**Sustaining Glasgow's Urban Networks: the Link Communities of Complex Urban Systems**

**Itova, I.**



https://doi.org/10.34737/vz7z8

# SUSTAINING GLASGOW'S URBAN NETWORKS:

# THE LINK COMMUNITIES OF COMPLEX URBAN SYSTEMS

IRENA ITOVA

A thesis submitted in the partial fulfilment of the

requirements of the University of Westminster

for the degree Doctor of Philosophy

Thesis submitted for examination in March 2022. Viva successfully passed in April 2022 with minor amendments. Final electronic copy submitted on the 16th of September 2022

# SUSTAINING GLASGOW'S URBAN NETWORKS: THE LINK COMMUNITIES OF COMPLEX URBAN SYSTEMS

IRENA ITOVA

PhD 2022

*Infrastructure, rendering nature to an artifice (Neuman, 2006) that humans are enabled to control by the use of technology (Heidegger, 1977), becomes its threat.*

# Declaration

This dissertation is an original work and a result of my own research and it is not produced as a result of any scientific collaboration with researchers from within or outside the University of Westminster. The work is not currently submitted to obtain any other degree or qualification at the University of Westminster or at any other University, including other research institutions within the United Kingdom or abroad.

Additionally, I declare that no part of my dissertation has been already submitted for obtaining a degree or a qualification at the University of Westminster or any other University, including other domestic or foreign research institutions.

I also declare that the work does not exceed the prescribed word limit as stated in the University's Research Degree Handbook 2021/22.

Irena Itova

University of Westminster

Marylebone Campus, London

August, 2022

# Acknowledgement

I would like to extend my gratitude to Ir. Jelle Rijpma for his unconditional and generous support, both motivational and financial, during the four years while writing this thesis. His passion for cities and improving urban living with the use of recent technological advances, was the true motivation to finance this research project. Next, I would like to thank my Director of Studies Prof. Michael Neuman for his patience, understanding and endless support throughout my PhD journey. He believed in my capability to conduct this research, even in the moments when I doubted myself, which were many(!) Further, special gratitude goes to my second supervisor Dr. Epaminondas Kapetanios, who was a true visionary, guiding and advising me patiently through the field of Complex Networks. The level of support he has offered me, as a novice in the field of Algorithmic Theory, Graph Theory and Combinatorial Programming, has helped me overcome the most basic yet fundamental concepts on which this research was built.

Further, equally I would like to thank my husband Alec, my sister Tatjana and my parents and all my friends, who have been patiently listening to all my stories of successes and failures, fears and hopes, that arose during the four years of the PhD journey. Especially Alec has been an exceptional inspiration and a beacon of light, convincing me over and over to pursue the topics in this thesis, as he too has a great passion for the environment, sustainability and cities.

There are many colleagues from two special research groups whom I'd like to thank. Those are the colleagues affiliated with the Active Travel Academy at the University of Westminster, and the Public Health Modelling group part of the MRC Epidemiology Unit, at University of Cambridge. While I was engaged in research conducted by these groups, I benefited greatly in learning new methods, skills and exchange of ideas, all of which helped me to finish my doctoral research.

In a final remark I would like to extend my gratitude to the Urban Analytics Lab at the Hogeschool van Amsterdam, for hosting me in the second semester of my PhD journey and where I got my initial training for RStudio.

# Summary

As cities grow in population size and became more crowded (UN DESA, 2018), the main future challenges around the world will remain to be accommodating the growing urban population while drastically reducing environmental pressure. Contemporary urban agglomerations (large or small) constantly impose burden on the natural environment by conveying ecosystem services to close and distant places, through coupled human nature [infrastructure] systems (CHANS). Tobler's first law in geography (1970) that states that "everything is related to everything else, but near things are more related than distant things" is now challenged by globalization.

When this law was first established, the hypothesis referred to geological processes (Campbell and Shin, 2012, p.194) that were predominantly observed in pre-globalized economy, where freight was costly and mainly localized (Zhang et al., 2018). With the recent advances and modernisation made in transport technologies, most of them in the sea and air transportation (Zhang et al., 2018) and the growth of cities in population, natural resources and bi-products now travel great distances to infiltrate cities (Neuman, 2006) and satisfy human demands. Technical modernisation and the global hyperconnectivity of human interactions and trading, in the last thirty years alone resulted with staggering 94 per cent growth of resource extraction and consumption (Giljum et al., 2015). Local geographies (Kennedy, Cuddihy and Engel-Yan, 2007) will remain affected by global urbanisation (Giljum et al., 2015), and as a corollary, the *operational inefficiencies* of their local infrastructure networks, will contribute even more to the issues of environmental unsustainability on a global scale.

Another challenge for future city-regions is the equity of public infrastructure services and policy creation that promote the same (Neuman and Hull, 2009). Public infrastructure services refer to services provisioned by networked infrastructure, which are subject to both public obligation and market rules. Therefore, their accessibility to all citizens needs to be safeguarded. The disparity of growth between networked infrastructure and socio-economic dynamics affects the sustainable assimilation and equal access to infrastructure in various districts in cities, rendering it as a privilege. Yet, the empirical evidence of whether the place of residence acts as a disadvantage to public service access and use, remains rather scarce (Clifton et al., 2016). The European Union recognized (EU, 2011) the issue of equality in accessibility (i.e. equity) critical for territorial cohesion and sustainable development across districts, municipalities and regions with diverse economic performance.

Territorial cohesion, formally incorporated into the Treaty of Lisbon, now steers the policy frameworks of territorial development within the Union. Subsequently, the European Union developed a policy paradigm guided by equal access (Clifton et al., 2016) to public infrastructure services, considering

their accessibility as instrumental aspect in achieving territorial cohesion across and within its member states.

A corollary of increasing the equity to public infrastructure services among growing global population is the potential increase in environmental pressure they can impose, especially if this pressure is not decentralised and surges at unsustainable rate (Neuman, 2006)[1]. This danger varies across countries and continents, and is directly linked to the increase of urban population due to; [1] improved quality of life and increased life expectancy and/or [2] urban in-migration of rural population and/or [3] global political or economic immigration.

These three rising urban trends demand new approaches to reimagine planning and design practices that foster infrastructure equity, whilst delivering environmental justice. Therefore, this research explores in depth the nature of growth of networked infrastructure (Graham and Marvin, 2001) as a complex system and its disparity from the socio-economic growth (or decline) of Glasgow and Clyde Valley city-region. The results of this research gain new understanding in the potential of using emerging tools from network science for developing optimization strategy that supports more decentralized, efficient, fair and (as an outcome) sustainable enlargement of urban infrastructure, to accommodate new and empower current residents of the city.

Applying the novel link clustering community detection algorithm (Ahn et al., 2010) in this thesis I have presented the potential for better understanding the complexity behind the urban system of networked infrastructure, through discovering their overlapping communities. As I will show in the literature review (Chapter 2), the long standing tradition of centralised planning practice relying on zoning and infiltrating infrastructure, left us with urban settlements which are failing to respond to the environmental pressure and the socio-economic inequalities. Building on the myriad of knowledge from planners, geographers, sociologists and computer scientists, I developed a new element (i.e. *link communities*) within the theory of urban studies that defines cities as complex systems. After, I applied a method borrowed from the study of complex networks to unpack their basic elements. Knowing the link (i.e. functional, or overlapping) communities of metropolitan Glasgow enabled me to evaluate the current level of communities interconnectedness and reveal the gaps as well as the potentials for improving the studied system’s performance.

The complex urban system in metropolitan Glasgow was represented by its networked infrastructure, which essentially was a system of distinct sub-systems, one of them mapped by a physical and the other one by a social graph. The conceptual framework for this methodological approach was

[1] i.e. conveying ecosystem goods at unsustainable rate results with failure of ecosystems’ remediation.

formalised from the extensively reviewed literature and methods utilising network science tools to detect community structure in complex networks. The literature review led to constructing a hypothesis claiming that the efficiency of the physical network's topology is achieved through optimizing the number of nodes with high betweenness centrality, while the efficiency of the logical network's topology is achieved by optimizing the number of links with high edge betweenness.

The conclusion from the literature review presented through the discourse on to the primal problem in 7.4.1, led to modelling the two network topologies as separate graphs. The bipartite graph[2] of their primal syntax was mirrored to be symmetrical and converted to dual. From the dual syntax I measured the complete accessibility (i.e. betweenness centrality) of the entire area and not only of the streets.

Betweenness centrality of a node measures the number of shortest paths that pass through the node connecting pairs of nodes. The betweenness centrality is same as the integration of streets in space syntax, where the streets are analysed in their dual syntax representation. Street integration is the number of intersections the street shares with other streets and a high value means high accessibility.

Edges with high betweenness are shared between strong communities. Based on the theoretical underpinnings of the network's modularity and community structure analysed herein, it can be concluded that a complex network that is both robust and efficient (and in urban planning terminology 'sustainable') is consisted of numerous strong communities connected with each other by optimal number of links with high edge betweenness. To get this insight, the study detected the edge cut-set and vertex cut-set of the complex network. The outcome was a statistical model developed in the open source software R (Ihaka and Gentleman, 1996). The model empirical detects the network's overlapping communities, determining the current sustainability of its physical and logical topologies.

Initially, an assumption was that the number of communities within the infrastructure (physical) network layer were different from the one in the logical. They were detected using the Louvain method that performs graph partitioning on the hierarchical streets structure. Further, the number of communities in the relational network layer (i.e. accessibility to locations) was detected based on the OD accessibility matrix established from the functional dependency between the household locations and predefined points of interest. The communities from the graph of the 'relational layer' were discovered with the single-link hierarchical clustering algorithm. The number of communities observed in the physical and the logical topologies of the eight shires significantly deviated.

[2] The graph representation of a network where both streets and intersections are represented by nodes and their topological connection by links.

# List of key abbreviations

| | |
|---|---|
| ABM | Agent-based Modelling |
| ABMs | Agent Based models |
| BA | Barabási – Albert model |
| CA | Cellular Automata |
| CHANS | Coupled Human and Nature Systems |
| CUD | Central Urban District |
| EB | Edge betweenness |
| EIA | Environmental Impact Assessment |
| ER | Erdős–Rényi model |
| EU | European Union |
| GC | Giant Component |
| GI | Green Infrastructure |
| GIS | Geographic Information Systems |
| GPS | Global Positioning System |
| GRASS | Geographic Resource Analysis Support System |
| GRG | Descriptive Generalized Random Graph |
| GUI | Graphical User Interface |
| K-NN | K number of nearest neighbours |
| LTSs | Large Technical Systems |
| LTNs | Low Traffic Neighbourhoods |
| NP-hard | Nondeterministic Polynomial time |
| OD | Origin-Destination matrix |
| OSM | OpenStreetMap |
| PARDLI | Priority Areas for Re-use of Derelict Land Index |
| QGIS | Quantum Geographic Information System |
| RS | Remote sensing |
| SLA | Street-Based' Local Areas |
| SNA | Social Network Analysis |
| SOC | Self-Organised Criticality |
| UHI | Urban Heat Island |
| WWII | World War II |
| WWW | World Wide Web |

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# Chapter 1: Introduction

The world is rapidly urbanizing. This growth is mainly taking place in 31 megacities, a number that could increase to 43 by 2050 (UN DESA, 2018). The definition of megacities varies across organisations. The United Nations Department of Social Affairs (UN DESA, op cit.) defines megacities as urban conglomerations that house more than 10 million inhabitants. If the areas with high probability to become megaregions undergo this transition by 2030, they will increase the urban land cover on Earth by 1.2 million squared km, resulting with significant loss of habitat in key biodiversity areas (Seto et al., 2012) and increased pressure on ecological services (Matthews et al., 2000). There is a growing fear that these changes will trigger abrupt shifts on planetary scale (Long et al., 2014). Therefore, close examination of key structural and functional shifts caused by the emerging polycentric urban regions is needed (Alberti, 2016). Much of the existing research that studies the way the physical form affects sustainability of any city (Jenks and Jones, 2010; Song et al., 2017) is focused on its physical aspect at a large scale, including spatial size, density, land use and shape, excluding the functional shifts triggered by the intertwined urban processes over the sustainability of its physical form.

For example, Bettencourt (2013) using socioeconomic output and infrastructure costs has already empirically explained the general interdependence among the socio-economic, technical and spatial facets in cities as aggregated socioeconomic forces. His study suggests that infrastructure costs and socioeconomic outputs are invariant across city sizes and they may be used as means to evaluate prospective urban strategies. The variables of Bettencourt's study include [1] the average number of local interactions per person, [2] the total average social output, [3] the total power spent in transportation processes to keep the city mixed and [4] the baseline area where productive interaction takes place. Subsequently, these variables resulted with an 'urban efficiency model' based on four model terms— [1] mixing population; [2] incremental (infrastructure) network growth; [3] bounded human effort and [4] socioeconomic output. Despite the very detailed empirical analysis made in Bettencourt's study, more holistic detailed empirical studies are still scarce (Jenks and Jones, 2010, p3; Bosch et al., 2019). Specifically, one very important variable missing in Bettencourt's definition of sustainable urban living is the environmental impact (i.e. environmental degradation) of the four variables analysed. The aspects of sustainability of the socioeconomic outputs is addressed in the following section.

Design and organization of space, the principle concern of urban morphology studies, for efficient movement of goods and people within cities is only one aspect of their sustainability. Sustainability of cities expands beyond their morphology, into their material flows, processes and movement of people, social interaction, goods, energy, money and ideas (Neuman, 2005). These fluxes are the challenging part whose inter-dependencies makes them difficult to be measured and explicitly linked directly to sustainability of urban living. Currently available methods for complex systems modelling, allow mapping and observing of dynamic processes of system components interactions, that involve millions of particles. The technical shortcomings of past research are eliminated in modern modelling practices, enabling contemporary models of cities to include their flows and processes. The method presented in the thesis tries to reduce the gap in existing models, by introducing new modelling approach of representing cities as complex systems, including in the same modelling process their 'physical' and 'logical' aspects (Hillier and Hanson, 1984). The model as such, can address the structural complexity behind how cities consume *ecosystem services* and in what way this impacts local and global ecosystems.

Until the 1970s, progressivist urban planning seemed to advocate socio-economic development that was supported by the industrial economy grounded in a linear production process— extract-use-dispose (Fiksel, 2006, p15). These processes to a very large degree resulted with denaturising the Earth (Matthews et al., 2000). Only recently general practices in the modern economy have emerged that deviate from the 'linear model'. These new practices, under the general category of circular economy are promoting an environmentally sustainable economy which decouples development from resource consumption by recycling, upcycling or cradle-to-cradle production principles (McDonough and Braungart, 2009; Ghisellini et al., 2016). This shift is slowly influencing planning practices as well.

Urban planning practices first in the early 70s started to include environmental issues and concerns in their strategic and design methods, by using accounting tools such as the Environmental Impact Assessment (EIA) (Wathern, 1990) or System of Environmental and Economic Accounts (Romanowicz et al., 2004). Both, the EIA method and the national accounts system have the goal to inform on the environmental impact of urban settlements at local, regional or national scale. They are used to evaluate, regulate and improve the outputs of projects, policies and development portfolios on the national and local ecosystem services.

Ecosystem services are defined (Lambin and Meyfroidt, 2010; Boerema et al., 2017) as benefits that the human gets freely from the natural environment and from properly functioning forest, grassland, agro- and aquatic ecosystems. The process is endogenous to the coupled socio-ecological system (Lambin and Meyfroidt, 2010). The complexity of movement in cities of materials spans beyond simple extraction of natural resource from any ecosystem and involves deeper issues related to global politics and trading. To produce any typical product or a service in the global economy, multiple ecological goods and ecological services are involved in the network spanning across geographical borders

(Ravalde and Keirstead, 2017). For example, offshore agriculture (Zoomers, 2010; Lambin and Meyfroidt, 2011) is a direct result of the globalization of trade, economic liberalization of land markets and the expansion of direct foreign investments in the agricultural and construction sectors. Further, water resources are another pressing concern enhanced by global trading. The international trade moves 'virtual' water[3] from more advantaged regions with soil water surplus, to more disadvantaged regions where soil water is scarce (Allan, 1996; 2005). Thus, sustainable consumption of resources is not only concerned with efficiently conveying and distributing of resources in the local ecological and economic system, but also the impact of their extraction across the countries concerned, which can either be centralized or decentralized.

Therefore, in the context of urbanisation and urban settlements, Neuman and Churchill (2015) extend sustainability to the ecology of the human presence in a place and its normative discourse—also known as urban ecology (Luck and Wu, 2002). The authors' definition of urban sustainability is extended to include not only the flows and processes within the urban environment, but also the second law of thermodynamics. Thus, they define urban sustainability as:

> *"...the degree to which an entity exists in coevolutionary process with its environment whose inherent condition enables it to continue evolving and developing without jeopardising its own life and livelihood and the lives and livelihood of those it affects, including the larger systems and networks in which the entity finds itself situated, now and in the foreseeable future."*
> (Neuman and Churchill, 2015, p470)

Thus sustainability of the contemporary urban form is conditioned by the ability of cities to grow at the right pace and within spatial boundaries and resources demand that minimise the invasion over ecological patches and spatial gradients[4] in the natural environment. Thus, their infrastructure systems need to infiltrate nature's ecosystems at sustainable rate and at a large extent decentralized, to which the nearby natural environments can recover the impact of conveyed resources.

Public infrastructure regulates the relationship between humans and natural resources and is a key variable that affects the resilience of ecosystems (Alberti, 2016, p42). The gradual and piecemeal growth of infrastructure (Bettencourt, 2013) to fairly connect people (Lorrain, 2005) as they join the city, is fundamental for supporting the population's vital functions in a sustainable way. Access to affordable public infrastructure (i.e. telecommunications, energy, transportation, housing and

---

[3] The term virtual water refers to water embedded in commodities such as food or clothes. I.e. water used to produce the commodities, which are usually produced in other parts of the world than where they are consumed.
[4] Such as air temperature, precipitation, soil fertility and acidity, moisture regime and frequencies of natural disturbances (e.g. fire, infestation or droughts).

sanitation) promotes regional competitiveness, equity and harmonious territorial development (EU, 2011). Cohesive incremental growth and efficient operation of infrastructure across agglomerations with diverse socioeconomic outputs is of great importance to promote sustainable consumption of ecological goods. If this process of infrastructure growth is not managed sustainably, its expansion proceeds to pose even greater threat to close and distant natural ecosystems.

Heidegger (1977) recognizes and warns us on the danger associated with the transformational power of infrastructure over nature, to steer this influence towards balanced role between the extractive and regenerative functions of infrastructure, to sustain the co-existence of humans with the environment.

## RESEARCH GAP

Numerous methods with the goal to alter the morphology of cities and improve urban sustainability by design already exist (Handy, 1996; Llewelyn-Davies, 1997; Jabareen, 2006). Other research (Williams et al., 2000) has demonstrated that multiple urban forms can be sustainable. However, most of the research is fragmented, focusing either on studying the relationship between morphology and travel behaviour (Song et al., 2017; Wu et al., 2017) or morphology and social-dynamics (van der Veer, 2000) or urban metabolism (Pulido Barrera et al., 2018; McArthur, 2018). Bosch et al. (2019) defines urban form beyond the morphology of places. Instead, they explain the urban form as the result of ever-developing and ever-changing overlapping interactions among governments, households, businesses and other 'agents', which makes sprawl and urban compactness difficult to assess empirically. However, they do conclude that sprawl is partially owed to those ever-changing interactions'. The goal of this thesis was to identify and quantify the 'ever-changing interaction' and characterise their topological structure.

Topology of a system describes the way the system's components are ordered, i.e. connected to each other by level of importance. (Cleveland State University-College of Business-DBA-CSI, unknown; Naimzada et al., 2009). Topology in complex systems such cities is more convoluted to map, since the different sub-systems are interconnected. Knowing which topological structure is found in different parts of urban systems will improve our understanding of the physical, economic, social and geological phenomena that shape our complex world. Thus, this thesis is motivated by the research question:

***In what way do certain [relational] predicates[5], either physical or logical, contribute to the aggregated behaviour of complex systems that further trigger morphological and environmental changes?***

---

[5] Spatial predicates are a formalism in any query language that represent the geometric aspects of the spatial relationships (Egenhofer, 1993) between spatial entities. Complex spatial data models relate to each other, exhibiting a large number of

Research reviewed in Chapters 5 and 6 has shown that systems which are more modular[6] are also the ones that have highly overlapping *communities,* lower network diameters[7] and are more resilient. These overlaps emerge in systems which have highly co-dependent functional components connected with each other by *functional links*. The literature review revealed scarce evidence of studying complex urban systems from this aspect.

The goal of this project was to find these functional links or corridors in the urban network of Glasgow and Clyde Valley that are embedded within the region's morphology (the arrangement of urban elements) connected by supporting infrastructure which provides daily access to opportunities, measured by each household's level of access (Geurs and van Wee, 2004) to different amenities in the wider city region. Thus, the functional dependency of complex urban interactions in the city-region was defined as the level of accessibility of its residents to various destinations and services. Preserving these functional corridors would allow to maintain overall accessibility of locations within the urban fabric even in cases of system's protuberances.

Although the project scope is narrowed to studying complexity through network science approach for measuring residents access to social infrastructure (i.e. the consumer points of some ecosystem services), the same method can be used to address networked infrastructure and fluxes' decentralization across urban sub-systems, or the urban and the natural interconnected system.

## 1.1 ORIGINALITY AND CONTRIBUTION

The thesis seeks to make contribution to the knowledge of urban studies by mapping part of the current [rather inflexible] state of infrastructure as a complex system, using theories and methods from the emerging field of urban science. This approach enabled analysis[8] of one fragment of the flows, processes and physical structure of the complex networked infrastructure at urban scale that spans across an entire city-region. Herein I analysed the spatial distribution of access to key infrastructure and amenities (i.e. ecosystem services access points) in the region of metropolitan

predicates. To improve their handling, they have been classified under two concepts; topological cluster predicates and topological predicate groups (Schneider and Behr, 2006, p40).

[6] Modular systems are comprised of many components (modules) that are semi-independent and in the case of failures of other system modules, they continue their core system operation.

[7] The network diameter is the largest step-length distance between the two network elements that are connected and furthest apart in the system.

[8] Due to time constraints and limited access to data sources and computational capacity, networked infrastructure, understood as the complex internetworked infrastructure consisted of many sub-systems that includes all the socio-economic and physical flows in the city in their entirety, is analysed with limited number of variables, considering only some aspects of urban living. Therefore, the findings are specific to this project and can be generalized with great care after considering context and transferability.

Glasgow. The research hypothesis assumed that nested community structures of urban processes[9] in the region have different topology than the physical layout of networked infrastructure that conveys them. This assumption was tested using a single-link clustering algorithm which grouped the two different type of links (i.e. physical and relation) into overlapping communities. The result of this algorithm uncovered the overlapping structure of places in their spatial arrangement and their supporting networked infrastructure. As it is presented in the results (see Chapters 8 and 9), some areas in the metropolitan region enjoyed better accessibility to places and infrastructure than others, demonstrating the potential of this new spatial analysis method to assist the planning and design practice addressing the issues of environmental and social injustice more systematically.

### *1.1.1 Theoretical originality*

The theoretical originality spans beyond the technical analysis, the quantification of infrastructure, and its efficiency within the urban environment. It offers cross-disciplinary approach by considering the resilience of infrastructure systems beyond their technical parts to extend to the resilience of the built and the natural environments. Resilience in hybrid human-nature systems (An et al., 2014; Alberti, 2016) is vital, but also challenging to reach. Due to human interactions with the natural environment, each urban change potentially affects connectivity (i.e. links) in numerous other systems, such as food chains (Janssen et al., 2006). Food webs may be disrupted by the construction of roads that run through an ecological patch. The balance in the ecosystem greatly depends on the density of nodes, thus ecological (i.e. functional) links with high *centrality* are needed to keep reachability among the nodes (e.g. plants, insects, animals, people, or places) within the hybrid human-nature systems. Even when a system is fully connected (i.e. all nodes are connected to each other by links), reachability of some parts (i.e. groups of remote nodes) to other more central network regions (i.e. nodes which are highly connected) may be difficult, due to the network's low density (i.e. low number of links). These long distances reduce the connectivity (i.e. integration) of remote nodes to the rest of the network. Adding new nodes[10] that reduce these distances, increases the overall reachability in the network. Further, connecting the newly added nodes with links that enjoy high *edge betweenness* (a concept introduced in 5.1.2) could improve the network's overall reachability. Thus, maintaining cohesive and integrated network topologies through links with high edge betweenness, helps to preserve the nodes with high community centrality to stay connected.

---

[9] In the absence of available data, mapped through the level of access of residents to infrastructure. This simplification is justified because the logical topology of flows will resemble the proximal access of the road network as usually infrastructure networks are collocated underneath carriage ways or footways in the public space.

[10] Which do not necessarily have to be hubs.

In urban systems, defining and understanding the co-dependencies across the links and nodes in networks is more ambiguous process. The challenge stems from the multiple co-existing infrastructure networks that constrain the use of space, defining the city as a complex system of multilayer networks. The infrastructure networks of cities, known as 'almost planar graphs' (Batty, 2004; Boeing, 2017), such as the transportation system and the water or waste distribution networks have crucial importance for our societies. Infrastructure networks are considered to be approximately planar because there exist links that cross each other, yet are not embedded in the same plane (they are grade separated). Such are roads networked with other roads, railways or mountains and rivers, crossing each other by over- and underpasses (e.g. forks, tunnels or bridges). They are designed to overcome separation between geospatial entities (Kuby et al., 2005). Thus, their seamless operation at a sustainable rate contributes to the balance of the co-dependent systems of urban settlements and natural ecosystems.

The thesis assumes that the creation of these over- and under- passes, along with infrastructure synergies, increase the hybrid human-nature network's modularity and robustness (e.g. against random failures). Grade separated overpasses such as eco-ducts, reduce the impact of manmade infrastructure that separate natural patches and enable species crossings and embrace biodiversity. Other over- or underpasses, which may or may not span in parallel with the eco-ducts, can segregate more vulnerable modes of travel from heavy motorised traffic and enable more pleasant, longer interregional 'active' journeys. Synergies across technical and social infrastructure and nature are created by adopting such flexible design solutions and multi-purpose use of surfaces, connections and buildings, that support both human and nature's needs (Derrible, 2017, Hanzl et al., 2021).

First, by reviewing large amount of interdisciplinary literature, I explained the complexities that emerge in urban settlements which are bound by their historical organisation and the complexities of individual interactions that contribute to aggregated behaviour of the city viewed as a complex system. Second, I analysed the origin of urban morphology and the interconnectivity of underlying infrastructure networks that support sustainable urban living. This allowed me to apply graph theory for mapping the places and infrastructure networks of greater Glasgow into graphs. Third, I presented the general theory of systems complexities which includes explanation on their evolution and phase transitions. Finally, I applied the system's theory to the studied area and discovered the complexities emerging within metropolitan Glasgow through their numerous overlapping communities.

### *1.1.2 Methodological originality*

The research presents two original approaches in the methodology. First, there have been limited attempts to use novel computing methods in urban science that go beyond Agent-Based Modelling

(ABM) or Cellular Automata (CA), in order to advance the scientific knowledge in contemporary urban studies and more specifically in urban science. This advancement is now possible due to the progress made in the global positioning system (GPS) tracking and remote sensing technologies, as well as powerful computing and the availability of open and interoperable coding practices and development environments. The first methodological contribution is the utilisation of advanced spatial statistical computing methods and software, already widely used in the sciences that study complex networks. Namely, the research used a statistical method widely popular in other geo-spatial (e.g. maritime or air transportation) and non-spatial[11] disciplines. The main goal of this research was therefore to contribute to the methodological knowledge base of contemporary urban studies. The methodology will be explained in Chapter 7 and its operationalisation is presented in Chapter 8. The novelty of my methodological approach is the application of the *community detection [in complex networks]* algorithm to capture the disparities between the physical and logical topologies of networked infrastructure. The goal of utilising this novel algorithm to reimagine the role of spatial organisation and infrastructure planning for equalities in both, environmental protection and human welfare.

Second, contrary to existing traditional aggregating gravity models, widely used in urban studies for transport planning or housing location-allocation analysis, the community detection algorithm applied in this study uses a ‘bottom-up’ method of discovering link communities inherited within the spatially constraining existing structure of land use and networked infrastructure. This method is also different than the ones widely used in Social Network Analysis (SNA) for example, which are based on presumed community structures, forcing nodes to only belong to one of the network’s numerous communities. The results from the novel link community detection algorithm are more intuitive (i.e. without imposing apriori biases) and aligned to the ‘demand side’ of urban living.

Namely, Ahn, Bagrow and Lehmann in 2010 introduced the pioneering link clustering algorithm that exposes inherited and highly overlapping communities in large complex networks. Their algorithm is different from the other ones known till then, by the technique applied to cluster links rather than nodes in groups based on the nodes’ similarity score [known as the Jaccard coefficient] that the two neighbouring links share (Ahn et al., 2010). This new method has great potential for application in numerous disciplines and its ability to advance the studies of urban science is acknowledged herein.

The built environment represents a large and highly complex network whose millions of nodes at the same time belong to several communities. Daily we can see examples of these overlapping community structures embedded in the way one person moves through and uses the city, and in the way that the

[11] E.g. social network analysis (SNA) vastly used in social and behavioural studies.

person consumes resources conveyed through public infrastructure. Consider a simple example observed in one person's routine over the period of one random day. Let's say this person first leaves their home and goes to the swimming pool, then off to work. On their way back home from work, they stop to buy groceries. The real complexity behind this 'tour'[12] presents itself when this person uses multiple modes to complete these trips (and legs). Not only the route between the origins and destinations of each trip is important, but also the locations of the mode interchanges. If we consider that the trips are made by bus and bicycle/or car through sharing schemes, then the node connectivity is of great importance. Namely, the intrinsic community in this situation consists of the nodes that define the origin and destination points of all stages of the trip as well as the locations of the bus stop and the car or bike share points throughout this trip. At present, trips in the UK are mainly car dominated, however since there is a great effort on national and international level to move to more shared and decarbonised modes of travel, the study presented in this paper shows a new method of conducting structural design analyses to decentralise and integrate their supporting networks.

We can expand this context to go beyond the mobility flow and apply it over multiple other resource flows the person relies on to perform these activities during the same day; such as the consumption of water, food and energy, and the subsequent waste. Then the nodes part of the same community include waste landfills and recycling points, grocery shops, food production farms, energy production points and water purification locations. These points are part of the functional community of the household this person belongs to and their underlying physical and logical topologies are structurally different from one another which makes their operation highly unsustainable. Two things are responsible for unsustainable behaviour in this system; First, the fact that their underlying networks are highly centralised (see 3.1.). Second, the many exogenous and endogenous socio-economic forces that make activities generated in locations to become self-fulfilled (i.e. the ability of activity to generate more activity). More discussion on this can be found in Chapters 3 and 4. Reducing the structural difference between these topologies, improves their sustainability.

Thus, systematic responses to the pressure to reduce human planetary impact and promote territorial cohesion across regions, is only possible if we are able to model cities in their entire complexity. This means, discovering rather than imposing, and modelling the functional interdependences of their communities. This methodological approach involves detecting instead of assuming the link communities of complex urban networks. The community discovery algorithm introduced in Ahn et al.'s study (2010), uses a *link similarity function* to 'organically' discover communities in any network,

[12] Tour in transport planning is defined as the order at which different trips performed by one person (or vehicle) using the same or different modes are organised.

and therefore this algorithm was suitable to use in the study. The processing of the single-link hierarchical clustering algorithm is opposite to the more traditional network partition algorithms, such as the Clique Percolation or the Girvan-Newman (Barabási, pp. 346-349). These methods discover communities by setting up presumptions of the network's underlying modularity. Moreover, these two well-known algorithms traverse the graph searching for communities by forcing nodes to belong to only one community. Ahn et al.'s (po cit., p 762) link clustering algorithm measures the link pair similarity based on the similarity of the neighbouring nodes connected by these links. Thus, their method considers communities as collections of closely interconnected links and the links that belong to several communities are the ones that have high edge betweenness. The algorithm allows also for nodes to belong to several communities at the same time, and with this criteria, it preserves the complexity in the network's overlapping structure.

The new approach to develop an urban model was structured in two parts; First, by utilising the former to map the infrastructure network components 'traditionally' defined as a collection of buildings, street and pipes. Second, by modelling inhabitants' accessibility, i.e. spatial interactions represented by the accessibility of home locations to various amenities. The former allowed me to analyse the physical topology of the networked infrastructure. The latter approach allowed me to study the logical topology— the emerging communities formed due to the functional dependency of urban locations. When combined for comparison, the two separate network analyses enabled me to confirm my hypothesis and verify the disparity between the physical and the logical topology of the complex underlying urban network that supports these interactions and subsequent processes/flows.

The outcome of this process was a novel model of the city-region that empirically evaluated the networked infrastructure modularity, and therefore its robustness measured as the systemic access of its nodes, i.e. by measuring their closeness across the system. This new evidence can help decision-making in respect to increasing networked infrastructure efficiency to make resources' consumption more sustainable.

## 1.2 RESEARCH SCOPE

### *1.2.1. Spatial scope: Glasgow, a model for sustainable city-region?*

The Glasgow metropolitan region spans on the area of 3.385 sq. km and is a home of nearly two million people (Goodstadt, 2006). This metropolitan region, sometimes referred to as West Central Scotland, is known in the United Kingdom as the region with nearly sixty years of highly varied experience in regional governance and strategic planning. It is located in the Scottish West Central Lowlands, with a densely settled population along the Clyde River.

Between the 70s and 90s Glasgow had witnessed significant population decline (Goodstadt, op cit.), which would naturally lead to assumptions that reduced human activities would result with decline in urban overheating (Krüger et al., 2013; Santamouris, 2016). Yet the contrary was true; the decreased population size led to increased urban overheating. This phenomena was linked mainly to the nature of the region's urban morphology, which contributed to the increase of the heat island effect (Emmanuel and Loconsole, 2015).

Urban Heat Island (UHI) is manifested through increase in ambient temperature which magnitude varies based on the morphological, physical and structural characteristics of the city, the types of construction materials used, the released anthropogenic heat and the urban synoptic climate (Santamouris, 2016, p. 66). Results from the Emmanuel and Loconsole (2015) study indicate that 20% increase of green infrastructure[13] by 2050, could decrease the effects of expected overheating somewhere between 30% to 50%, lowering urban surface temperature by 2 °C.

Land use planning and urban transport policies, due to their long-lasting impact on the city's size and shape, have a share in this influence. However, their combined effects are complex and strongly depend on the indicators used to quantify heat-wave risk (Lemonsu et al. , 2015). Alternatively, structural mapping of self-fulfilled urban interactions and readapting networked infrastructure and built-up areas to have not only extractive, but also regenerative relationship with the natural habitat, offers new design avenues for rolling out green infrastructure and tackling UHI effects.

### *1.2.2 The social construct of infrastructure*

Infrastructure projects have wider goal to deliver public services to all inhabitants, while ensuring they are equitable, cost-effective and sustainable. More recently, greater focus in executing infrastructure projects is placed on their balanced impact and regenerative influence over nature and pollution. The potential of reducing inequalities in urban flood susceptibility risk, ambient temperatures or air and soil pollution, should be well planned using structured and holistic system-approach analyses.

Contemporary design and component analyses for upgrading and expanding infrastructure networks have greatly improved, designing new infrastructure to perform both, extractive and regenerative functions. However these designs lack the methodological scale and the ability to quantify near and far components of infiltrating networked infrastructure with one common measurement. Using novel tools such as the link communities detection algorithm presented herein, enable new avenues of advancement of strategic urban and territorial planning. The method presented in this thesis shows that it can address spatial equality in social accessibility of urban areas from a new social perspective, complementary to the traditional technical and environmental analyses, and support future design and enlargement of infrastructure networks with empirical evidence from the technical, environmental and more importantly, the social concern.

[13] Introducing more green surfaces in variety of forms in the Glasgow Clyde Valley region.

Increasing body of evidence shows the risks unhealthy urban environments pose on residents as short-term and long-term effects. In 2010, a study (Walsh et al., 2010) carried out by the Glasgow Centre for Population Health proved a positive correlation between deteriorated mental health and growing up in an urban environment. Nearly ten years after the research was publicly available, a news article in The Guardian (Macdonald, 2019) communicated the results to the wider public. The Macdonald's article described this emerging social phenomenon as 'The Glasgow Effect'. Walsh et al.'s (op cit.) study strongly positioned one theory to be the main contributor to The Glasgow Effect— the radical urban planning decisions made onward the 1950s have subjected not just the physical, but also the mental health of Glaswegians to the adverse effects of deindustrialization, job losses and associated poverty.

Following these findings, another research team (Maantay and Maroko, 2015) looked further into the potential correlation of specific urban morphology types (i.e. land use structure) and population's wellbeing. More specifically, Maantay and Maroko's study examined the potential association between concentration of derelict land and prevalence of mental health disorders such as anxiety and depression. The study confirmed significant positive correlation ($r=.521,\ p < .001$) between deprived neighbourhoods and density of derelict and vacant land. Thus, deprived communities with unfavourable socio-economic, environmental and health inequalities measured on the PARDIL index, were disproportionately burdened with the psychosocial stress and environmental impact associated with this type of land use.

The Scottish Vacant and Derelict Land Survey (2012) defines derelict land as land which soil quality has been damaged by previous use that is no longer beneficial for future use without soil remediation. Good deal of the present derelict and vacant land in Glasgow is former industrial sites and based on the study's estimates (Maantay and Maroko, 2015, p5), nearly 50 per cent has predominantly 'dirty' and contaminated (i.e. hazardous) soil. Thus, regardless that current land-use is assigned to housing, the soil where these houses built during and post WWII-period are located, could be hazardous to human health. Furthermore, the same study stressed that children and young adolescents could be most vulnerable to adverse health effects, since they use the vacant land as improvised playgrounds.

After Walsh and associates published their research findings, the following year another research (Popham and Boyle, 2011) confirmed that those born in Scotland and living in either Scotland or England and Wales had higher premature mortality expectancy than the English living in England and Wales. These effects did not diminish after adjusting the models for access to cars and tenure type. Thus, leading the researchers to conclude that adjusting the models for only household-level differences and socio-economic deprivation does not fully explain the 'Scottish effect'. They recommend future research study in this area to include *a life course approach* (Braveman, 2014) in order to reveal the real cause behind premature deaths of the Scottish population.

A *life course approach* tends to reveal the connections among individuals and between individuals and their historic, cultural and socioeconomic context in which they conduct their daily routines. For example, persistent socioeconomic residential segregation or financial mechanisms that support public schools based on local property taxes, are direct outcomes of zoning regulations and the accompanying planning practices. Increasing body of evidence (Smith et al., 2017) shows positive relationship between accessibility interventions such as provision of built environment walkability components (e.g. quality parks and playgrounds) and the increase in active travel, physical activity and frequency of use of public places. Nonetheless, lower income households in segregated urban areas are usually less likely to benefit from these accessibility interventions. Evidence from a systematic review (Smith et al., 2017) of twenty-eight international studies recorded tendencies for the spatial distribution of active travel infrastructure improvements to predominantly benefit more affluent socioeconomic groups.

While equity of wide access to various infrastructure interventions in metropolitan Glasgow remains unknown, there is empirical evidence of predominant prevalence of derelict land, polluted industrial soil or poor river water quality in close proximity to deprived areas. One extremely valuable study on these issues conducted by Fairburn and associates (2004) confirmed the presence of spatial inequalities of the built environment quality in metropolitan Glasgow.

Equity within the built environment measures the fairness in spatial distribution of infrastructure and amenities against the socioeconomic and cultural context of the areas. However, provision of amenities, not always guarantees to positively benefit local residents whose neighbourhoods were the targeted beneficiary of improved public spaces.

In summary, equity in high quality outdoor environment and the effects of residential segregation on people's opportunities has received strong study interest in recent decades (Rokem and Vaughan, 2019; Light and Thomas, 2019). As discussed in this sub-chapter, infrastructure equality and accessibility is easier to quantify. While other spatial inequalities, such as the susceptibility to the adverse effects of climate change or health inequalities induced by the neighbourhoods' outdoor environment are more challenging (Lindley et al., 2011; Maantay and Maroko, 2015; Majekodunmi et al., 2020). Challenges in measuring health exposure differences and disparities (Braveman, 2014) arise from the variety of socio-economic factors[14] next to the myriad life events of successes and stresses encountered at the different stages of individuals' lifetime. All these life-course [dis]advantages have a combined effect on individuals' health and well-being at their early and later stages of life, only measurable with the life-course approach.

Furthermore, health disparities by default create compound disadvantages by putting already socioeconomically disadvantaged households into further disadvantage (op cit.). In order to reduce this effect, it is essential to address the problem of spatial inequality in the strategic planning phase of urban environments.

[14] such as household's income or education level.

Lastly, accessibility interventions not always deliver the anticipated outcomes, thus performing spatial analyses that reveal embedded socio-economic dependencies (i.e. the link communities) of studied areas where these interventions should take place, have a great importance.

## 1.3 RESEACH QUESTION

### *1.3.1 Research question and sub-questions*

From the stated research gap, originality and domain relevance, the research question was generated:

**Can the sustainability of expanding 'networked infrastructure' be improved by increasing the number of edges with high edge betweenness[18] connecting communities from the physical with the ones from the 'logical' sub-networks?**

To answer the main question I developed two sub-questions and subsequent objectives.

#### Research sub-question [1]

**As the technical network grows by acquiring new nodes with new links, what is the relationship between the overall physical topology of the network and the topology of the two sub-networks?**

Objective 1: To answer the first sub-question, I constructed two graphs, one of the physical network and one of the relational network. Observing the structural differences of the two separate sub-networks allowed me to measure the statistical correlation between their network properties (for example on various network measures see 5.1). Afterwards, I applied the community detection algorithm, to determine the optimised partition density[19] of the relational (logical) sub-network at which the most relevant communities were obtained. The partition density was then compared with the expected percolation threshold[20] (see 6.1. for more detail) of the physical sub-network.

Unlike the majority of well-known community detection algorithms that apply network-divisive strategy that detects communities in a 'top-down' fashion, this thesis implemented the novel link clustering algorithm introduced by Ahn, Bagrow and Lehmann (2010). The method uses link similarity hierarchical clustering to discover the inherited community structure in networks. There are two innovative methodological approaches in this algorithm. Rather than clustering nodes, the algorithm clusters links and rather than forcing nodes to belong to only a single community (which in real life is never the case), the nodes can belong to more than one community, inheriting all the membership indices. The algorithm measures link similarity to discover community structure in networks, which is

---

[18] Edges with high betweenness are the ones that accommodate the largest number of shortest paths connecting any pair of nodes in the network. The edges with high betweenness that tend to connect nodes that belong to different communities within the complex system are of special strategic importance.
[19] The measure of the quality of the link communities discovered by the algorithm is introduced in greater detail in Chapter 8.
[20] The level of connectivity in the system (i.e. network) at which point all the nodes in the system are connected in a Giant Component.

contrary to the popular network partition strategies usually applied to community discovery network studies. The link similarity criteria is based on two principles: [1] the two links analysed have to share a node and [2] there has to be high similarity between the other two nodes (i.e. the ones that are not shared) the links connect to. Essentially, those two nodes must be each other's neighbours by sharing other nodes in their immediate neighbourhood. The similarity between the nodes in the immediate neighbourhood of the shared node is a function of those two nodes' immediate neighbours that are shared between the two and the total number of nodes within the community. Detailed discourse on the concepts introduced in this paragraph follow in Chapters 5, 6 and 8.

Knowing the current number of edges with high betweenness is of a critical importance for any system (i.e. in this study networked infrastructure), since it allows to detect underserved parts of the complex network where these types of edges are absent. Increasing the number of such edges in these parts of the network ensures stress dispersion and resource flow optimization.

### Research sub-question [2]

**As the population becomes more dense in the city— and more nodes are added to the sociometric graph— how does the logical topology of the relational network develop?**

Objective 2: To answer the second sub-question, I empirically determined the correlation coefficient between the infrastructure systems' technical connectivity and the accessibility of places, that served as a proxy for the logical (i.e. relational) topology of local ecosystem services demand (i.e. pressure) points.

I achieved this by determining the correlation coefficients between four of the traditional node centrality metrics (betweenness centrality, degree, constraint and closeness) [21] and the new link-community based metric at the city-region scale. The outcomes from the single-link hierarchical clustering algorithm presented the relationship between the tendency of the network's behaviour once a new node becomes structural hub in increasing or decreasing the value of link betweenness in the communities it is part of.

Topological self-similarity of the logical network is a signature of optimal organization of that system and usually the lack of self-similarity as in the case of scale-free[22] networks is influenced by the central

[21] All of these measures are defined in 5.1.

[22] This concept is introduced in detail in 3.4. For reading clarity, scale-free networks are the networks that lack consistency in the number of connections their elements have with each other, thus there is no universal average number of node connections to other nodes.

units (highly connected nodes/hubs) that are different from the peripheral units (nodes with low connectivity but potentially high strategic importance in keeping the network modular).

### *1.3.2 Hypothesis*

***The structural importance of links that connect functional communities in the network is overlooked when detecting and increasing the presence of hubs. If the number of links with high edge betweenness increases, then the flows running through the infrastructure systems decentralise and distribute localised pressure points on ecosystem services.***

## 1.4 Thesis structure and chapters overview

The literature review on the theory of urban studies is organised in two parts. The first part introduces analysis of the contemporary urban form, namely the city-region and its nature, physical characteristics and sustainability concerns. The second part introduces the nature of complex systems and the basic elements of their structural topology, connecting this theory to the dynamics observed in urban systems. The introduction of Part 3 includes the literature background on the methodology, the process of collecting the datasets, defining the variables within the study scope and their measures.

Specifically, Part 1 is consisted of three chapters. Chapter 2 discusses the nature of the city region and presents relevant arguments on the lack of proper approach to its sustainable growth. Chapter 3 introduces the two rather distinct types— the 'social' and 'physical planning' models observed in the built environment which set out the substance for the overall methodological approach of reconstructing them as two sub-networks of one complex system. Chapter 4 discusses the nature of the topology (i.e. connectivity) of the built environment and the large technical systems enabling it.

Part 2 consists of two chapters, Chapter 5 analyses the theory of network science, defining the main elements of a network, their properties and the network's evolution. It offers a comparative assessment of the scale-free network models vs random models through their main similarities and dissimilarities. Chapter 6 offers in-depth discussion on the structural topology and regime shifts of complex systems and examines in-depth the structural topology of general hybrid systems, along with the application of the tools developed in network science for structural analysis of their complexity.

