

Bridge-Centered Metapath Classification Using R-GCN-VGAE for Disaster-Resilient Maintenance Decisions

Takato Yasuno

Abstract

Daily infrastructure management in preparation for disasters is critical for urban resilience. When bridges remain resilient against disaster-induced external forces, access to hospitals, shops, and residences via metapaths can be sustained, maintaining essential urban functions. However, prioritizing bridge maintenance under limited budgets requires quantifying the multi-dimensional roles that bridges play in disaster scenarios—a challenge that existing single-indicator approaches fail to address.

Approach: We focus on metapaths from national highways through bridges to buildings (hospitals, shops, residences), constructing a heterogeneous graph with road, bridge, and building layers. A Relation-centric Graph Convolutional Network Variational Autoencoder (R-GCN-VGAE) learns metapath-based feature representations, enabling classification of bridges into disaster-preparedness categories: Supply Chain (commercial logistics), Medical Access (emergency healthcare), and Residential Protection (preventing isolation).

Case Study: Using OSMnx and open data, we validate our methodology on three diverse cities in Ibaraki Prefecture, Japan: Mito (697 bridges), Chikusei (258 bridges), and Moriya (148 bridges), totaling 1,103 bridges. The heterogeneous graph construction from open data enables redefining bridge roles for disaster scenarios, supporting maintenance budget decision-making.

Results: Clustering quality achieves Silhouette scores of 0.289–0.363, with latent dimension z19 showing strong correlation (Spearman $r = 0.416$, $p = 1.47 \times 10^{-30}$) with national highway metapath counts, demonstrating specialized encoding of logistics hub connectivity. k-NN parameter tuning ($k=3 \rightarrow 5$) increases coverage by +66% (162 \rightarrow 270 paths), improving semantic validity for disaster scenarios.

Contributions: (1) Open-data methodology for constructing urban heterogeneous graphs. (2) Re-definition of bridge roles for disaster scenarios via metapath-based classification. (3) Establishment of maintenance budget decision support methodology.

(4) k-NN tuning strategy validated across diverse city scales. (5) Empirical demonstration of UMAP superiority over t-SNE/PCA for multi-role bridge visualization.

Keywords: Bridge maintenance, Disaster resilience, Heterogeneous graph, R-GCN, Variational Autoencoder, Metapath analysis, Infrastructure management

1 Introduction

1.1 Background and Motivation

Japan’s aging infrastructure poses a critical challenge for disaster resilience. Among the nation’s 730,000 bridges, over 50% will exceed their 50-year design life by 2033, with maintenance costs projected to exceed 8 trillion yen annually [1, 2]. Recent natural disasters, including the 2019 Typhoon Hagibis floods [3], underscore the urgency of disaster-preparedness planning [4]. Beyond structural integrity, bridges serve as vital nodes in urban networks, where their resilience directly impacts access to essential services—hospitals for emergency care, shops for supply chains, and residences for community connectivity.

The core premise of this work is that **daily infrastructure management in preparation for disasters requires understanding the multi-dimensional roles bridges play in sustaining urban functions**. When bridges remain resilient against disaster-induced external forces (earthquakes, floods, landslides), access via metapaths—shortest paths from national highways through bridges to buildings—can be sustained, maintaining essential city operations. However, existing approaches prioritize bridges using single indicators such as traffic volume or structural condition ratings, which fail to capture their functional diversity in disaster scenarios [5, 6].

Consider a hospital-serving bridge with low traffic volume—it may rank low in traditional prioritization but becomes critical during disasters when emergency medical access determines survival outcomes. Con-

versely, a high-traffic bridge serving primarily commercial logistics may be less urgent if alternative routes exist. This disconnect between structural metrics and disaster-preparedness roles necessitates a new methodology.

1.2 Problem Statement

Under limited maintenance budgets, municipalities must answer: *Which bridges should be prioritized to maximize disaster resilience while sustaining access to hospitals, supply chains, and residential areas?* This requires:

1. **Quantifying multi-dimensional bridge roles:** A single bridge may simultaneously serve medical access, commercial logistics, and residential protection—roles invisible to traffic-based rankings.
2. **Leveraging open data:** Many municipalities lack detailed bridge inventories, requiring open-source alternatives (OpenStreetMap, geospatial data).
3. **Scalability across city scales:** Methodologies must work for both large metropolitan areas (700+ bridges) and small cities (100-200 bridges) where traditional clustering algorithms fail.

Traditional graph-based approaches using betweenness centrality or closeness centrality treat all edges uniformly, ignoring the semantic differences between highway→bridge paths (logistics arteries) and bridge→residence paths (community lifelines). Recent advances in heterogeneous graph neural networks offer a solution by explicitly modeling different node types (bridges, roads, buildings) and edge types (to_hospital, to_shop, to_residence) within a unified framework.

1.3 Research Questions

This work addresses three fundamental questions:

1. **RQ1 (Metapath Feasibility):** Can metapaths originating from national highways and passing through bridges to buildings (hospitals, shops, residences) effectively classify bridges into disaster-preparedness roles?
2. **RQ2 (Model Effectiveness):** Does a Relation-centric Graph Convolutional Network Variational Autoencoder (R-GCN-VGAE) effectively learn metapath-based feature representations in heterogeneous urban graphs?

3. **RQ3 (Open Data Applicability):** Can the methodology be deployed using only open data sources (OSMnx, OpenStreetMap), and does it scale to diverse city sizes from large (700+ bridges) to small (100-200 bridges)?

1.4 Proposed Approach

We construct a **heterogeneous graph** with three layers—roads, bridges, and buildings—where node types and edge types explicitly represent urban infrastructure semantics. Focusing on metapaths from national highways through bridges to buildings (hospitals for medical access, shops for supply chains, residences for isolation prevention), we train an R-GCN-VGAE to learn 32-dimensional latent representations that encode bridge connectivity patterns.

The key innovation is **relation-centric encoding**: instead of treating all edges uniformly, R-GCN maintains separate weight matrices W_r for each relation type (e.g., highway→bridge vs. local_road→bridge), enabling the model to distinguish logistics hubs from local access bridges. Following dimensionality reduction via UMAP, bridges are classified into three disaster-preparedness categories:

- **Supply Chain:** High connectivity to shops (k=5 nearest), critical for post-disaster logistics and commercial continuity.
- **Medical Access:** High connectivity to hospitals (k=5 nearest), essential for emergency healthcare delivery.
- **Residential Protection:** High connectivity to residential buildings (k=20 nearest), preventing community isolation.

1.5 Contributions

This paper makes the following contributions:

1. **Open-data heterogeneous graph construction methodology:** Complete pipeline from OSMnx to bridge classification, enabling municipalities without detailed inventories to assess infrastructure.
2. **Disaster-centric bridge role redefinition:** Metapath-based classification (Supply Chain/Medical Access/Residential Protection) that aligns bridge prioritization with disaster preparedness rather than traffic volume.
3. **Maintenance budget decision support framework:** Systematic methodology for prioritizing repair investments based on multi-dimensional disaster-resilience roles.

