

ARTICLE OPEN



Urbanity: automated modelling and analysis of multidimensional networks in cities

Winston Yap¹, Rudi Stouffs¹ and Filip Biljecki^{1,2}✉

Urban networks play a vital role in connecting multiple urban components and developing our understanding of cities and urban systems. Despite the significant progress we have made in understanding how city networks are connected and spread out, we still have a lot to learn about the meaning and context of these networks. The increasing availability of open data offers opportunities to supplement urban networks with specific location information and create more expressive urban machine-learning models. In this work, we introduce Urbanity, a network-based Python package to automate the construction of feature-rich urban networks anywhere and at any geographical scale. We discuss data sources, the features of our software, and a set of data representing the networks of five major cities around the world. We also test the usefulness of added context in our networks by classifying different types of connections within a single network. Our findings extend accumulated knowledge about how spaces and flows within city networks work, and affirm the importance of contextual features for analyzing city networks.

npj Urban Sustainability (2023)3:45; <https://doi.org/10.1038/s42949-023-00125-w>

INTRODUCTION

Cities are increasingly being recognised as complex urban systems^{1–5}. From this viewpoint, numerous interdependent social, economic, and environmental components merge from the ground-up to comprise what we come to acquaint with as ‘urban’. As the lifeblood of cities, urban street networks play a vital role in connecting multiple urban components and developing our understanding of cities.

In recent years, having more detailed information about cities and ways of analyzing that data have led to many city network analytical use cases—spatial homogeneity⁶, human development⁷, active mobility^{8,9}, colocation and social bonds¹⁰, retail demand estimation¹¹, traffic forecasting¹², and urban wealth scaling¹³. All these changes have greatly helped us understand how the spatial structure and connectivity of networks influence specific parts of city systems. In reality, city networks are complex because they involve multiple city elements interacting and changing together^{14,15}. For instance, how many people choose to walk or bike along a certain route can depend on many factors, like how easily they can reach their destination, how pedestrian-friendly the streets are, and how many people live in the area (which has become even more important because of Covid-19). In this scenario, standard methods for studying city networks that only focus on the structure and flow within these networks, such as how many connections a point has or how movement happens in the network, only capture a part of the rich diversity found in city landscapes¹⁶.

Current studies show that city networks have a high degree of similarity in their structures due to real-world physical constraints, which makes it hard for many computational algorithms to learn from them¹⁷. To fix this problem, researchers have suggested using more detailed deep learning models that use location-specific features^{18–20}. While the idea of using diverse city indicators to better represent the complexity of a city isn't new, existing solutions (e.g. Knowledge Graphs, Planning Support Systems, City Information Models, Digital Twins) pose significant

barriers to adoption due to immense computational and human resource requirements^{21,22}. Not discounting the importance of aforementioned approaches, the lack of accessible tools persistently limits our ability to conduct open science and reason with the complexity of our urban environments^{23–26}.

To address these challenges, we introduce Urbanity, a network-based Python package that helps create detailed city networks worldwide in an automated way. This network-based method has several advantages: (1) it is a scalable and lightweight way to represent data, (2) it can be used with many different analysis and modelling methods, including graph machine learning and complex network analysis, and (3) it can be extended to work with both within-city and across-city analysis tasks. With Urbanity, our goal is to help make the study of complex city systems easier and more comprehensive. It gives urban researchers and practitioners an easy way to (1) automatically build typical city networks at any scale, (2) add diverse city indicators like building shapes, street views, population counts, and points of interest to network nodes, and (3) get summary city indicators for any geographical area they're interested in. At its core, Urbanity aims to bring our computational tools and theoretical understanding of cities as complex urban systems one step closer to each other^{27,28}. We proceed with a brief background on how city networks are commonly represented. In the process, we discuss and compare Urbanity with related tools for urban network analysis. Subsequently, we describe data acquisition and preprocessing strategies underlying the core modules of our package. We'll also present a unique set of city network data from five major cities across Asia, Europe, and North America. We use this data to (1) study the differences in city networks within and across cities, and (2) test our tool's ability to predict road categories for different cities. Our experiments show that adding these different types of contextual features significantly improves predictive accuracy for graph deep-learning tasks. We wrap up with a discussion about what this means for complex city systems, current limitations, and our plans for future development.

¹Department of Architecture, National University of Singapore, Singapore, Singapore. ²Department of Real Estate, National University of Singapore, Singapore, Singapore. ✉email: filip@nus.edu.sg

Urban networks manifest themselves most visibly through our streets. Uncoincidentally, archetypal representations of urban networks are street-centric—primal planar and dual networks. Primal planar networks are the mainstay of urban networks and represent street intersections as nodes and street segments as non-intersecting edges. On the other hand, dual networks consider nodes as street segments, and edges represent contiguity between connected streets. Primal planar graphs are often translated into dual graphs (see Supplementary Fig. 1). At its simplest, urban networks can be represented by undirected, distance-weighted graphs. More sophisticated network models may introduce additional properties on nodes and edges of the graph—edge direction or multi-typed nodes/edges. Without loss of generality, we define a network as $G = (N, E)$, where N and E refer to the set of nodes/vertices and edges/arcs of the graph, respectively. Both types of network representations are closely related functionally but show different aspects of city networks^{29,30}.

Both primal and dual network representations are well-represented by urban network analytical tools. We overview the landscape of related tools, discuss their functionality, and justify our development efforts (see Table 1). Our review excludes tools for general network analysis (e.g. NetworkX, igraph, Gephi, graph-tools), reflecting a consistent focus on the urban planning applications and use cases.

We observe that most tools focus primarily on distance-based measurements (metric) and connectivity (topological) network indicators. While some tools such as OSMnx, Place Syntax Tool, and cityseer allow users to compute contextual indicators (e.g. building footprints and land use), it is not the primary objective of those tools. On functionality, urban accessibility and network clustering form the central component among tools.

Our development efforts extend the current eco-system of urban network tools on three fronts: (1) user interface support; (2) enhanced network feature representation; (3) efficient benchmarking and comparative functionalities. Firstly, Urbanity offers a network-based mapping interface that is easy to understand and navigate for urban planners. This mapping interface helps users to quickly specify and confirm site boundaries, which can be a complicated coding process when dealing with complex site boundaries. Secondly, Urbanity supports subsequent analytical tasks by automatically integrating context-based and semantic indicators into city networks. With improved feature representation, Urbanity can assist in various applications like multi-criteria site analysis, graph predictive modelling, and geospatial visualisation of networks. For instance, Urbanity can help planners identify key areas for age-friendly planning and design by filtering network locations to include areas with a high proportion of older adults, poor streetscape conditions (like a lack of greenery), and a lack of amenities. Last but not least, Urbanity supports benchmarking and comparative studies between urban networks by allowing rapid, consistent extraction of aggregate spatial information for any geographic area of interest.

RESULTS

Descriptive summary

On initial impression, urban networks seem relatively small and low-dimensional when compared to their counterparts in the biological or social sciences³¹. Yet, traditional measures of network size (e.g., count of nodes and edges) belie the multitude of distinctions that define cities and their inhabitants^{32,33}. Notably, as our ability to capture and process large scale urban data improves, the inherent complexities within urban systems become more apparent^{34,35}. In this vein, we introduce the Urbanity global network dataset (an ongoing initiative) that aims to promote complex network analysis and machine learning for global urban

Table 1. Landscape of tools for urban network analysis.

Reference	Tool	OS ^a	Functionality				Indicators											
			Primal	Dual	Wayfinding	Urban accessibility	Cluster and segment	Network interpolation	Network construction	Metric	Topological	Semantic	Contextual					
77	OSMnx	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	
90	Pandana	●	-	-	●	●	●	●	●	●	●	●	●	●	●	●	●	●
91	spaghetti	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
92	SANET	-	-	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
93	QNEAT3	●	-	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
94	UNA	-	-	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
95	cityseer	●	●	-	●	●	●	●	●	●	●	●	●	●	●	●	●	●
96	Place Syntax Tool	●	●	-	●	●	●	●	●	●	●	●	●	●	●	●	●	●
97	COINS	●	-	-	●	●	●	●	●	●	●	●	●	●	●	●	●	●
98	depthmapX	●	●	-	●	●	●	●	●	●	●	●	●	●	●	●	●	●

^aOS—Open-Source.

