GRAPHING THE CITY

EXTRACTING THE DNA OF THE CITY AS AN IDENTITY GRAPH

LAURENCE ELSDON
NORTHUMBRIA UNIVERSITY
JANUARY 2017
9,885 WORDS
4,879 LINES OF CODE

CONTENTS

00 ABSTRACT

Using structured data from the Wikidata and Google Knowledge Graph Search APIs, this paper introduces a computational method to diagrammatically reduce the places of interest within a city to an *identity graph*. This experimental research extends Hillier & Hanson's (1984) concept of extracting the genotype of a dwelling from its topological graph, taking it from the microscopic scale of the house to the macroscopic city scale. The paper begins with an exploration of the existing literature on many relevant topics including urban identity, space syntax, social network analysis and volunteered geographic information. The literature then leads the development of a data collection and analysis application which is used to analyse five cities using graph theoretic social network analysis formula to quantitatively and objectively explore the commonalities and unique elements of their architectural identities.

00 GLOSSARY

Algorithm	"A process or set of rules to be followed in calculations or other problem-solving operations, especially by a computer" (Oxford Dictionaries, 2016)
API	Application Programming Interface – "A set of functions and procedures that allow the creation of applications which access the features or data of an operating system, application, or other service" (Oxford Dictionaries, 2016).
Application	"A program or piece of software designed to fulfil a particular purpose" (Oxford Dictionaries, 2016)
Array	A "set of related elements" (Oxford Dictionaries, 2016).
Endpoint	An entity – typically a URL – to which specific API requests are made.
Get request	An HTTP request that returns data, as opposed to a Post which sends data.
HTTP	Hypertext Transfer Protocol – The standard request protocol used for sending and receiving data and documents over the internet.
Imageability	"The ease with which [something] gives rise to a sensory mental image" (Paivio, Yuille & Madigan, 1968)
Intelligibility	"The state or quality of being intelligible" (Oxford Dictionaries, 2016) or comprehensible.
JSON	JavaScript Object Notation – a lightweight data-interchange format that is easy for humans to read and write and for machines to parse and generate. (JSON, n.d.)
Key: Value pair	*See Object*
KML	Keyhole Markup Language – "a file format used to display geographic data in an Earth browser such as Google Earth. You can create KML files to pinpoint locations, add image overlays, and expose rich data in new ways" (Keyhole Markup Language, n.d.).
Metadata	"A set of data that describes and gives information about other data" (Oxford Dictionaries, 2016).
MVC	Model-View-Controller – a software design pattern to separate application code, user interface elements and data storage methods.
Object	A set of related elements where each value is paired with a key.
Parameter	Data sent as part of a request typically to make a request query more specific.
Post request	An HTTP request that sends data, as opposed to one which only returns data.
Processing	"A series of [computational] actions or steps taken in order to achieve a particular end" (Oxford Dictionaries, 2016).
Recursive	*See recursive*
Request	The data sent to a server during an HTTP request
Response	The data returned following an HTTP request
Script	"An automated series of instructions carried out in a specific order" (Oxford Dictionaries, 2016).
Structured data	"Data with a high level of organization, such as information in a relational database" (Google Search, n.d.-c)
Topology	"The study of geometric properties and spatial relations unaffected by the continuous change of shape or size of figures" (Oxford Dictionaries, 2016).
Unstructured data	Data without a high level of organisation, such as an encyclopaedia article
URL	Uniform Resource Locator – "The address of a World Wide Web page" (Oxford Dictionaries, 2016).
User Interface	"The means by which the user and a computer system interact, in particular the use of input devices and software" (Oxford Dictionaries, 2016).

01 INTRODUCTION

The concept of a genotype within the literature of space syntax was first posited by Hillier & Hanson (1984) in *The Social Logic of Space* and then further developed in *Ideas are in Things (*Hillier, Hanson & Graham, 1987). The authors identified that the domestic house could be diagrammatically reduced to a discrete mathematical graph where each room is a node and the openings between rooms are syntactically represented by edges connecting between nodes (Figure 1).

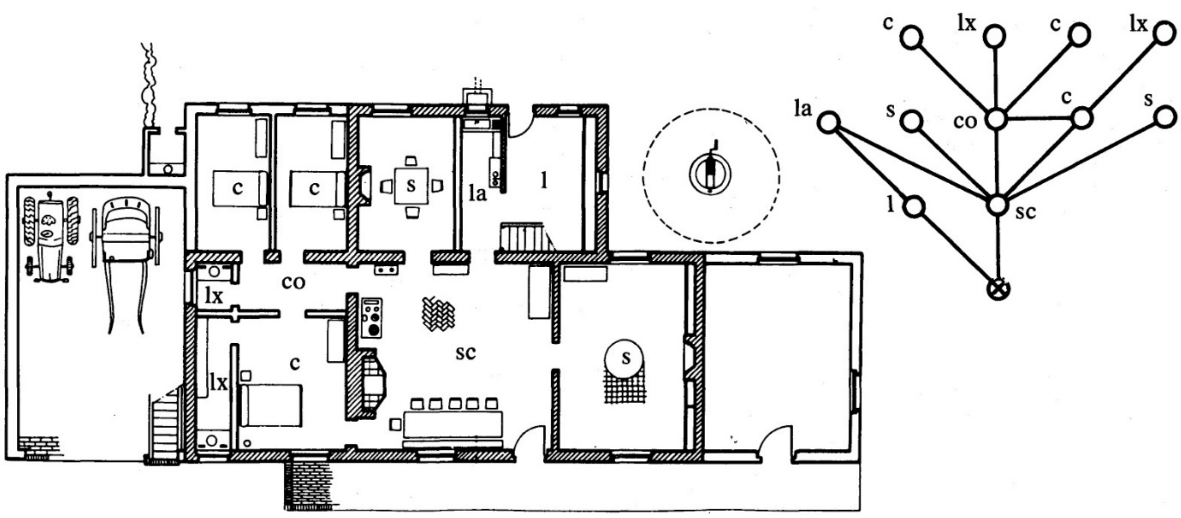


Figure 1 – House plan and genotype graph of Le Cormier (Hillier et al., 1987)

Space syntax has developed significantly over the decades since with concepts such as the axial line (Hillier & Hanson, 1984), isovist (popularised by Michael Benedikt at University of Texas) or the urban web (Salingaros, 1998) being applied to the city scale, however, these theories have focused only on the topological connections within the city.

Within this paper, a computational method of urban analysis will be introduced with a view to defining the factors that contribute towards the identity of a city, and an application will be developed to objectively define and represent this identity. The method will be applied to cities in order to identify the commonalities and unique elements which define their identities, and finally, it will be discussed whether the resulting *identity graphs* can be considered as a visual representation of the *DNA of the city*.

02 BACKGROUND

Image of the City

The Image of the City (Lynch, 1960) presented a new method for evaluating the form of cities. Through the vocabulary of paths, edges, districts, nodes and landmarks, Lynch investigated how citizens formed mental maps of their city (Figure 2). His primary focus was on the *legibility* of the city – "the ease with which its parts can be recognized and can be organized into a coherent pattern" (Lynch, 1960). Lynch claims that if an environment is to be considered legible then it must be imageable and imageability is "that quality in a physical object which gives it a high probability of evoking a strong image in any given observer" (Lynch, 1960).

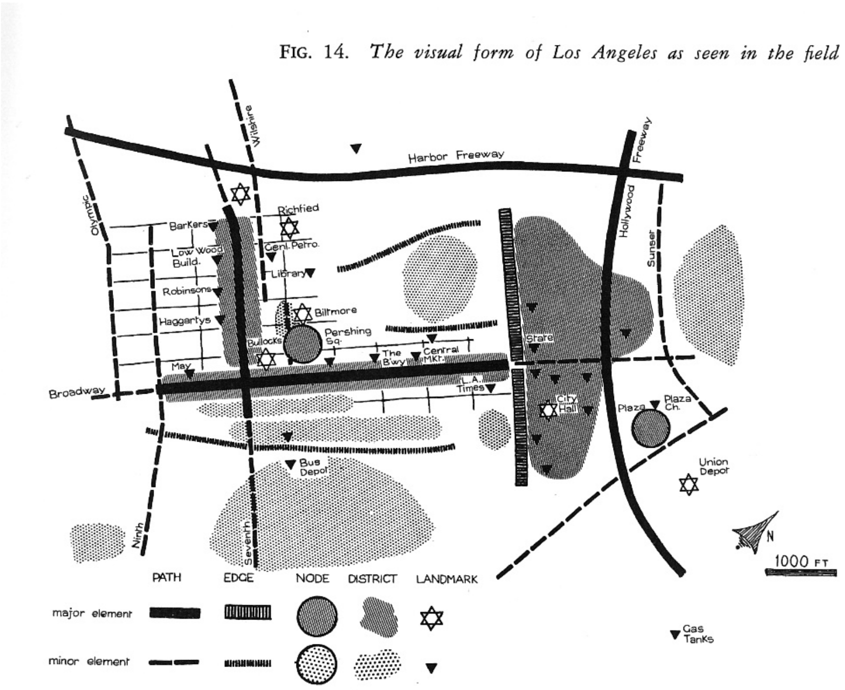


Figure 2 – Lynch's image of Los Angeles (Lynch, 1960, p. 33)

Space Syntax

The intention of space syntax – an avenue of research into the relationship between human societies and spatial layout (Bafna, 2003) – is to describe the configuration of inhabited spaces including buildings and settlements. A number of syntactical descriptions have been used such as the convex map, axial map, isovists or visibility polygons (Figure 3).

graph model	nodes	edges	pictogram
occupancy grids	xy-intervalls with obstacle probability	predefined, non-directional	
place graph	local position information at places	local navigation rules	
view graph	local position information at place transitions	local navigation rules, directional	
access graph	spaces	connectivity	
axial map	lines of sight	intersections	
isovist field	viewshed polygon	mutual visibility	
visibility graph	xy-intervalls	mutual visibility	

○ = node ◁ = directional edge ◁▷ = non-directional edge

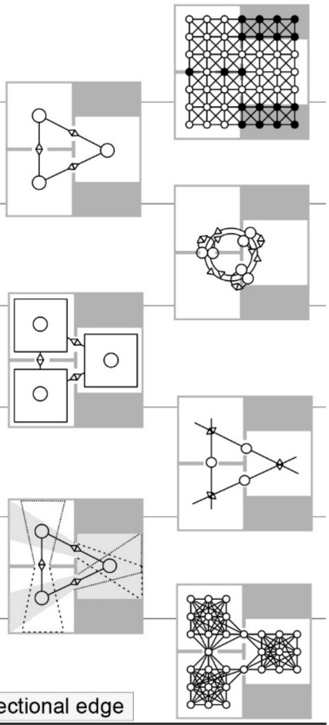


Figure 3 – Overview of the different graph models (Franz et al., 2005)

The Social Logic of Space (Hillier & Hanson, 1984) explores a wide range of topics within the realms of space syntax, including a topological analysis of settlement layouts (Hillier & Hanson, 1984, p. 82). The authors then proceed to apply the same analytic graph concepts from the urban scale to the building interior. The spatial configuration and topological relationships between internal spaces are illustrated through a diagrammatic reduction of the building form to nodes and edges – the language of graphs – where the nodes represent rooms and the edges represent adjacencies between those rooms (Figure 1).

Hillier's later paper *Ideas are in Things* (Hillier, Hanson & Graham, 1987) applies graph analytic techniques to the building graph in order to "demonstrate that cultural ideas are objectively present in artefacts as much as they are subjectively present in minds" (Hillier, Hanson & Graham, 1987). The authors study the rank order of integration value of spaces within samples of vernacular farmhouses in Normandy. "The integration value of a space expresses the relative depth of that space from all others in the graph" (Hillier, Hanson & Graham, 1987). The integration value is a numeric indicator of the function of a space and spaces of differing activities are integrated into the house to differing degrees. Thus if there is an identifiable pattern amongst the numerical differences between spaces it can be said there is a cultural pattern. This pattern is defined by Hiller, Hanson & Graham as the "*inequality genotype*" (Hillier, Hanson & Graham, 1987). The authors' analysis of the 17 samples identifies that while there is no obvious prototypical house "there is evidence of at least one underlying spatial-functional 'genotype'" (Hillier, Hanson & Graham, 1987) which is identifiable by consistencies within how spaces are configured and related to one another, these are seen in differing 'phenotypical' arrangements.

Bafna (2001) proposes that rather than considering each individual space within the house as a node on the graph we should consider clustering spaces of the same function together. This idea is supported by the concept of sectors in Amorim (1997).

Building upon Hillier and Hanson's earlier works Dalton & Kirsan (2008) introduce a method for determining the *genotype signature* from a collection of buildings, their method provides "mathematical clarity and a reproducible method to the hitherto theoretical concept of genotype and phenotype" (Dalton & Kirsan, 2008). Rather than studying the integration value Dalton & Kirsan propose identifying the *median graph* from a sample through the graph theoretic method of graph matching. The authors make the distinction that this median graph is a signature of the genotype – not the genotype itself – "since a genotype is a purely theoretic construct; it is not realizable whereas a genotype signature is an evidence-based example deemed to be most denotative of its theoretic genotype" (Dalton & Kirsan, 2008). Although in its infancy, graph matching has been successfully applied to problems such as molecular comparison, object recognition, and shape analysis. The method works by quantifying the number of operations it takes to transform one graph into another; with the genotype signature being identified as the sample with the lowest distinction value – the lowest average number of acts of transformation per graph element. In their method, the genotype signature is, therefore, a representative graph – a radical shift from the established genotypic indicator: the rank order of integration values.

In a later paper, Bafna (2012) discusses the use of the terms genotype and phenotype within space syntax, suggesting that Dalton & Kirsan's *genotype signature* would be better described as the *prototype*.

> "The difference is that where the concept of genotype refers to a core set of attributes which are shared amongst all members of a class and which help define it … the concept of a prototype (originating in some highly influential studies in human cognition by Eleanor Rosch and her colleagues in the 1970s [Rosch, 1978]) refers to an individual that exemplifies a class without completely defining it" (Bafna, 2012).

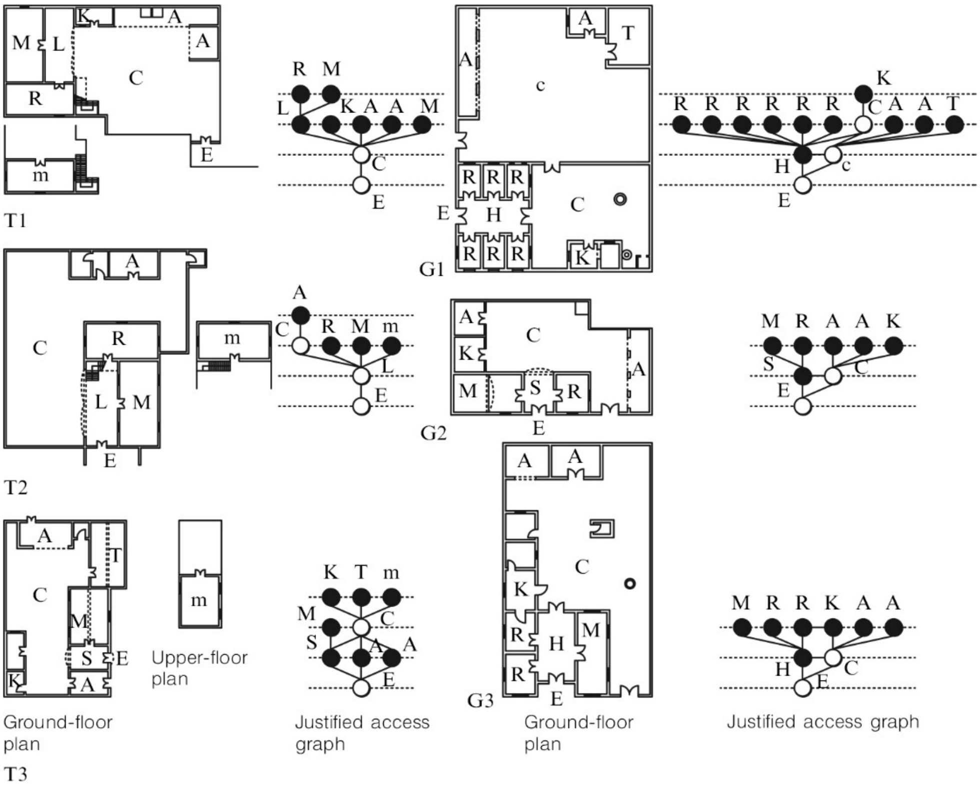


Figure 4 – House plans, where T and G denote Turkish and Greek, respectively. The labeled spaces in the plans and graphs are represented as follows: C courtyard; c animal courtyard; L loggia; R room; M main room; K kitchen; A animal shed; E ˆ exterior; m upper main room; S semiclosed central space/hallways; H closed central space/hallway; T straw store. (Dalton & Kirsan, 2008)

Dalton & Kirsan (2008) apply their graph matching method to a sample of six traditional Cypriot houses, three from Greek areas and three from Turkish (Figure 4). By calculating the number of steps it takes to transform the topological graph of each building into the graph of each other, the distinction value of that building can be determined – that is the number of operations required to transform the graph into any other. Then by comparing the mean normalized distances it can be determined that "on average, it takes 0.25 acts of transformation per graph element to transform any Greek house into any other Greek house. Equally, 0.24 transformations are needed to transform any Turkish house into any other Turkish house. However, almost double the number of transformations is needed to transform a Greek house into a Turkish house (0.44)" (Dalton & Kirsan, 2008). The authors are therefore able to quantitatively conclude that there is a notable cultural difference between the two sample groups, which matched people's subjective intuition in an informal study. In closing, Dalton & Kirsan mention that graph matching can be laborious and time consuming task to perform by hand and is highly prone to error, thus it would be ideal if the method could be computed automatically (Dalton & Kirsan, 2008).

Hillier (1996) introduced the concept of the intelligibility of the city – a measure of its comprehensibility and navigability.

> "The property of 'intelligibility' … means the degree to which what we can see from the spaces that make up the system - that is, how many other spaces are connected to - is a good guide to what we cannot see, that is, the integration of each space into the system as a whole. An intelligible system is one in which well-connected spaces also tend to be well-integrated spaces. An unintelligible system is one where well-connected spaces are not well integrated, so that what we can see of their connections misleads us about the status of that space" (Hillier, 1996).

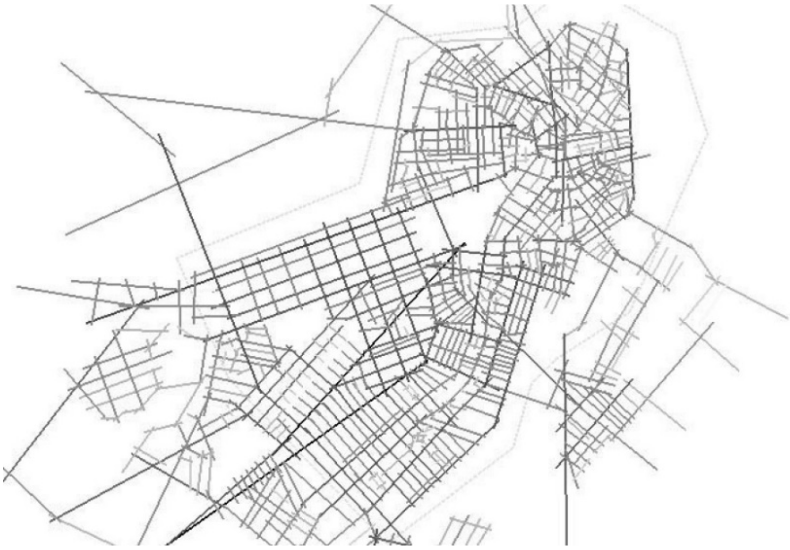


Figure 5 – Axial map of Boston; darker lines indicate higher integration (Dalton & Bafna, 2003)

Dalton & Bafna (2003) who attempt to redefine Lynch's (1960) theory of legibility of the city with space syntactical representations including the isovist and axial line (Figure 5). In their discussion they begin to explore the relationship between Lynch's imageability and Hillier's intelligibility, they conclude: "all imageable cities must be intelligible, but all intelligible cities need not be imageable" (Dalton & Bafna, 2003). This suggests that more cities are capable of being comprehended through space syntax than can be defined by imageability.

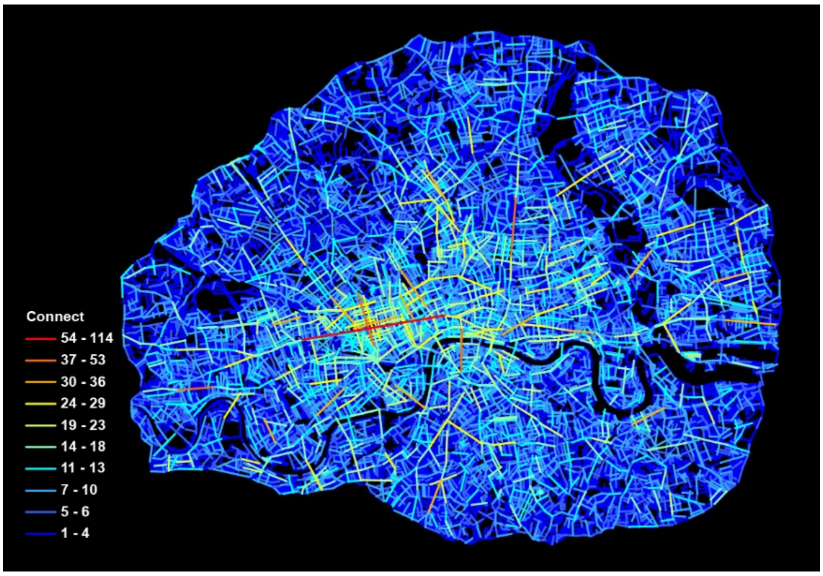


Figure 6 – Ten hierarchical levels for the London open space represented by the axial line; the lines within the red spectrum constitute a part of the London mental map (Jiang, 2012)

Jiang (2012) builds upon Dalton & Bafna's work by creating an algorithmic method for the derivation of the image of the city (Figure 6). He concludes that his algorithm defines a new era for imageability, previously studies were qualitative but now, due to the availability and depth of geospatial databases, the image of the city can be evaluated quantitatively.