Part 3 reflects on the findings from the literature review of the methodological approach adopted in this research. It consists of four chapters. Starting with Chapter 7, there is brief introduction of current contemporary modelling methodologies, their benefits and shortcomings. The aim of this chapter was to critically and rigorously justify the most suitable existing algorithms that were applied in the statistical inference part of the thesis. Based on the analyses presented in this chapter, the single-link hierarchical clustering and the Louvain algorithms were selected to measure the modularity and discover the community structure of the studied complex network. This chapter also highlights the limitations of the research carried out.

Chapter 8 discusses the process of collecting the datasets, defining the variables within the scope of the study and the measures used. Chapter 9 presents the main findings of the analysis for all eight shires separately and discusses the results within the context of the study. Chapter 10 gives brief overview of the hypothesis tested in the research, the main topics studied, the results, as well as a proposition on a direction for future research focus.

# PART 1: SUSTAINABILITY OF THE URBAN FORM

## Chapter 2: The emergence of the new urban form: towards sustainable city-region

### 2.1 INTRODUCTION

Many studies indicate that [effective] compact urban form (Brelsford et al., 2017; Ahlfedlt and Pietrostefani, 2017; Song et al., 2017) along with good economic practices including waste re[up]cycling and circular industrial processes (McDonough and Braungart, 2009; Ghisellini et al., 2016) are the two key approaches to plan and build more dense and resource-cautious cities. Governmental organisations such as the European Commission are prompting sustainability of cities (Lehmann, 2016) by designing optimal compact urban forms. These design strategies are based on reducing sprawl, agricultural land protection and reuse of previously developed land. The Commission's sustainable urbanisation efforts resulted with the creation of numerous successful policies among member countries (Jenks and Jones, 2010) that promote massive reuse of brownfield land. Nevertheless, this brownfield development came at the expense of losing open unbuilt space and amenities (Jenks and Jones, op cit.). Reduced open (natural) space affects the biodiversity and the provision of ecosystem services as amenities. It also affects, as explained earlier, local temperatures and increases pollution. As a corollary, increased compactness affects social well-being of people and their quality of life (Breheny, 1997).

### 2.2 THE SUSTAINABLE URBAN FORM

Neuman (2005) in an earlier study has demonstrated that compact urban areas are not necessarily more sustainable. The study argues that the design of good urban form should be governed by the flow of various socio-economic and ecological processes. This recommendation is supported by empirical evidence (Bettencourt, 2013) that there is an upper value of the city's physical capacity, at which people come together in balanced socio-economic interactions. After reaching this upper bound, the dissipation costs are overcoming the social benefits of the central places and force centralities to restructure into a cluster (Hall and Pain, 2010). This cluster is consisted of separate but co-dependent socio-economic conglomerations, forming a city-region (Neuman and Hull, 2009).

Bounded effort— defined below— was first registered among spatially unconstrained networks such as the cell phone communication network (Schläpfer et al., 2014) or the e-mail communication network (Guimerà et al., 2003). The bounded effort measured in these networks was a function of the human constraint restricted by time/or financial investment needed to maintain a relationship and the structure and service costs of the supporting physical network. Specifically, bounded effort is a primary characteristic of humans that naturally tend to preserve energy and therefore effort.

However, more evidence exists that this phenomenon appears as well in spatially constrained networks. Bettencourt (2013) conducts quantitative analysis of a model sample of a city, built on the characteristics generalised from the observation of forty cities varying across sizes and geographic positions. This analysis is very insightful. He classifies the social interactions in the city as local and the effort to engage in social life depends on the cost and easiness associated with reaching places. This implies that there is a capacity limit of the catchment area of locations servicing a city and once this capacity is exceeded, the city becomes overrun and destinations of interest became difficult and more expensive to reach. This results with people's activities dissipating away from large cities and concentrating at secondary central places in adjacent locations; the dissolving of large, overcrowded cities (or agglomeration of several socioeconomically co-dependent cities) into city-regions (Neuman and Hull, 2009; Castells, 2010; Hall and Pain, 2010).

This new urban form is characterized by distribution of the functional centrality of the main urban core into hierarchically decentralized specialization of functions among different urban centres. "Joint processes of both central flow and central place have led to the creation of the mega-city region" (Pflieger and Rozenblat, op cit., p2731). Evidence of this transformation is registered in Hall and Pain's research (2010) with a main hypothesis that the specialization in advanced services, located in some of the old or new centres of major world cities, serves as "the source of centralization in some areas of the world" (cited in Castells, op cit., p2740). These mega cities gradually develop in "the pivotal nodes of the networked management capacity in our society" (cited in Castells, op cit., p2740). When cities like those became the pivotal nodes of several networks, they grew into powerful cities that are largely connected to many other smaller ones. These 'powerhouses' are then defined as hubs in the relational networks of global socio-political (i.e. decision-making) and economic exchanges (Pflieger and Rozenblat, 2010). These hubs (i.e. nodes with large number of connections) of the two global networks of labour and finance need to be served by 'multidimensional' spatial infrastructure of connectivity (Sassen, 2001). Connectivity by multimodal transportation and telecom networks are creating the main base for knowledge generation and attraction of highly skilled labour (Castells, 2010). A practical example is the city of Zurich, which was able to restructure itself and adjust to the new economy manly due to the supporting technical infrastructure— i.e. the intra-metropolitan transport system (Dessemontet et al., 2010). This allowed the region broader access to several markets of highly skilled labour that could now easily commute to work.

Thus, the globalization of the world's economy has followed the organization of key advanced services, rendering some cities as main economic powerhouses and major contributors to environmental degradation, affecting cities and natural habitat in their own countries (Sassen, 2001).

### 2.3 CONCLUSION

Urbanisation (Luck and Wu, 2002) is positively correlated to the dispersion of land, fragmentation of ecosystems through the restructuring of patches and the complexity of land shape (Luck and Wu, op cit.; Liu and Wu, 2016; Bosch et al., 2019). Pflieger and Rozenblat (op cit.) state that while city-regions drive urbanization, in the same time, their clusters and functional hierarchies can run opposite of sustainable development of urban settlements. The main reason for this according to Wheeler (2013) is that this newly emerged urban form of a city-region lacks a coherent approach to sustainability supported with a framework and metrics. This is grounded in two macro dynamics.

First, the key features of this new urban form are the networking and dispersal of people and activities that enhances the growth of various centres in a hierarchical order of functional specialization (Castells, 2010). As a corollary, people increase their traveling, communication and consumption patterns on the expense of the environment. The strong positive relationship between multi-nuclei urban form and socio-economic dynamics, and its environmental impact can be observed in this process.

Second, the communication infrastructures and transportation routes become the 'nervous system' of the polycentric metropolitan region (Neuman, 2006; Neuman and Hull, 2009; Castells, 2010; Hall and Pain, 2010). Economic activities together with diversification of urban centrality functions are evidenced to be decentralized, mainly along the transportation lines, which before were a mixture of highly dense residential areas only (Hall and Pain, 2006). This results with some areas of the region to have better access to this road network and therefore became more in demand, rising the standard of living and creating social inequalities and spatial exclusion of remote, less desirable communities elsewhere (Lorrain, 2005).

# Chapter 3: The social and the physical models

Modelling became attractive in the discipline of urban planning by holding great promise to support decision-making in the spatial arrangement of land use and infrastructure, while superseding the more conventional approach in organizing cities through the virtue of architectural and aesthetic skills. The vast application of linear models of cities was becoming mainstream in the 1960s (Batty, 1983). These models, based on gravitational and potential theories, presented a cross-section of the city at a point in time with its relationships between the mobility flows that were simulated as inversely proportional to travel costs or distance between the “origin” and “destination”. In this way, the models were very static and aggregated, unable to reflect the diversity and heterogeneity of contemporary cities emerging through individual micro dynamics (Batty, 2005b, p8). In the late 1970s and 1980s a shift appeared in the building of models that captured the bottom-up micro dynamics and transformed modelling from hypothesis testing to frameworks that facilitated formal and informal dialogs. Nevertheless, this new modelling became very intensive in data and computation, which revealed issues of model parsimony, triviality and quality.

However, in recent years urban modellers have witnessed the beginning of a ‘data-rich’ era (Kitchin, 2017), that is characterized with unprecedented availability of large, fine-grain datasets next to advances in computing power (Heppenstall, 2016). The emergency of urban science (Kitchin, 2017) is a response to these developments. Scientific inquiries in this novel field include mapping and modelling the urban *dynamics* (Kitchin, 2017, p2) represented by the *flows* and *processes* in cities. The scientific outcome of this emerging discipline is the production of new theoretical insights that support the science of cities (Kitchin, op cit.). For example, Hellervik et al. (2019) showed new way to infer the distribution of urban economic activity. Their original ‘preferential centrality’ measures are based on the traditional gravity model of the PageRank algorithm[24]. The new way of integrating urban and transport models by developing new network measurements seems to be able to respond to the parsimony criticism of the classical gravity models. The following literature review on the main theoretical framework in urban studies, develops the main definitions of the fundamental elements that are the building blocks of the complex network studied in the thesis and explained in detail in Chapters 7 and 8.

## 3.1 THE “TREE”

Urban studies in the period of the mid 20 century were grounded in trying to understand the social nature of the city and the use of space by its residents in an analytical manner, with the goal to inform new development and arrangement of land use. For example, Alexander (1965) conducted analysis of top-down[25] developed master plans developed by various architects of nine metropolitan areas that inherited the structure of what he calls a ‘tree-like’ spatial design. His results concluded that neither the original blue-prints of Columbia, nor the one of the Greenbelt (both in Maryland, USA) correspond to the ‘social realities’ of the respective cities. Their physical layout plans suggested ‘tree-like’ hierarchy of institutions ordered by district zones. Houses of each district linked via the street network

[24] It is an algorithm originally introduced to rank the relevance (i.e. number of unique visitors) of webpages based on the number and quality of links they have to other webpages, which are evaluated equally.
[25] To a great extent organized and designed through the virtue of architectural and aesthetic skills.

to a neighbourhood centre to which they were (perceived by the planners) to be 'functionally' depend on. Further, the different neighbourhood centres were linked in accordance to their functional dependence, to a village (or district) centre comprised of core institutions— for education, medical care and other public services. Thus, these various village (or district) centres were at the top of the hierarchy directly linked with the principle cluster— the central urban district (CUD) — which typically, contained the chief business and governmental institutions.

However, this kind of organization in the city illustrates clustering of stronger and stronger isolated social groups and segregated communities. This is contrary to today's modern societies, where virtually closed social groups do not exist. One of many examples is the spatial layout of the city of London where the individual neighbourhoods— boroughs and wards— have no reality as self-contained functional units (i.e. link communities). Alexander (op cit.) concluded that almost no person living in a city of the size of London or even smaller, manages to find work within the area they are living in. Recent evidence supporting this notion are the 3.7 million workers in the UK that commute for over two hours daily to their jobs (Gayle, 2016). This implies that this type of deterministic planning and zoning, that is based on 'tree-like' patterns predicting that people will meet at one (or several) central point(s) to do business or enjoy leisure time, are not validated. The social transformation of how people use contemporary cities can be observed in the recently popular forms of remote working and the resulting digital 'working nomads' (i.e. people working remotely from home or anywhere else in the world) or online shopping and other social activities.

Similarly to Alexander's findings, another case study of the redevelopment plan of a smaller English town did not register a 'tree-like' social structure in the town's neighbourhoods functional dependency (Glass, 1998) either. The social structure discovered in British Middlesbrough was contrary to the 'tree-like' design of zones for designated uses— i.e. zones that are physically connected with one another through the hierarchy of social infrastructure and services. Namely, Glass (1998) conducted socio-spatial analysis of 29 neighbourhoods that accommodated 200.000 inhabitants. She studied whether the 'social infrastructure'— i.e. the relationships of people to places of interest— follows the physical infrastructure of the neighbourhoods they conducted their daily activities in. Her findings concluded that the two 'different infrastructures' are not coherent with each other. Although they are inter-dependent, they follow different dynamics. While Glass (op cit.) did not explicitly create the dichotomy between the two networks (i.e. the physical and logical) observed in the city, the general conclusion of her study reflected disparity between planned locations of services that rendered the neighbourhood as a single independent functional unit, and the actual services individuals used on daily bases, often located in other areas.

The conclusion from the two studies Is that the links between individuals and their places of interest observed at the scale of the entire city (i.e. including all inhabitants) are not bounded by the administrative neighbourhood's limits (Alexander, 1965; Glass, op cit.). People from different neighbourhoods tended to visit 'functional centres' of other neighbourhood; in that way creating what in graph theory is known as *overlapping communities* (Newman and Girvan, 2004; Guimerà, 2003).

These overlaps in *graph theory* are known as the vertices and edges with high *betweenness centrality*[26] (Barabási, 2016) that decrease the graphs' diameter and therefore improve the global efficiency of the entire network. The disparity between expecting individuals to visit planned central places of a local community and the actual places of interest visited by these individuals across the same community can be mathematically illustrated by *a lattice.* In graph theory the same representation is called *a graph*. Sub-chapters 3.4 and 3.5 define the main concepts in graph theory and their application.

## 3.2 THE "NETWORK"

From the literature review (Alexander, 1965; Moudon, 1997; Evans et al., no date; Portugali, 2011; Song et al., 2017) studying urban morphology, I concluded that scholars trying to understand public space and the public realm, strive to map and treat elements of the city together with people and institutions. Such representation of the spatial and temporal elements as part of the same network fails to capture the complexity and diversity of urban forms in contemporary cities. This emerging complex form embedded in both physical and 'virtual' (which throughout this thesis is called logical) space is defined by the "multiple overlapping interactions between households, firms, governments and other agents impossible to be assessed by a single point in time approach" (Bosch et al., 2019, p2). These overlaps define the functional dependences of the social processes embedded in Euclidean space of physical constraints. However, large number of studies (Snijders, 2001; König et al., 2007; König and Battiston, 2009; Casetlls, 2010; Bosch et al., 2019) undertaken by network scientists studying social and economic networks in and between cities, show that the two types of elements (streets and buildings vs. functional relationships between people and places) are not necessarily part of the same network (Boccaletti et al., 2014). They both are interdependent networks (also called *layers*) that together create a larger *multidimensional network* (Boccaletti et al. 2014, p 9) representing the *hybrid* human-nature system (An et al., 2014; Alberti, 2016).

Thus, from the literature review I establish that there is a main difference between [1] physical networks in urban planning terminology better understood as *networked infrastructure* (Graham and

[26] A concept introduced in Chapter 5.

Marvin, 2001) and [2] socio-economic networks represented by the various social and business interactions elicited between social groups or other institutional entities in cities.

The measures of nodal and edge betweenness centrality (both explained in 5.1.2) are evidenced in recent studies (Rosenthal, 1960; Girvan and Newman, 2001; Holme et al., 2007; Onnela, J.P. et al. 2007; Blondel, et al. 2008; Lin, Y.-R. et al., 2009; Guimerà, R. et al., 2003) analysing various networks reconstructing information-based interaction or functional dependence and collaboration (e.g. phone calls, e-mails, affiliations to biological processes). Janssen et al. (2006) studied the importance of betweenness centrality in socio-ecological systems. In the world's mega-regions that house globalized trade, international markets and remote labour with their respective distinctive micro-dynamics (Castells, 2000), the access to universal services (i.e. accessibility to places of interest, water and solid waste management, water and energy supply etc.) is further challenged. Detailed and comprehensive empirical analysis at the level of Large Technical Systems (LTSs) that explores the relevance of node betweenness or centrality and edge betweenness, in networked infrastructure is needed to better understand the physical properties of the networks' interactions and the way their underlying topologies influence each other.

## 3.3 CITIES: SOCIETY, URBAN FORM AND INFRASTRUCTURE

In its most basic form, the central role (Lorrain, 2005, p 15) of the built environment is to sustain human activities (individuals, organizations and institutions) and it does so through its fundamental components— the physical infrastructure and buildings. The basic morphological structure of the urban form is arranged around a hierarchy of levels that mutually connected compose a functional agglomeration. Morphology of cities is the study of the hierarchical ordinance of different sub-centres around key economic functions and this form defines most of the complex human interventions (Moudon, 1997; Luck and Wu, 2002; Batty, 2008). The theory of morphology in its most elementary level is based on the principles of *form*, *scale* and *time*. Hierarchies in urban planning are basic methods of representing spatial organization based on the level of importance of urban functions across multiple scales (Batty and Longley, 1994). Their application spans from systems of elements within cities (Batty and Longley, op cit.; Batty, 2005a; 2008) to systems of cities (Batty, 2006; Castells, 2010). The hierarchy of levels in cities is engineered by the organizing principle of urban elements—street/block, plot series, plot, building, cell and structure—and these levels are not interchangeable (Luck and Wu, 2002; Ahvenniemi et al., 2017). Such hierarchy of arranging elements in graph theory is known as a *tree (or centralized) network*.

Urban sociology, with the exception of Ildefons Cerda's plan of Barcelona, came later to the attention of planners and spatial geographers. Urban sociology highlights that cities are not isolated bodies with functions. They rather represent a complex support system for human interactions comprised of many distinct networks (Jacobs, 1961) — i.e. economic, social, political and technical. These networks intersect at a given point creating urban spaces with their own specificities (Dessemontet et al., 2010). The technical networks comprise the urban infrastructure, a fundamental part of the urban form. Infrastructure goes beyond physical support of buildings, enabling their transformation through the conveyance of flows and therefore its role, rather than exploitational, is in fact transformational and empowering (Neuman, 2006).

> *"Infrastructure is the physical network that channels a flux (water, fluid, electricity, energy, material, people, digital signal, analog signal, etc.) through conduits (tubes, pipes, canals, channels, roads, rails, wires, lines etc.) or a medium (air, water, soil) with the purpose of supporting a human population, usually located in a settlement, for the general or common good." (p6)*

Neuman (2006) extensively assessed the nature of infrastructure as an umbrella term defined differently by the various practices in the public and private sector. His research makes two main contributions. First, he recognizes infrastructure as a long-lasting network, typically over several hundred years, connecting producers, service providers and large number of users. According to my analysis, the longevity is the fundamental property that differentiates the topology of spatial physical networks to the one of relational networks. Social and economic interactions among actors have short-term nature defined by time and space (Batty and Torrens, 2005)— lasting between couple of months to several years. Second contribution is that the mediums (i.e. air, water and soil) are acknowledged as integral network components. Although the overarching term defining infrastructure refers to many infrastructure networks, which Neuman (2006) categorizes under six main types— utilities, public works, community facilities, telecommunications, transport, and knowledge networks— he distils fifteen main characteristics these infrastructure networks share. These categories do not exist in isolation but interrelate with one-another into networked infrastructure (Graham and Marvin, 2001), creating synergies to support the functions of human settlements (Neuman, 2006; Neuman and Smith, 2010, Derrible, 2017).

Derrible (2017) explains that exactly this integration can assist in creating more decentralized, or in the author's own words '*semilattice-like*' infrastructure. Decentralization of a network in network science is observed by the network's edge betweenness[27] or betweenness centrality[28], which results with e.g. the small-world property of a network. Decentralization is a direct measurement of the network's efficiency (Latora and Marchiori, 2001). The more the network becomes decentralized, its efficiency grows. Both network properties, high level of edge betweenness and node betweenness centrality increase network decentralization, while eliminating redundant links or hubs.

## 3.4 LATTICE: ELEMENTARY MATHEMATICAL FORMS

The more technical sciences such as infrastructure engineering and resource management or quantitative geography, used many linear, non-linear, static and dynamic mathematical methods and models to study urban systems and analyse the social interactions and exchanges within them. They do so in order to undertake planning, design and optimization of these large technical structures that support cities. The methods and models are rooted in the sub-fields of pure and applied mathematics that use order theory, set theory, combinatorics and probability theory (among the most prevailing ones) to explain relationships between elements in bounded or unbounded abstract space. The *lattice* as a main abstract form is used by these sub-disciplines to visualise the relationships between elements. Since the research deals with multiple complex and interlinked subjects of observation, the lattice defined herein represents the back-bone and unifying approach of the research enabling consistency in the scientific inquiry of the growth and operations within cities as complex systems.

*Lattice* (or *graph*) is an abstract form used in order theory (mathematics) and graph theory to represent repeated arrangement of points (i.e. elements) and their relations in a network (Latora and Marchiori, 2001, p 1). Depending on the arrangement of relationships between the points, lattices can take many forms. The representation of a *tree* structure (Fig. 1) indicates that two points called *vertices*, or in network science known as *nodes*, are linked to form a pair by exactly one path to vertices (or nodes) in the level below or above the hierarchy. This attachment follows clearly defined and ordered (preceding or superseding) relationship. The *tree lattice* is abstract structure of a connected graph with no cycle and can be ordered or unordered. It usually has one designated vertex (or node— both terms used interchangeably) which is called the *root* (Nakano, 2014). The unrooted tree lattices and their characteristics are beyond the scope of this research and are to be revisited in a later study.

[27] Edge betweenness measures the number of shortest paths (in a step-length) connecting pairs of nodes that pass through that edge.

[28] Betweenness centrality of a node measures the number of paths (in a step-length) that pass through that node to connect pairs of nodes. Nodes with high betweenness centrality in network science are called hubs. Both measurements are explained in detail in Chapter 5.

The representation of a *semilattice* in order theory indicates a partially ordered set in which any two elements have a unique upper bound—join—and unique lower bound—meet (Fig. 2).

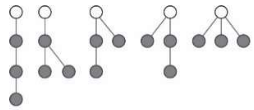


Figure 1: Rooted ordered tree lattice (Author's illustration)

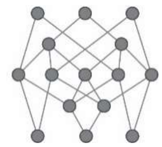


Figure 2: Semilattice(Author's illustration)

*Regular* lattice is a representation of a grid-like network, *semilattice* represents the interaction among nodes in a random network, while *inter-linked stars*— also known as a *hub-and-spoke/or scale-free network* (Fig. 2 right illustration) is represented by linking together several *tree networks* at their roots.

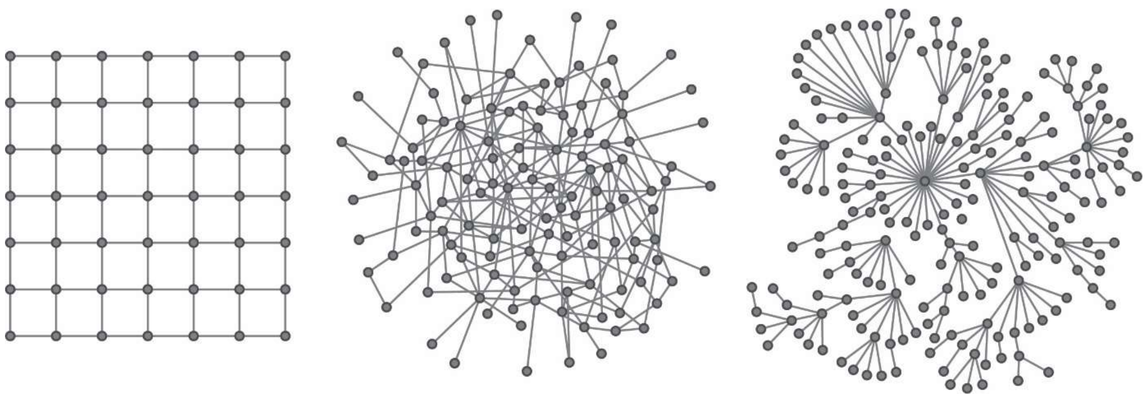


Figure 3: Three typical lattice representations: from left to right—regular, random and inter-linked stars (hub-and-spoke) (Author's own illustration)

A *hybrid* (*multilayer*) *network* is a network constructed of two or more networks co-dependent with each other (Janssen et al., 2006). These networks individually are not isolated and closed-off systems on their own, but represent layers within the hybrid network (Latora and Marchiori, 2001, p 4) and in some cases; they share nodes or links among each other (Boccaletti et al., 2014). Such an example in the built environment is the multi-modal transport network where trains, busses and cycling networks are representing the respective layers of the hybrid transportation system. A person using several modes of transport to complete a single trip (e.g. bicycle-train-bus-bicycle to reach point B from point A) forms a [functional] community within this hybrid network. The links in this community are usually different type of infrastructure components (e.g. cycling lane/railway/road) and so are the nodes (e.g. bike-share point/train station/bus station/bus stop). In this case the train stations can be co-located with the bus station and the bike-share hub, thus the three respective sub-networks are sharing theis 'node'. Additionally, the biking infrastructure and the bus-road network can be sharing links, e.g. by having their lanes being allocated along the same street.

## 3.5 GRAPH: BASIC COMPONENTS AND TECHNICAL APPLICATION

### *3.5.1 Defining a graph*

A *graph* in graph theory is a set of objects that graphically represents the relationship of the objects—meaning some pairs of objects are in some sense related. The Swiss mathematician Euler (1735) introduced graph theory in 1735 to visually represent the well-known mathematical problem— *The bridges of Königsberg*. This visualization helped to explain that the existence of a path does not depend on the ingenuity of the problem-solver to find it. Its existence is rather the property of the graph representing the phenomenon under study. Thus, the ability of a traveller to find the path between two places does not affect the physical existence of paths (i.e. bridges) between them. Between 1956 and 1968, the mathematicians Pal Erdős and Alfred Rényi published series of papers on a model merging probability theory and combinatorics, with graph theory. Their result was the introduction of the *random graph model.* This model defined the theory used frequently in the design of infrastructure networks in the decades that followed.

As previously explained, the graph's main components are nodes (vertices) and links (edges). For the purposes of the present study, which applies these components to define the new element of "link communities" within the theory of urban studies, these components are defined as follows:

[1] *Node* of a network is the single component in the network that is [not] connected to the network. For example, in a physical computer network such as the Internet, a node is one computer or a router, in the case of the digital network of the world wide web (WWW) it is a single document. In the co-acting network analysis of the movie stars, a node is one actor. In the cases of more complex networks (Boccaletti et al., 2014) such as the power grid, a node would represent several different entities— a power generator, a switching unit, a transformer, a low-voltage line or a consumer, etc. Nodes that have many connections to other nodes in the same network are called *hubs.* Hubs are of special interest to study in regards to network robustness, community formation and degree correlation. In most real networks studied so far, any new node that 'joins' the network under the *preferential attachment* property is more likely to connect to one of those hubs. Those highly connected nodes are representing a signature of a deeper organizing principle in networks which exibit a *scale-free property.* This property is encountered in most real networks such as the WWW, actors network, the Internet, the citations network, e.coli metabolism, protein interactions, linguistics network, electric circuit, publication co-authorships network, citation network, phone and mobile calls, e-mail, software, energy landscape, Twitter and Facebook (social networks), few among many others.

[2] *Link* is the representation of the relationship between any two nodes in the graph. In the respective examples of networks, the links are the physical connections of for instance copper or optical cables that connect routers between different locations, the URLs that connect the different web pages, a relation of an acquaintanceship if the actors acted in the same movie. In the complex network of the power grid, the links are the transmission lines and the weight is their carrying capacity of energy distribution. Links in a network can be directed or undirected, which may not have implication in the case of the WWW, the actor network, or the power grid, but may have big implications in the water supply network, the rainwater discharge system or the waste disposal network.

By using probabilistic mathematical modelling, random models attempt to prove the existence of graphs (i.e. connected objects) in real-life networks that follow certain principles of adding/removing nodes and/or links. According to Barabási (2016), the random network as a (graph) theory has existed for a very long time despite its poor agreement with reality. Although it was originally invented to serve as a *model for designing real systems,* random networks are regarded as the simplified replica of the evolution of some physical network infrastructures — i.e. the railway, the road, the electricity etc. The main shortcoming of network simulations using linear or non-linear optimization modelling such as the location-allocation problems[29] projected on a random graph, treated flows and processes as a secondary problem (Scott, 1971, p 95) in network optimization. The main argument of this thesis is that the flows conveyed within the networked infrastructure constitute the logical topology of these networks. The thesis proposes new approach in using statistical modelling tool for prediction and optimization of their flows. The proposed approach analysed decentralisation of the networked infrastructure and its design, aiming to improve sustainable sourcing of resources, their efficient transfer throughout cities and optimal consumption by end users.

### *3.5.2 Defining the topology of a network*

[3] *Topology* explains the principle under which the nodes are organised (linked together) in a connected network and can be *physical* and *logical* (Cleveland State University-College of Business-DBA-CSI, unknown; Naimzada et al., 2009). Topology of a network is a principal concern of network design and optimization problems and in modern technical practices, the study of network's topology finds application in many areas of living. *Physical topology* refers to the principle of organizing the components (nodes) of a network in a physical connection (e.g. the physical link between two cities

[29] The location-allocation mathematical problems (in depth explained in 8.3.2) arise in many sorts of practical planning situations such as planning of geographical systems for instance hospitals, schools or transport planning and optimization of the movement of people, goods and resources.

can be a railway or a motorway, or the physical link between the water distribution centre and the households are the supplying pipes). *Logical topology* of a network represents the process (the flux of matter or items) in the network (e.g. the way resources travel through the food or the water network, the way cars utilize the road network). The physical and the logical topologies of the graph representation of the complex network are the main aspects of interest in this thesis.

The logical topology in a number of studies (Snijders, 2001; Albert and Barabási, 2002; König et al., 2007; König and Battiston, 2009; Tan et al., 2018; Bosch et al., 2019) published over the past few decades has been established to exhibit properties different than the ones of a random graph— e.g. scale-free or small worlds. The hypothesis posed herein argues that this topological differences (i.e. the physical topology vs. the logical topology) within networked infrastructure, result with unsustainable consumption of ecosystem services and lead to the general unsustainability of infrastructure systems. For example, according to Barabási (2016), most nodes in random graph models (used at large to plan and design infrastructure networks) have similar node degree[30] and detached nodes from the network [usually] do not exist. In the context of services that support liveability of any globally connected city— the principal concern of this study— the technical design of critical infrastructure systems (e.g. 'global' agricultural or waste handling and disposal networks) often are not designed to include weight optimization[31] problems of remote nodes that act as a source or a sink[32]. In location-allocation optimization problems, these remote nodes (if ever considered!) during modelling, are highly unlikely to appear as connected to the rest of the network. This occurs for two reasons; the used algorithms and their probability rules predominantly distribute links to the nodes randomly, so the nodes are randomly connected. The second algorithmic logic is the uniformed probability and relative short step-length distances applied in the modelling process that would not include remote places as part of the network. We have established this earlier in this chapter. Due to the short step-length constraint, remote nodes that have long step-length distances and usually only one link to the rest of the network, are modelled as 'disconnected'. This remains a limitation for the presented study as well.

The central argument is that physical properties of infrastructure networks are different from utilization of resources (i.e. flows), which are reconstructed through the logical topology of the same network. While the earlier concerns the physical connectivity of points (places), the latter concerns the socio-economic dynamics (flows) that take place along these connections. The flows are the logical

[30] A degree of a node is the number of links that node has to other nodes in the network.
[31] Node weight refers to the maximal capacity of a node to source or accept a flow.
[32] Here the source refers to source that induces a flow. For example in the food network, a parcel of agricultural land in Spain supplying food to Denmark which would be the sink of the flow.

layer conveyed by physical networks and this layer aggregates and dissipates *functional* communities governed by laws different from the laws followed by the *hierarchical* communities of their physical counterpart.

## 3.6 CONCLUSION

Both, urbanisation (Choay, 1996; Luck and Wu, 2002) and sustainability of urban settlements (Neuman and Churchill, 2015) are continuous processes which will exist as long as liveability on this planet is possible. Therefore, the comprehensive literature analysed in this chapter recognised the urgent need for undertaking more refined empirical research focused on the relationship between sustainable use of natural resources in cities (*represented by the logical topology of flows*[33]) and *the physical topology of infrastructure systems* (Lorrain, 2005). Empirical study of the relations' dynamics between urban actors in a combination with e.g. multiplicative growth (i.e. exponential growth) models can reveal many potentially important internal structural micro-dynamics in the networks of Large Technical Systems. These structural dynamics are not able to be observed with the macro-level models (e.g. the edge-city model or the Fujita-Ogawa model) developed around location theories (Andersson et al., 2006; Barthélemy, 2016) and urban economics (e.g. the Alonso-Muth-Mills model) (Barthélemy, 2016, pp 201-214) or the Gibrat's law (Gabaix, 1999). The inability of these models to register fine-grain dynamics is due to their global and deterministic approach of explaining complex micro-scale urban phenomena using macro-scale constants and exponents. Tools of network science for geographical analysis enable modelling the topologies of communities (as smallest co-dependent units) of complex urban networks.

# Chapter 4: Topology of the urban form: land use and Large Technical Systems (LTSs)

## 4.1 INTRODUCTION

To study the relationship between the physical topology of LTSs and the logical topology of their flows induced by socio-economic processes, we need to make a fundamental dichotomy in the structure of networked infrastructure. Nodes and their links have both spatial and relational dimensions. This results with a main distinction between *spatial physical networks* such as the ones that form parts of the physical infrastructure and *spatial relational networks* such as the ones that represent interaction between agents– people or institutions (Barabási et al., 2000; Crucitti et al., 2006; Song et al., 2017).

[33] *Logical topology* of a network represents the process (the flux of matter or items) in the network; the way resources travel through the food or water network, the way people, cyclists and motorised traffic travel through the road network.

This study considered that flux of a matter is a spatially explicit relational network, in which the relational link is represented by any functional dependency of agents or items, regardless of the actual existence of a physical link.

Spatial physical networks have spatial structure (Ducruet and Lugo, 2011; Ducruet and Beauguitte, 2014), and both nodes and links are embedded in three-dimensional Euclidean space. The fourth dimension (the temporal dimension) is also considered as relevant to these networks, capturing the changes emerging in their topology and growth — i.e. adding or removing nodes and links. Thus, physical networks are subordinated to number of constraints imposed among other by the geographical landscape. When the nodes and connections have a spatial dimension and are dependent on the geography, many links between nodes may not be 'allowed' to emerge (Buhl et al., 2004) due to various constraints related to inefficiencies (e.g. capital cost, running costs or time constraints). The results are long-range (Ducruet and Lugo, 2011) and expensive connections, causing structural limitation, connecting highly clustered communities with each other through a single link (Guimerà et al., 2003, p2). Thus, efficiency of the network by increasing the edge betweenness is conditioned by the physical landscape. I.e. the possibility to establish new link with high betweenness to increase the network's efficiency is determined by the geography in which the network is constructed. As a corollary, the edge betweenness of these networks is very low and there is no emergence of the *small-world behaviour* or the *scale-free* degree distribution (Buhl et al., 2004; Crucitti et al., 2006; Li et al., 2006; Naimzada et al., 2009).

The presence of edges (links) with high betweenness— which network property is named by Barabási (2016, p 334) as *link betweenness*– is a signature for the presence of *community structure* in complex (multilayer) networks, indicating that these networks have some kind of self-organised flow optimization structure. This thesis combines edge betweenness optimisation with the *maximal steady state flow strategy,* which in the case when the network's capacity is ignored, I found them to be interchangeable. The outcome is a novel analytical approach that empirically evaluates the networked infrastructure modularity, and therefore robustness as systemic access of the nodes, by measuring the nodes closeness in the system. This new evidence can help decision-making in respect to increasing networked infrastructure efficiency to make resources consumption more sustainable. The maximal steady state strategy as a planning model of networked infrastructure is discussed in detail in section 7.3.3.

However, the distance between any two geospatial entities recreated with a graph, is typically not represented by the simplified[34] Euclidean distance (Nourian et al., 2018) but a geodesic one. Euclidean or metric distance is the straight-line distance between any two points in a three dimensional space. The geodesic distance[35] in graph theory represents the number of links that connects two nodes with each other. Therefore, geodesic distance of any node pair in a network is a summation of the number of links that a path connecting the two nodes goes through. One way of representing the actual physical length unit of each link of nodal pair is to add it as a weight to that link. This variable is important when setting up constraints to the both physical and/or the logical links in solving optimization problems (see e.g. sub-chapters 8.3.5.2 and 8.3.5.3).

Spatial relational networks involve nodes embedded in physical space, maintaining a virtual link with each other that is not affected by any geography. Empirical findings concluded that in spatial interactions (or relational networks), the spatial proximity only affected the network's spatial structure (physical topology), but not its logical topology (Illenberger et al., 2013). The logical topology depicts functional dependence between nodes, same as the functional dependence characterizing the communities in complex networks. Unlike topology of relational networks, for (near-planar) physical networks, topology refers to their physical embedding in Euclidean space, same as the 'spatial structure' (Ducruet and Beauguitte, 2014) of relational networks.

The relationships between the nodes in spatial relational networks are usually represented by a *sociometric graph* (Scott, 1971, p96). A sociometric graph (or in short sociogram) is a tool used in sociometry to study the social aspect of space (Hillier, 2007). The method is used for the analysis of choices and preferences within a group, representing their interaction's structure and patterns. Maintaining a virtual link is associated with non-economic costs (e.g. time or effort) that always have to be lower than the benefits of keeping that link existent (Guimerà et al., 2003, Bettencourt, 2013). According to Guimerà et al. (2003), this constraint is responsible for the emergence of the scale-free property of social networks. The nodal betweenness centrality (i.e. the existence of hubs) has vital structural importance in these networks and is responsible for the appearance of the scale-free property. Nodes with high centrality (i.e. popular nodes) represent the points of coordination and control of the network, but may also reduce the network's diversity. For example, in the case of sharing an information, when too many of the nodes in the network are hubs some of the hubs could become redundant. In the context of news updates, this is desirable as it is efficient and covers the entire network, but in the case of innovation, it is obsolete since the same information reaches the

[34] The more accurate geospatial distance between any two geospatial entities is usually not a straight line.
[35] The geodesic distance is also known as a step-length distance explained in 7.4.2.

nodes from multiple paths. Hubs also make networks more vulnerable under targeted attacks (Albert et al., 2000) and affect the system's robustness (Alberti, 2016). Natural systems do not have a central control, but a decentralized one— and the decentralized control is an emergent property of the interaction between the system's components (Garnett, 2018, p 687).

[Nodal] Betweenness centrality, edge betweenness, small-world and scale-free properties are defined and further explained in Chapter 5.

## 4.2 LOCATION THEORIES AND EARLY MODELS: MAPPING THE TOPOLOGY OF CITIES

To understand the contemporary topology (i.e. morphology) of cities, we need to understand the evolution of urban settlements and the theories that guided urbanisation of their growing population. Since the study will focus on Scotland, I briefly accounted the major historical developments that in recent time guided the urbanisation process in Europe.

The first modern era book on urban planning is the work of *Ildefonso Cerdá, Teoría general de la urbanizació*n (Choay, 1996). In his work, borrowing from physics the notions of flux and inertia, he announces the new age of universal communication that will change the role of spatial proximity. He recognizes that the new modes of communications will allow dispersed modes of settlements, transforming the patters of urbanisation of ancient cities. The functional definition of urbanisation (the term is coined by Cerdá) is a generic description relevant for settlements of all sizes:

> *"...Urbanisation resides in nothing other than the relation between <u>rest</u> and <u>movement</u>, the space that accommodate human repose and those that facilitate movement, that is, buildings and the network of streets..."* (Choay, op cit., p 237, Dupuy et al., 2008)

To this day, these two basic concepts remain the operative axes of urban planning.

The traditional theories of land use and land transitions in spatial economics are mainly based on the *location theory,* introduced in the mid-19$^{th}$ century by the inaugural work of the German economist Johann Heinrich von Thünen in his book *Isolated State* (Batty, 2005a; Portugali, 2011). *Location theory* is a group of theories that resulted from a combined effort of economists and settlement geographers to include space as a variable in the political economic models of that time. It explains the patterns of land use as a very simple contest of space market models, where land is valued according to the ability of the user to pay. In these spatial models, land use designation tends to follow radial hierarchical and mainly concentric organization of space around key commercial and industrial centres, which results with hierarchical morphology (of importance) of the physical space. The location theory was a guide for industries to reach strategic decisions of optimal location for production plants, reducing transport time and costs between their market and primary resources. The location theory ultimately treated cities as systems that eventually reach an equilibrium state, positioning them as [1] central places of their agricultural fringes; [2] mediators between their agroscapes and other cities; and [3] hierarchical systems of central places.

Although Weber's (1929; 1971) *industrial location theory* was developed as a complementary to the one of von Thünen, it actually mirrors it. Weber's theory was not explicitly developed as an urban theory, yet includes urban components and it is implicitly related to three spatial properties. Developing multiple tools from his location triangle concept, he studied several phenomena, among which are the agglomeration of industries and the impact of labour. With this, he could explain the consequences associated to the industries' tendency to move from their optimal location to another agglomeration, a city or even a country. Following the industrial revolution and the age of globalization, the major attractors of urbanisation in Weber's theory are: [1] the location of industries, [2] the phenomena of agglomeration [economies] and [3] the migration of industries to inexpensive labour countries.

The [industrial] location theories supported the emergence of many urban economic models (Lee, 1973; Gabaix, 1999; Benenson and Torrens, 2004; Glaeser, 2008; Ellison et al., 2007; Ponzetto, 2012; Power, 2014). They all treated cities as systems that eventually mature and reach equilibrium (Glaeser and Gottlieb, 2009) and the spatial structure (the morphology) and not their dynamics, had the dominant focus (Batty, 2005a, p 19). Approximating a system to an equilibrium state (Benenson and Torrens, 2004, p 52; Batty, 2005a, p 31) is the same as regarding it as one without any or with a very small change. Yet, there are evidence of fundamental similarities in the behaviour between ecological and social systems in cities, which indicate that cities are constantly evolving and changing (Berry, 1964; Bourne and Simmons, 1978; Lambin and Meyfroidt, 2011).

In the seventies of the past century, by analysing the abrupt changes of urban and regional events, many geographers and urban economists attempted to incorporate dynamics into the static economic models (Dendrinos and Rosser, 1992). Their work to a great extend was based on the models developed by Alonso (1964) and Wheaton (1974). This attracted much attention to the issue of land use and discontinuous urban growth (or decline) from a macro-level perspective. The models such as the *Casetti's model, the Papageorgiou model* and the multiple *Dendrinos models,* included the possibilities to simulate cyclical change and a discontinuous change of the city size. Thus, in this period, continuities and discontinuities were at the core of understanding urban evolution.

Dendrinos and Rosser (1992) advanced the thesis introduced in successive papers in the 1980s to a comprehensive dynamic theory of the discontinuities in urban population size. They argue for their extended theory to be consistent with the notions found in Darwin's evolution theory. That is, evolution takes place when discontinuities occur and as a gradual and time-consuming process, the main urban dynamic forces are either *endogenous* forces based on the internal net growth or decline of an area (e.g. internal agglomeration effects or generated innovation) or *exogenous* (e.g. trading of final or bi-products or natural resources). Especially, the growth of the long-distance trade is a major contributing factor to urban growth (or decline) (Dendrinos and Rosser, op cit., 139). Thus, they add the spatiotemporal comparative [dis]advantage as the fundamental dynamic factor, constantly changing in time and space, to invoke the endogenous or exogenous forces. Precisely the comparative advantage resembles the one factor found in Darwin's process of evolution through natural selection— *fitness* (Kauffman, 1993).

The new global (i.e. aggregated) models emerging in that decade included land use transition, a reversible process— object of study in gradient analysis (McGarigal and Marks, 1995; Luck and Wu, 2002), as a part of an open and complex human-environment system (Neuman, 2005; Portugali, 2011; Neuman and Churchill, 2015; Alberti, 2016). This complex system was for the first time acknowledged to include predominantly long distance flows of commodities, capital and people (Lambin and Meyfroidt, 2011). The long distance flows, due to external trade relations, in the models were relying on the comparative advantage of the region, which in turn depended on the regional transport costs (within, towards- and outwards- the region itself). However, the discussed global models only speculatively included the changes in comparative advantage such as technology innovation or rates of resource exploitation, that affects labour demand and immigration. Moreover, the everyday dynamics of the social and cultural realm, as potential forces that can aggregate to act globally and influence urbanisation at large, were entirely overlooked. The influence of these micro-dynamics was rigorously and empirically proven later on, through the development of local spatial interaction models (that first appeared as non-spatial ones).

What follows is a brief overview of the spatial interaction models and their evolution from static to dynamic urban economics models.

### *4.2.1 Global spatial models*

It first came to the attention of urban economists to study the city as a proximity of people that based on its productivity, has specific density that is a function of the income variation and the housing prices (Glaeser and Gottlieb, 2009). However, the limitations of this approach was that investigators developed the field's central tool around spatial equilibrium, which predominantly guided many urban models of housing prices and agglomeration economies (Lambooy, 1998; Rosenthal et al., 2003; Glaeser and Gottlieb, op cit.), and much of the empirical work done on urban growth and the urban wage premium (Bettencourt et al., 2007). A particular set of theories (Krugman, 1991) of agglomeration economies focuses on the benefits of reduced costs while moving goods across space, or the colocation of firms near other firms that are their costumers or suppliers. These theories are predominantly reflected in the new zoning planning practices widely used in organising contemporary western cities. The gravity model developed by Marshall (1920) and supporting empirical evidence introduced in economics (Isard, 1954) embraced modelling the attraction of firms as a function of the robust market demand and the continued importance of transport costs for goods, people and ideas (cited in Ellison et al., 2007). Following the industrial boom and the post-WWII reconstruction period, this became the predominant way of organising urban land to accommodate people and services. Soon enough, the zoning practice triggered global acceptance which translated into 'Westernization' of cities across the world, regarding it as the most efficient way of organising urban life.

The generality of the early developed global spatial models were heavily criticized. Namely, Glaeser and Gottlieb (2009) denoted that a complete urban model has at least to include three crucial variables dependent on the scale of the area; [1] *population density,* [2] *income level* and [3] *housing prices.* This urban model assumes that the variables are determined by the three equilibrium conditions— in the housing market, the wages corresponding to the workers' marginal productivity and the real wages adequate to the real costs of living in the area. In the period following the spatial equilibrium theory, the most significant development was the assumption that housing prices will decline with decline in commuting costs. Not only equilibrium in dynamic systems such as the labour or the housing market is unattainable (Lambooy, 1998), the linear functions used in explaining some of the early-developed agglomeration economies models were too simplistic. Inverse variation between the physical length (or volume) of infrastructure and corresponding increase in some socio-economic rates (i.e. income, patents and GDP) is empirically presented by Bettencourt et al., (2007) and later revisited with more detail (Bettencourt, 2013). Therefore, if anything, negative collinearity is recorded instead. Empirical evidence in both studies (ibid.) conclude that the balance between costs of infrastructure and socioeconomic outputs are independent of the physical size of a city, thus proving the existence of (exponential) scaling relationships of the variables.