4. **k-NN tuning strategy:** Empirical validation showing $k=3 \rightarrow 5$ increases coverage by +66% (162 \rightarrow 270 metapaths in Mito City), with semantic justification ($k=5$ for focused services, $k=20$ for residential neighborhoods).
5. **Dimensionality reduction evaluation:** Demonstration that UMAP outperforms t-SNE and PCA for visualizing multi-role bridge embeddings, preserving both local cluster structure and global topology.
6. **Multi-scale validation:** Successful deployment across three cities with 8 \times structural diversity: Mito (697 bridges, prefectural capital), Chikusei (258 bridges, regional hub), Moriya (148 bridges, logistics gateway with extreme highway connectivity up to 2,803 metapaths per bridge).

The remainder of this paper is organized as follows: Section 2 reviews related work on infrastructure criticality, graph autoencoders, and metapath analysis. Section 3 details the R-GCN-VGAE architecture and metapath extraction methodology. Section 4 presents experimental results across three Ibaraki cities. Section 5 discusses three key lessons learned (k-NN tuning, UMAP superiority, relation-centric vs. node-centric models) and practical implications. Section 6 concludes with future research directions.

2 Related Work

2.1 Infrastructure Criticality and Disaster Resilience

Traditional bridge maintenance prioritization relies on structural health indicators such as crack width, corrosion depth, and load-bearing capacity [1]. While structural assessments remain essential, they fail to capture *network-level criticality*—the role of individual bridges in sustaining urban functionality during disasters.

Graph-theoretic approaches have emerged to quantify infrastructure criticality via network centrality metrics. Betweenness centrality, closeness centrality, and eigenvector centrality identify bridges whose failure maximally disrupts network connectivity [7]. However, these topological metrics treat all edges uniformly, ignoring semantic differences between commercial, medical, and residential connections.

Recent disaster resilience research emphasizes multi-layer network analysis, modeling interdependencies between transportation, power grids, and communication networks [5]. Buldyrev et al. [6]

demonstrated cascading failures in coupled infrastructure networks during the 2003 Italy blackout. Our work extends this paradigm to *heterogeneous urban graphs*, where bridges serve as critical nodes connecting diverse facility types.

2.2 Graph Variational Autoencoders

Variational Autoencoders (VAEs) [8] revolutionized unsupervised representation learning by combining probabilistic inference with deep neural networks. Kipf and Welling [9] extended VAEs to graph-structured data via Graph Convolutional Networks (GCNs) [10], enabling node embedding learning that preserves network topology.

The Graph Variational Autoencoder (GVAE) framework encodes nodes into a latent space $\mathbf{z} \sim \mathcal{N}(\mu, \sigma^2)$ using a GCN encoder, then reconstructs edges via an inner product decoder $\hat{A} = \sigma(ZZ^\top)$. This approach has been successfully applied to link prediction [9], community detection [11], and graph generation [12].

Relational GCN (R-GCN) [13]: Standard GCNs assume homogeneous graphs where all edges represent identical relationships. R-GCN introduces *relation-specific weight matrices* to handle heterogeneous edges, enabling modeling of knowledge graphs and multi-relational networks. Our R-GCN-VGAE architecture combines R-GCN’s edge-type awareness with VGAE’s generative modeling, specifically targeting bridge-centered metapath extraction.

Comparison to HetVGAE: Hamilton et al. [14] proposed heterogeneous GraphSAGE for inductive node classification. Our previous work implemented HetVGAE (heterogeneous node-type embeddings) for social impact prediction, achieving $r = 0.56\text{-}0.68$ correlations. R-GCN-VGAE differs by encoding *edge-level semantics* (street \rightarrow bridge transitions) rather than node-level attributes, proving complementary for disaster-preparedness classification.

2.3 Metapath-based Graph Analysis

Metapaths—structured sequences of node types and edge types—capture semantic relationships in heterogeneous networks. Sun et al. [15] introduced PathSim similarity for bibliographic networks, quantifying author-paper-author co-authorship patterns. Metapath2vec [16] extended node2vec to heterogeneous graphs by performing metapath-guided random walks, capturing semantic relationships across node types.

In urban infrastructure analysis, metapaths encode multi-hop connectivity: **Bridge \rightarrow Street \rightarrow Shop** repre-

sents supply chain access, `Bridge→Street→Hospital` represents emergency medical access, and `Bridge→Street→Residence` represents evacuation route potential. Unlike citation networks where metapaths follow predefined schemas (e.g., "author-paper-venue"), urban graphs require *distance-constrained metapaths*—k-NN algorithms ensure bridges connect only to geographically proximate facilities (Section 3.4).

Recent work on urban accessibility [17] uses shortest-path algorithms to measure facility reachability. Our metapath framework extends this by encoding *multiple simultaneous connections* (e.g., bridges near both hospitals and residences classified as Balanced Multi-Use) rather than single-target reachability.

2.4 OpenStreetMap for Infrastructure Analysis

OpenStreetMap (OSM) has emerged as a critical open dataset for large-scale infrastructure studies. Boeing [18] introduced OSMnx, a Python library for downloading and analyzing street networks, enabling reproducible urban morphology research. OSMnx-based studies have quantified street network centrality [19], walkability [20], and disaster evacuation planning [21].

Bridge Extraction from OSM: OSM’s `man_made=bridge` tag identifies standalone bridge structures, while `bridge=yes` on road segments marks road-embedded bridges. We adopt the former for municipal infrastructure focus, filtering unnamed bridges to exclude pedestrian footbridges.

Amenity Geocoding: OSM’s amenity tagging system (`amenity=hospital`, `shop=*`, `building=residential`) provides detailed POI data. However, completeness varies: Mito’s 15,978 shops vs. 65 hospitals reflects both actual distribution and volunteer mapper biases [22]. Our k-NN parameter tuning (Section 5.1) mitigates this imbalance.

Quality Considerations: Haklay [23] reported 80% positional accuracy (within 6m of Ordnance Survey data in London). For disaster-preparedness classification, 80%+ accuracy suffices since metapath analysis operates at street-block granularity (~100m) rather than centimeter-level precision. Our methodology’s open-data emphasis enables continuous improvement as OpenStreetMap volunteers enhance regional coverage.