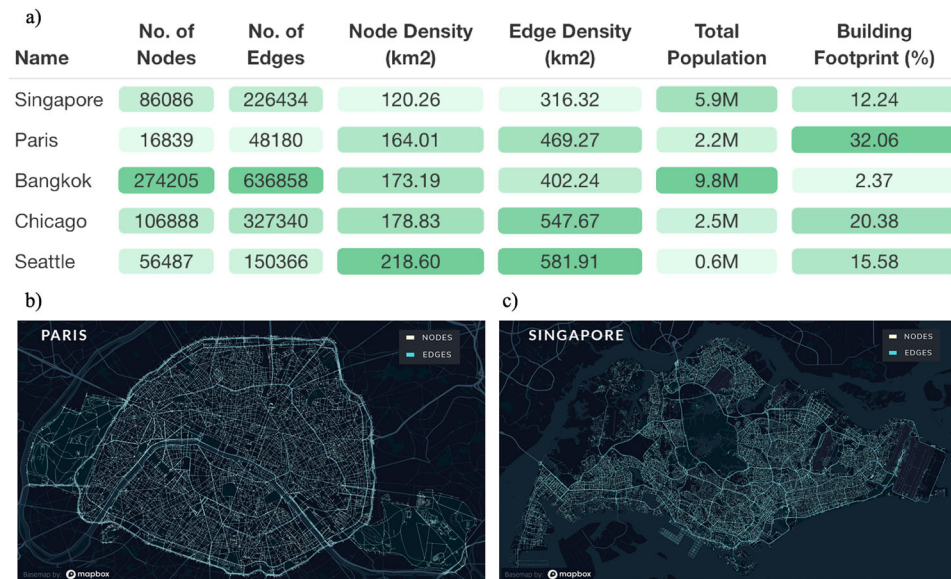


Fig. 1 Aggregate network indicators and network spatial structure. **a** Aggregate indicators for five global cities. Urbanity automates the computation of contextual and semantic network indicators (e.g. population count and building morphology), producing a feature-rich urban network. Indicators help to support efficient multi-criteria site analysis. **b** Spatial network of Paris reveals a concentric ring pattern of emergence. **c** Spatial network of Singapore with hub-and-spoke transit structure. Basemap: [OpenStreetMap](#) and [Mapbox](#).

networks. An overview of selected descriptive attributes and network visualisations for each city is presented in Fig. 1.

At first glance, we observe palpable differences between cities. For example, Bangkok has the highest population count and the lowest proportion of building footprints among cities, indicating evidence of urban sprawl³⁶. On the other hand, Paris, Chicago, and Seattle have less than half as many nodes and edges as Bangkok, but their network density is relatively high, reflecting a dense and intricately planned urban mobility network. Compared to other global cities, Singapore stands out as a city-state with a high population counts coupled with low building footprint and network density. Notwithstanding, Singapore is one of the densest cities to live in with a population to built-up area ratio of approximately 68,000 persons per km². In comparison, Paris and Chicago measure at 66,600 persons per km² and 20,600 persons per km², respectively.

The following section demonstrates how indicators derived from Urbanity can be used for descriptive, and predictive urban analytical tasks to better understand urban complexity within and between cities.

Segregation within urban networks

Urban areas closer to one another are more likely to be related than areas further apart³⁷. However, cities commonly exhibit patterns deviating from spatial homogeneity and often display signs of segregation^{38,39}. In this regard, urban networks provide a powerful lens for understanding how semantic values correlate throughout the urban fabric between connected locations. Network assortativity measures the extent to which nodes with the same properties are joined to one another⁴⁰.

Using Singapore as a case study, we observe a high and consistent network-based correlation (assortativity) in green view index (GVI) throughout Singapore (see Fig. 2). Our observation suggests that urban greenery between adjacent locations in Singapore is generally well balanced, reflecting the success of early planning efforts pursued by the city state^{41,42}.

On the other hand, Fig. 2 shows urban greenery to be sporadically distributed throughout Bangkok. Sites with high assortativity for GVI coincide with the location of existing green zone regulations in Bangkok³⁶. These findings suggest the

importance of green zone protection strategies in maintaining equitable urban greenery access in Bangkok.

The World Health Organisation predicts that by 2030, one in every six people globally will be over the age of 60. As population aging accelerates, cities will face an increased prevalence of dementia among older adults. Planning and designing dementia-friendly neighborhoods will become crucial, with a focus on creating walkable communities. Green streets encourage more walking, providing a comprehensive set of health benefits for older adults⁴³. Urbanity, with its capacity to provide detailed street-level information, can facilitate more precise and effective urban planning strategies. This is particularly beneficial for communities striving to meet the challenges presented by a rapidly aging population and a rise in dementia cases.

Network and building complexity across cities

Here we look beyond individual cities and examine how complex network indicators can help to benchmark and compare urban structures within and between cities. Figure 3 presents a comparison of node density against mean building complexity for five cities and their subzones. Both indicators relate directly to the complexity and density of cities, suggesting implications for a city's resilience and sustainability.

We observe several distinct patterns between cities. For example, subzones in Paris display high homogeneity which corresponds to a long history of centralised planning under the 'Haussmann period'⁴⁴. On the other hand, subzones across Singapore show the largest variance. This observation makes intuitive sense as Singapore faces diverse land use requirements as a global city state without a hinterland.

Another interesting observation relates to a striking positive relationship between mean building complexity and node density across subzones in Bangkok, Chicago, and Seattle which could be explained by the presence of a centralised and economic-driven planning context⁴⁵. Among this group, American subzones are identified by low average mean building complexity—a characteristic of the well-known urban block morphology. As an exception in this group, the Magnificent Mile district (a premier arts and commercial district) in Chicago, displays high mean building complexity.

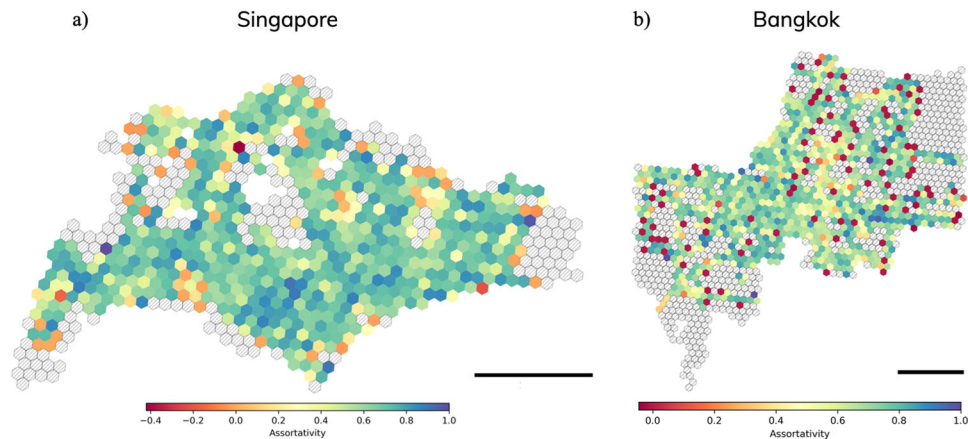


Fig. 2 Co-similarity of urban greenery across city networks. Scale bars correspond to 10 kilometres. Grid cells coloured dark blue indicate high similarity of urban greenery between adjacent locations. On the other hand, red grid cells suggest segregation of urban greenery between adjacent locations. The white cells and the grey hashed cells correspond to areas with no spatial overlap with the underlying network, and to areas reporting an observation count of less than 30, respectively. **a** Urban greenery throughout Singapore is generally well-balanced with high assortativity, indicating the success of Singapore’s “city in a garden” policy. **b** Bangkok shows signs of segregation with sporadically distributed urban greenery pattern. Sites with high assortativity correspond to existing green zone regulations, indicating their importance for promoting equitable urban greenery access in Bangkok.