Krugman in 1991 developed an intellectually consistent model that explained the desire of firms to concentrate spatially to reduce shipping costs. His model was further superseded by the more sophisticated New Economic Geography (NEG) models, which adjusted the aggregation economies model, incorporating the relationship between the size of the city and the productivity of workers. The costs of transport in these models are explicitly considered and they confirm that some places are more productive than others. This indicates that the cost of using infrastructure affects the productivity of a city rather than its size (length or volume of physical infrastructure). Same as Krugman (op cit.), Glaeser and Gottlieb's (2009) work acknowledges that agglomeration economies do exist strongly related to existence of reduced transport costs of resources, people and the cost of transmitting ideas. Although more comprehensive in the modelling of the variables, both agglomeration economies and the NEG are global models based on global spatial equilibrium. Furthermore, several factors are omitted from these models: over-crowding, social services availability (e.g. health care, education or culture), $CO_2$ pollution and congestion (i.e. commuting times) (Barthélemy, 2016, p72). These were included later on by revisiting the well-known Fujita-Ogawa model (Fujita and Ogawa, 1982) by Louf and Barthélemy (2013; 2014).

### *4.2.2 Local spatial models*

A significant progress from the classical urban economics models was the recognition that the spatial structure of a city is not monocentric. The spatial composit variable is no longer only the distance to the city centre as a single disutility term, and there are other dominant forces influencing growth. Those are the secondary activity centres whose number usually scales sub-linearly with the population size. The number of secondary activity centres is dependent on the attributes of the transportation network (Barthélemy, 2016). Thus, if agglomeration economies are the driving force that create cities, congestion is the main force that brings them apart. The Alonso-Muth-Mills model (the bid-rent gradient) and its many variations included the utility variable to be a product of the rent cost, transport cost and all other costs of living in the city. This model for the first time empirically explained the statistical regression between the aggregated individual choices and the size variability of a city. This model extended to numerous transport models and models that coupled the social network with individuals' spatial locations. The models were at last able to empirically explain the dynamics behind the social segregation and income structure of cities.

The secondary activity centres are explained as 'feedback loops' of individuals' interaction in the Fujita-Ogawa (local) model (Barthélemy, 2016, p57). This model explains bottom up the agglomeration economies model; using location choices of individuals and firms as variables affecting growth and productivity in cities. The model proves that at a global equilibrium, monocentric organization (e.g. the Central Urban District introduced in Chapter 3) is not stable. Thus, the equilibrium of firms (their location choice based on profit optimization— leaning towards large-density regions) and the labour market equilibrium (people that are looking for jobs and housing) are the ones responsible for creating the effect of polycentricity. Both 'agents' try to optimize their position; companies aim to optimize profits and individuals are looking to optimize their *composite commodity* (Barthélemy, 2016, p57). Composite commodity is the amount of money left after paying rent and transport costs. Including these variables, at the labour market equilibrium, the individual at a fixed location $\boldsymbol{x}$ will commute to work at a location $\boldsymbol{y}$ only when the disposable income for land and composite commodities is at a maximum level. Congestion here is neglected and the model represents the total transport cost as a liner function of the commuting cost per unit of distance $\boldsymbol{t}$ and the Euclidean distance from the origin to the terminal node $\boldsymbol{d_{(x,y)}}$ . At the firm equilibrium, the optimizing of the firm's profit located at $\boldsymbol{y}$ is a function adding the benefit for coming to the location, the number of working individuals at that location, minus the level of payroll at that location.  From the aforementioned, the equilibrium of both workers and firms is competing with each other. Companies will always be looking to pay the least in the interest of increasing their profits and workers will always be looking to maximise their composite commodity.

### *4.2.3 What the models uncovered*

The remarkable clustering of human activity in small number of areas, initially seeking empirical explanation in the location theory models of von Thünen (1826) and Marshall (1890), supported the conclusion that there are big advantages from clustering of people, consumer amenities, or housing supply, that positively affect the economic productivity of cities. However, the empirical evidence whether the increased productivity in cities such as San Diego or New York is supported by the actual clustering of people per se, is still scarce (Florida, 2002). The clustering of people due to consumption reasons could as well potentially act as an endogenous driver that increases the supply of housing and urban amenities (e.g. booming restaurants or creative culture like a thriving theatre scene) or the workers' productivity.

For example, as Glaeser and Gottlieb explain, after the bombarding during the Second World War, which resulted with the near-complete population loss and destruction of critical infrastructure in major Japanese cities, almost all of them returned to their original growth. Thus, if agglomeration economies were of a single crucial importance, these cities "may have been derailed from their long-run growth paths" (2009, p 1000). Although for a population density to act as an endogenous variable is counterintuitive, the authors assume this to be the case, with two variables as its prior accelerators. The natural advantage of the cities' geography and the political climate[36] assisted the renewal of their infrastructure. Their theory claims that using spatial equilibrium approach in adjusting the three main

[36] both challenging to measure empirically.

variables that would lead to stability of the system, cannot be ignored even by the agglomeration economies and the NGE modellers, due to the in-migration and out-migration of people. They argue that rising productivity attracts more people, and people increase the productivity further. Therefore "agglomeration economy acts as a 'multiplier' that enhances the relationship between exogenous productivity-enhancing factors, that further increase the housing prices, wages and population density" (Glaeser and Gottlieb, 2009, p1002).

## 4.3 THE SOCIO-ECONOMIC DYNAMICS

As population grows and technological development takes new currents, cities grow larger to facilitate a division of labour that generates economies of scale. Established studies (Neuman, 2005) argue that the best places to locate growth are smaller cities, reflecting the trade-offs between economies of scale and congestion. Another concern is density distribution (Lehmann, 2016), an essential result from the urban economic models analysing trade-offs between travel costs or distance and the cost of space (e.g. rent, house prices, land value etc.). Thus, other more recent empirical studies from rapidly urbanising countries (Song et al. 2017) discovered that, the cities that accommodate new growth in modern urbanising nations are the ones of mega sizes. Main benefits associated with economic [urban] density is related to improved access to jobs, green spaces and services, which attracts highly qualified labour and improves labour productivity (Ahlfedlt and Pietrostefani, 2017). In a world of doubling urban population, the question of how we should urbanise sustainably to optimise space and resources, remains a paradox.

The general properties of cities and the way they affect the biological and ecological systems they are dependent on, can be observed by the statistical relationship between the disruption of ecosystems measured in the loss of habitat and species diversity (Turner, 1989; Collins et al., 2000) against increased rates of urbanising land, expanding infrastructure and raising socioeconomic and pollution activities (Sveikauskas, 1975; Sassen, 1990; Luck and Wu, 2002; Rosenthal and Strange, 2003). But these naïve and not so parsimonious models (Hellervik et al., 2019) that follow geometric logic and obey strict rules of morphological placement of functional units, rarely succeed to represent the quality of life expected from their inhabitants. In response, long-term planning approaches that translate into masterplans, in modern planning era are being replaced with short-term intervention planning strategies.

However, these short-term planning approaches are not completely novel to the planning practice. Analysing the work of the famous French architect Patte, Choay (1996) recognized the potential of strategies as a short-term response to urban challenges. Precisely, Patte's recommendations can be summed up as strategies of planning stating that:

- All the elements of the city must intercommunicate—thus, the significance of infrastructure and its integration
- Removal of pollutants from the city including manufacturing, trades, hospitals and cemeteries—the acknowledgement of the importance of land use planning and spatial organization

To reduce this naivety of the models, short-term planning strategies should rely on more robust empirical evidence of variable, largely observed in scaling laws. Empirical evidence gathered through studying morphology and growth of cities by the use of multiplicative (exponential) growth models (Gabaix, 1999; Batty, 2006; 2008; Brelsford et al., 2017; Song et al., 2017) shows that physical scaling of urban form and related infrastructure do not necessarily follow the growth of consumed resources, population growth, financial services innovations growth, nor patents and scientific products (Barabási et al., 2000; Bettencourt et al., 2007; Bettencourt, 2013). Scaling laws are found in many different fields such as mathematics, physics or biology, and are indicators of something that has great importance (Barenblatt, 1996). The scaling laws are explained by the scaling relationships. Scaling (or agglomeration) relationships (Bettencourt and Lobo, 2016) are the 'forces' that pull people to concentrate in settlements in special locations due to socio-economic advantages and associated costs.

Urban scaling refers to the way certain properties of cities change in a scale-invariant manner between one and another quantity of interest. For example, the scale transformations on infrastructure and socio-economics are a function of the population size in cities; these transformations are scale invariant and are common to all urban systems (Bettencourt, Lobo and Youn, 2013). However, designating land use, amenities and infrastructure in cities to enable equal access and well-being to all residents could rarely follow the 'social logic' (Hillier, 2007) and scaling (Barenblatt, 1996) of public space. For example, O'Sullivan in his thesis highlights the exceptional contribution of the Watts and Strogatz's *small-world* network model (2016, p 111) used to demonstrate that the faster rates of dynamic processes in these graphs, unlike in the rates of sparse 'regular' systems, are tributed to their small-world's system topology. With this model, Watts and Strogatz directly linked the structure and processes in systems.

O'Sullivan's work (po cit.) is an important contribution to the understanding of current urban (geographic) systems through the analysis of Watts' CA rules applied on small-world networks[37]. Through applying the density classification problem (i.e. the *majority problem*) on small networks, O'Sullivan confirmed Watts' finding by showing that although CA rules were used on the system with the goal to deform its structure, near nodes remained near other nodes (2016, pp162-163). This readdressed Watts' idea that high-performance CAs could be developed by coupling their topology and not the transition rules. O'Sullivan's resulting irregular lattices were not small worlds and the formation of strong strategic locations of nodes as in the small-world networks, were not observed. Two attributes appeared to be typical for these networks; [i] regions with lower path lengths and [ii] the most central nodes in these graphs were not part of the cliques. Are overlapping communities dominating this type of networks?

There is a growing debate among scholars that although many infrastructure systems are designed and built as a tree (hierarchical) lattice (Alexander, 1965; Andersson et al., 2006; Derribe, 2017), their utilization—the flows of goods and movements— follows semilattice structure (Batty, 2006; 2008; Bettencourt et al., 2007; Bettencourt, 2013; Brelsford et al., 2017; Ahlfedlt and Pietrostefani, 2017).

Abstracted from the aforementioned, the urban space can be explained as two distinct layers of a complex network— physical and rather inflexible[38] layer and virtual and very flexible layer. Channels for communication, connecting the buildings of habitation and commerce, represent the physical sub-network. The second virtual layer maps the interactions (links) of the different agents (resource, movement and energy flows of individuals, private companies and public institutions) in the city. Although these networks exist in a spatio-temporal parallel, they are governed by different principles of evolution. Cities of different sizes do have different social quantities such as innovation and wealth creation that improve quality of life, and these quantities scale exponentially (Bettencourt et al., 2007), but their positive feedback effect, as a distribution, is not equal. In addition, they have one thing in common; their infrastructure decreases exhibiting sublinear scaling (Bettencourt, 2013). The volume of infrastructure scales faster than the area occupied by urban land and both are sublinear to the population size. This produces cities that are vertically densified (Bettencourt, 2013). On one hand, it can be argued that densification benefits from agglomeration economies by spatial concentration and maximization of infrastructure potential (Rosenthal and Strange, 2003; Brelsford et al., 2017). On the other hand, it signalizes overcrowded places and threat to the local and global ecosystem services by unsustainably exploiting ecosystem goods (Zoomers, 2010; Seto et al., 2012). As local ecosystem

[37] small-worlds CA (i.e. six or less step-lengths are measured between the cellular spaces furthest apart from each other in the CA system) are not yet confirmed to appear without defining specialized rules on them.
[38] The inflexibility of infrastructure refers to currently highly centralised infrastructure engineering design practices.

services became the pressure points and unable to carry the requirements of crowded cities in fair manner, remote ecosystem services are pressured further, since overcrowded successful places look beyond their proximity to satisfy their needs for resources.

## 4.4 DISORDERS IN CITIES AND LTSS

Evidence (Rosenthal and Strange, 2003; Bettencourt, 2013; Brelsford et al., 2017) suggests that the scaling laws of population growth embedded in the socio-economic dynamics of cities and the contest for space guiding morphology (Batty, 2008), follow different principles and the topology—the arrangement of the elements in a network— is the reason for this variance. Growth of infrastructure systems, similar to urban population growth and land use (Batty, 2006), is highly contested process (Coutard et al., 2005). On the other hand, to establish controlled and sustainable use of ecological goods, universal public infrastructure (Coutard et al., 2005) should offer equal access to ecosystem services. Thus, to understand the discrepancy between *the growth rates* of the spatial physical network such as physical infrastructure and the socioeconomic relational networks, we need to map and understand their different networks' topologies that guide the nature of their growth.

*Large technical systems (LTSs)* are defined as systems that include large number of users, multi-user components (sub-systems) and have complex interdependency between them. Due to their complexity, they are also referred to as *networked infrastructure* (Graham and Marvin, 2001). It is assumed that decentralised design of LTSs can overcome some of the issues associated with the contested process of expanding infrastructure with highly centralised (or hierarchical—tree) topology (Derribe, 2017; Hanzl et al., 2021). Synergies and co-dependencies of networked infrastructures are well registered (Neuman, 2006; Neuman and Smith, 2010; Derrible, op cit.) and they can encourage more decentralised and fair access. These synergies are also known as the 'networked nature' of infrastructure (Graham and Marvin, 2001; Neuman, 2006; Coutard et al., 2005) which is empirically shown to increase the global efficiency of the complex network (Latora and Marchiori, 2001).

*The relational networks* facilitate the built environment based on an unprecedented pattern of interactions (Jacobs, 1961; De Nadai et al., 2016). However, this pattern is not random (Barabási, 2016; Song et al., 2017). Analysis of historic behaviour, with some level of certainty, can predict future behaviour. Urban dynamics create networks of different kinds (Castells, 2010), each characterized with their own nodes and links that do not necessarily overlap (Derribe, op cit.). For example, the network of scientific collaboration does not necessarily coincide with the network of technological innovation (König et al., 2007; König, 2009), nor does the financial network with the technology innovation network (König and Battiston, 2009, p 25; Schweitzer et al., 2009). However, when the two different networks share a node, the phenomenon in urban studies is called 'economies of synergy'

(Castells, op cit.) and these nodes became *mega nodes*. They are not necessarily global places, but places that simultaneously facilitate interactions of several networks, creating one large network of *inter-linked stars* (i.e. hub-and-spoke, see Fig.3). Zeller (2010) in his empirical research of the international pharmaceutical industry finds that synergies between different networks do exist to some degree in same cities (cited in Pflieger and Rozenblat, 2010). The study is limited to one industry studied at macro level, excluding the presence of supporting infrastructure at local level and the way its [un]availability affects sustainable urban growth.

As demonstrated by previous research, urban population growth is exponential and takes place at unpredictable locations and temporalities, mathematically represented by a power law[39] distribution (Batty, 2005a). Nodes (agents) from the social network get attracted to nodes (cells) from the physical spatial network; however, the attraction appears in the relational (socio-economic interactions) network. Contrary to this, the growth of infrastructure in cities is sublinear (Bettencourt, 2013), a sign of densification to harness economies of scale. Andersson et al. (2006) explain that for example, the physical expansions of cities by developing new lots (i.e. parcels) of land don't depend on the activities generated on those empty lots, but on their potential to be connected to the road network[40]. This and several other urban change factors (Batty, 2005a, pp 21-23) are further discussed in the next chapter.

This dichotomy creates the disparity between demand and supply in planning. People may decide to move their residence based on their job location or social connections, which is a direct result of the activities generated on the land. Thus, the attractors for nodal growth in the respective networks are different. That is, "in the random (infrastructure) network the node (cell) selection mechanism is uniform per growing unit" (e.g. per land lot) while in scale-free (socio-economic) networks "the node selection mechanism is uniform per unit size of the growing unit" (e.g. per unit activity on the land lot) (Andersson et al., 2006, p 1946). The latter corresponds with the well-known Gibrat's law (Gabaix, 1999). There are general principles under which this growth happens. In network science they are known as principles of growth in evolving networks (König et al., 2007; König and Battiston, 2009; Naimzada et al., 2009; Barabási, 2016). They will be reviewed in the next chapter.

Empirical modelling of the interdependence among infrastructure, the urban form and sustainability can result in structural knowledge of the relationship of their co-dependent underlying networks (Batty, 2005b; Lambin and Meyfroidt, 2010; Alberti, 2016). The aim of morphology is to abstract tangible results from the social and economic forces that influence the city (Moudon, 1997, p 2). Morphology defines 'the physical elements and their related open public spaces and streets' (op cit., p

[39] Known as the scale-free property found in most real-life social networks.
[40] Thus, the reason for networked infrastructure and land use to exhibit random network properties.

7), understood historically. Thus, the study of morphology examines the physical relationship of buildings through their public spaces and street networks only retroactively. Indirectly, morphology offers understanding of the hierarchy of networked infrastructure, both above[41] and underneath[42] the urban surface. Much about the current rates and the laws of scaling of networked infrastructure (Graham and Marvin, 2001) in cities, and the future growth of their population can be better studied and predicted with tools available from network science (Snijders, 2001; König et al., 2007; König and Battiston, 2009; Barabási, 2016), reviewed in the next chapter.

In the past, this type of analysis was constrained due to the inaccessibility or delayed access to sufficient information on the flow of people and materials throughout cities. The availability of information today owed to the hyper-connectivity of communication tools that track human activities in cities, enables in-depth study of the networks emerging from these interactions.

# PART 2: THE STRUCTURAL TOPOLOGY OF COMPLEX SYSTEMS AND HYBRID NETWORKS

## Introduction

Modern society is filled with complex networks (Boccaletti et al., 2014) whose seamless and resilient operation is dependent on the complete integration of their communication infrastructure comprised of several co-dependent systems. The system of interaction between mobile phones, computers and satellites or the transport network and many other systems, are collectively known as *complex systems (*Albert and Barabási, 2002) and *networks* are in their heart. *Complex systems* are defined as systems whose collective behaviour is challenging to derive from the system's components (Portugali, 2011; Barabási, 2016). Cities are among the largest complex systems (Neuman, 2005; Batty, 2005a; Portugali, 2011; Lambin and Meyfroidt, 2011) and the complex interactions of their underlying networks were studied in this thesis.

Barabási (2016) explains that we cannot interpret a system's behaviour without full understanding of the network behind it. Ultimately, determining the dynamics of any network under observation leads to better understanding of the structure— i.e. *topology*— of that network. Thus, improving

[41] A street network and its associated buildings, lots, blocks etc.
[42] In most cities, other physical 'pipeline-based' infrastructure systems are collocated underneath street networks and therefore much of their topology can be understood by the topology of the street network per se.

sustainability of cities naturally requires understanding and improving the interaction between cities and nature. This means developing technical understanding of the interaction of its components (Alberti, 2016) and their behaviour towards the natural habitat.

*Complex networks* are networks whose arrangement of nodes is not necessarily limited to the form of a simple tree— i.e. parent and child nodes— or path, or circuit. Complex networks may contain multiple trees, circuits or paths (Scott, 1971). Analysis of complex networks has a great importance for understanding the basic structuring of large number of planning design and organization problems in various transportation or communication systems. Even more so, recently complex networks have found special application in urban planning (Masucci, et al. 2009) when establishing unified theory in planning of urban accessibility (Batty, 2009; Batty, 2022). Kitchin (2017) in a recent primer briefly introduces *urban science* as a contemporary scientific field that complements urban studies. This newly emerging discipline is established from the *network science* discipline.

# Chapter 5: Network science

In the 1990's a new scientific discipline slowly began to take its course, stemming from the need to build systematic approach using *graph-theoretical models* to compare the properties of *real networks.* The network paradigm was established into a new scientific discipline named *network science,* a study area responsible for introducing numerous ground-breaking network theories such as *random networks, scale-free networks, evolving networks, small-world networks* and others. These theories systematically study and explain well-known social phenomena for instance; *economic networks, communication networks, the power grid, innovation networks, spatial networks, social networks (via sociograms),* to name a few. Network science is defined both by its subject matter— complexity— and the methodology applied *(*Albert and Barabási, 2002). It originates from *graph theory,* which studies the pairwise relationship of objects— i.e. graphs. Graph theory assumes that the connection between the nodes in these networks— i.e. the way they are linked— is completely random. The distinction between graph theory and network science lies in the empirical nature of network science focused on data, function and utility, and the developed tools being tested on data derived from the real world.

The 2003 New York blackout is a good example that can explain the great importance of understanding the topology of a complex network through the application of network science methods, which show the typical network behaviour in the case of cascading failures in these systems. When a link or a node fails in the power grid, its load shifts to adjacent nodes. If the load is negligible, this failure goes unnoticed. However, if the load is too much for the neighbouring nodes to accept, as they have limited physical carrying capacity, they will also crash— initiating a cascading effect. The plausible relation between the topological (proximity of the nodes in the network) and non-topological (the capacity of each node) reliability measures indicate that the topology may influence the robustness of the power grid, during the flow of electricity through the network (Solé et al., 2008). Increasing body of scholars have evidenced the mutual influence between *dynamic behaviour* (logical topology) and *topological structure* (physical topology) of complex networks (Albert and Barabási, 2002; Solé et al., 2008, Boccaletti et al., 2014). The plausible relationship between logical and physical topology in transport planning can be for e.g. best described by differentiating multimodal transport routes from different modes of transport. The former represents the logical topology of the transport network (e.g. the way people travel through the network) and the latter represents its physical topology (e.g. the physical design of rail, road and cycling network as collocated networks in a mobility system, to support each other).

Naturally, if we want to improve the ecological behaviour of technical systems in cities, it is of a great importance to conduct integrative empirical studies of the physical and logical topology of their LTSs; i.e. their physical structure, the observed dynamics of the ecosystem goods and processes conveyed in these inter-networked systems and their estimated impact on the ecosystem services.

## 5.1 STRUCTURAL TOPOLOGY OF A NETWORK: PROPERTIES AND DYNAMICS

To determine the way nodes of a network interact with each other in any system, network science uses graph theory which scope of study has a goal to develop knowledge of the elementary characteristics of networks. These elementary characteristics assist in understanding the interaction among the components (nodes and links) of complex systems. The nodes represent a single agent in a system and the links represent the direct interaction of the agents in the system (Janssen et al., 2006, p 3; Barabási, 2016, p 45). The links between the nodes usually refer to the representation of a chosen attribute that can be either *physical* — measurable in physical units — or *relational* — measured in rates of exchange. Topological relationships predominantly have some hierarchical structure (Guimerà et al., 2003). Some examples of links between nodes include:

- flow of any natural resource, energy, information, contamination, movement, etc. for representation of the (functional) logical topology in a network, and
- personal/business relationships, object-object physical dependence and interaction or pipes, roads and wires in the technical infrastructure etc. for physical topology in a network

This network representation of any system creates a chance to study the interaction of the components of the system and the properties that the components or the system exhibit. These properties can reveal great amount of information on the system's dynamics or stability. Some systems are stable and do not evolve, which can be measured by the number of [new] nodes that were [not] added to the systems or removed from the system. Other systems may not acquire new nodes, but their existing nodes create new and remove existing links to other nodes; which behaviour shows dynamics in closed system and provides information on which nodes interact with each other and the nature of their relationship (i.e. functional linkages). A link between nodes can be *directed* or *undirected* (Barabási, 2016, p 48, p 116). A directed link denotes a direction in the interaction (e.g. a movement from point A to point B). This direction between two nodes can be incoming or outgoing for a single directed link, or in/out-going for bi-directional link. When the link between the nodes is undirected, the interaction is mutual (e.g. face-to-face communications, business relationship or friendship between individuals).

The links that are constructing the network of a system is determined by the network's attribute that is under observation. For example, if we wish to examine the water supply system and quantify its efficiency and accessibility, we need to perform quantitative network analysis of this system, and the link between the system's nodes will be the physical existence of water supply pipes and their carrying capacity from the central supplying unit to any household. In addition, the link is represented by the actual flow of water through the pipes, which can be bi-directed, in the pipes from higher hierarchy where the water supply circulates, and in- going for the pipes that distribute clean water to a single household. The amount of consumed water per node (household) is the weight of the logical link, while the physical volume and distance[43] between the pipe segments from the source to the sink is the weight of the physical link.

The reminder of the chapter defines the main concepts used in the project.

### *5.1.1 Properties of a node*

*Degree* is the key property of every node and represents the number of links it has with any other node in the network (Barabási, 2016, p 47).

*Nodal centrality* measures the importance of a node on the basis of its position in the network. There are different centrality measures that incorporate different aspects of the nodes' positions in a network (Freeman, 1978; König and Battiston, 2009, p 33). The four measures below are all part of the more generalized centrality measure of a node:

---

[43] Although the physical distance between nodes in network science is replaced by path length (Barabási, 2016, p. 56), in spatial networks the physical (geodesic) distance conveys significant information associated with the cost of building and maintaining a physical network and should not be ignored.

*Betweenness centrality* measures the number of paths passing through a node that connect pairs of nodes (Crucitti et al., 2006, p 2; König and Battiston, 2009, p 33). Betweenness centrality is important in decreasing the network's distance— e.g. establishing new links between any two nodes by introducing new shared node to shorten their distance (Janssen et al., 2006, p 6). Nodes with high betweenness centrality are the ones that grow into hubs.

*Degree centrality* represents the number of links incident to a node (König and Battiston, 2009, p 33).

*Closeness centrality* measures "the number of steps needed to reach any other node in the network" (König and Battiston, 2009, p 33).

*Eigenvector centrality* measures "the importance of a node calculated as a function of the importance of its neighbours." There are networks where a node that has very small value for degree centrality, can have high eigenvector centrality value due to the fact that some of its neighbours are highly important (highly connected nodes) hubs (König and Battiston, 2009, p 34).

*Clustering coefficient* (or *transitivity*) is the fractions of pairs of neighbours of a node which are likely to be neighbours themselves (König and Battiston, 2009, p 32; Barabási, 2016, p 63, p 93). Transitivity is calculated by the ratio of the number of triplets that appear in the graph and the number of all possible triplets. High clustering coefficient is a suggestion of modularity (also known as *community structure*) of a network (Ravasz et al., 2002).

### *5.1.2 Properties of a network*

*Average degree* is a general property of the network and represents the average number of links distributed throughout the nodes in the network (Barabási, 2016, p 48). In a directed network, the average degree can be incoming and outgoing (Barabási, 2016, p 116). The incoming degree represents the number of links that are directed towards the node, while the outgoing degree is represented by the links that are emerging from the node. In the case of spatial physical networks, the direction of links has great importance, since it represents the direction of flows.

*Average path length/or distance* in a network is the average step-length distance between all pairs of nodes in a network (Barabási, 2016, p 59).

*Degree distribution* shows the probability that a randomly selected node will have a given degree (Barabási, 2016, p 78, p 104).

*Small-world property* is observed in a network where the average diameter (i.e. the average shortest path in the network) depends logarithmically on the system's size. Thus, as the system size increases the shortest path decreases (Barabási, 2016, p 78, p 91).

*Scale of a network* indicates that nodes in a random network with equal distribution have comparable degrees and the average degree *<k>* of the network serves as the scale of any random network (Barabási, 2016). Unlike random networks, most real networks do not have average degree due to their hubs, which leads to their scale-free property (Barabási, 2016, p 124).

*Scale-free property of a network* is typical for real networks that do not have average degree distribution— the average number of links that any randomly selected node in the network is likely to have (Barabási, 2016, p 122, p 125). The number of links that nodes of scale-free networks can have varies greatly. These networks have nodes with many links— i.e hubs— but also nodes with few or no links at all. The scale-free property is a signature of deep order in the network, following a specific principle of growth, called *preferential attachment*.

*Network centrality* is different from nodal centrality. While the latter refers to the importance of a node for the network, the former refers to networks which relay on one or few highly connected nodes— i.e. hubs. This makes the network vulnerable (Janssen et al., 2006, p 15) to targeted attacks (targeted nodal removal). In targeted nodal removal, the nodes that are connected to the rest of the network only via the targeted hub, became isolated (Albert et al., 2000).

*Edge betweenness (EB)* is 'the number of minimum paths connecting pairs of nodes going through that link' (Guimerà et al., 2003, p 2). Barabási (2016, p334) further defines the abstract notion of edge betweenness. He emphasizes that links/edges with high edge betweenness are the ones that have high number of shortest paths going through them that connect different communities, rather than the links/edges found within the communities per se, that connect node pairs belonging to the same community. The discovery of these links is the main task of the Girvan-Newman (GN) algorithm (Guimerà et al., 2003). Unlike the GN algorithm, Ahn et al. (2010a) algorithm focuses on discovering community edges with high edge betweenness using a community detection rather than network partition strategy. This means that instead of forcing nodes to belong to pre-defined number of communities, the latter algorithm focuses on discovering the inherited community structure of the network. Specifically, the Ahn et al. (op cit.) algorithm discovers edges that are part of more network pathways, thus these edges/links are expected to belong to more than one community.

*Degree correlations* capture the relationship between the degrees of nodes that link to each other and the nodes connectivity can be correlated (i.e. assortative), uncorrelated (i.e. disassortative) or neutral.

*Assortativity* (i.e. network with assortative degree correlation) is typical for nodes in evolving networks that tend to connect to similar degree nodes, when the network acquires new nodes or links—hubs connect to hubs and low degree nodes connect to low degree nodes (Barabási, 2016, p 236):

> *Assortativity coefficient*[44]— is the Pearson correlation coefficient of the degrees of two nodes that are connected to one another.

*Disassortativity* opposite to assortativity, is the tendency of the nodes in evolving networks to connect to nodes with degree different than theirs— i.e. hubs connect to low degree nodes and vice versa (Barabási, 2016, p 236).

*Neutrality* is observed when nodes in the network connect to each other randomly (Barabási, 2016, p 236). It is also considered a signature of resilient and highly efficient decentralized (or distributed) networks (Solé et al., 2008).

*Transitivity* measures the presence of tightly connected communities in the network overall. This measure is often referred to as the *clustering coefficient.* It is a function of the total observed number of triplets and the maximum possible number of triplets in a given graph. The networks that have the small-world property often have very small network diameters and very high transitivity value (nearing 0.75).

*s-Metric* is a summary statistic of the node's interconnectivity, linearly associated with the assortative coefficient. It unifies many aspects of complex networks since it is closely related to betweenness and degree correlation (Albert and Barabási, 2002; Wang and Provan, 2009).

### *5.1.3 Network evolution*

Barabási in the early 90s started to question the randomness explained behind the logic of nodes' interaction and took interest in exploring the *organizing principles* governing the numerous 'everyday' networks. He has concluded that several <u>general properties</u> influence the various processes in *evolving* networks, which make them to have unique *topology* and exhibit *evolution.* Namely, these properties are defined as in the following:

1. *Fitness* represents the node's importance relative to other nodes;
2. *Internal links* are links created between nodes already part of the network;
3. *Aging* is the process of nodes slowly reducing the rate at which they acquire new links and is linked to their limited resource when acquiring new links;

[44] Also known as assortativity degree.

4. *Accelerated growth* appears in some real networks when the number of links grows faster than the number of new nodes;
5. *Preferential attachment* is the behaviour observed in new nodes that tend to attach to more connected nodes (i.e. hubs);
6. *Initial attractiveness* is proportional to the probability that a new node acquires its first link in the next time-step;
7. *Node deletion* is the disappearance of a node from the network and can affect the entire network.

The topologies of both scale-free and random network models are predominantly determined by the abovementioned properties and they are responsible for the network's regime shifts. The regime shifts are discussed in the following section 5.2. Detailed theoretical underpinnings and discussion on the topic of network's evolution and the way this evolution affects the regime shifts and state transitions in complex systems follows in Chapter 6.

## 5.2 SCALE-FREE AND RANDOM NETWORKS

The main difference between the models of scale-free and random networks is in the process of obtaining or removing new nodes and links. The process of acquiring or removing nodes and links is responsible for the topological regime shifts of the network. Real (scale-free) networks have a supercritical regime and are sparse (links between nodes are less dense) (Del Genio et al., 2011). This means that the network becomes fragmented only after large number of links are being removed. When networks are sparse, they tend to have high marginal costs, resulting with network's distributional inefficiency (König, 2009). Random network models follow an evolution process when creating links— from subcritical (only isolated nodes), critical (most nodes grouped in small components), supercritical (numerous isolated components co-exist) and ending with a connected regime, known as the *giant component (GC)* (Barabási, 2016, pp85-86). The GC observed in the models of random networks means that very large portion of nodes are connected with each other, usually these connections contain loops and cycles. Many numerous isolated components (node clusters) co-exist with the GC (Barabási, 2016, p 86). Most nodes in random networks have similar degree. Hubs are forbidden in random networks, while in networks with the scale-free topology, hubs are expected.

In general, random networks follow binominal and Poisson degree distributions and the length of the shortest path connecting two nodes (vertices) grows very slowly—logarithmically— with the size of the network (Barrat and Weigt, 2000). Real networks that exhibit the scale-free topology, mostly follow power law degree distribution, and have an entire hierarchy of hubs (Barabási, 2016). These hubs in a network of LTSs and ecosystem services are the nodes that are expected to increase the efficiency of the network. In the same time, they are also the source of pressure points and concentrated high resource demand, localizing environmental pressure around some settlements more than others.

The main distinction between the physical and the relational network is that the physical infrastructure is embedded in Euclidean space and the relational network is Euclidean space independent. By measuring the phase transition points and scaling laws of the two networks[45], the two distinct systems' complexities— their scale at a specific point in time— was captured in this project. There is great distinction between the notion of *scale* used in the studies of urban systems— the *hypothesis of urban scaling* (Bettencourt, Lobo and Youn, 2013) and the one used in network science— the *scale of a network* (Barabási, 2016). The non-linear scaling— growth or decline— of infrastructure and socioeconomics properties in cities is a function of their population, they are similar for all cities and shared across spatial units (e.g. urban block, neighbourhood, district or greater spatial unit) and time. On the other hand, scale of a network in graph theory is representing the topology of the system's components (nodes) in a simple measure— the way nodes create links with each other. Presence of scale indicates that the node degree and distribution is comparable across the network. For consistency purposes and avoiding confusion between terms, this research will use *growth* and *decline* when referring to system adding and removing nodes respectively and the notion of scale will be used to describe the unit of spatial observation (i.e. household, block, neighbourhood or a district).

## 5.3 CONCLUSION

The existence of the small-world property indicates network's efficiency (Latora and Marchiori, 2001). Although the small-world property is intriguing concept, it is not a sign of particular organizing principle in the system (Barrat and Weigt, 2000). Both random and scale-free networks are registered to exhibit the small-world property (Barabási et al., 2000; Albert and Barabási, 2002). This conclusion highlights the existing shortcomings of space syntax as a tool, vastly used for the analysis of street accessibility in the studies of urban morphology.

Degree correlation[46] is also observed in both random and scale-free networks (Barabási, 2016). In random graphs, the clustering coefficient is related to the randomly distributed edges, whereas for real (scale-free) networks is much larger than random graphs. Thus, the clustering coefficient can lead

---

[45] Scaling laws for cities try to explain the relationship between the increase of the population size and their physical growth with socio-economic output and spatial interactions.

[46] The variety in the nodes' degree that tend to connect to each other (e.g. high degree nodes (hubs) connect only to small-degree nodes; or nodes of comparable degree connecting to each other).

to particular organizing principle different for random and scale-free networks. In a network with high clustering coefficient, the two neighbours of a node are highly likely to be each other's neighbours (König and Battiston, 2009). Random graphs are likely to be dispersed and tend to have small diameters, even when the probability ***($p_k$)*** that randomly selected node having ***$k$*** edges is not too small. Albert and Barabási (2002) discovered that the clustering coefficient ***$C_i$*** [47]and the average size in real (scale-free) networks, when represented as a function of their size, does not decrease as ***$N^{-1}$*** and appears to be independent regardless the graph's size ***$N$***. This relationship is also evidenced in the Zipf's scaling law for cities (Gabaix, 1999) explaining the sublinear relation between the physical size of cities and their population size. Thus, real networks do not follow the predictions of random graphs.

Interestingly enough, two networks, out of the plethora studied by network scientists, behaved differently than expected. Wang and Provan's observation (2009) using the Growing Random Graph (GRG) model could not reproduce the topology of the Internet. This spatial network is predominantly a subject to technological and economic constraints. Instead of optimizing the overall wire-costs, the physical geography only plays marginal role in the network formation. The deployment of this network focuses on optimizing local connections, known as the "*last mile*". Thus, the organizing principle of the Internet is different. High-degree nodes do not appear at the network's backbone and exist only in local (sub)networks— i.e. communities at the periphery of the network. This results with the Internet's high performance and robustness to failures. If a hub at the periphery malfunctions, the reminder of the network does not bear any consequences. The internet traffic may be re-routed to the neighbouring hubs that could take on the additional load, a network design relying on good clustering of nodes. However, the GRG model is a good simulator only of networks with high s-Metric[48].

This led to the first step of empirically explaining the research gap from which the thesis argument was formulated. Infrastructure (spatial) networks are designed in a way that their topology is affected by betweenness centrality (Crucitti et al., 2006) and degree correlation. If the s-Metric is linearly associated with the networks assortativity, it means that these networks will have high assortativity for high s-Metric, thus hubs in these networks will connect to hubs and nodes with low degree will tend to connect with other nodes with low degree. However, in Barabási (2016) we can see that the power grid is the only neutral (uncorrelated) real network not following any degree correlation, thus it has no s-Metric. Its unique topology that stems from the network design may be responsible for the *randomness* of the nodal connectivity. Can other infrastructure networks learn from the design of

[47] *$C_i$* denotes local clustering coefficient, a measure of the neighbourhood of a node in a network. The clustering coefficient for the entire network is represented by the average clustering coefficient ***$\langle C \rangle$***. It measures the likelihood that two neighbours of a randomly selected node would have a link between them.
[48] s-Metric is the variable measuring how good the nodes with high degrees (hubs) are connected with each other in the network.

these two networks? I argue that potentially high number of links with high edge betweenness are responsible for the resilience of both networks, their efficiency and sustainability.

# Chapter 6: A system; definition, regime shifts and phase transitions

This chapter includes a summary of the reviewed literature explaining what constitutes an evolving system and in what way the system components can interact in the context of a city. Based on the decades spent in developing the theory of cities, a common definition is now accepted that describes cities as complex systems whose functional boundaries span beyond their physical limits. The sub-disciplines within urban studies, that explore the functional dependence of urban components, are predominantly concerned with understanding the topologies of the city's physical aspect and its relational counterparts (the flows of goods, people and information).

The broader definition describes the system as an assembly of many small components, which can be represented by many connected and/or isolated components and nodes. The fundamental distinction of evolving complex systems, such as cities, is that they are bound to change. Over a period of time, cities evolve from one state into another, at some threshold point. This evolution happens due to an endogenously or exogenously induced change. During this change, at a certain threshold, the system consisted of isolated components can transition into a subcritical regime, where the majority of nodes are bundled in numerous small isolated clusters. The clusters can further group into larger functional components (i.e. communities) and this iteration can take place until the system is fully connected — i.e. comprised of one giant component and trivial number of small isolated clusters and nodes.

Kauffman (1990) explains that the nodes in any physical (Euclidean-bounded) system that can be represented on a 2D grid, have two Boolean states— active and inactive. When certain *rules of transition*[49] are applied to any node in the system, the node changes its state based on the critical value of the system's *percolating threshold*. The percolation theory applied in spatial problems answers the following question; for a given average network degree, what is the probability that the furthest parts of the network are connected with each other? For physical systems, this percolating threshold is based on the prediction that increasing the probability of a randomly selected node to change its state[50], the cluster size of disconnected communities does not change gradually. On the contrary, the cluster size changes abruptly at the system's percolation threshold. Thus, the way the node will behave depends on the state of the nodes it is connected to and on the rules applied to it. In

[49] Example of transition rule in spatial problem is the land use change of the parcel, once enough neighbouring parcels are found in the new state (undeveloped/developed) typical for modular systems close to their transition points.
[50] From connected to disconnected or the reverse.

other words, in spite of the local rules applied to a node, there is a global correlation among the states of the cells in the node's immediate neighbourhood. If the state of its adjacent nodes is different from the state imposed on the node by the rule, then the node remains in a *frozen* state. The node can transition from this state only when numerous small changes in the system are large enough to make the system percolate, a condition known as the *critical value*. If the system is too inflexible (i.e. not modular), it will have a difficulty to increase its fitness surface, a property unique to every system. This is one of the main reasons Ahn et al.'s algorithm (2010) was selected. Their community search method is based on measuring nodes similarity by the number of neighbourhoods they share. More discussion on the mathematical logic of the algorithm follows in 8.1.

Systems with similar structure are confirmed to have different fitness level. 'Cellular' cities that undergo a small change should have relatively simple geometry and topological self-similarity (i.e. *fractals*) is not observed. Kauffman's model simulations show that after stochastic perturbation effects, despite its simple deterministic dynamic components, the system exhibited all sorts of behaviours stemming from the variations of the critical percolation threshold. A dichotomy between the physical and socio-economic realm of a city in his conclusion is not made. The limitation in his approach was addressed in this thesis and the distinctive percolating threshold points between the physical and logical topologies of the city were statistically analysed. Follows general observation of the percolating thresholds of physical and the ones of logical (i.e. relational) systems.

For physical systems, such as the road network, that are often mathematically represented as a 2D square (grid) or triangular lattices, the critical values[51] are 0.593 and 0.5 respectively (Barabási, 2016, pp 273-277) and the critical thresholds of a random network failure in these network equals to *1 – probability of the critical value* (i.e. 0.407 and 0.5 respectively). If the underlying topology of the physical network is not as regular, then at a random node removal, the breakdown of the network into isolated components is the same as the one found in *infinite-dimensional percolation*. Bianconi, Kryven and Ziff (2019) proved perfect positive correlation with the value of 1 between the inter-layer network degree and the percolation probability of multiplex (multi-layer) networks. Thus complex networks with high average degree are expected to have very high critical threshold. This is close to the behaviour of relational networks.

For a relational (usually scale-free) system, the percolating threshold is different than in systems whose topology can be represented as a 2D lattice. It follows the Molloy-Reed Criterion[52] and

[51] Critical value measures the probability that new node randomly placed on the lattice will connect the network's existing small clusters into a Giant Component (GC).
[52] Molloy-Reed Criterion states that in any arbitrary network expected to contain a Giant Component, every node of that network must be connected on average to at least two other nodes.

approximates to the value of 1. Due to the existence of the hubs, scale-free networks can endure an arbitrary number of random node crashes before they break apart. Therefore, increasing betweenness centrality of these networks enables their further decentralization towards network's neutrality (as opposed to the disassortativitiy typically observed in these scale-free networks), which could affect the links of their hubs, making some of them redundant.

In conclusion, as Kauffman (1990) and later White and Engelen (1993) extensively explained in their work, 'evolvability' (the ability to evolve) of physical systems requires them to be just at the transition point of order and disorder (i.e. chaos). For this reason, we need to know the threshold points of these systems and how close (or far) they are from these points, in order to understand their ability to evolve. According to several researchers (Bak and Chen, 1989; Bak, Chen and Creutz, 1989; Kauffman, 1990), systems that evolve from order into chaos (and vice versa) by reaching this threshold point, are the same ones that have the 'fractal dimension' (i.e. topological self-similarity) and features that fit power-law distributions.

## 6.1 THE COMPLEXITY OF CITIES

Cities as complex (hybrid) systems (Alberti, 2016), including numerous biological and ecological sub-systems, are far from reaching equilibrium stage and their evolution is best described through the theory of percolation. According to Alberti (op cit.), their complexity stems from their partially known, highly unpredictable and highly uncontrollable components. The interactions of these components are heterogeneous, nonlinear and multi-scalar. Non-linear processes in complex adaptive systems can intensify the microscopic heterogeneity hidden within the system, resulting with multiple possibilities of a system's response to the environment (Allen, 2001). This is the main signature of complexity; local actions have the ability to generate a global order of a kind. O'Sullivan's (2000, pp 284-287) doctoral thesis confirms the potential of geographical models (suggesting that Graph-AB could be better fit for the purpose of spatial and social sciences) to support the observations in geography theory to a limited extend[55] and explain complex system's behaviour derived from local actions. Admittedly, he too cautions on the inconclusiveness of the modelling results due to model overfitting or lack of inclusivity of exogenous constraints such as; public policy, the state of the economy, or any other large scale hazards such as the current global public health crisis.

When systems undergo a change, this happens at the level of their local cells, rather than globally (Batty and Longley, 1994, p 44). This manifests a degree of global resistance and tendency of the system to preserve its stability and form (Simon, 1969). The larger the number of cells that undergo

[55] Mainly due to the multi-faceted nature of the observed phenomenon which could not possibly be fully explained by any single theory.

this change, the more dominant the change is. At certain threshold, the system collectively changes its regime— its behaviour. In the new regime, the system will tend to support the stability (equilibria) until new predominant change emerges (Batty, 2005a; Alberti, 2016). Thus, complex systems exhibit two main states: a state of resilience (or equilibria; resistance) and a state of co-evolution (or change). According to Gunderson and Holling (2002), the high connectivity (such as the giant component) between nodes of a network leads to the system's inability to evolve and be tolerant to dynamic changes. The necessary pre-condition that allows seamless operation of the human-nature system is innovation, which is enabled through establishing connections across multiple scales (connections across multiple layers of the network). Multiple scale connectivity (i.e. high edge betweenness across link communities) creates complex interactions and requires the system to be modular and heterogeneous (Alberti, 2016). The relationship between nodes of complex network structures, including the network's robustness and resilience, is contributed to the network's modularity and heterogeneity. This is observed from their dissasortative topology (Alberti, 2016).

All changes in the physical and social systems are related to the size of the object or phenomenon under consideration. Growth or decline in these systems occurs as a *positive* or *negative feedback* on their existing state (Batty, 2005a). Complex interactions cause emergent properties and not one, but multiple equilibria in cities that involve regime shifts (Alberti, 2016). Regime shifts are the phenomena that appear due to interaction between human (socio-political, economic and physical expansion) and natural (climate, hydrology and soil) drivers, including their habitats. They are expressed as shifts in dominant feedbacks which can be positive (Batty, 2005a; Lambin and Meyfroidt, 2010) or negative (Lambin and Meyfroidt, op cit.). In large complex (hybrid) systems, the resistance to change is determined by the system's tendency to maintain its basic functions. Negative feedback is normally related to natural disasters such as flooding, losing estuary biodiversity or algal bloom (Alberti, op cit., p36). Alternatively, negative feedback can also occur through human-induced changes to the natural landscape (Lambin and Meyfroidt, 2010) such as the construction of elevated freeway without accompanying storm water management system, can cause floods to surrounding natural or manmade habitus. According to Garnett (2018), even the society itself is an emergent property of positive and negative feedback in a complex system of individuals interacting with each other.