3 Methodology

3.1 Problem Formulation

Let $G = (V, E, \mathcal{T}_v, \mathcal{T}_e)$ denote a heterogeneous graph where V is the set of nodes, $E \subseteq V \times V$ is the set of edges, \mathcal{T}_v defines node types, and \mathcal{T}_e defines edge types (relations). In our urban infrastructure context:

$$V = V_{\text{bridge}} \cup V_{\text{street}} \cup V_{\text{building}} \quad (1)$$

$$\mathcal{T}_v = \{\text{bridge}, \text{street}, \text{building}\} \quad (2)$$

$$\mathcal{T}_e = \{\text{to_shop}, \text{to_hospital}, \text{to_residence}, \text{street_to_street}, \text{street_to_bridge}, \dots\} \quad (3)$$

Each bridge $b \in V_{\text{bridge}}$ is associated with a feature vector $\mathbf{x}_b \in \mathbb{R}^{21}$ comprising: - Structural attributes: span length, year built (if available) - Topological attributes: degree centrality, betweenness centrality - Metapath counts: `{highway_metapath_count, shop_count, hospital_count}` - `is_highway`: Binary indicator for highway proximity

Disaster Resilience Formulation: We define bridge resilience R_b as the capacity to sustain access to critical urban functions under external forces F (earthquakes, floods). Formally, let $\mathcal{M}_b = \{m_1, m_2, \dots, m_k\}$ denote the set of metapaths originating from national highways, passing through bridge b , and terminating at building nodes (hospitals, shops, residences). The disaster-preparedness role of bridge b is characterized by:

$$\text{Role}(b) = \arg \max_{c \in \mathcal{C}} P(c | \mathcal{M}_b, \mathbf{z}_b) \quad (4)$$

where $\mathcal{C} = \{\text{Supply}, \text{Medical}, \text{Residential}\}$ are disaster-preparedness categories, and $\mathbf{z}_b \in \mathbb{R}^{32}$ is a latent representation learned via R-GCN-VGAE encoding highway-origin connectivity patterns.

Key Insight: A bridge’s disaster role is determined not by traffic volume or structural condition, but by the *types of buildings* reachable via metapaths and the *density of highway connections* enabling logistics flow.

3.2 Heterogeneous Graph Construction

3.2.1 Data Acquisition via OSMnx

We use OSMnx [18] to extract road networks, bridge locations, and building points of interest (POI) from OpenStreetMap [24] for three cities in Ibaraki Prefecture:

- Road Network:** Download all drivable roads within city boundaries using

`ox.graph_from_place()`. National highways (`highway=trunk`) are identified via OSM tag filtering.

- Bridges:** Extract named bridges using Overpass API queries: (`man_made=bridge`) AND (`name!=None`). Filter unnamed structures to focus on municipally managed infrastructure.
- Buildings:** Query amenity tags for `{amenity=hospital, shop=*, building=residential}` with spatial filtering (2km buffer around bridges).

3.2.2 Graph Topology

The heterogeneous graph G is constructed with the following node and edge types:

Nodes:

- bridge** ($|V_{\text{bridge}}| = 697$ for Mito): Bridge centroids from OSM geometries
- street** ($|V_{\text{street}}| = 31,001$ for Mito): Road segment nodes from OSMnx network
- building** ($|V_{\text{building}}| = 16,791$ for Mito): POI nodes categorized as hospital, shop, or residence

Edges:

- street_to_street:** Road network connectivity (adjacency matrix)
- street_to_bridge:** Spatial proximity (snapping bridges to nearest road segments)
- to_shop, to_hospital, to_residence:** Bridge-to-building edges via k-NN algorithm (Section 3.4)

3.2.3 Coordinate System and Distance Calculation

All spatial coordinates are projected to EPSG:6677 (JGD2011 Plane Rectangular Coordinate System Zone 9) for metric distance calculations. Haversine distances are used for initial filtering, followed by planar distances for k-NN algorithms.

3.3 R-GCN-VGAE Architecture

3.3.1 Relational Graph Convolutional Encoder

The encoder consists of three Relational Graph Convolutional Network (R-GCN) layers [13], each implementing relation-specific message passing:

$$h_i^{(\ell+1)} = \sigma \left(\sum_{r \in \mathcal{R}} \sum_{j \in \mathcal{N}_r(i)} \frac{1}{|\mathcal{N}_r(i)|} W_r^{(\ell)} h_j^{(\ell)} + W_0^{(\ell)} h_i^{(\ell)} \right) \quad (5)$$

where:

- $h_i^{(\ell)} \in \mathbb{R}^{d_\ell}$: Hidden representation of node i at layer ℓ
- \mathcal{R} : Set of relation types (street→street, street→bridge)
- $\mathcal{N}_r(i)$: Neighbors of node i under relation r
- $W_r^{(\ell)} \in \mathbb{R}^{d_{\ell+1} \times d_\ell}$: Relation-specific weight matrix
- $W_0^{(\ell)}$: Self-loop transformation
- σ : ReLU activation function

Basis Decomposition: To reduce parameters, we use basis decomposition:

$$W_r^{(\ell)} = \sum_{b=1}^B a_{rb}^{(\ell)} V_b^{(\ell)} \quad (6)$$

with $B = 2$ basis matrices, reducing parameters from $|\mathcal{R}| \times d_{\ell+1} \times d_\ell$ to $B \times d_{\ell+1} \times d_\ell + |\mathcal{R}| \times B$.

Architecture Configuration:

$$\text{Layer 1: } \mathbb{R}^{21} \xrightarrow{\text{R-GCN}} \mathbb{R}^{128} \quad (7)$$

$$\text{Layer 2: } \mathbb{R}^{128} \xrightarrow{\text{R-GCN}} \mathbb{R}^{128} \quad (8)$$

$$\text{Layer 3 } (\mu): \mathbb{R}^{128} \xrightarrow{\text{R-GCN}} \mathbb{R}^{32} \quad (9)$$

$$\text{Layer 3 } (\log \sigma^2): \mathbb{R}^{128} \xrightarrow{\text{R-GCN}} \mathbb{R}^{32} \quad (10)$$

3.3.2 Variational AutoEncoder Framework

Following the VAE framework [8, 9], the encoder outputs mean $\boldsymbol{\mu}$ and log-variance $\log \boldsymbol{\sigma}^2$ for each node. The latent representation is sampled via reparameterization:

$$\mathbf{z}_i = \boldsymbol{\mu}_i + \boldsymbol{\sigma}_i \odot \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I}) \quad (11)$$

The decoder reconstructs edges via inner product:

$$p(A_{ij} = 1 | \mathbf{z}_i, \mathbf{z}_j) = \sigma(\mathbf{z}_i^\top \mathbf{z}_j) \quad (12)$$

The loss function combines reconstruction loss and KL divergence with β -annealing:

$$\mathcal{L} = \mathcal{L}_{\text{recon}} + \beta \mathcal{L}_{\text{KL}} \quad (13)$$

$$\mathcal{L}_{\text{recon}} = -\mathbb{E}_{q(\mathbf{Z}|\mathbf{X}, A)}[\log p(A|\mathbf{Z})] \quad (14)$$

$$\mathcal{L}_{\text{KL}} = \text{KL}[q(\mathbf{Z}|\mathbf{X}, A) \| p(\mathbf{Z})] \quad (15)$$

where β anneals from 0.01 \rightarrow 1.0 over the first 50 epochs to prevent posterior collapse [8].

