverage multi-dimensional continuous node attributes. Consequently, EdgeCNN exhibits superior generalization and robustness in differentiating graphs.

Discussion

Research implications

The paper explores the non-isomorphism of road networks across different cities, analyzing its relationship with various aspects of urban socio-economic development, resource allocation, and future urban planning. It provides both academic and practical contributions to urban studies, including infrastructure fairness assessment, quantitative urban analysis, and applications in transfer learning for urban computing.

- *Assessment of Infrastructure Equity (Urban Science and Policy)* In an era of rapid urbanization, sociologists can use this metric to measure differences in road networks within the same city, thereby promoting the development of equitable evaluation standards⁴³. This approach not only aids in formulating fairer urban policies but also improves the rational allocation of urban resources.
- *Quantitative Urban Analysis (Urban Archaeology)* Leveraging the research findings on inter-city road network differences, urban archaeologists can effectively trace the historical evolution of global road networks and socio-economic environments. This provides crucial insights into the trajectories and patterns of past urban development.
- *Transfer Learning (Urban Computing)* Transfer learning can facilitate the adaptation of policies and regulations from one city to another. However, the universal applicability of transfer performance between different cities remains uncertain. In-depth studies of URN non-isomorphism can enhance the generalization capability of transfer learning models, aiding in the development of more adaptive urban management strategies.

Looking ahead, the availability of more detailed URNs and urban traffic data (such as road grades and traffic volumes) will aid in quantifying the differences in traffic networks. User-friendly non-isomorphism computation packages will enhance the practical implementation of machine learning. Fractal dimensions quantify the self-similarity of URNs at different resolution levels and characterize their evolutionary patterns⁵², while non-isomorphism measures the differences in the overall topological structure of inter-city road networks. By combining fractal methods, we can gain a deeper understanding of the URN complexity and their differences across cities.

Broader impact

The inter-city road network non-isomorphism we proposed has two significant advantages. Firstly, it allows for quantitative comparison of inter-city road networks. Secondly, it reveals the transferability of various urban planning policies. Due to these two characteristics, the indicators proposed in this study have wide applications in disciplines such as urban science, urban computing, and road network science. Our indicators are based on a global understanding of road network changes. Here are a few potential issues related to practical applications: First, the road network data used in this study comes from the publicly available OpenStreetMap data source. While widely used in academia, this dataset may not include all the latest road segments. Therefore, we recommend that readers verify the accuracy of this road network data with other sources when adopting our method. Second, training neural networks inevitably consumes energy. As network size and training data volume increase, computational costs and resource consumption also rise. Compared to other AI tasks that require hundreds of GPUs or hundreds of hours of GPU runtime, our method has relatively low computational costs. Specifically, in our practice of running 10,361 samples using small GNNs or graph kernels, we spent over a dozen hours training each model, using only a GPU on a personal computer (CUDA). Third, to mitigate sample bias, we used a latitude and longitude interval of $0.01^\circ \times 0.01^\circ$ for URNs, balancing the bias caused by different sample areas and shapes. We used graph classification accuracy to evaluate model performance, and multiple experiments demonstrated that the model is robust to various URNs, with stable results.

Research limitations

This study has four limitations. First, although GNNs and graph kernels can effectively capture the complex multi-hop node relationships in URNs, the accuracy of graph classification tends to decrease after multiple iterations, leading to gradient vanishing and over-smoothing issues during model training⁵³. Adopting more advanced learning architectures and training algorithms may improve classification performance⁵⁴. Second, the reliability of URN classification is affected by the training set and test set ratio (80%/20%), and parameter selection poses inevitable challenges during model training. Third, we defined urban areas using a $0.2^\circ \times 0.2^\circ$ latitude and longitude grid, which, while convenient, results in rigid urban boundaries. In theory, urban boundaries should be based on administrative divisions, but these divisions lead to differences in the area and shape of the study objects. Fourth, this study is limited to the extraction and analysis of URN features and does not investigate the intricate relationship between URNs and socio-economic indicators. Moreover, the influence of natural features, including topography, river networks, and water systems, on URNs requires further investigation.