## 6.2 STATE OF CO-EVOLUTION

In the state of co-evolution, the hybrid system replaces the stability with flexibility to adapt to uncertain outcomes (Alberti, 2016, pp 49-50). Cities are governed by forces on macro-scale related to political and economic concerns and on micro-scale of sociological nature. On macro-scale, Benenson and Torrens (2004) define six principal urban [temporal] dynamics; [1] land use, [2] social segregation,

[3] cultural segregation, [4] urban morphology, [5] urban spatial economy and [6] mobility [movement] in cities. The list is not conclusive. The micro-scale forces are defined as *weak forces of* urban dynamics and include the daily routines of individuals governing the daily use of space (Portugali, 2011). Although these micro-scale forces are weak on their own, their aggregated impact (e.g. unsustainable individual resource consumption) can have global consequences— e.g. climate change. According to the social urbanist Jane Jacobs, exactly these micro-interactions are defining the city as a conglomeration of organised complexities (Jacobs, 1961).

The extensive longitudinal study carried out by Lambin and Meyfroidt (2010, 2011) registered two fundamental forces produced by the agglomerated effect of the social dynamics (coupled with the ecology or the economy respectively) to influence land use transition decisions— one of the six macro-scale dynamics:

1. Endogenous forces which are defined as negative socio-ecological feedbacks and involve transitions triggered by natural ecosystems' decline in flow of goods and services (Lambin and Meyfroidt, 2010, p 108);
2. Exogenous forces including the various socio-economic dynamics that cause a shift in the expansion of land use from/to recovery of natural ecosystems (Lambin and Meyfroidt, 2010, p 108).

Although one of their studies focuses on forestland transition (2010) and the other on economic globalization and imminent land scarcity (2011), they present evidence that growth or decline of ecosystem services (or the growth and decline of national and international trade) can affect land use transition decisions and consequently land use change and vice versa.

Batty (2005a) elaborates in detail the specific drivers to urban development change in the city, which also affects urban change overall. He argues that the baseline for development change in cities is usually random, but when utility, human intention and constraints of spatial geography are added to this process, highly ordered structures— such as *fractals* (Batty, Longley and Fotheringham, 1989)— can emerge. According to the same study, the five drivers listed below induce this change. While the first four are spatially conditioned factors and endogenous to the process of land use transition, the last one, as an exogenous factor, is constantly shifting nexus of the developing city:

1. *Randomness* – "alluding that some part of the decision may be determined by whim, as partial ignorance of the decision-maker, but it is often associated with other more obvious reasoning" (Batty, 2005a, p 21).

2. *Historical accident* — "refers the coincidence by which the founding fathers decided on locating the historical growth of cities" (Batty, 2005a, p 21).
3. *Physical determinism* — construction of a structure constrained by the geographical landscape (Batty, 2005a, p 22).
4. *Natural advantage* — "refers to choice of industries to locate in close proximity to ecosystem services and as they continue to deplete resources, this process contributes to the *negative feedback-loop"* (Batty, 2005a, p 22).
5. *Comparative advantage* — measures the "amenities that support any location and how easy is to access them. In general terms, they are part of the [positive] *feedback-loop* and can suggest to market potential or physical accessibility, spatial distances or the cost-benefits ratio of gains from the utility" (Batty, 2005a, p 23).

In this regime, change in development is considered to be proportional the development's size and typically occurs by a constant growth rate parameter (Batty, op cit.), which can be positive for growth or negative for decline. This mathematical model, considering the distance decay effect from the central urban district (CUD), was the foundation of many traditional monocentric models in the past, used to simulate decline in density, commuting trips or rent variations.

### *6.2.1 System resilience*

The resilience of a system is defined by the system's capacity to adapt to internal and external disruptions, through changing its mode of operation and structure. In the state of resilience, the core activities of the system need to be successfully shifted away from the system components undergoing the disruption. For example, the heterogeneity (diversity of the system components) known in network science as *overlapping communities* (Palla et al., 2005) of ecological and sociological systems and their modularity (i.e. incomplete connectedness of parts) enables them to better adapt than homogeneous and highly connected systems.

Overlapping communities are the ones that share nodes and in some cases links across their different communities (e.g. see Fig. 4). The yellow community in Figure 4 shares four nodes with the purple community and one common link. The shared nodes and links are coloured in red. The turquoise community shares two nodes with the yellow community and doesn't share any link. The purple and turquoise communities share only one node and have no links in common.

Real life systems are constructed of multilayer networks that contain overlapping communities. One example is the social network of one individual who may have neighbours that are also going to the same gym and some of them are work colleagues or friends. The layers in this network are the different social groups. The shared nodes are the people that belong to more than one group, while a shared link is the two people that are neighbours but also work colleagues (this is the functional dependency depicted by the link). Another example from urban planning are the mixed-use neighbourhoods. One such neighbourhood could be incorporating commercial buildings, housing and leisure and be used by some people as a work place, by others as their domicile or entertainment.

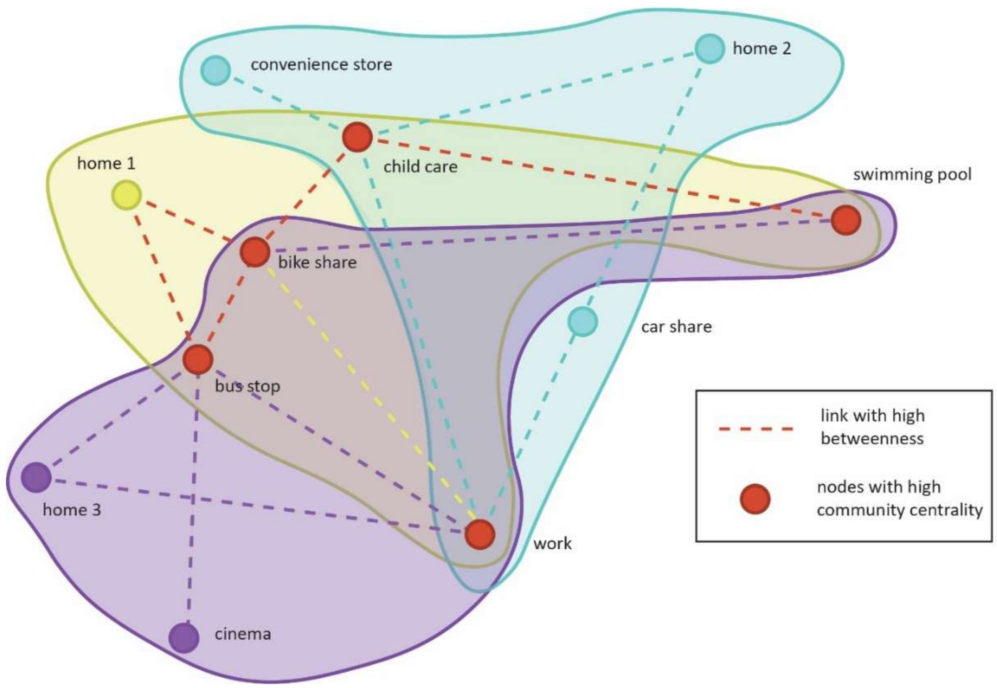



**Figure 4: Author's own illustration that shows three link communities within a multi-layer network that share some nodes and links, defining the communities' overlaps.**

In the examples above, the physical topology is defined by the location of buildings in which these interactions take place and the hierarchy of streets connecting them— i.e. the urban morphology defined and discussed in Chapter 4. The logical topology of cities is defined by the way people use these places— (i.e. their spatial functional dependencies which induce the flow of people through the physical space)—or in Hillier's words, "the social logic of space" (Hillier and Hanson, 1984). In the example in Figure 4, the functional dependency in the yellow community is between home 1—bus stop—bike share—work—bike share—child care—swimming pool. In the turquoise community the functional dependency is between home 2 – car share—work – child care—convenience store. In the purple community the functional dependency of nodes spans home 3—work – bus stop – bike share—swimming pool – bike share – bus stop – cinema. The link that connects two of the nodes with heigh community centrality (coloured in red) and also has the highest number of shortest paths going through it (coloured in red) is the one that has high link betweenness and belongs to the cut-set of the graph.

Next, if we expand this observation at a larger scale, situating the built environment in the context of the planet, example of overlapping communities in these Coupled Human and Nature Systems (CHANS) are found in abundance. Based on the nature of human-nature coupling, Janssen et al. (2006) categorises the CHANS (An et al., 2014) in three different types of complex social-ecological systems:

- Ecosystems that people connect through flows or materials — e.g. strategically positioning new habitat patches to encourage eco-system exchange between existing patches that would otherwise be disconnected;
- Ecosystems that people with their actions disconnect and fragment— e.g. construction of a railway or motorway disconnects the natural habitat in smaller patches;
- Artificial ecological networks which are created by people — e.g. irrigation systems designed to distribute water to agricultural land from a river or a lake in close proximity;

Each of these types of complex social-ecological systems have components that are networked with each other by multiple layers to form one global functional system (i.e. like the city-region analysed in this project). Their layers contain number of overlapping communities that contain nodes and links from the natural and the built environment. Furthermore, these layers share nodes and links across each other.

From the aforementioned it follows that in order to successfully study the resilience of CHANS, it is of a fundamental importance to include network components (nodes and links) with high heterogeneity and consider these networks as layers of a dynamic, evolving system (Janssen et al., 2006). Without this integrative and comprehensive approach, the study of a coupled system would be bounded and fragmented in two segregated systems (the natural and the built environment). The links with high heterogeneity[56] are usually the ones through which for e.g. multiple functional food-chains are passing (i.e. the logical topology of an ecosystem) and their existence and optimization is crucial, to enable system's adaptation to internal or external disruptions.

In the context of urban agglomerations, Alberti develops a hypothesis that variable patterns of urbanisation (i.e. land use) and modular urban infrastructure may be the key to cities' resilience (Alberti, 2016; Neuman, 2022). Thus, similarly to the conclusions by Janssen et al. (2006) and An et al. (2014), the modularity of the city perceived as a complex network and the heterogeneity of infrastructure links (e.g. collocation of cycling lanes next to roads and light railway infrastructure) with

[56] Highly heterogeneous functional links of the coupled natural-built environment are the ones that connect nodes from the natural with nodes from the built environment and often are encouraging synergies among infrastructure networks (e.g. irrigation system of channels that run along agricultural patches could at the same time serve as a storm water management system).

high edge betweenness[57], enable the city's resilience to abrupt internal or external changes. Modularity in her hypothesis refers to the loose connectivity among network's components (resembling a hub-and-spoke structure, see Fig. 3), which allows for the component to autonomous function in case of system's disruption. This is equivalent to the phenomenon of community structure studied in network science.

The resilience paradigm considers multiple stable states of a system and its main design goal is to maintain the system's core functions (Alberti, 2016, p 49) during any abrupt system turbulences. The relationship between nodes of complex network structures, including the network's robustness and resilience, is contributed to the network's modularity and heterogeneity. The network's modularity is observed when both of the networks' topologies (i.e. logical and physical) are exhibiting disassortativity. It spans between the value of -0.5 and 1, where a value of 0 is typical for networks with no modularity and 1 is the highest value for networks with perfect modular structure. Values below 0 and down to -0.5 suggest networks with non-modular clustering structure.

Networks are disassortative when their nodes connect to nodes of diverse degrees. E.g. nodes with low number of links in a network with disassortative structure have tendency to connect to highly connected nodes (hubs) and vice versa. Heterogeneity of a network is observed when the nodes that comprise the functional communities of a complex network belong to different network layers (e.g. the yellow, turquoise and purple layers—different neighbourhoods—in Fig.4). Thus, the existence of overlapping communities is a signature of increased resilience of the system and the higher the number of overlapping communities is, the better the system's sustainability. Nevertheless, the number of nodes with high betweenness should be not too high. Otherwise, the network becomes over-connected and some nodes become redundant, thus making the network 'cost' inefficient.

Therefore, to improve the sustainability of a system and in the same time reduce redundancies and inefficiency, optimizing the number of links with high edge betweenness that connect central nodes is of primary concern in the networked infrastructure design problem. The concept of communities and their relationship to modularity of a network are the third subject of investigation in this study and a crucial component of the thesis. Modularity and communities are in-depth introduced in sub-section 8.2.4 and the results of these network metrics for Glasgow and Clyde Valley are presented in 8.6.

[57] Edges with high betweenness in the networked infrastructure are the links that connect high number of points of interest observed from Origin-Destination (single trip) matrices. For example, if many people use the same street to go between different locations, then it is said that this street has high edge betweenness since it connects several origin-destination pairs.

### 6.2.2 System robustness

Robustness according to Alberti (2016) is determined by the quality of the highly connected nodes and the disassortativity of the network—the variance in the degree correlation— nodes connecting to other nodes with high or low degree. Solé et al. (2008) proved that networks with better topological robustness are also more reliable. However, topological robustness is fragile under attacks (Barabási, 2016) or abrupt system changes such as natural hazards. Network robustness can be designed either by designing network topologies that are robust to both random failures[58] and attacks or by interventions that limit the spread of cascading failures. A system is said to be robust if it can maintain its basic functions in the presence of internal or external disruptions or malfunctions. Barabási (2016, p 298) explains that a network to be robust to both random attacks and failures, has to be constructed of two type of nodes— a single node with a maximum degree which is a function of the system's size and the rest of the nodes with a consistent minimum degree.

Ecosystem services that are supporting the urban ecosystem may reach an upper threshold level[59], which drives the hybrid system into unstable state, and eventually the ecosystem shifts to a new state in which the ecological processes are highly compromised. Dendrinos and Rosser (1991) explain that similarly to Darwin's evolution in biology, *urban* evolution occurs when discontinuities take place. This unifying theory underlies that sudden environmental catastrophes rather than the continuity[60] of processes are the reason that cities are defined as complex systems characterized by self-organised criticality. As a result of any environmental perturbation, a novel ecosystem state evolves (Alberti, op cit., p 40). An example of such a process is the hurricane Katrina in New Orleans that appeared as an external shock and triggered completely novel complex system and new system's behaviour. The loss of wetlands due to human intervention and land use changes, resulted with land cover changes that led to urban ecosystem now vulnerable to extreme storm events (op cit., p 38). A *historical state* refers to a state of the system which is stable, a *hybrid state* refers to a state in which new process makes major changes in the system's components, but the system is able to reverse their novelty. *Novel ecosystems* have crossed the line of reversibility (ibid., p25) since the new process overcame the system's major components beyond the capacity of the remaining system components to reverse this novelty. This is vastly studied in *percolation theory*, a sub-field of statistical physics, elaborated in greater detail in the following sub-section.

---

[58] In the real world, a random failure can refer to a natural hazard such as a hurricane or a land slide that either occurred naturally or was caused by human induced changes to the landscape or the climate.
[59] Upper threshold level i.e. carrying capacity refers to the physical capacity of the planet to timely provide resources needed to sustain life. For further discourse on the issue please consult Rees (1996).
[60] Gradual, everlasting, continuous processes.

For further reading on the distinction between historic, hybrid and novel states of ecosystems, the reader is directed to Hobbs, Higgs and Harris (2009).

### *6.2.3 Modelling regime shifts and phase transitions in a system of networks*

Network science has developed numerous tools (i.e. modelling techniques and algorithms) that enable empirical analysis of these regime shifts, their statistical properties and the emergent states of a system. In the reminder of this sub-chapter the two predominantly used theories are explained.

#### 6.2.3.1 Percolation theory

Percolation theory is a sub-field in statistical physics and mathematics, which is highly developed (Barabási, 2016, p 273), and has a predominant goal to use mathematical models to discover clusters within a random graph. Percolation is the simplest and not precisely solved model, which displays a phase transition of a complex system (Essam, 2002, p 834). The phase transition in percolation theory is explained by the comparison between the removal of a single node vs the removal of several nodes in a network.

The theory is used to model the robustness and integrity of many technical networks. Robustness of a network is principally concerned with the impact on the network's integrity after removing a node or a link (Barabási, 2016, p275). A single node removal may have almost no impact on the network's integrity and main functions. However, if several highly important nodes (i.e. nodes with high betweenness centrality) are removed, the network could potentially break into many isolated components. A network can become disconnected by the removal of several highly important links (i.e. links with high edge betweenness) as well. These links create the 'cut-edge' set (Urban and Keitt, 2001, p1206) that once removed, the network breaks down into multiple disconnected components. The number of components increases as the connected graph becomes disconnected. Based on this, Urban and Keitt (2001) introduce three new measurements that can assess the network's integrity and the overall connectivity of its graph— *the number of components, order (the number of nodes) of the largest component* and *the largest component's diameter*.

Modelling the number of critical nodes or edges that are removed is related to the problem of mathematically calculating the network's robustness and failure. The percolation threshold is defined to be "the concentration (*occupation probability* $p_c$) at which an infinite cluster appears in the system" called a percolating cluster (Essam, 2002, p 835). The existence of this cluster makes the network to transitions from many small isolated clusters to a percolating one, which permeates the entire lattice.

The percolation threshold concept in cellular space— i.e. cellular automata— can be explained by testing the probability of for e.g. randomly coloured black and white cells to change their initial colour, when displayed on a two-dimensional (2D) grid system. The probability for any cell in this system to switch from their original 'white' to 'black' state and vice versa, ranges somewhere between the value of 0 and 1. For the probability value of zero, the cell remains in its current state (either black or white). At the probability value of one, the cell changes to a new state. For the critical threshold to be reached in the entire grid system, at some point in time enough cells need to switch their state to black or white respectively. Consequently, the black or white cells (whichever predominates the grid) will percolate the grid's opposite boundaries, either from top to bottom, or from left to right.

In a random *spatial Poisson process*[61] *test*, for small probability $\boldsymbol{p_c}$, most of the black cells are isolated and randomly placed on the 2D grid. For a probability of 0.59, a percolating cluster of black cells appears for the first time, meaning that the system for the first time displays a giant component[62] or a bond[63]. This measure can also be interpreted as the occupational probability at which an infinite cluster is obtained for the first time (Essam, op cit., p 836). At the probability value of 0.65, most of the black nodes (cells) belong to a single cluster (i.e. bond or giant component), meaning they are connected with each other through other nodes (cells) in the same colour, part of the system. At this probability value, the connected surface of black nodes permeates the opposite boundaries of the system (top-to-bottom, or left-to-right). In this phase, only a very small number of white nodes (cells) remain in the original state, thus they are now isolated. At a percolation threshold of 0.75, almost all the nodes are connected with each other and the number of unchanged (white) nodes is trivial. I expected the percolation threshold of systems with community structure to be much lower, as presented by O'Sullivan (2016, p162), since cells (nodes) are not only influenced to change their state based on the spatial proximity of their neighbours, but are also dependent on the size of their community overlaps.

From the aforementioned follows that cells in the same colour contain the same process, or they are in the same variable state (Barabási, 2016). This is equal to the connectivity of nodes in a graph where two nodes are connected with each other when they share a link, and disconnected otherwise. There have to be enough number of cells that transitioned from their original state into the new state for the system to reach its percolating threshold and this threshold point is expected to be much lower in systems with overlapping communities.

---

[61] If the random process is not spatially related, it is called a *stochastic process.*

[62] A concept introduced in graph theory.

[63] The same concept introduced in percolation theory.

It can be concluded that in evolving complex networks, new nodes connect to existing ones by two principles; nodal *attractiveness* (the number of node's immediate neighbours that contain the same process) and the ability of the network to grow (whether new nodes are added to the network — *growth*). As stated in the earlier sub-chapter, these two principles differentiate scale-free from random networks. Following, for infrastructure systems to become sustainable conveyors of flows, their networks need to be able to support synergies and decentralization of the system components (e.g. multiple uses of one or more components of any infrastructure network, or collocation of several infrastructure networks in close proximity to form a functional system, such as the multi-modal transportation system).

#### 6.2.3.2 Molloy-Reed Criterion

The Molloy-Reed criterion is the reverse of percolation and is concerned with measuring the condition for the existence of a giant component in a randomly wired network, when hubs are absent. By this criterion, a randomly wired network to have a giant component needs to be consisted of nodes most of which are connected to at least two other nodes in the network. This criterion can be utilized to calculate the critical point at which the network breaks apart and most of the connected nodes become isolated network components. The individual (sub-networks) within the networked infrastructure (Graham and Marvin, 2001) by design complies with the Molloy-Reed Criterion[64] and therefore it was tempting to use this to explore new way of developing optimization strategies when creating new or removing existing links for support synergies in the giant component of the networked infrastructure[65]. However, at the higher level of abstraction of the network's complexity, a transit system for example is composed of more than just road or rail segments, and includes central stations (hubs) where different modes of transit intersect and booking systems that enable their use. Therefore, critical thresholds are hard to determine in complex networks. As I argue in this thesis, the goal instead should be to find link communities and optimize their presence, rather than using the traditional node-based measures (such as the average degree, betweenness centrality, preferential attachment, assortativity or the clustering coefficient).

### *6.2.4 Conclusion*

The threshold points and phase transitions of a city as a complex coupled system, the subject of inquiry in this thesis, are relevant in order to empirically assess and critically evaluate the sustainability of its flows and processes bounded by the city's morphology and networked infrastructure. The

---

[64] most of the street segments connect to two intersections, unless we consider the buildings that these streets connect as network nodes, which only connect to the street segment itself.
[65] Networked infrastructure refers to many inter-linked infrastructure networks.

consumption of local and remote natural resources is directly reflected in the flows and processes and their sustainable or unsustainable intake are responsible for the phase transitions of the human-nature coupled system. Therefore, use and replenishing of ecosystem services at a sustainable rate— adequate to what the natural environment can provide at a time— is crucial for sustaining life on this planet. The modularity (presence of a community structure) and heterogeneity of components (presence of overlapping communities), including the number of links with high betweenness shared among the components of the city with the components of the natural habitat, determines the state in which the city will co-exist with nature. These overlapping components regulate the resilience of human-nature systems, their robustness and co-evolution.

Both, Molloy-Reed and Percolation theory determined the algorithm used in the analytical phase of the research. The algorithm was used to discover the links part of the edge cut-sets of the studied urban networks, through modelling their existing overlapping communities and their respective threshold points (i.e. partition density) of phase transitions.

## Conclusion: relevance of the complex systems approach for the spatial planning practice

By understanding the theory of network science, and the numerous interdisciplinary studies (Snijders, 2001; König et al., 2007; König and Battiston, 2009; Castells, 2010, among others) reviewed herein, first I have concluded that although some networks like the social, the economic and the physical infrastructure share nodes, they are distinctive networks and these nodes belong to several distinct functional communities that can span across all of the networks at the same time. Their distinction appears in the interaction of their nodes, also known as the logical topology of the network. These many networks collectively comprise the complex human-nature system— Coupled Human and Nature System (CHANS) (An et al., 2014). Different laws govern the urban scaling of these interconnected networks. The dynamics of cities are not only governed by the local relationships between their physical elements and their properties (e.g. land value is to a large extent determined by the land value of the parcels' immediate neighbours and proximity to infrastructure) (McGarigal and Marks, 1995; Portugali, 2011), but by many local and non-local relationships (relational networks) among individuals and companies (Snijders, 2001; König et al., 2007; König and Battiston, 2009; Naimzada et al., 2009). White and Engelen (1993, p 1195), referring to Kauffman's work, explain the difference between the wiring of these diverse networks. Specifically, their topology — the links between the network's nodes and the nodal *fitness* – leads to the creation of two separate systems in the Euclidean space (the physical and the logical). These systems are co-dependent, yet have different laws under which their topology develops, which determines the phase transitions of the system from one state to another — e.g. from chaotic to stable and vice versa. Network science, borrowing the underlying theory from the science of graphs (West, 2001; Li et al., 2006), to a certain extent explains the topology of these networks and some of their underlying principles of growth.

Second, an important distinction should be made between scale as the level of functional differentiation that takes place in different sizes at different locations— *functional scale or size of a spatial unit* — and scale as the level of resolution which we impose in our observation over the city— *map or spatial scale*. In this context, *scaling* refers to the proportionalities in the growth and decline of the size of the spatial unit to their respective population size. Thus, in spatial (or near-planar) network representations (depending of the map scale), I have observed two main differences using the notion of a *node*; a node representing a city or a single land unit (i.e. cadastral parcel).

Completely different (third) meaning of the notion of scale is prevalent in network science, which measures the number of connections a node has with the rest of the nodes in the network. Which brings me to the last concluding point. Third point, the degree distribution of any scale-free network highly reliant on hubs, follows a power law. In this thesis I argued that while hubs are good for inter-connectivity in collaboration and exchange (Baran, 1964; Wang et al., 2015), they do not necessarily positively contribute in connecting physical infrastructure. They make it vulnerable to bottlenecks and single point failures or attacks. Due to the underlying centralized topology of operations, daily activities may be inflexible and occasional periods of high demands may be impossible to handle. For example, transportation hubs for freight may be desirable due to benefiting of sorting and consolidation operations on these locations, but may be critical for time-sensitive cargo and passengers. There is an average estimated 23% travel time loss in the UK due to connections of trips conducted by more than one mode (Gallotti and Barthélemy et al., 2013). Distances travelled throughout hubs became longer and have larger dependencies. For example, a delay at one part of the route may result in missing the connections for the reminder of the route.

The comprehensive and interdisciplinary analysis of an overwhelming body of knowledge discovered the importance of designing networks with overlapping and heterogeneous communities. The new design approaches have great potential to overcome the challenges of the hub-and-spoke network design and improve sustainability of infrastructure services. Thus, a strategy of increasing links with potentially high edge betweenness rather than increasing nodes with high betweenness centrality could be more beneficial for improving the sustainability of infrastructure in urban agglomerations with booming population.

Distilled from the argument above, it remains a predicament whether semilattice design of spatial physical networks could improve sustainable use of ecological goods and encourage ecological services. Therefore overlapping communities became the main scientific inquiry in this study. Ecological functions (Janssen et al., 2006) (or services) are encouraged e.g. by promoting activities that purify polluted air and water resources, preserve and promote bio-diversity, decompose and repurpose waste of materials, among many others (Alberti, 2016). In urban setting this would be equivalent to introducing more natural green areas (beyond manicure landscaping) that improve eco-diversity (Hanzl et al., 2021). More recently designed and built infrastructure to a large extent supports ecological services by adopting green solutions (Schaffler, 2018). Another lower-level technical example of Green Infrastructure (GI) are the development and application of multifunctional materials such porous pavement for roads. However, most cities have inherited infrastructure adapted for users that creates many threats to ecological services, such as polluting the air with carbon dioxide induced by fossil fuel-based transport activities, coal-based energy produced and distributed to industry and households, or polluting freshwater sources with plastic coming from waste disposal systems. Therefore, large focus remines to be adapting legacy infrastructure with more Green-Blue infrastructure components.

The examples above are simplistic representation of issues that are of a highly complex nature. However, they address the urgent necessity for more holistic perspective and 'complex system thinking' in terms of networked infrastructure (Graham and Marvin, 2001) and in what way flows passing through them are discharged in the mediums (air, water and soil). Neuman (2006), as discussed in the dedicated sub-chapter (see 3.3), explicitly defined these mediums as integral components of infrastructure networks, that are negatively affected by the current economic and social practices and further challenged (Matthews, E. et al., 2000) by their limited physical carrying capacities (Rees, 1996).

Additional point of discussion that is distilled from the literature review and further elaborated in the methodology chapters is the past and present approaches in measuring, modelling and predicting the future of urban settlements in the pursuit of planetary sustainability. Planners still tend to design neighbourhood forms that follow 'tree-like' network topology, while the literature review herein concluded that neighbourhoods which evolved organically over time follow a network organization resembling a semilattice (Glass, 1998) and are more resilient to changes. Although the practice of zoning and single use neighbourhoods is being phased out from the contemporary planning practice in some parts of the world[66]. Major factor that makes the morphology of cities more 'tree-like' is the

[66] Many settlements in other parts of the world are organised more informally, therefore having more organic morphology.

centralized hierarchies of their supporting infrastructure. The tree-like network design could be one of the drivers to unsustainable access and usage of universal public services by creating 'premium networked spaces' and bypassing number of 'less-powerful locations' (Coutard et al., 2005, p 49; Graham and Marvin 2001, p249). Such premium-networked spaces are the business improvement districts that thrive on foreign investments, or the elitist (gated) residential communities. The less-powerful spaces refer to the sprawling suburban residential neighbourhoods and their strip malls acting as an artificial generator of 'social life'. Other countries are fostering policies of using publicly owned urban and undeveloped land as a capital gain by converting green areas in central zones into building parcels, or converting cycling lanes to parking places. This makes cities even more car-dominated and in that way reversing legacy semilattice urban networks into tree-like ones.

A tree-like structure of relations clashes with the relations that emerge in a natural course between people and places in cities (Batty and Longley, 1994, p 45). Alexander (1965) in his work *City is not a Tree* argues that elements in organic cities do not follow tree-structured lattice. Unlike them, infrastructure of universal public services are still designed as separate entities with tree-like design logic, conflicting with the large complex system that they are part of (Neuman and Smith, 2010; Derrible, 2017). The complexity related to planning is that the network of connections conditions future events (relationships) among elements in the urban network. If the [possible] connections of the same set of elements are to be visualized by a random network of interactions (Batty and Torrens, 2005), they can be represented by multiple variations of a semilattice structure (Batty and Longley, 1994, p 46). The speculative nature in predicting use of public infrastructure and urban space (i.e. *speculative* social logic of space) is one more reason in support of infrastructure design with high level of overlaps for future-proofing cities with inherent adaptability, changing their traditionally highly centralised layouts. The overlaps[68] known as edge betweenness and node betweenness centrality respectively are creating the semilattice structure of networks.

In historic cities, links with high edge betweenness have been observed (Alexander, 1965; Glass, 1998) to emerge through a more organic process. In recent decades, urban and landscape designers (Hanzl et al., 2021) and architects started to develop design solutions including infrastructure links that belong to multiple communities[69], e.g. collocation of a light rail transit with cycling routes and stormwater regulatory canals (such as the Delftweg connecting Delft and the Hague), aiming to increase their number throughout the complex system. However, overlaps of communities (i.e. networked infrastructure links with high edge betweenness) produced just for the sake of overlapping

[68] Overlaps refers to a shared edge among many pairs of nodes or a shared node in the pair of three nodes— *closed triplets.*
[69] Links with high edge betweenness.

is not enough (Alexander, op cit.). They must occur through natural processes (i.e. flows) stemming from the way people facilitate infrastructure, convey resources through it, access amenities in the city and use public space. As the relationship between their demand and supply changes, the links and nodes belonging to the overlap of communities must change as well, to facilitate the new way of interacting. This is rather challenging to execute with the physical dimension of infrastructure that tends to be inflexible (Neuman, 2006, 2022). However, most recent examples of infrastructure re-use projects such as New York city's *High Line* or Rotterdam's *Hofbogen* and London's *Low Line* suggests otherwise.

Semilattice is a representation of overlapping hierarchies. For example, Batty and Torrens (2005) explain that when given three independent variables, seven possible processes determine three dependent variables. Thus, the complexity of such probability estimation of processes grows exponentially and the dependent-independent variable relationship- i.e. lattice, has the form of a semilattice and not a tree (i.e. hierarchy). Based on the computational complexity of this model, the prediction effort becomes a computational problem known as *spatial Poisson process*. Moreover, complex systems cannot be closed-off from the environment and therefore determining the exact number of independent and dependent variables is impossible. When there is more dependent than independent variables, the independent data has to be specified with as much information as possible, which affects the parsimony (Batty and Torrens, 2005) of the model. These issues result with the dilemma whether spatial [urban] processes and their sustainability can be modelled altogether and the very least, derive predictions of their future behaviour.

Thus, the aforementioned underlines the relevance and need in executing fine-grain statistical inference to discover edge cut-sets, rather than trying to model and predict complex and dynamic multi-part systems of networked infrastructure in their entirety. Confirming the existence of the edge cut-sets and increasing their number could potentially improve the sustainability and robustness of networks. Until now only the degree distribution of one LTS, the European power grid network has been fully registered. First at the scale of Italy and in another attempt at the scale of the European continent (Solé et al., 2018). Its average degree distribution indicates a topology lacking preferential attachment, a general property of scale-free networks (Barabási, 2016). Thus, the power grid— the way in which electricity is generated and flows through the network (i.e. logical topology)— is the only real-life truly *neutral* (uncorrelated) network, contributing to its flexibility and decentralization. The power grid is different from its underlying transmission and distribution network (i.e. the physical topology). This is the first empirical indication that infrastructure networks can exhibit scaling behaviour other than a power-law, which is typically observed in social and economic networks (Snijders, 2001; König et al., 2007; König and Battiston, 2009). The neutral nature of the power grid also serves as initial evidence of the difference in the scaling laws describing the physical infrastructure (explained with Poisson distribution) to the ones observed in relational networks (explained by power-law distribution).

The main conclusion of the literature review narrowed down the scope of the study and informed the research design and methodology. From the theory on urban morphology and the sustainability of the urban form, I have concluded that many forms can be sustainable and focused only on analysing the structural characteristics of their supporting networks. Second, clear separation was established behind the complex nature of cities viewed as a complex system of networks with distinct physical and logical topologies. Third, what is common for complex systems is that they constantly undergo perturbations and transition from one state to another. To maintain the system's functions uninterrupted, the system heavily relies on modularity of its components. While modularity is crucial for maintaining operative state in the system, overlapping communities and links with high edge betweenness are the ones that keep the system from becoming disconnected.

Part 3 explores the most relevant and recent developments in existing methods and tools from operations research and urban science, that informed the decision for selecting the algorithms to test the main research hypothesis. Therein, I present the methodology, the data collected and the results. After, follows discussion on the results, limitations of the study, concluding remarks and the relevance of the findings for future research.

# PART 3: THE MODELING FRAMEWORK

## Introduction: Urban Science

The challenges aforementioned in the conclusion of Part 2 and the advancements in network science discussed in Chapter 5 encouraged social geographers, physicists and urban planners to advance the contemporary urban studies with the development of a new scientific field named *urban science*. Kitchin (2017) in his short primer defines urban science as an interdisciplinary approach that uses statistical analysis and advanced analytical computational methods— e.g. data mining techniques, or machine learning for predictive and prescriptive analysis or data visualisation. This field of scientific inquiry aims to find causal relationships and predict processes in cities viewed as complex systems.

Complex system models, that to an extent predict complex behaviour, are becoming more popular in many disciplines such as social studies, biology and physics (Batty, 2006). They are used and advanced despite the full awareness of modellers that these models can only be partially tested; and therefore not able to be completely validated. Their advancement comes from discoveries specified in Chapter 5— e.g. the knowledge developed around network and node properties such as hubs, preferential attachment, assortativity and disassortativity, six degrees of separation, small world or the critical threshold and the formation of a Giant Component (GC)[70]. Scientists using complex system modelling are understanding how to model targeted attacks or the reasons for cascading failures. Based on their findings, they learn how to design and build more robust networks and systems.

As discussed in the conclusion of Part 2, using complexity to observe urban systems is the only way to understand the distinct topologies of their physical and relational networks, in urban studies better known as the structure and the dynamics cities (Ducruet and Lugo, 2011). A complex network can represent the topologies of these dynamics as two separate yet coupled multilayer networks— i.e. a system of systems.  Using complexity we can discover the different laws that govern the systems' individual topologies. These laws determine the systems' respective phase transitions— e.g. from chaotic to stable state and vice versa, which in urban processes, I hypothesise, to be different from each other. Phase transitions appear at a point when the exogenous or endogenous changes imposed on the system are prevalent enough among the system's components, that lead the system to transition from one state to another. Several system states are known so far and were discussed in detail in Chapter 6.

---

[70] The GC is introduced throughout this thesis and sometimes referred to as the percolating cluster or the bond from the theory of system percolation. To my best knowledge, the three concepts are synonyms, simultaneously used in different fields.

In this Part, Chapter 7 introduces the theoretical underpinnings and technical characteristics of three of the six general classes of models, based on linear or nonlinear programming, that find vast application in spatial statistics and spatial analyses. The brief overview of the main groups of modelling tools which informed the modelling paradigm, was necessary to underline the fundamental difference between existing modelling approaches and the new approach proposed with this thesis. The goal of the thesis was to advance the practice of urban modelling, and the new urban science discipline, with an original model that can help urban planning practices worldwide to improve equity of access and sustainable resource consumption, through more decentralized and robust infrastructure network planning and design.

Since the principle contribution of the thesis was the methodological advancement, the main concepts to be studied by applying the single link clustering algorithm (i.e. defining the elements of the city-region, defining spatial interaction and defining the system of systems) are extensively described in this chapter. The literature review in the earlier Chapters 2 throughout 6 narrowed the selected variables which were measured in the thesis. In Chapter 7 their measures are defined to be a product of their spatial dimension and accessibility. The primal and dual syntax of their mathematical representation as graphs is also introduced and critically explained, since accessibility of places is a measure originating from the graph's dual representation.

Chapter 8 discusses the process of designing the model, accessing the main datasets, cleaning and manipulating them, creating the variables and the method used to measure their variance. Chapter 9 presents the results from the model, followed by wider discussion of their interpretation within the sustainability framework of the city-region. Chapter 10 presents the summary of the research findings and returns to the research questions and explains how the results helped answering them.

# Chapter 7: Methodology

As shortly introduced, models built to represent the cellular dynamics (Batty, 2016) are powerful tool enabling planners and geographers to tackle many urban problems such as population growth and associated land use changes, or traffic management (Heppenstaal et al., 2016). In the introduction of the forthcoming chapter I analysed the current widespread use of cellular automata based on its advantages to simulate micro-agents behaviour, especially for modelling land use transitions (Shafizadeh-Moghadam, 2017) and address its limitation under uncertainty (complexity). Then, alternative approach to CA-based modelling is proposed using the tools from network science.

Langton (1986; 1990) uses cellular automata to explain the way complex systems embedded in Euclidean space, at a specific point of time, transition from organised into chaotic state, or vice versa. He reasons that the evolvability of physical systems, which is crucial for their survival, emerges due to a phenomenon known as *phase transition* (White and Engelen, 1993, p1195). The behaviour of the nodes in the organised state is defined in *percolation theory*, describing the behaviour of connected clusters in a random graph, while the behaviour of the chaotic system state is well defined in *graph theory* as the *scale-free network.* In scale-free networks, the behaviour of the nodes is guided by two of the network's properties—*growth* and *preferential attachment*. Further, similar to Barreira-Gonzále et al.'s (2015) advance of this technique is proposed by O'Sullivan (2016) through combining CA with graph theory. The new tool is able to simulate cellular growth/nodal growth (or cellular change) that leads to exponential change, fully dependent on the system's state of phase transition, by increasing or decreasing the edge betweenness in the physical network.

Percolation theory finds application in urban planning as a tool to evaluate the connectivity of urban systems (e.g. Arcaute et al.'s study (2016) on street intersections) or occupational probability of cellular space based on its neighbouring cells. The formation of clusters (i.e. communities) is always the outcome of a process that pushes the system to go over the threshold into a different state or when this process is less influential, the system maintains its stability since the number of connected cells remains under the threshold. Arcaute et al. (op cit.) demonstrate that the hierarchical organization of the street networks in major British cities is outcome of complex political, socio-historical and geographical processes. This organization was explained using percolation theory. Another spatial study using percolation theory (Molinero et al., 2015) discovered strong positive correlation between the regional fractures and political polarization. They concluded that the morphology of the city is inherently formed of many fragmented communities organized in hierarchical structure and the street network holds them together in a connected system. When the percolating cluster appears, in fact all of the studied systems exhibit hierarchy which shows the organization of their underlying network into communities. Furthermore, the critical threshold point of these spatial systems is affected by the geometrical properties of the space. At the minimum threshold, evident boundaries appeared at which the cities' cores could be geographically identified. At the maximum critical threshold, the clusters observed that defined cities were in good alignment with other proxies for city boundaries obtained from satellite images.

Form the review of the application of percolation theory in urban studies, I defined the main thresholds to be observed. In this thesis, the percolation threshold of the physical network[71] is

[71] i.e. the value below which the arrangement of buildings, public spaces and streets become disconnected from one another into isolated clusters.

measured by calculating the percolating threshold of their geospatial topology. The network's geospatial topology is built by the observed accessibility of 400 meters or less, between each household in the region to various destination categories.

A system in this study was defined as a cluster of units comprised of all the houses in spatial proximity that are connected to a bike share, a car share, a car charging, a recycling location, the store, a fitness/wellness centre, a café or a land with natural greenery at the cut-off distance. All the remaining building units were considered as the disconnected nodes in the system. Thus, the percolating threshold of this system measured how many nodes (links) can be removed before the system breaks into many small disconnected components. In the case when some of the nodes were disconnected from the Giant Component, the percolating threshold measure gave the minimum number of links needed to connect the nodes in a GC within a walking distance.

## INTRODUCTION: MEASURING AND MODELLING IN SPACE

Spatial statistics and analysis often used as synonym to quantitative geography, in a narrower sense uses formal techniques to study the spatial relationships between geospatial entities, using their geometric, topological and geographic properties. This field first emerged by using methods borrowed from other aspatial disciplines (Fotheringham et al., 2005), later developing its own methods (Barthélemy, 2016) aiming to analyse spatial data. One of the main advancements in the field of quantitative geography is shifting the focus from global regularities to local exceptions and the production of local statistics, rather than global ones (e.g. mean or median distribution). Another advancement is the acknowledgement that in spatial regression models, spatial data is positively spatially autocorrelated, contrary to the common understanding in general regression models that the data consists of independent observations (Fotheringham et al., op cit., 12). Subsequently, spatial statistics methods were developed to perform both global and local spatial interaction analysis.

Urban models (Lawry, 1965; Batty and Torrens, 2005; Batty, 2005b; Barthélemy, 2016) in their essence represent an abstraction of urban systems delivered through a computable logical and mathematical formalism in order to enable 'scientific' predictions (Batty, 2016). They are the essential guides in understanding the reality of complex systems (Batty, 2005a; Barabási, 2016) such as cities (Batty, 2005a; Batty and Torrens, 2005; Barthélemy, 2016) and the processes governing them.

Decades after their first implementation, despite their intensive criticism, scientific models have progressed significantly (Couclelis, 2005). Such scepticism is based on past experience when much of the land use decisions, were based on the models' predictions, and failed the ultimate goal to enhance the strategic mission and 'future-proof' planning. Instead, their result was contemporary behavioural patterns (Marshall and Gong, 2009) of land uses that created new urban issues of sprawl, congestion,

$CO_2$ pollution, social segregation and spatial injustice (i.e. accessibility inequality). Much of these issues are not tributed to the segmented approach to planning alone, in fact they were at large induced by the structure of the consumer-oriented economy and its dominant role in politics and consequentially reflected in planning the living environment.

On the other hand, the issues associated with the poor outcomes of the early models (Fotheringham et al., 2005) of urban phenomena are credited to the shortcomings related to using aspatial tools predominantly borrowed from econometrics. With the major progress made over the four periods of transition and the merger of the study of social physics with macroeconomics, the new methods and tools developed can generate various forms of spatial interactions. As such, they are used to produce models of aggregated or disaggregated cross-sectional studies.

Based on the qualitative or quantitative knowledge urban models generate, they can be categorized as *quantitative* and *qualitative*. In this thesis the focus is on quantitative models which are described in the coming section.

Quantitative knowledge (Cecchini and Rizzi, 2001; Couclelis, 2005) is generated by mathematical and numerical representation of the process of changes on a geographic surface over the natural landscape by the built environment in a given time-span. The early version of these models are the foundations of the geographic information systems (GIS), which advanced the numerical representation of space by the use of digitally coded geometries (Batty, 2016). Common misconception is that quantitative researchers try to explain everything about cities. Quite the opposite is true; with empirical observations and consequent predictions, quantitative models aim to embrace the broader dimension and theory developed by social scientists (Fotheringham et al., 2005).

The greatest drawback of existing quantitative models (Barthélemy, 2016, pp43-45; Hellervik et al., 2019) is their reliance on statistical estimates rather than exact measures of observed variables. Two particular issues are strongly emphasized: statistical estimation and absence of clear understanding of the city's boundaries. One example in the absence of direct measures is the use of estimates on the total population commuting distance per city size, in the effort to calculate the variation in $CO_2$ emissions when population size increases. An example of the latter issue is the challenge related to defining city's boundaries and subsequently estimating internal (endogenous) and external (exogenous) socio-economic interactions.

In the face of the challenges above, it is difficult to draw any conclusions from aggregated measures, often considering that e.g. the observed nonlinearity (i.e. exponential function) can possibly be an error of the statistical inference (Barthélemy, op cit.). Barthélemy (op cit.) identifies that this is related to the "absence of any mechanistic insights about the observed scaling". Hellervik et al.'s model

(2019) addresses this issue of parsimony by introducing the concept of preferential centrality, by replacing the application of the gravity model in urban studies with the preferential attachment model from network science. As such, in their alternative model the level of urban activity is a direct product of the accessibility and attractiveness of urban locations measured as the power-law for linear processes and approximately power-law after regime shifts in the non-linear systems. Thus their model closed the loop between activity and spatial interaction of origin and destination areas back to activity, as activity in itself stimulates attractiveness and the traditional preferential attachment becomes nonlinear[72].

The modelling approach in the thesis, similar to O'Sullivan's (2016) graph Cellular Automata, is a blend of agent-based and graph theory models, therefore the main theoretical underpinnings of both modelling paradigms are explained in the following sub-chapter.

## 7.1 AGENT BASED MODELS (ABMS)

Agent Based models (ABMs) are based on the Agent-based modelling (ABM) technique, which consists of describing, computing and simulating emerging attributes typical for self-organizing complex systems. The mathematical relationships they represent are always dynamic and can be either linear or non-linear. They are best known at handling nonlinearity and representing complexity (Kuby et al., 2005; Batty and Torrens, 2005). They traditionally find a widespread use in engineering and physics to model the dynamic processes and temporal changes in spatial structures such as the infrastructure networks (Ducruet and Lugo, 2011). In urban planning and infrastructure development, agent-based modelling has emerged as a leading method (e.g.UrbanSim, ActivitySim, SILO) for predicting the coupled development of land use and transportation network (Kuby et al., 2005, p35). The main goal of this technique is to model the interaction of the individual system components called agents and the emerging relationships between them and their environment. They use simple rules of individual interaction with each other, which produce unexpected collective behaviour. In their overview of different modelling techniques that can be used to enhance SDSS (or GIS), Kuby et al. (2005) specify that different individual agents can be combined in a hierarchy of higher and lower order classes, which combined in a variety of ways could answer many different research questions.