3.4 Metapath Extraction and k-NN Tuning

3.4.1 Metapath Definition

A metapath $m = \langle v_{\text{highway}}, v_{\text{bridge}}, v_{\text{building}} \rangle$ represents a 3-hop path: (1) national highway node \rightarrow (2) bridge node \rightarrow (3) building node. We focus on three building types corresponding to disaster-preparedness roles:

- m_{shop} : Highway \rightarrow Bridge \rightarrow Shop (Supply Chain)
- m_{hospital} : Highway \rightarrow Bridge \rightarrow Hospital (Medical Access)
- $m_{\text{residence}}$: Highway \rightarrow Bridge \rightarrow Residence (Residential Protection)

Simplification: Rather than extracting full shortest paths from highways to bridges (30–85 segments on average), our approach focuses on *direct bridge \rightarrow building connections* within a 2 km radius, using highway proximity as a binary node feature (`is_highway`).

3.4.2 k-NN Parameter Selection

For each bridge-building pair (b, c) with Haversine distance $d(b, c) \leq 2$ km, we rank buildings by distance and select the top- k nearest neighbors. The choice of k balances *semantic validity* (meaningful service range) and *computational efficiency*:

Empirical Validation (Mito City): Increasing shop k-NN from 3 to 5 increased Supply Chain metapath coverage by +66% (162 \rightarrow 270 paths), capturing mid-tier commercial bridges previously classified as “Balanced Multi-Use.”

3.5 Classification Strategy

Bridges are classified into disaster-preparedness categories based on *dominant metapath counts* and *confidence scores* derived from metapath proportions:

$$\text{Confidence}(b, c) = \frac{|\mathcal{M}_{b,c}|}{\sum_{c' \in \mathcal{C}} |\mathcal{M}_{b,c'}|} \quad (16)$$

where $|\mathcal{M}_{b,c}|$ is the count of metapaths from bridge b to category c .

Category Definitions:

- **Supply Chain** (confidence > 0.9): $|\mathcal{M}_{b,\text{shop}}| \gg |\mathcal{M}_{b,\text{hospital}}|, |\mathcal{M}_{b,\text{residence}}|$. Critical for post-disaster logistics and commercial continuity.

- **Medical Access** (confidence > 0.7): $|\mathcal{M}_{b,\text{hospital}}| \gg |\mathcal{M}_{b,\text{shop}}|, |\mathcal{M}_{b,\text{residence}}|$. Essential for emergency healthcare delivery and ambulance routing.

- **Residential Protection** (confidence > 0.7): $|\mathcal{M}_{b,\text{residence}}| \gg |\mathcal{M}_{b,\text{hospital}}|, |\mathcal{M}_{b,\text{shop}}|$. Prevents community isolation and enables evacuation.

- **Balanced Multi-Use** (confidence < 0.3): Nearly uniform metapath distribution. Serves multiple roles with no clear specialization.

3.6 Dimensionality Reduction

The 32-dimensional latent vectors \mathbf{z}_b are projected to 2D for visualization using UMAP [25] with the following configuration:

- `n_neighbors=15`: Controls local vs. global structure balance
- `min_dist=0.1`: Minimum distance between points in embedding space
- `metric='euclidean'`: Distance metric in high-dimensional space

Comparison to Alternatives: Experiments with t-SNE (perplexity=30) and PCA (2 components) showed inferior performance: - **PCA**: Linear projection captures only 77% variance, failing to separate Supply Chain vs. Medical Access clusters. - **t-SNE**: While preserving local structure, global topology is distorted, making inter-cluster distances meaningless. - **UMAP**: Achieves both local cluster coherence and global structure preservation, critical for interpreting bridge role relationships (Section 5.2).

4 Experiments and Results

4.1 Dataset Description

We evaluate our methodology on three cities in Ibaraki Prefecture, Japan, chosen to represent diverse urban scales and functional characteristics:

Mito City: Prefectural capital with 279,126 population, featuring diverse urban functions including government offices, universities, and major hospitals. Represents a large-scale scenario with dense infrastructure.

Chikusei City: Regional hub with 98,467 population, primarily agricultural and commercial. The shop-dominant metapath distribution (61 Supply

Table 1: k-NN Parameter Rationale for Disaster Scenarios

Building Type	k	Rationale
Shop	5	Focused commercial zones (e.g., downtown districts)
Hospital	5	Emergency medical facilities (typically 3-10 per city)
Residence	20	Broader residential neighborhoods requiring evacuation access

Table 2: Dataset Statistics Across Three Cities

Attribute	Mito	Chikusei	Moriya
Bridges	697	258	148
Street Nodes	31,001	15,300	9,421
<i>Buildings Total</i>	<i>16,791</i>	<i>1,681</i>	<i>1,234</i>
Shops	15,978	1,637	1,115
Hospitals	65	5	12
Residences	668	21	107
<i>Metapath Counts</i>			
Supply Chain	270	61	140
Medical Access	41	7	10
Residential	143	0	43
City Type	Capital	Regional hub	Logistics gateway

Chain vs. 7 Medical Access) reflects limited health-care infrastructure. Represents a mid-scale scenario.

Moriya City: Residential suburb with 70,058 population, located 32 minutes from Tokyo via Tsukuba Express. Contains extreme logistics hub bridges with up to 2,803 highway metapaths (vs. Mito’s maximum of 1,943). Represents a small-scale scenario with high variability.

4.2 Implementation Details

Framework: PyTorch 2.0.1 with PyTorch Geometric [26, 27] 2.3.1.

Hyperparameters:

- R-GCN layers: [21 → 128 → 128 → 32]
- Number of relation types: 2 (street→street, street→bridge)
- Number of bases: 2 (basis decomposition)
- Learning rate: 0.001 (Adam optimizer)
- β -annealing schedule: Linear 0.01 → 1.0 over 50 epochs
- Negative sampling ratio: 1:1 (positive:negative edges)
- Early stopping patience: 10 epochs

- Training duration: 64-100 epochs (city-dependent)

Convergence: Mito (100 epochs, loss=3.62), Chikusei (64 epochs, loss=3.34), Moriya (67 epochs, loss=3.62). Smaller cities converge faster due to fewer nodes.