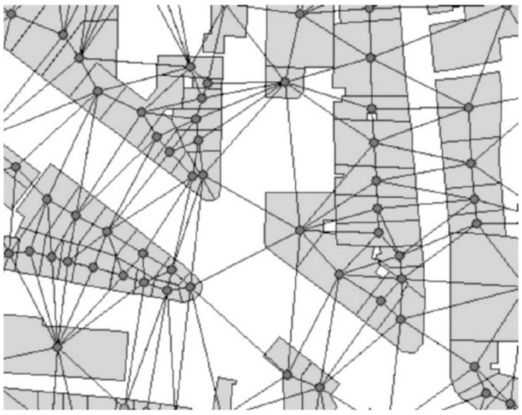


Figure 7 – A portion of a graph-CA model where cells are individual buildings and graph structure represents cell neighbourhoods (O'Sullivan, 2001)

Studies have been conducted at the city scale including O'Sullivan (2001) who theorises combining graph theory and cellular automata to illustrate the dynamics of the city. His research involves each individual building within a neighbourhood being represented as a node with edges drawn to both facing and adjacent properties (Figure 7).

Urban Design

In his paper *Theory of the Urban Web* Salingaros (1998) applies a graph method – the urban web – in order to consider planning principles mathematically. He concludes that his principles are satisfied by all successful urban environments and violated by "environments that fail, that are unfriendly, that are unpleasant, that are isolating and that are dehumanizing" (Salingaros, 1998). The nodes within the *Urban Web* are not buildings but areas of human activities and the edges are defined by topological connective elements such as pedestrian paths or green spaces. The terms used are more specific versions of Lynch's Image of the City (1960). It is noted empirically that the stronger the connections between nodes the more life there is within a city (Alexander, 1965) and that complexity is a requirement of a city, without it the city is dead, however "without sufficient organization, a city becomes chaotic and unliveable" (Salingaros, 1998).

Salingaros discusses current planning policy identifying what contributes to successful urban design and what detracts from it. For example, it can be shown that the urban web is damaged by the use of non-linked design styles (Batty & Longley, 1994). He also chastises the insistence upon geometric regularity in urban plans as these are damaging to the nodes and connections of the urban web. It is not possible to successfully force human activity into a geometrically simple form and, in fact, a successful, strongly connected urban design "usually looks irregular from the air" (Salingaros, 1998; Gehl, 1987; Hillier, 1996).

> "Architects and urban planners became infatuated with visual simplicity and ignored the fundamental process of organization, which is not visually simple. We now have many examples of urban regions where the complexity has been eliminated altogether by suppressing connections (Batty & Longley, 1994). The visual purity in the plan has severely curtailed human activities that led to urbanism in the first place" (Salingaros, 1998).

Salingaros also discusses the rectangular grid in urban design, he states in theory it has organisational advantages but the block rigidity of a rectangular plan limits the number of connections that could be achieved if diagonals were permitted to cut across city blocks. He also references Camillo Sitte who verified that "curved streets of medieval towns give the greatest pleasure" (Salingaros, 1998), and identifies that such curving streets increase the likelihood of increased connections, this is then expanded upon by Bafna (2003) who notes that correlations between connectivity and integration values of cities can be predictors of the intelligibility of an environment.

"An urban setting such as Manhattan with a regular grid network should have a very low intelligibility, whereas a more traditional market town, such as Mytilini in Greece (Peponis et al., 1989) with its labyrinthine street network, should have a moderate to good intelligibility" (Bafna, 2003).

There are two reasons often referenced for this, the first is that while randomly wandering the streets of a medieval town the probability of finding an area of high integration – a more intelligible area – is greater than if one were randomly wandering a grid layout. The second is that a medieval area provides more "cognitive cues for both local and global properties" (Bafna, 2003), which is to effectively state that a medieval town simply feels more intelligible.

Social Network Analysis

Park (2015) proposes the use of social network analysis (SNA) techniques in order to find appropriate urban design patterns (Alexander et al., 1977) for a design problem. By representing each of the 253 empirical design patterns (Alexander et al., 1977) as a node with edges connected between linked patterns (Figure 8) Park is able to identify key patterns through the SNA statistical methods of degree centrality, closeness centrality, and betweenness centrality.

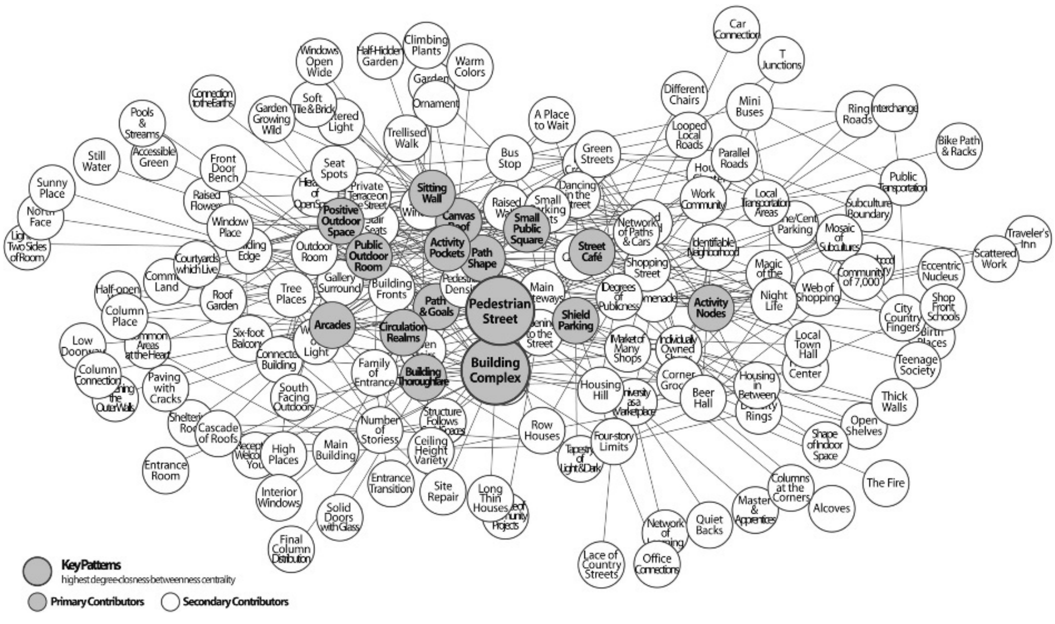


Figure 8 – The network of patterns for a downtown design (Park, 2015).

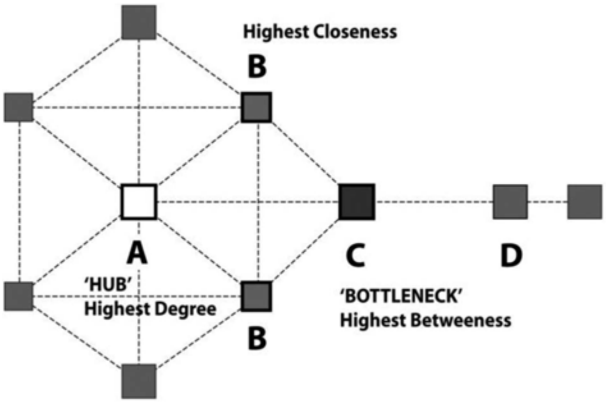


Figure 9 – An example of centrality measures in a network (Krebs, 2011; Park, 2015)

> "Degree centrality, the simplest way to measure centrality, refers to which nodes have higher numbers of direct connections (Freeman, 1977). … A node with high degree centrality, which means it has a high number of direct relations with other nodes, tends to influence actively other nodes in the given network (Kim et al, 2011). In [Figure 9] node A holds the highest number of direct connections and thus has the highest degree centrality score and functions as a hub in this network. Closeness can be regarded as a measure of how long it will take to spread information sequentially from the starting point to all other nodes (Newman, 2005). Closeness centrality is the inverse of the sum of the distance from one node to all other nodes. Thus, when a node becomes more central, its total distance to all other nodes becomes less (Freeman, 1977). For instance, a node with the highest closeness has the shortest paths to others, which allows quicker access to other nodes. Nodes B have the shortest paths to all the connected nodes, which means they have the highest closeness centrality.
>
> Betweenness centrality assigns a high centrality score to a node that holds large numbers of the shortest paths that link it to other nodes. This score represents the importance of the node in the network. If a node with a high betweenness is removed, the network can be divided and may lose its systematic function. It plays a broker role, one that acts as an intermediary, in the network. Node C is located at a critical point in the network which can split the network in half, isolating node D from A and B. This connecting role gives node C the highest score for betweenness centrality."
>
> Park (2015)

Park concludes that while her method can't be used to identify patterns that are appropriate to a particular design problem what it can do is provide "a rational and statistical basis for identifying the less significant patterns that can be dropped from the list" (Park, 2015). Additionally, it provides an indication of which patterns should be prioritised when time or budget require the number of patterns to be limited.

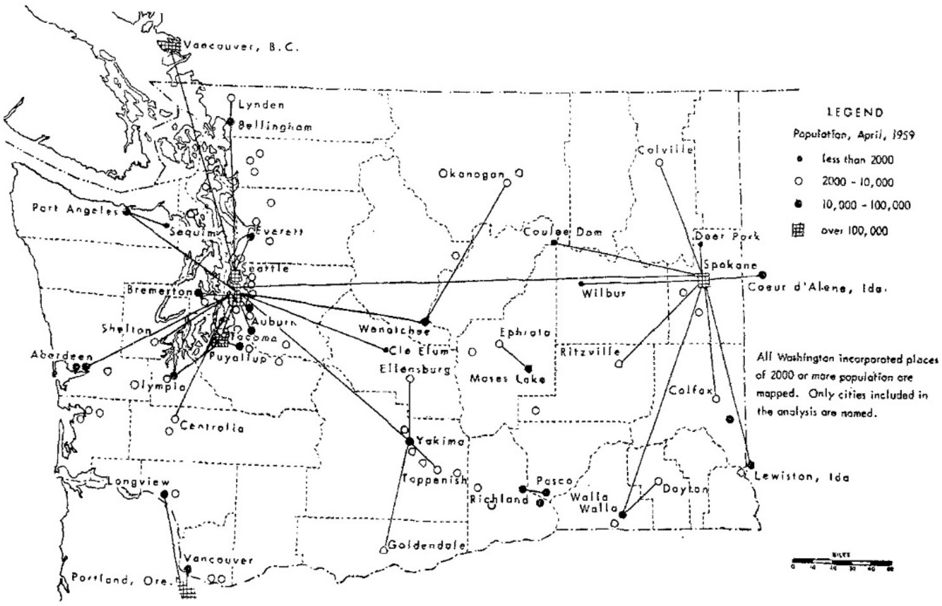


Figure 10 – Nodal structure based on telephone data, Washington state, 1958 (Nystuen, 1961)

Earlier papers such as Nystuen (1961) work at a regional scale and use graph theory to map inter-city relationships (Figure 10), in this case by the volume of inter-city telephone calls and in the case of Smith & Timberlake (1995) by volume of air travel.

Identity of the City

In *Livable Cities Observed,* Crowhurst Lennard & Lennard (1995) propose that the liveable city has an intrinsic DNA encoded within the architecture and spatial characteristics and these elements contribute towards the sense of identity of the city. It is the combination of materiality, colour, form, scale, and detailing of public and semi-public spaces that this image is derived from and that gradually the DNA is being eroded as developers are permitted to build structures that have no relation to their context, or to the character and traditions of the city. "Buildings compete for attention but do not pay attention to each other. The dialogue among building[s] is too often characterized by fragmentation and discontinuity, and the collage of buildings and public spaces creates a profound sense of anomic dissociation" (Crowhurst Lennard & Lennard, 1995). The authors argue that key to protecting the sense of identity – the DNA – is dialogue between architects and the community going so far as to suggest that design guidelines should be stricter and tailored more thoroughly to the unique heritage of the city. They claim that within London Docklands a lack of rigorous design guidelines "encouraged overbuilding and lack of a coherent urban fabric" (Crowhurst Lennard & Lennard, 1995) whereas the policies for a number of small American cities protected their identity and urban DNA. Crowhurst Lennard & Lennard state that the identity of the city stems from the street façade, the entity that is the interface between the public image and private functions (Park, 2015).

The authors make a number of claims regarding contributors to a healthy city, in their view cities should be a continuous built fabric with buildings of similar scale, form and materiality with only minor variance within the detailing of each structure. Functional entities within the city such as offices, homes, shops and schools should be integrated together with each at the same level of development. If these requirements are not met, then the cultural decay will put the identity and health of the city in jeopardy. Suburban sprawl is one of the most notable threats to cities at present (Crowhurst Lennard & Lennard, 1995). Consistent with Salingaros (1998) the authors claim that the suburban sprawl reduces the integration between contrasting city functions by separating them entirely. Instead of being able to easily walk to work many cities in America have become reliant upon vehicular transport to commute. "For suburban tissue to be healed it is clear that the suburbs need to receive high-density urban nuclei that provide all the necessary functions within walking distance" (Crowhurst Lennard & Lennard, 1995). For, it is the close proximity between living, working and socialising that successful settlements are based upon.

The authors chastise architects for their vanity projects that follow the international fashion of the moment and by doing so ignore the traditions of the local culture. These views are supported by Riza, Doratli & Fasli (2011) who discuss the contributions of four iconic buildings to the image of the city and how these pieces of architecture affect the quality of life of the residents. Riza et al. conclude that when iconic buildings pay attention to their context – such as in the examples of Frank Gehry's Dancing House and I. M. Pei's Louvre Pyramid – the residents benefit from improved quality of life because the buildings support the identity of the city. However, when context is ignored for example Norman Foster's Swiss Re Office or Gehry's Guggenheim Museum, Bilbao the immediate surroundings are diminished and the city's identity damaged.

Riza et al. argue that cities have three ways of promoting themselves and establishing their brand or identifiable identity: they can promote the restoration of historic buildings and heritage assets, construct iconic buildings or commit to cultural mega events. The authors cite a number of papers which confirm a positive relationship between the identity of a city and the satisfaction of tourists and residents (Bigne et al., 2001; Petrick, 2004; Chen et al., 2007; Chi et al., 2008).

The work of Zhou et al. (2014) presents an application to automatically identify the identity of a city by analysing geotagged images from Flickr. A database of 2 million photos is downloaded for the 21 cities and each is run through a *scene attribute classifier* to recognise elements within the image. These attributes – detailed in Zhou et al. (2014b) – are grouped and merged into 7 *city attributes*: green space, water coverage, transportation, architecture, vertical building, aesthetic activity, and social activity. The distribution of these classified and geotagged images can then be mapped across the city (Figure 11).

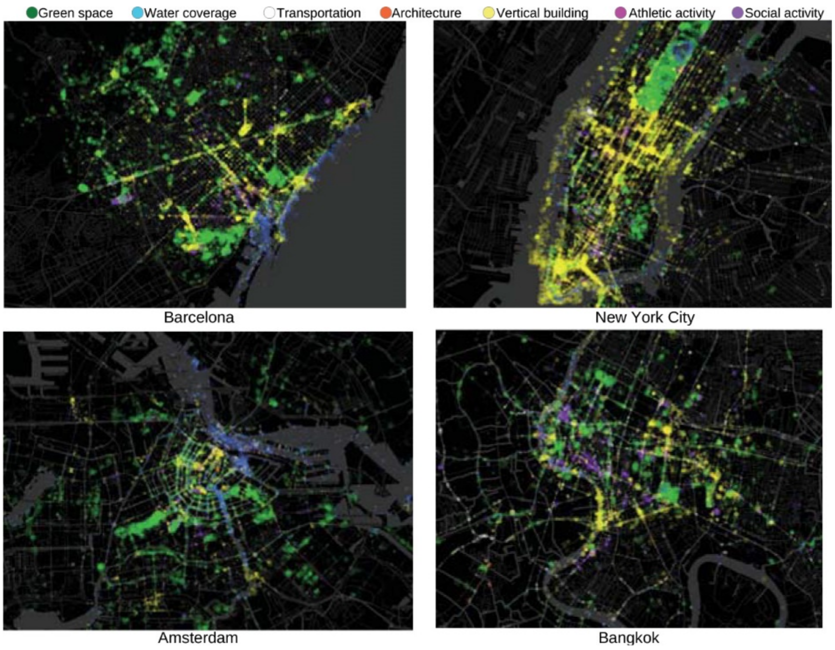


Figure 11 – City perception map of Barcelona, New York, Amsterdam, and Bangkok. Each coloured dot represents a geo-tagged image detected with one city attribute (Zhou et al., 2014)

The next stage of their paper investigates whether it is possible to train the classifier to do the reverse: can the city be determined simply by analysing a city? The authors recognise that the success is not particularly high, however, the application has a "good enough discriminative ability compared to random chance" (Zhou et al., 2014). They hypothesise that the misclassification rate of images can be used as an objective measure of the similarity between two cities, with their results validating "that the geographical distance between cities plays an important role in determining the similarity of cities. Indeed, historically there was more trade and cultural exchanges between spatially-neighbouring cities, so they would share similar elements in their form and culture" (Zhou et al., 2014). They also note that there is a decreased distinction between metropolises such as London, Tokyo, Paris and New York as they mix “cultures from all over the world” (Zhou et al., 2014).

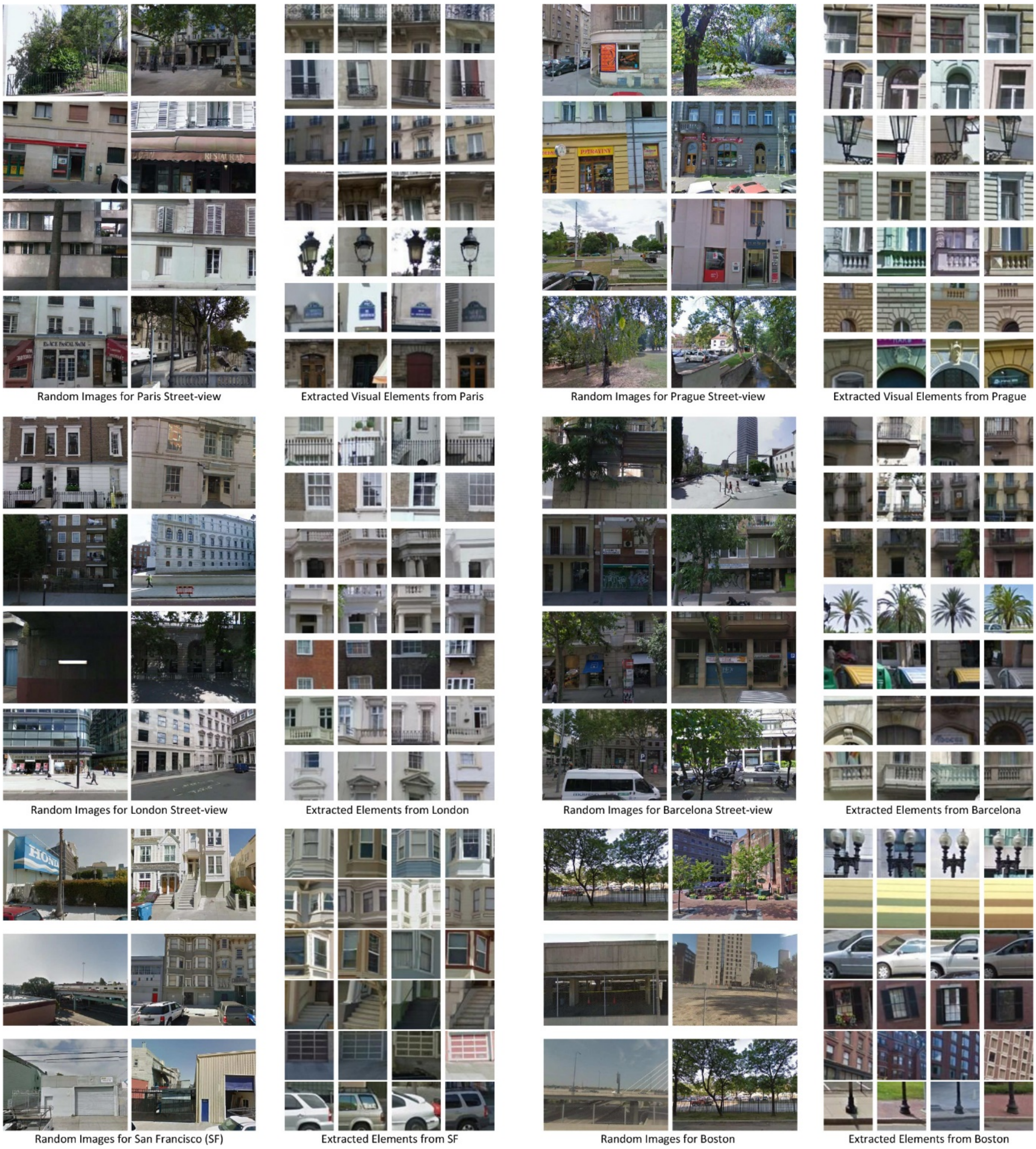


Figure 12 – Google Street View vs. geo-informative elements for six cities. Arguably, the geo-informative elements (right) are able to provide (Doersch et al., 2012)

While Zhou’s work investigates how “City identity emerges in every aspect of daily life and implicitly exists in people’s perception of the city” (Zhou et al., 2014) Doersch et al. (2012) explores the distinguishing architectural elements that capture the aesthetic identity or “look and feel” of a geographic area. They create an application to inspect Google Street View images and automatically identify clusters of *geographically-informative elements*. The visual elements identified by their application (Figure 12) “convey the stylistic feel of the city” (Doersch et al., 2012). The authors were then able to investigate whether there were connections between the discriminative features of European and non-European cities (Figure 13), finding that “while arches are common in cities across Europe, double-arches seem rare in London. Similarly, while balcony railings in Paris, Barcelona and Milan are all made of cast iron, they tend to be made of stone in London and Prague.” The algorithm is also able to detect the regions from which architectural influence has been drawn. Figure 14 “shows images from the 5th arrondissement of Paris, pointing out which elements are specific to that arrondissement, which are Paris-specific and which are pan-European. For example, the stone balcony railings and arches are pan-European, windows with collapsible shutters and balconies with iron railings are Parisian, and the grooves around the windows are typical of the 5th arrondissement” (Doersch et al., 2012). Visual patterns are also noted in the scale of objects, for instance, the floor to floor height of Parisian buildings is typically equal, whereas in London the floor heights differ, “with the first floor much taller and more stately” (Doersch et al., 2012).