ABM are opposite to system modelling (simulation) and are a representation of a collection of agents, each of which has their own *attributes* and behaves to specific *rules* (Kuby et al., 2005). Their collective interrelationship represents large-scale system level structures. Attributes (Ducruet and Lugo, 2011) are the intrinsic characteristics of an agent (e.g. name, age, gender or location) and behavioural rules

[72] Attraction as directly dependent on the level of activity transforms into a continuous type of the preferential attachment.

(Benenson and Torrens, 2004; Batty, 2005b; Ducruet and Lugo, 2011) are the mechanisms that enable agents to change their original state, whatever that is (e.g. from undeveloped site to developed site, from active to inactive etc.). ABM explicitly represents the decision process of each agent and in the process-response agents can be programmed to 'learn' based on feedback of previous experience (Kuby et al., 2005). Thus, the modelling can incorporate three different processes: agent decision making, agents changing decision making rules (due to learning) and agents interacting with each other and the environment around them. ABMs can use spatial and aspatial data and in the latter case they are widely known as cellular automata (CA). The main benefit of linking AMB to spatial information system such as GIS is mapping of the agent's individual behaviour to their location and the various attributes associated with that location.

The cellular automata models are developed to model complex systems and are based on processes that take place in the local neighbourhoods of cells. They are distance-weighted models and the state changes of cells at a given distance is automatically adopted (Batty and Torrens, 2005). They are predominantly focused on constructing and demonstrating the way rich *urban dynamics* are generated, omitting the process of validation. As such, models that cannot be validated are no different from intuition or qualitative reasoning (Batty and Torrens, 2005). This is the major drawback of CA application in urban modelling. CA have five main characteristics (Barreira-González et al., 2015, p120): [1] discrete space representation; [2] discrete time representation; [3] from the set of all possible states, a cell can only take one state; [4] the state of the neighbouring cells define the transition rules; [5] the transition rules are the same for the cell and its neighbours.

The local rules applied during the state transition of a cell have the capacity to generate complex patterns (Barreira-González et al., 2015) observed on a global scale. This leads to the conclusion that the system in its entirety is more than the sum of its parts (Torrens, 2000). However, the strict grid and regular structure of CA imposes limitations to simulate urban growth at its 'chaotically irregular' reality (Couclelis, 1985) or to elicit representative local rule-based models, since neighbouring cells are identically located (O'Sullivan, 2016, p 106) which is not the case for land use patterns of urban settlements. Over the years these limitations have been partially resolved. For example, a progress has been made by changing the <u>possibility of the cell states</u> from being simply '*dead*' or '*alive*' (i.e. either developed or undeveloped), to simulations that e.g. at the same time allow multiple land uses (Barredo and Gómez-Delgado, 2008; Engelen et al., 2007). This is much closer to the reality observed in mixed-use development, thus simulations that allow this flexibility register distinct uses as a qualitative knowledge— i.e. percentage of each use.

Further, progress based on Tobler's rule (1970) was made by changing the influence of a cell on its neighbourhood when it goes through a state change, which strength of influence to a larger extent includes the distance decay function. Additionally, the time and space regularity in CA has been replaced with the possibility to model change, over irregular time intervals and irregular cells. The improvement (O'Sullivan, 2016; Barreira-González et al., 2015, p 122) in the irregular cells representation are done by partitioning the space using one of the following methods: [i] Voronoi polygons, [ii] Delaunay triangulation extracted from plot centroids, [iii] land use parcels, [iv] census blocks, [v] a set of irregular geographical object that permit change in their structure or [vi] cadastral parcels defined based on property ownership and structure.

There are two different methods of conducting ABM application: generative (Epstein, 1999) and degenerative (Ducruet and Lugo, 2011). In short, the generative method uses the local network's information on the level of nodes. It analyses and explains the way nodes acquire new links with other nodes over time and the phase transition points of the system. This method studies the properties of the nodes. The degenerative method analyses the complete network, considering all nodes that are part of the network and all of their links. Each of these links represents a possible path that can be used in the system. The importance of links are determined by the number of agents using them and the cost of maintaining them, as well as the importance (i.e. betweenness centrality) of the nodes that connect to them. To be able to develop any of the methods, social and economic factors context is necessary. For example, the evolution of the transport network is closely related to the growth and change of its central places (Ducruet and Lugo, 2011), so they need to be one of the input criteria.

To better understand the coupled human and natural systems (CHANS), agent-based modelling (ABM) is perfect tool that can capture and integrate data across scales, considering the three dimensions of spatial data; i.e. spatial, temporal and hierarchical (An et al., 2014). For example, using this modelling scheme, An et al. (op cit.) represented the society as a construct connected within its structure at levels of hierarchy —individual agents; persons, household and community (a cluster of ten to twenty close proximity households that share similar community context) vs. environmental agents; neighbourhood and the world. The *similar context* is defined as similar access to markets, employment opportunities, schools, public transport stations and healthcare facilities. In summary, the ABM tool can assist in the decision-making process, capturing the relationships between individual agents of complex systems and their local influence on their surroundings. Agents must either represent human actors or an institution—entity (e.g. government), that could constraint and shape the processes that are creating changes or affecting the corresponding landscape (An et al., op cit.).

On the other hand, coupled human natural systems (CHANS) are mainly defined by the reciprocal relationships (An et al., 2014) between their agents or the components these agents are part of (Alberti, 2016). An et al. (2014) elaborate on six main features that are representing CHANS, those are:

1. Reciprocal effects and feedback loops;
2. Non-linearity and existing thresholds;
3. Surprises;
4. Legacy effects and time lags;
5. Resilience;
6. Heterogeneity.

These complex feedback loops are well handled ABMs. When coupled together as a simulation tool, ABM and CHANS can best represent the non-linear relationships between the multiple agents, next to their collective impact on the local surrounding (system-components) and the way the change of the system-component(s) features' does[not] influence part or the entire system. The limitation of CHANS ABMs are that they are very situation specific (An et al., 2014) and the comparison of results between different locations and situations is challenging. This makes CHANS ABMs very unique from one-another and they should normally be used as a supplement to other top-down approaches (An et al. 2005). ABMs are a good supplement to top-down approaches when modelling multicomponent processes with many feedback loops between the different components. For example, An et al. (2012) in their research presented modelling of the feedback loops between land use ABMs and ABMs simulating population processes (birth, death, marriage etc.).

From the forementioned methodological analysis of the current application and advancement in CA, it was concluded that capturing the evolutionary complexity of the coupled human-nature system in its entirety is a devious task. Therefore, CA rules of transition were not applied in this research. Instead, the percolation threshold expressed as the system's modularity of its existing size was measured. The system's modularity measure was used to retrieve the network's existing value for modularity and to analyse the potential of improvement of this measure across the various regions on the network. I was not interested in changing transition points and rules for the system studied, but the discovery of overlapping communities.

## 7.2 GRAPH THEORY MODELS

The branch of graph theory analysing 'static' networks offers tools to researchers to study the topology (structure) of unchanging static networks for over 60 years (Kuby et al., 2005). Since the nature of the network measures (e.g. centrality, scale, degree distribution etc.) are treating networks as unchanging, this branch is also known as static graph theory. However, in the last two decades, this field has progressed and new measures are constantly being developed to analyse 'evolving' networks (Albert and Barabási, 2002; Costa et al., 2005; Boccaletti et al., 2006; Barabási, 2016). Graph theory, as explained in the Chapter 5, is the elementary theory, which any tool or method used to analyse network topologies is based on. As demonstrated in the Barreira-González et al. (2015) study, graph theory shows great potential to operate with vector space. Although the static approach to complex and dynamic networks has been superseded by the introduction of the Barabási – Albert model with measures of preferential attachment and growth (Barabási, 2016, pp 169-192), the main disadvantage of graph theory still remains to be the process of explaining cause and effect. Thus, the models operate in a 'black-box' manner. Static graph theory is deterministic, while random graph theory is a stochastic tool (Kuby et al., 2005, p 16). Static graph network is finite with fixed set of nodes and links, not scaling and not exhibiting any randomness in the behaviour. Opposite to the static graph, a random graph includes randomness and growth or decline of the number of nodes and creates new or removes existing links between them.

The combination of CA and random graph theory models allow a level of flexibility, since graph theory models are not too restrictive, one of the main drawbacks when using CA alone. Second challenge is the resolution of the output which is either aggregated measure explaining the property of the entire network or the properties of individual nodes. This shortcoming in the thesis is addressed by shifting from the traditional graph partitioning to a clustering method which, next to the traditional global measures such as the graph's partition density or modularity, the algorithm produces local network measures which are saved in the form of a dendrogram.

The research tried to supersede this drawback by combining it with CA system, which in principle is highly transparent. One main characteristic used in CA that was applied in this research and with that makes the model deterministic; was the geodesic distance between origin and destination points as determinant of whether two nodes have a link. Thus, the cut-off value was the only criteria in defining the presence of a link. The second deterministic rule was the undirected links. The third characteristic that used the advances made in CA was the cellular division of space in irregular vector space based on the buildings' footprints. Since no new links or nodes were added to the graphs after the network was built, the rule of how nodes link with one another was well-defined (see 8.4) from the onset and fixed, thus the graphs in this study were static. In future research, the evolution of the network will be based on defined interventions of adding new links to the network in regions where the link community overlaps were low. Thus, there was no element of stochasticity in the evolving network analysis either.

O'Sullivan (2016), as discussed in 4.3 concluded that CA rules applied to transform the system did not affect the nodes' proximity to each other, which meant that small worlds did not characterise the CA system he studied and the hubs (i.e. the highly-connected central[73] regions) were not part of the cliques. This means keeping the spatial dimension of the elements (i.e. nodes and links) in the graphs is very important for spatially explicit systems. With this modellers can use the spatial proximity of nodes to investigate whether nodes are linked with each other. Thus, in the graph construction process I assumed that the transitivity (i.e. the clustering coefficient) of the system was very low. And as I will show in the results in 8.6. this was the case for all sixteen analysed graphs. To reduce the complexity of the model and based on O'Sullivan's (2016) recommendations (see 4.3), I didn't use any transition rules in the systems studied in the thesis.

### *7.2.1 Scale-free network: Albert-Barabási model*

#### Preferential Attachment

Preferential attachment is a probabilistic mechanism (Barabási, 2016, p170) that implements the rule of preference in the graph model. According to the rule, a new node connects with any node in the network regardless of the existing node's degree. However, the likelihood that the new node connects to a hub is twice as large compared to the chance that it will connect to a low degree node. Unlike the random model where the assumption is that, any node chooses its neighbours randomly, preferential attachment in the Barabási – Albert model links new nodes to more popular i.e. established ones.

#### Growth

All real networks are a result of growth by adding new nodes. The main characteristic of growth is that earlier added nodes tend to have higher degrees since the nodes that arrived later in the network tend to connect to existing ones following the rule of preferential attachment. Over time, the existing nodes compete for links. In business and marketing this phenomena is called *first-mover advantage.* For transport networks, this can be observed as noted by Parshani et al. (2010), in the case of the inter-similar networks of seaports and airports, where well-connected airports tend to connect with well-connected seaports. Ducruet and Lugo (2011) argue that the expansion of any infrastructure network is first bult to complement the existing network, while later it competes with it. This is also evidenced by the Barabási-Albert model showing that growing networks (Barabási, 2016, pp.165-188) result with linear preferential attachment and the oldest nodes in the beginning are complementing parts of the growing network, and in time, they turn to compete for the new links (or to connect to newly added nodes).

---

[73] Not necessarily found in the centre of the spatial region (in geometric terms), but in the sense that they were of central importance to the network (i.e. network hubs).

## Graph clustering- modularity and communities

The problem that studies the clustering of graphs (Brandes et al., 2008) is one of the most fundamental problems that analyse relational data (such as information, physical flows and chemical interactions). Also known as *graph partitioning problem*, it has been used in a variety of applications for a very long time. The problem subsets a graph (e.g. a network) into communities with the goal to better study the dynamics of complex networks, including their formation and evolution. It uses statistically observed variables (Guimerà et al., 2003) that describe the properties of the network; e.g. *average distance* between the nodes within the network and the degree distribution of the nodes; or the properties of the node— the clustering coefficient and betweenness centrality. The representation and empirical study of graph's *modularity* (Newman, 2005; Branders et al., 2008) finds important application in many scientific problems concerning large and complex networks. The outcome from studying modularity in a network is a mathematical model that replicates the universal mechanisms, which are the drivers for the formation and evolution of relational networks in self-organizing systems, such as social networks. The main parameter of this mathematical model, that measures the structural features of the network's communities is the <u>clustering index</u> (i.e. the modularity) of the novel graph (Brandes et al., 2008).

To study these statistical properties of novel networks and their nodes, the clustering method has a goal to use an algorithm which will best optimize the modularity of the network, producing optimal partitions—*communities* (Brandes et al., op cit., p1). The presence of communities in a network is the sign of a hierarchical structure (Palla et al., 2005) in the complex system under study. However, in complex networks a single node can belong to several different communities (Ahn et al., 2010; Bocchaletti et al., 2014; Afsarmanesh and Magnani, 2016). Therefore, it is important in the output of the graph clustering process, to allow for nodes to belong to more than one community.

Palla et al. (2005) explain that cohesive and overlapping communities in large networks create challenges for scientists to interpret their global organization. They consider current methods to only partition networks in separate communities, ignoring the fact that many of them overlap. The Girvan-Newman (GN) algorithm, the Clique Percolation and the Louvain algorithm[74] are all based on the network partition principle, forcing nodes to only one community. According to Palla et al. (op cit), overlaps are meaningful to the network's features. With this regard, they introduced a novel algorithm that captures the degree distribution of communities while allowing nodes to belong to multiple communities. When their algorithm was applied over several scale-free networks, the resulting degree distribution was not a power-law curve, typically expected for the topology of scale-free networks.

[74] The third one is used in this study.

Each of the three studied large-scale networks exhibited novel degree distribution. It was comprised of two distinct parts; an exponential decay of characteristic community degree, followed by a power law tail. The exponential decay showed that the probability of randomly selected node within the communities to have a certain number of links, is decreasing at a rate proportional to that of the community size distribution. At the first part of the community size distribution, most communities have *k*-order (comparable) sizes, so the distribution predominantly relies on the predetermined minimum cohesiveness of the community (i.e. whether the communities are weak or strong[75]). When the distribution of the cumulative overlap sizes was compared, there was no typical overlap size and limited range, each network had power law distribution with a large exponent. With this, Palla et al. (op cit.) confirmed the importance of capturing communities' overlaps, and therefore the greatest importance herein is given to discovering communities and their overlaps in the context of accessibility within the Clyde Region.

A real life example of nodes belonging to multiple network communities in urban planning is the functional representation of a multimodal single trip of one agent moving from point of origin to destination located in two neighbourhoods, interchanging at neighbourhood locations to different modes of transport. In this case, the respective layers are the different infrastructure sub-networks (i.e. communities) (e.g. cycling lane, walkway, railway and bus lane). Another example is the different neighbourhoods that an agent visits to carry out their day-to-day activities.

In the example below (Fig. 5), the top layer represents the street footway network, the layer in the middle represents the cycling network layer and the bottom represents the railway network layer and the nodes in red are the origin, the mobility hub interchanges and the destination of the trip all found in different locations. Third real life example from transport planning of overlapping communities is the inter-similar networks of seaports and airports (Parshani et al., 2010), where one city belongs to several sub-layers of the multi-layer logistics network.

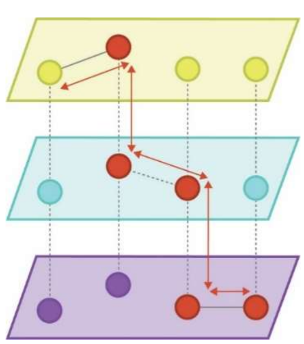


**Figure 5: Simplified example of a multilayer network depicting the travel pattern from origin to destination.**

[75] In strong communities individual nodes inside the community have links to many of their internal neighbours, while in weak communities two nodes have more than one unique link with each other, but are not required to connect with other nodes inside the community.

The algorithms of communities (modularity) detection allow "zooming" into the structure of large and complex networks and they are a powerful instrument that enables predicting the changes in the modularity of the system when a community (i.e. module) is removed.

Comparing the best performing community detection algorithms, Guimerà et al. (2003) empirically show that the Girvan-Newman (GN) algorithm (see Figure 6) is the most efficient one to detect communities in complex networks. The algorithm detects and removes the links of the network that have high *edge betweenness (EB)* and groups of nodes that remain after the removal of these links, belong to the same community.

Links with high edge betweenness (Figures 5 and 6 coloured in red) are the links that connect highly clustered communities with other highly clustered communities. A *good* community is expected to have small cut size, i.e. small number of links with high EB (Fortunato, 2010). Closely connected communities share resources faster than sparse ones. It also means that they are strongly reliant on the links with high EB. In terms of networked infrastructure (Graham and Marvin, 2001), neighbourhoods that are closely connected with each other and to central resource supply systems are known as premium locations with better accessibility than the rest. These neighbourhoods are usually sparsely connected with the rest of the network through scarce number links with high EB.

Based on the conclusions from the literature review, the thesis assumed that their flows — i.e. resource consumption or production of waste — have different rates than the underlying technical systems. For this reason, we are more likely to observe *topological self-similarity*[76] in the former. Contrary to the flows' self-similar topology, the technical networks that distribute them are best mapped by a random graph that lacks topological self-similarity, as shown in the example in Figure 9c, for the email network generated by the GN algorithm from the random exponential network.

*Topological self-similarity* is the geometry of spanning graph, which overall tree structure is comprised of subtrees that share alike organization (Fig. 7). The scientific discourse on complexity and self-organised criticality (SOC) argues that the topological self-similarity is partially responsible for the self-organised efficiency of some natural systems. Topological self-similarity is a signature of replication of the network's structure and it is scale invariant (i.e. independent of the size of the system). Guimerà et al. (op cit.) argue that efficiency in social networks is the balance between "the need to cooperate and the costs of keeping active links". This is the reason that self-similarity is observed in the structure of

[76] such as the example presented in Figure 7.

their systems, which in fact is the product of their self-organized criticality (2003).

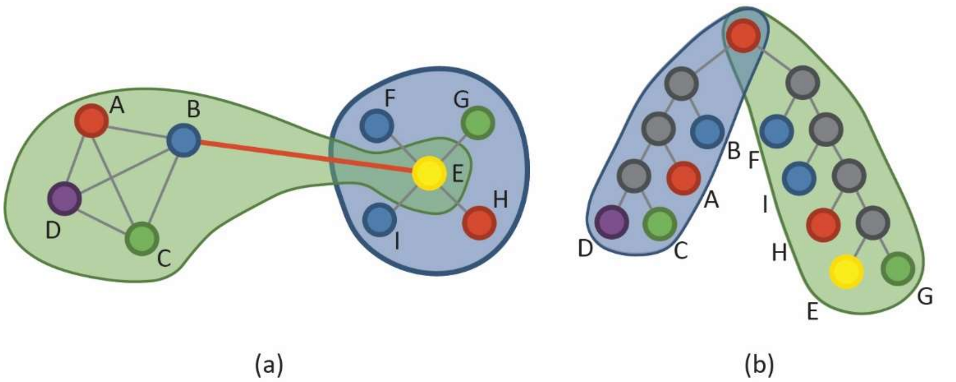



Figure 6: The links in the binary tree (b) represent community structure and not a link between two neighbouring nodes. Each branch of the binary tree (b) represents a community detected by the GN from (a). The most central nodes of both communities are at the lower levels of the tree (e.g. nodes B and E) (Guimerà et al., 2003, p2) (reused with permission).

Topological self-similarity is also observed in networks that have fractal (Bak and Chen, 1989; Batty, 2005a) and multifractal characteristics, since fractality (Batty, Longley and Fotheringham, 1989) is observed only in optimal structures (Rodriguez-Iturbe et al., 1994). Thus, social systems are inherently optimal. In that respect, <u>the topological self-similarity and fractals currently [not]observed in the flows of urban systems is an artifact of the centralized nature of their technical networks</u>.

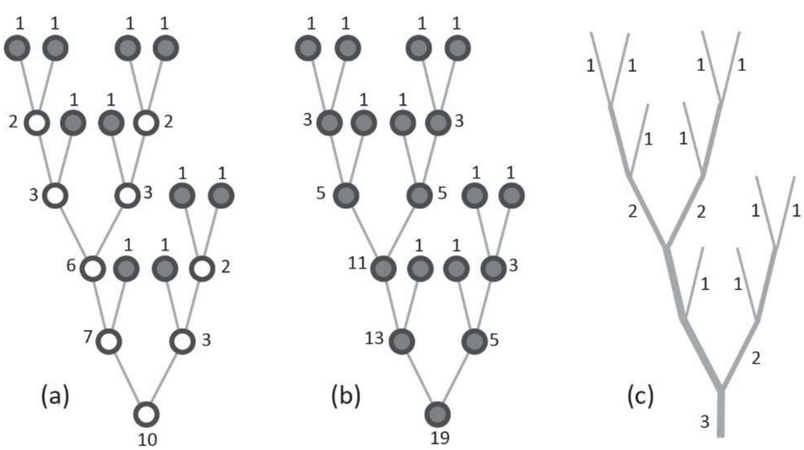



Figure 7: Example of topological self-similarity represented in a binary (hierarchical) tree generated with the GN algorithm (Guimerà et al., 2003, p 3) (reused with permission).

The black nodes in Figure 7 exhibit (a), represent the actual nodes in the network and the white nodes graphically represent the communities they belong to (e.g. nodes A and B belong to a community of size 2— i.e. two members— and together with node E, belong to a community of size 3— three members). Exhibits (b) and (c) are representations of the drainage capacity of a river network's basin to illustrate the application of this algorithm in planning and designing LTSs. The hierarchy of clusters, such as the ones in the spanning graph in Figure 7a, is often depicted by a dendrogram. The root of the dendrogram tree (the solid red line in Figure 8) connects all the nodes in a single cluster and each of the leaves correspond to a single node belonging to a specific hierarchical level. The Author's own illustration in Figure 8 is a graphical representation of the way the GN algorithm partitions the graph from Figure 6 in hierarchically arranged communities. The GN algorithm separates the central nodes (or hubs)[77] at the last iteration, and therefore they always appear at the lowest levels of the tree branches. We observe this in the example in Fig. 6b where the node E, which is the 'root' (i.e. the largest hub) of the blue community, is found in the lowest level of the branching tree[78].

Without understanding and presenting the topology of communities, it is impossible to explain the network's full topology based solely on the network's degree distribution (Barabási, 2016, p324) and that is the reason why the link communities topology model was central to this research.

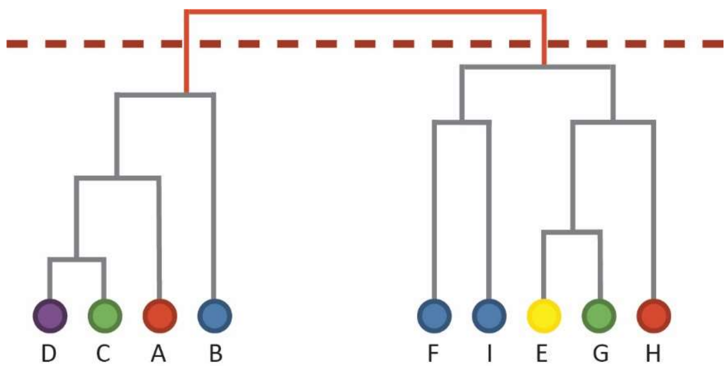



Figure 8: The hierarchy of the clusters of the graph from Fig.6 partitioned by the GN algorithm in two communities.

Communities of complex networks interact in tight clusters of agents (Barabási, 2016), each carrying a specific cellular function (Hartwell et al., 1999) that in a collective way the cluster of agents/cells distinguishes the community as a self-functional unit also named a functional module (Ravaz et al., 2002). The presence of community is solely conditioned by the principle by which agent/cell connects to which agent/cell. The community structure of a network is uniquely encoded in the community's topology that is captured at the different levels of the dendrogram as shown in Figure 8. The empirical analysis of the dendrogram's heights at which each community is discovered does empirically explain the network's wiring. Ahn et al.'s (2010) single-link hierarchical clustering algorithm, used to analyse the logical topology in this study, capitalises on the information encoded in the heights of the dendrogram and as illustrated in Figure 8, rather than guessing the community membership of nodes, uses a link partitioning strategy and removes first the links with highest betweenness value in the

[77] i.e. the ones that connect the two sub communities.

[78] instead as per the traditional representation of branching trees, where the root is typically found at the top of the tree, which makes the results a bit counterintuitive to interpret.

network (such as the link in red in Figure 8). It then iterates the link removal process based on their modularity value onto the lower levels of the dendrogram, until it reaches the nodes at the lowest that have no more links.

A community of nodes means that all of the nodes in the same bundle must be reached through members that belong in the same community. This quality forms the first feature of a community, which is *connectedness.* The second feature of a community is *density*, measured by the high probability for nodes that belong to the same community to connect with each other. When all of the nodes are connected with all other nodes in the same community, this condition in graph theory is known as a *clique (complete subgraph).* There are two main divisions (O'Sullivan, 2016, pp 51-53) of communities in complex networks:

- strong communities- also referred to as *Luccio and Sami (LS) sets* (O'Sullivan, op cit., p 51). The LS sets are communities where the number of links of each node within the community[79] is larger than the nodes' *external degrees*[80] (i.e. the number of links to external communities).
- weak communities- also known as *lambda sets*. The concept of lambda sets was first introduced in Borgatti et al. (1990) and in spatial sciences observed in O'Sullivan's thesis in the application of graph cellular automata (op cit., p 52). At the conceptual level, lambda sets are a version of LS sets with the difference that the demand of connecting every node within the subgraph (i.e. community) to the rest of the nodes in the subgraph is relaxed. These communities are more inwardly connected. Meaning their nodes have more links with each other than to the rest of the network. I.e. the total internal degree exceeds the total external degree. Nodes in a network form a lambda set only when more than one unique path connects these nodes in different ways to each other, than to the rest of the nodes. So, lambda sets are able to generate series of groupings hierarchically nested within each other, based on different qualitative aspects (i.e. categorical attributes) of the link type (e.g. race, gender, ethnicity etc.). In the lambda community the inequality of internal to external connections of nodes applies to the whole community rather than each individual node. Thus, the nodes' individual degrees in these communities are not important, and instead their collective degree is the relevant measure.

From the aforementioned Barabási (2016, p327) and O'Sullivan (2016, p 52) conclude that each clique is a strong community and each strong community is, by default a weak community. Furthermore,

---

[79] Mathematically represented by the *internal degree* $k_i^{int}$ of each node in the sub-graph.
[80] Mathematically represented by $k_i^{ext}$, which denotes the number of links that connect each node in that sub-graph to the nodes of the rest of the network.

although there may be situations where each weak community is a strong community, this does not mean that that strong community is by a default, a clique. From this can be stated that complex multilayer networks with many (if not all) strong communities are not by default also *complete*[81] graphs.

The size of the community and the size of the overlaps of the different communities (Palla et al., 2005) are another relevant property of the network. The *cumulative community size* distribution (Guimerà et al., 2003) denoted with $\boldsymbol{P}_{(\boldsymbol{s})}$ is the probability that a community has a size that is equal or larger than $\boldsymbol{s}$. The usual size of communities' overlaps in large networks such as; the co-authorship network, the network of word associations or the protein-protein interaction, follow [near] power law distributions, meaning that characteristic overlap size does not exist. A high fraction of shared nodes between communities represents the importance of the overlaps between these communities (Palla et al., op cit.). According to the same authors, when the network of communities, rather than the network of nodes, is considered in the analysis of the statistical properties of large real networks, a fat-tail distribution is still observed. The fat-tailed distribution is related to a characteristic scale, below which the distribution becomes exponential.

Already three decades ago was proven (Amaral et al., 2000; Newman et al., 2002) that the *six-degrees of separation* phenomenon (also known as *small world*), observed in scale-free networks, is in fact an artifact of the limitations (or costs) associated with establishing a connection — e.g. the costs of maintaining the social acquaintances active. The last argument presented here relates to the earlier discourse on topological self-similarity, fractals and self-emerging criticality. Thus, communities (and subsequently small worlds) are observed as a result of optimised costs.

The effect that the presence or absence of a community structure has on the network's diameter, is visually presented in Figure 9a-c. In Figures 9a and 9b we observe very short distances throughout the entire graph, which means low costs to get from one node to another. Unlike the dense connectivity of the social network Figure 9b, its random graph counterpart in Figure 9c has long connections (and therefore expensive since they costs more effort/time), which are a result of the lack of community organization.

[81] A complete graph is a graph where every node is connected to every other node in the graph. For example, a complete graph with 16 nodes has 120, or a complete graph with $\boldsymbol{N}$ number of nodes has $\boldsymbol{N} * (\boldsymbol{N}-\mathbf{1})/\mathbf{2}$ links.

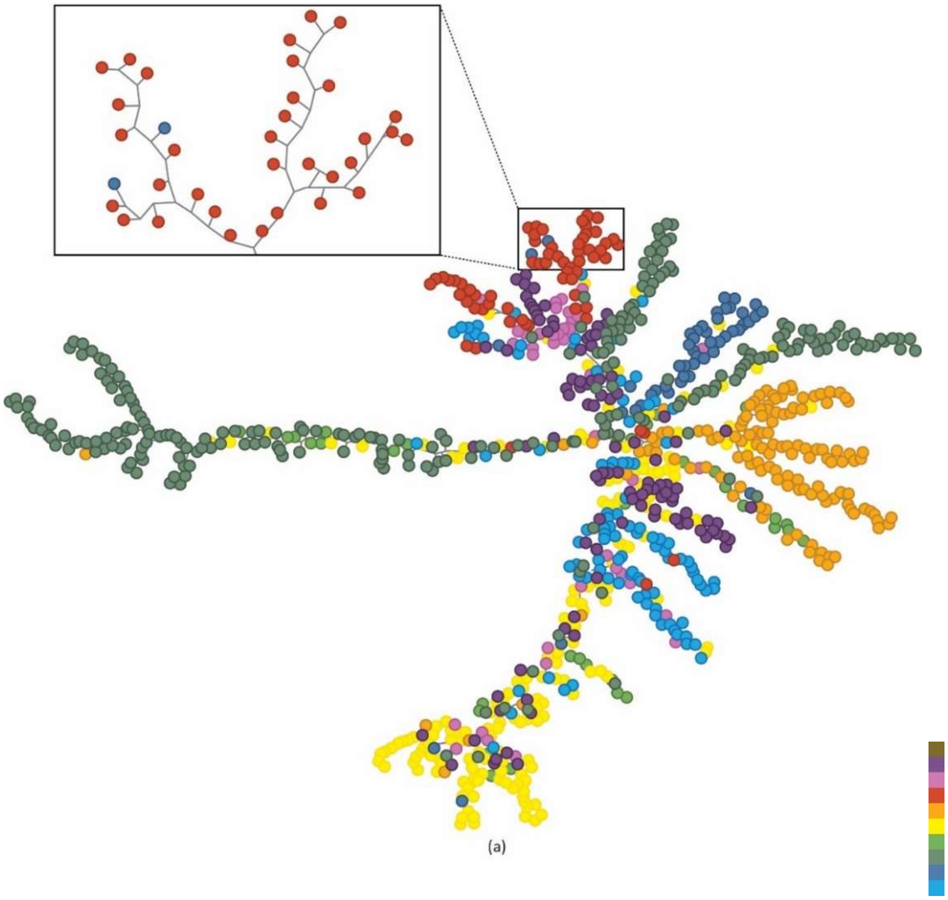



Figure 9a: Illustration of the topology of a network based on its communities' structure (Guimerà et al., 2003, p 2) (reused with permission).

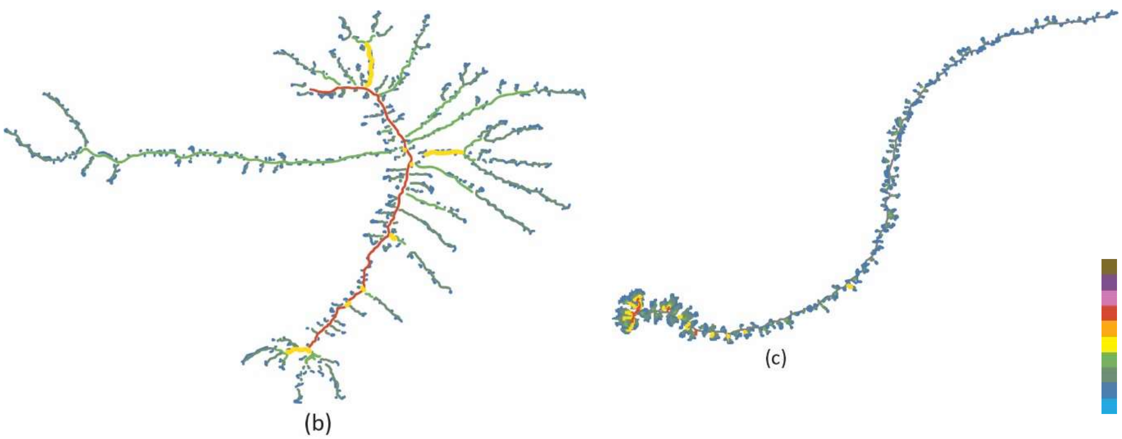



Figure 9b-c: Comparative illustration between the topology represented through the communities' of the social network (left) and a random graph (right) (Guimerà et al., 2003, p 2) (reused with permission).

Figure 9b illustrates the same graph from exhibit 9a without the leaves, for simplified visual representation. Figure 9c illustrates the hierarchical tree (dendrogram) generated by the same GN algorithm, describing the community structure of a random exponential network, with an exponent similar to the one observed in the real network converted into the graph, presented in Figures 9a-b.

From these images it's clear that the random wiring of the network barely has any community structure, indicating strong hierarchy with [nearly] single root (i.e. main source) — **confirming the neutrality of the network's degree distribution**. The graph in Figure 9c has very long tailed structure and only one dominant community which acts as the graph's root. It visually shows the significant effect absence of community organization has on the graph's size, which would have otherwise reduced the graph's size.

### *7.2.2 Random graph models: application in urban planning*

Random graph models are used to capture the topologies of various technological networks (Wang and Provan, 2008). They were originally introduced with the *Erdős–Rényi (ER)* model (Erdős and Rényi, 1959) as a tool for planning and designing infrastructure networks. The original model was able to predict the probability of observing a connected graph when the largest component has $\boldsymbol{k}$ number of links. The model calculates the number of iterations needed which are bounded by some rules, as explained earlier (i.e. preferential attachment or nodes' age) to add new links to the graph, until it becomes fully connected. However, this model and its early successors failed to implement the spatial dimension (Kaiser and Hilgetag, 2004; Masucci et al., 2009) of real networks. In reality, the topology of a spatial network is strongly limited by its geographical embedding (Wang and Provan, op cit.).

Topology model analyses of real-world networks are important to improve planning for costs and optimization of established and new connections, and moreover improve the organization and accessibility across spatial cells (i.e. nodes). Explanatory models with spatial constraints such as the *Spatial Preferential Attachment* (Hansen et al., 1999) or the *Optimization Model* (Li et al., 2006) could provide the underlying principles that are shaping the topologies of specific complex networks (Boccaletti et al., 2014) and have better predictive and rescaling power for topology generation (Song et al., 2017) (e.g. the addition of new nodes or links). However, using these models in urban planning is still emerging, limiting their application over few existing urban systems, usually with the goal to upgrade or improve them. In those use cases, often there is a little room to experiment on alternatives due to already limited land use or other spatial and technical restrictions, bound to the existing conditions.

Recently, Wang and Provan (2009) introduced the *Descriptive Generalized Random Graph (GRG) Model,* which generality of use allows the model to be extended in various ways to represent real-world networks. The GRG model reproduces the degree distribution of the real network directly and unlike the other two models mentioned above, matches the topology of the mobility flows of the German highway system much better. All three models are independent from any systemic growth process and reproduce prescribed degree distributions, which are shaped and constrained under domain-specific spatial restrictions. The GRG model outperforms the former two, since its input respects the original sequence of the real network. Moreover, the results from the GRG model showed that the German highway system demonstrates a non-small-world property, and as a subsequence it lacks community structure. This ties to the conclusion in p. 71 that the absence of self-similarity in the topology of flows across infrastructure networks is an artifact of the technical network that is highly centralised.

In conclusion of the study, social systems are inherently optimal and the fractal dimension is typically observed in them, thus the mobility network of the German motorway system if not bounded by the centralised nature of the road network should have confirm this behaviour, yet it did the opposite.

### *7.2.3 Conclusion*

Form the review in this subchapter I understood that prescriptive models, such as the GRG, can support studying the success rate of modelling the future growth of infrastructure networks based on their current and future topologies. Furthermore, the random models developed this far, predominantly focused on capturing of the structure of processes (e.g. mobility flow from A to B), treat supporting infrastructure as static system. I addressed in this study this limited view of infrastructure; infrastructure as a complex system infiltrates the natural habitat and therefore cannot be treated as a static entity. Networked infrastructure conveys flows of resources to support the coupled human-nature system and precisely these flows comprise its dynamic dimension. The physical topology of these complex networks, although computationally expensive, can better explain, predict and optimize their overlapping community structure.

## 7.3 COMPLEX (MULTILAYER) NETWORK MODELS

In spatial analysis, network data is considered to be a variant of “line data” and is used to express connected finite set of lines that represents some connected object — e.g. road network, a network of rivers or rail network (Fotheringham et al., 2005, p 22). This spatial entity is also known as complex line (Schneider and Behr, 2006, p8). However, the understanding of a network and its current representation with GIS-based tools in such a narrow sense is very limited (Albert and Barabási, 2002; Costa et al., 2008). First, any spatial networked entity not only carries linear data, it also carries nodal data. Although nodes are located on the intersection of line segments, they are spatial entity on their own account, defined by a point with unique spatial and attribute components. Second, the physical networked infrastructure (Graham and Marvin, 2001) represented by “line data” is a part of more complex, multi-layered network system (Boccaletti et al., 2014). This system, as already briefly introduced in Chapters 3 and 4 has both, physical and social properties. This thesis was dedicated to connecting and modelling the complexity of these systems, not only represented by their multiple layers but by their simple and complex objects. Ford Jr. (1956) defines a network—also called *linear graph*— as a finite collection of nodes that may [or may not] be joined together by a link. He also acknowledges the associated direction of a link, its length and capacity. Every link has an origin and terminal node and is limited to some sort of constraint; [1] either physical capacity constraint; or [2] cost constraints of building and maintaining the link or of transporting flows through the link.

Complex networks are spatial constructs with unique statistical mechanics (Albert and Barabási, 2002) whose layers are mutually affecting their distinct dynamics. Boccaletti et al. (2014) in their study summarised the findings of biology scientists, which for the first time were able to present the neural network of a nematode C.elegans as a multiplex network comprised of two layers. Each layer shares the same neurons; however, their connection is by either a chemical link or an ionic channel. The wirings displayed different network dynamics (i.e. flows). A node of high centrality in one layer appeared of marginal centrality in the other one. This is one of the simplest examples of complex multilayer networks.

*Multilayer networks* (Boccaletti et al., 2014; Afsarmanesh and Magnani, 2016) explicitly incorporate numerous channels of connectivity and constitute complex systems interconnected through several categories of linkages. Each layer represent the various interactions of the same nodes at different levels. Depending on the interaction layer, one node can have multiple different neighbours that are not neighbours with one another.

As specified in the extensive review by Kuby and associates (2005), there are three main drawbacks or limitations so far to using complex network theory in spatial planning and decision-making. The first is related to the lack of incorporating the link and/or hub capacities in the network modelling process. The principal object of investigation in complex network theory is the [non]existence of a link between two nodes. Physical systems, unlike the social network or the WWW have limits to the capacity they can deal with. When modelling complex systems that are inheritably uncertain and infinitely complex the only way (Batty and Torrens, 2005) is to know the limits of the system that is being modelled and to develop the model in the constraints of these limits. The second limitation is the limited understanding of this method to its potential for planning a cost-efficient network. For example, complex network theory can model the existence of a link without the temporal attribute of the link. Finally, complex network lacks a method to predict the volume of interactions, necessary for planning of infrastructure networks. Kuby et al. (op cit.) conclude that without considering node sizes or interaction distances, complex network theory is not appropriate tool to use for modelling

[near]planar networks such as infrastructure. This was addressed later in the thesis in the modelling phase, by modelling *spatial accessibility* as a proxy to the infrastructure (link) capacity on the network.

### *7.3.1 The spatial nature of infrastructure networks*

Infrastructure networks (known also as '*almost planar graphs'*) such as the transportation, the power grid and the water and waste distribution networks are crucial to our societies and their seamless operation is sustaining the balance between urbanized areas and ecosystems. These networks are considered to be approximately planar because there exist links that cross each other yet not embedded in the same elevation, such as roads networked with other roads, railways or mountains and rivers by forks, tunnels or bridges. Therefore, links pass each other in the third dimension without intersecting (Boeing, 2017, p 83). They are designed to overcome separation between geospatial entities (Kuby et al., 2005). In the geospatial representation, these networks are normally considered as connected systems (Kuby et al., op cit., 9) of 'points' and 'lines' that enable 'movement'.

Most of the properties of infrastructure networks are guided by the spatial constraints (Barthélemy, 2011) among which some are more prevalent than others (Barthélemy, 2016, p 83):

- They have peaked degree distribution, due to the way these networks are designed. Thus, if their physical topology is the only aspect being considered, large number of junctions (Batty, 2004) lead to maximum four different street segments, and some to even less than that (i.e. mainly two). E.g. most of the train and metro stations lead to two other towns/metro stops. In rare cases, street intersections have more than four road segments. And for the ones that do, in the perspective of the entire network, are underrepresented in the distribution plots. I.e. the overwhelmingly predominant number of three and four way intersections, determines the global measure of degree distribution.
- Their assortativity and the clustering coefficient are very large; highly connected nodes connect to other highly connected nodes and nodes with low connectivity connect to other nodes with low connectivity. Contrary to this, networks whose high and low degrees node pairs are not correlated, such as the power grid, are more resilient (Solé et al., 2008).
- The most relevant information in infrastructure networks is encoded in the spatial distribution of their nodal *betweenness centrality*; the number of paths passing through a node that connect pairs of nodes. Betweenness centrality is important in decreasing the network's distance—e.g. establishing new links between any two nodes by introducing shared node to shorten their distance (Janssen et al., 2006, p 6). The way the nodes located in-between the hubs are connected to the rest of the nodes in the network is of crucial importance in these sparse networks.

Infrastructure networks are near-planar and are graphically represented by *planar maps*. Planar maps are the combinatorial embedding of planar graphs[82] (Stanford University, 2000).

Bogart (2009), examining the development and interdependence of canals, roads and ports, over the period of the English industrial revolution, discovers significant interdependence among nodes belonging to different networks based on their functional and spatial proximity. This is evident (Parshani et al. 2010) for example from the newly coined measure of networks' *inter-similarity* that was developed from analysing the structure of the coupled worldwide sea and air transport networks. The networks' inter-similarity showed that well-connected airports tend to connect with well-connected seaports. From the perspective of infrastructure networks, the argument made by (Ducruet and Lugo, 2011) that very often new network extensiosn are first built to support and later compete with the original network was challenged by the Parshani et al. (op cit.) and needed to be empirically validated. Therefore, the scientific question posed in this thesis was whether these coupled networks are more sustainable due to their underlying physical structure that follows their logical topology or the inter-similarity of the coupled networks' topological structures[83]?

### *7.3.2 The optimal network problem and its limitations to plan sustainable infrastructure*

Complex network analysis is of great importance for the basic structuring of large number of dynamic and organizational problems. There are two basic classes of complex network problems (Scott, 1971, p 95):

1. The problems of optimizing networks with physical expression that have a consequent investment— such as the road or the rail network;
2. The problems of optimizing networks that do not require any capital investment since their topology portrays a relationship rather than a physical connection. Examples of such networks are the airline routes[84] or social group formations. The relationships between the nodes in these networks are usually represented by a *sociometric graph.*

In the case of solving the problem of optimizing a complex network in regards to the commodity needed to be conveyed and the stages of the network construction, there have been developed

---

[82] Planar graph is a graph embedded on the surface of a plane.

[83] i.e. topological inter-similarity means that the topology of the seaports network to a great extent resembles the topology of the airports network.

[84] Although these networks as well have capital expenditure in terms of monetary or time costs conditioning their existence—e.g. airport personnel capacity, terminal building capacity and passenger demand conditioned by their budget. The distinction is made in the path-finding problem solving method, which separates them from the physically constrained networks since their relationship is conditioned on materialistic presence of roads made of stones and sand, paved with asphalt.

several variations of network structures. The outcomes of the network structures are arc (i.e. edges/links) capacities and flow patterns that are in perfect equilibrium (Scott, op cit.). In these problems, the arc capacities are defined as *continuous variables* (variables that have unlimited number of categories) and the solutions are reached by linear programming. Physical networks are usually planned to be optimized by applying constraints to the arcs that limit their capacity to pass on flows.

On the other hand, in the case of relational networks, the optimization strategy generally includes applying a constraint to the nodes of the system to absorb, pass on or initiate any flow (Scott, 1971, p115). Relational networks are also considered to be special case, subsumed within the first network class. Thus, the optimisation strategies are conflicting from the onset of the problem solving and therefore complex systems such as the networked infrastructure (Graham and Marvin, 2001) in urban agglomerations are challenged in handling sustainable distribution of resources. Moreover, it is known that infrastructure networks have high initial investment requirements and considerable accumulation of operating costs, so the capacity constraint on the node is open-ended[85]. The three main and mutually complementing strategic approaches to optimising infrastructure networks include [1] minimizing of total operation costs constrained by total capital expenditure; [2] minimizing total capital expenditure on the expense of least permissible system performance or [3] minimizing jointly operating and capital costs.

This type of network problems is of great importance to the analysis and planning of flows in the systems that are very likely to be greatly influenced by the behaviour of their flows. Such systems are the traffic systems and the majority of ‘pipe-line’ infrastructure systems. Specific basic principles in the mathematical representation (modelling) of these types of complex networks are applied. However, they only consider the flow behaviour to be of secondary importance (Scott, op cit., p 95), prioritizing the topological and geographical structure of the network.

Therefore, the thesis addressed the behaviour of flow by explicitly recreating a separate ‘logical’ graph of the physical network, in that way making this a primary problem to be solved in the modelling process. Secondary importance was given to the model of the network’s physical topology, ignoring completely the capacity constraints in either of the models due to time limitations. The following section has more discourse on the scientific reasoning to why the capacities of the network were ignored in this thesis.