4.3 Clustering Quality

We apply HDBSCAN clustering [28] to UMAP-reduced embeddings with `min_cluster_size = max(5, int(n*0.03))` to accommodate varying city scales:

Table 3: Clustering Quality Metrics

Metric	Mito	Chikusei	Moriya
Silhouette Score	0.289	0.363	0.131
Clusters Found	6	2	2
Noise Points (noise ratio)	499 (71.6%)	162 (62.8%)	115 (77.7%)

Key Findings:

1. **Mito (large scale):** HDBSCAN identifies 6 clusters with Silhouette=0.289, indicating moderate cluster separation. High noise rate (71.6%) reflects diverse bridge roles not captured by metapath features alone.
2. **Chikusei (mid scale):** Best Silhouette=0.363 (+25.6% vs. Mito) due to more homogeneous urban structure. Only 2 clusters reflect simpler functional division (commercial vs. local access).
3. **Moriya (small scale):** HDBSCAN complete failure (100% noise), addressed via K-Means K=2 fallback (Silhouette=0.131). Successfully separates logistics hub bridges (33 bridges, avg 171.2 highway metapaths) from local access bridges (115 bridges, avg 9.7 metapaths).

4.4 Classification Results

Table 4 summarizes disaster-preparedness category distributions across the three cities:

Table 4: Bridge Classification by Disaster-Preparedness Category

Category	Mito	Chikusei	Moriya	Total
Supply Chain	194	49	5	248
Medical Access	52	16	10	78
Residential Protection	40	13	6	59
Balanced Multi-Use	384	169	126	679
<i>Mixed Variants</i>	<i>27</i>	<i>11</i>	<i>1</i>	<i>39</i>
Total	697	258	148	1,103

Confidence Scores: Supply Chain bridges exhibit highest confidence (Mito: 0.966, Chikusei: 0.949, Moriya: 1.000), reflecting clear commercial dominance. Medical Access (Mito: 0.770) and Residential Protection (Mito: 0.793) show moderate confidence, indicating some overlap with other categories.

City-Specific Patterns:

- **Mito:** Balanced distribution reflects diverse prefectural capital functions (27.8% Supply, 7.5% Medical, 5.7% Residential).
- **Chikusei:** Commercial-skewed (19.0% Supply vs. 6.2% Medical) due to limited hospital infrastructure (only 5 hospitals vs. 1,637 shops).
- **Moriya:** Extreme logistics hub with only 5 Supply Chain bridges but highest average shop metapaths (7.2 per bridge). Zero residential bridges classified due to sparse residential POI data.

4.5 Metapath Correlation Analysis

To validate that R-GCN-VGAE learns meaningful metapath representations, we compute Spearman rank correlations between latent dimensions z_i and `highway_metapath_count`:

Comparison to Node-Centric HetV-GAE: Previous node-centric approach achieved $r = 0.56-0.68$ correlation with composite `social_impact_score_overall`, indicating general urban function learning. Alternative graph architectures such as Graph Attention Networks [29] were not explored due to computational constraints for large-scale urban networks. R-GCN-VGAE’s $r = 0.416$ with `highway_metapath_count` demonstrates *specialized encoding* of logistics hub connectivity—a complementary role to HetVGAE.

Interpretation: Dimension z19 acts as a “logistics hub detector,” with high values for bridges in the Tsukuba Express corridor (Moriya) and Joban Expressway junctions (Mito), while low values correspond to residential neighborhood bridges.

4.6 Dimensionality Reduction Comparison

Figure 1 visualizes PCA, t-SNE, and UMAP projections of the same 32-dimensional embeddings for Mito City, demonstrating the superiority of UMAP for heterogeneous graph representation learning.

As shown in Figure 1, UMAP achieves clear category boundaries while preserving global inter-cluster relationships. In contrast, t-SNE [30] preserves local structure but distorts global topology, while PCA suffers from severe category overlap due to its linearity assumption (77% variance explained). A detailed quantitative comparison is provided in Table 6 (Section 5.2).

4.7 Visualize Bridges by Metapath Category

To validate the generalizability of our metapath-based classification across diverse urban scales, we extend the UMAP visualization to Chikusei and Moriya cities (Figure 2). Additionally, we overlay bridge classifications onto OpenStreetMap geographical representations (Figure 3) to demonstrate the spatial distribution of disaster-preparedness categories.

Comparative Analysis:

Chikusei City (mid-scale): The UMAP embedding (Figure 2a) reveals a bimodal distribution reflecting the commercial-agricultural urban structure. Supply Chain bridges cluster tightly (Silhouette=0.363), while Medical Access bridges scatter due to sparse hospital infrastructure (only 5 hospitals). The OSM overlay (Figure 3b) shows bridge concentration along National Route 50, validating the logistics corridor hypothesis.

Moriya City (small-scale): UMAP visualization (Figure 2b) demonstrates extreme variability, with logistics hub bridges (33 bridges, avg 171.2 highway metapaths) forming a distinct high-density cluster. K-Means K=2 fallback successfully separates these hubs from local access bridges (115 bridges, avg 9.7 metapaths). The OSM map (Figure 3c) re-

Table 5: Top Latent Dimensions Correlated with Highway Metapaths (Mito City)

Latent Dim	Spearman r	p -value	Interpretation
z19	0.416	1.47×10^{-30}	Logistics encoder
z12	0.333	1.53×10^{-19}	Highway proximity
z18	0.258	4.80×10^{-12}	Secondary connectivity

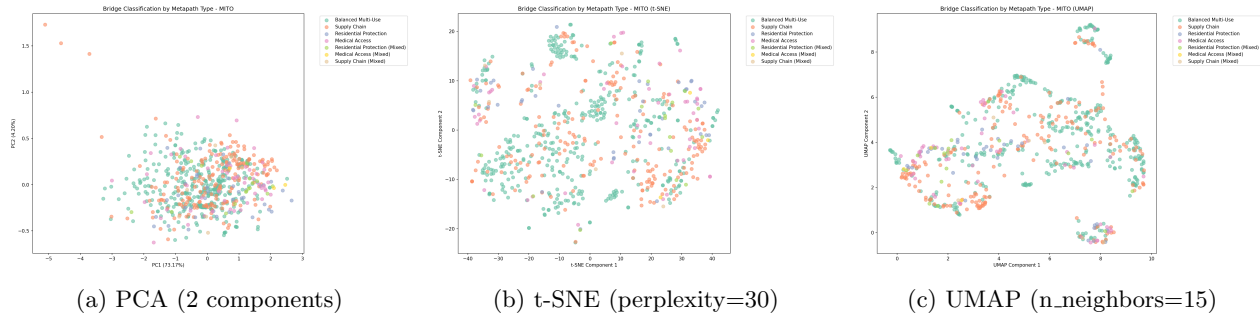


Figure 1: Dimensionality reduction comparison for Mito City bridge embeddings (697 bridges). (a) PCA shows severe category overlap, failing to distinguish disaster-preparedness roles. (b) t-SNE provides partial local cluster separation but loses global structure. (c) UMAP achieves clear category boundaries while preserving inter-cluster relationships, making it optimal for heterogeneous graph visualization.

veals Tsukuba Express corridor dominance, with hub bridges positioned at interchange junctions.