Classification of urban streets

Urban streets are intrinsically complex, showing self-similarity and self-organisation across scale. On a more fundamental level, streets function as multi-faceted urban elements that feature a rich diversity of social, economic, and environmental activities. Correspondingly, the task of predicting street category (e.g. primary, secondary, or arterial) is a complex endeavour since street categories depend not only on the geometric properties of streets but also on their semantic and contextual dimensions. Reliable estimates for street categories provide many useful applications and use cases, not limited to, understanding urban hierarchy, clustering and segmentation of road networks, estimating urban air and noise pollution, modelling urban mobility flows, and assessing disaster response.

We specify a transductive (within-city), graph neural network edge classification task to predict road classification labels for each city in the Urbanity global network dataset. Based on OSM tag information, we categorised road type into five main hierarchical categories—national (1); regional (2); precinct (3); neighbourhood (4); and local access roads (5). For each city, we randomly split network edges into a training and validation set with 80:20 ratio. To obtain edge feature embeddings, we adopt the cross-attention method which appends adjacent node features. We evaluate feature importance for prediction through ablation studies and report mean classification accuracy across each model configuration. We run each model configuration for 500 epochs with hidden dimension size = 64 and learning rate = 0.01. All models were implemented with the PyTorch Geometric graph deep learning framework⁴⁶. Table 2 lists feature ablation results for each city and corresponding feature set.

Feature ablation shows that the inclusion of semantic and contextual indicators provides clear, consistent improvements in model predictive performance across all cities and model architecture. For instance, Paris and Singapore saw the largest increase in mean classification accuracy by up to 11.7% and 7.9%, respectively. On the other hand, Chicago, Seattle, and Bangkok saw marginal increase of between 3 to 5%. In line with graph machine learning literature, we find that GAT and GraphSAGE architectures consistently outperform standard GCN in terms of predictive accuracy. Figure 4 shows multi-class performance for models with the highest accuracy. In general, neural network (GNN) models deliver consistent performance across all categories. For Bangkok, poorer predictive performance on neighbourhood

and local access categories might be attributed to data quality issues.

Another interesting observation is the sharp drop in model performance across all cities when metric and topological measures are removed. Intuitively, this makes sense as existing graph machine learning architectures such as GCN⁴⁷, GraphSAGE⁴⁸, and GAT⁴⁹ are designed to aggregate information from graph neighbourhood structure. These results present opportunities to explore various graph aggregation and learning architectures that improve the utility of contextual and semantic attributes.

Our experiments further underline the importance of context-based analysis. For example, smaller node buffer radius led to an increase in predictive performance for Paris, Singapore, and Chicago, while performance decreased for Seattle and Bangkok. Findings show that cities are fundamentally different, providing support for 1) comparative analysis of cities; and 2) context-based feature representation, and model parameter tuning for machine learning models. Our findings suggest that there is no one-size-fits all solution—parameters such as buffer bandwidth should be tuned according to geographical context.

DISCUSSION

We maintain our position that urban networks form a powerful, multi-scalar focal lens to examine the complexities embedded within and across cities. Streets, as multi-faceted conduits of urban life, encompass much more than linear movement—sustaining numerous vital human connection and functions. Throughout this paper, we argue for a more comprehensive approach to urban network analytics that accounts for diversity of urban functions embedded within our streets. In the face of complex urban challenges, an interdisciplinary approach to designing systemic solutions for cities is fundamental⁵⁰.

As a way forward, Urbanity aims to promote theoretical and empirical consolidation in the study of complex urban networks on four aspects—(1) reconciling ongoing network studies with traditional planning theory; (2) extending current planning approaches beyond urban physicalism; (3) exploring linkages with the emerging field of GeoAI; and (4) improving benchmarking practices for the empirical evaluation of urban networks.

On reconciliation with planning theory, current street network studies and tools prejudice a technical interpretation of urban

Comparison across Cities and Subzones Node Density vs Mean Building Complexity

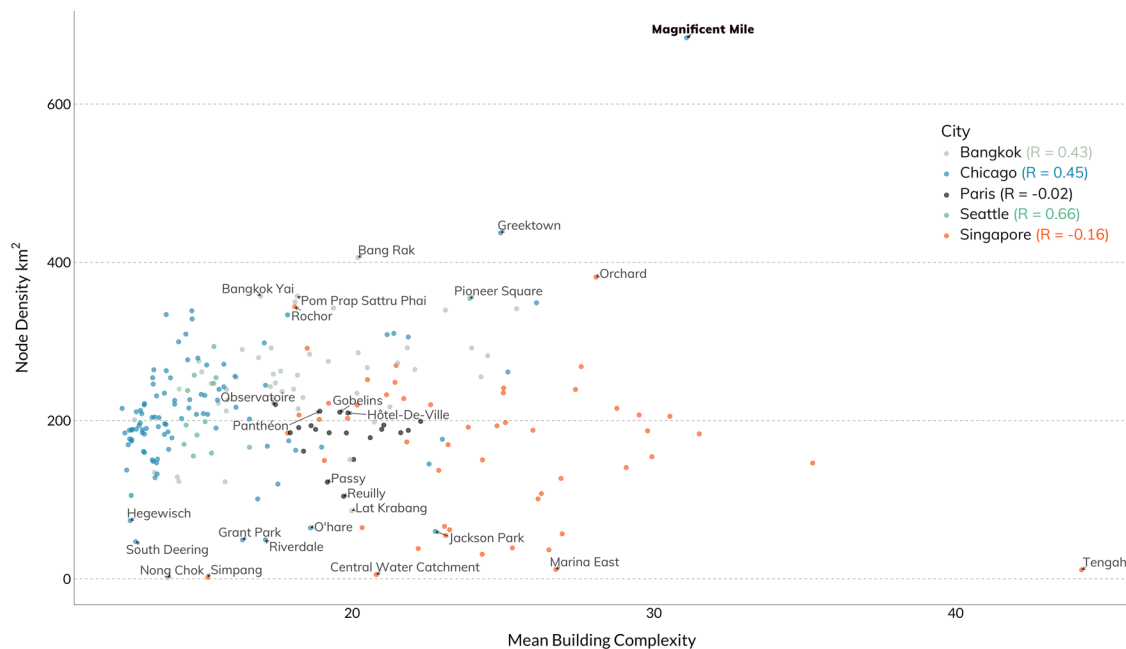


Fig. 3 Comparison of node density and building complexity across five cities. Each point on the scatter plot indicates a city's subzone. Points are coloured according to their respective city. Clear patterns and clusters emerge across different cities that correspond intuitively with each city's planning history and context. Parisian subzones share high co-similarity, corresponding to a long history of centralised planning. On the other hand, Singaporean subzones have the highest diversity, reflecting the diverse national land use requirements faced by the global city state. Seattle, Chicago, and Bangkok exhibit a positive association between node density and mean building complexity which hints at implicit similarities between these cosmopolitan cities.

streets that is largely based on graph theory or the network sciences (see Table 1). While such approaches have proven useful to understand the structure and connectivity of streets, they remain disconnected to the qualitative, ground-up interpretations of streets that is characteristic of traditional urban studies^{51,52}. As argued by⁵³: "there is nothing simple about that order (of streets) itself, or the bewildering number of components that go into it. Most of those components are specialized in one way or another. They unite in their joint effect upon the sidewalk, which is not specialized in the least". Towards a fuller contemplation of urban networks in their multi-faceted entirety, it is important to reconcile this disconnect and recognise the bottom-up, self-organised complexity of streets^{54,55}. In this regard, Urbanity offers an extensible basis to promote integration between the physical and social aspect of complex networks and their interconnected interactions. Specifically, our framework will lay the foundation to develop cross-over models⁵⁶ that consider social (actor-based) interactions within physical networks. These developments will bring network analytics closer to urban studies by allowing for fuller engagement with the human and experiential dimension of streets.

Secondly, while the physical form of cities is undeniably a key aspect in our understanding and modelling of urban environments, it is clear that a host of other contextual factors play a crucial role in forming a well-rounded understanding of cities. Emerging evidence suggest that form is not guaranteed to follow function⁵⁷. Cities are intricate and constantly evolving systems that don't settle into a predictable, static condition. Urbanity, by linking a variety of urban indicators (uniting factors of urban demand and supply), offers an exploratory framework for urban planners. This framework goes beyond the physicalism of cities and paves the way for methods to envision, measure, and uncover

connections between diverse, interlinked components of urban environments.