Model dependency

When we use a certain model to classify the road network set of a certain city pair, the classification accuracy is closely related to both the model performance and the URN structural differences between cities. For each city pair, the better the model performance, the higher the classification accuracy. For each model, those city pairs exhibiting relatively higher classification accuracy always show comparatively higher classification accuracy in other models. It is important to note that different models may yield different classification accuracies, introducing some subjectivity in quantifying the non-isomorphism of the road networks. However, this subjectivity does not affect our ability to assess the relative differences in road network structures between city pairs, nor does it impact our evaluation of the relative performance of the models.

Data dependency

When the training data is increased, the classification accuracy generally improves. However, this improvement gradually slows and ultimately reaches a saturation point. This is because the model can extract more structural features from a richer dataset, but as the model becomes more refined, the improvement diminishes and approaches an upper bound. When the quality of the training set is poor, we cannot rely on the classification accuracy as a proxy for non-isomorphism. In cases of insufficient data diversity or excessive noise, the model may learn ineffective features or be influenced by noise, leading to an inaccurate reflection of structural differences between graphs. Therefore, ensuring both the quality and diversity of training data is crucial to prevent model overfitting and to enhance model effectiveness. In this study, the selected road network samples encompass a diverse array of cities differing in age, regions, locations, and development status. Furthermore, the dataset, sourced from OpenStreetMap, underwent effective preprocessing, significantly reducing data noise.

Over-smoothing issue

In Graph Neural Networks (GNNs), the over-smoothing problem refers to the phenomenon in which, as the number of network layers increases, the node representations become progressively more similar, and even identical, leading to a gradual loss of distinctiveness between node representations and making them difficult to differentiate. This issue is closely related to the spectral properties of normalized matrices, such as the adjacency matrix or the Laplacian matrix. Since the eigenvalues of these matrices lie within the range $[-1, 1]$, repeated convolution operations (matrix exponentiation) cause the smaller eigenvalues to quickly decay towards zero, while the largest eigenvalue remains at 1. As the number of layers increases, the node representations progressively converge to the subspace defined by the eigenvectors corresponding to the largest eigenvalues, ultimately resulting in over-smoothing. To explore the impact of varying network layers on graph classification tasks, we conducted experiments using GCN and EdgeCNN models with different layer configurations. The experimental results indicate that when the number of layers is set to 1, 2, 3, 4, or 5, no significant over-smoothing phenomenon was observed. Moreover, the results show that classification accuracy stabilizes at a relatively fair level when the number of layers is set to 3. Based on these findings, and to ensure a fair comparison of model performance, we standardized the number of layers to 3 for all GNN models. This choice of fewer layers not only prevents over-smoothing but also ensures training efficiency.

Methods

Datasets

OpenStreetMap (OSM), as a prominent initiative within the realm of Volunteered Geographic Information (VGI) platforms⁵⁵, is renowned for its rich data, extensive coverage, free access, and rapid update speed. In recent years, a large number of studies leveraging OSM have achieved significant results. In this study, we utilized Python's OSMnx library to download global URN data from OSM. World cities often serve as transportation hubs and epitomize URN planning paradigms for their respective regions. Therefore, this study selected the top 30 cities⁵⁶, as released in 2020 by the Globalization and World Cities Research Network, as research subjects (Fig. 1a). These samples encompass a diverse array of cities differing in age (ancient and modern cities), regions (Europe, North America, and the Asia-Pacific), locations (coastal and inland cities), and development status (cities in developed and developing countries).

The study area for each city was defined as a $0.2^\circ \times 0.2^\circ$ (approximately $20\text{km} \times 20\text{km}$) region centered around the city's centroid, which was subdivided into 400 grid cells, each measuring $0.01^\circ \times 0.01^\circ$ (approximately $1\text{km} \times 1\text{km}$). Such grid cells have been widely adopted in existing urban studies. Regardless of the research objectives, a fundamental principle is that the size of the grid cells should be appropriate for the specific task. In this study, we found that grid cells of $0.01^\circ \times 0.01^\circ$ contain a sufficient number of intersections and road segments to effectively train graph classification models. The specific steps are as follows:

- Construct a $0.2^\circ \times 0.2^\circ$ (approximately $20\text{km} \times 20\text{km}$) region around the city's centroid and subdivide it into 400 grid cells, each measuring $0.01^\circ \times 0.01^\circ$ (approximately $1\text{km} \times 1\text{km}$).
- Use the function $G = \text{ox.graph_from_bbox}(\text{north}, \text{south}, \text{east}, \text{west}, \text{network_type} = \text{"drive"}, \text{simplify} = \text{True})$ to extract road network data within each grid cell. Here, the variables north, south, east, and west represent the latitude and longitude of the grid boundaries, and the variable "simplify=True" indicates the simplification of the graph topology by removing redundant nodes and non-intersection nodes.