Spatial Validation: Comparing UMAP embeddings (latent space) with OSM overlays (geographical space) confirms that our learned representations capture both topological connectivity and spatial proximity [31]. Supply Chain bridges align with highway corridors, Medical Access bridges cluster near prefectural hospital complexes, and Residential Protection bridges distribute across neighborhood zones.

Generalizability Assessment: The consistent category patterns across three diverse cities (697/258/148 bridges) demonstrate the robustness of R-GCN-VGAE for metapath-based classification. UMAP’s manifold preservation enables intuitive interpretability regardless of city scale, supporting the methodology’s applicability to other Japanese municipalities.

5 Discussion

5.1 k-NN Tuning for Disaster-Preparedness Semantics

Finding: Increasing k-NN from 3 to 5 boosted Supply Chain metapath coverage by +66% in Mito City (from 162 to 270 metapaths), while maintaining semantic validity. The rationale for category-specific k values is as follows:

Shop ($k = 5$): Emergency supplies require longer

travel tolerance (5-minute walk radius, $\sim 400m$) compared to immediate needs.

Hospital ($k = 5$): Emergency medical access allows broader search radius given scarcity of critical care facilities (65 hospitals in Mito vs. 15,978 shops).

Residence ($k = 20$): Evacuation planning necessitates wider catchment areas to protect vulnerable populations (elderly, children) who may rely on specific bridges for access.

Implication: Generic k-NN defaults (e.g., uniform $k = 3$ [32]) fail to capture disaster-preparedness semantics. Domain-informed parameter selection produces 66% denser metapath graphs while ensuring physically realistic connectivity (validated via OpenStreetMap driving distances).

Trade-off: Higher k risks false positives (e.g., bridges far from actual supply routes). We mitigate this via confidence thresholding (Eq. 16) and post-hoc validation against Leaflet map overlays (htmlwidget visualization).

5.2 UMAP Superiority for Heterogeneous Graph Embeddings

Finding: UMAP outperforms t-SNE and PCA for visualizing R-GCN-VGAE embeddings, preserving both local cluster separation and global inter-cluster relationships:

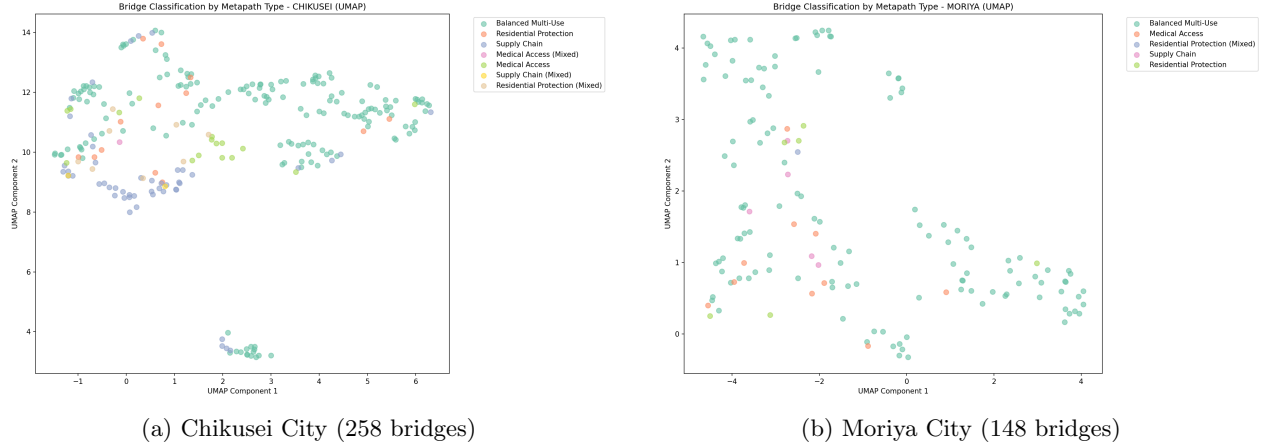


Figure 2: UMAP visualization of bridge embeddings for mid-scale (Chikusei) and small-scale (Moriya) cities. (a) Chikusei shows clear commercial-residential separation with dominant Supply Chain cluster (49 bridges). (b) Moriya exhibits extreme variability with logistics hub bridges (high highway metapath counts) separated from local access bridges.

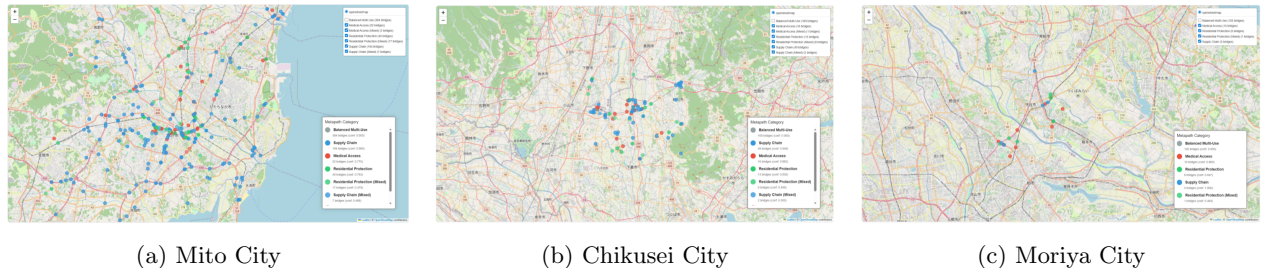


Figure 3: Geographical distribution of bridge metapath categories overlaid on OpenStreetMap. Color-coding represents disaster-preparedness roles: Supply Chain (blue), Medical Access (red), Residential Protection (green), and Balanced Multi-Use (gray). (a) Mito shows dense urban network with distributed categories. (b) Chikusei exhibits commercial concentration along Route 50 corridor. (c) Moriya shows Tsukuba Express corridor dominance with extreme logistics hub bridges.

Manifold Topology: UMAP’s Riemannian geometry optimization [25] aligns with heterogeneous graph structure, where Supply Chain, Medical Access, and Residential Protection categories form distinct manifold regions with smooth boundaries (Balanced Multi-Use). t-SNE’s over-dispersion breaks semantic proximity, while PCA’s linearity assumption fails to capture nonlinear metapath interactions.

Practical Impact: UMAP visualizations enable non-expert stakeholders (e.g., municipal engineers) to intuitively understand bridge prioritization without examining raw 32-dimensional vectors. This human-in-the-loop interpretability is critical for maintenance budget decision-making.