Figure 13 – Architectural patterns across Europe. While arches (A) are common across all Europe, double arches (B) seem rare in London (Doersch et al., 2012)

Figure 14 – Detecting architectural influences. Each image shows confident detections for architectural styles at different geographic scales (Doersch et al., 2012)

The *World Cities Culture Report* (World Cities Culture Forum, 2016) uses quantitative measures of cultural infrastructure to assess the culture of world cities, these metrics include the number of UNESCO World Heritage Sites, art galleries, concert halls and more. The report also takes into account the number of visitors or admissions to venues in order to assess public engagement with cultural institutions.

Volunteered Geographic Information

Previously geospatial data was only accessible to individuals within the academic realm through geographic information systems (GIS) databases, but in recent years, there have been profound changes to the ecosystem in the availability of data; primarily through advances in web technologies and the open culture ingrained within the internet.

> "The development of the Web has seen a wide range of legal, regulatory, political and cultural developments surrounding the control, access, and rights to digital content. However, the Web has also always had a strong tradition of openness: working with open standards, using open source software … making use of free data, re-using data and working in a spirit of open innovation" (Anderson, 2007, p. 25).

Crowdsourced geospatial data is known within academia as volunteered geographic information (VGI) (Ford & Graham, 2016) and is a new democratic method by which individuals voluntarily collate, moderate, and correct geospatial data. This culture has allowed projects such as Google Maps, Open Street Map, Wikipedia and Wikidata to thrive. These services collate a mixture of authoritative and crowdsourced data which they then make accessible via Application Program Interfaces (APIs) for developers to freely and openly integrate into applications within the limitations of the terms of use.

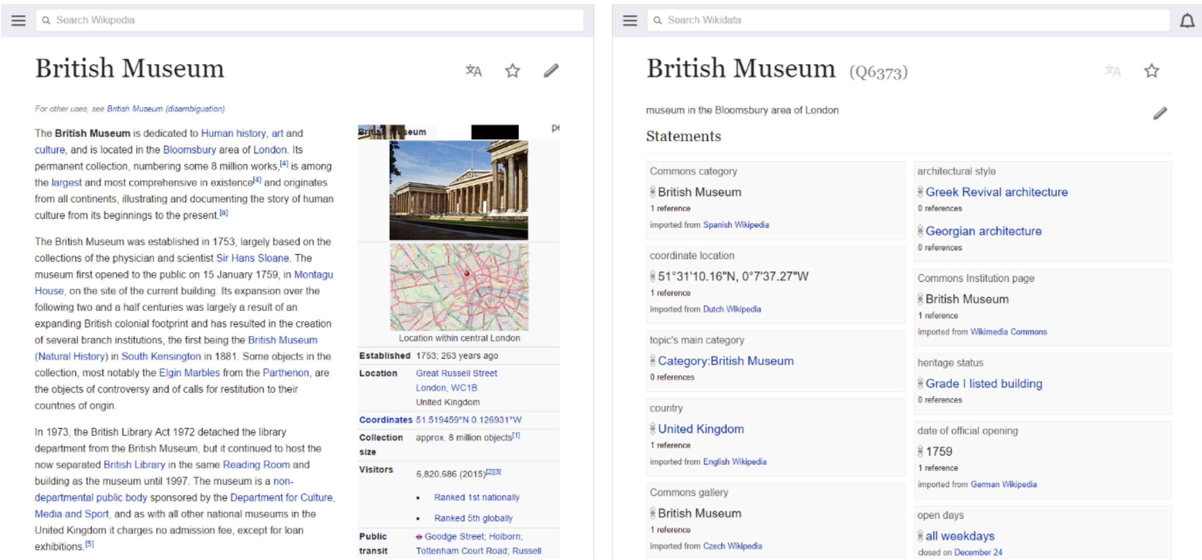


Figure 15 – Comparison between unstructured Wikipedia article (left) and structured Wikidata statements (right)

APIs return two types of data: structured and unstructured. Wikipedia, for instance, provides unstructured data, by sending an HTTP *Get request* to the API endpoint users can retrieve – depending on the request –

the latest or past revisions of any Wikipedia article. In the same manner, as if the user went directly to the URL of a Wikipedia article the API request will return the complete article – albeit in a data format more digestible by an application. The article itself is unstructured data because although it contains a lot of information the application is unable to understand that data without processing it in some manner. On the other hand, if we send a request to the Wikidata API we will retrieve structured data that it is organised such that it can be interpreted, formatted or displayed by an application without the need for further processing (Figure 15).

The ease of availability of such vast quantities of collective geospatial data from APIs have led to a number of academics arguing the case for using this data within research despite its non-authoritative source; these include Batty et al. (2010), Antoniou et al. (2009), Du et al. (2011a), Du et al. (2011b) and Anand et al. (2010).

Many research projects have already been published built upon public APIs which use crowdsourced data, for example, Toman & Olszewska (2014) who propose an algorithm to build graphs of ski pistes from KML data extracted from Google Maps. Other projects attempt to combine structured data retrieved from APIs with unstructured data mined from blog articles such as Jiang, K. et al. (2012) and Ji et al. (2009). There is also research into the automatic estimation of tourist attractions locations by the analysis of geotagged photos on websites such as Flickr: Li et al. (2013), Han & Lee (2014), Deeksha et al. (2015).

Ford & Graham (2016) discuss two structured data APIs: Google Knowledge Graph Search and Wikidata and notes four elements that are lost when data becomes stored in structured data APIs: nuance, provenance, user control and alternative perspectives (Ford & Graham, 2016). Using the example of the Google Knowledge Graph that is displayed along with results during user searches (Figure 16), and in particular the results for politically contested entities such as "Jerusalem" the authors illustrate that not all answers are as defined as presented by the Knowledge Graph. For instance, it is stated that Jerusalem is the "capital of Israel" whereas in Wikipedia it is further qualified that while currently administered by Israel the city is claimed by both Israel and Palestine (Ford & Graham, 2016).

The authors conclude that structured data APIs "eliminate some of the scope for culturally contingent representations of places, process people, and events, but also depoliticises and obfuscates some of those processes of representation" (Ford & Graham, 2016). They believe that "structured data is not conductive to the representation of minority opinions" (Ford & Graham, 2016). They fear that applications that retrieve information from these APIs will fail to relay the nuance and provenance of the data such that "messy political informational layers of cities [will be made] more transparent to machines and more opaque to humans" (Ford & Graham, 2016).

Figure 16 – Results of Google's Knowledge Graph for "Jerusalem" (4 December, 2014) (Ford & Graham, 2016)

Summary

As discussed in the introduction, the intention of this paper is to present a computational method for defining and representing the *DNA of the city* as an *identity graph*, inspired by Hillier & Hanson's (1984) syntactical representation of houses as graphs. This literature review has explored how this avenue of research has developed over time such as Dalton & Kirsan's (2008) introduction of the *genotype signature* or *prototype* of a dwelling derived through graph matching. Methods of urban analysis which syntactically or diagrammatically represent cities were introduced such as the image of the city (Lynch, 1960) and Salingaros's (1998) urban web where the city is reduced to a series of human activity nodes connected by edges representing paths. The application of social network analysis methods to Alexander's (1977) urban design patterns by Park (2015) was then discussed as a method of objectively analysing graphs.

Next, the concept of the identity of the city was explored with a number of interpretations and computational methods of deriving the identity introduced. As culture and identity are expansive subjects the focus within this study has been on the architectural identity of the city. Crowhurst Lennard &

Lennard (1995) argue that the identity of the city is found within the materiality, colour, form, scale, and detailing of the street façade – the interface between the public realm of the city and the private interior. While Zhou et al. (2014) compare cities objectively by the presence or absence of cultural attributes in photographs and then objectively measure the similarity of cities by the rate at which their algorithm misclassifies a photograph as showing one city rather than another. The work of Doersch et al. (2012) to automatically distinguish the unique geo-cultural architectural elements that capture the identity of the city was then introduced, which reaffirmed that it was the form and materiality of subtle architectural features such as balconies, railings and window details that make a city or region distinguishable. Riza et al. (2011) looked at identity from another perspective and relate identity to iconic structures similar to the landmarks of the Image of the City (Lynch, 1960).

Finally, the literature review discussed the use of volunteered geographic information in academia and how such data can be accessed through structured data APIs such as Google Knowledge Graph Search and Wikidata. Examples of experimental methodologies where a computational method is presented have been included throughout this literature review and have guided the methodology of this research.

03 METHODOLOGY

This research has followed an experimental methodology, which in accord with the principles of Amaral et al. (2011) was divided into two phases. First, the exploratory phase evaluated candidates for the nodes and edges of the graph, defined the cities to be studied, and developed the application. The evaluation formed the second phase, where the graphs of the case study cities were analysed and compared using graph theoretical techniques.

Nodes

From the literature, it became apparent that *places of interest* are the ideal candidates for the graph node. Similar to Lynch's (1960) landmarks a *place of interest* is a reference to a geospatial entity that is of interest to city residents and visitors alike. It is within the subtleties of architectural style of *places of interest* that Crowhurst Lennard & Lennard (1995), Riza et al. (2011) and Doersch et al. (2012) believe the contributors to the identity and DNA of the city lie.

At first, this research experimented with retrieving *places of interest* from the Google Places API (Google Places API, n.d.-a). The Place Search endpoint (Google Places API, n.d.-b) permitted the search to be limited to specific *types* such as "*museum*", "*park*" or "*city_hall*" (Google Places API, n.d.-c), however, the places that were retrieved were lacking quality. Results that one would consider a place of interest – such as notable museums and train stations that are significant landmarks within a city – were mixed in with less interesting results, such as individual bus stops, and often places were found to be entirely miscategorised, such as the office of a local taxi firm being identified as a train station.

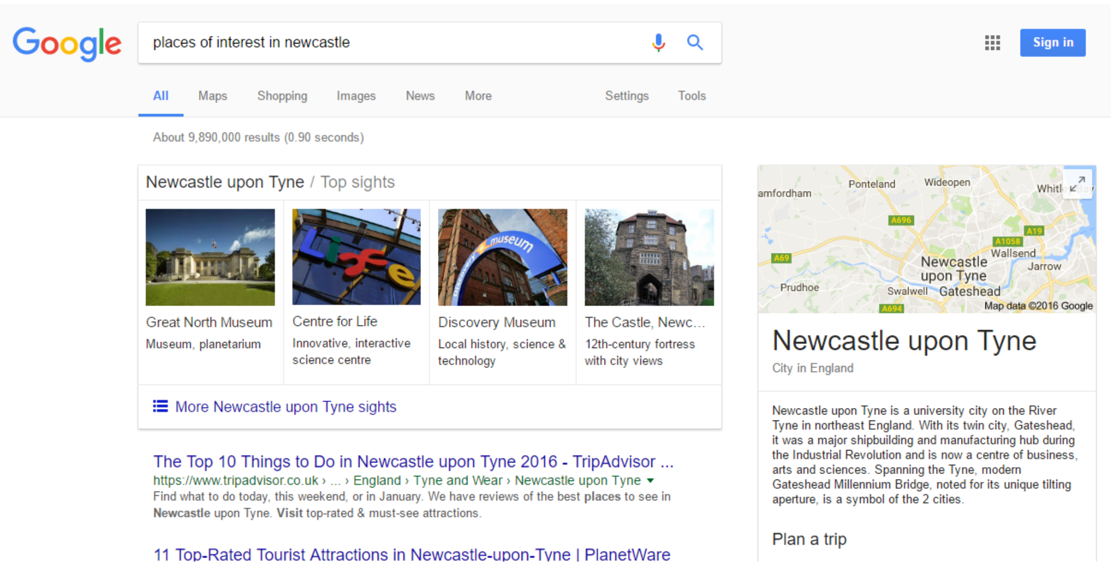


Figure 17 – Google Knowledge Graph Search picks out specific "top sights" when searching for places of interest and displays them alongside regular results

The developed application now uses the Google Knowledge Graph Search API (Google Search, n.d.) to determine a list of relevant *places of interest* for a named city. By specifying the *query* parameter as the name of the city and the *types* parameters as *LandmarksOrHistoricalBuildings*, *TouristAttraction,* and *CivicStructure* – based on Schema.org specifications (Schema.org, n.d.) – we can retrieve a list of *places of interest* within the city that are ranked by their relevance. The ranking is determined by a proprietary algorithm which determines the relevance of the results to the query based upon metrics such as relatedness, notability, and popularity derived by analysis of linked structured data (Shen et al., 2012; Starr, 2015; Ford & Graham, 2016). This results in the returned data being significantly more relevant to the identity of the city compared to those returned by the Google Places API. Rather than the unfiltered list of all local businesses and properties tenuously linked to type criteria that were retrieved from Google Places the results from the Google Knowledge Graph Search API are landmark and iconic structures that are ordered by their relevance and notability as *places of interest*.

Edges

Within a graph, the edges are the connective elements between nodes. Within space syntax, the edges are typically a topological feature – such as the axial line, or adjacent and opposite buildings (O'Sullivan, 2001) – however topology is not necessarily relevant to defining the identity of the city, it was, therefore, important to explore alternative connective elements.

Crowhurst Lennard & Lennard (1995) and Doersch et al. (2012) discuss how the architectural façade – the interface between the public and private parts of the city is the major contributing factor to the public identity of the city, this is expressed through the materiality, detailing and architectural style of the building. Crowhurst Lennard & Lennard (1995) also identifies that one of the key tasks of an architect is to understand the heritage and cultural aspects of a city with which its residents identify, thus it can be interpreted that as architects develop more projects within a city they will have a deeper understanding of the culture of that place. Finally, the *World Cities Culture Report* (World Cities Culture Forum, 2016) measures the culture of cities by quantifying the cultural infrastructure of the city. Therefore within this study, the place nodes within the graph will be connected, where they share a type, architectural style, or architect.

Using these connections, the *identity graph* will be an objective representation of the architectural identity or *DNA of the city* and it should be possible to quantitatively compare cities as the more connected the nodes within the graph the more defined the identity of the city must be.

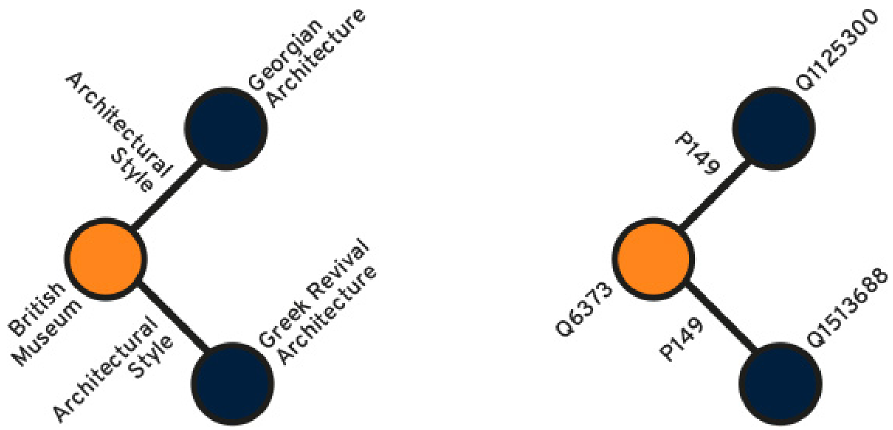


Figure 18 – Wikidata entities linked together by properties

In order to do so, we must gather the metadata for each *place of interest*. The Wikidata API (Wikidata, n.d.-a) is the structured data counterpart to Wikipedia, it is a curated repository of facts with each claim referenced to their primary sources (Vrandečić & Krötzsch, 2014). It was developed following the principles of the graph; each entity has a unique identifier, for instance, the *British Museum* is *Q6373* (Wikidata, n.d.-b). Each entity then has a series of claims which are property and value pairs. For example, the *British Museum* has the claim property *Architectural Style* with two values *Greek Revival architecture* and *Georgian architecture (*Figure 18*)*. For the benefit of legibility, the associated labels have been used, in reality, programmatically speaking property *P149* (Architectural Style) connects entity *Q6373* (British Museum) to both *Q1513688* (Greek Revival architecture) and *Q1125300* (Georgian architecture).

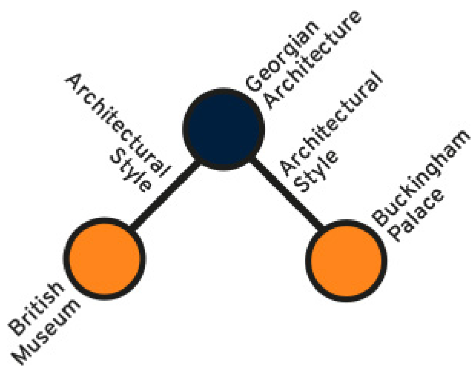


Figure 19 – Example of two places connected by a shared claim entity

By matching places with alike properties – such as the *British Museum* to *Buckingham Palace* through their shared architectural style of *Georgian architecture (*Figure 19*)* – we will build a graph of the identity of the city. From this graph, it will be objectively apparent which factors contribute most to that identity and which places are perhaps not as integrated into the identity of the city as they should be.