Data preprocessing

To enhance the quality of URN data and improve the accuracy of experimental results, data preprocessing is essential. This is especially important considering that initial OSM road data, despite its good integrity, can still be somewhat rough. The following are the specific preprocessing steps undertaken: First, we checked data consistency to ensure both completeness and accuracy. Next, we excluded overlapping roads and invalid data to reduce redundancy and eliminate errors. Following this, double-line roads were converted into single lines to simplify their representation. Complex road networks were then transformed into simple graph structures, facilitating easier subsequent analysis and processing. Additionally, topology checks were performed to ensure the logical structure of the road network. Isolated roads were deleted to maintain network connectivity and consistency. By adhering to these steps, we ensured that the preprocessed road data possess robust topological relationships, thus providing high-quality data support for our experiments.

GNNs

This paper employs the following six types of GNNs: *Graph Convolutional Network (GCN)*: GCN stands as one of the pioneering models in the GNN realm. It utilizes local convolution operations to aggregate the features of nodes, thereby extracting comprehensive global features from the graph⁵⁷. *Graph Attention Network (GAT)*: GAT integrates an attention mechanism that dynamically adjusts the weights of relationships between nodes, enabling the model to capture crucial information within the graph structure with greater precision⁵⁸. *Feature-wise Linear Modulation (FiLM)*: FiLM introduces conditional information to linearly transform internal network features. By learning scale and shift parameters, it fine-tunes intermediate features, thereby enhancing the model's adaptability to specific tasks⁵⁹. *Graph Sample and Aggregate (GraphSAGE)*: GraphSAGE samples a fixed number of neighbors for each node, then aggregates the features of these neighbors to generate a new node representation. This approach is particularly effective in handling large-scale graph data⁶⁰. *Graph Isomorphism Network (GIN)*: GIN refines the contributions of node features and their neighbors through learnable parameters, achieving precise node representation calculations. GIN's ability to recognize and distinguish different graph structures, owing to its graph isomorphism property, makes it highly effective for graph classification tasks^{19,49}. *Edge Convolutional Neural Network (EdgeCNN)*: EdgeCNN enhances node representations by performing convolution operations on the edges between nodes, directly leveraging edge features^{45,46}.

In this study, our objective is to classify the road networks of 30 global cities. URN was abstracted as an undirected graph $G = (V, E)$, where the start and end points of road segments and intersections constituted the node set V . The attributes of these nodes were normalized latitude and longitude coordinates, and the road segments connecting these nodes formed the edge set E . Utilizing the `train_test_split()` function, we partitioned 10,361 road network grid cells into 80% for training and 20% for testing. GNNs update each node representation through message passing, where h_i^{l+1} results from a nonlinear transformation of the representations of neighboring nodes and the node itself from the previous layer:

$$h_i^{l+1} = \sigma \left(\sum_{r \in R} \sum_{j \in N_i^r} W_r^l h_j^l + W_0^l h_i^l \right)$$

Here, l denotes the network layer, R represents the set of relations, N_i^r is the set of neighboring nodes of v_i under relation r , and W_r^l and W_0^l are learnable parameter matrices. We used the Adam optimizer to minimize cross-entropy loss. Through recursively applying the message passing mechanism, GNNs construct high-dimensional feature representations of nodes layer by layer, thus achieving effective graph encoding.

- **Model Performance** We trained the aforementioned six types of GNNs on the road networks of each city pair and evaluated their performance based on classification accuracy using the test set. The results show that the EdgeCNN achieves a classification accuracy of nearly 85%, and the GIN exceeds 75% accuracy and continues to improve.
- **Quantifying Non-Isomorphism** Road network non-isomorphism describes the degree of difference between URNs. We quantified this non-isomorphism using classification accuracy and illustrated the differences among the road networks of 30 cities using a 30×30 matrix. The higher the classification accuracy, the greater non-isomorphism between URNs.
- **Hyperparameters** The hyperparameters of the GNNs include the number of layers, epochs, neurons, and learning rate. To optimize these hyperparameters, we conducted standard five-fold cross-validation. Ultimately, we determined the optimal configuration to be 3 layers, 20 epochs, 128 neurons, and a learning rate of 0.001. These settings were chosen to maximize the model's classification performance and training efficiency.