Current developments in the fields of GeoAI and machine learning provide opportunities to examine urban networks at an unprecedented scale⁵⁸. Although such computational methods allow us to make sense of a fast emerging urban data landscape, they also implicitly influence our perceptions and thought processes of the city⁵⁹. In this regard, it is important to realise that optimization might not be a universal/sustainable goal across urban systems. For instance, cities planners often have to recognise and make seemingly 'non-optimal' decisions to ensure that planning is carried out in an equitable, inclusive, and collaborative manner. A notable example is Bogotá, the first city in South America to address needs related to care work with its innovative 'Block of Care' framework, which targets support and development for systematically marginalised women caregivers. Urbanity supports such use cases by allowing planners to pinpoint optimal care block locations with high proportion of women and children across neighbourhoods. Data on streetscapes, like greenery and visual information, can also be used to evaluate traffic safety around care blocks. To promote equitable and sustainable planning, the integration of machine learning should be both transparent, explainable, and accountable to all stakeholders. Ideally, this involves crafting a clear implementation technology adoption roadmap that outlines specific goals, trade-offs, model biases, and the extent of stakeholder involvement. We must stay vigilant against criticisms of technological determinism; computational methods, in the quest for efficiency, should neither undermine established good planning practices such as participatory and communicative elements nor override traditional planning wisdom. Domain knowledge and theory play a fundamental role in safeguarding against model predictive biases. The inclusion of these critical aspects facilitates a more nuanced urban

Table 2. Feature ablation for edge classification across different cities.

Model ^a	All ^b (MA)	only M/T (MA)	w/o M/T (MA)	w/o Bldn (MA)	w/o Popn (MA)	w/o POIs (MA)	w/o SVI (MA)
<i>Singapore</i>							
GCN _{100M}	72.4%	67.2%	57.7%	70.3%	72.0%	72.0%	71.8%
GAT _{100M}	77.0%	69.7%	64.9%	74.6%	77.1%	76.1%	75.1%
GraphSAGE _{100M}	75.1%	69.8%	53.7%	72.3%	74.4%	74.5%	74.2%
GCN _{200M}	72.5%	67.5%	56.6%	70.4%	71.8%	72.0%	71.4%
GAT _{200M}	76.9%	69.0%	63.6%	74.5%	75.5%	75.0%	75.2%
GraphSAGE _{200M}	75.3%	70.7%	48.6%	72.9%	74.0%	74.1%	74.4%
<i>Paris</i>							
GCN _{100M}	71.0%	64.3%	64.2%	70.8%	70.3%	69.8%	70.1%
GAT _{100M}	78.9%	67.8%	70.7%	77.9%	79.3%	75.9%	77.3%
GraphSAGE _{100M}	74.0%	66.1%	64.4%	73.1%	73.2%	72.1%	72.6%
GCN _{200M}	69.6%	62.8%	62.8%	68.9%	68.4%	68.1%	68.8%
GAT _{200M}	77.9%	67.2%	65.3%	75.2%	76.9%	74.6%	75.8%
GraphSAGE _{200M}	72.4%	65.6%	60.3%	70.0%	70.9%	70.7%	70.9%
<i>Bangkok</i>							
GCN _{100M}	73.8%	71.0%	68.8%	72.9%	73.2%	73.6%	73.3%
GAT _{100M}	76.8%	72.5%	70.3%	76.3%	76.8%	76.3%	76.5%
GraphSAGE _{100M}	75.9%	72.5%	67.7%	74.7%	75.0%	75.0%	75.0%
GCN _{200M}	73.2%	70.5%	68.4%	72.7%	72.3%	72.8%	73.1%
GAT _{200M}	77.3%	72.1%	69.5%	74.9%	76.3%	75.9%	75.5%
GraphSAGE _{200M}	74.9%	71.9%	67.7%	74.0%	74.6%	74.2%	74.5%
<i>Chicago</i>							
GCN _{100M}	64.5%	64.6%	55.7%	64.0%	64.1%	64.0%	63.7%
GAT _{100M}	74.6%	72.8%	58.7%	74.0%	75.4%	74.5%	74.5%
GraphSAGE _{100M}	76.5%	73.9%	59.0%	75.3%	76.5%	75.8%	76.2%
GCN _{200M}	63.8%	63.5%	54.1%	63.4%	63.2%	64.6%	63.9%
GAT _{200M}	73.4%	72.4%	56.9%	74.2%	74.1%	74.3%	73.8%
GraphSAGE _{200M}	75.9%	74.2%	58.3%	75.0%	75.7%	75.9%	75.9%
<i>Seattle</i>							
GCN _{100M}	71.4%	69.4%	58.8%	71.2%	71.8%	71.7%	71.5%
GAT _{100M}	79.1%	72.6%	63.6%	77.8%	79.1%	78.1%	77.4%
GraphSAGE _{100M}	76.1%	72.6%	56.2%	75.6%	76.2%	75.5%	75.6%
GCN _{200M}	72.4%	69.8%	58.0%	70.7%	72.6%	71.5%	71.2%
GAT _{200M}	79.3%	73.2%	62.7%	77.2%	78.4%	77.1%	76.6%
GraphSAGE _{200M}	76.5%	73.0%	55.0%	75.5%	75.9%	75.6%	75.6%

^aModel column: GCN—Graph Convolution Network⁴⁷; GAT—Graph Attention Network⁴⁸; GraphSAGE—Graph Sample and Aggregation⁴⁹; 100M/200M—Euclidean buffer radius.

^bFeature columns: MA—Mean Accuracy; M/T—Metric and topological features; Bldn—Building; Popn—Population; POIs—Points of Interest; SVI—Street View Imagery.

Bold values refers to the top 4 best-performing models that were selected for visualisation in Fig. 4.

analysis, aligning with the broader, emancipatory goals of urban planning.

Towards developing a science of cities, we posit the importance of urban benchmarking datasets. They provide a standardised and transparent basis for researchers to measure progress between different analytical methods, providing direction to advance the current state-of-the-art. Consistent feature and data representation is especially important in the study of complex urban systems since system components interact with one another in dynamic and non-linear ways. Through modular and extensible design, Urbanity eases the data construction process and helps to reduce data consistency limitations.

Future development opportunities largely involve enhancing modelling capabilities to capture richer contextual representation

of networks. At present, information like sidewalk condition and barrier-free accessibility at the pedestrian scale are frequently missing from global city networks, presenting challenges to pedestrian planning initiatives⁶⁰. Incorporating such local network information is critical in aiding marginalized local communities and enabling robust evaluation of local spatial entities. In light of these gaps, data sharing and interoperability modules that permit various stakeholders to dynamically modify or enrich networks could prove significantly beneficial. Implementing these advancements not only promises to boost collaborative planning endeavours but also aims to enhance the spatial-temporal data frequency. Another promising direction of exploration lies in assessing the impact of actor-based interactions on urban networks. Recent efforts employing reinforcement learning