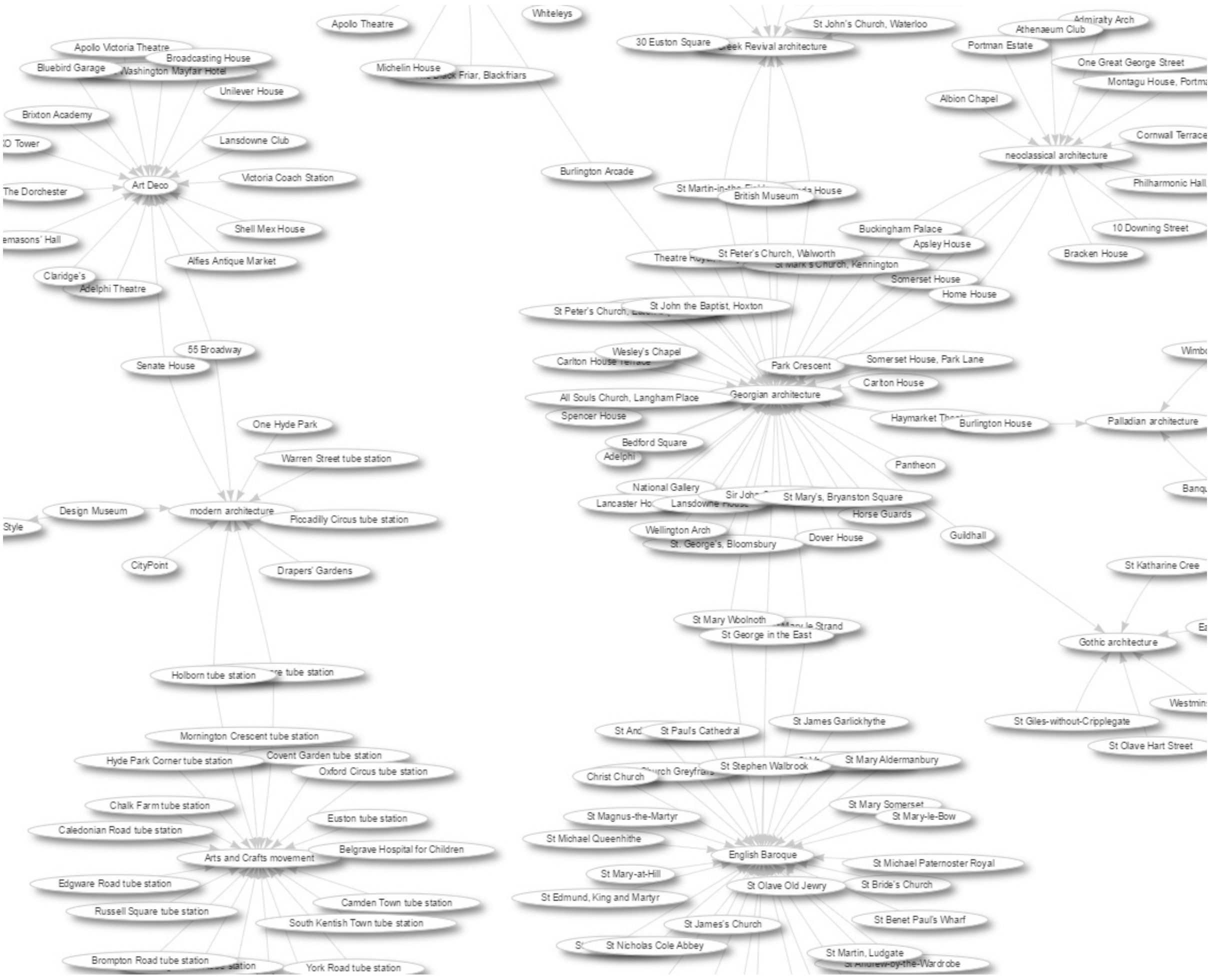


Figure 20 – Extract of Results for Wikidata Query Service script

In order to quickly test the viability of this method of linking nodes a test script was written for the Wikidata Query Service in SPARQL – an RDF markup language favoured by the Query Service, the script (Elsdon, 2016c) is included in Appendix A. It retrieves a list of geolocated places which have an architectural style property within 5 km of the centre of a city – in this example, London – and then connects each place to its architectural style. The resulting graph is shown in Figure 20 and an interactive model can be found at http://bit.ly/2hVMtL1

Cities

In order to demonstrate the capabilities of the final application, a number of case studies of European cities have been selected. The candidates have been chosen based upon their comparative visitor numbers. Barcelona, Vienna, and Venice have 6.0 million, 5.4 million and 5.2 million visitors respectively (Euromonitor

International, 2016) and each has a high level of architectural cultural identity (JLL, 2016; World Cities Culture Forum, 2016) which should be represented within the graph. Additionally, comparisons between London and Paris – 17.4 million and 15.0 million (Euromonitor International, 2016) – are included in order to discuss other findings disclosed by the application with regards to more populace, global cities with high visitor numbers.

Application Development

The application has been developed as a frontend web application in the AngularJS JavaScript framework. The differentiation between a frontend and backend application lies in where the application logic is processed. For a frontend application, the full application runs within the user's web browser, whereas a backend application is one that runs on a server. As illustrated by Figure 21, the application makes a series of requests to the external APIs – Google Knowledge Graph Search and Wikidata – the responses are processed, then the data is stored in the browser's local storage and the resulting graph is output to the user within the browser.

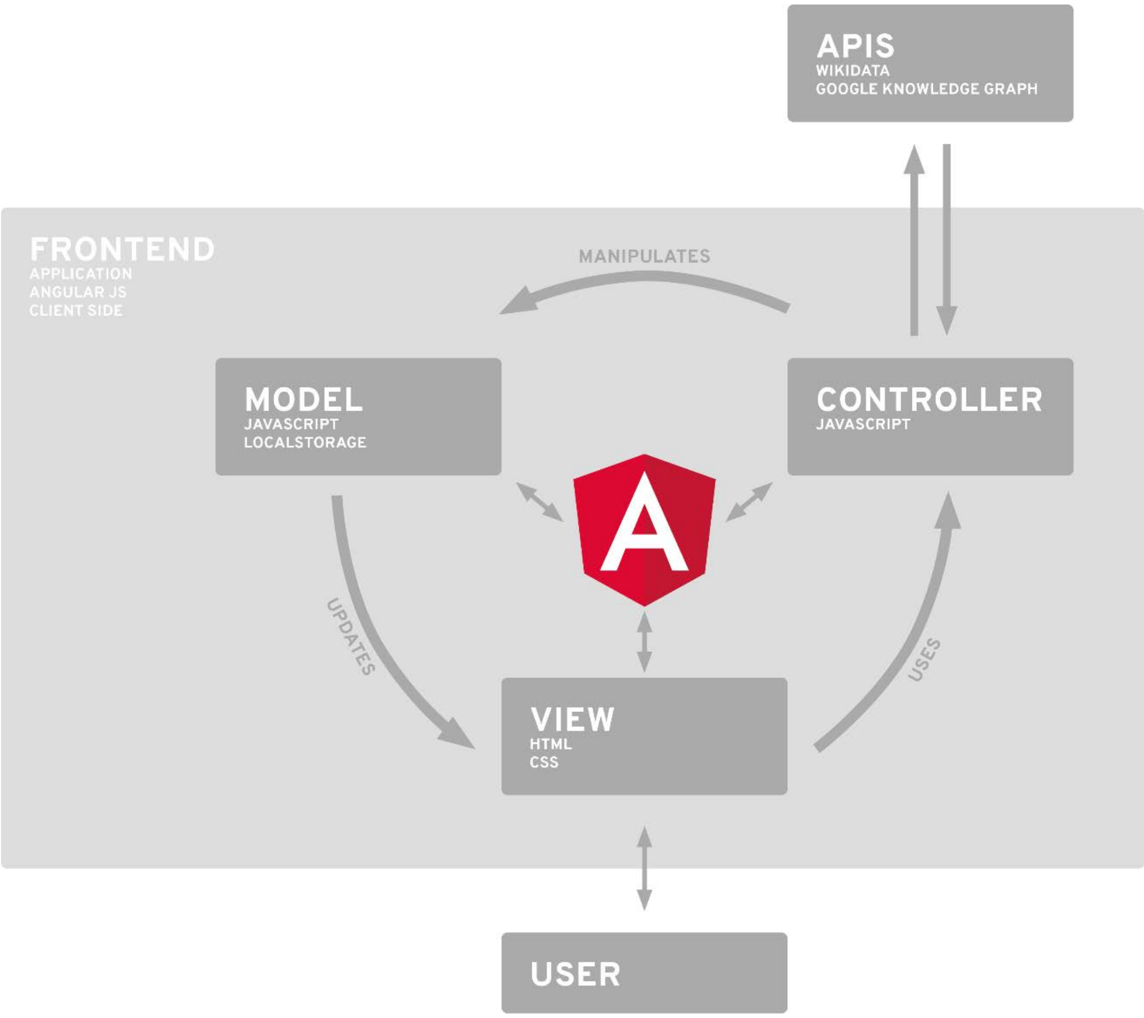


Figure 21 – Application diagram showing MVC pattern and separation between frontend and APIs

JavaScript, alongside HTML and CSS, is one of the core technologies used in websites, more specifically it is used for interactive elements within those pages, and therefore is the natural language of choice for interactive web applications. The AngularJS framework was selected for the application primarily due to familiarity – an often overlooked factor in selecting appropriate technologies – but also due to its rapid speed of development (AngularJS, n.d.), extensibility, and modularity through pre-written libraries. The framework follows the Model–View–Controller (MVC) pattern which separates the model (data storage), the view (user interface) and controller (data processing).

As Figure 21 illustrates, the user interacts with the User Interface (UI), these interactions are processed within the controller which stores the results in the model which then returns the results back to the view where it is formatted and displayed in the user interface.

Cytoscape.js (Cytoscape.js, n.d.-a) has been chosen as the graph visualisation library for the application, D3.js (D3.js, n.d.) was also considered but following some initial experimentation Cytoscape.js was deemed to be more graph focused, better suited to the task and has the additional benefits of inbuilt graph theory methods and ease of export to common graph analysis and network analysis software such as Cytoscape Desktop (Cytoscape, n.d.).

Social Network Analysis

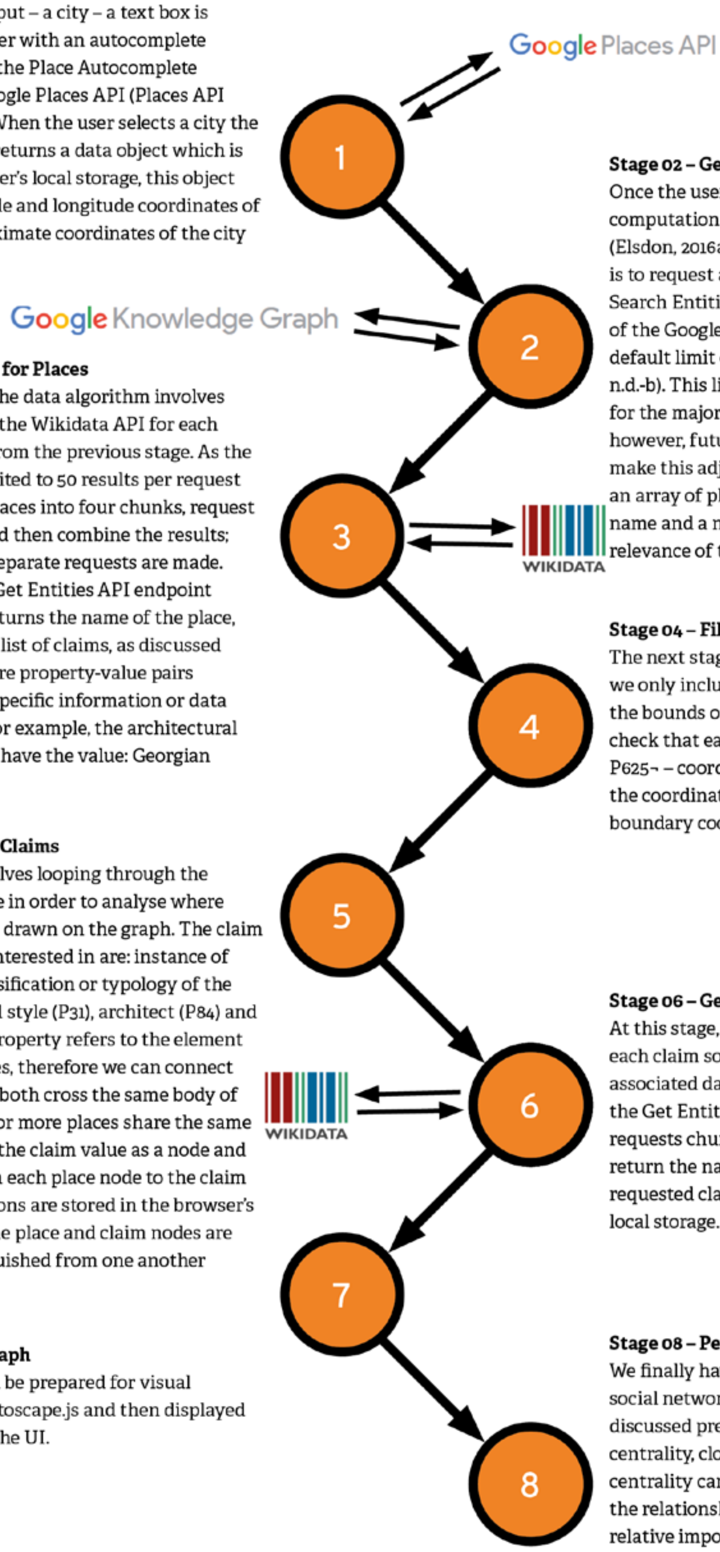
Social network analysis techniques – such as those used by Park (2015) – will be performed in order to quantitatively analyse the *identity graphs*, and this analysis will be performed using the inbuilt analytic functions within Cytoscape.js. The function for degree centrality (Cytoscape.js, n.d.-b) is based on the work of Opsahl et al. (2010), closeness centrality (Cytoscape.js, n.d.-c) by Bavelas (1950), and betweenness centrality (Cytoscape.js, n.d.-d) is from Brandes (2008). For both the degree and closeness centrality the normalised function variants will be used, in this form the figures are divided by N - 1, (where N is the number of nodes in the graph), this means that the values can be compared between graphs of different numbers of nodes. The normalised figures will be presented to 4 decimal places consistent with Park (2015).

Application Logic

The majority of the computational logic of the application is completed within the *ComputeCtrl* (Elsdon, 2016a) controller. The process is dynamically illustrated step by step within the UI in an attempt to make the process as transparent as possible to the user (Figure 22).

Stage 01 – Define Criteria

The first stage of the application is the criteria definition; at present, the application only requires a single input – a city – a text box is presented to the user with an autocomplete functionality – via the Place Autocomplete endpoint of the Google Places API (Places API Web Service, n.d.). When the user selects a city the Google Places API returns a data object which is stored in the browser's local storage, this object contains the latitude and longitude coordinates of the city, and approximate coordinates of the city boundaries.



Stage 02 – Get list of Places

Once the user submits the input form the computation begins within the ComputeCtrl (Elsdon, 2016a) controller. The initial step of which is to request a list of places of interest from the Search Entities endpoint (Google Search, n.d.-b) of the Google Knowledge Graph Search API. The default limit of results is 200 (Google Search, n.d.-b). This limit has been found to be sufficient for the majority of cities so has not been adjusted, however, future iterations of the application could make this adjustable. The request will return an array of place objects including the place name and a number representing the presumed relevance of this place to the search criteria.

Stage 03 – Get data for Places

The third stage of the data algorithm involves fetching data from the Wikidata API for each place in the array from the previous stage. As the Wikidata API is limited to 50 results per request we break the 200 places into four chunks, request them separately and then combine the results; thus, in total four separate requests are made. For each place the Get Entities API endpoint (Wikidata, n.d.-c) returns the name of the place, a description and a list of claims, as discussed earlier, the claims are property-value pairs which are links to specific information or data about that place. For example, the architectural style property may have the value: Georgian architecture.

Stage 04 – Filter Places

The next stage is to filter the places, ensuring that we only include places that are located within the bounds of the specified city. To do so, we first check that each place of interest has the property P625¬ – coordinate location – then we check that the coordinates are within the city, using the city boundary coordinates we stored in stage 01.

Stage 05 – Analyse Claims

The fifth stage involves looping through the claims of each place in order to analyse where connections will be drawn on the graph. The claim properties we are interested in are: instance of (P31, this is the classification or typology of the place), architectural style (P31), architect (P84) and crosses (P177, this property refers to the element that a bridge crosses, therefore we can connect two bridges if they both cross the same body of water). Where two or more places share the same claim we will store the claim value as a node and create an edge from each place node to the claim node. The connections are stored in the browser's local storage and the place and claim nodes are graphically distinguished from one another within the graph.

Stage 06 – Get data for Claims

At this stage, we only know the identifier for each claim so the next step will be to request the associated data from the Wikidata API. Again, the Get Entities endpoint will be used with the requests chunked into groups of 50. The API will return the name and description for each of the requested claims, which will then be stored in local storage.

Stage 07 – Draw Graph

Next the graph will be prepared for visual presentation by Cytoscape.js and then displayed to the user within the UI.

Stage 08 – Perform Social Network Analysis

We finally have the data required to begin running social network analysis (SNA) on the graph. As discussed previously the SNA algorithms: degree centrality, closeness centrality, and betweenness centrality can be used to quantitatively analyse the relationships between nodes and identify the relative importance of each node (Park, 2015). The SNA algorithms built into Cytoscape.js will be used to perform this analysis and the data stored in the local storage to be presented within the user interface.

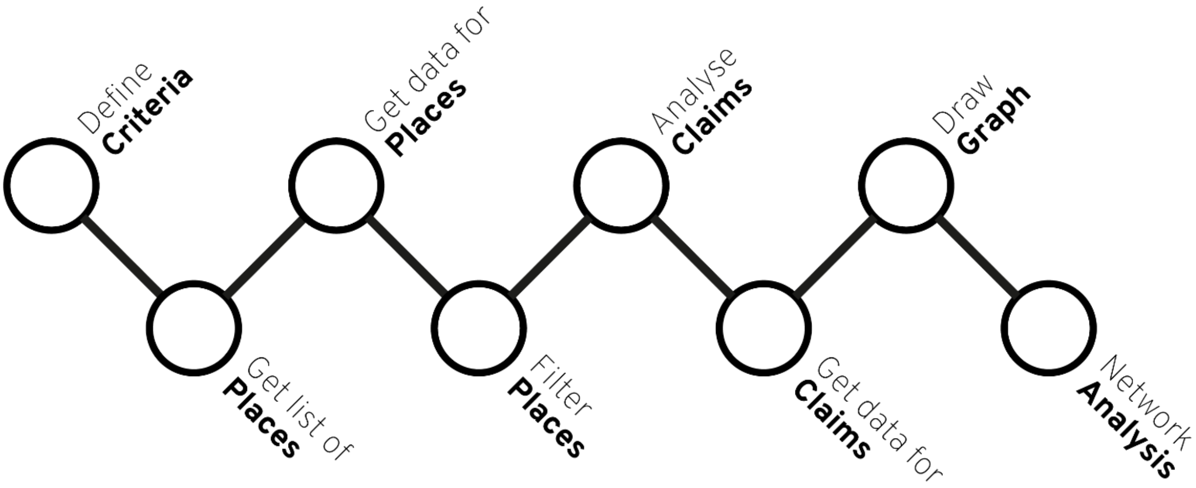


Figure 22 – Process diagram

04 APPLICATION

The application is hosted publicly and available for immediate use at:

https://graphingthecity.studiole.uk

The application has only been tested in the latest version of Google Chrome which at the time of writing is Version 55.0.2883.87 m. As such there is no guarantee that the application will function correctly in other browsers, it is therefore recommended that you use the application in Google Chrome.

If you encounter any issues with the application or find it to be temporarily unavailable, please contact the author directly to arrange a demonstration.

The source code has been released under an open source software license and is publicly available at the following code repository:

https://github.com/StudioLE/GraphingTheCity

Additional details on the features of the application are available in Appendix B.

05 RESULTS

The purposes of the application are two-fold, first to produce the identity graph, and second, to provide an interactive rich media interface to view and analyse the results. Subsequently, rather than viewing the static images in this paper it is important to view the graph and tables dynamically within the application.

The following results were recorded on 23rd December 2016 using version 0.3.0 of the application.