Graph Kernels

Graph kernels map the local substructures of a graph to a Hilbert space, generating a set of high-dimensional feature vectors that describe the graph's overall structural characteristics. Their primary goal is to address the graph isomorphism problem. Given that graph kernels typically require substantial computational resources, selecting an appropriate graph kernel is crucial. Commonly used graph kernels include the Weisfeiler-Lehman (WL) Kernel^{40,42}, the Random Walk (RW) Kernel⁶¹, the Shortest Path (SP) Kernel⁶², the Regularized Graph Kernel (REGK), and the Multiscale Laplacian Graph Kernel (MPGK). Among these, the WL kernel is considered one of the most advanced algorithms in the field of graph classification. Based on the 1-WL graph isomorphism test, the WL kernel iteratively relabels nodes and their neighborhoods to detect graph isomorphism, demonstrating high computational efficiency and powerful isomorphism detection capabilities.

In this study, our task is to classify the road networks of 30 global cities. URN is abstracted as an undirected graph $G = (V, E)$, where V is the set of nodes (representing the start and end points of road segments and intersections) and E is the set of edges (representing road segments). We partitioned 10,361 road network grid units into 80% training set, 10% validation set, and 10% test set. The specific steps of color refinement in the WL kernel are as follows:

- **Initial Coloring**: Assign an identical initial label to each node.
- **Neighbor Aggregation**: For each node, aggregate the labels of its neighboring nodes to form a multiset. This multiset reflects the node's position and role within its local structure.
- **Label Compression**: Use a hash function to sort, compress, and map the labels in the multiset, thereby assigning a new, more detailed label to each node:

$$c_i^t(v) = \text{HASH}(c_i^{t-1}(v), \{c_i^{t-1}(u) | u \in N(v)\}),$$

where t denotes the iteration count, $N(v)$ denotes the set of neighbors of node v , and $c_i^t(v)$ denotes the label of node v at iteration t .

- **Iteration:** Repeat the neighbor aggregation and label compression steps multiple times. With each iteration, the node labels become increasingly detailed, capturing higher-level structural information.

Through multiple iterations, color refinement produces stable label assignments such that two isomorphic graphs will eventually have identical label distributions. The WL kernel uses these label distributions as feature vectors for graph classification, employing a support vector machine *sklearn.svm.LinearSVC* to perform the classification and evaluating model performance using classification accuracy on the test set.

- **Model performance** We evaluated three graph kernels (WL kernel, RW kernel, and SP kernel) for the graph classification task. However, we found that the RW kernel requires approximately 23 minutes to classify each pair of city road networks, taking nearly seven days to classify the 435 pairs of city road networks for 30 global cities. The SP kernel also demands significant computational resources. Considering the time cost, we ultimately abandoned the RW kernel and the SP kernel. The WL kernel completes all the road network classifications in just a few hours, with its running time only linearly related to the number of nodes and edges. The WL kernel achieves a classification accuracy of nearly 80%, representing a significant advancement in isomorphism detection in this field.
- **Hyperparameters** The hyperparameters of the WL kernel include the number of WL layers and the number of iterations. We determined the optimal hyperparameter settings through standard five-fold cross-validation. Ultimately, we set the number of layers to 5 and the number of iterations to 10.

Model performance

The paper employs six different GNN models along with the WL graph kernel for training and testing on the data. The URN classification task presents a highly imbalanced machine learning challenge, characterized by a predominance of grid-type instances over radial or free-type instances. We evaluated model performance using graph classification accuracy. Overall, the EdgeCNN achieves the highest accuracy, approaching 85%; the WL kernel reaches nearly 80%; and the GIN exceeds 75%, demonstrating continuous improvement. This underscores the robust capability of EdgeCNN, WL kernel, and GIN in capturing the structural characteristics of URNs. The non-isomorphism rates among the 30 cities vary significantly, with Chicago having the highest non-isomorphism rate at 0.931 and Amsterdam the lowest at 0.777. Notably, if the same city's road network is used for graph classification, such as London and its own road networks, the theoretical classification accuracy should be 0.500.