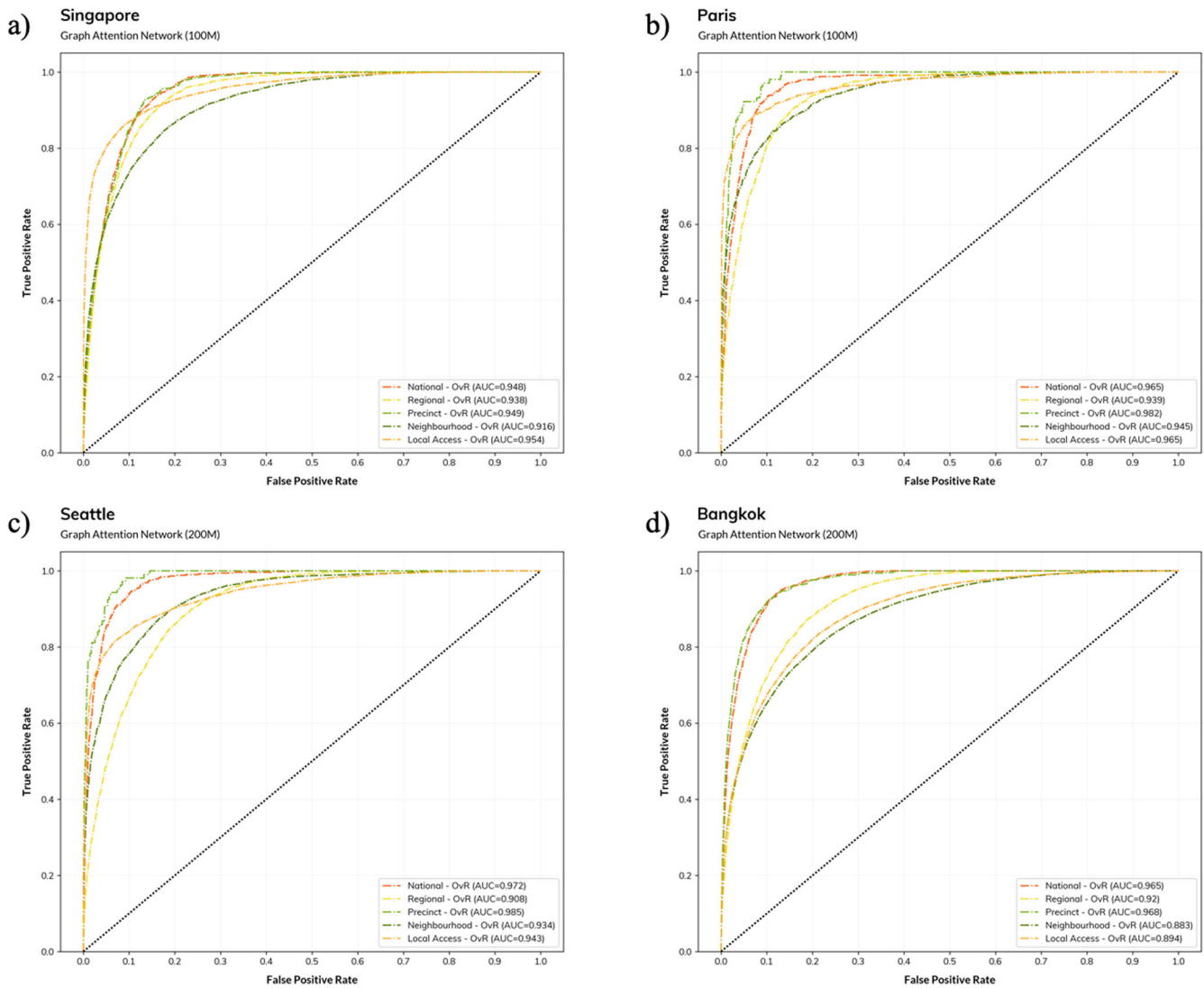


Fig. 4 GNN multi-class predictive performance on road categories. Each curve is a One versus Rest (OvR) receiver operating curve that reports classification performance for each road category. Area under curve (AUC) provides a measure of model performance by aggregating across all possible decision thresholds. For our predictive task, graph attention networks, while taking longer to train, demonstrate better classification accuracy over classical graph convolutional and graph sampling and aggregation (SAGE) architectures. **a** Graph attention model is able to accurately predict road categories from adjacent node attributes in Singapore. **b** Similarly, model performs well across all road classes for Paris. **c** Model displays strongest performance for national and precinct road categories for Seattle. **d** For Bangkok, the model has some difficulty predicting neighbourhood and local access roads.

algorithms demonstrate potential in dynamically modelling urban processes on networks^{61,62}. Ultimately, these advancements could set the stage for scrutinizing dynamic processes within urban systems, discerning the emergent behaviour of urban networks, and deepening our comprehension of urban complexity.

In this work, we developed an open-source urban network analysis tool to automate construction of feature rich urban networks, and demonstrated the value of contextual network features for multi-scalar descriptive and predictive urban analytical tasks. Our findings show that contextual network features form the foundation for comparative urban studies and are vital in the creation of scalable and expressive urban machine-learning models. Towards advancing our understanding of complex urban systems and supporting evidence-based planning, comprehensive modelling of both the physical and social facets of networks will be of paramount importance. The sustained advancement and application of accessible open-source tools will serve as a key contributor to this endeavour.

METHODS

Urbanity was developed with the Python programming language. Python was chosen because it provided an open-source, general-purpose, and high-level programming interface for package development. Urbanity's main modules are built upon existing packages (see Supplementary Table 1).

Data acquisition

Urbanity provides a high-level interface to read, extract, and pre-process global urban data. We evaluate datasets for inclusion based on several criteria: (1) global coverage (to facilitate comparative studies); (2) spatial granularity (finer spatial resolution is preferred); (3) open access (non-proprietary access which allow liberal usage for analytical purposes). These conditions meant that some popular, proprietary datasets (e.g., Google Street View and WorldPop) were excluded from our study.

Urbanity utilizes data elements with consistent global jurisdictional coverage to carry out comprehensive comparisons of cities

around the world. There has been extensive research into the reliability and soundness of urban open data. As laid out in their [guidelines](#), the OSM community typically authenticates OSM data. A significant amount of work has gone into examining the existence and standard of OSM data related to aspects like road networks^{63–67}, points of interest⁶⁸, and building footprints^{69–72}. Likewise, several studies have assessed the quality and coverage of street view imagery on crowdsourced platforms like Mapillary and KartaView^{73–75}. Furthermore, the geographical precision of high-density population maps has been rigorously validated against population census data in a technique-oriented paper⁷⁶.

Population data. We collect population data from Meta's high-resolution population density maps, which provide spatially detailed population data in 30-m spatial resolution for 200 countries⁷⁶. Based on implementation checks conducted on October 7, 2022, we found population data for all countries except Ukraine to be available. Urbanity's application programming interface (API) fetches population data in a dynamic manner, bypassing data storage. Since access to population data is provided at a national scale, this poses challenges for micro scale (e.g. precinct or neighbourhood level) queries for large countries (e.g., the USA). To allow API queries at a finer geographical resolution, we slice the original national level dataset into equal-sized vector tiles. Lastly, where data is available for multiple time periods, we report the most recent population figures.

Street network, points of interest, and building footprints. We extract street network, points of interest (POIs), and building footprints from OpenStreetMap (OSM). OSM is an open collaborative mapping platform that hosts the most comprehensive global crowdsourced collection of geometric features including building footprints, urban amenities, and street networks. OSM data access is provided through Pyrosm API, which provides access to raw, daily updated OSM data from GeoFabrik. This approach prevents bottlenecks resulting from continuous querying of OSM's Overpass API. We pre-process raw urban networks to simplify network edges and include connected boundary nodes⁷⁷.

POIs correspond to OSM primary key tags (amenity, shop, tourist, leisure). For each primary key tag, not all tags correspond to urban amenities. To address this issue, we manually inspect and choose relevant tags under each primary key tag. For example, we extract 'museum', 'gallery', 'artwork', and 'attraction' from the tourist primary key tag. In addition, we found observations to be tagged under multiple POI tags. For example, an observation might be tagged as both amenity and shop. To prevent double counting, we apply procedural selection across each observation. More specifically, we first check if the amenity field is empty, and if it is, we check for values in the order of tourist, leisure, and shop. Finally, we relabel the list of amenities according to eight main categories—civic, recreational, entertainment, food, healthcare, institutional, social, and commercial.

For buildings footprints, we implement a procedural script to ensure that all buildings correspond to valid polygons. We first check the geometric type of each building row and convert line objects into polygons (see Supplementary Figure 2). For objects with multiple lines (e.g., compounds with inner courtyards), we build polygons in a two-step process: (1) identify the exterior building perimeter by geographic extent; (2) build polygon with building perimeter as bounds and interior lines as open space within each building.