As the data from external APIs is likely to change over the time these results will not be precisely reproducible in the future, as such a method to directly reload the data discussed within this paper has been introduced. To do so, visit the following URLs:

Barcelona
https://graphingthecity.studiole.uk/#/load/barcelona

Venice
https://graphingthecity.studiole.uk/#/load/venice

Vienna
https://graphingthecity.studiole.uk/#/load/vienna

London
https://graphingthecity.studiole.uk/#/load/london

Paris
https://graphingthecity.studiole.uk/#/load/paris

	Nodes	Connections	Place Nodes	Claim Nodes	Visitors
Barcelona	109	128	83	26	6.0
Venice	130	166	104	26	5.2
Vienna	138	154	114	24	5.4
London	220	209	162	58	17.4
Paris	179	184	135	44	15.0

Table 1 – Node and connection counts for the identity graph of each city

Barcelona

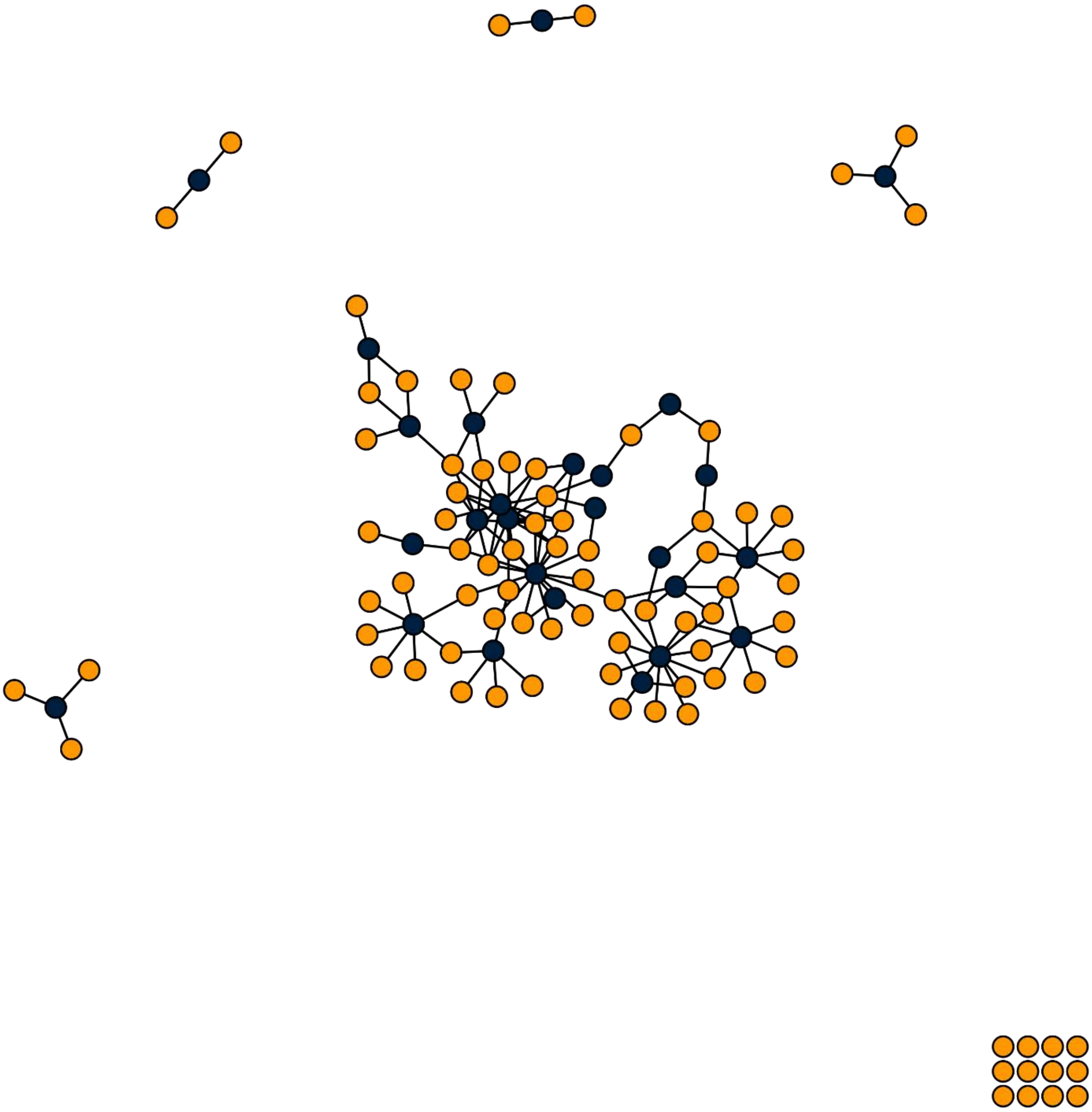


Figure 23 – Identity graph for Barcelona exported directly from the application

https://graphingthecity.studiole.uk/#/load/barcelona

	ID	Property	Label	Direct Connections	Degree Centrality	Closeness Centrality	Betweenness Centrality
1	Q41176	instance of	building	15	1.0000	1.0000	9442.67
2	Q34636	architectural style	Art Nouveau	12	0.8000	0.8665	3911.92
3	Q11303	instance of	skyscraper	12	0.8000	0.8610	2941.70
4	Q1122677	architectural style	Modernisme	11	0.7333	0.8369	3150.23
5	Q25328	architect	Antoni Gaudí	7	0.4667	0.7433	1106.09
6	Q207694	instance of	art museum	7	0.4667	0.6785	980.85
7	Q18142	instance of	tower block	7	0.4667	0.6410	1112.70
8	Q27686	instance of	hotel	7	0.4667	0.6067	1140.70
9	Q845318	architectural style	high-tech architecture	5	0.3333	0.7233	1326.45
10	Q33506	instance of	museum	5	0.3333	0.5905	655.30
11	Q1200957	instance of	tourist destination	4	0.2667	0.5851	461.74
12	Q16970	instance of	church	4	0.2667	0.5727	935.00
13	Q317876	architect	Lluís Domènech i Montaner	3	0.2000	0.6372	25.18
14	Q176483	architectural style	Gothic architecture	3	0.2000	0.4146	242.00
15	Q174782	instance of	square	3	0.2000	0.0840	6.00
16	Q8502	instance of	mountain	3	0.2000	0.0840	6.00
17	Q2070359	architect	Pere Domènech i Roura	2	0.1333	0.6332	417.10
18	Q16917	instance of	hospital	2	0.1333	0.6185	10.81
19	Q74156	architectural style	Moorish Revival architecture	2	0.1333	0.6026	7.66
20	Q811979	instance of	architectural structure	2	0.1333	0.5848	162.00
21	Q12518	instance of	tower	2	0.1333	0.5476	126.17
22	Q333585	architect	Itō Toyoo	2	0.1333	0.4877	1.00
23	Q483110	instance of	stadium	2	0.1333	0.4735	194.14
24	Q1003305	architect	Francesc Mitjans i Miró	2	0.1333	0.4621	122.14
25	Q79007	instance of	street	2	0.1333	0.0560	2.00
26	Q483453	instance of	fountain	2	0.1333	0.0560	2.00

Table 2 – Social network analysis of claims for Barcelona

Venice

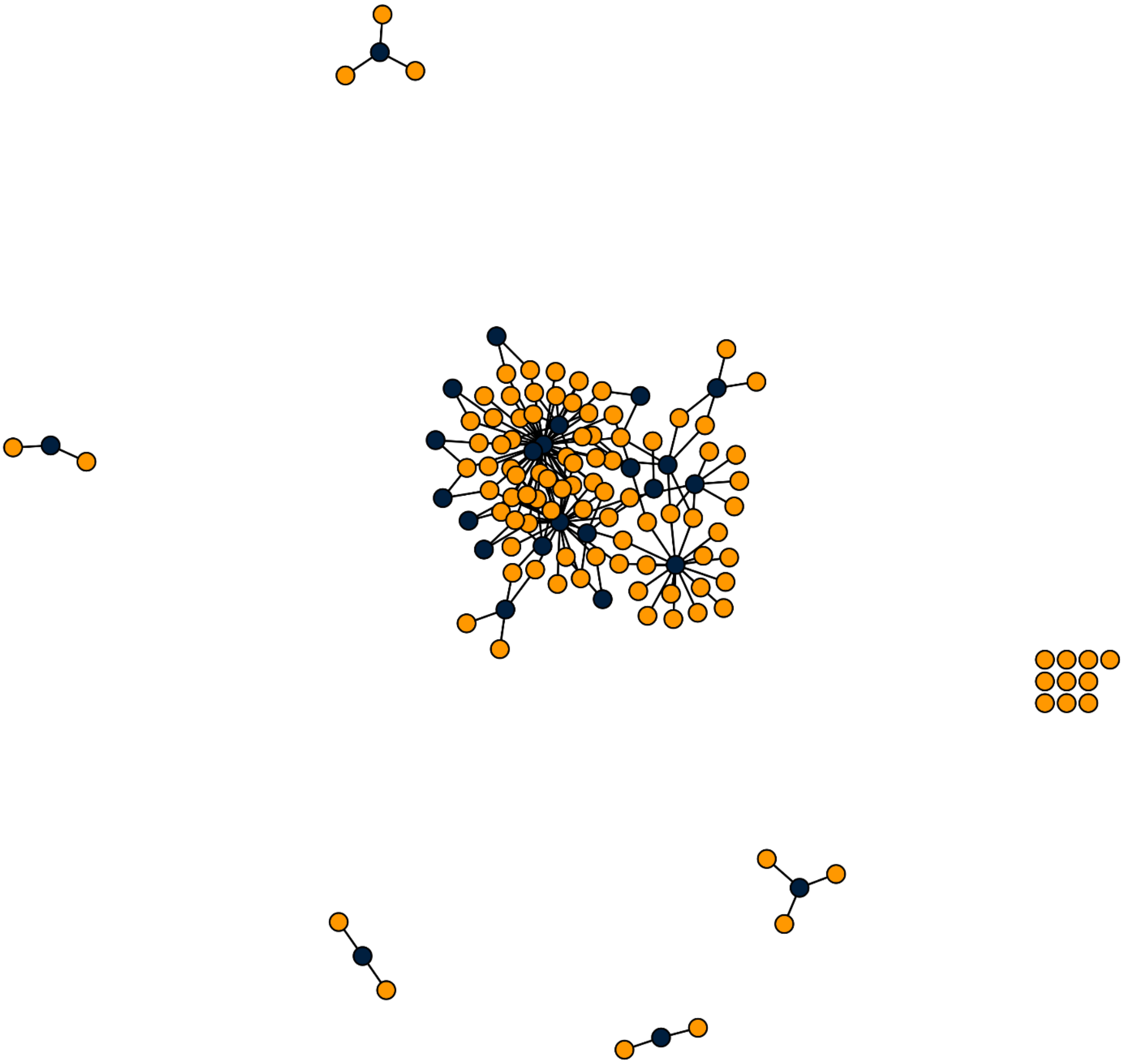


Figure 24 – Identity graph for Venice exported directly from the application

https://graphingthecity.studiole.uk/#/load/venice

	ID	Property	Label	Direct Connections	Degree Centrality	Closeness Centrality	Betweenness Centrality
1	Q16970	instance of	church	45	1.0000	1.0000	28644.56
2	Q236122	architectural style	renaissance architecture	30	0.6667	0.8702	24640.21
3	Q16560	instance of	palace	17	0.3778	0.6518	4537.97
4	Q6092662	architectural style	Italian Baroque architecture	7	0.1556	0.5408	42.90
5	Q207694	instance of	art museum	7	0.1556	0.5375	1248.56
6	Q316297	architect	Baldassarre Longhena	6	0.1333	0.5663	2493.71
7	Q176483	architectural style	Gothic architecture	5	0.1111	0.5216	478.97
8	Q177692	architect	Andrea Palladio	4	0.0889	0.5257	80.73
9	Q734053	architect	Mauro Codussi	4	0.0889	0.5090	4.67
10	Q2651004	instance of	palazzo	4	0.0889	0.4389	599.00
11	Q19899465	instance of	former church	4	0.0889	0.3694	599.00
12	Q202267	architect	Jacopo Sansovino	3	0.0667	0.5024	3.53
13	Q27686	instance of	hotel	3	0.0667	0.0476	6.00
14	Q23442	instance of	island	3	0.0667	0.0476	6.00
15	Q3766471	architect	Giovanni Antonio Scalfarotto	2	0.0444	0.4879	6.58
16	Q1347426	architect	Alessandro Vittoria	2	0.0444	0.4879	0.95
17	Q599457	architect	Antonio Abbondi	2	0.0444	0.4839	0.67
18	Q840829	architectural style	Baroque architecture	2	0.0444	0.4656	8.16
19	Q54111	architectural style	neoclassical architecture	2	0.0444	0.4368	1.67
20	Q2245489	architectural style	Italian Gothic architecture	2	0.0444	0.4328	1.00
21	Q6092863	architectural style	Italian Neoclassical architecture	2	0.0444	0.4328	1.00
22	Q811979	instance of	architectural structure	2	0.0444	0.4156	202.00
23	Q41176	instance of	building	2	0.0444	0.4135	1.00
24	Q2935533	instance of	campo	2	0.0444	0.0317	2.00
25	Q1068842	instance of	footbridge	2	0.0444	0.0317	2.00
26	Q55488	instance of	railway station	2	0.0444	0.0317	2.00

Table 3 – Social network analysis of claims for Venice

Vienna

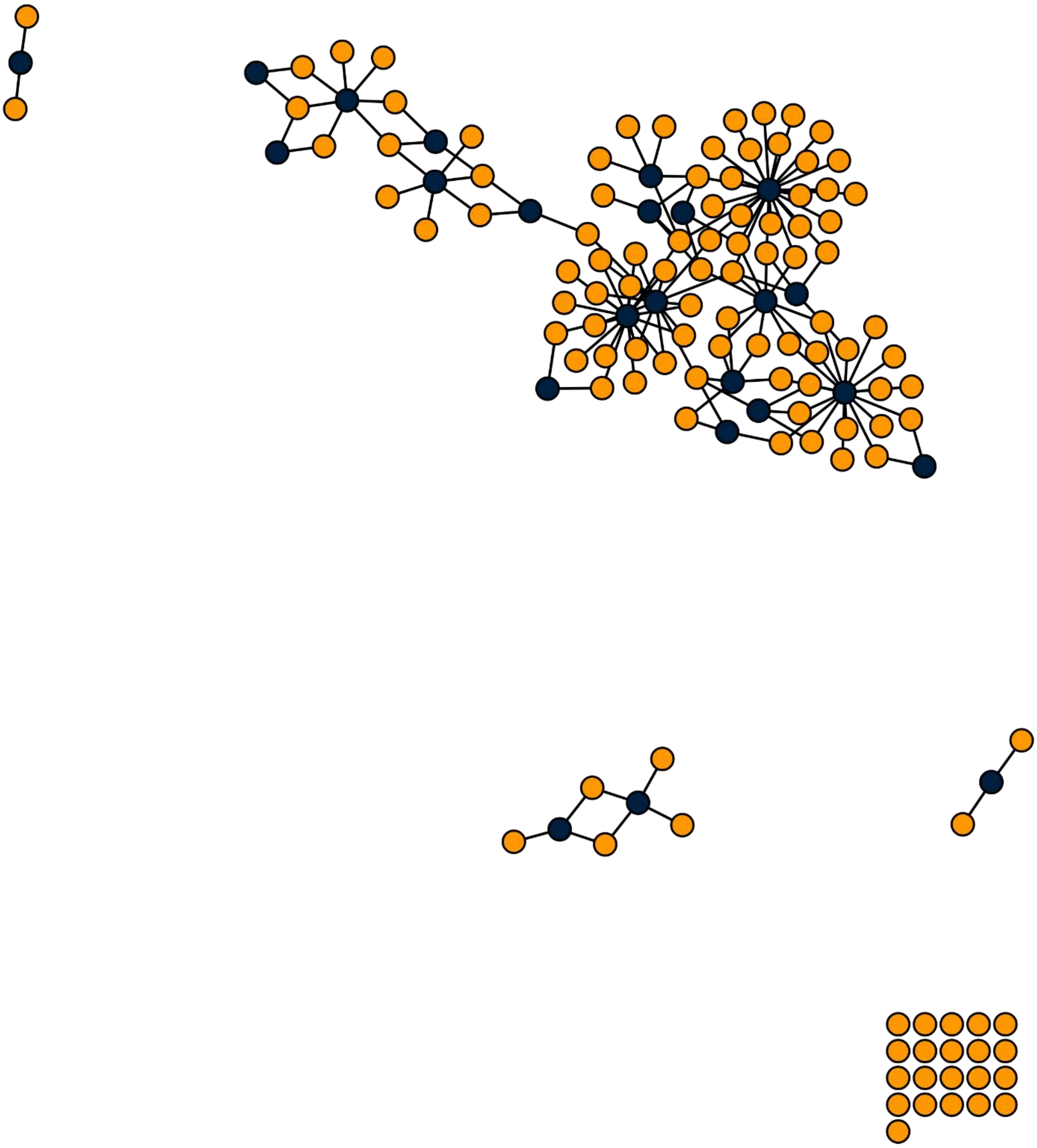


Figure 25 – Identity graph for Vienna exported directly from the application

https://graphingthecity.studiole.uk/#/load/vienna

	ID	Property	Label	Direct Connections	Degree Centrality	Closeness Centrality	Betweenness Centrality
1	Q16560	instance of	palace	26	1.0000	1.0000	8072.87
2	Q33506	instance of	museum	18	0.6923	0.8418	7294.91
3	Q16970	instance of	church	18	0.6923	0.7716	7277.24
4	Q207694	instance of	art museum	14	0.5385	0.8609	11797.75
5	Q840829	architectural style	Baroque architecture	11	0.4231	0.8007	6042.34
6	Q24354	instance of	theater	7	0.2692	0.4438	2197.17
7	Q317557	instance of	parish church	6	0.2308	0.6507	715.92
8	Q811979	instance of	architectural structure	6	0.2308	0.4900	2120.58
9	Q1200957	instance of	tourist destination	5	0.1923	0.6203	856.48
10	Q84507	architect	Johann Lukas von Hildebrandt	4	0.1538	0.6793	732.32
11	Q751876	instance of	château	4	0.1538	0.6512	439.57
12	Q176483	architectural style	Gothic architecture	4	0.1538	0.6008	872.97
13	Q174782	instance of	square	4	0.1538	0.1032	23.00
14	Q46261	architectural style	Romanesque architecture	3	0.1154	0.5865	263.74
15	Q84305	architect	Johann Bernhard Fischer von Erlach	3	0.1154	0.5815	87.67
16	Q34636	architectural style	Art Nouveau	3	0.1154	0.5477	4044.17
17	Q694854	architect	Fellner & Helmer	3	0.1154	0.4473	1306.25
18	Q79007	instance of	street	3	0.1154	0.0890	14.00
19	Q7075	instance of	library	2	0.0769	0.4547	1.00
20	Q186363	architectural style	Gothic Revival architecture	2	0.0769	0.4215	1.00
21	Q78968	architect	August Sicard von Sicardsburg	2	0.0769	0.3059	1.67
22	Q153562	instance of	opera house	2	0.0769	0.3059	1.67
23	Q1154710	instance of	football stadium	2	0.0769	0.0427	2.00
24	Q41176	instance of	building	2	0.0769	0.0427	2.00

Table 4 – Social network analysis of claims for Vienna

London

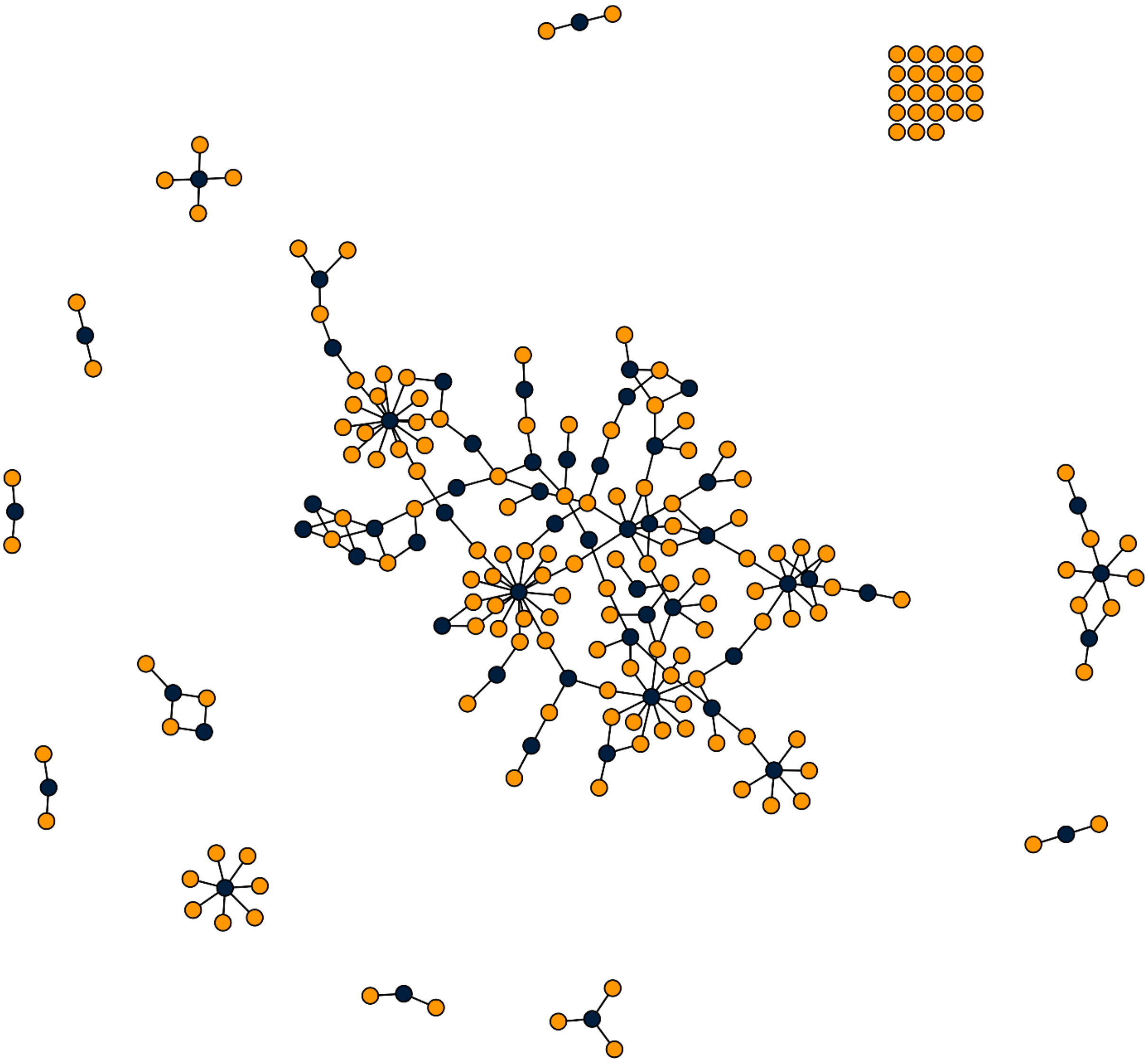


Figure 26 – Identity graph for London exported directly from the application

https://graphingthecity.studiole.uk/#/load/london

	ID	Property	Label	Direct Connections	Degree Centrality	Closeness Centrality	Betweenness Centrality
1	Q24354	instance of	theater	17	1.0000	1.0000	8367.17
2	Q27686	instance of	hotel	15	0.8824	0.8260	6034.04
3	Q811979	instance of	architectural structure	11	0.6471	0.8830	6251.48
4	Q207694	instance of	art museum	9	0.5294	0.7227	3664.85
5	Q1125300	architectural style	Georgian architecture	8	0.4706	0.9323	11858.09
6	Q3917681	instance of	embassy	7	0.4118	0.1653	42.00
7	Q41176	instance of	building	6	0.3529	0.5569	1420.00
8	Q2755753	instance of	area of London	6	0.3529	0.1810	85.00
9	Q33506	instance of	museum	5	0.2941	0.7695	3139.78
10	Q2087181	instance of	historic house museum	5	0.2941	0.7177	5225.16
11	Q173782	architectural style	Art Deco	4	0.2353	0.6746	2594.78
12	Q483110	instance of	stadium	4	0.2353	0.6633	1967.16
13	Q16970	instance of	church	4	0.2353	0.6631	1657.98
14	Q19686	crosses	River Thames	4	0.2353	0.5423	1946.83
15	Q1007870	instance of	art gallery	4	0.2353	0.5116	6.00
16	Q34442	instance of	road	4	0.2353	0.0945	12.00
17	Q543654	instance of	city hall	3	0.1765	0.7491	4457.59
18	Q1443739	architect	Frank Matcham	3	0.1765	0.7334	3100.45
19	Q1208016	architect	Foster + Partners	3	0.1765	0.6711	2584.49
20	Q16560	instance of	palace	3	0.1765	0.6403	574.00
21	Q1595639	instance of	local museum	3	0.1765	0.6379	1134.00
22	Q32815	instance of	mosque	3	0.1765	0.5700	431.00
23	Q170373	architect	Christopher Wren	3	0.1765	0.5191	408.27
24	Q537127	instance of	road bridge	3	0.1765	0.4401	8.83
25	Q55488	instance of	railway station	3	0.1765	0.4250	574.00
26	Q3957	instance of	town	3	0.1765	0.1248	26.00
27	Q928830	instance of	metro station	3	0.1765	0.0826	8.00
28	Q3918	instance of	university	3	0.1765	0.0708	6.00
29	Q1524274	architect	Giles Gilbert Scott	2	0.1176	0.7217	683.21
30	Q7789267	architect	Thomas Edward Collcutt	2	0.1176	0.6858	2817.16
31	Q845318	architectural style	high-tech architecture	2	0.1176	0.6723	2633.24
32	Q176483	architectural style	Gothic architecture	2	0.1176	0.6679	810.68
33	Q1513688	architectural style	Greek Revival architecture	2	0.1176	0.6633	321.76
34	Q245188	architectural style	modern architecture	2	0.1176	0.6418	1159.09
35	Q104898	architect	Norman Foster	2	0.1176	0.6345	2448.00
36	Q1060829	instance of	concert hall	2	0.1176	0.5879	288.00
37	Q5079376	architect	Charles J. Phipps	2	0.1176	0.5847	1.00
38	Q1082849	architect	Populous	2	0.1176	0.5842	1508.45
39	Q11303	instance of	skyscraper	2	0.1176	0.5632	67.33