Traditional analysis

We conducted a comprehensive analysis of the topological characteristics of the road networks in 30 world cities. Specifically, we used the *ox.basic_stats(G_projected)* function to extract fundamental attributes of the road networks, including *k_avg* (average node degree), *edge_length_avg* (average edge length), *streets_per_node_avg* (average number of streets per node), *street_length_avg* (average street length), *circuitry_avg* (average circuitry), *degree ≤ 2* (percentage of nodes with degree less than or equal to 2), *degree = 3* (percentage of nodes with degree equal to 3), *degree = 4* (percentage of nodes with degree equal to 4), and *degree ≥ 5* (percentage of nodes with degree greater than or equal to 5). We employed cosine similarity and Spearman correlation analyses to evaluate the similarity and correlation of the intercity road networks.

Data availability

We used the publicly available road network data from OpenStreetMap (<https://www.openstreetmap.org>) via the OSMnx Python package. The road network data is available at the online data warehouse: <https://github.com/TinaGina/Urban-Road-Networks>.

Code availability

Source codes for the training and testing results are available at the online data warehouse: <https://github.com/TinaGina/Urban-Road-Networks>.

Received: 27 June 2024; Accepted: 17 February 2025

Published online: 22 February 2025

References

1. Serra, M. & Hillier, B. Angular and metric distance in road network analysis: A nationwide correlation study. *Comput. Environ. Urban Syst.* **74**, 194–207 (2019).
2. Nigam, R., Sharma, D. K., Jain, S. & Srivastava, G. A local betweenness centrality based forwarding technique for social opportunistic iot networks. *Mob. Netw. Appl.* **1**, 1–16 (2022).
3. Shang, W.-L. et al. Statistical characteristics and community analysis of urban road networks. *Complexity* **2020**, 1–21 (2020).
4. Akbarzadeh, M., Reihani, S. F. S. & Samani, K. A. Detecting critical links of urban networks using cluster detection methods. *Physica A* **515**, 288–298 (2019).