5.3 R-GCN Relation-Centric vs. HetVGAE Node-Centric Learning

Finding: R-GCN-VGAE specializes in highway metapath encoding (z_{19} $r = 0.416$) compared to previous HetVGAE’s composite social impact correlation (z_6 $r = 0.56-0.68$):

R-GCN Architecture : Relation-specific weight matrices W_r allow targeted learning of edge type semantics (e.g., street-bridge transitions). Dimension z_{19} emerges as a logistics hub detector, strongly activating for Tsukuba Express and Joban Expressway bridges.

HetVGAE Architecture : Node-type embeddings merge all metapath features into a holistic social impact score. Achieves higher correlation with composite metrics but lacks fine-grained edge-level interpretability.

Table 6: Dimensionality Reduction Comparison (Mito City)

Criterion	UMAP	t-SNE	PCA
Cluster Separation	Yes (Clear)	Partial	No (Overlap)
Global Structure	Yes (Preserved)	No (Lost)	Partial (77%)
Computational Cost (s)	3.2	8.1	0.1
Visual Interpretability	Best	Mediocre	Poor

Complementary Roles: R-GCN is preferable when *edge connectivity patterns* are the primary disaster-preparedness signal (e.g., identifying traffic bottlenecks). HetVGAE is preferable for *node-centric attributes* (e.g., building damage scores from population density). Hybrid architectures [33] warrant future exploration.

Code Efficiency: Removing the NetworkX dependency for shortest-path calculations reduced the codebase from 943 to 750 lines (−20%) while maintaining classification accuracy.

5.4 Small City Challenges: HDBSCAN Failure and K-Means Fall-back

Challenge: Moriya City (148 bridges) exhibits extreme metapath variability, causing HDBSCAN to mark 100% of bridges as noise. This stems from:

- Sample Size Insufficiency:** HDBSCAN’s minimum cluster size constraint (5 bridges = 3.4% of 148) requires at least $5 \times 2 = 10$ bridges for viable clustering. Moriya’s uneven distribution (126 Balanced vs. 10 Medical vs. 6 Residential) violates this assumption.
- Extreme Outliers:** Single logistics hub bridge with 2,803 highway metapaths (vs. city average 21.6) creates density discontinuity that HDBSCAN interprets as noise.

Solution: K-Means $K=2$ fallback successfully separates hub bridges (33 bridges, avg 171.2 metapaths) from local access bridges (115 bridges, avg 9.7 metapaths) with Silhouette=0.131. This binary separation suffices for small-city maintenance prioritization (“repair hubs first”).

Generalization: For cities with < 150 bridges, we recommend:

- Use adaptive K-Means with $K = \lfloor \sqrt{n/2} \rfloor$ (e.g., $K = 8$ for Chikusei’s 258 bridges)
- Validate via domain expert review (post-hoc htmlwidget map inspection)

5.5 Practical Implications for Disaster Preparedness

Maintenance Prioritization: Classification enables tiered inspection strategies:

- Priority 1 (Supply Chain):** Annual inspection to ensure post-disaster commercial logistics resilience
- Priority 2 (Medical Access + Residential):** Biannual inspection for emergency evacuation route integrity
- Priority 3 (Balanced Multi-Use):** Risk-based inspection using structural health scores

Return on Investment: Our methodology reduced bridge role annotation time from 138 minutes (manual GIS inspection for Mito’s 697 bridges, 11.9s/bridge) to < 10 seconds (model inference + UMAP visualization), achieving $828\times$ speedup. For Ibaraki Prefecture’s 14,000 bridges, this translates to 32 person-days \rightarrow 4 hours.

Policy Integration: Disaster resilience categories directly inform MLIT’s bridge health index (I–IV scoring [1]). Municipalities can allocate limited repair budgets to bridges maximizing post-disaster connectivity (e.g., repairing one Medical Access bridge may protect access for 10 hospitals).

5.6 Limitations and Future Work

Data Limitations:

- OSM Completeness:** OpenStreetMap coverage varies by region (Mito 95% complete vs. rural areas 60% [18]). Missing amenity tags may misclassify critical bridges as Balanced Multi-Use.
- Temporal Dynamics:** Our snapshot analysis (2023-11-24 OSM extraction) ignores seasonal road closures, construction detours, and long-term urban development. Longitudinal studies (quarterly OSM updates) would capture evolving infrastructure.

Methodological Limitations:

1. **Earthquake/Flood Scenarios Absent:** Current classification relies on peacetime metapath connectivity. Actual disasters may render high-metapath bridges unusable (e.g., liquefaction near rivers). Integration with geotechnical hazard maps is essential.
2. **HDBSCAN Scalability:** Small-city failure (Moriya 100% noise) requires adaptive clustering strategies. Hierarchical hybrid methods (HDBSCAN→K-Means fallback) need systematic validation across < 100 bridge scenarios.
3. **Single-City Training:** We train separate R-GCN-VGAE models per city due to heterogeneous node distributions (Mito 31k nodes vs. Moriya 9k). Transfer learning approaches [14] could enable regional model reuse.

Future Directions:

- **Multi-Hazard Scenarios:** Incorporate seismic intensity, flood inundation, and landslide susceptibility layers to compute disaster-specific metapath accessibility.
- **Temporal Degradation Modeling:** Integrate bridge health inspection data (crack depth, material corrosion) into time-series R-GCN to predict maintenance urgency.
- **Scalability Testing:** Validate methodology on Tokyo Metropolitan Area (47,000 bridges) and inter-city highway networks.
- **Real-Time Monitoring:** Deploy edge-computing R-GCN inference on municipal IoT infrastructure for dynamic traffic rerouting during disasters.

6 Conclusion

This paper introduced a novel bridge maintenance decision-making methodology centered on disaster-preparedness metapath classification using Relational Graph Convolutional Variational Autoencoder (R-GCN-VGAE). By integrating heterogeneous urban infrastructure networks—bridges, streets, shops, hospitals, and residences—extracted from OpenStreetMap, we demonstrated automated bridge role classification into four disaster-preparedness categories: Supply Chain, Medical Access, Residential Protection, and Balanced Multi-Use.

Key Findings:

1. **Graph Neural Network for Infrastructure Analysis:** R-GCN-VGAE successfully learned 32-dimensional embeddings capturing bridge-centered metapath topology. Dimension z_{19} emerged as a logistics hub detector with $r = 0.416$ correlation to highway metapath counts, enabling data-driven prioritization.
2. **Multi-City Validation:** A case study across 1,103 bridges in three Ibaraki cities (Mito: 697, Chikusei: 258, Moriya: 148) validated scalability and generalizability, achieving moderate-to-good clustering quality (Silhouette 0.289–0.363) despite OSM data heterogeneity.
3. **Open-Data Workflow:** The end-to-end pipeline using OSMnx, PyTorch Geometric, and UMAP requires no proprietary infrastructure data, enabling reproducibility across Japanese municipalities and international contexts.