40	Q1154710	instance of	football stadium	2	0.1176	0.5252	288.00
41	Q2977	instance of	cathedral	2	0.1176	0.5249	294.35
42	Q681007	architect	George Gilbert Scott	2	0.1176	0.5218	1128.00
43	Q622425	instance of	nightclub	2	0.1176	0.5205	288.00
44	Q2574110	architectural style	English Baroque	2	0.1176	0.5034	72.27
45	Q12570	instance of	suspension bridge	2	0.1176	0.4990	207.00
46	Q1497364	instance of	building complex	2	0.1176	0.4900	288.00
47	Q3469910	instance of	performing arts venue	2	0.1176	0.4833	288.00
48	Q2516357	instance of	transport museum	2	0.1176	0.4700	288.00
49	Q158438	instance of	arch bridge	2	0.1176	0.4185	0.50
50	Q3397526	instance of	stone bridge	2	0.1176	0.4185	0.50
51	Q4989906	instance of	monument	2	0.1176	0.1090	18.00
52	Q14562709	instance of	London Underground station	2	0.1176	0.0669	1.00
53	Q1059324	instance of	teaching hospital	2	0.1176	0.0472	2.00
54	Q17350442	instance of	venue	2	0.1176	0.0472	2.00
55	Q1378975	instance of	convention center	2	0.1176	0.0472	2.00
56	Q17431399	instance of	national museum	2	0.1176	0.0472	2.00
57	Q875538	instance of	public university	2	0.1176	0.0472	2.00
58	Q43501	instance of	zoo	2	0.1176	0.0472	2.00

Table 5 – Social network analysis of claims for London

Paris

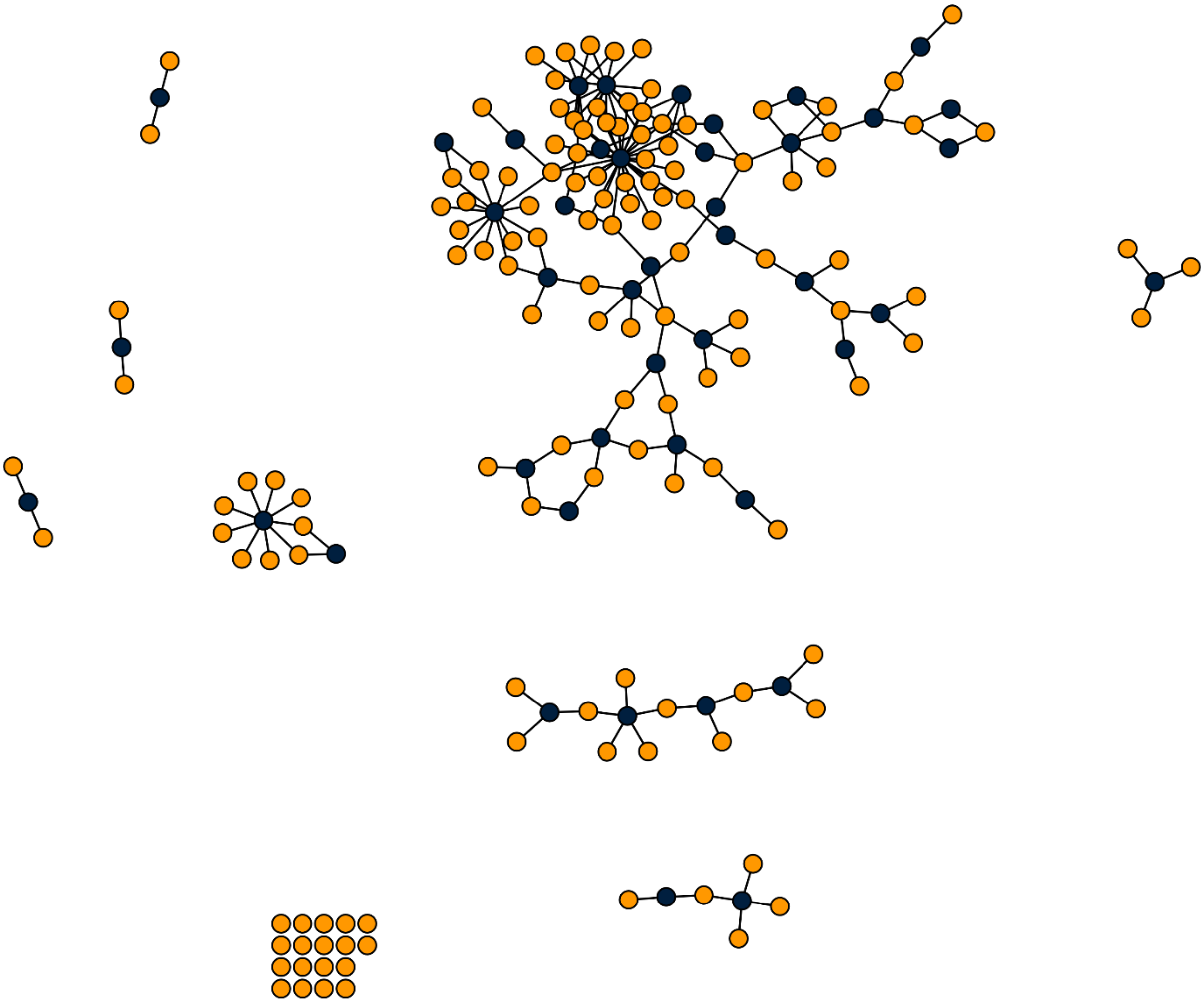


Figure 27 – Identity graph for Paris exported directly from the application

https://graphingthecity.studiole.uk/#/load/paris

	ID	Property	Label	Direct Connections	Degree Centrality	Closeness Centrality	Betweenness Centrality
1	Q33506	instance of	museum	29	1.0000	1.0000	13090.21
2	Q207694	instance of	art museum	16	0.5517	0.7331	4649.78
3	Q24354	instance of	theater	13	0.4483	0.7053	3769.44
4	Q11303	instance of	skyscraper	9	0.3103	0.1864	92.00
5	Q17431399	instance of	national museum	7	0.2414	0.5829	767.40
6	Q16560	instance of	palace	6	0.2069	0.5191	3073.00
7	Q27686	instance of	hotel	5	0.1724	0.5436	2082.04
8	Q22746	instance of	urban park	5	0.1724	0.1566	128.00
9	Q3330843	instance of	musée de la ville de Paris	4	0.1379	0.5687	65.36
10	Q173782	architectural style	Art Deco	4	0.1379	0.5018	1244.64
11	Q1365179	instance of	hôtel particulier	4	0.1379	0.4810	672.00
12	Q120560	instance of	minor basilica	4	0.1379	0.4200	1522.00
13	Q16970	instance of	church	4	0.1379	0.4125	999.00
14	Q55488	instance of	railway station	4	0.1379	0.0948	24.00
15	Q840829	architectural style	Baroque architecture	3	0.1034	0.4990	3083.00
16	Q1060829	instance of	concert hall	3	0.1034	0.4084	1510.00
17	Q19860854	instance of	former building or structure	3	0.1034	0.4017	1522.00
18	Q652952	architect	Pierre-François-Léonard Fontaine	3	0.1034	0.3818	19.00
19	Q483110	instance of	stadium	3	0.1034	0.3315	450.00
20	Q2977	instance of	cathedral	3	0.1034	0.3283	339.00
21	Q252916	instance of	administrative quarter	3	0.1034	0.1305	98.00
22	Q174782	instance of	square	3	0.1034	0.1165	50.00
23	Q123705	instance of	neighborhood	3	0.1034	0.1112	50.00
24	Q7543083	instance of	avenue	3	0.1034	0.0589	6.00
25	Q1268134	architectural style	Beaux-Arts architecture	2	0.0690	0.5918	360.62
26	Q1064558	architect	Charles Girault	2	0.0690	0.5822	1182.35
27	Q18674739	instance of	event venue	2	0.0690	0.5822	1182.35
28	Q8053	architect	Jules Hardouin Mansart	2	0.0690	0.5778	2792.74
29	Q153562	instance of	opera house	2	0.0690	0.5414	226.00
30	Q2772772	instance of	military museum	2	0.0690	0.5373	63.99
31	Q214317	architect	Jean Nouvel	2	0.0690	0.5188	1890.00
32	Q34636	architectural style	Art Nouveau	2	0.0690	0.4829	1074.30
33	Q979605	architect	Gabriel Davioud	2	0.0690	0.4107	1.00
34	Q641226	instance of	arena	2	0.0690	0.3185	226.00
35	Q186363	architectural style	Gothic Revival architecture	2	0.0690	0.3152	113.00
36	Q40357	instance of	prison	2	0.0690	0.3056	112.00
37	Q23413	instance of	castle	2	0.0690	0.3056	112.00
38	Q54111	architectural style	neoclassical architecture	2	0.0690	0.3049	226.00
39	Q149621	instance of	district	2	0.0690	0.3007	226.00

40	Q3384840	architect	Pierre Dufau	2	0.0690	0.0948	1.00
41	Q643621	architect	Jacques Ignace Hittorff	2	0.0690	0.0687	10.00
42	Q7075	instance of	library	2	0.0690	0.0392	2.00
43	Q16917	instance of	hospital	2	0.0690	0.0392	2.00
44	Q176483	architectural style	Gothic architecture	2	0.0690	0.0392	2.00

Table 6 – Social network analysis of claims for Paris

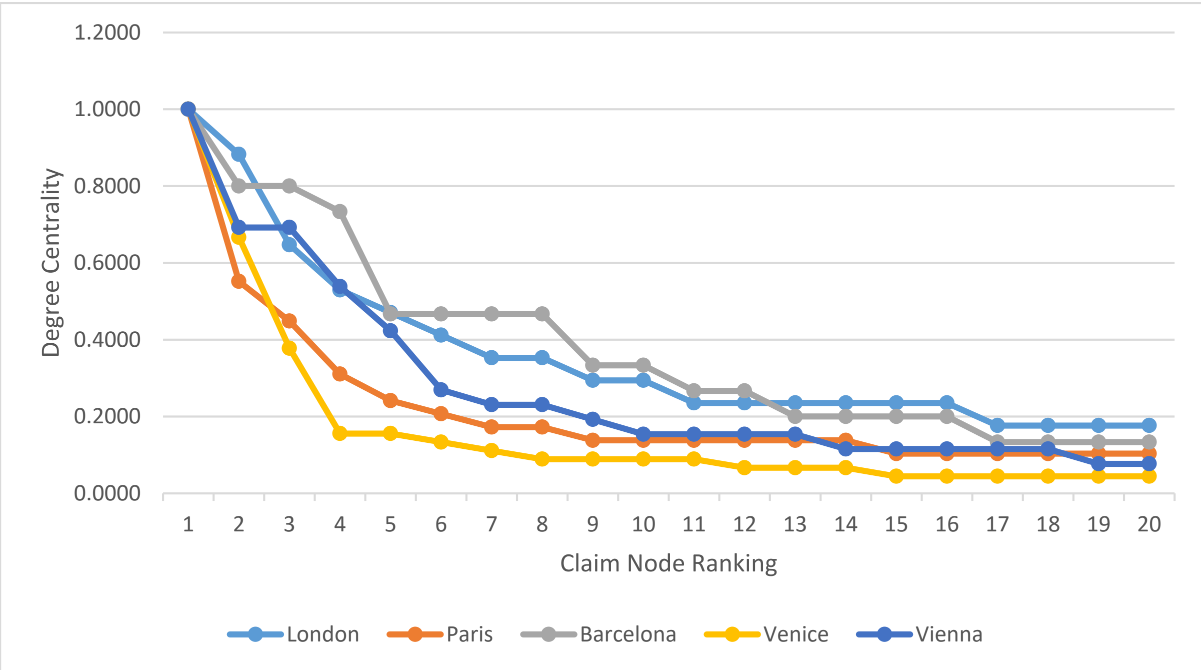


Figure 28 – Chart of degree centrality against ranking

06 DISCUSSION

Findings

Table 1 suggests there is a correlation between increasing visitor numbers and the number of nodes & connections on the *identity graph*, however, this correlation is less substantial when we consider only the number of place nodes. While there is still a correlation between increased visitor numbers and increasing numbers of place nodes it is less pronounced. There is, however, a clear correlation between increasing numbers of claim nodes and increased visitor numbers, starting from Venice with 5.2 million visitors and 26 claim nodes to London with 17.4 million visitors and 58 claim nodes. There are two potential hypotheses for this correlation, first, the more visitors a city has the more likely people are to have recorded claims on Wikidata. Second, the greater the number of people within a city, the more diverse the identity of the city. That is to say that an increase in the number of claim nodes suggests there must be more architectural styles, architects, or place classifications within the graph so the identity is more dilute.

The concept of the more dilute identity is particularly of note to London (Figure 26) and Paris (Figure 27), a quick glance at their graphs shows a much wider spread of nodes than for those of the three cities with lower visitor numbers (Figure 23 - Figure 25). This could be due to a number of reasons, first, as suggested previously it could simply be that Wikidata claims are more likely to be apparent in a larger city, or, second, it could possibly be due to a more diverse level of culture within London and Paris as Zhou et al. (2014) suggest large metropolises typically mix “cultures from all over the world”.

As discussed within the methodology social network analysis has been used in order to analyse the *identity graphs* for each city with expansive definitions of the functions explained within the background. As a brief summary, degree centrality is a measure of the number of direct connections a node has, closeness centrality indicates the inverse sum of the distance from a node to all other nodes in a network and betweenness centrality is assigned based upon the importance of a node to keeping the network connected. By using the normalised functions for degree centrality and closeness centrality we are able to compare the figures between the graphs despite the variation in the quantity of nodes. Figure 28 shows a comparison between the degree centrality for the first 20 highest ranked claim nodes of each city, as normalised degree centrality is presented as a proportion of the number of direct connections of the most connected node the first ranked element will always equal 1.000.

Venice

Venice, in particular, has a strongly focused graph with high connectivity between the place nodes. This is shown quantitatively within the social network analysis where *churches* and *renaissance architecture* have comparatively high number of connections, 45 and 30 respectively, and the number of connections – indicated by degree centrality – sharply dropping for the subsequent claim nodes as shown in Figure 28. *Churches* and *renaissance architecture* also have high betweenness centrality – 28644.56 and 24640.21 respectively from these statistics we can easily conclude that these claims are integral to the identity of Venice. Similarly, we can see that *palaces* and *art museums* are relevant to the architectural identity of Venice, however, they do not stand out to the high degree of *churches* and *renaissance architecture*.

Vienna

The classification of the places of interest within Vienna bear significant similarity to those of Venice. *Palace*, *museum*, *church* and *art museum* are the top four results for degree centrality and each holds high values for betweenness. While Venice has significantly fewer *museums* the other three values have comparable numbers of connections. It should be noted at this stage that typically an *art museum* will be classified on Wikidata as being both an *art museum* and a *museum*, however, this doesn't appear to be the case for the *art museums* in Venice hence the absence of *museums* within the data for Venice.

From the SNA data within the application for Vienna, *Baroque* appears to be the most significant architectural style in Vienna with a degree centrality of 0.4231 followed by *Gothic* (0.1538) and *Art Nouveau* (0.1154). *Johann Lukas von Hildebrandt* appears as the most significant architect, having designed three palaces and a church.

Barcelona

Disappointingly, the highest ranked claim by degree centrality in the graph for Barcelona is *building* while this is technically correct this is not something that we would consider an aspect of identity. The following claims are however considerably more culturally descriptive. *Art Nouveau*, *Modernisme*, and the classification *skyscraper* have 12, 11, and 12 connections respectively, indicating themselves as significant to the identity of Barcelona. Antoni Gaudi – with 7 connections and a degree centrality of 0.4667 – is the most significant architect on the city.

Barcelona is home to a significant number of UNESCO World Heritage Sites (UNESCO, n.d.) including Park Güell, Palau Güell, Casa Milà, Casa Vicens, Sagrada Família, Casa Batlló, Palau de la Música Catalana, and

Hospital de Sant Pau. These honours have been awarded based upon their historical or cultural significance and thus should be considered integral to the identity of the city. Therefore, the fact we can see these places so integrated within the identity graph adds credibility to this methodology.

Disconnected Graphs

Barcelona also indicates a few place classifications that are culturally important but do not integrate into the body of the graph. Both *square* and *mountain* have a degree centrality of 0.2000 but their low betweenness centrality of 6.00 indicates that they are disconnected from the rest of the identity graph. The same disconnect can be seen in the graphs of Paris and Vienna, although within their graphs the *square* claim is slightly more connected this strain is still independent of the major chain of the *identity graph*.

In the same way, bridges typically aren't connected into the larger part of the identity graph however for London the *Millennium Bridge* is connected by the shared architect – *Norman Foster* – to *City Hall* which in turn is connected into the major part of the graph. This is reflected quantitatively in the high betweenness centrality of the *Norman Foster* node compared to other nodes of the same degree centrality.

London

As it stands at present, in London the architects *Frank Matcham*, *Foster + Partners,* and *Christopher Wren* equally share the highest rank for most significant *architect*, however, they do so with a significantly lower degree centrality than *Antoni Gaudi* has in Barcelona – 0.1765 compared to 0.4667. The highest architectural style within London is *Georgian* (0.4706), and *theater* is the highest ranking entity with a total of 17 connections.

The classification of *embassy* ranks highly in the graph of London (0.4118) but as indicated by the figure for betweenness these entities are completely disconnected from the rest of the graph. This feels counter intuitive as typically embassies are grand buildings, many of which – within London at least – are listed buildings so one would expect for them to have defined architectural styles integrating them into the major part of the graph.

Paris

Paris is of interest as it is the only city in this study with no significantly ranked *architectural style*. *Art Deco* is the highest but with only 4 connections it ranks in joint 9th place when listed by degree centrality with a value of 0.1379.

Similar to London we see a high number of *museum*, *art museum*, *theater* and *hotel* within the results, they also both exhibit similar degree centrality for r*ailway stations* that are included as places of interest – London: 0.1765, Paris: 0.1379. It is outside the scope of this paper to begin to suggest whether these could be considered elements of the genotype of a global city but certainly it would be an interesting topic for future research.

Limitations

Claim Consistency

Within the graph for London, there are three places connected to *Foster + Partners* by *architect* and two different places connected to *Norman Foster*, this is a clear discrepancy between the way Wikidata entities are linked. Either all entities should be connected to the firm *Foster + Partners* – which in turn is connected to *Norman Foster* through the *Founder* property – or each place should be directly linked by *architect* to both *Norman Foster* himself and the firm *Foster + Partners*. If all five places were connected to the firm, we would see a significantly higher degree of connectivity within the graph and then it could be objectively suggested that Foster was the most significant architect with regards to *places of interest* in London.

Inaccuracy and Potential Subjectivity of Wikidata

Skyscrapers are another highly ranked entity within Paris with 9 connections; the results are close to those for Barcelona which has 12 skyscrapers; by comparison, London is significantly lower with only 2. Intuitively, these statistics do not reflect the true nature of the cities, while both London and Paris have significantly lower skylines than American cities skyscrapers, by the Emporis Standards Committee definition of any building greater than 100 m we would expect to see significantly more skyscrapers in evidence. Certainly, the skyscraper is becoming an important element of the identity of London in both Canary Wharf and the City of London.