5. Zheng, X. & Yang, J. Urban road network design for alleviating residential exposure to traffic pollutants: Super-block or mini-block?. *Sustain. Cities Soc.* **89**, 104327 (2023).
6. Wu, N., Zhao, X. W., Wang, J. & Pan, D. Learning effective road network representation with hierarchical graph neural networks. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, 6–14 (2020).
7. He, S. et al. Roadtagger: Robust road attribute inference with graph neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence* **34**, 10965–10972 (2020).
8. Jana, D., Malama, S., Narasimhan, S. & Tacioglu, E. Edge-based graph neural network for ranking critical road segments in a network. *PLoS ONE* **18**, e0296045 (2023).
9. Chang, Y., Tanin, E., Cao, X. & Qi, J. Spatial structure-aware road network embedding via graph contrastive learning. *Adv. Database Technol.-EDBT* **26**, 144–156 (2023).
10. Wang, M.-X., Lee, W.-C., Fu, T.-Y. & Yu, G. On representation learning for road networks. *ACM Trans. Intell. Syst. Technol. (TIST)* **12**, 1–27 (2020).
11. Zhang, L. & Long, C. Road network representation learning: A dual graph based approach. In *ACM Transactions on Knowledge Discovery from Data* (2023).
12. Xue, J. et al. Quantifying the spatial homogeneity of urban road networks via graph neural networks. *Nat. Mach. Intell.* **4**, 246–257 (2022).
13. Grohe, M. & Schweitzer, P. The graph isomorphism problem. *Commun. ACM* **63**, 128–134 (2020).
14. Bell, E. W. & Zhang, Y. Dockrmsd: An open-source tool for atom mapping and rmsd calculation of symmetric molecules through graph isomorphism. *J. Cheminform.* **11**, 1–9 (2019).
15. Lipinska, A. P., Collén, J., Krueger-Hadfield, S. A., Mora, T. & Ficko-Blean, E. To gel or not to gel: Differential expression of carrageenan-related genes between the gametophyte and tetrasporophyte life cycle stages of the red alga *Chondrus crispus*. *Sci. Rep.* **10**, 11498 (2020).
16. Wefers, A. K. et al. Isomorphic diffuse glioma is a morphologically and molecularly distinct tumour entity with recurrent gene fusions of mybl1 or myb and a benign disease course. *Acta Neuropathol.* **139**, 193–209 (2020).
17. Ding, K. et al. Feature-enhanced graph networks for genetic mutational prediction using histopathological images in colon cancer. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part II* **23**, 294–304 (Springer, 2020).
18. Ranjan, S. et al. Isomorphism: Molecular similarity to crystal structure similarity in multicomponent forms of analgesic drugs tolfenamic and mefenamic acid. *IUCrJ* **7**, 173–183 (2020).
19. Peng, Y. et al. Enhanced graph isomorphism network for molecular admet properties prediction. *IEEE Access* **8**, 168344–168360 (2020).
20. Lim, J. et al. Predicting drug-target interaction using a novel graph neural network with 3d structure-embedded graph representation. *J. Chem. Inf. Model.* **59**, 3981–3988 (2019).
21. Wieder, O. et al. A compact review of molecular property prediction with graph neural networks. *Drug Discov. Today Technol.* **37**, 1–12 (2020).
22. Dou, B. et al. Machine learning methods for small data challenges in molecular science. *Chem. Rev.* **123**, 8736–8780 (2023).
23. Xiong, Z. et al. Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism. *J. Med. Chem.* **63**, 8749–8760 (2019).
24. Korolev, V., Mitrofanov, A., Korotcov, A. & Tkachenko, V. Graph convolutional neural networks as “general-purpose” property predictors: The universality and limits of applicability. *J. Chem. Inf. Model.* **60**, 22–28 (2019).
25. Wang, F. et al. Graph attention convolutional neural network model for chemical poisoning of honey bees’ prediction. *Sci. Bull.* **65**, 1184–1191 (2020).
26. Kriege, N. M., Johansson, F. D. & Morris, C. A survey on graph kernels. *Appl. Netw. Sci.* **5**, 1–42 (2020).
27. Siglidis, G. et al. Grakel: A graph kernel library in python. *J. Mach. Learn. Res.* **21**, 1993–1997 (2020).
28. Guo, Z. & Wang, H. A deep graph neural network-based mechanism for social recommendations. *IEEE Trans. Ind. Inf.* **17**, 2776–2783 (2020).
29. Min, S. et al. Stgsn-a spatial-temporal graph neural network framework for time-evolving social networks. *Knowl.-Based Syst.* **214**, 106746 (2021).
30. Kumar, S., Mallik, A., Khetarpal, A. & Panda, B. Influence maximization in social networks using graph embedding and graph neural network. *Inf. Sci.* **607**, 1617–1636 (2022).
31. Cummings, D. & Nassar, M. Structured citation trend prediction using graph neural networks. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 3897–3901 (IEEE, 2020).
32. Liu, J., Xia, F., Feng, X., Ren, J. & Liu, H. Deep graph learning for anomalous citation detection. *IEEE Trans. Neural Netw. Learn. Syst.* **33**, 2543–2557 (2022).
33. Yang, C. & Han, J. Revisiting citation prediction with cluster-aware text-enhanced heterogeneous graph neural networks. In *Proceedings of the IEEE International Conference on Data Engineering (ICDE)* (2023).
34. Zhang, X.-M., Liang, L., Liu, L. & Tang, M.-J. Graph neural networks and their current applications in bioinformatics. *Front. Genet.* **12**, 690049 (2021).
35. Yi, H.-C., You, Z.-H., Huang, D.-S. & Kwok, C. K. Graph representation learning in bioinformatics: trends, methods and applications. *Brief. Bioinform.* **23**, bbab340 (2022).
36. Pfeifer, B., Saranti, A. & Holzinger, A. Gnn-subnet: Disease subnetwork detection with explainable graph neural networks. *Bioinformatics* **38**, ii120–ii126 (2022).
37. Wu, S., Sun, F., Zhang, W., Xie, X. & Cui, B. Graph neural networks in recommender systems: A survey. *ACM Comput. Surv.* **55**, 1–37 (2022).
38. Gao, C., Wang, X., He, X. & Li, Y. Graph neural networks for recommender system. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, 1623–1625 (2022).
39. Yang, L., Liu, Z., Dou, Y., Ma, J. & Yu, P. S. Consisrec: Enhancing gnn for social recommendation via consistent neighbor aggregation. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2141–2145 (2021).
40. Huang, N. T. & Villar, S. A short tutorial on the weisfeiler-lehman test and its variants. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 8533–8537 (IEEE, 2021).
41. Brooksbank, P. A., Grochow, J. A., Li, Y., Qiao, Y. & Wilson, J. B. Incorporating weisfeiler-leman into algorithms for group isomorphism. <http://arxiv.org/abs/1905.02518> (2019).
42. Kiefer, S. & Neuen, D. The power of the Weisfeiler–Leman algorithm to decompose graphs. *SIAM J. Discret. Math.* **36**, 252–298 (2022).
43. Nello-Deakin, S. Is there such a thing as a ‘fair’ distribution of road space?. *J. Urban Des.* **24**, 698–714 (2019).
44. Zheng, Y. et al. Spatial planning of urban communities via deep reinforcement learning. *Nat. Comput. Sci.* **3**, 748–762 (2023).
45. Ding, C. et al. A cloud-edge collaboration framework for cognitive service. *IEEE Trans. Cloud Comput.* **10**, 1489–1499 (2020).
46. Kumar, A. et al. Mobihisnet: A lightweight cnn in mobile edge computing for histopathological image classification. *IEEE Internet Things J.* **8**, 17778–17789 (2021).
47. Morris, C. et al. Weisfeiler and leman go neural: Higher-order graph neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence* **33**, 4602–4609 (2019).