Street view imagery. Mapillary is a free and open crowdsourcing platform that provides high-resolution SVI for cities and urban regions. Till date, Mapillary's coverage has penetrated most cities around the world⁷⁵. Compared to other popular data sources, such as Google Street View, Mapillary images are hosted under a CC-BY-SA 4.0 licence, which allows users to freely share, use, and adapt

images. The latest access point is provided by Mapillary API Version 4.0, which allows location-based query of image vector tiles. As dynamic computing of images would be resource intensive, we adopt a pre-compute and store approach to integrate SVI information into our package. Accordingly, we first query the vector tiles according to each city's geographic extent. Subsequently, we extract meta-information for all street view imagery located within each vector tile. To ensure consistency in image segmentation results, we filter out non-frontal facing images. The process was carried out in two steps: (1) computing the bearing of the closest network edges to each image; (2) removing images where the bearing angle differs from the compass angle by more than 20°. As daylight visibility is important for image segmentation, we further filter the set of images to those captured between 9 am till 5 pm (local time) to ensure optimal lighting. Last but not least, where there are many images within a tile, we reduce computational workload by randomly sampling 2000 images. An overview of the image screening and selection process is enumerated for each city (see Supplementary Table 2).

Data integration

A variety of spatial methods have been used to delimit catchment areas and measure access coverage for urban locations. Popular methods include uniform Euclidean^{78,79}, network-based distance⁸⁰, network voronoi⁸¹, and spatial modelling approaches^{13,82}. Euclidean methods are the most straightforward and delineate catchment areas through the creation of a radial buffer around each network node. Similarly, network-based distance methods create catchment zones along networks that correspond to all points that are accessible within specified distance from the starting node. Network voronoi methods extend on distance-based approaches by splitting the network into continuous subgraphs where all points in each subgraph belong to its closest node. For example, network voronoi are commonly used to evaluate the network coverage of medical emergency facilities and identify infrastructure shortage along networks⁸³. However, we found that not all spatial methods lend themselves readily to the integration of urban data due to (1) data interoperability; and (2) standardisation issues. For our application, we adopted the uniform Euclidean approach as it provided a flexible basis to harmonise heterogeneous geospatial data types, including building polygons, vector point data, and population raster maps. In particular, Euclidean buffers provide an intuitive and straightforward mechanism to compute building footprint area around network nodes. On the other hand, network-based methods provide a less intuitive interpretation of building footprint proportion since most buildings do not occur directly on networks. Accordingly, we construct Euclidean buffers around each network node and, for each indicator, compute the spatial intersection with the corresponding spatial element (see Supplementary Table 3). For example, building footprint proportion corresponds to the ratio between the building area and the buffered area for each node⁸⁴. While we specify attributes for 100-metre and 200-metre intervals, Urbanity allows users to freely specify buffer distance depending on their application and use cases.

SVI model architecture. In recent years, Transformer models have emerged as one of the most exciting innovations for deep learning⁸⁵. Originally conceived in the field of Natural Language Processing (NLP), transformer architectures have since demonstrated high scalability and applicability, setting benchmarks in other domains⁸⁶. Building on these developments, we adopt the 'Mask2Former' approach by⁸⁷, which is a universal, end-to-end architecture applicable to a wide range of image segmentation tasks. Mask2Former is trained and validated on the Mapillary

Version 1.2 validation dataset, which comprises 65 semantic classes⁸⁸, reporting state-of-the-art performance (MIoU=63.2%). Compared to other transformer architectures, Mask2Former features several notable innovations for computer vision: 1) masked attention mechanism which allows the model to concentrate and utilise local information in images; and 2) multiscale feature decoder which allows the model to extract features of different sizes. In this regard, Mask2Former offers two main advantages for our purposes: (1) improved accuracy to pick out different semantic categories in images (previous models tend to ignore small objects); and (2) lightweight and scalable computation. The latter point is especially important given the computational intensity of segmenting images for entire cities. Readers interested in the specifics of Mask2Former architecture and training are referred to^{87,89}.

On hardware, we segment SVI images with a NVIDIA Geforce RTX 3090 GPU. For each city, computation time took approximately three days. As a benefit of pre-computation, users can access and compute SVI indicators within seconds. A descriptive list of selected cities with pre-computed SVI indicators is provided in Fig. 1. Upcoming updates will incorporate more pertinent streetscape indicators and broaden the present city list. These changes will be integrated into subsequent versions of Urbanity and detailed in our package documentation as well as our city discussions [page](#).

Package design

Under the hood, Urbanity is constructed in an object-oriented fashion and designed to suit planning workflows. For instance, Urbanity parallels the inchoation of the urban planning process by allowing planners to interact with their target planning boundary in a dynamic and visual manner. At this stage, users have the flexibility to explore, customise, and visualise their area of interest. Once the location is set, users can specify geographic bounds for network or aggregate statistical extraction by either (1) manually drawing on a digital map with the provided draw tools; or (2) uploading their own geometric shape files.

Network extraction. Urbanity's network extraction API offers a flexible and extensible interface for users to request various combinations of contextual information, not limited to metric/topological, demographic, building morphology, points of interest, and street view imagery. Internally, we employ pre-processing, pre-computation, dynamic memory handling, and web scraping to provide a simple and lightweight data extraction experience. The computation time depends on both the number of indicators requested and the size of the target area. For the computation of areal and metric indicators, geographic projection is automatically computed and applied based on the chosen geographic boundary location. As output, Urbanity provides users with a Networkx graph object, and relational data tables corresponding to attributes for network nodes and edges respectively.

Aggregate statistics. It is not uncommon for planners to require aggregate statistics (e.g. total population, building footprint proportion) for target sites. Currently, the process to extract such information is manual, time-consuming, and not straightforward (especially for boundaries that do not align with the census tract). To address this concern, Urbanity provides a function that computes aggregated statistics for any arbitrary geographic bounding box. Users can specify a common shape file with single or multiple areas of interest and quickly obtain relevant information, facilitating rapid comparative analyses across different urban scales and geographic contexts.

The outputs from the network extraction and statistical aggregation steps serve as a bridge towards downstream

descriptive and predictive urban analytical tasks. Urbanity offers detailed installation instructions, documentation, and example use cases for users to get started.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

DATA AVAILABILITY

All spatial data is publicly available. Urban street network, building footprint, and points of interest data are obtained from OpenStreetMap. Crowdsourced SVI data are obtained from Mapillary. Population data are available from Meta's high-resolution population density maps: <https://dataforgood.facebook.com/dfg/tools/high-resolution-population-density-maps>. Data to reproduce analytical results: https://github.com/winstonyym/urbanity_examples. Re-tiled population data for fast geospatial querying: https://figshare.com/articles/dataset/USA_TILE_POPULATION/21502296.

CODE AVAILABILITY

Source code for Urbanity is hosted under an open source MIT license at: <https://github.com/winstonyym/urbanity>.

Received: 27 January 2023; Accepted: 15 July 2023;