[85] Usually the greater the node capacity is, viewed from the eyes of urban and territorial planning for example as population density or undeveloped land availability, the more justified the infrastructure spending.

### 7.3.3 Network flow problem: maximal amount vs. maximal steady state vs. phase transition

To understand the way networked infrastructure (Graham and Marvin, 2001) are planned and operate, and furthermore, to empirically investigate the organization principle behind these networks in both their physical and relational aspects, I first reviewed the basic mathematical formulation of the optimisation problem. In this section, I briefly elaborate on the two out of the three known basic problems in network flow theory (Ford Jr., 1956):

- the maximal amount of goods transported from origin to terminal node in a time constraint;
- the maximal steady state flow of goods transported from origin to terminal node, regardless of the cost constraints.

For ease of reference, I call the first problem *maximal flow* problem and the second problem *maximal steady state* problem. The second problem appears as a sub-problem in a large number of transportation problems and its goal is to maximise either the flow instigating from the origin node or ending at the terminal node. In this form, this problem is constructed by linear programming since the relationship between the variables is linear. Moreover, the construction of the problem focuses on maximising carrying capacity of the physical part of the network rather than optimising the source's capacity. The capacities of nodes and links in the network were beyond the scope of the study. Thus I chose to set up the logical graph models as a modified version of the maximal steady state problem. And rather than calculating maximum carrying capacity, I used the cut-off distance as a proxy[86] to determine the carrying capacity of links for the existing network. The links did not have an upper capacity limit in this model. The maximum steady state of the network at hand was then solved only by finding their s-t cuts, which essentially are the links with the highest betweenness (as presented in Figures 21-29).

### 7.3.4 Modularity of complex networks

Modularity is a function that measures the quality of the structure of any existing community within a graph (Law, 2017). Modularity was first introduced by Newman and Girvan (2004) and calculates the difference between the existing and expected number of edges in the subgraph, part of the graph i.e. community that has tightly connected nodes with each other. Modularity can be computed after the community structure of the graph is known and its value can span anywhere on the scale between -0.5 and 1. After their initially proposed, and now well-known GN algorithm, many more efficient

[86] The cut-off distance was pre-defined from the onset of the graph representation of the flows as the logical graph. It was defined as the 400 meters network-based distance between OD points which determined their accessibility—i.e. weather a sink (destination) node is directly accessible from a source (origin) node.

algorithms have been proposed (Law, 2017, p 7) that can be applied to multilevel networks to optimize computing time. Such is the *multi-level modularity optimization algorithm*, often also referred to as the Louvain algorithm, one of the most recognized methods for its accuracy and efficiency (Blondel et al., 2008; Lancichinetti and Fortunato, 2009). In the spatial context of social systems, Law (op cit.) successfully applied this community detection technique in order to explain the socio-economic homogeneity of neighbourhoods (i.e. neighbourhoods consisted of people from similar or the same socio-economic class).

Since determining the graph's modularity is considered to be an NP-hard problem, solving the modularity problem using an optimization strategy for large datasets, especially if this would include both link and node capacity constraints, remains to be impossible. Due to this reason, despite its reliance on heuristics, I considered the multi-level method (Blondel et al., 2008) as the most appropriate one to use in this thesis that approximates the physical infrastructure. Second reason for favouring this algorithm was its successful application against the modularity function used in another spatial study. In Law's study (2017), the use of the multi-level methods for community detection were successfully applied to detect Street-based Local Areas (SNA) as socio-economic cohesive units that offer alternative and more appropriate unit scale of spatial analysis. The heuristic method was also used by Blondel et al. (2008) in different spatial study that performed Social Network Analysis of 2.6 million Belgian mobile phone network users.

Despite the fact that this algorithm relies on heuristics and furthermore, was proven (Lancichinetti and Fortunato, 2009) to have second-best performance speed after the Infomap algorithm, its success in accounting for the spatial autocorrelation of the error term (Law, op cit.) was the third reason for which this algorithm was preferred over Infomap. Spatial autocorrelation has a great importance in spatially explicit graph analyses, since geographies of spatial proximity tend to be more similar with each other than geographies far away.

The results presented in Chapter 8 show the value for which the logical topologies in the eight graphs studied in the thesis was maximised with the use of this algorithm. The results further show that at the maximum value of the current network's modularity, for which reasons explained further in the text excludes large part[87], the eight graphs have very low number of communities and links with high edge betweenness. These areas of the graphs are the areas where nodes are shared across multiple communities and due to the existence of the areas it is said that the communities in these graphs overlap (Figures 17-20). This suggests that these networks have great potential in increasing their

[87] More specifically it excludes over half of the total number of nodes in the entire study area.

community overlaps and presence of links with high edge betweenness with the goal to improve their sustainability.

## 7.4 DEFINING THE SPATIAL INTERACTION IN THE GRAPH MODEL

Urban processes are defined (Cheng, 2003) as a sequence of changes in time and space and can be spatial and temporal. The sequence of changes measured over time is called temporal process and the sequence of changes measured in space is called spatial process. In reality, both processes are interconnected and any urban process is bound to have both spatial and temporal dimension. Urban flows and processes are used interchangeably. The regularities in the logic of space are called space patterns while the regularities in changes over time are called temporal patterns (Cheng, op cit., p 6). Many studies (Ford Jr., 1956; Batty, 2004; Oberoi et al., 2018; Nourian et al., 2018) had developed a mathematical system to model urban interactions, representing the topological relationships in networks in a mathematically comprehensible manner. This system can be easily transferred into network science and is known as the primal and the dual syntax problem.

The primal and dual mathematical problems are of a specific interest to this research, since they allow to statistically study the accessibility and distance of locations in a city measured by a step-length[88]. The step-length as a measure counts the number of street segments that have to be 'walked through' by an agent to reach location B from location A (i.e. node B from node A). Thus, the variable that is measured in the primal and dual syntax is the number of total 'directions' changed, rather than the segment's geodesic lengths. The primal problem is the basis of angular analysis, a method that allows quantification of space (Turner, 2000). Angular analysis is also the mathematical foundation of *space syntax* (Hillier et al., 1976; Hillier, 2007). Turner (op cit.) defines angular analysis as one of the many methods used to predict and simulate movements through spatial systems.

Space syntax was developed as a more sophisticated tool of measuring street accessibility in a city (Yamu and van Nes, 2017). Although more sophisticated, one of its shortcomings is basing the analysis on step-length measures traversing the network, resembling the small-world property in network science, which was concluded in Chapter 5, is not a signature of any particular organization in a system. Although space syntax can successfully predict movement of people and traffic through urban areas mainly in two-dimensional space, angular analysis has a goal to introduce fine-grain detail for

[88] Simple step-length in the context of space syntax does not refer to the geodesic nor the Euclidean distance between points. It is a measurement of the number of steps needed to reach a destination point from a start location. Simple spatial example is one person standing at the corner of street C, aiming to reach street E. In the computation of the shortest path between these two points, if the individual goes through two other streets (A and H), it is said that the step-length distance of this trip is three. Namely, [1] she walks from C to the intersection of C with A, to [2] reach A. From A, she goes through the intersection between A and H to [3] reach H which intersects with E.

the same analysis considering the third dimension (Wang, Zhu and Mao, 2007). This is mostly of interest in behavioural and cognitive studies to predict interaction rules of the way people use physical space, based on their cognitive capacity (op cit.).

### *7.4.1 Defining cellular (spatial) relationships as a primal graph*

The longstanding research that interprets the urban form by visualising topological relations firmly embedded in Euclidean space using graphs, dates back to Nystuen and Dacy (1961). They used the representation of the *primal graph* to measure the gravitational potential and centrality of places. The gravitation of nodes in graph representation is measured trough the collective weigh of all the links that begin or terminate at that node. In network science, these are called *in*- and *out*- degrees of a node and the metric is the degree centrality of a node. The primal graph problem (Batty, 2004) is the mathematical foundation of the traditional *space syntax* (Hillier et al., 1976; Hillier, 2007) tool, that with the construction of a bijection graph, the tool studies the abstract relation of streets through their shared junctions. It needs to be noted that this type of graph representation is no longer embedded in Euclidean space (i.e. is not spatially explicit), since it only represents the relational dependence between two sets of spatial entities— the relation between streets through their intersections (Batty, 2004, p4). In the primal graph syntax, each street [segment] is represented by a link and the origin and terminal nodes are the intersections of the street segment with adjacent street segments. If two street segments share a node, then that indicates that there is direct access from the one to the other, unless the link (street segment) is restricted by one-way movement. In the primal problem, accessibility explains the accessibility of the street segments; e.g. how many street segments a street segment of interest can be accessed from.

In space syntax, the focus is on streets, rather than intersections, that are analysed by their depth and accessibility (Hillier, 2007; Ratti, 2004; Van Nes and Yamu, 2018).  The level of accessibility is calculated as a function of the nodes in- and out- degrees. As the number of nodes associated with the street gets greater, the street importance increases. The value of access is expressed by an index, also called *integration*, by taking and inverting each mean value of the distance measured in simple step-length (Batty, 2004, p13). However, considering accessibility in a simple step-length distance ignores the relative importance and strength of paths' length in the graph (i.e. the Euclidean or the geodesic length of the link). In the case the step-length distance increases, the relative importance of a path gets smaller (Kaiser and Hilgetag, 2004).

This was addressed for the logical graph in the study by setting up the 400 meters cut-off distance to define the existence of a link.

### 7.4.2 Defining cellular (spatial) relationships as a dual graph

The dual graph problem (Batty, 2004; Boeing, 2017) is the more usual morphological representation of the relationship among junctions trough their street segments and the accessibility of any place on the junctions. It models the reachability of any junction, from any other junction, through the connecting street [segment] of the network seen as a connected graph. This problem measures the individual level of accessibility of a junction. The shift of focus, proposed by Batty (2004), from the primal problem that is the principle concern of space syntax analysis, to the dual one, has great importance when the accessibility is projected as a bipartite graph. The dual graph is sometimes named as *line-graph* (Evans and Lambiotte, 2010; O'Sullivan, 2016). The bipartite graph of the dual problem of the street network, counts the number of paths between any pair of junctions in terms of their common streets. In any other case, except for a restricted traffic network, this matrix is symmetric.

In the representation of the street network, it is counterintuitive to use nodes to represent linear - entities and lines to represent point entities. Therefore, the use of bipartite graph allows for easier navigation from the primal into the dual problem. "The primal and dual problems interlock with one another and this has a practical implication of the way how point accessibility can be translated to line accessibility and vice versa" (Batty, op cit., p12). Additionally, the dual problem is interoperable with Global Positioning System (GPS) tracking data, since both the junction data model and the GPS tracking data model use the global positioning coordinate pair (Nourian et al., 2018).

Essentially, the original space syntax accessibility matrix (and its respective graph representation) is usually sliced, a practice aiming to reduce the computational complexity (Batty, 2004). The slicing of the graph loses any information on the relation between a pair of lines (e.g. street segments) and information regarding the strength of self-loops (Batty, op cit.). This is especially important when considering restriction of movement (e.g. one-way streets or dead ends).

The dual graph, also known as a *visibility graph*, is associated with the primal graph, which since it is embedded in a plane is also called a *planar graph* (Batty, op cit.). *Plane graph* on the other hand is the graph already embedded in a plane such that none of its edges are crossing curves. The definition of the dual graph depends on the possibility to embed the original graph as a plane graph, and therefore the dual graph is the property of a plane graph rather than then the planar. Street networks not always can be represented by planar graphs and therefore are considered as almost-planar. These networks can sometimes contain overpasses, forks or underpasses, information only conveyed once we construct their dual graphs without slicing the matrix.

By developing more comprehensive framework, Batty (op cit.) enabled the shift between the two problem forms and with that he “relates the accessibility measures between each problem” (Batty, op cit., p 9). The morphological, or also known as topological predicates between two distinct set of objects represented in their most elementary form, are essential in developing unified framework between the two problems. In order to study the unique topological relationship of two sets of spatial entities to each other (e.g. streets and intersections), this relationship which is a function of bijection, must be unambiguous. I utilised this framework and the existence of a relationship between the nodes in the physical topology was constructed by representing this network with its dual graph. Thus the focus here was on the level of accessibility of buildings and intersections via the street segments.

In conclusion, the primal representation of the street network crates the foundation of space syntax (Hillier et al., 1976; Hillier, 2007; Batty, 2004; Boeing, 2017). Although space syntax-based tools are regarded as highly useful in analysing the accessibility of streets in the street network by the use of angular analysis, recently have been regarded as “less-than-ideal” representation (Boeing, 2017, p 87) of a street network. Unless the network is fully planar (when it doesn’t contain any bridges and tunnels), the primal representation of street network could potentially produce inaccurate metrics, underestimating length of edges and their betweenness, or centrality of nodes (by misrepresenting nodes on locations where roads only pass over each other through tunnels and bridges).

### *7.4.3 Discussion on the Primal and Dual Syntax*

From the discourse on the primal and dual syntax, several conclusions for the new study were made. First, it was noted that the construction of any type of bijection[89] to classify planar graphs, does not contain any information on the geometrical shape of that planar graph (Barthélemy, 2016). Two spatial network patterns may have the same topology, but have completely different spatial structure and topology comparison cannot be made based on their bijection graphs. Therefore, it was important to take into account not only the topology of the planar graph, but also the geographical position (coordinate pair) of the nodes, therefore all the graphs were generated to be spatially explicit.

Second, considering the accessibility in a simple step-length distance ignores the relative importance and strength of paths’ length in the graph. For spatial graphs, empirical evidence (Kaiser and Hilgetag, 2004) suggested that as the step-length distance increases, the relative importance of a path gets smaller. In practice, the average depth of ‘space syntax’ graphs is small. In network theory, it is calculated as the degrees of separation and does not carry any information on the geodesic distance given two nodes (Barabási, 2016, pp 89-92). Thus, the graphs are sparse and the observed number of

[89] Bijection is the process of mapping each element of the first set to exactly one element in the second set.

links (and not their actual length) is much smaller than the maximum possible connections the nodes in the network could have. This is the general observation of all real networks such as the WWW, the neural network of the C.elegans worm or the actor network (Barabási, 2016, p 53). The exception from this general observation for the real networks modelled so far are the space syntax networks, which have large diameters. This is due to their spatial dimension. The large diameters were as well observed for the eight physical graphs in this study. From the results presented in 9.1 we see that indeed all of them had extremely large diameter, especially compared against their line-graphs counterparts.

Following the aforementioned, if a network has extremely low depth or distance, rarely being more than six or seven steps, the size of the step-length (i.e. its length) herein bears key information and therefore its geodesic distance should not be disregarded when computing the street integration in space syntax (Batty, op cit., p 22). The consideration of the distance as a factor gives the interpretation of the conventional space syntax 'line diagrams' a sense of trends within the whole system of networked streets. As already addressed in the previous chapters, the integration of places (i.e. the construction of the logical graph) in this study was measured by setting up a cut-off distance beyond which a link was not existing between two places.

Third, as explained in 8.4.2, in most cases when constructing the dual syntax, the relation of presence or absence of a link is considered as symmetric and its value is always binary[90]. The element of symmetry in the relation between two nodes in the street network is already misleading. Simple example is a one-way street segment that starts at a junction (node) A and terminates at B. A car could reach node B from node A, but not vice versa. However, a car flow would not have symmetric bipartite graph, yet the same network could be different for other types of flows such as pedestrians and bikes which usually are symmetric. Therefore, the construction of one bijection per flow[91] (i.e. separate logical graphs) (Sim et al., 2015) was considered the most appropriate approach when dealing with accessibility and reachability of nodes in the network, such as the ones in this thesis.

Batty's (op cit.) second conclusion (i.e. the importance of the step-length geodesic distance) concerns the representation of both node and link accessibility matrices. Although the primal and dual problems have been long known to scientists, some studies (Solé et al., 2008; Ducruet and Lugo; 2011; Boeign, 2017; Boeing, 2018) that use graph theory to represent the relationships (or functional links) between nodes, still confuse lines that demarcate relational link (interaction or a spatial predicate)

---

[90] Binary value can take one of the two possible states — value of one when a link exists between two nodes or value of zero otherwise.

[91] Although in this thesis technically I only modelled one flow, this was worth including in the conclusion from the literature review as a reference for future research.

between two spatial objects, for a 1-D spatial object (a linestring or a linestring segment) that has the length property. An example of this confusion is a street and its two intersections. The street is represented by a linestring as a 1-D spatial entity, while its intersections at the opposite ends are two separate spatial entities represented by a 0-D object, a point. Both spatial entities have distinctive attribute values (see for example Figures 11 and 12); the 1-D object consists of a length, while the 0-D object has a coordinate pair as an attribute. This approach results with misunderstanding of the main difference between spatial physical and relational networks, thus confusing the relational (i.e. functional or logical) and the physical network models as the same thing.

Regardless of the primal or dual problem, when representing the street segments as links and the intersections as nodes, the abstract relation between pairs of street segments or intersections is lost. Street segments do not need to be adjacent with each other (to meet at an intersection) to be reachable from one another. Thus, the relational link in the graph representation of the street network demarcates a functional dependence between intersections or street segments respectively. To avoid future misrepresentation and confusion, this study adopts uniformed approach (see Fig. 12), reproduced after Wand, Zhu and Mao's (2007) example.  Namely, both street segments and building polygons are represented as lines in the primal syntax (see Fig. 11a and 11b) and thereafter they are converted to nodes (see Fig. 11c), in the dual syntax of the problem which allowed measuring the accessibility of places through their street connections.

## 7.5 CONCLUSION

The first conclusion from the review and discussion presented in this Chapter is the methodological difference between the community detection and the graph partitioning problems (Barabási, 2016, p327). The latter one divides the network into pre-defined number of subgraphs, while the former detects natural communities within the network— thus the inherent community structure of the network under study is maintained. In the graph or network partitioning problem, one must know the exact number of subgraphs that the network is comprised of, while in the community detection problem, the final result shows the actual functional communities of the network's compartments, or modules whose core operation continues to exist even though some links are removed from the network. These links are also known to be the ones that have high *edge betweenness centrality.*

Detecting the number of ways in which nodes can be grouped in communities within a network is often done using the simplest community-finding problem also known as *graph bisection*. This problem aims to minimize the network's *cut size*, while the network is divided in two non-overlapping subgraphs. A <u>*cut*</u> in a network is any set of links whose removal disconnects the origin and terminal nodes and the <u>*value of the cut*</u> is the sum of all link capacities in that set (Ford Jr., 1956). Any cut determines a cut-set, which in flow network (called the *s-t cut*) requires the origin and terminal nodes to be in different subsets and the direction of links is explicitly defined— from the origin to the terminal node. The smallest possible cut (i.e. the minimum cut size) of a network (i.e. a graph) is the one that divides the network into subgraphs (i.e. communities), however these subgraphs are still sharing some nodes and in that way are not completely disjoined from one another. In terms of computation complexity, detecting the minimum cut size of a graph with large number of nodes is high demanding problem. The brute-force strategy in such network problems is bound to fail (Barabási, 2016, p329). Therefore other methods were reviewed extensively and both, the single hierarchical link clustering and the multi-level modularity optimization algorithms were selected as the most efficient ones to use.

From the aforementioned, this research applied a version of the maximum steady state problem by using the community detection algorithm. In the first graph model type representing the logical topology of the eight shires under study, I found the cut-sets of the respective graphs (Figures 17-20). Additional regression analysis showed the correlation between the nodes' parameters as part of their community, regressed against their metrics as independent network components (i.e. degree, betweenness centrality, closeness and constraint) presented in Table 5 found in Chapter 9.

# Chapter 8: Data collection and pre-processing

The complex network model of Glasgow and Clyde city region included eight shires; the City of Glasgow, East Dunbartonshire, East Renfrewshire, Inverclyde, North Lanarkshire, Renfrewshire, South Lanarkshire and West Dunbartonshire. The land use characteristics of the neighbourhoods by shire are further presented in Table 1. After filtering the buildings based on their uses from the datasets, for the purpose of the study four main use categories were created. Table 1 presents an overview of the total number of buildings included in this study grouped in the following categories:

i. Category I— residential buildings, residential units with retail on the ground floor and community-institutional and communal accommodation buildings. All spatial features in this category include both single and multi-household units, and each is represented with one spatial count;
ii. Category II— important buildings or important locations (e.g. schools, hospitals, community centres, parks etc.);
iii. Category III— buildings that are of commercial or other cultural importance (e.g. theatres, cinemas, shops etc.);

iv. Category IV— All other remaining buildings including the ones designated for commercial use, as well as the ones that didn't belong in any category in the original data source (e.g. agricultural-unclassified, office buildings, recreation and leisure, transport etc.)

| **Council Area** | **Residential buildings** | **Important buildings** | **Points of Interest** | **Commercial and other** |
|---|---|---|---|---|
| City of Glasgow | 123.391 | 606 | 30.223 | 13.978 |
| East Dunbartonshire | 38.983 | 108 | 4.855 | 645 |
| East Renfrewshire | 31.370 | 84 | 4.092 | 474 |
| Inverclyde | 19.438 | 101 | 4.522 | 1.204 |
| North Lanarkshire | 113.917 | 375 | 15.839 | 2.751 |
| Renfrewshire | 56.437 | 128 | 9.684 | 2.085 |
| South Lanarkshire | 112.767 | 375 | 191.196 | 3.279 |
| West Dunbartonshire | 32.413 | 105 | 4.984 | 1.464 |
| **Total** | **528.716** | **1.882** | **265.395** | **25.880** |

Table 1: number of buildings that were mapped into the nodes of the graph reproducing the urban system of Glasgow and Clyde Valley as a complex network

## 8.1 INTRODUCTION: THE SOCIAL AND PLANNING CONTEXT LEADING TO THE 'GLASGOW EFFECT'

The planning period of the Glasgow metropolitan region (Figure 10) which determined the land use designations in Table 1 is divided in three distinctive phases. The Clyde Valley Plan (1944-75), characterized with centrally planned industrial zones and urban growth. After which followed the period of deregulation and retrenchment mandated by the Strathclyde plan (1975-96) and the period that planned for post-industrial restructuring and growth of the Glasgow and Clyde Valley metropolitan area (onwards 1996).

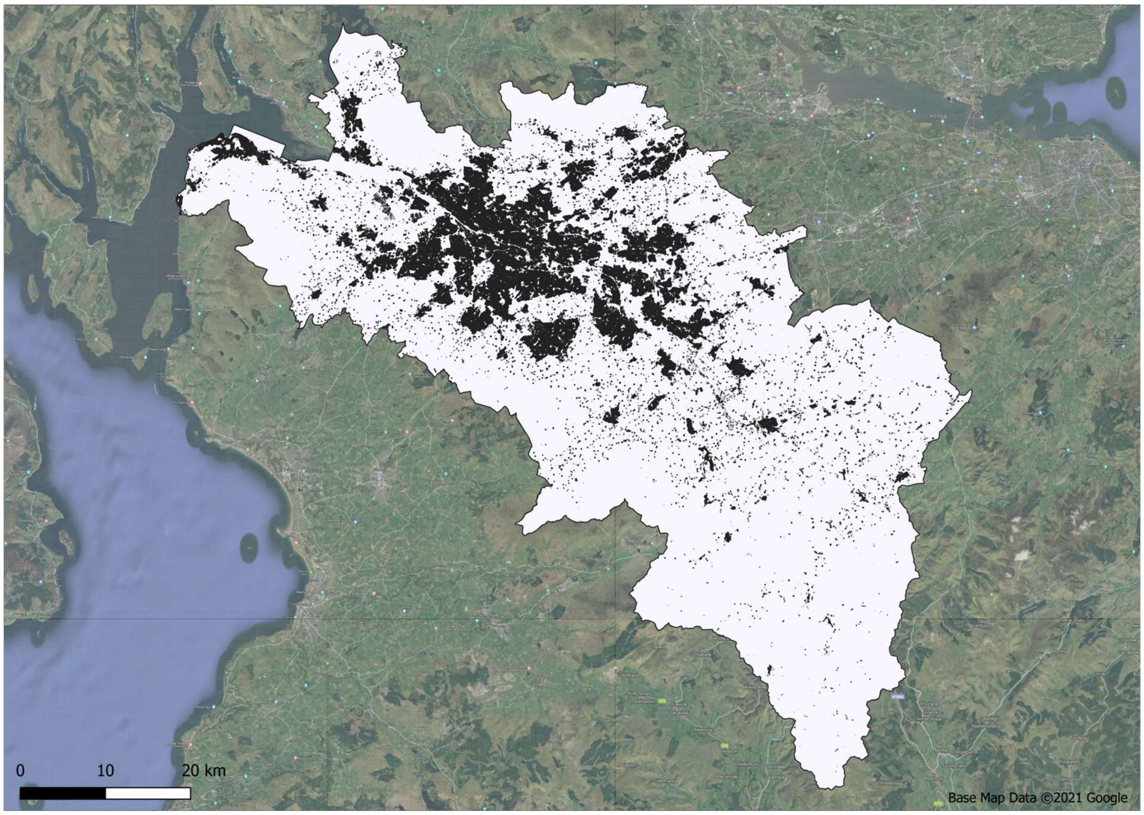



Figure 10: Map of the study area showing Glasgow and Clyde Valley region.
Base map data copyrighted OpenStreetMap contributors and available from https://www.openstreetmap.org.
Map produced using QGIS and EDINA Digimap Ordnance Survey datasets, May 2020.

The Clyde Valley Plan was the first effort of the Scottish Government to tackle the concentration of Victorian slum housing in the overcrowded industrial Glasgow. By the 70s, nearly 40 per cent of the workforce worked in manufacturing (Goodstadt, 2006, pp 318-320). The manual labour industry mainly included shipbuilding, textile production, steel making, dye works and the manufacturing of other worldwide-distributed commodities. This meant relocation of over half a million people away from the crowded and 'dirty' industrial city centre where most of the production factories were located (Macdonald, 2019), to the underdeveloped peri-urban fringes of the city, leading to growth of New Town-like communities. The outcome of this sprawl were the long daily commuting patterns of the relocated workforce to and back from central Glasgow. The period that superseded this trend intensively focused on urban renewal, leaving a legacy of underdeveloped inner city vacant land. After the inner-city renewal movement run out of support, the four newly created communities (Easrerhouse, Castlemilk, Pollok and Drumchapel) became a symbol for deprivation and urban poverty. In spite of these failures, the Plan held a sense of success, setting an example ranging from the clearance of the worst housing conditions across the United Kingdom, to the establishment of two new towns— East Kilbride and Cumbernauld. During this period, it is notable to mention that some of the former industrial land was converted to housing, which soil contamination levels are still questionable (Maantay and Maroko, 2015).

The following strategic plan (Goodstadt, op cit., pp320-322) developed by the Strathclyde Regional Council steered away from planning for sprawl and adopted a strategy of urban renewal and revival of social deprivation and employment opportunities. The main driver of this plan for regional development was the delivery of social and community infrastructure mainly through channelling private investment into brownfield housing development projects. The plan achieved between 65 to 70 per cent of net increase in new builds compared to the previous period, and this rate was sustained at that level for over twenty years.

The greater relevance of the plan in terms of sustainable communities were its pioneering policies that enabled environmental management of the urban fringe and the river valley. This was delivered through the indicative strategies for rural management and the creation of new regional parks. The positive societal gain this plan had in terms of sustainable economy was the impact it had on changing the 'out-of-town' retail experience promoted in the earlier planning period. This spiked regional economic rebound. The European Council of Town Planning with the '91 year's award, recognised the exceptional achievement the Strathclyde Structure Plan had. To this day, this plan remains as an example for stewardship and 'place-making' planning in post-war Britain.

After 1986, the Council was the sole remaining regional planning authority in Britain, which was decommissioned in 1996. The eight individual councils of metropolitan Glasgow, following the structural changes in 1986 decided to establish a statutory joint committee, which would have the power to prepare the structural plan, and a liaison technical team to monitor and review the one. The loss of power of the Regional Council to service new developments created uncertainty in new infrastructure investments. The decisions on water supply, new road extensions and affordable housing were left to the government agencies. The key transport, economic and health agencies ensured delivery of the 2000 structure plan by linking their adopted policies and strategies to programmes and delivery mechanisms. The 2000 plan continued to be effective on urban renewal, and additionally, between 1996 and 2004, it achieved a 10 per cent rate increase in new built houses. The ability of the councils' partnership enabled Glasgow's strategic planning and delivery of projects that spanned across councils, making it a strong functional region, comparable to the one in Copenhagen, Stockholm and Barcelona.

Glasgow City Council as a local planning authority, continues to promote planning strategies that support 'place making' in new and existing urban developments. In this mission, the Council relies heavily on its partnerships with the Shawlands and Strathbungo Community Council and the

Shawlands business community to deliver the *Action Plan;* a collaboration project between the City Council, the Community Council and the Business Community that won the 2019 Quality in Planning Scottish Award (Scottish Government, 2019). The Glasgow City Council - Shawlands Town Centre Action Plan was recognised for its comprehensive background knowledge on the historical changes of transport which consequently disconnected the Shawlands communities from the town centre. The project has demonstrated bottom up approach in community involvement for place making, but also agility to move with the market.

This progressive approach of metropolitan Glasgow in regards to urban and territorial planning which included economic revitalisation, through place making by bottom up approach, was the reason the study area was chosen to test the new modelling approach.

## 8.2 DATA SOURCES

To measure the variation in spatial access to services by every household within the Clyde Region, I collected spatially referenced datasets that included the exact location of every building or service point[92] and their use category. These datasets were assembled from three main data repositories. Namely, Ordnance Survey (OS), Open Street Map (OSM) and Improvement Services Scotland, found in several different collections. From Ordnance Survey information on land use within the built environment was collected from several spatially referenced collections:

- Geomni UKBuildings— which contained the use and locations of buildings which for the purposes of this study were categorised into two categories: residential [1] and commercial including other uses [2] (see Table 1);
- VectorMap— from this collections I retrieved the geospatial location of every important building such as a school or a hospital [3], electric vehicles' (incl. bicycles) charging points [4], electric vehicles' (incl. bicycles) share points [5], functional sites [6] and woodland areas [7];
- VectorData— I extracted the access points for parks [8];
- Points of Interest— from this dataset I extracted the locations of bus stops [9], playgrounds and parklets [10], repair and recycling services [11].

Three additional datasets were created using available information from Improvement Service Scotland, established in 2005 as the national improvement organisation for the Local Government in Scotland:

[92] i.e. places which do not technically have a building such as bus stops.

- Vacant Land [12]— a dataset derived from the Scottish Vacant and Derelict Land Survey, which collection contains the extent and state of derelict and vacant land;
- Public Waste Recycling Places [13]— locations of waste bins and waste treatment centres;
- Public Waste Recycling Sites [14]— larger areas for waste management, including transfer centres and active or historic landfill sites.

The remaining datasets were created by downloading from the crowdsourced platform OpenStreetMap (OSM) (OpenStreetMap contributors, 2021) various datasets using the *osmdata* package (Padgham et al., 2017) available for the open source statistical software R (Ihaka and Gentleman, 1996):

- Recycling locations [15]— to complement locations missing in the datasets obtained from Improvement Services Scotland;
- Bike share locations [16]— to complement locations missing in the datasets obtained from Ordinance Survey;
- Car share locations [17]— to complement locations missing in the datasets obtained from Ordinance Survey; and
- Electric car charging locations [18]— to complement locations missing in the datasets obtained from Ordinance Survey.

After the data collection from the three main data repositories and the data cleaning process, the green spaces access points and the parks and playgrounds datasets were merged in one, as well as the recycling locations datasets. Duplicated information that came from multiple sources was controlled to avoid double-counting and overlapping features were manually removed. For simplicity of the modelling process, all polygon datasets were converted to point datasets. For the smaller polygons such as the building footprints, geometric centroids were created and for the larger polygons such as the public waste recycling sites, multiple points with grid regular spacing of 200 meters were generated.

### *8.2.1 Defining the elements of a city-region*

As already discussed in Chapter 2, that sustainable form is closely related to the upper value of the physical capacity of the city, also known as the bounded effort (pp. 1-2 in this thesis) beyond which the city-region becomes decentralised in a conglomeration of several central places, the main focus is placed on exploring the arrangement of elements in the city that make this upper constraint sustainable.

Alexander et al. (1977) in their work define the main elements important for the design of well-balanced regions. More importantly, their work also defines the city-region as a function of the capacity of a group of people to govern themselves in a 'human way' (i.e. in the Greek style of democracy) which cannot exceed a certain number $\boldsymbol{N}$. At the very minimum, among a population size of $\boldsymbol{N}$-people, there exist at least $\boldsymbol{N^2}$ number of person-to-person connections (i.e. forming the logical network) that need to always be maintained, to keep the channels of communication open. For the lower and upper population limits of a city-region, they define population size between two and ten million. Thus, not only the city-region is bound in size by the functional dependency and underlying neoclassic economic forces (discussed in 4.3 and 4.4), furthermore it is bounded in size by the limits of democratic governance. Moreover, the placement of people across the region in social and ecological balance cannot be properly maintained, unless people are fairly spread out across the region, and they are given the opportunity to live in different types of settlements[93] while caring for the surrounding natural ecosystem services. These settlements as part of the city-region can only have few big and many small towns, as defined with the Zipf's law for cities (Gabaix, 1999), which have even spatial distribution (i.e. not concentrated only in few parts). The even distribution of the population of the city-region across space is important for the ecology of the region of concern (Alexander et al., op cit.).

Two aspects need to be considered to allow the city region to be evenly distributed in space: the ecology and the economy.

The *economic argument* is that people move where jobs are, and new jobs are created in areas with high economic activity. Thus, cities become more congests as new people come in search for jobs, economically devastating and declining remote areas in the city fringes. This affects the *ecological argument* stating that overconcentrated population in one place pressures the ecosystem of the region disproportionately. The region then begins to experience air pollution, water shortages, housing shortages, overcrowded public services, strangled transportation systems, and inhumane living densities. If the size of the population crosses the threshold point, the per capita cost of the city grows exponentially. For example, the cost of transportation routes and roads increase with the increasing number of commuters on them. Similarly the cost of housing increases in overly densified urban neighbourhoods, since the capital costs of high-rise accommodation are much higher than the one of traditional houses. Similar to this trend, the public services expenditure such as distribution of food, or the treatment of waste and sewage is significantly higher for people living in crowded cities, compared to people living in small towns. Thus, decentralization of population should increase self-sufficiency and minimize the pressure on ecosystem services.

[93] Cities, towns, villages and farms.

The buildings and spaces of interest in this study were selected based on the principles found in Alexander et al.'s *A Pattern Language* (1977). The pattern language describing how cities, buildings and construction practices are interlinked into one spatial system is very comprehensive, consisting of 253 well defined patterns. The work defines both, general and meticulous classes of patterns. I extracted only 49 general patterns, and based on the time and data availability constraints, left the small-scale detailed ones for future work.

The general patterns studied in this thesis were the ones explaining the physical and social relationships arising between people and places in the spatial context. Specifically, I looked into the 'Global Patterns Defining a Town or a Community' chapter, in which section were found the following groups of buildings and places/spaces/points that define an independent region.

[1] Connecting communities to one another— this classification includes supporting components that connect the city back into a cluster which is de facto divided by transport systems of regional roads (i.e. trunks) and motorways into local transport areas, to efficiently maintain the flow of goods and people across the city-region and the rest of the country:

- Web of public transport— all the various public transportation modes such as shared bicycles and e-cars, busses, trams, metro, trains, boats, ferries, taxies, moving footways etc. available to individuals can only work if all these components are connected together[94];
- Ring roads— highspeed roads (i.e. motorways) as essential infrastructure that connect transport modes into a system, should not divide and destroy communities, but shield them from noise, offering unobstructed access to the waterfront or the countryside;
- Network of learning— the city, all over the space, should allow a network of thousands interconnected opportunities that enable a society that emphasises learning over teaching, encouraging young people to learn;
- Web of shopping— new shops should be placed in locations that fill in gaps in underserved communities;
- E-bicycles and e-cars (replacing the original component of mini-busses)— serving the in-between areas, i.e. connecting the local transport areas which rely heavily on foot traffic and bicycles, with connected city sub-regions[95] which rely heavily on trains, planes and coaches;

[2] Encourage the formation of local centres both in neighbourhoods and communities— this is the land-use component which is typically determined by the master plans, explicitly restricting the use of buildings and public spaces[96] that are in proximity of one another:

[94] Which usually are not because different governing bodies run each mode of transport.
[95] The smaller administrative sub-regions that collectively comprise the city-region.
[96] Such as green public areas, e.g. urban parks, urban forests or public squares.

- Eccentric nucleus— the community centre should lie at the boundary of the community[97], nudging into the community just about enough, since after all, people do tend to need their community centre to be located towards the geometric centre of gravity;
- Density rings— people prefer to be in proximity to services and shops for excitement and entertainment, and away from centres for access to green spaces and rest.
- Activity nodes— in this group belong 'activity' facilities that attract people in a congregation and contribute to the life of the city. They are also called 'centres of action';
- Promenade— an area in the city where people go to walk up and down and meet their friends. Place that acts as a street theatre and offers people-watching;
- High-street (or Shopping street)— such streets should be quiet, convenient and comfortable, away from traffic and have good access to pedestrian pathways. They should be accessible by major transport hubs such as train and coach stations at a walking distance;
- Nightlife— most of the places that are open through the night need to be in close proximity of each other (i.e. a cluster of night life hot-spot) to offer bustling night life to the community;
- Interchange— places where various webs of transport services meet each other. These interchanges should surround workplaces and housing locations and their interior should maintain the continuity of pedestrian pathways, with mode transfer distances no longer than 100 to 180 meters;

[3] Encourage the formation of work communities between the cluster of residences, the activity centres and between neighbourhood boundaries:

- Work community— the workplace becomes as important as the home community, since large part of the day is spent with colleagues;
- Industrial ribbon— zoning laws that separate industrial parks away from housing areas create 'plastic' realities, thus these places should be distributed in the form of ribbons subdivided in long blocks that at every edge have a place where locally residing communities can meet and benefit from the industrial activity slipovers;
- University as a marketplace— the university becomes institution open towards the city and to people of all ages and walks of life, meaning that people can attend its programs in full-time, part-time or on course-work basis;
- Local townhall— each community is equally represented and has its own political control only if it has its own local physical townhall which is located near major interchanges, this plays a similar role to the Greek agora;
- Necklace of community projects— surrounding the townhall where community ideas and projects get public attention, critics, input, support and feedback;
- Market of many shops— instead of congregating all household goods and foods that people need under one roof (and one management) essentially making them blend and uninteresting, distribute clusters of small specialised and independent shops in the form of marketplaces to replace the modern supermarket;
- Health centres— a network of health centres offering treatment and preventative care for physical and mental well-being;

[97] A community is defined by the circularly buffered catchment area with the buffer's geometric centre found at the location filling-in the gaps of underserved communities.

- Housing in between— urban infill of new houses in the urban fabric of shops, public services, crafts workshops and university buildings.

[4] Informal growth of local road and path network: between house clusters and work communities:

- Looped local roads— roads between houses that are placed in a way that no other path taken would offer a shortcut, while in the same time they discourage high volumes of high speed traffic;
- T-junctions— traffic accidents are more frequent when two roads meet (creating sixteen collision points) compared to a T junction which has only three collision points;
- Green streets— local roads don't have to be paved with asphalt which reflects heat. Instead, they can be designed with two parallel rows of stones (or concrete blocks) to create a flat surface for the wheels and the reminder of its surface (in-between the rows of stones) should be green;
- Network of paths and bicycles, e-bicycles and e-cars (here I replaced the original component of cars)— pedestrian paths should intersect perpendicularly with minor roads and they should be arranged separately, frequently meeting at several focal points;
- Main gateways— boundaries that are important in the city should be visibly marked and in the places major paths are penetrating these boundaries, gateways should mark the points of transitions;
- Road crossing— the crossing points that create more than two seconds delay in crossing should have a refuge island between the lanes;
- Raised walk— walkways should be elevated in a way to keep cars below the level of pedestrian visibility and span only on one side of the road, as wide as possible;
- Bike paths and racks— a system of designated bike paths that are marked clearly in easily recognisable colour and run to every building, ending with a bike rack near the building's main entrances. Where possible, they should run in parallel with major pedestrian paths (grade separated) and local roads (on the same level);
- Children in the city— if the children cannot explore the city independently and are constantly observed by the adults, they cannot grow up. Therefore paths in the biking network should have one part that is extra safe from motorised traffic. The paths for exploration should be well lit, with many crossings and bridges along and bustling street life offering many eyes. Such paths can run along pedestrian streets, through workshops, warehouses, bakeries, assembly plants and other invisible life of the town, offering children to roam and explore the city freely on bikes.

[5] Public spaces where people can relax, spontaneously meet each other and renew themselves:

- Carnival— there should be space in the city where people can hold shows, competitions, street theatre, dancing events, among others;
- Quiet backs— the city needs to have quiet backs away from the busy and fast city, where people can escape busy streets and noisy traffic, to stroll down the river, where the mood is slow and reflective;

- Accessible green— people feel good when they have access to nature (i.e. the countryside) and experience open fields, agriculture and access to the wild birds, plants and animals;
- Small public squares— public squares are the most public 'rooms' of a city, however if they are too large, they feel and look deserted;
- Common land— social systems need common land to survive, and there must be enough common land that allows children games and small gatherings. Around 25 percent of public land would make the private land not over-dominating the neighbourhood;
- Connected play— common land, paths, bridges and gardens must be laid out in a way that at least a cluster of 65 households are connected continuously by these patches without crossing traffic;
- Public outdoor room— public space in every neighbourhood and workplace where people can hang out comfortably for a longer period of time;
- Grave sites— cemeteries should be small pieces of land scattered throughout the community as grave sites;
- Still water— each neighbourhood should have still water such as pond or pool for swimming;
- Local sports— places for team and individual sports should be scattered across every work community and neighbourhood;
- Adventure playgrounds— each neighbourhood should have a children playground with raw materials of all kinds, so children can recreate their own playground over and over;
- Animals— animals are as important part of nature as flowers, trees and grass. Direct contact with animals helps children's emotional development. Some locations in the neighbourhood should provide unpaved fenced space for animals such as horses, sheep and cows in the community to graze and roam freely;

[6] The local shop and gathering places:

- Individually owned shops— large shops controlled by corporations or absentee owners become too bland, plastic and abstract, therefore the city should support the formation of small independent shops and any large commercial property should include the option of small business owners to rent very small places in it;
- Street café— the street café provides unique setting where people can be lazy, relax and people-watching. Local cafés should be present in every neighbourhood as intimate places, with only few sittings, always one (or where possible more) row of tables stretching onto the footpath;
- Corner shop— each neighbourhood should have at least one corner shop, specifically placed on corners so they can be accessible from all directions and positioned among the houses, so owners can live above them;
- Beer hall—place where people can drink, sing and let go of their sorrows. A place in the city where few hundred people can gather around beer, wine, music and other activities;
- Traveller's inn— places that offer travelling people to be part of the community for the duration of their stay and meals are offered in communal areas;
- Bus stop— bus stops should be easily recognisable, forming small centres of public life. They should be located in a way that they work together with several other activities such as corner shop, tree places, special road crossings, cafés, public toilets;

- Food stands— offering simple and inexpensive way of getting food on people's way to work, school or shopping. They should be either portable or small huts, clustered where foot paths and bike paths meet;
- Sleeping in public— the public realm should be filled with ample benches, elevated or on the ground, relatively sheltered offering comfortable sittings, so people can be able to read the paper or just relax and doze off.

### *8.2.2 Defining the system of systems in the context of the study*

Depending on the social phenomenon being analysed within the context of urban morphology, urban networks and their accessibility, the networked infrastructure can be modelled in various ways to perform different types of network-based analysis. Therefore, before the network construction process, I needed to define what was being represented by a node and what was being represented by a link in the network's graph representation. Thus, it was important to decide the focus of the analysis. As already discussed in section 4.4, the study analysed the dichotomy between the growth of the physical and the logical topologies of the networked infrastructure. Therefore two graphs were constructed— a spatial physical graph, which represented the urban elements of LTSs and buildings under study, and a spatially-explicit relational (i.e. logical) graph, in part representing some of the countless spatial interactions grouped in three categories to define their functional dependency; mobility, social and ecosystem services— all defined in the sub-sections bellow.

Defining the nodes and links of the physical network was a straightforward process. In this instance, the nodes represented all the household locations in the city-region and all the services and amenities in the city that were available to the households, to the extend and the accuracy of the dataset. Intersections and buildings' street access points were as well represented as nodes of the graph. The links in the physical network were the existing streets that connected the household units to access various amenities. In the end, the entire city-region's infrastructure was mapped as a graph whose nodes were all of the buildings in the city, including the street intersections, and the links were the street segments.

Opposite to this, defining the elements of the logical network was more ambiguous process. The nodes 'participating' in the logical network (i.e. the ones for which the functional dependency and accessibility was sought after) remained the same as in the physical graph, except the street intersections and buildings' access points in this phase were excluded. The links in this graph were defined by the accessibility of these amenities in terms of the spatial proximity between the two. Spatial proximity within a network-based distance is different from the geospatial topological relationship as it will be discussed in the forthcoming 8.3.5 section. Therefore, the cut-off distance of 400 meters was essentially the restriction criteria used to define the neighbourhood[98] around the nodes in the logical graph. Thus, nodes were immediate neighbours with one another (although not spatial neighbours per se) if the network-based distance between the two was 400 meters or less.

[98] In this instance, node's neighbourhood is not the same as two buildings being spatial neighbouring. Please refer to 8.1 for the meaning of a neighbourhood around a node.

With the goal to simplify the mapping process of the large complex system represented by the two graphs and reduce computing time while only focusing on the research question, I removed any further qualitative information (such as demographic characteristics) attached to the nodes and links (e.g. direction or width). The direction of links in this study as a qualitative information was not important, since most of the roads on the network (unless they are highways) would have a two-directional sidewalk. Removing any attribute information of the nodes and links meant that the network's topology was the only retained information. As earlier explained, in the case when the only available information on the graph was its topology, then the nodes' immediate neighbours carry the graph's most fundamental characteristics. From the literature review in Chapter 3, specifically based on the conclusion in section 3.6, I studied the difference between the growth process of the physical and logical spatial networks. Following the theoretical evidence outlined in Chapters 5 and 6 and the clarified distinction between the existence of physical and logical network which simultaneously coexist within any given urban context, link communities were represented by various sets of links that were in some way topographically (i.e. functionally) related. Definition of topological relationships as a functional dependency between spatial objects follows in 8.3.5.