This could be evidence of a level of subjectivity present on Wikidata, as Lennard Crowhurst (1995) notes London Docklands lacks a “coherent urban fabric” and subsequently are not a part of London’s identity. Therefore, the absence of places of interest within the Docklands area could be due to the subjectivity of those who have submitted data. If this is the case, then rather than it being a weakness it could act as a positive filtering mechanism. The data that is recorded on Wikidata is what the contributors consider important, and thus intrinsically reflects the identity of the city.

Lack of Nuance

As reported by Ford & Graham (2016), there can be a tendency for structured data to lack the nuance that is contained within unstructured articles. Take for example *King's Cross station:* Wikidata defines the architect as *Lewis Cubitt*, however, this lacks the nuance that the station was heavily refurbished in 2012 with a new concourse designed by John McAslan. Subsequently, only the architect Lewis Cubitt is recorded within the *identity graph* and not the renovation works by McAslan.

Figure 29 – Left: Extract from Wikidata entry for King's Crossing station (Wikidata, n.d.-d), Right: Reichstag building (Wikidata, n.d.-e)

Of course, this could be due to the subjective bias discussed earlier, McAslan's contribution to King's Cross may in time be added to Wikidata. As shown in the comparison in Figure 29, Norman Foster's contribution to the Reichstag Building is included with a qualifier alongside the original architect Paul Wallot and thus would be evident on the *identity graph* of Berlin.

Insufficient Data

In its present evolution, there are a number of limitations to the application, particularly with regard to the availability of claim data. For the 162 places of interest within London, only 55 have an *architect* defined and 25 have an *architectural style*, the averages across the five cities are slightly higher with 35% having an *architect* and 25% an *architectural style (*Table 7*).*

	P31	P84	P149
	Instance of	Architect	Architectural Style
Venice	96%	32%	49%
Vienna	96%	28%	21%
Barcelona	99%	41%	31%
Paris	98%	39%	19%
London	96%	34%	15%
	97%	35%	25%

Table 7 – Percentage of places with claim data defined for each property

As Wikidata is an open and ongoing project, the availability and quality of data will gradually improve over time as more contributions are added and existing data is refined.

The initial mechanisms for populating the Wikidata database was through automated mining of data from Wikipedia articles (Vrandečić & Krötzsch, 2014) and extracting data from projects such as DBpedia, Freebase or Yago (Bizer et al., 2009; Hoffart et al., 2013). Additional data has later been added by voluntary contributors and by automated scripts as can be seen by viewing the revision history of any item (Wikidata, n.d.-f).

Improving the data

Improving Wikidata Claims

In order to expedite the rate at which relevant architectural identity claims are added to Wikidata, an approach similar to the reCAPTCHA system by Von Ahn et al. (2008) could be implemented.

CAPTCHA is a method of preventing bots from submitting web forms, by asking the user to type the words or characters in an image. The reCAPTCHA system (Figure 30) takes this a step further, by asking the user to type two words, one of which is already known by the system, the other is a word that an Optical Character Recognition (OCR) program has failed to interpret. Thus, during its lifetime reCAPTCHA has been able to improve the AI behind OCR and result in the direct digitisation of millions of books and articles for Google Books (Google reCAPTCHA, n.d.).

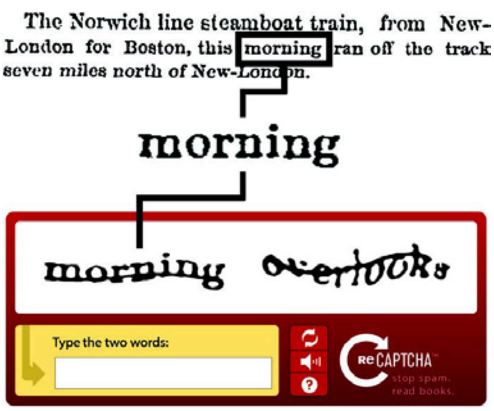


Figure 30 – reCAPTCHA (Von Ahn et al., 2008).

A similar methodology could be applied to the categorisation of *architectural style* for Wikidata, by presenting the user with a picture of a building and asking them to identify the architectural style – or perhaps more thoroughly asking them what architectural detailing is present – a dataset of subjective opinions can be collected. The architectural style of a building will then be objectively evident where a majority of users have consistently defined the image as a particular style. By including some images for which the architectural style is already known it will be possible to test the accuracy of the participants. This methodology would quickly improve the quantity of available data while maintaining a consistent academic integrity.

If each architecture school were to have students use the proposed architectural style application to teach their students variations in architectural styles and test their students' recognition ability to recognise the styles of local buildings, we could quickly improve the availability of Wikidata for cities.

An alternative approach would be to extract existing information stored within academic databases such as University of Washington (n.d.) Cities and Buildings Database which associates photos of buildings with their construction materials and architectural styles. Or, Doersch et al.'s (2012) algorithm could be applied to automatically identify the presence of architectural elements in geotagged photos.

At present, there is also a limitation in the size of city that can be analysed. If we take *Newcastle upon Tyne* as an example, at the time of writing there are 44 places of interest on the graph but only 29 connections, resulting in a significant number of disconnected strains. By applying one – or all – of these three methodologies to improve the availability of data for the city we would find a more complete *identity graph*.

Proposing Additional Wikidata Claims

As Crowhurst Lennard & Lennard (1995) discuss the identity of a city can be found within not just the style but the materiality, detailing, form, scale, and colour. Doersch et al. (2012) build upon this by identifying traits in geo-cultural architectural elements such as balconies and window detailing. Therefore, by connecting nodes only by *architectural style* we are in fact missing out on a lot of nuance of architectural identity. To improve the intricacy of connections it could be interesting to propose *materiality* and *architectural detail* as new Wikidata properties. These would allow us to connect buildings which share materials but not a particular *architectural style,* and the *architectural detail* properties could begin to pick out some of the subtleties of building detailing, such as the classical order of columns which at present are not noted within the structured data.

Further Research

Graph Matching

Dalton & Kirsan (2008) proposed a new method of deriving the genotype signature of buildings through graph matching. The same methodology could be applied to the urban *identity graph* in order to identify the genotype signature of the city within a sample. The samples for particular regions could then be quantifiably compared to illustrate the distinct identities of particular regions.

Additional Quantifiable Metrics

Within this paper, social network analysis has been used as a quantitative mathematical method but additional tools are available and could be applied in order to gain additional insights. These include the rank order of integration values (Hillier, 1987) and real relative asymmetry (RRA) (Bafna, 2003) which allows for the comparison of graphs of differing numbers of nodes.

Correlations with Wellbeing and Liveability

There is scope to perform further research looking for correlations between the *identity graph* and measures of livability, wellbeing, quality of life and intelligibility in cities. Such correlations have already been raised between the imageability and intelligibility (Dalton & Bafna, 2003), organisation (Gehl, 1987; Hillier, 1996; Salingaros, 1998), identity and quality of life (Doratli & Fasli, 2011), identity and satisfaction of residents and tourists (Bigne et al., 2001; Petrick, 2004; Chen et al., 2007; Chi et al., 2008), and identity and liveability Crowhurst Lennard & Lennard (1995).

Applications in Other Fields

The methodology used within this paper, and the subsequent application that has been developed could very easily be applied to other fields. By simply adjusting the connection criteria defined within the application we can easily gain insight into urban metrics other than identity and, likewise, if the city criteria were to be replaced entirely future researchers could identify relationships between an entirely different typology – such as people – without too much modification to the existing application.

Real World Applications

There are a number of real world applications for the *identity graph* particularly in architecture, urban planning, and urban design.

As Crowhurst Lennard & Lennard (1995) note, as part of the design process an architect must be able to quickly evaluate the context of the city, and the *identity graph* may provide a simple method of doing so: by providing immediate, objective feedback on which places and elements currently contribute to a city's identity.

The model may also be used in the planning, process, a method could be derived to manually add new nodes to the graph, this would allow architects and planners to test how their proposed buildings integrate into the graph, it can then be objectively determined whether a building positively contributes to the identity of the city or whether it negatively detracts, diluting the city's identity.

07 CONCLUSION

The intention of this paper was to develop a computational method of urban analysis in order to extract the DNA of the city as a graph through the analysis of structured data. Following a thorough literature review – investigating the metaphor of DNA and genotypes in space syntax research, the concept of urban identity, and the use of volunteered geographic information and structured data in research – an application that connects *places of interest* by elements of architectural identity in order to objectively derive and visualise an *identity graph* for that city has been developed.

Similar to the work of Dalton & Kirsan (2008) the identity or DNA in this instance is formulated as a representative *identity graph* a "conceptual shift" (Dalton & Kirsan, 2008) from the *inequality genotype* indicated by Hillier and Hanson (1984) in the rank order of integration values.

The use of publicly accessible structured data APIs means that this application can be intuitively used by anyone to analyse any city within seconds and the use of social network analysis provides quantitative methods for analysing the graph data. Finally, the open source release of this application and public hosting of the application (Elsdon, 2016b) will make the methodology easily accessible for future researchers to experiment, adapt and evolve.

The limitations of the application – such as the availability of data in some areas – have been discussed and solutions to these issues have been proposed – including three methods by which the architectural identity data stored on Wikidata could be improved and expanded. Finally, avenues of future research that have been opened by the development of the *identity graph* have been proposed, most notably, the potential use of Dalton & Kirsan's (2008) graph matching methods to compare the similarity of two cities and to determine the *genotype signature* or *prototype* of a sample of cities.

Running the Application with Unrestricted Properties

It was also discussed how the application and graph methodology can be applied in other fields to gain insight into metrics other than identity. So, in closing an example is included for London where the properties are not restricted to only the architectural identity metrics – architect, architectural style and type – and instead the place nodes are connected by all Wikidata properties.

	Nodes	Connections	Place Nodes	Claim Nodes
London (All Properties)	265	659	161	104

Table 8 – Node and connection count for the identity graph

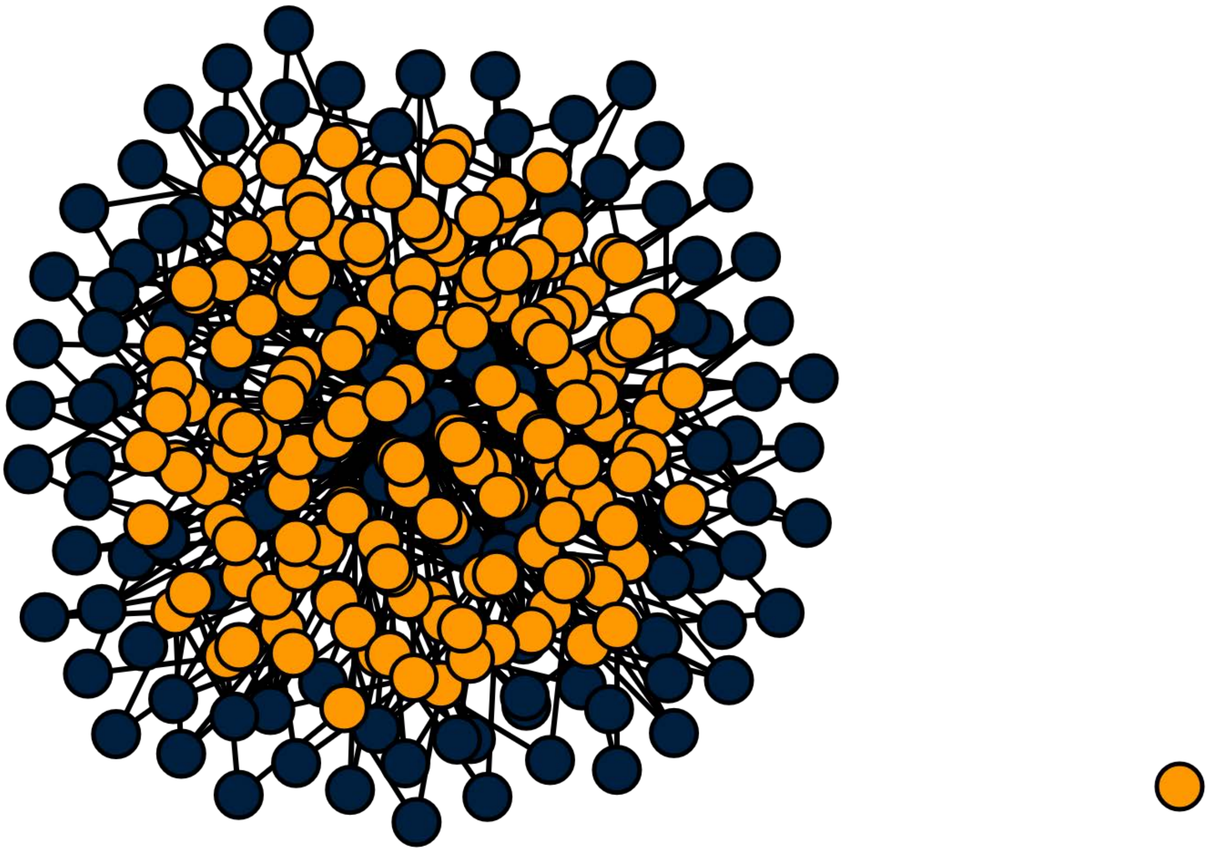


Figure 31 – Identity graph for London with all Properties enabled exported directly from the application

				Direct Connections	Degree Centrality	Closeness Centrality	Betweenness Centrality
1	Q145	country	United Kingdom	160	1.0000	1.0000	132884.69
2	Q179351	located in the administrative territorial entity	City of Westminster	53	0.3312	0.5952	4875.52
3	Q15700818	heritage status	Grade I listed building	33	0.2062	0.5462	2934.17
4	Q15700834	heritage status	Grade II listed building	25	0.1563	0.4870	666.51
5	Q15700831	heritage status	Grade II* listed building	19	0.1187	0.4763	573.36

6	Q84	located in the administrative territorial entity / location / headquarters location	London	17	0.1062	0.4550	406.28
7	Q24354	instance of	theater	17	0.1062	0.4439	243.82
8	Q23311	located in the administrative territorial entity	City of London	16	0.1000	0.4629	512.30
9	Q27686	instance of	hotel	15	0.0938	0.4352	181.67
10	Q202088	located in the administrative territorial entity	London Borough of Camden	13	0.0813	0.4372	231.61
11	Q811979	instance of	architectural structure	11	0.0688	0.4309	148.57
12	Q730706	located in the administrative territorial entity	London Borough of Southwark	9	0.0563	0.4293	186.67
13	Q207694	instance of	art museum	9	0.0563	0.4186	75.25
14	Q19686	located next to body of water / crosses	River Thames	8	0.0500	0.4333	232.30
15	Q1125300	architectural style	Georgian architecture	8	0.0500	0.4190	70.38
16	Q3917681	instance of	embassy	7	0.0437	0.4017	30.63
17	Q188801	located in the administrative territorial entity	Royal Borough of Kensington and Chelsea	6	0.0375	0.4103	56.13
18	Q208152	located in the administrative territorial entity	London Borough of Tower Hamlets	6	0.0375	0.4068	42.02
19	Q41176	instance of	building	6	0.0375	0.4056	31.59
20	Q2755753	instance of	area of London	6	0.0375	0.4044	34.77
21	Q202059	located in the administrative territorial entity	London Borough of Lambeth	5	0.0313	0.4048	36.98
22	Q33506	instance of	museum	5	0.0313	0.4048	27.93
23	Q19861084	member of	ORCID, Inc.	5	0.0313	0.4013	29.97
24	Q2087181	instance of	historic house museum	5	0.0313	0.4001	14.49
25	Q20075	part of	London Underground	4	0.0250	0.5545	4.10
26	Q483110	instance of	stadium	4	0.0250	0.3993	23.90
27	Q173782	architectural style	Art Deco	4	0.0250	0.3993	14.69
28	Q557141	genre	public art	4	0.0250	0.3981	16.84
29	Q16970	instance of	church	4	0.0250	0.3981	10.75
30	Q34442	instance of	road	4	0.0250	0.3934	6.49
31	Q1007870	instance of	art gallery	4	0.0250	0.3934	5.75
32	Q543654	instance of	city hall	3	0.0187	0.4044	30.59
33	Q385378	significant event	construction	3	0.0187	0.4032	25.47
34	Q82337	material used	Portland limestone	3	0.0187	0.4009	17.67
35	Q716737	structural engineer	Arup	3	0.0187	0.3985	18.92
36	Q537127	instance of	road bridge	3	0.0187	0.3985	12.47
37	Q1208016	architect	Foster + Partners	3	0.0187	0.3973	23.90
38	Q14946379	diocese	Diocese of London	3	0.0187	0.3961	8.20
39	Q170373	architect	Christopher Wren	3	0.0187	0.3949	7.45

40	Q32515	located in the administrative territorial entity	London Borough of Richmond upon Thames	3	0.0187	0.3938	11.80
41	Q2736	sport	association football	3	0.0187	0.3938	11.08
42	Q1595639	instance of	local museum	3	0.0187	0.3938	7.64
43	Q16560	instance of	palace	3	0.0187	0.3926	5.94
44	Q3957	instance of	town	3	0.0187	0.3914	10.08
45	Q4741072	owned by	Ambassador Theatre Group	3	0.0187	0.3914	5.18
46	Q590631	member of	Russell Group	3	0.0187	0.3914	4.78
47	Q928830	instance of	metro station	3	0.0187	0.3910	4.10
48	Q3918	instance of	university	3	0.0187	0.3902	5.17
49	Q1443739	architect	Frank Matcham	3	0.0187	0.3902	3.42
50	Q40478	located in the administrative territorial entity	London Borough of Hammersmith and Fulham	3	0.0187	0.3890	6.52
51	Q55488	instance of	railway station	3	0.0187	0.3890	3.80
52	Q32815	instance of	mosque	3	0.0187	0.3878	3.67
53	Q124184	location	Mayfair	3	0.0187	0.3867	2.21
54	Q12570	instance of	suspension bridge	2	0.0125	0.3930	4.92
55	Q104898	architect	Norman Foster	2	0.0125	0.3930	4.81
56	Q2977	instance of	cathedral	2	0.0125	0.3918	4.34
57	Q5284447	main building contractor	Mace	2	0.0125	0.3918	4.03
58	Q845318	architectural style	high-tech architecture	2	0.0125	0.3918	4.03
59	Q1093862	maintained by	Bridge House Estates	2	0.0125	0.3918	3.26
60	Q5880816	main building contractor	Holland, Hannen & Cubitts	2	0.0125	0.3918	2.75
61	Q739384	owned by	London County Council	2	0.0125	0.3918	2.75
62	Q152245	named after	Albert, Prince Consort	2	0.0125	0.3906	3.63
63	Q4989906	instance of	monument	2	0.0125	0.3906	3.43
64	Q1513688	architectural style	Greek Revival architecture	2	0.0125	0.3906	1.71
65	Q158438	instance of	arch bridge	2	0.0125	0.3906	1.52
66	Q3397526	instance of	stone bridge	2	0.0125	0.3906	1.52
67	Q9439	named after	Queen Victoria	2	0.0125	0.3902	4.37
68	Q1524274	architect	Giles Gilbert Scott	2	0.0125	0.3894	3.70
69	Q176483	architectural style	Gothic architecture	2	0.0125	0.3894	3.61
70	Q245188	architectural style	modern architecture	2	0.0125	0.3894	2.59
71	Q219538	heritage status	scheduled monument	2	0.0125	0.3894	2.41
72	Q2574110	architectural style	English Baroque	2	0.0125	0.3894	1.09
73	Q1082849	architect	Populous	2	0.0125	0.3882	4.80
74	Q11303	instance of	skyscraper	2	0.0125	0.3882	2.33
75	Q1060829	instance of	concert hall	2	0.0125	0.3882	2.24
76	Q3469910	instance of	performing arts venue	2	0.0125	0.3882	1.82
77	Q40861	material used	marble	2	0.0125	0.3882	1.32
78	Q481289	use	official residence	2	0.0125	0.3882	0.79
79	Q1154710	instance of	football stadium	2	0.0125	0.3870	4.39
80	Q693450	located in the administrative territorial entity	Royal Borough of Greenwich	2	0.0125	0.3870	4.36
81	Q6574	located in time zone	UTC±00:00	2	0.0125	0.3870	4.09

82	Q2928594	structural engineer	BuroHappold Engineering	2	0.0125	0.3870	4.06
83	Q5849	sport	rugby union	2	0.0125	0.3870	3.39
84	Q875538	instance of	public university	2	0.0125	0.3870	2.83
85	Q10344038	part of	Palace of Westminster and Westminster Abbey including Saint Margaret's Church	2	0.0125	0.3870	1.71
86	Q43501	instance of	zoo	2	0.0125	0.3870	1.54
87	Q170027	part of	University of London	2	0.0125	0.3870	1.48
88	Q6635653	part of	list of public art in the City of Westminster	2	0.0125	0.3870	1.16
89	Q7789267	architect	Thomas Edward Collcutt	2	0.0125	0.3870	0.91
90	Q14562709	instance of	London Underground station	2	0.0125	0.3867	0.73
91	Q30192	located in time zone	Greenwich Mean Time	2	0.0125	0.3859	3.74
92	Q681007	architect	George Gilbert Scott	2	0.0125	0.3859	1.88
93	Q2516357	instance of	transport museum	2	0.0125	0.3859	1.88
94	Q55019	shares border with	Covent Garden	2	0.0125	0.3859	1.35
95	Q3308692	located on street	Euston Road	2	0.0125	0.3859	1.32
96	Q5079376	architect	Charles J. Phipps	2	0.0125	0.3859	0.60
97	Q1378975	instance of	convention center	2	0.0125	0.3847	3.52
98	Q8577	occupant	2012 Summer Olympics	2	0.0125	0.3847	1.84
99	Q1497364	instance of	building complex	2	0.0125	0.3847	1.80
100	Q1059324	instance of	teaching hospital	2	0.0125	0.3847	1.29
101	Q151048	located in the administrative territorial entity	London Borough of Barnet	2	0.0125	0.3835	1.79
102	Q208139	located in the administrative territorial entity	London Borough of Newham	2	0.0125	0.3835	1.76
103	Q17350442	instance of	venue	2	0.0125	0.3835	1.11
104	Q622425	instance of	nightclub	2	0.0125	0.3835	0.95

Table 9 – Social network analysis of claims for London with all properties enabled

08 APPENDIX A

```
# Wikidata Query Service at:
# https://query.wikidata.org/
#defaultView:Graph

SELECT ?place ?placeLabel ?location ?style ?styleLabel ?instanceLabel
# SELECT ?style ?styleLabel (count(*) as ?count)
WHERE {

  wd:Q84 wdt:P625 ?loc . # Geo-coords of Newcastle upon Tyne: Q1425428; London:Q84
  SERVICE wikibase:around {
      ?place wdt:P625 ?location . # Places with Geo-coords
      bd:serviceParam wikibase:center ?loc .
      bd:serviceParam wikibase:radius "5" . # Within radius 5 km around center
  }
  ?place wdt:P149 ?style . # Has architectural style
  OPTIONAL {
    ?place wdt:P31 ?instance . # Optionally include the instance of
  }
  SERVICE wikibase:label {
    bd:serviceParam wikibase:language "en" .
  }
  BIND(geof:distance(?loc, ?location) as ?dist) # Calculate distance in km
}
# GROUP BY ?style ?styleLabel
ORDER BY ?dist # Order by distance
LIMIT 1000
```

09 APPENDIX B

During the development of the application, a significant effort was spent on developing the user experience (UX) and a series of prototypes were developed before the UI was finalised with what intends to be an intuitive and informative interface.