Published online: 25 July 2023

REFERENCES

- Batty, M. *Cities and Complexity: Understanding Cities with Cellular Automata, Agent-based Models, and Fractals* (The MIT Press, 2007).
- Batty, M. *Cities as Complex Systems: Scaling, Interaction, Networks, Dynamics and Urban Morphologies* (Springer, 2009).
- Bettencourt, L. M. The origins of scaling in cities. *Science* **340**, 1438–1441 (2013).
- Lai, S.-K. *Planning within Complex Urban Systems* (Routledge, 2020).
- Portugali, J. (ed.) In *Handbook on Cities and Complexity* 13–27 (Edward Elgar Publishing, 2021).
- Xue, J. et al. Quantifying the spatial homogeneity of urban road networks via graph neural networks. *Nat. Mach. Intell.* **4**, 246–257 (2022).
- Chirigati, F. Gauging urban development with neural networks. *Nat. Comput. Sci.* **2**, 216–216 (2022).
- Steinacker, C., Storch, D.-M., Timme, M. & Schröder, M. Demand-driven design of bicycle infrastructure networks for improved urban bikeability. *Nat. Comput. Sci.* **2**, 655–664 (2022).
- Yap, W., Chang, J.-H. & Biljecki, F. Incorporating networks in semantic understanding of streetscapes: contextualising active mobility decisions. *Environ. Planning B: Urban Analytics City Sci.* **50**, 1416–1437 (2023).
- Carmody, D. et al. The effect of co-location on human communication networks. *Nat. Comput. Sci.* **2**, 494–503 (2022).
- Sevtasuk, A. & Kalvo, R. Patronage of urban commercial clusters: a network-based extension of the huff model for balancing location and size. *Environ. Planning B: Urban Analytics City Sci.* **45**, 508–528 (2018).
- Zhao, T. et al. Coupling graph deep learning and spatial-temporal influence of built environment for short-term bus travel demand prediction. *Comput. Environ. Urban Syst.* **94**, 101776 (2022).
- Alves, L. G., Rybski, D. & Ribeiro, H. V. Commuting network effect on urban wealth scaling. *Sci. Rep.* **11**, 22918 (2021).
- Raducha, T. & San Miguel, M. Emergence of complex structures from nonlinear interactions and noise in coevolving networks. *Sci. Rep.* **10**, 15660 (2020).
- Kirkley, A. Spatial regionalization based on optimal information compression. *Commun. Phys.* **5**, 1–10 (2022).
- Boeing, G., Batty, M., Jiang, S. & Schweitzer, L. In *Handbook of Spatial Analysis in the Social Sciences* (eds Rey, S. J. & Franklin, R. S.) 503–516 (Edward Elgar Publishing, 2022).
- Gharaee, Z., Kowshik, S., Stromann, O. & Felsberg, M. Graph representation learning for road type classification. *Pattern Recognition* **120**, 108174 (2021).
- Janowicz, K., Gao, S., McKenzie, G., Hu, Y. & Bhaduri, B. Geoai: spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond. *Int. J. Geogr. Inform. Sci.* **34**, 625–636 (2020).
- Mai, G. et al. A review of location encoding for geoai: methods and applications. *Int. J. Geogr. Inform. Sci.* **36**, 639–673 (2022).

20. Liu, P. & Biljecki, F. A review of spatially-explicit geospatial applications in urban geography. *Int. J. Appl. Earth Observation. Geoinform.* **112**, 102936 (2022).
21. Pettit, C. et al. Planning support systems for smart cities. *City Culture Soc.* **12**, 13–24 (2018).
22. Chadzynski, A. et al. Semantic 3d city database—an enabler for a dynamic geospatial knowledge graph. *Energy AI* **6**, 100106 (2021).
23. Harris, R. et al. More bark than bytes? reflections on 21+ years of geocomputation. *Environ. Planning B: Urban Analytics. City Sci.* **44**, 598–617 (2017).
24. Anselin, L. & Rey, S. J. Open source software for spatial data science. *Geogr. Anal.* **54**, 429–438 (2022).
25. Boeing, G. et al. Using open data and open-source software to develop spatial indicators of urban design and transport features for achieving healthy and sustainable cities. *Lancet Global Health* **10**, e907–e918 (2022).
26. Yap, W., Janssen, P. & Biljecki, F. Free and open source urbanism: software for urban planning practice. *Comput. Environ. Urban Syst.* **96**, 101825 (2022).
27. Gahegan, M. Fourth paradigm giscience? prospects for automated discovery and explanation from data. *Int. J. Geogr. Inform. Sci.* **34**, 1–21 (2020).
28. Singleton, A. & Arribas-Bel, D. Geographic data science. *Geogr. Anal.* **53**, 61–75 (2021).
29. Barthélemy, M. Spatial networks. *Phys. Rep.* **499**, 1–101 (2011).
30. Batty, M. Integrating space syntax with spatial interaction. *Urban Informatics* **1**, 4 (2022).
31. Leskovec, J., Lang, K. J., Dasgupta, A. & Mahoney, M. W. Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Math.* **6**, 29–123 (2009).
32. Comunian, R. Rethinking the creative city: the role of complexity, networks and interactions in the urban creative economy. *Urban Stud.* **48**, 1157–1179 (2011).
33. Batty, M. Mumford’s recurring challenge: What is a city? *Environ. Planning B: Urban Analytics City Sci.* **49**, 387–390 (2022).
34. Kandt, J. & Batty, M. Smart cities, big data and urban policy: towards urban analytics for the long run. *Cities* **109**, 102992 (2021).
35. Engin, Z. et al. Data-driven urban management: mapping the landscape. *J. Urban Manag.* **9**, 140–150 (2020).
36. Khamchiangta, D. & Dhakal, S. Future urban expansion and local climate zone changes in relation to land surface temperature: Case of bangkok metropolitan administration, thailand. *Urban Clim.* **37**, 100835 (2021).
37. Tobler, W. R. A computer movie simulating urban growth in the detroit region. *Econom. Geogr.* **46**, 234–240 (1970).
38. Bokányi, E., Juhász, S., Karsai, M. & Lengyel, B. Universal patterns of long-distance commuting and social assortativity in cities. *Sci. Rep.* **11**, 20829 (2021).
39. Tóth, G. et al. Inequality is rising where social network segregation interacts with urban topology. *Nat. Commun.* **12**, 1–9 (2021).
40. Newman, M. E. Mixing patterns in networks. *Phys. Rev. E* **67**, 026126 (2003).
41. Yuen, B. Creating the garden city: the singapore experience. *Urban Stud.* **33**, 955–970 (1996).
42. Tan, P. Y., Wang, J. & Sia, A. Perspectives on five decades of the urban greening of singapore. *Cities* **32**, 24–32 (2013).
43. Hou, Y., Yap, W., Chua, R., Song, S. & Yuen, B. The associations between older adults’ daily travel pattern and objective and perceived built environment: a study of three neighbourhoods in singapore. *Transport Policy* **99**, 314–328 (2020).
44. Barthélemy, M., Bordin, P., Berestycki, H. & Gribaudo, M. Self-organization versus top-down planning in the evolution of a city. *Sci. Rep.* **3**, 2153 (2013).
45. Sotomayor, L. & Danieri, A. The dilemmas of equity planning in the global south: a comparative view from bangkok and medellín. *J. Planning Educ. Res.* **38**, 273–288 (2018).
46. Fey, M. & Lenssen, J. E. Fast graph representation learning with pytorch geometric. Preprint at <https://arxiv.org/abs/1903.02428> (2019).
47. Kipf, T. N. & Welling, M. Semi-supervised classification with graph convolutional networks. Preprint at <https://arxiv.org/abs/1609.02907> (2016).
48. Hamilton, W., Ying, Z. & Leskovec, J. Inductive representation learning on large graphs. *Adv. Neural Inform. Process. Syst.* **30** (2017).
49. Veličković, P. et al. Graph attention networks (2018). Preprint at <https://arxiv.org/abs/1710.10903> (2018).
50. Shi, W. et al. Prospective for urban informatics **1**, 1–14 (2022).
51. Jacobs, J. *The Death and Life of Great American Cities* (Random House, 1961).
52. Whyte, W. H. *The Social Life of Small Urban Spaces* (Conservation Foundation, 1980).
53. Jacobs, J. In *The Exploding Metropolis* (ed Whyte, W. H.) Vol. 168 (Doubleday, 1958).
54. Marshall, S., Gil, J., Kropf, K., Tomko, M. & Figueiredo, L. Street network studies: from networks to models and their representations. *Netw. Spatial Econom.* **18**, 735–749 (2018).
55. Perrone, C. ‘downtown is for people’: The street-level approach in jane jacobs’ legacy and its resonance in the planning debate within the complexity theory of cities. *Cities* **91**, 10–16 (2019).
56. Weijnen, M. P. & Bouwmans, I. Innovation in networked infrastructures: coping with complexity. *Int. J. Critical Infrastruct.* **2**, 121–132 (2006).
57. Batty, M. The conundrum of ‘form follows function’. *Environ. Planning B: Urban Analytics. City Sci.* **49**, 1815–1819 (2022).
58. Grekousis, G. Artificial neural networks and deep learning in urban geography: a systematic review and meta-analysis. *Comput. Environ. Urban Syst.* **74**, 244–256 (2019).
59. Boeing, G. & Arribas-Bel, D. Gis and computational notebooks. Preprint at <https://arxiv.org/abs/2101.00351> (2021).
60. Hosseini, M., Sevtsuk, A., Miranda, F., Cesar Jr, R. M. & Silva, C. T. Mapping the walk: a scalable computer vision approach for generating sidewalk network datasets from aerial imagery. *Comput. Environ. Urban Syst.* **101**, 101950 (2023).
61. An, L. et al. Modeling agent decision and behavior in the light of data science and artificial intelligence. *Environ. Modelling Softw.* **166**, 105713 (2023).
62. Turgut, Y. & Bozdog, C. E. A framework proposal for machine learning-driven agent-based models through a case study analysis. *Simul. Modelling Practice Theory* **123**, 102707 (2023).
63. Haklay, M. How good is volunteered geographical information? a comparative study of openstreetmap and ordnance survey datasets. *Environ. Planning B: Planning Des.* **37**, 682–703 (2010).
64. Neis, P. & Zielstra, D. Recent developments and future trends in volunteered geographic information research: the case of openstreetmap. *Future Internet* **6**, 76–106 (2014).
65. Karduni, A., Kermanshah, A. & Derrible, S. A protocol to convert spatial polyline data to network formats and applications to world urban road networks. *Sci. Data* **3**, 1–7 (2016).
66. Barrington-Leigh, C. & Millard-Ball, A. The world’s user-generated road map is more than 80% complete. *PLoS ONE* **12**, e0180698 (2017).
67. Higgs, C. et al. Policy relevant health related liveability indicator datasets for addresses in Australia’s 21 largest cities. *Sci. Data* **10**, 113 (2023).
68. Zhang, L. & Pfoser, D. Using openstreetmap point-of-interest data to model urban change—a feasibility study. *PLoS ONE* **14**, e0212606 (2019).
69. Fan, H., Zipf, A., Fu, Q. & Neis, P. Quality assessment for building footprints data on openstreetmap. *Int. J. Geogr. Inform. Sci.* **28**, 700–719 (2014).
70. Zhang, Y., Zhou, Q., Brovelli, M. A. & Li, W. Assessing osm building completeness using population data. *Int. J. Geogr. Inform. Sci.* **36**, 1443–1466 (2022).
71. Milojevic-Dupont, N. et al. Eubucco v0. 1: European building stock characteristics in a common and open database for 200+ million individual buildings. *Sci. Data* **10**, 147 (2023).
72. Biljecki, F., Chow, Y. S. & Lee, K. Quality of crowdsourced geospatial building information: A global assessment of OpenStreetMap attributes. *Building. Environ.* **237**, 110295 (2023).
73. Juhász, L. & Hochmair, H. H. User contribution patterns and completeness evaluation of mapillary, a crowdsourced street level photo service. *Trans. GIS* **20**, 925–947 (2016).
74. Quinn, S. & Alvarez León, L. Every single street? rethinking full coverage across street-level imagery platforms. *Trans. GIS* **23**, 1251–1272 (2019).
75. Hou, Y. & Biljecki, F. A comprehensive framework for evaluating the quality of street view imagery. *Int. J. Appl. Earth Observation Geoinform.* **115**, 103094 (2022).
76. Tiecke, T. G. et al. Mapping the world population one building at a time. Preprint at <https://arxiv.org/abs/1712.05839> (2017).
77. Boeing, G. Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Comput. Environ. Urban Syst.* **65**, 126–139 (2017).
78. Marselle, M. R. et al. Urban street tree biodiversity and antidepressant prescriptions. *Sci. Rep.* **10**, 22445 (2020).
79. Nori-Sarma, A. et al. Impacts of the choice of distance measurement method on estimates of access to point-based resources. *J. Exposure Sci. Environ. Epidemiol.* **33**, 237–243 (2022).
80. Sevtsuk, A. & Basu, R. The role of turns in pedestrian route choice: a clarification. *J. Transport Geogr.* **102**, 103392 (2022).
81. Chen, B. Y., Teng, W., Jia, T., Chen, H.-P. & Liu, X. Transit voronoi diagrams in multi-mode public transport networks. *Comput. Environ. Urban Syst.* **96**, 101849 (2022).
82. Spadon, G., de Carvalho, A. C., Rodrigues-Jr, J. F. & Alves, L. G. Reconstructing commuters network using machine learning and urban indicators. *Sci. Rep.* **9**, 11801 (2019).
83. Billhardt, H., Lujak, M., Sánchez-Brunete, V., Fernández, A. & Ossowski, S. Dynamic coordination of ambulances for emergency medical assistance services. *Knowledge-Based Syst.* **70**, 268–280 (2014).
84. Biljecki, F. & Chow, Y. S. Global building morphology indicators. *Comput. Environ. Urban Syst.* **95**, 101809 (2022).
85. Vaswani, A. et al. Attention is all you need. *Adv. Neural Inform. Process. Syst.* **30** (2017).
86. Khan, S. et al. Transformers in vision: a survey. *ACM Comput. Surv. (CSUR)* **54**, 1–41 (2022).