Unlike the specified neighbourhood set of the graph-based Cellular Automata introduced by O'Sullivan's thesis (2016), the neighbourhood of each node belonging to one of the two respective graphs, was part of the discovery process. I.e. uncovering the inherited community structure of the urban complex network. Consequently, to define a functional dependency that will form a link community, the study distinguished three different sub-constructs:

- Mobility infrastructure— there are numerous ways nodes (places of significant interest within the built environment) can be arranged with each other to represent functional dependency in the context of social accessibility. Thus, many combinations can be made in creating the origin and destination (OD) points that induce the flow of people. This study specifically looked at the combinations of places defined in the 49 patterns earlier, as a connected unity equally distributed across the city-region. For this purpose, the accessibility matrices included points between residences and transport points, including mode interchanges. Their proximity to residences and one another enables positive experience of traveling, especially when combining multiple modes (te Brömmelstroet et al., 2021). To create these accessibility matrices I used the household locations, public bus stop points[99], public mode shares points and supporting electric infrastructure locations. Specifically, the accessibility to mobility infrastructure was modelled by measuring people's network-based access from [a] their usual place of residence to each point in the destination set: [b] car sharing locations, [c] car charging spots, [d] bike sharing facilities and [e] bus stops.

- Ecosystem and social services— Due to time and data availability restriction, the accessibility matrices for the infrastructure of ecosystem services and social services were constructed by mapping [a] the residential locations as origins and [b] commercial buildings, [c] recycling points and [d] vacant land as destination points. There are many other locations part of the

[99] taken as a proxy for public transport due to data availability restrictions. Rail and underground transport was not included in this research due to limitation in data access.

complex human-nature system that comprise a functional dependency in ecological terms such as agricultural land, parks, forests, floodplains and water bodies that play eminent role in supporting ecosystems' diversity and remediation, and therefore form a link community in their own right. These locations are not included in this model and should be considered in a future study to extend the model presented in this thesis.

– Social infrastructure— the OD accessibility (i.e. adjacency) matrices between places of social exchange were constructed from [a] the residential buildings as origin locations and [b] all other important buildings (such as educational, day-care or healthcare facilities), [c] urban parks and woodlands, [d] parklets and playgrounds and [e] commercial buildings as destination points. Access to these places in the urban environment is measured separately from the access to mobility services since in traditional ABM, modelling discretionary trips is the most challenging task. By separating these from the mandatory tours[101] this model is able to help predict mode choice based on the activity specific destination (Giles-Corti and Donovan, 2002).

This form of 'loose' modelling approach excluding typically used information from Travel Survey and socio-demographics is justified since the model constructed in this research did not aim to estimate the number of people that actually went to specific places in metropolitan Glasgow. Instead, its purpose was to evaluate whether social, mobility-enabling and ecosystem-remediating infrastructure was equally accessible to local residents within a walking distance from their usual home of residence.

## 8.3 DATASETS MANIPULATION

The OD datasets from which the accessibility matrices were constructed for the purposes of the study, were created by measuring the network-based distance from each residence to all locations of interest below 400 meters. This included a pathfinding task performed from every home location on the network to every destination as defined in the section above. Given the computing intensity needed to measure access from each home location to all relevant amenities (e.g. food stores, restaurants, cinemas, GP, day-care, schools etc.), a reduction criteria was used to significantly narrow the search space.

First the OD matrices were generated by beeline first-nearest-neighbour (1-NN) search with a cut-off distance of 400 meters. Once the beeline distance filtered out all non-pairs, another network-based pathfinding search was performed with a 400 meter cut-off distance which generated the final OD pairs. The building inventory data was used as the origin input and the remining datasets were used as destination locations for the diverse types of facilities. This process created the measure of access to various uses within the network-based cut-off distance from each household. Thus, the access of each household to the first point of interest regardless of its use (which varied across commercial, recreational or technical infrastructure) was measured. The goal of the research was not to produce

[101] Mandatory tour is a frequency sub-model in ABM that predicts the number of work and school trips. In the mandatory tour the agent is committed to at least one trip outside the home location that has a mandatory activity attached to it, such as work, studies or schooling. This tour can include non-mandatory intermediate stops.

accessibility metric for each individual home location, but to test the hypothesis that using the network science algorithm for discovering link communities, provides a new perspective and understanding on the sustainability of different areas in the city, which opens new avenues for future research. This approach although not comprehensive was justified by three reasons:

[1] The computing time for all-to-all pathfinding would be extremely large and beyond the time constraints of the project;

[2] The research did not have the objective to produce comprehensive accessibility metric of each individual home location;

[3] To fulfil the second research objective which aimed to discover the link communities of Clyde Valley and with this give fresh perspective and open new future research avenues.

## 8.4 DATASETS CLEANING

Complementary to accessing and using existing proprietary datasets, the study used open source datasets as described earlier. Since these datasets are volunteer-generated, they are not always 100% accurate, and therefore are less reliable. The open source datasets had to undergo additional manual cleaning and pre-processing. The following section describes the process of cleaning these datasets and creating the physical graph. thereafter, the following section presents the methodology applied to construct the logical graph of spatial relationships. In the last section of this sub-chapter I presented methodological evidence from the reviewed literature on the disparity between the physical and logical topologies which in this thesis is captured by their separate representation as two graphs.

### *8.4.1 Creating the road network dataset (the physical network)*

The spatial data of the road network was retrieved in 2020 and 2021 from OpenStreetMap (OSM) (OpenStreetMap contributors, 2021) using the *osmdata* package (Padgham et al., 2017) available for the open source statistical software R (Ihaka and Gentleman, 1996). OpenStreetMap was chosen over Ordnance Survey's MasterMap data collection due to its more detailed representation of the road network. While the MasterMap collection had better qualitative information including level of service, road width and other road assets, the dataset has coarse spatial resolution and omits many lower level residential and service roads. Since the research had a goal to measure the network-based distance from home locations to various facilities, these types of roads were very important and the OSM dataset was chosen as more appropriate.

The network in all eight shires was cleaned to ensure all network segments were connected and each housing, commercial or other building types had access to the road, which was represented by adding new links to the base network (OpenStreetMap contributors, 2021). Adding these new links connected all building units with a *linestring* to the network file. Houses and other building units were 'integrated' in the network using the Quantum Geographic Information System (QGIS) plugin tool *Networks* (Palmier, 2021). The network was cleaned both manually in QGIS (QGIS.org, 2021) and automatically by executing commands from the Geographic Resource Analysis Support System (GRASS) suite (GRASS Development Team, 2020) within the QGIS Graphic User Interface (QGIS GUI). The pathfinding task was performed using the *stplanr* package (Lovelace and Ellison, 2018).

### *8.4.2 Constructing the spatial logical and physical graph models*

A data model is defined as an abstract object in a spatial database registry consisted of: [i] a topological entity, [ii] the entity's geospatial location, [iii] its non-spatial properties and [iv] the relationship it has with other topological entities in a defined proximity. The goal of the spatial data model is to provide formal means for presenting and using georeferenced information (Bolstad, 2012). The spatial data model can contain both spatial and aspatial data.

Topological entities or objects, in the domain of quantitative geography are unique features on the Earth's surface (Kuby et al., 2005). The topological entities of the graph mapping the logical topology of the eight shires within metropolitan Glasgow were both, the residential buildings and all remaining non-residential buildings in the network which were represented as the graph's nodes. The links between these nodes were defined as the network-measured distance of 400 meters connecting an origin point (i.e. a household) to a destination point (i.e. any of the non-residential buildings), and this rule was applied to all residential buildings. The logical graph was constructed by linking each node pair with a route from the origin to the destination. The theoretical rationale by which the OD accessibility matrices were constructed was defined in 8.2.2.

The nodes of the physical graph extended to include the building locations and intersections of the street network. Additionally, 'pseudo nodes' were added between each building and its access to the street segment (see Figure 11 b). The links were represented by the street segments spanning from intersection to intersection[104] and their length was registered as the links' attribute. The resulting graph was undirected and weighted, excluding loops and multi links. This graph was then converted to its dual (i.e. line-graph) representation (see Figure 11 c).

[104] Intersection in the scope of the study refers to a point where a street intersects to another street or a house footpath intersects to a street (see Figure 11.b).

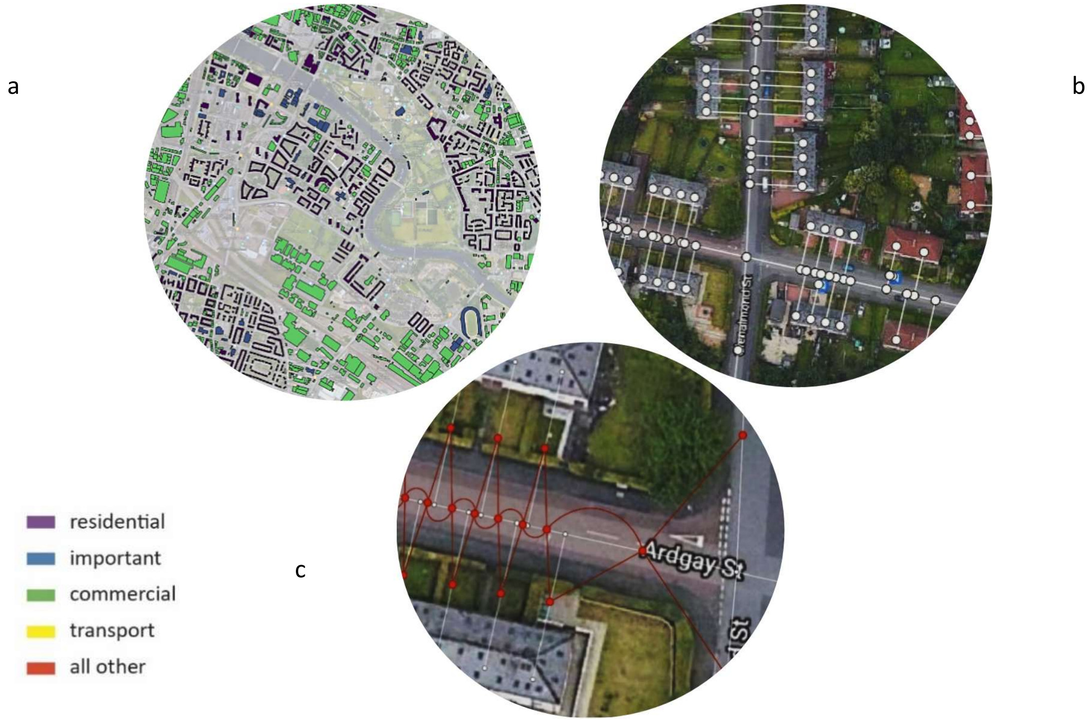



Figure 11: Navigating from the primal to the dual graph: mapping Glasgow and Clyde Valley's buildings and streets as network nodes (Author's own maps)
Base map data copyrighted OpenStreetMap contributors and available from https://www.openstreetmap.org.
Map produced using QGIS and EDINA Digimap Ordnance Survey datasets, May 2020.

Figures 11 and 12 present the overarching method by which the buildings' datasets were reconstructed into the primal and dual[105] graph models of the physical network for the eight council authorities in metropolitan Glasgow.

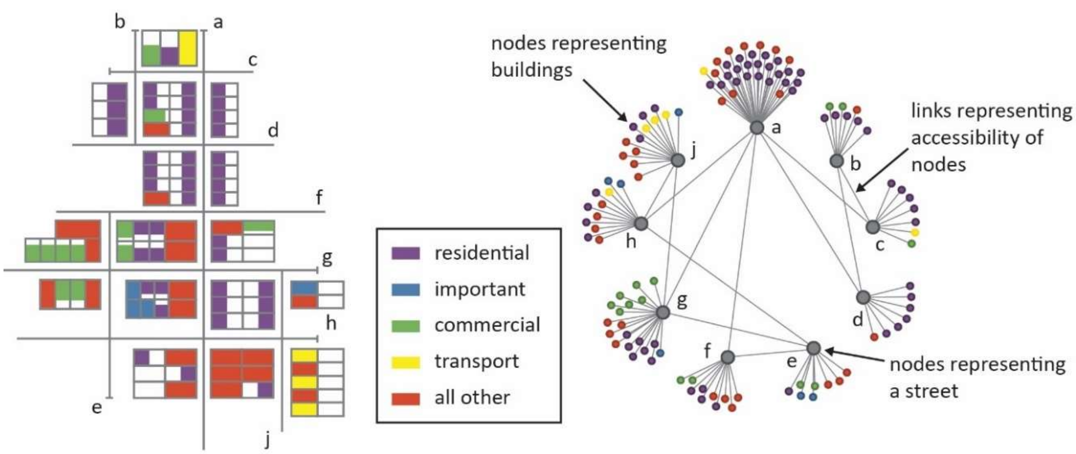



Figure 12: An illustrative example for the conversion of the primal to the dual graph of the physical topology of the network (Author's own diagrams)

105 The primal and dual problems are defined in 8.4.1 and 8.4.2 following an extensive discussion in 8.4.3.

### 8.4.3 Topological relationship between 'cellular' spatial objects

Chapters 4, 5 and 7 established a clear distinction between the notion of *topology* in the study of network science and the spatial statistics sub-fields of quantitative geography. From the broad review of definitions (Egenhofer, 1993; Barabási, Albert and Jeong, 2000; Schneider and Behr, 2006; Naimzada, Stefani and Torriero, 2009; König and Battiston, 2009; Bolstad, 2012; Barabási, 2016; Shen, Chen and Liu, 2018), a general conclusion for this study was made. Topology in network science displays the [non]existence of a link between two selected nodes in the network and the evolution and dynamics of connected nodes[106] is encoded by the existence of this link. Whereas, topology in spatial analysis represents the geographical position of entities and their spatial relationship with other neighbouring entities. The first network shows the "dependence" of links based on some form of functional relationship (Urban and Kaitt, 2001), displayed in two or three dimensional space (property of the planar graph). The latter one conveys information of the type of topological relationship between neighbouring spatial entities. In the first one, although nodes connect with each other they do not have to be in physical proximity. In the latter one, if they are not adjacent to other nodes or to other spatial entities[107] (lines or regions) they do not share any topological relationship, even though they may be in close proximity. In the context of this study proximity was relevant for the graph of the physical topology, while its was irrelevant for the logical topology and instead, the topology was encoded as the geodesic distance between the OD adjacency pairs.

Contrary to this, Campbell and Shin (2012) define *geospatial topology* as the set of rules that models the relationship between spatial entities such as points and lines/curves, determining the way in which the objects share geometry, observed in Euclidean space. An example of geospatial topology is the common border of two adjacent cadastral parcels (plots), demarcated by two polygons sharing a common line. In this case, although the parcels share a border it is very clear which side belongs to which parcel and their geospatial topology is defining them as bordering plots (or also *adjacent*).

In conclusion, it should be noted that the topological relationship of spatial objects is different from the topological relationship studied in graph theory, network science or statistical mechanics. In the latter, topology refers to the [non]existence of relations among nodes in the network. Thus, network's measurements can only be characterised and analysed, after obtaining the informative features of the network's topology. Some of these are; *vital nodes, clustering coefficient, preferential centrality, betweenness centrality, preferential attachment* or *scale* (Costa, 2008; Barabási, 2016).

[106] i.e. how quickly the nodes create, maintain and loose existing links.
[107] E.g. intersections are connected by street segments.

To summarise this relevance for the thesis, as illustrated in Figure 12, topological relationship between the buildings and streets was represented by the direct access of buildings to the street segment. Different than this, the topological relationship between the nodes in the logical graph of the same network existed between origin and destination points if the geodesic network-based measured distance from the origins to destinations was 400 meters or less. This connectivity requirement determined the opportunity available to people within a walking distance from their places of residence, to any destination of a specific interest for this study, as defined in the following section.

## 8.5 VARIABLES

The following section defines the main variables studied in the research. Namely, the accessibility matrices used as input in the community-finding task were constructed based on the variables defined in 8.4.1 and operationalised in 8.4.2. The following variables were defined as input to generate the accessibility matrices.

### *8.5.1 Accessibility*

Various definitions of accessibility exist in the literature of urban planning, transport planning and social geography, which make this concept very fluid and its operationalisation can take many forms. It usually measures very context specific value of a geodesic distance between one land use to another. Most commonly, the broad understanding of accessibility is ‘the proximity of individuals to locations’. Essentially, accessibility intrinsically measures the ‘potential of opportunities for interaction’ (Hansen, 1959). However, other definitions of accessibility include the ease of reaching (by the means of public transport) any activity that is constraint by land use, or the freedom that any individual residing in a certain location has to decide whether to take part in different activities (Geurs and van Wee, 2004).

Recently, accessibility became more complex index which includes four main elements; land use component, transportation disutility (usually valued by cost and/or distance), a temporal (time-of-the-day availability of activities) and an individual utility-based component. Considering these elements, the accessibility index includes four measures; [1] infrastructure-based, [2] location-based, [3] person-based and [4] utility-based. Each of which is further composed of several [sub]measures. Accessibility in the scope of this thesis was limited to only include the social indicator as a measurement.

Social accessibility measures the level of access of individuals to food stores, jobs, shops, healthcare and education services, as well as opportunities to meet family and friends. The limited definition of accessibility in the context of this study was justified by [1] the very large area under study and [2] the limited availability of disaggregated socio-economic data to the level of the households. The need to not overlook last reason was based in the first objective of this study. To remind the reader on the first objective; discovering link communities using the novel method will revealing the structural disparities of the relational (i.e. logical) and physical graphs in the context of social accessibility. The link communities discovery process doesn't rely on its ability to measure accessibility of places as a holistic index. Instead, clustering link communities in the spatial context can be applied to test accessibility's sub-measurements and once the success of the algorithm is confirmed, the method can be upscaled to the more comprehensive accessibility index.

Therefore, the study's main focus was to spatially disaggregate and present the social accessibility of all households within metropolitan Glasgow, by measuring their equal access to social infrastructure. Thus the accessibility metric accounted for households' access to [1] food stores and other commercial services, [2] health, social and education services, [3] derelict land, [4] public transport and shared mobility hubs, [5] outdoor recreational and green spaces and [6] recycling facilities.

Stemming from the conclusion presented in the next section 8.5.3, I considered adding the geodesic distance to the step-length one[108] as extremely important. Following this criteria, a link community was defined as the functional union of locations in the city that are connected at the furthest of 400 meters network-based distance with one another. This criteria was applied to all 528.716 household units[109] and locations that generate urban activities and infrastructure links that distribute spatial interactions to each household locations. The study considered the 400 meters network-based distance as a cut-off value, based on the results of the mean walking distance registered in the National Travel Survey (DfT, 2014). The NTS reported that people in 2014 on average walked 0.29 miles (0.464 km) to a bus stop. This is also the lowest average distance when walking was one of the modes of multi-stage trips. Since the NTS survey assumed that the mean walking length in the statistics is overestimated, for the purpose of this study the walked distance was reduced to 400 meters. According to the NTS specifications, the overestimation appears due to a sampling error that omits the extremely short walking trips, which are likely to occur among the general population.

The average walking distance to a bus stop is a good proxy to measure accessibility of places and services to all inhabitants in the city equally, while encouraging social and technical infrastructure synergies, because it represents the average distance that majority of the population is willing and able to walk. Beyond this distance, some population groups may be excluded.

[108] Based on the conclusions from the empirical study referenced in this section.
[109] Some of the households are multi-residential buildings. Due to incompleteness of the buildings dataset in relation to the detailed information of all households on the property, a simplification in this thesis was made regarding each building location as one household.

### *8.5.2 Social Equity*

As already discussed in the Introduction of this paper, the EU Commission by adopting territorial cohesion as the leading strategy for territorial development, pursues equity in accessibility to both social and technical infrastructure. Equity of infrastructure is seen as the foundations for spatial equality, which are one of the key support systems for cohesion across EU regions. Equity in accessibility means living environment where individuals from disadvantaged socioeconomic groups are provided with fair access to *public life* as the more advantaged households.

Talen (1999) defines social equity as the equitable provision of accessible social goods and services to all the members of the community. Talen (op cit.) explains that the fair delivery of social infrastructure has the ability to encourage social interaction and exchange across members of neighbourhoods and communities. I adopted this definition of social equity for the purpose of the study.

### *8.5.3 Link communities*

Link communities in the study context are defined as the functional units of the urban system that support infrastructure and social synergy and enables people to go about their daily lives. These synergies are said to be at optimal capacity, by optimising and equally distributing the access to resources in the city-region, while distributing the negative feedback loops (i.e. waste distribution) and enabling remediation of ecosystem services at a sustainable rate[112].

Link communities have a great potential to be applied in spatial analyses, offering new scale to study the built environment. In urban science they are proven to overcome some of the general shortcomings like resolution limitations (Kevin and Krizek, 2003; Strominger et al., 2016), associated with the more traditional ad-hoc area-based administrative units of analyses. For example, Law (2017) empirically showed that the hierarchical clustering of neighbourhoods' socio-economic similarity is linked to the hierarchical clustering of the 'Street-Based' Local Areas (SLA) and not their administrative area boundaries. SLA as an outcome from the community detection process, was strongly associated with the housing prices and spatially constrained by the spatial topology of the local street network. Specifically, property prices in close proximity tended to cluster in tightly knitted communities, whose hierarchy resembled the hierarchy observed in the arrangement of the local street network (i.e. the physical network). This phenomenon complied with Tobler's first law in geography and with other usually observed morphological and infrastructure hierarchies, extensively reviewed in Chapter 3.

[112] I.e. a rate at which the natural environment can repair its ecosystems disrupted by human actions.

Law's study (op cit.) concluded that SLAs in their most basic definition are communities (i.e. the property bundle[113]) within the street network, that emerged based on the functional dependency (i.e. functional link) between the properties' locations and the properties' market prices. This 'functional dependency' is the term that separates the construct of a link community within the system of the city-region, from the rest of the spatial arrangements. If these same links are also observed in other functional dependencies between the SLAs, for example positive correlation between property market prices and the amount of school or hospital subsidies received, then we say that we have observed an overlapping link communities (i.e. link synergies) within the complex urban network.

Synergies across technical infrastructure (Derrible, 2017) and social infrastructure are based on the principle of flexible, multi-use of urban amenities and places, supporting both human and nature's needs. These design principles (Hanzl et al., 2021) are recognised as key practices in new developments, or for redevelopment of existing neighbourhoods. These synergies are also the overlapping communities in the context of urban studies.

Figure 13 below shows a footway that is shaded by a tree canopy cover along a bicycle path and a carriageway with reduced width, serving bi-directional motorised traffic. The design solution to collocate multi-purpose carriage ways with different surface qualities (pavement with greater friction for the motorised traffic compared to the smooth surface of the cycleway) is essential in delivering the synergies between different infrastructure sub-networks. Furthermore, another technical synergy is achieved by designing the footway pavement to slope outwardly, guiding rain water away from the footway and towards the trees. In this case the footway is not only serving pedestrians but also as water collection system to irrigate plants.

[113] Similar houses in the same neighbourhood influence each other's values creating a sub-market in their own right or a bundle of locational, physical, and neighbourhood social quality attributes as characteristics of the competing units (Watkins, 2001).

Figure 13: Synergy of both nodes and links in the networked infrastructure in Eindhoven, The Netherlands.

Other examples of synergies are the links covered in green patches of permeable surfaces underneath tram tracks (as shown in Figure 14) which span alongside brick paved motorised transport roads and 'slow' lanes, that help drain excessive amounts of atmospheric water while accommodating mass transit, protected low speed motorised traffic and other more active forms of micro-mobilities. In Figure 15 we can observe rain-water management canal collocated with a footway which helps drain rain water. In the same parallel setting we also observe strips of green unmanicured patches that help create habitat for small insects such as butterflies and bees that are well known key insect groups in maintaining a healthy ecosystems.

An example of synergies of nodes (i.e. places) are the multipurpose[114] buildings or public spaces such as parks, playgrounds or underground parking garages that at the occasion of for e.g. heavy rain, are used as flood protection tanks[115]. The Museumpark aside of being a large scale technical infrastructure, at its surface accommodates the historic central public park of Rotterdam. Museumpark is located in the nexus of Kunsthal, Netherlands Architecture Institute (NAI), Museum Boijmans van Beuningen and Natuurhistorisch Museum Rotterdam. This multipurpose structure daily attracts numerous visitors, both locals and tourists.

[114] Here multipurpose refers to both land use classification for mixed use of land/buildings, but also spaces reused for different activities in different hours of the day.
[115] Reference of this type of location is the Museumpark in Rotterdam.

Figure 14: Synergy of links in the networked infrastructure in downtown Rotterdam, The Netherlands.

Another more generalised examples of nodal synergies are industrial or residential buildings which next to their primary use serve as urban farms; or abandoned sites and former parking lots that are reprogrammed into public social infrastructure facilities such as urban pocket parks, simultaneously serving as renewable energy sources, shared mobility hubs and electric charging points.

Figure 15: Synergy of links in the networked infrastructure in one of the most central streets in Eindhoven, The Netherlands.

### 8.5.4 Community centrality

Community centrality is a novel metric applied in complex networks analysis which holds great promise in detecting and explaining the structural role of community overlaps, responsible for network's resilience, robustness and efficiency. This measure is very important once we have discovered the nodes with low number of internal links[116], but large number of links towards the communities found in the rest of the network. As explained earlier, these nodes are the ones whose external links are part of the edge cut-set. The edge cut-sets of overlapping communities (Figure 4) are comprised of links that connect the node with two or more communities, and high number of shortest paths go through these links. The nodes belonging to these cut-sets are also said to have high community-based node centrality if their external links are the ones that have high edge betweenness. They are of special interest since when removed, the communities in the network become isolated from the rest, while the network becomes fragmented. Measuring nodes' community centrality is relevant for this research, since discovering nodes which do not act as network hubs but connect to links with high edge betweenness as already discussed in Chapter 6, are the ones that increase the network's resilience and robustness.

## 8.6 THE HIERARCHICAL SINGLE LINK CLUSTERING ALGORITHM

Considering the previous main methodological underpinnings, I assumed that the spatial system in its entirety is mathematically represented by a graph $G$ that consists an edge set $E_{(g)}$ and a node set $V_{(g)}$. This main complex urban system was further represented by many different subgraphs which were either physically or functionally (or both) connected with one another. I denoted these subgraphs by $G_{(i)}$ and their node and edge sets were denoted by $E_{(g,i)}$ and $V_{(g,i)}$ respectively. Both $g$ and $i$ were integers sets defined as $i = \{1,2, \dots, n\}$ and $g = \{1,2, \dots, m\}$.

At the first instance of the single hierarchical link clustering process, each link is initially assigned to its own community. In the next iteration, the algorithm searches for similarity between links that share at least one node. Similarity is determined based on the links' other two non-shared immediate nodes. With this approach, the search space of the algorithm is significantly reduces. Once the link similarity between two neighbouring links was determined only for the logical graph using Equation 1, the links were merged in the same link community at every level of the hierarchy of the link dendrogram. Since one link can appear only in one link community, redundant links in the merge process were removed respectively. The process was repeated until all the links dissolved into a single cluster. The algorithm

[116] Internal links refer to the links the node of one community has with nodes belonging to the same community (i.e. sub-graph).

writes several outputs among which is the dendrogram heights at which communities are discovered, the community link and node IDs, and the levels at which links are merged in link communities (see Tables 2-4 for example outputs). Another output are the values for the local partition density of each link community and the global modularity.

Once every partition density value for each community of the network was discovered, the correlation between the traditional node centrality and the centrality of the node considered as a part of the community was examined. This new measure of community-based node centrality was first introduced by Newman (2006) and adapted by Kalinka and Tomancak (2011). This comparison allowed me to empirically confirm the difference between the network's physical and logical topologies, which will also be relevant for any future research building on this findings, that will explore the graph's behaviour when growth and aging are simulated in the graphs which in reality are evolving, but in this study due to time constraints were examined in their static representation.

$$\boldsymbol{S}(\boldsymbol{e}_{\boldsymbol{ik},}\, \boldsymbol{e}_{\boldsymbol{jk}}) = \frac{|\boldsymbol{n}_{+(\boldsymbol{i})} \cap \boldsymbol{n}_{+(\boldsymbol{j})}|}{|\boldsymbol{n}_{+(\boldsymbol{i})} \cup \boldsymbol{n}_{+(\boldsymbol{j})}|}$$

**Equation 1 Introduced in Ahn et al. (2010b, p 5)**

The result of the link community discovery process was a link dendrogram that kept a record on all the meta communities at various levels of the dendrogram's height determined by the similarity value $S$ (Eq. 1). The similarity value is considered as the strength of the merged community, and is available for all of them. Due to the large number of communities found in the eight spatial logical graphs, the link dendrograms were not visually presented. Instead, the descriptive statistics of the community densities for all eight council areas are presented in Table 5 in Chapter 9.

The link clustering algorithm for this study was used through its implementation within the *linkcomm* package (Kalinka and Tomancak, 2011) available for the open-source language for statistical programming R (Ihaka and Gentleman, 1996). The package is created for general application of discovering link communities in the nested and overlapping structure of complex networks.

In addition to generating the dendrogram plot (which for the reason explained earlier is not presented in this thesis), the outcome of the algorithm was the value for the partition density of the entire network where the link clusters membership was optimised. The partition density was calculated using Equation 2.

Further benefit of the algorithm is the transparent process used to discover the communities. Unlike the majority graph theory models that operate in a 'black-box' manner as explained in 7.2, the outcome of the hierarchical link clustering algorithm is fully transparent and explainable.

The transparency of the process is in the product of several records on the link communities' merges. Namely, while merging (or clustering) link community members, all the links that belong in the same community are registered by the algorithm at the specific height of the dendrogram where the communities were merged in the new community. The order by which the nodes were merged in communities was also recorded. If a node belonged to multiple communities at different hierarchical levels, the same node was recorded in all of them. The algorithm keeps track of all node and link indices at every height of the dendrogram. The results of this process are further explained and partially presented in 8.7.3.

Two main benefits are notable. First, unlike the node dendrogram that allows for one node to appear only at one level of the dendrogram, in the link dendrogram, nodes that belonged to several communities were allowed to appear at several levels. Thus, the nodes that belong to several communities were labelled with all of the link community membership IDs.

Second, the optimal level of the hight at which the dendrogram was cut (i.e. the partition density measure) was maximised. Partition density of a link community is the weighted sum of all the partition densities of the joined communities that share the node (Lee et al., 2017). In its essence, the measure evaluates the quality of the link communities discovered by the algorithm.

Partition density is the opposite of the modularity measure[117]. It does not have a resolution limitation and the algorithm returns a global value for the entire network at which height of the dendrogram's cut, the number of link communities is optimised. Modularity contains an intrinsic scale (Fortunato and Barthélemy, 2007, p 38) which dependents on the number of links of the entire network and therefore limits the number and size of modules to be discovered (the value is always determined a-priori by the modeler). Opposite to this, the link clustering algorithm produced all local partition density values for the various heights of the link dendrogram, at which link communities were discovered, satisfying the granularity requirement.

Thus, each term in Equation 2, part of the hierarchical link clustering algorithm, is a local measure of every specific community part of the network.

[117] A concept introduced earlier in Chapter 6.

$$D = \frac{2}{M} \sum_c m_c \frac{m_c - (n_c - 1)}{(n_c - 2) - (n_c - 1)}$$

Equation 2 Introduced in Ahn et al. (2010a, p 764)

Here $M$ is the total number of links in the network, $m_c$ is the number of links and $n_c$ the number of nodes in the specific cluster being analysed. The clusters are part of the network's communities set $C$, and for each of them the partition density $D$ is being calculated individually. The limitation of this measure is that it doesn't use any statistical null model, so it is improper function for global optimisation (Lee et al., 2017), but this was not relevant for this study, since the goal was not optimization of the networked infrastructure, but detection of the s-t sets.

## 8.7 THEORETICAL AND MATHEMATICAL FRAMEWORK ON LINK COMMUNITIES

This section using a mathematical framework defines more specifically the notion of overlapping communities in the context of networked infrastructure, and the measures set up to define them.

As discussed earlier, within the scope of network science, there is no one unified definition to what constitutes a community. However, there is a general acceptance that communities are sets of related nodes and usually have more links that span inwards the community to connect internal nodes, than to connect the community's nodes to other nodes external to that community. I.e. nodes belonging to the same community are densely linked with one another. Opposite to this, real-life communities often are highly overlapping and have more external than internal links (Ahn et al., 2010). Therefore conventional community detection methods based on discovering nodes' community membership, fail to capture these complexities.

In the last decade, particular attention has been given to novel methods and algorithms that offer community detection allowing communities to overlap (which is the main result when communities share nodes and links). This algorithmic approach is opposite to the ones based on graph partitioning principles or agglomerative hierarchical clustering of groups of nodes based on their similarity (e.g. the Ravasz algorithm (Ravasz, E. et al., 2002)) or based on their clique-clique overlap matrix (Palla et al., 2005). Such algorithms restrict nodes to belong to only one community, whereas many real-life communities overlap and one node belongs to multiple groups (Ahn et al., 2010).

The approach introduced by Ahn, Bagrow and Lehmann (op cit.) detects nested communities based on link similarity. Considering how similar two links are by comparing the nodes these links connect to, the algorithm hierarchically clusters the links with high similarity in a community. In this process, the algorithm allows for nodes to belong at the same time to several communities. Precisely these overlaps of links between communities of large scale spatially-explicit networks are significant, and their spatial distribution in the networks disclose universal features such as the community-based node centrality, including their spatial positions in the network.

### *8.7.1 Measures of Social Accessibility*

In this study, social accessibility was quantified through the land-use component of accessibility and empirically measured as the network-based distance of each household access to:

- playgrounds or allotments as destinations;
- recycling points;
- bus stops;
- repair shops;
- other functions (emergency services; central and local government; health facility; institutional and communal accommodation; general commercial mixed use; industry manufacturing; office space; offices with retail on ground floor; recreation and leisure; petrol stations; retail; retail with office/residential above; transport related infra such as parking facilities; unclassified);
- vacant land;
- car charging stations;
- functional sites (including higher education; primary or secondary school; leisure or sports centre; medical care; place of worship; police station; coach station);
- open access point to green space;
- forest;
- bike share locations;
- e-car charging stations;
- e-car sharing locations;
- recycling or landfill sites.

### *8.7.2 Measures of Equity*

In this research, I measured the equal opportunity available to an individual to access any of the specific locations of social infrastructure within 400 meters measured as a network-based distance, regardless of whether a person uses this opportunity to get there or not. Thus, equity in social accessibility in the scope of the research was measured by constructing OD accessibility matrices from every household in the study area to the various destinations defined earlier. Household whose network-based distance was beyond 400 meters were not connected to the logical graph, and these parts of the network were considered to have unequitable access (i.e. no social access).

### *8.7.3 Measures of Link Communities*

The link community clustering algorithm follows a mathematical logic in three steps. First, two links are considered for the analysis if they share a key stone node $\boldsymbol{k}$. Then the similarity between the pair of links is calculated using the Jaccard coefficient. The Jaccard coefficient measures the similarity between two communities as a proportion of the cardinality[118] of their intersection to the cardinality of the communities' union.

To determine the links' similarity value (using Eq. 1) the algorithm measures the similarity of the immediate (i.e. first-order) neighbourhoods of the two remaining nodes found at the other end of the links as presented in Figure 16 below.

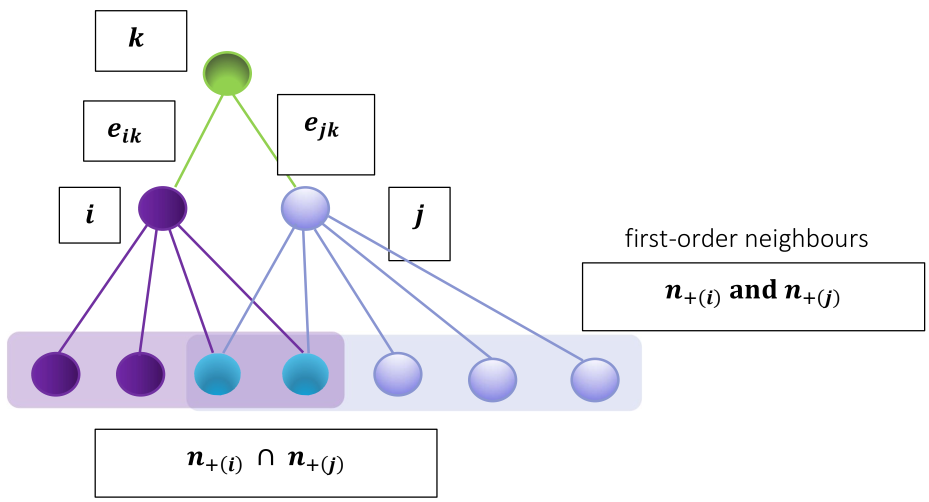



Figure 16: Illustrative diagram of the process by which the algorithm determines link similarity (Author's own graphics).

In Equation 1, $\boldsymbol{S}$ represents the similarity measure of the two links $\boldsymbol{e_{ik}}$ and $\boldsymbol{e_{jk}}$ that are connecting nodes $\boldsymbol{i}$ and $\boldsymbol{j}$ respectively, with their commonly shared node $\boldsymbol{k}$. The $\boldsymbol{n_{+(i)}}$ notation represents the number of links in the set connecting node $\boldsymbol{i}$ to all the nodes that belong to its first-order neighbourhood (i.e. its immediate first tier neighbouring nodes). In the same manner, the notation $\boldsymbol{n_{+(j)}}$ represents the first-order neighbourhood link set of node $\boldsymbol{j}$. After the pair-wise similarity is scored using Equation 1, in the second step the links are hierarchically clustered in a dendrogram. After, the link dendrogram is partitioned (cut as illustrated in Fig. 8) at the level where the density of links in every community found in the network is maximised. The optimal value of this measure is called partition density $D$. The lowest value for $D$ is -0.67 and indicates a community of two completely disconnected links. The algorithm's setup by default extracted link communities by partitioning the link dendrogram at multiple threshold levels. Every height of the dendrogram at which

118 The number of links found in the communities overlap.

new community was discovered was also registered. Then, the partition density was normalized against the minimum and maximum numbers of links that could belong to each community.

The principle of calculating link similarity of only selected number of links, rather than between each link in the network, is justified by the probability that a pair of links will be similar to each other if these links share a common node. In this way the computation time needed for comparing link similarity in the entire network is significantly reduced. After the pairwise similarities of all the links in the network were calculated, links were clustered hierarchically where each link belonged to exactly one community, while nodes were allowed to belong to multiple communities at once. This corresponded to the real-life examples of complex urban systems explained in detail in the Introduction of Part 2 of this thesis. The interested reader is directed to find more detailed explanation of this principle in Ahn et al. (2010a; 2010b). The same study also offers validation of the method's outperforming capabilities compared against some of the traditional algorithms, such as the Girvan-Newman or the Clique Percolation.

### *8.7.4 Measures of Community-based Node Centrality*

The community-based centrality measure essentially calculates the node's centrality relative to the similarity between the communities the node is part of. The measure defines a type of importance of a node in all the communities it belongs to. A high value indicates that the node has more general and dominant role in the entire network, contrary to nodes that appear in fewer communities and tend to have more specialised role. For example, in social networks, it is well accepted that individuals who are not very central to communities, belong to the community's 'boundary' (the edge effect). However, they are extremely important to the community observed as a group, since these individuals connect many unconnected people. Thus, these nodes (or agents) effectively act as gatekeepers of information flow and enjoy unexpected level of control. Measured in network science, this situation is observed when the network is disassortative. However, the disassortativity metric is typical node measure and not a link measure, thus although it can identify a network with potentially high modularity, it does not capture link memberships.

Intuitively, in assortative networks we would expect hubs to have the greatest importance in the entire network, but this is not the case either. Instead, as argued earlier and confirmed by the results in the Appendix of this thesis, high community centrality was typical for nodes that had more connections to other nodes outside their community and they contributed more to the network's modularity. These nodes were neither the network's hubs in any of the eight modelled graphs. Thus nodes that acted as 'community hubs' were less important and had very small or no effect to the functional integrity of the community as a unit. Furthermore, the hubs global relevance within the network from the aspect of modularity was trivial. This resulted from the strong influence of the nodes with high community-based centrality. In future modelling work, I intend to investigate whether these nodes will still maintain the community's central role in the network, even in the case of potentially removing the community's hubs.

As explained in the earlier paragraph of this sub-chapter, the focus of this research was to confirm the disparity between the physical and logical topologies of the network and therefore the analysis involved measuring the corelation coefficients for several of their structural differences.

First, for all eight shires I calculated the correlation between the node centrality and the community-based node centrality (see the summary results of their linear models in the Appendix). The community-based centrality was calculated by weighting the similarity of each community a node belonged to, to all other communities the same node was part of. Equation 3 calculated the community weights over the sum of $N$ communities in which node $i$ was found and the $S_{(j,k)}$ term represented the similarity value between communities of nodes $j$ and $k$. Essentially, $S_{(j,k)}$ is the Jaccard coefficient between the communities being compared. Then, the sum is averaged with the $m$ number of communities. The community-based centrality measure shows unique valuation of the node's importance in the network.

$$C_{c(i)} = \sum_{i \in j}^{N} \left(1 - \frac{1}{m} \sum_{i \in j \cap k}^{m} s_{(j,k)}\right)$$

Equation 3: Community weighing measure introduced in Kalinka and Tomancak (2011, p 2012)

## 8.8 LIMITATIONS OF THE STUDY

First limitation of this study is using the one dimensional approach to measuring access to public transport by mapping only bus stops and car/bike share locations. Further limitation to measuring public transport access is the frequency of service and availability of sharing utilities at the time of demand. Second limitation of this study that should be addressed in future research is the generation of the OD accessibility matrices constrained only to the first nearest neighbour search. Future study should analyse all-to-all accessibility of each point of interest, not grouped in the four main categories presented here and map them individually based on their use class into separate logical and physical graphs. Additionally, accessibility matrices should be also generated between each point of interest as well. Third limitation is defining accessibility to only measure the social facet of access to infrastructure, ignoring competition and temporal constraints.

Fourth limitation in this study was in the technical capacity to represent the primal and dual graph of infrastructure networks. The number of nodes which are redundant (pseudo-nodes) can be significantly reduced, which will probably affect the overall network size (diameter), considerably reducing the diameter sizes for the physical graphs as shown in Table 6.

The fifth limitation is linked to my theoretical and practical understanding and capacity of modelling network growth. To fully test the research hypothesis introduced in this thesis in future I need to simulate growth on the networks.

The sixth limitation of the study has technical aspect. Namely, large areas of forest land are converted to point datasets, and their street access location is unknown. There is a great chance that for some households the distance to forest land was measured to a point that in reality is not physically accessible.

Last technical limitation of the study stems from the restricted access to good quality with spatial granularity of the data. Namely, the pedestrian network of Glasgow metropolitan region is not fully mapped in OSM. Thus, it is represented only by a fraction of footways which are greatly disconnected and are largely missing in many parts in the network. Thus, to reproduce the pedestrian infrastructure network I used the network dedicated to motorised traffic. That means that in some parts of the analyses, there is a great chance that the pathfinding process traversed parts on the motorways, which are not suitable for walking.

# Chapter 9: Findings

The regression analyses between the traditional node degree and the community centrality showed that nodes which are hubs in the network are not the ones that have centrality (i.e. importance) within their own community. This is observed by the low correlation coefficients as shown in Table 5. In the results from the linear regression models presented in the Appendix (Tables 7-10) we can see that most of the models had well balanced symmetrical residuals (median value around 0) which meant that the predictions of those models were well in line with the observations. The exception in the models is the one for betweenness centrality and community-based node centrality. All of the models for the dependent variable (i.e. betweenness centrality) had extremely skewed upper range residual values, which meant when using the community-based node centrality as the predictor, the models were not well predicting the high range of nodes with high betweenness centrality. Similarly, looking at the intercept and slopes (Tables 11-14), we can see that the intercepts are reasonable in all shires' linear regression models, except the linear regressions of the community-based node centrality to the betweenness centrality. For these models, all eight shires show to be highly intercept-driven. The standard error for these models is also relatively heigh, which tells that there is very high uncertainty in the predictions.

Next to the regression models, to answer my research question, I calculated the value for the partition density at the level of the dendrogram's cut[119] where the densities of links within the communities was maximised. This value carried greater information for this research compared to the conventional dendrogram plots, which typically only presents the community hierarchy within the network. The value for the partition density was more important since this measure assesses the quality of the link communities found at each level of the link dendrogram.

The dendrogram contains information on several local measures. For example, Table 3 is an excerpt from the large table containing information on every height in the link hierarchy at which communities appear and the value of partition density for each of them. Another valuable information maintained in the output of the community clustering algorithm is the connection between the unique node IDs and community IDs they belong to. As presented in Table 2, we can see that one node ID can be found in more than one community. Opposite to this, every link ID is linked to only one community ID, to which the link belongs. Already discussed extensively throughout this thesis, the strength of this algorithm is that it clusters links which can only belong to one community.

---

[119] The level at which the dendrogram is partitioned (i.e. sliced) to optimize the number of link communities.

| Node ID | Cluster ID |
|---|---|
| 1 | 2222 |
| 1 | 4445 |
| .. | .. |
| 1 | 7964 |
| 10 | 5774 |
| 10 | 7176 |
| 10 | 7613 |
| 10 | 7964 |
| 100 | 1607 |
| 100 | 5776 |
| .. | .. |
| 100 | 7964 |

Table 2: Example of the nodes with ID1, ID10 and ID100 found in the node communities in East Dunbartonshire.

From the algorithm results we can easily access the community-based node centrality of each node. For example, the community centrality for node ID 1 presented in Table 2 is 8.46, the community centrality for node ID 10 is 4.17 and the community centrality of node ID 100 is 7.43. This measure is normalised by both, the number of links the communities to which the node belongs, established to other communities and the number of nodes found inside the community, multiplied by the average degree of the entire graph.