48. Meng, L. & Zhang, J. Isomorph neural network for graph representation learning and classification. <http://arxiv.org/abs/1907.09495> (2019).
49. Kim, B.-H. & Ye, J. C. Understanding graph isomorphism network for rs-fmri functional connectivity analysis. *Front. Neurosci.* **14**, 630 (2020).
50. Boeing, G. Off the grid... and back again? The recent evolution of American street network planning and design. *J. Am. Plann. Assoc.* **87**, 123–137 (2021).
51. Boeing, G. Urban spatial order: Street network orientation, configuration, and entropy. *Appl. Netw. Sci.* **4**, 1–19 (2019).
52. Zhang, H., Lan, T. & Li, Z. Fractal evolution of urban street networks in form and structure: A case study of hong kong. *Int. J. Geogr. Inf. Sci.* **36**, 1100–1118 (2022).
53. Chen, D. et al. Measuring and relieving the over-smoothing problem for graph neural networks from the topological view. In *Proceedings of the AAAI Conference on Artificial Intelligence* **34**, 3438–3445 (2020).
54. Shi, M. et al. Genetic-gnn: Evolutionary architecture search for graph neural networks. *Knowl.-Based Syst.* **247**, 108752 (2022).
55. Feldmeyer, D., Meisch, C., Sauter, H. & Birkmann, J. Using openstreetmap data and machine learning to generate socio-economic indicators. *ISPRS Int. J. Geo Inf.* **9**, 498 (2020).
56. Tian, L., Rao, W., Zhao, K. & Vo, H. T. Analyzing world city network by graph convolutional networks. *Sci. Rep.* **14**, 18933 (2024).
57. Chiang, W.-L. et al. Cluster-gcn: An efficient algorithm for training deep and large graph convolutional networks. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 257–266 (2019).
58. Brody, S., Alon, U. & Yahav, E. How attentive are graph attention networks? <http://arxiv.org/abs/2105.14491> (2021).
59. Brockschmidt, M. Gnn-film: Graph neural networks with feature-wise linear modulation. In *International Conference on Machine Learning*, 1144–1152 (PMLR, 2020).
60. Zhou, J. et al. Graph neural networks: A review of methods and applications. *AI Open* **1**, 57–81 (2020).
61. Nikolentzos, G., Siglidis, G. & Vazirgiannis, M. Graph kernels: A survey. *J. Artif. Intell. Res.* **72**, 943–1027 (2021).
62. Nikolentzos, G., Meladianos, P., Rousseau, F., Stavarakas, Y. & Vazirgiannis, M. Shortest-path graph kernels for document similarity. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 1890–1900 (2017).

Author contributions

L.T. conceived the experiments, L.T. and W.R. conducted the experiments, W.R. and K.Z. analysed the results, W.R. and H.V. reviewed the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to L.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025