87. Cheng, B., Misra, I., Schwing, A. G., Kirillov, A. & Girdhar, R. Masked-attention mask transformer for universal image segmentation. In: *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition* 1290–1299 (2022).
88. Neuhold, G., Ollmann, T., Rota Bulò, S. & Kotschieder, P. The mapillary vistas dataset for semantic understanding of street scenes. In: *Proc. IEEE International Conference on Computer Vision* 4990–4999 (2017).
89. Cheng, B., Schwing, A. & Kirillov, A. Per-pixel classification is not all you need for semantic segmentation. *Adv. Neural Inform. Process. Syst.* **34**, 17864–17875 (2021).
90. Foti, F., Waddell, P. & Luxen, D. A generalized computational framework for accessibility: from the pedestrian to the metropolitan scale. In: *Proc. 4th TRB Conference on Innovations in Travel Modeling* (eds Srinivasan, S. & Sabina, E.) *Transportation Research Board* 1–14 (2012).
91. Gaboardi, J., Rey, S. & Lumnitz, S. spaghetti: spatial network analysis in pysal. *J. Open Source Softw.* **6** (2021).
92. Okabe, A., Okunuki, K.-i & Shiode, S. Sagnet: a toolbox for spatial analysis on a network. *Geogr. Anal.* **38**, 57–66 (2006).
93. Raffler, C. Qneat3—qgis network analysis toolbox 3 (2018). <https://root676.github.io/>.
94. Sevtsuk, A. & Mekonnen, M. Urban network analysis toolbox. *international. J. Geomatics. Spatial Anal.* **22**, 287–305 (2012).
95. Simons, G. The cityseer python package for pedestrian-scale network-based urban analysis. *Environ. t and Planning B: Urban Analytics. City Sci.* **50**, 23998083221133827 (2022).
96. Stahle, A., Marcus, L. & Karlstrom, A. In *New Developments in Space Syntax Software* (ed. Turner, A.) 35–42 (ITU Faculty of Architecture, 2007).
97. Tripathy, P., Rao, P., Balakrishnan, K. & Malladi, T. An open-source tool to extract natural continuity and hierarchy of urban street networks. *Environ. Planning B: Urban Analytics City Sci.* **48**, 2188–2205 (2021).
98. Turner, A. Depthmap: a program to perform visibility graph analysis. In: *Proc. 3rd International Symposium on Space Syntax* Vol. 31, 31–12 (Citeseer, 2001).

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the contributions of the open-source community. We thank the members of the NUS Urban Analytics Lab for the discussions. The first author thankfully acknowledges the NUS Graduate Research Scholarship granted by the National University of Singapore. This research is part of the project Large-scale 3D Geospatial Data for Urban Analytics, which is supported by the National University of Singapore under the Start Up Grant R-295-000-171-133.

AUTHOR CONTRIBUTIONS

W.Y.: Conceptualisation and design, software development, data acquisition and analysis, data interpretation, and manuscript drafting. R.S.: Conceptualisation and design, data interpretation, manuscript drafting, and research supervision. F.B.: Conceptualisation and design, data interpretation, manuscript drafting, research supervision, and project funding.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42949-023-00125-w>.

Correspondence and requests for materials should be addressed to Filip Biljecki.

Reprints and permission information is available at <http://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 4.0 International License, which permits use, sharing, adaptation, 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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023