Criteria

https://graphingthecity.studiole.uk/#/criteria

The initial criteria screen prominently features the city input field with an intuitive autocomplete feature. Enabling the user to easily define the precise city that is of interest. Alternative links to load existing data are included beneath the input box, and advanced settings can be revealed by clicking the appropriate toggle button.

Under the advanced settings, it is possible to add or remove Wikidata properties which the application will use to connect places. By removing all properties and submitting the criteria the application will connect by all properties without restriction. To add a new property simply enter its Wikidata property identifier (e.g. P149).

Advanced settings also permit the user to change the layout algorithm used to display the graph, with cose-bilkent being the default.

Compute

https://graphingthecity.studiole.uk/#/compute

Once the user hits submit on the criteria they are taken to the compute page where the application logic is performed in the background. As the process takes a few seconds to complete the user is kept updated by the process diagram (Figure 22) with the nodes filling as the compute stages are completed.

Graph

https://graphingthecity.studiole.uk/#/graph

Once computation has completed the user is automatically redirected to the graph page, here the process diagram remains visible for a few seconds while Cytoscape.js renders the graph in the background and the social network analysis is completed. Once these operations are complete the graph is displayed and laid out with the algorithm defined under advanced settings – the default being cose-bilkent.

The graph is a dynamic, interactive element which can be interacted with intuitively by anyone familiar with the Google Maps or similar application interfaces. The graph can be panned by clicking and dragging the canvas or zoomed with the scroll wheel.

The nodes themselves are also interactive, they can be moved around the interface by click and dragging. Details of the node are displayed in the infobox on the left, which gives details of the name, description and centrality measures for the node. Claims nodes, in addition, include information on the property through which they connect, and place nodes have information on which claims they hold. All property and claim labels are linked to their respective Wikidata entries for additional data.

By default, the infobox is filled when hovering over a node with the mouse, but a node can also be selected by clicking on it and then deselected by clicking another node or clicking on the empty canvas in order to return to the hover mode.

Analysis

https://graphingthecity.studiole.uk/#/analysis

The analysis page is opened by clicking on the table icon beneath the criteria input at the top of the infobox. Two tables are included on the page, the first lists all the claim nodes and, beneath, the second lists the place nodes. The tables can be resorted by clicking the column headings and use the scroll wheel on your mouse to scroll through the table. Again all Wikidata entities are linked to their respective pages.

Map

https://graphingthecity.studiole.uk/#/map

The map page is opened by clicking the map icon beneath criteria input on the infobox. The map displays the coordinates of every place node and provides information in the infobox when each node is clicked.

Export Methods

https://graphingthecity.studiole.uk/#/graph

Returning to the graph page – by clicking the chain icon beneath the criteria input on the infobox – a number of export methods are available. These are shown at the bottom of the infobox when no nodes are selected. These methods have been included to ensure that the data derived by the application remains portable so that researchers can analyse it as they see fit, whether that be importing it into Microsoft Excel, Cytoscape Desktop or even GIS software.

.json

Exports all the data used by the application to display the graph for the criteria city, by downloading the JSON data it can be re-imported to the application at a later date.

.cyjs

Once the graph has been rendered the .cyjs data can be exported and the file can be imported into the Cytoscape Desktop application to perform deeper graph analysis.

.png & .jpg

Once the graph has been rendered these links will export the graph layout as images in the respective format. By default, these are exported with black edges and borders for display on a white background, however, this can be disabled and the white edges and borders retained by changing the checkbox in advanced settings.

10 BIBLIOGRAPHY

Alexander, C. (1965). A city is not a tree. In Architectural forum (Vol. 122, No. 1, pp. 58-62).

Alexander, C. (1977). A pattern language: towns, buildings, construction. Oxford University Press.

Amaral, J. N., Buro, M., Elio, R., Hoover, J., Nikolaidis, I., Salavatipour, M., Stewart, L., & Wong, K. (2011). About Computing Science Research Methodology.

Amorim, L. (1997). The sectors paradigm. In Proceedings of the first Space Syntax international symposium (Vol. 2, pp. 18-1).

Anand, S., Morley, J., Jiang, W., Du, H., Hart, G., & Jackson, M. (2010). When worlds collide: combining ordnance survey and open street map data.

Andersen, P. (2007). What is Web 2.0?: ideas, technologies and implications for education (Vol. 1, No. 1, pp. 1-64). Bristol: JISC.

AngularJS (n.d.). HTML enhanced for web apps!. Available at: https://angularjs.org/ (Accessed 29 October 2016)

Antoniou, V., Morley, J., & Haklay, M. (2009). The role of user generated spatial content in mapping agencies. In Proceedings of the GIS Research UK 19th Annual Conference (pp. 251-255).

Bafna, S. (2001). Geometrical intuitions of genotypes. In 3rd Space Syntax Symposium, College of Architecture, Georgia Institute of Technology, Atlanta.

Bafna, S. (2003). SPACE SYNTAX A brief introduction to its logic and analytical techniques. Environment and Behavior, 35(1), 17-29.

Bafna, S. (2012). Rethinking genotype: Comments on the sources of type in architecture. The Journal of Space Syntax, 3(1), 69-80.

Batty, M., & Longley, P. A. (1994). Fractal cities: a geometry of form and function. Academic Press.

Batty, M., Crooks, A., Hudson-Smith, A., Milton, R., Anand, S., Jackson, M., & Morley, J. (2010). Data mash-ups and the future of mapping. JISC, Bristol.

Bavelas, A. (1950). Communication patterns in task-oriented groups. Journal of the acoustical society of America.

Bigne´, J. E., Sanchez, M. I., & Sanchez, J. (2001).Tourism image, evaluation variables and after purchase behavior: Interrelationship. Tourism Management, Vol.22 No.6,pp. 607-616

Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., & Hellmann, S. (2009). DBpedia-A crystallization point for the Web of Data. Web Semantics: science, services and agents on the world wide web, 7(3), 154-165.

Brandes, U. (2008). On variants of shortest-path betweenness centrality and their generic computation. Social Networks, 30(2), 136-145.

Chen, C., & Tsai, D. (2007). How destination image and evaluative factors affect behavioral intentions. Tourism Management, Vol.28 No.4,pp.1115-1122.

Chi, C. G. Q., & Qu, H. (2008). Examining the structural relationships of destination image, tourist satisfaction and destination loyalty: An integrated approach. Tourism Management, Vol. 29No.4, pp.624-636.

Cytoscape. (n.d.). Cytoscape. Available at: http://www.cytoscape.org/ (Accessed: 29 October 2016)

Cytoscape.js (n.d.-a). Cytoscape.js. Available at: http://js.cytoscape.org/ (Accessed: 29 October 2016)

Cytoscape.js (n.d.-b). closeness-centrality.js. Available at: https://github.com/cytoscape/cytoscape.js/blob/127575d9814ebed2f3976f84ed1026ae27ee337b/src/collection/algorithms/closeness-centrality.js#L66 (Accessed: 12 December 2016)

Cytoscape.js (n.d.-c). degree-centrality.js. Available at: https://github.com/cytoscape/cytoscape.js/blob/127575d9814ebed2f3976f84ed1026ae27ee337b/src/collection/algorithms/degree-centrality.js#L109 (Accessed: 12 December 2016)

Cytoscape.js (n.d.-d). betweenness-centrality.js. Available at: https://github.com/cytoscape/cytoscape.js/blob/127575d9814ebed2f3976f84ed1026ae27ee337b/src/collection/algorithms/betweenness-centrality.js#L8 (Accessed: 12 December 2016)

D3.js (n.d.). D3 Data-Driven-Documents. Available at: https://d3js.org/ (Accessed: 29 October 2016)

Dalton, R. C., & Kirsan, C. (2008). Small-graph matching and building genotypes. Environment and Planning B: Planning and Design, 35(5), 810-830.

Dalton, R., & Bafna, S. (2003). The syntactical image of the city: a reciprocal definition of spatial elements and spatial syntaxes.

Deeksha, S. D., Ashrith, H. C., Bansode, R., & Kamath, S. (2015, July). A spatial clustering approach for efficient landmark discovery using geo-tagged photos. In Electronics, Computing and Communication Technologies (CONECCT), 2015 IEEE International Conference on (pp. 1-6). IEEE.

Doersch, C., Singh, S., Gupta, A., Sivic, J., & Efros, A. (2012). What makes paris look like paris?. ACM Transactions on Graphics, 31(4).

Du, H., Anand, S., Morley, J., Hart, G., Leibovici, D., & Jackson, M. (2011a). Developing open source based tools for geospatial integration.

Du, H., Jiang, W., Anand, S., Morley, J., Hart, G., Leibovici, D., & Jackson, M. (2011b). An ontology based approach for geospatial data integration of authoritative and crowd sourced datasets. In Proceedings of the 25th International Cartographic Conference.

Elsdon, L. (2016a). Graphing the City Repository. Available at: https://github.com/StudioLE/GraphingTheCity/blob/master/src/controllers/compute.js (Accessed: 1 November 2016)

Elsdon, L. (2016b). Graphing the City. Available at: https://graphingthecity.studiole.uk (Accessed: 1 November 2016)

Elsdon, L. (2016c).Wikidata Graph: Connect places by architectural style in city. Available at: https://gist.github.com/StudioLE/c1e5fbb5358d72e8cfb5cdb65a564acf (Accessed: 21 December 2016)

Euromonitor International. (2016). Top 100 City Destinations Ranking. Available at: http://go.euromonitor.com/Top-100-City-Desinations-Free-Report.html (Accessed 29 October 2016)

Ford, H., & Graham, M. (2016). Provenance, Power and Place: Linked Data and Opaque Digital Geographies. Ford, H., and Graham, M.

Franz, G., Mallot, H., Wiener, J., & Neurowissenschaft, K. (2005). Graph-based models of space in architecture and cognitive science-a comparative analysis. In Proceedings of the 17th International Conference on Systems Research, Informatics and Cybernetics (Vol. 3038).

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. Sociometry, 35-41.

Gehl, J. (1987). Life between buildings. Using Public Space, New York.

Google Places API. (n.d.-a). API Features. Available at: https://developers.google.com/places/ (Accessed: 29 October 2016)

Google Places API. (n.d.-b). Text Search Requests. Available at: https://developers.google.com/places/web-service/search#TextSearchRequests (Accessed: 29 October 2016)

Google Places API. (n.d.-c). Place Types. Available at: https://developers.google.com/places/web-service/supported_types (Accessed: 29 October 2016)

Google reCAPTCHA. (n.d.). The reCAPTCHA advantage. Available at: http://www.google.com/recaptcha/intro/index.html (Accessed: 31 October 2016)

Google Search. (n.d.-a). Google Knowledge Graph Search API. Available at: https://developers.google.com/knowledge-graph/ (Accessed: 29 October 2016)

Google Search. (n.d.-b). Method entities.search. Available at: https://developers.google.com/knowledge-graph/reference/rest/v1/ (Accessed: 29 October 2016)

"Google Search. (n.d.-c).

Introduction to Structured Data. Available at: https://developers.google.com/search/docs/guides/intro-structured-data (Accessed: 22 November 2016)"

Han, J., & Lee, H. (2015). Adaptive landmark recommendations for travel planning: Personalizing and clustering landmarks using geo-tagged social media. Pervasive and Mobile Computing, 18, 4-17.

Hanson, J. (1998). Decoding homes and houses. Cambridge University Press.

Hiller, B. (1996). Space is the Machine: A configurational theory of architecture.

Hillier, B. (2007). Space is the machine: a configurational theory of architecture.

Hillier, B., & Hanson, J. (1984). The social logic of space, 1984. Cambridge: Press syndicate of the University of Cambridge.

Hillier, B., & Leaman, A. (1974). How is design possible?. Journal of Architectural and Planning Research, 3(1), 4-11.

Hillier, B., Hanson, J., & Graham, H. (1987). Ideas are in things: an application of the space syntax method to discovering house genotypes. Environment and Planning B: Planning and Design, 14(4), 363-385.

Hoffart, J., Suchanek, F. M., Berberich, K., & Weikum, G. (2013). YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia. Artificial Intelligence, 194, 28-61.

Ji, R., Xie, X., Yao, H., & Ma, W. Y. (2009). Mining city landmarks from blogs by graph modeling. In Proceedings of the 17th ACM international conference on Multimedia (pp. 105-114). ACM.

Jiang, B. (2012, June). Computing the image of the city. In Proceedings of the 7th International Conference on Informatics and Urban and Regional Planning (pp. 111-121).

Jiang, K., Liu, L., Xiao, R., & Yu, N. (2012). Mining local specialties for travelers by leveraging structured and unstructured data. Advances in Multimedia, 2012, 15.

JLL. (2016). Benchmarking the Future World of Cities. Available at: http://www.jll.com/research/176/jll-benchmarking-the-future-world-of-cities (Accessed: 22 December 2016)

JSON (n.d.) Introducing JSON. Available at: http://json.org/ (Accessed: 22 November 2016)

Keyhole Markup Language. (n.d.) What is KML? Available at: https://developers.google.com/kml/ (Accessed: 22 November 2016)

Kim, Y., Choi, T. Y., Yan, T., & Dooley, K. (2011). Structural investigation of supply networks: A social network analysis approach. Journal of Operations Management, 29(3), 194-211.

Lennard, S. H. C., & Lennard, H. L. (1995). Livable cities observed: a source book of images and ideas for city officials, community leaders, architects, planners and all other committed to making their cities livable. Gondolier Press.

Li, J., Qian, X., Tang, Y. Y., Yang, L., & Mei, T. (2013). GPS estimation for places of interest from social users' uploaded photos. Multimedia, IEEE Transactions on, 15(8), 2058-2071.

Lynch, K. (1960). The image of the city (Vol. 11). MIT press.

Newman, M. E. (2005). A measure of betweenness centrality based on random walks. Social networks, 27(1), 39-54.

Nystuen, J. D., & Dacey, M. F. (1961). A graph theory interpretation of nodal regions. In Papers of the Regional Science Association (Vol. 7, No. 1, pp. 29-42). Springer-Verlag.

Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. Social networks, 32(3), 245-251.

O'Sullivan, D. (2001). Graph-cellular automata: a generalised discrete urban and regional model. Environment and Planning B: Planning and Design,28(5), 687-705.

O'Sullivan, D. B. (2000). Graph-based cellular automaton models of urban spatial processes (Doctoral dissertation, University College London (University of London)).

Oxford Dictionaries. (2016). Available at: https://www.oxforddictionaries.com/ (Accessed: 22 November 2016)

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. Journal of experimental psychology, 76(1p2), 1.

Park, Y. (2015). The network of patterns: creating a design guide using Christopher Alexander's pattern language. Environment and Planning B: Planning and Design, 42(4), 593-614.

Peponis, J., Hadjinikolaou, E., Livieratos, C., & Fatouros, D. A. (1989). The spatial core of urban culture. Ekistics, 43-55.

Petrick, J. F. (2004). The roles of quality, value, and satisfaction in predicting cruise passengers' behavioural intentions. Journal of Travel Research, Vol.42 No.4, pp.397-407.

Places API Web Service. (n.d.). Place Autocomplete. Available at: https://developers.google.com/places/web-service/autocomplete (Accessed: 29 October 2016)

Riza, M., Doratli, N., & Fasli, M. (2012). City branding and identity. Procedia-Social and Behavioral Sciences, 35, 293-300.

Roach, E., & Lloyd, B. B. (1978). Cognition and categorization. Hillsdale, New Jersey, 47.

Salingaros, N. A. (1998). Theory of the urban web. Journal of Urban Design,3(1), 53-71.

Schema.org. (n.d.). Full Hierarchy. Available at: http://schema.org/docs/full.html (Accessed: 29 October 2016)

Shen, H., Huynh, D. F., Chung, G., Zhou, C., Huang, Y., & Li, G. (2012). U.S. Patent Application No. 14/651,332.

Smith, D. A., & Timberlake, M. (1995). Conceptualising and mapping the structure of the world system's city system. Urban Studies, 32(2), 287-302.

Starr, B. (2015). Structured Data & The SERPs: What Google's Patents Tell Us About Ranking In Universal Search. Available at: http://searchengineland.com/structured-data-serps-googles-patents-tell-us-ranking-universal-search-219205 (Accessed: 22 December 2016)

The Economist Intelligence Unit. (2016). Global Liveability Ranking 2016. Available at: http://www.eiu.com/public/topical_report.aspx?campaignid=Liveability2016 (Accessed: 30 October 2016)

Toman, J., & Olszewska, J. I. (2014). Algorithm for graph building based on Google Maps and Google Earth. In Computational Intelligence and Informatics (CINTI), 2014 IEEE 15th International Symposium on (pp. 55-60). IEEE.

Turner, A., Penn, A., & Hillier, B. (2005). An algorithmic definition of the axial map. Environment and Planning B: planning and design, 32(3), 425-444.

UNESCO. (n.d.). World Heritage List. Available at: http://whc.unesco.org/en/list/?search=barcelona (Accessed: 29 October 2016)

University of Washington. (n.d.). Cities and Buildings Database. Available at: http://content.lib.washington.edu/buildingsweb/index.html (Accessed: 28 December 2016)

Von Ahn, L., Maurer, B., McMillen, C., Abraham, D., & Blum, M. (2008). recaptcha: Human-based character recognition via web security measures. Science, 321(5895), 1465-1468.

Vrandečić, D., & Krötzsch, M. (2014). Wikidata: a free collaborative knowledgebase. Communications of the ACM, 57(10), 78-85.

Wikidata. (n.d.-a). Linked Data interface. Available at: https://www.wikidata.org/wiki/Wikidata:Data_access#Linked_Data_interface (Accessed: 29 October 2016)

Wikidata. (n.d.-b). British Museum (Q6373). Available at: https://www.wikidata.org/wiki/Q6373 (Accessed: 29 October 2016)

Wikidata. (n.d.-c).MediaWiki API help. Available at: https://www.wikidata.org/w/api.php (Accessed: 29 October 2016)

Wikidata. (n.d.-d). King's Cross station (Q219867). Available at: https://m.wikidata.org/wiki/Q219867 (Accessed: 27 December 2016)

Wikidata. (n.d.-e). Reichstag building (Q151897). Available at: https://m.wikidata.org/wiki/Q151897 (Accessed: 27 December 2016)

Wikidata. (n.d.-f). Revision history of "British Museum" (Q6373). Available at: https://www.wikidata.org/w/index.php?title=Q6373&offset=&limit=500&action=history (Accessed: 28 December 2016)

World Cities Culture Forum. (2016). The World Cities Culture Report 2015. Available at: http://www.worldcitiescultureforum.com/publications/world-cities-culture-report-2015 (Accessed: 19 April 2016)

Zhou, B., Liu, L., Oliva, A., & Torralba, A. (2014). Recognizing city identity via attribute analysis of geo-tagged images. In European Conference on Computer Vision (pp. 519-534). Springer International Publishing.

Zhou, B., Liu, L., Oliva, A., & Torralba, A. (2014b). Supplementary Materials. Available at: http://people.csail.mit.edu/bzhou/project/eccv2014/eccv2014supp.pdf (Accessed: 21 December 2016)