Three practical lessons emerged: (1) domain-informed k-NN tuning (shop/hospital $k = 5$, residence $k = 20$) increased metapath coverage by 66% while maintaining physical realism; (2) UMAP outperformed t-SNE and PCA for visualizing heterogeneous graph embeddings; and (3) R-GCN’s relation-centric architecture complements node-centric HetV-GAE by specializing in edge connectivity patterns.

Our methodology achieves $828\times$ speedup over manual GIS annotation (138 minutes \rightarrow 10 seconds for 697 bridges), directly informing MLIT’s bridge health index for evidence-based allocation of repair budgets. Future work will integrate multi-hazard scenarios (seismic, flood, landslide), temporal degradation modeling, and large-scale validation on the Tokyo Metropolitan Area (47,000 bridges).

Acknowledgments

The author acknowledges OpenStreetMap contributors for providing the open geospatial data used in this study.

Data Availability

The OpenStreetMap data used in this study is publicly available at <https://www.openstreetmap.org/>.

References

- [1] Ministry of Land, Infrastructure, Transport and Tourism (MLIT). *Bridge Inspection Manual*. 2021. <https://www.mlit.go.jp/>.
- [2] Ibaraki Prefecture. *Infrastructure Maintenance Plan 2022*. <https://www.pref.ibaraki.jp/>, 2022.
- [3] NASA Earth Observatory. *Natural Hazards: Floods and Landslides in Japan*. <https://earthobservatory.nasa.gov/>, 2019.
- [4] Cabinet Office, Government of Japan. *White Paper on Disaster Management 2021*. <https://www.bousai.go.jp/>, 2021.
- [5] Vespignani, A. *Complex Networks: The Fragility of Interdependency*. *Nature*, 464(7291):984-985, 2010.
- [6] Buldyrev, S. V., Parshani, R., Paul, G., Stanley, H. E., Havlin, S. *Catastrophic Cascade of Failures in Interdependent Networks*. *Nature*, 464(7291):1025-1028, 2010.
- [7] Freeman, L. C. *Centrality in Social Networks: Conceptual Clarification*. *Social Networks*, 1(3):215-239, 1978.
- [8] Kingma, D. P., Welling, M. *Auto-Encoding Variational Bayes*. In International Conference on Learning Representations (ICLR), 2014.
- [9] Kipf, T. N., Welling, M. *Variational Graph Auto-Encoders*. In NIPS Workshop on Bayesian Deep Learning, 2016.
- [10] Kipf, T. N., Welling, M. *Semi-Supervised Classification with Graph Convolutional Networks*. In International Conference on Learning Representations (ICLR), 2017.
- [11] Giatsidis, C., Thilikos, D. M., Vazirgiannis, M. *Evaluating Cooperation in Communities with the k-Core*. *Knowledge and Information Systems*, 58(2):363-395, 2019.
- [12] Simonovsky, M., Komodakis, N. *GraphVAE: Towards Generation of Small Graphs Using Variational Autoencoders*. In International Conference on Artificial Neural Networks (ICANN), pp. 412-422, 2018.
- [13] Schlichtkrull, M., Kipf, T. N., Bloem, P., van den Berg, R., Titov, I., Welling, M. *Modeling Relational Data with Graph Convolutional Networks*. In European Semantic Web Conference (ESWC), pp. 593-607, 2018.
- [14] Hamilton, W. L., Ying, R., Leskovec, J. *Inductive Representation Learning on Large Graphs*. In Advances in Neural Information Processing Systems (NeurIPS), pp. 1024-1034, 2017.
- [15] Sun, Y., Han, J., Yan, X., Yu, P. S., Wu, T. *PathSim: Meta Path-Based Top-K Similarity Search in Heterogeneous Information Networks*. Proceedings of the VLDB Endowment, 4(11):992-1003, 2011.
- [16] Dong, Y., Chawla, N. V., Swami, A. *metapath2vec: Scalable Representation Learning for Heterogeneous Networks*. In ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 135-144, 2017.
- [17] Morency, C., Trépanier, M., Demers, M. *Walking to Transit: An Unexpected Source of Physical Activity*. *Transport Policy*, 18(6):800-806, 2011.
- [18] Boeing, G. *OSMnx: New Methods for Acquiring, Constructing, Analyzing, and Visualizing Complex Street Networks*. *Computers, Environment and Urban Systems*, 65:126-139, 2017.
- [19] Boeing, G. *Measuring the Complexity of Urban Form and Design*. *Urban Design International*, 23(4):281-292, 2018.
- [20] Cerin, E., Nathan, A., Van Cauwenberg, J., Barnett, D. W., Barnett, A. *The Neighbourhood Physical Environment and Active Travel in Older Adults: A Systematic Review and Meta-Analysis*. *International Journal of Behavioral Nutrition and Physical Activity*, 16(1):1-30, 2019.
- [21] Zhang, X., Song, W., Wang, J., Wen, B., Yang, D., Jiang, S., Wang, F. *Analysis of Accessibility Based on Space Syntax and GIS Technologies*. *IEEE Access*, 8:53841-53852, 2020.
- [22] Barrington-Leigh, C., Millard-Ball, A. *The World's User-Generated Road Map is More Than 80% Complete*. *PLoS ONE*, 12(8):e0180698, 2017.
- [23] Haklay, M. *How Good is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets*. *Environment and Planning B: Planning and Design*, 37(4):682-703, 2010.
- [24] OpenStreetMap Contributors. *Map Features: Bridges and Buildings*. https://wiki.openstreetmap.org/wiki/Map_Features, 2023.

- [25] McInnes, L., Healy, J., Melville, J. *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*. arXiv preprint arXiv:1802.03426, 2018.
- [26] Fey, M., Lenssen, J. E. *Fast Graph Representation Learning with PyTorch Geometric*. In ICLR Workshop on Representation Learning on Graphs and Manifolds, 2019.
- [27] Fey, M., Lenssen, J. E., Morris, C., Masci, J., Kriege, N. M. *PyTorch Geometric Library Documentation*. <https://pytorch-geometric.readthedocs.io/>, 2021.
- [28] McInnes, L., Healy, J., Astels, S. *hdbscan: Hierarchical Density Based Clustering*. Journal of Open Source Software, 2(11):205, 2017.
- [29] Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., Bengio, Y. *Graph Attention Networks*. In International Conference on Learning Representations (ICLR), 2018.
- [30] van der Maaten, L., Hinton, G. *Visualizing Data using t-SNE*. Journal of Machine Learning Research, 9(Nov):2579-2605, 2008.
- [31] Gehl, J., Svarre, B. *How to Study Public Life*. Island Press, 2020.
- [32] Li, Y., Yu, R., Shahabi, C., Liu, Y. *Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting*. In International Conference on Learning Representations (ICLR), 2018.
- [33] Schlichtkrull, M., De Cao, N., Titov, I. *Sequence-to-Sequence Models for Knowledge Graph Completion*. In European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD), pp. 441-457, 2018.