In table 3 I presented an example of the partition density measured for each of the dendrogram's heights at which communities appeared. For example if we take the lowest height 0.99996 presented in Table 3 in the link dendrogram of East Dunbartonshire, the community's partition density found at this height has the value of 0.000584. This measure as mentioned earlier, calculates the quality of the link community at that dendrogram height. We can see that the value of the partition density can vary and is very informative. For example at the highest level of the link dendrogram 1, where all the links belong to the Giant Component, the community quality is nearly 2.5 times worse than the one found four levels below. This is also intuitive since the communities mashed up into one GC are very unlikely to have links that are similar to one another, unless the graph is complete, or at the very least it is extreneky dense.

| Heights | Partition density |
|---|---|
| 1.00000 | 0.0002392641 |
| 0.99999 | 0.0005835717 |
| 0.99998 | 0.0005836048 |
| 0.99997 | 0.0005836158 |
| 0.99996 | 0.0005837082 |

Table 3: The top five levels of the East Dunbartonshire dendrogram's partition density table

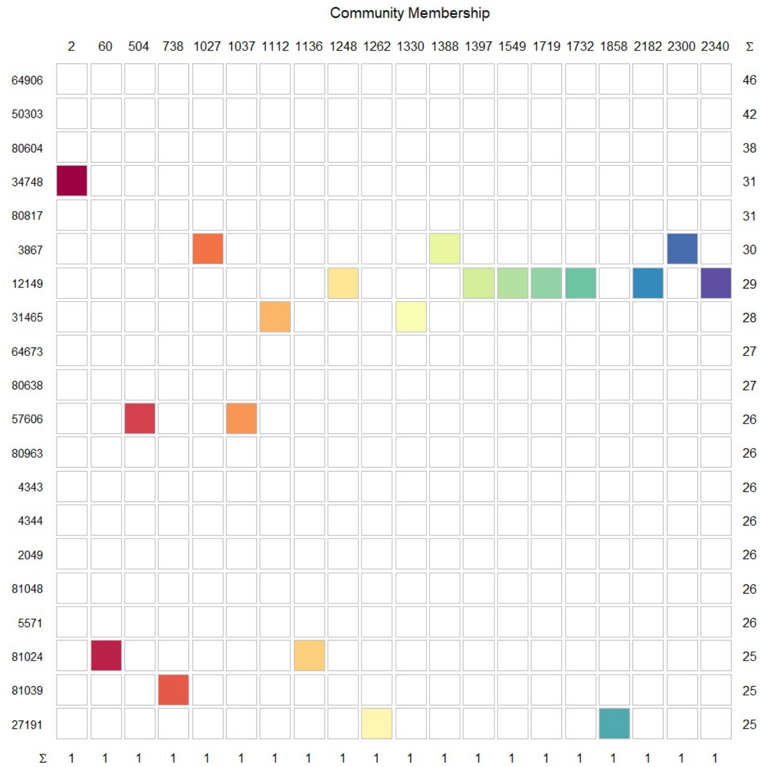
| Dend. height | merge 1 | merge 2 |
|---|---|---|
| 5.651219e-09 | -11372 | -13989 |
| 5.651219e-09 | -37417 | -37425 |
| 5.651219e-09 | -71748 | -71756 |
| 5.651219e-09 | -80794 | -80921 |
| 5.651219e-09 | -109729 | -110465 |
| 5.651219e-09 | -177457 | -178593 |
| 5.651219e-09 | -228176 | -229090 |
| 7.044226e-09 | -3497 | -8036 |
| 7.044226e-09 | -65661 | -65746 |
| 7.044226e-09 | -110350 | -110578 |

**Table 4: Example of the first ten hierarchical clustering of link communities at the specific dendrogram heights**

The other important result from the link clustering algorithm is the levels at which the link communities themselves are hierarchically clustered. This measure shows the nested hierarchy of the meta-communities per se, and provides the opportunity to explore the communities' relatedness at more abstract level in the graph.

In Table 4, for example, we can see that at level 5.651219e-09 of the link dendrogram, the meta-communities at link community merges -11372 and -13989 are grouped. The relatedness is generated by the Dynamic Tree Hybrid algorithm (Langfelder et al., 2009), which uses the link dendrogram generated in the earlier stages. The algorithm respects the order of the link communities clustered on the dendrogram's tree. It is usual to observe positive or negative values in the output vector of the Tree Hybrid, such as the negative values of -11372 and -13989. The positive values are found on the vector's scalar between the point of 0 and the breakpoint, and after the transition point, while the negative values are found between the breakpoint and transition on the same scalar vector. For more detail on the algorithm, the reader is referred to Langfelder et al.'s publication (op cit.). The Dynamic Hybrid algorithm aside the information retrieved from the link dendrogram, additionally uses the dissimilarity of the communities (i.e. the meta-communities relatedness). The meta-community relatedness is calculated by calculating their Jaccard index; the ratio between the number of shared nodes between two meta-communities and their nodes' union.

The other very important results from the hierarchical clustering algorithm are the community overlaps. These are crucial since these overlaps, as established in earlier chapters, are the ones that increase the network's sustainability and reduce its vulnerability to external shocks. To present the discovered community overlaps in the graphs, I constructed the community-membership matrix which showed the community overlaps of the 20 most frequent nodes that appeared in different communities. The left hand side in Figures 17-20 displays the node's IDs, the right hand side displays the total number of times the node appears in various communities.



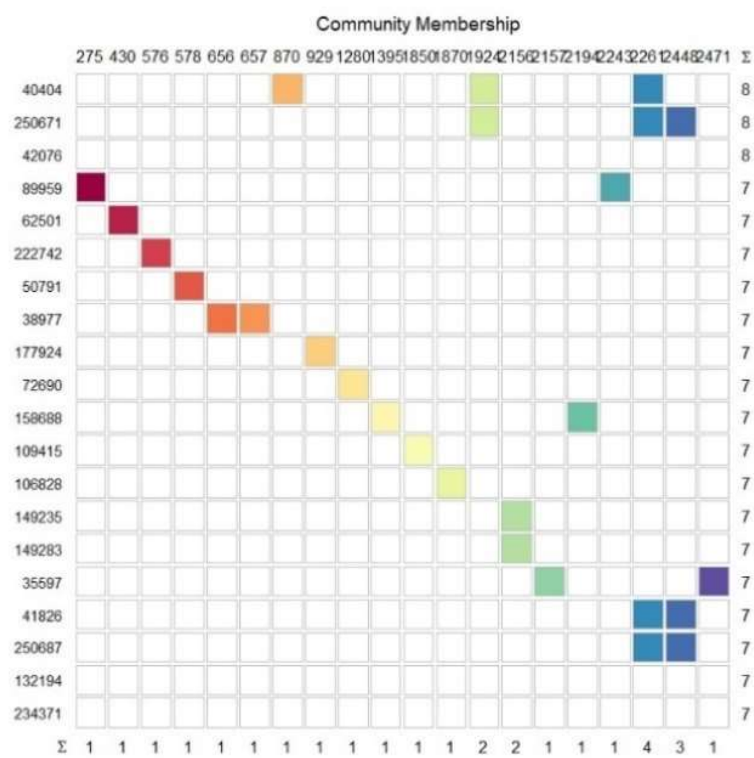
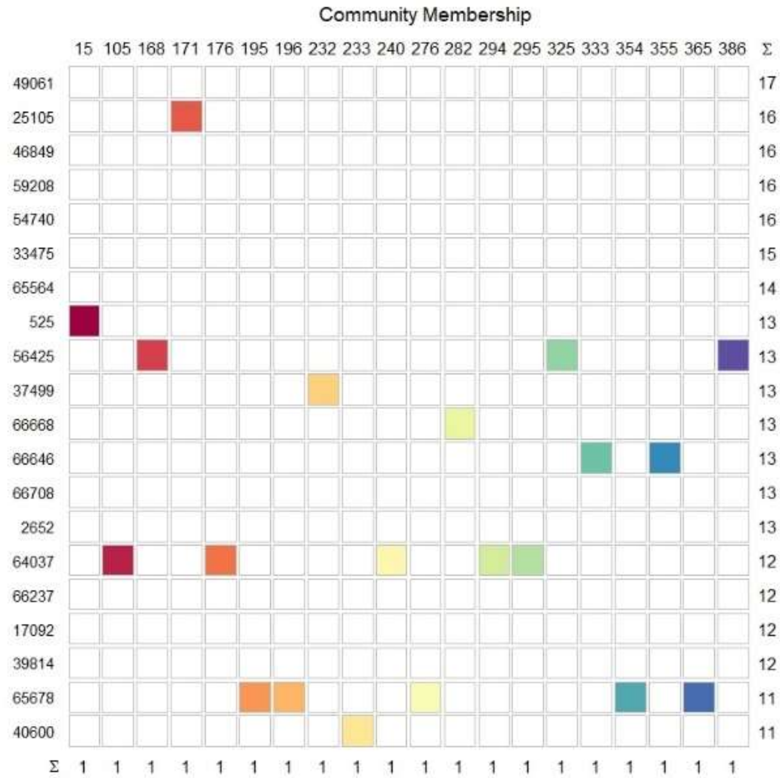




Figure 17: Community membership number of nodes for the link communities in East Dunbartonshire (top) and Glasgow (bottom) (Author's own graphics).



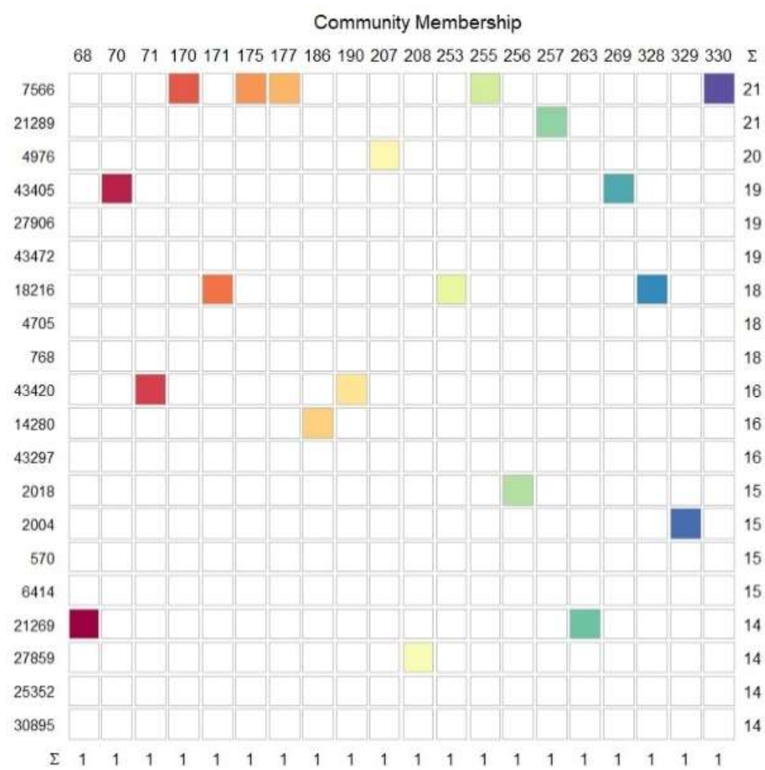
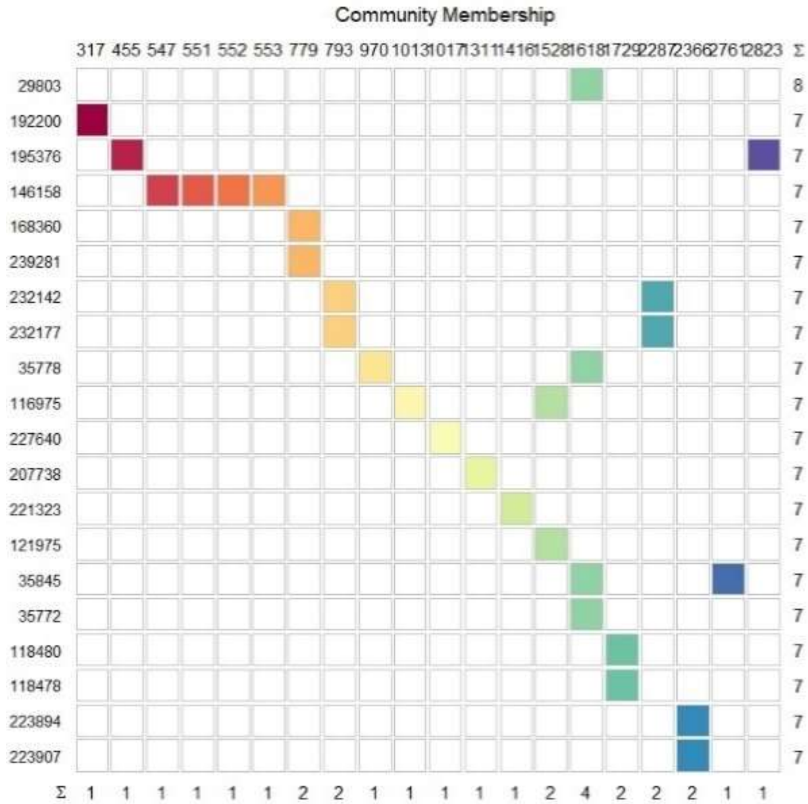




Figure 18: Community membership number of nodes for the link communities in East Renfrewshire (top) and Inverclyde (bottom) (Author's own graphics).



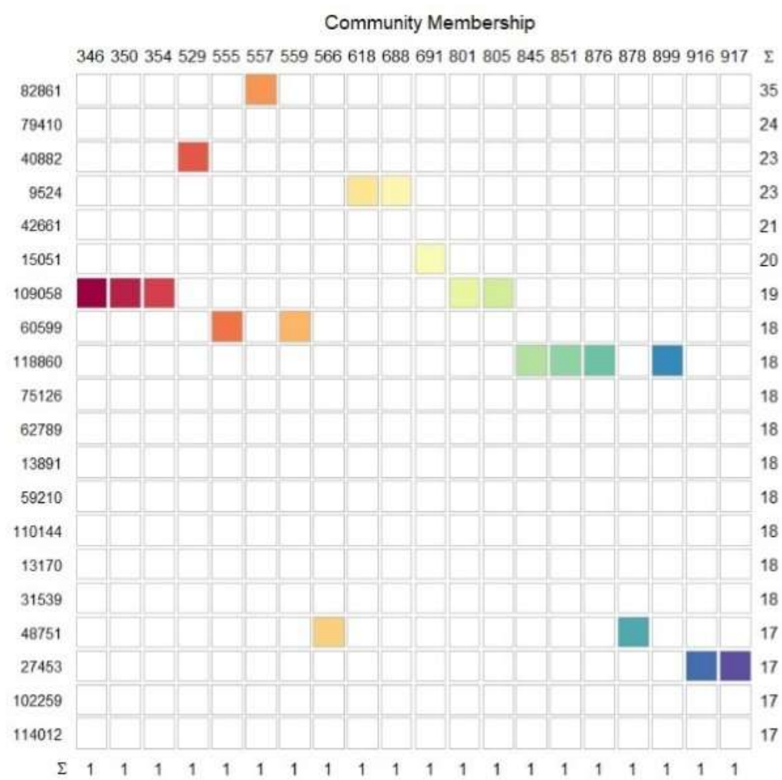
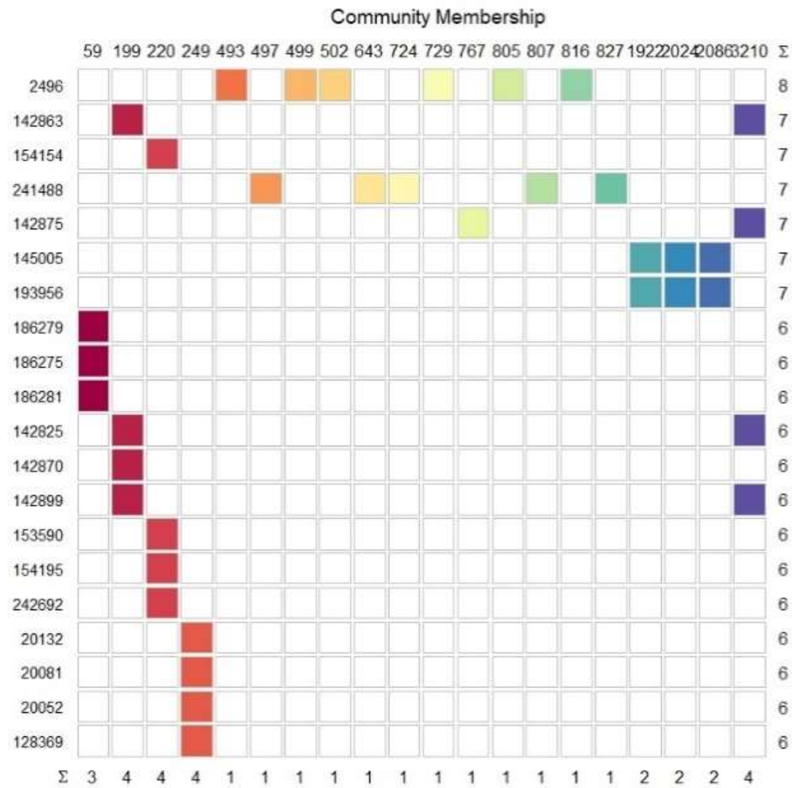




**Figure 19: Community membership number of nodes for the link communities in North Lanarkshire (top) and Renfrewshire (bottom) (Author's own graphics).**



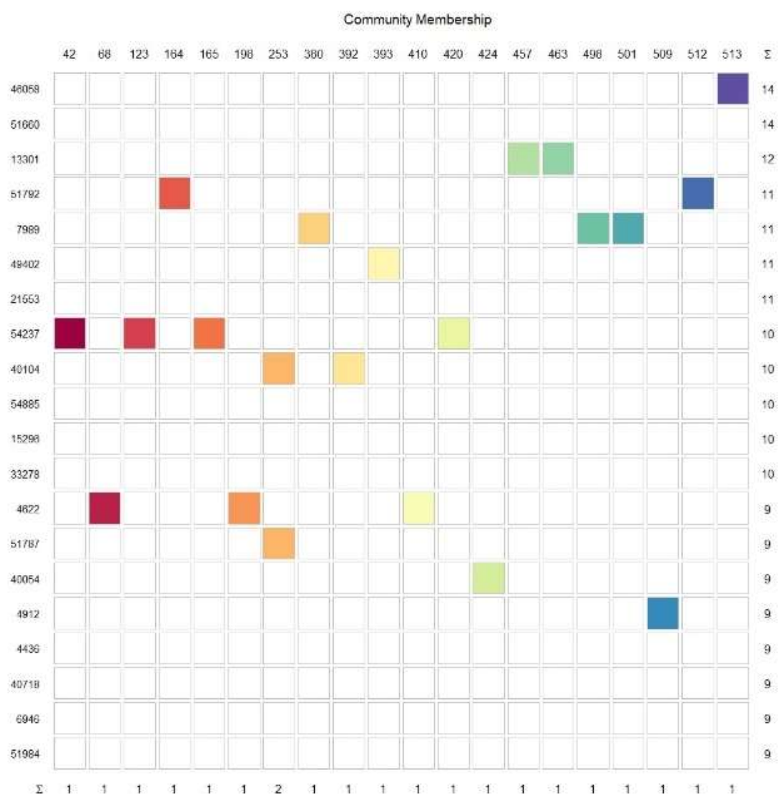




**Figure 20: Community membership number of nodes for the link communities in South Lanarkshire (top) and West Dunbartonshire (bottom) (Author's own graphics).**

At the top side of the matrix we can see the IDs communities in which each shared node is found and at the bottom we see the sum of the top 20 nodes present in the same community ID. With this visualisation of the shared nodes found in the overlaps of the communities I could see how many of the most frequently shared nodes were shared across the communities' overlaps. The matrices in Fig. 17-20 visually represent the measure equivalent to the community size (Palla et al., 2005) and the size of the community overlaps as discussed in 7.2 (p.74) in this thesis. The colouring of the grids in the membership matrix indicates that same colour grids (i.e. nodes) belong to the same community. The colour scheme ranges within the RGB spectrum and doesn't serve any purpose other than a indicative. For example in Figure 17, the community with ID 1618 (ID found on the top side of the matrix) has four from the top 20 nodes that are found within the communities' overlaps (see the matrix bottom side). Namely, the nodes with IDs 29803, 35778, 35845 and 35772 (see left matrix side) are found in eight (ID 29803) and seven community overlaps (35778, 35845 and 35772). We can find this information on the right side of the same matrix.

As presented in the matrices below, the logical topologies of the eight shires had very trivial overlaps i.e. the size of the communities' overlaps (Palla et al., 2005) was negligible, which is measured by the cardinality of the node cut-sets. The very low (in the best case only four) number of shared nodes across the communities shows that they are not very sustainable. Specifically, they do not show great points of synergy between the technical and social infrastructure as defined in this study of access across the different levels. The different levels of social infrastructure synergies are captured in the overlapping (i.e. co-located) access to greenspace, public transport, or mode shares such as cars, bikes or e-charging units. The full list of the fourteen measures to access can be found in 8.7.1.

The lack of large overlap sizes, i.e. large number of shared nodes, in addition to the low values for the communities' similarity observed in the results, indicates that the logical graph constructed from the fourteen accessibility measures only very marginally capitalises on the potential synergies that can be created among these fourteen networked infrastructure sub-systems.

Increasing the number of links with high betweenness centrality, as well as increasing the number of nodes that have high community centrality will improve the equity to access of services in the region, and in the same time optimise their flows' rate to be more sustainable; by spatially decentralising them. For example, in the output of the results, not only the physical, but as well the logical graph is spatially referenced. The edge list of the input adjacency matrices includes all node 'from' and 'to' IDs for the spatially referenced logical graph, and they are kept in the link clustering process when the link communities are discovered. This means that we can trace back in space the exact nodes and links that belong to these overlaps. We can also trace back in space link communities with very low similarity values. The communities with low similarity values have ‘tree-like’ structure. Opposite to this, communities with high similarity value have ‘semi-lattice’ structure and allow high number of overlaps. This can be examined for any of the nodes that belong to the different social infrastructure separately or in a combination. The process is done in the case when we are looking to increase the overall number of communities that overlap, or only increase the size of the communities that already overlap. For example, if we would like to add more greenspace, but in the same time increase the access of residences to greenspace, we can take the car and bike share social infrastructure and create synergy (i.e. new community overlap) across the sub-systems, by adding more bike or car share locations that are within a proximity of 400 meters or less to residences and within a cut-off distance between 1 km and 2 km for cycling and 1 km and 5 km for driving. The cut-off distances for cycling and driving in this example are only illustrative, more suitable distances should be chosen based on local travel survey results.

Finally, the graphs’ diameters and their degree distributions were calculated and compared. Table 6 below summarises the findings for the eight logical graphs tested. The correlation between the community centrality and the classical node centrality measures indeed revealed discrepancies between the node’s importance when it was considered as individual member in the network, against its importance when it was considered as an integral part of one (or several) of its communities. Moreover, the significance was relatively heigh for many of the graphs.

## 9.1 THE LINK COMMUNITIES OF GLASGOW AND CLYDE VALLEY

This research examined the sustainability of the networked infrastructure in the Clyde Valley city-region through mapping the spatial organisation and access of infrastructure networks to residents. Through the representation using physical and logical graphs, the model measured the accessibility of social infrastructure observed as a complex multilayer network. The results of the logical and physical topologies of the complex networks separately mapping the eight shires are summarised below.

As I presented in this section, none of the eight logical topologies in the council areas showed high clustering coefficient and shared very few nodes across their communities. Further, these networks had typical degree distributions as also reported in many of the studies reviewed in Chapters 4 and 5. In Table 5 below we can see the summary of the regression analyses performed on the eight graphs, studying the structural correlation between their physical and logical topologies.

From these findings I concluded that the network holds great potential to improve its overall performance in respect to both social accessibility and sustainability. As already concluded in Chapter 6 on p64, lack of system level sustainability affects the system's overall resilience. The resilience of the complex system could be improved by increasing the number of overlapping nodes that have high value for the community centrality measure. Further, system's efficiency can be increased by reducing the networks' extremely large diameters (see for example the diameters of the physical graphs in Table 6) through establishing new links with high edge betweenness that connect the nodes with high community centrality across the communities, rather than within the same community.

| | Node 'link-community based' centrality vs. node degree Corr. coefficient | Node 'link-community based' centrality vs. node betweenness centrality Corr. coefficient | Node 'link-community based' centrality vs. closeness centrality Corr. coefficient | Node 'link-community based' centrality vs. Burt's constraint Corr. coefficient |
|---|---|---|---|---|
| East Dunbartonshire | 0.5442148 | 0.2751505 | 0.1576403 | **-0.4746445** |
| East Renfrewshire | 0.4489936 | 0.1733935 | 0.0317823 | **-0.7348904** |
| Glasgow City | **-0.1228974** | **-0.0006983233** | 0.05512692 | **-0.2685352** |
| Inverclyde | 0.5442349 | 0.1912694 | 0.127343 | **-0.6243545** |
| North Lanarkshire | **-0.05813179** | 0.001913787 | 0.1558255 | **-0.4443357** |
| Renfrewshire | 0.6149454 | 0.1047212 | 0.07314386 | **-0.6200792** |
| South Lanarkshire | 0.02518971 | 0.01485589 | 0.201031 | **-0.5764703** |
| West Dunbartonshire | 0.436065 | 0.1531385 | 0.3608804 | **-0.7490534** |

Table 5: The correlation coefficient *r* of the nodes' community-based centrality to the classical node centrality measures: degree, betweenness, closeness and Burt's constraint

Table 6 below summarises the main structural characteristic of the graphs' sizes. The networks' diameters, which measure the maximum step-length distance between the two nodes in the network that are furthest apart, varied significantly across the topologies. The logical graphs had relatively small diameters. This was due to the fact that these graphs were not fully connected (see Figures 21-29). I.e. the clusters of nodes that were tightly connected in small sub-graphs showed the importance of including the geodesic distance when calculating the path lengths.

The high differences in the diameter sizes between the physical graphs and their line-graphs showed that the diameter of the graph that represents the accessibility of streets is not as large either. Although the diameters of their spatially constrained (i.e. plane) networks were characterised with a substantial size, their planar graphs had significantly reduced the path lengths. This requires future exploration.

| | Logical graph | Physical graph (primal) | Line-graph (dual) |
|---|---:|---:|---:|
| East Dunbartonshire | 27 | 29753.75 | 681 |
| East Renfrewshire | 62 | 30540.34 | 631 |
| Glasgow City | 114 | 24809.79 | 641 |
| Inverclyde | 57 | 32541.88 | 587 |
| North Lanarkshire | 156 | 43139.86 | 787 |
| Renfrewshire | 92 | 27842.35 | 448 |
| South Lanarkshire | 81 | 76115.06 | 766 |
| West Dunbartonshire | 53 | 40653.64 | 458 |

Table 6: Overview of the networks' diameters

Moreover, the results presented in the Appendix confirmed that nodes which were measured as structural hubs in the network were not as important when they were observed as part of the larger community they belonged to. Instead, other nodes with lower average degree that were identified to have more external connections than within the community that they belonged to, were the ones that affected (i.e. increased) the modularity of the system. These findings are very promising. They are particularly encouraging for future research in this area and offer two unique methodological advancements for the field of urban science:

[1] improved and comprehensive, yet not too restrictive, integration analysis of the networked infrastructure (i.e. integrated analysis of LTSs as part of one complex system);

[2] opportunity to test the accessibility of locations from an integrated aspects of networked infrastructure, by user defined criteria where the links of the graph can represent any socio-economic process we'd like to explore.

In the following sub-section I summarised the main findings in all eight shires in detail.

### *9.1.1 East Dunbartonshire*

The total number of elements of the undirected graph of the logical network in East Dunbartonshire included 40.604 nodes and 237.824 edges. The algorithm using single-linkage hierarchical clustering discovered 8.389 communities in this graph. The largest cluster within these communities had 1.157 nodes. The maximum similarity value $D_c$ at which the last two clusters were merged was 0.02141042. This similarity value was recorded at the maximum height of 1 for that relevant dendrogram branch for that community. The minimum value for both the dendrogram height and partition density in this graph was 0, while the arithmetic mean of the link dendrogram was 0.0058513 at a height of 0.468699 and the median partition density for all of the communities in the East Dunbartonshire logical network was 0.001946059 at height of 0.41104 of the link dendrogram. Essentially, the maximum partition density measures how 'semi-lattice' opposed to 'tree-like' each link community is. The relatively low median and mean values for partition density of the logical graph show that overall, the graph is very 'tree-like'. The hight at which the dendrogram was partitioned (the value of $D$) to maximise the number of communities was 0.84228.

The physical topology (the infrastructure) of the network was constructed from 81.095 nodes and 82.166 links that were part of the fully connected graph (i.e. part of the Giant Component). The dual graph of the physical network was constructed of 82.166 nodes and 129.466 edges. The large difference between the number of nodes in the logical compared to the one of the physical graph meant that only half of the houses in East Dunbartonshire were connected to the basic amenities[140] as per the study's definition, within a 400 meters network-based distance.

The relatively low value of $D_c$, which measures the quality of the link partition, suggests that each of the communities' stricture in the network is 'tree-like'. This result coincided with findings in preceding notable researchers (Alexander, 1965; Glass, 1998) which relevance for the field of urban studies were discussed in Chapter 3. Thus, the empirical results that I obtained in this thesis empirically confirmed the early theoretic and qualitative analyses that cities are planned and organised in a tree-like pattern.

Further, the modularity of the logical network's graph was calculated using network partition strategy implemented in the well-known multi-level modularity algorithm. This algorithm is also known as the Louvain algorithm and its mathematical logic is based on a greedy strategy, without any a-priori input. The greedy algorithm randomly organised the nodes of the physical network by partitioning its graph into 72 modules, to reach the optimised modularity value of that network at 0.83. The optimal modularity of the line-graph of the physical network was as heigh as 0.99 and the network was divided

[140] As defined for the purposes of this research in the Methods.

using the same multi-level modularity algorithm into 236 communities. The result of lower number of discovered communities in the line-graph confirmed that applying the link clustering algorithm allows for significantly better resolution of the community structure detection in the network.

The measured values for transitivity and the assortativity degree of the logical graph were respectively 0.01517217 and -0.3407669, opposed to its physical counterpart for which the transitivity was measured at 0.002879146 and the assortativity degree had a value of -0.3294544. The transitivity of the dual graph of the physical network, also known as line-graph, was 0.4904256 and its assortativity degree was 0.4699974. The transitivity measurement is a property of the network that measures the likelihood for the first-tier neighbours of a randomly selected node to be neighbours themselves. The transitivity of the logical topology, although relatively low as expected based on the conclusion from previous studies on spatially unbounded growth, nonetheless was significantly higher for the logical topology of the network, compared to the one of its physical counterpart.

As presented in 7.4.3, new evidence and discourse on the graph's transitivity was introduced by Kaiser and Hilgetag (2004), who concluded that the high clustering coefficient of spatially bounded graphs (such as the ones in this thesis) is a signature of high modularity of those networks and the presence of communities. Additionally, even though the growth[141] in these spatially explicit graphs was spatially restricted, the high density in these networks remained constant. Following this, the clustering coefficient and the network's density, also as proposed by the Barabási – Albert (BA) model, only decreased when modelling spatially unbounded growth. The conclusion from Kaiser and Hilgetag's study can be used to extend the research of modelling Glasgow's infrastructure networks and urban population growth in future projections and scenarios.

The negative value of the assortativity degree registered in the primal graph of the physical network shows that the network is in fact disassortative and has relatively high value[142]. For the value of 1, the network is perfectly assortative and each node in this network links only to nodes with exactly the same degree. For the maximal negative value of -1, the network is perfectly disassortative and the reverse is true for the nodes in disassortative networks.

The results for East Dunbartonshire show that in both the logical and the physical graphs, hubs are present and they tend to link to nodes with lower degree which, indicative highlights the role of

[141] Adding or removing nodes or links.

[142] The assortativity value (i.e. degree-degree correlation) ranges between 1 and -1. The positive value symbolizes that hubs connect to hubs in the network and the negative value shows that hubs tend to link to low degree nodes. The value of 0 shows that the network is neutral and the pattern of nodes linking with each other is random and not dependent on their degrees.

communities in these networks. Although just by the transitivity value itself, we know nothing specific about where these communities are or what their composition is. However, the assortativity degree for the line-graph of the physical network has positive value and it is very high. It shows that indeed streets with large number of connections (in this case building units) are connected to other streets with large number of building units found on them.

The negative value of the assortativity degree of the physical graph is not the typically expected result for its topology, however it was a confirmation of the role of communities in the network. While the positive and relatively heigh value for the degree-degree correlation coefficient of its line-graph is something what is usually expected for these physical networks. These extremely polarised values (in the primal vs. dual problem) measured for the physical network's transitivity, once again reiterates the importance of navigating seamlessly between the primal and dual problem, as extensively argued in 7.4.3. Essentially, the primal and dual problem enables us to study the abstract relations and accessibility solely among streets or other linear infrastructure components; or solely among places in the urban context.


Figure 21: Link communities of the logical topology of East Dunbartonshire.

The different colouring scheme of the links in the map shows the count of flows running through each of the links. The ones in dark blue are the links with highest edge betweenness in the logical graph of the network and the ones in yellow are the ones with the lowest value for edge betweenness.

## 9.1.2 East Renfrewshire

East Renfrewshire's logical graph included 31.663 nodes, 89.873 edges grouped in 4.626 communities where the largest one had 422 nodes. The maximum similarity value $D_c$ recorded for that relevant dendrogram branch of the community at the dendrogram's height of 1 was 0.03343377. The minimum value for both the dendrogram height and partition density again, as in East Dunbartonshire, had the value of 0. The arithmetic mean of this graph found at the link dendrogram's height of 0.6983914 was 0.01864945. The median partition density for all of the communities in the East Renfrewshire's logical network was 0.02057306 at height of 0.7427150 of the link dendrogram. Again, the low value of $D_c$, suggested that each of the communities' stricture in the network is 'tree-like'.

The physical topology of the network was constructed from 66.931 nodes and 72.246 links that were part of the fully connected Giant Component. This meant that 35.268 households did not have access within the 400 meters cut-off distance to the basic amenities in the network. The dual representation of the same physical graph was constructed of 72.246 nodes and 169.836 links.

The modularity detected with the Louvain algorithm partitioned the logical graph's nodes at optimal modularity value of 0.96 into 303 clusters. The same algorithm partitioned the dual representation of the physical graph at the optimal modularity level of 0.99 into 244 clusters. The transitivity of the logical graph was 0.02845898, of the physical graph was 0.002953222 and of its dual representation was 0.7931871. The assortativity degree of the logical graph was -0.3643124, the one of the physical graph was 0.5667559 and 0.9022486 of its dual representation. This meant that the physical network was almost perfectly assortative, approaching the value of 1. Almost all of the subgraphs in this network were k-cliques.

The logical graph had ten times higher transitivity value for its network topology, although relatively low, suggesting that the likelihood for neighbouring nodes of a randomly selected node in the network to be neighbours themselves, was ten times lower. The values of the topologies' assortativity degree show that the logical network is disassortative and the physical network is assortative. Thus, they both have hubs, however in the logical topology of the network hubs tend to link to low-degree nodes whereas in the physical topology nodes with similar degree tend to link with each other (i.e. following the principle of preferential attachment). These results align with the expectations as discussed in the theoretical part of this research.

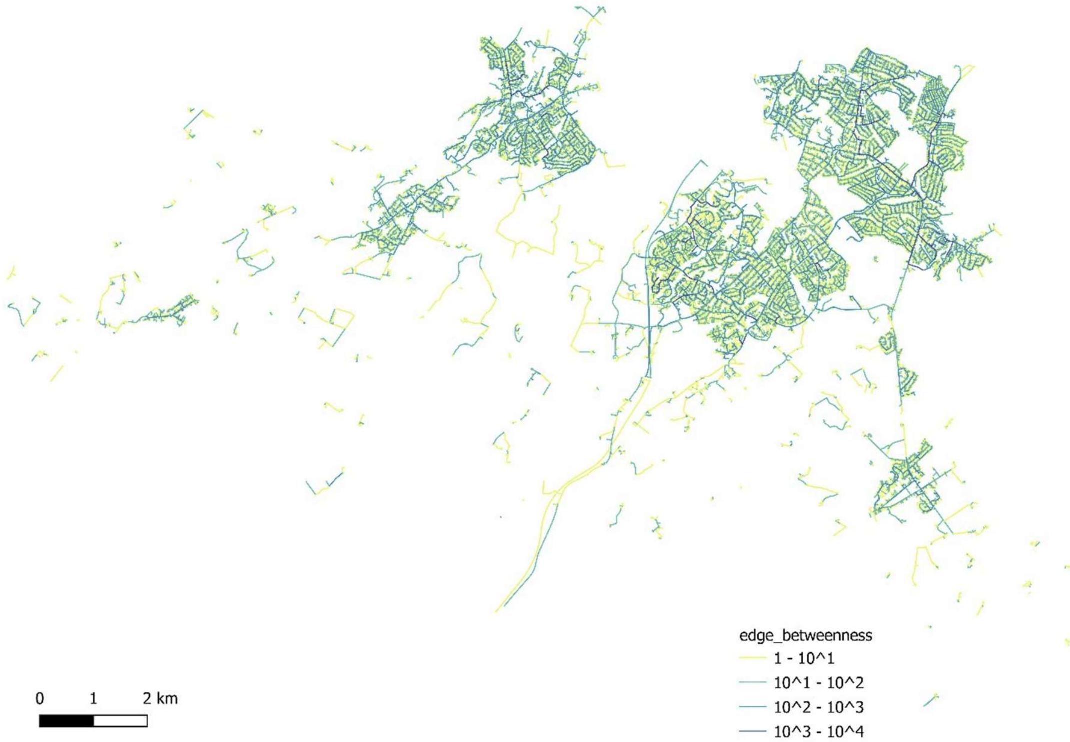



Figure 22: Link communities of the logical topology of East Renfrewshire (south-west region).


Figure 23: Link communities of the logical topology of East Renfrewshire (north-east region).

### *9.1.3 Glasgow City*

The logical topology of Glasgow's network was built of 131.716 nodes and 484.614 links, grouped in 7.007 link communities with the largest one of them containing 964 nodes. The physical topology of the network had 271.124 nodes and 278.345 links part of the GC. The dual representation of the physical graph had 278.345 links and 443.702 nodes.

The highest level at which the dendrogram in the logical network was partitioned to maximise the number of communities was 0.774, and the median and mean values of the different dendrogram levels were both close to this value; 0. 768685 ($D$ = 0.00802334) and 0. 7454483 ($D$ = 0.007134777) respectively. The maximum value for the partition density at the dendrogram height of 1 was 0.01203014. The modularity of the logical network using the greedy optimizing algorithm was 0.96, organising the nodes of the logical graph in 101 clusters.

The same multi-level clustering algorithm at the optimal modularity value of 0.99 found 335 communities in the dual representation of the physical graph. Again, the transitivity of the logical topology was very low; 0.02297802 and the transitivity of the physical topology was 0.004753205 and 0.479659 for its dual representation. Thus the nodes in the logical layer were 5 times more likely to be neighbouring in the case when they shared a node, compared to the physical graph arranged in the primal syntax. And both topologies, the physical and the logical were not 'clique-ish'. While the topology in the dual representation of the physical network, measuring the accessibility of streets and residence access paths via their intersections, found the graph to be very clique-ish. I.e. if two nodes (i.e. streets or access paths) in the dual representation were neighbours, there was very high likelihood (~ 0.48) that they share a third neighbouring node (i.e. street/or access path). This result is expected for the physical network.

Both the logical and physical networks had disasssortativity degree of -0.3807454 and -0.2725315 respectively, however the physical network was less disassortative and was more close to becoming neutral. The dual representation of the graph of the physical network had assortativity degree of 0.5316048.

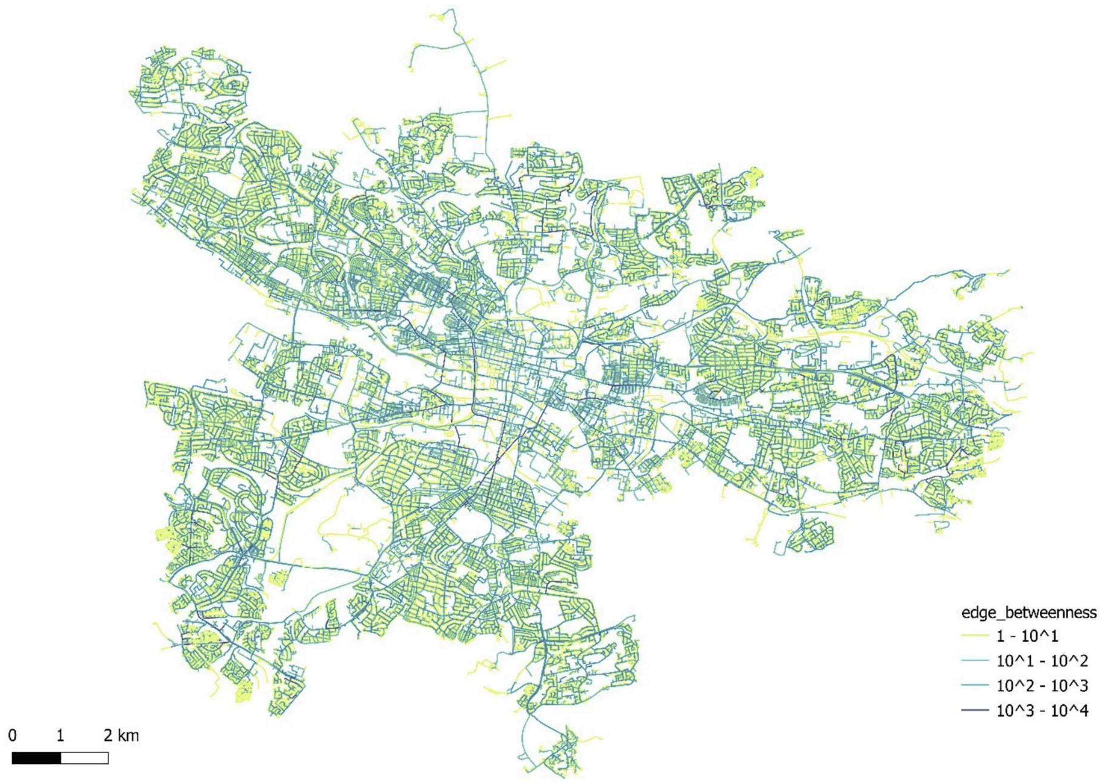



Figure 24: Link communities of the logical topology of Glasgow City.

## *9.1.4 Inverclyde*

Ahn et al.'s algorithm (2010) hierarchically clustered 4.628 link communities out of the logical graph constructed from 20.695 nodes and 73.656 links. The largest community had 285 nodes and the height of the dendrogram at which the link communities' number was maximised was 0.497. The mean value of the communities' partition density measured on the link dendrogram was 0.02489461 at the height of 0.648, and the median value was 0.03109 at the height of 0.6713. The maximum partition density at the top of the link dendrogram was 0.04120711. The optimised modularity of the logical graph at the level of 0.95 contained the maximum number of 210 node communities. The same multi-level clustering algorithm, at the optimised modularity level of 0.99, partitioned the dual representation of the physical network in 203 communities.

The clustering coefficient (i.e. transitivity) of the logical graph was 0.03541291 and the one of the physical network was 0.003447137. This meant that in both networks the likelihood of finding triplets was very low, but ten times lower in the physical than in the logical topology. In its dual representation, the physical graph was highly assortative, with an assortativity degree of 0.885203 and the transitivity in the dual representation of the same graph was 0.7016227.

The physical topology (the infrastructure) of the network was constructed from 43.535 nodes and 46.877 links that were part of the fully connected graph (i.e. part of the Giant Component). The dual representation of the physical graph had 46.877 nodes and 101.854 links.

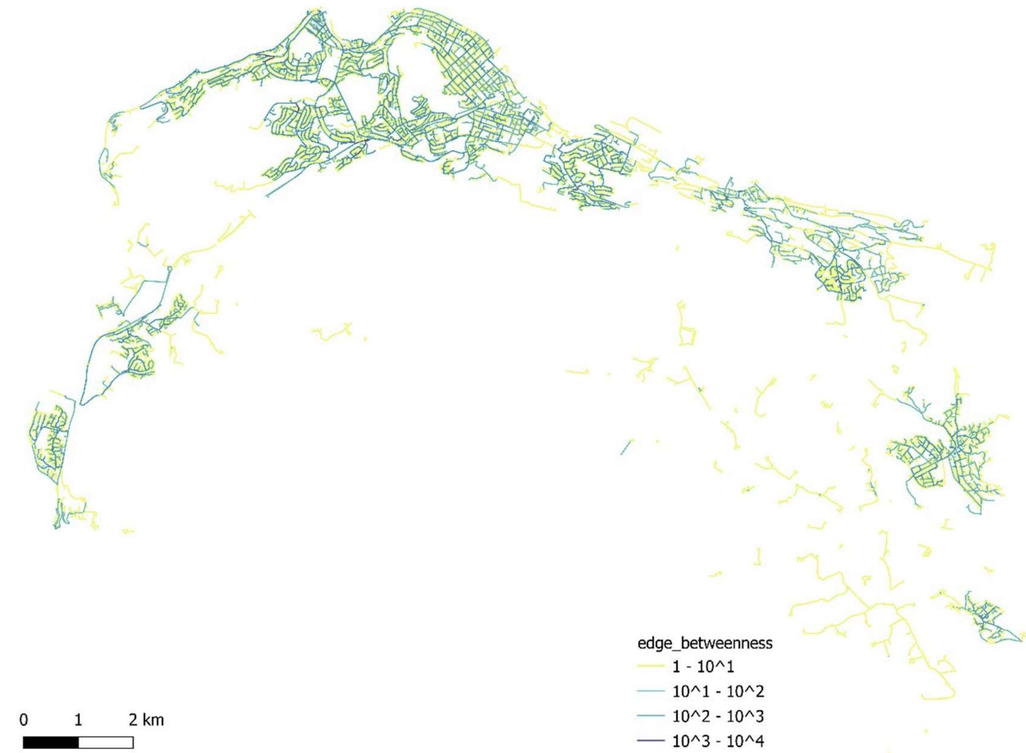



Figure 25: Link communities of the logical topology of Inverclyde.

## *9.1.5 North Lanarkshire*

The logical topology of North Lanarkshire's network consisted of 119.654 nodes connected to each other by 403.561 links. The single-link clustering algorithm found 7.243 hierarchically clustered communities of which the largest community was composed of 704 nodes. The level at which the dendrogram was partitioned to maximise the number of communities was 0.749. At the mean height of the network's dendrogram of 0.7447 the value of the partition density was 0.007969. At the median height of the network's link dendrogram 0.7695 the partition density was 0.008513 and at the height of 1, the partition density was 0.012555.

The modularity of North Lanarkshire's logical topology was 0.97 and the greedy multi-level modularity clustering algorithm at this optimum network modularity level found 393 communities. The same greedy algorithm partitioned the dual representation of the physical graph into 535 communities, for the graph's modularity maximum value of 1.

The physical topology was mapped into a graph of 243.826 nodes and 266.483, which comprised the Giant Component. The dual representation of this graph had 266.483 nodes and 703.431 links. The disassortativity of the network for the logical topology had a negative value of -0.3857842 and the assortativity of the physical topology was 0.5481344. The assortativity degree of its dual representation was 0.8810247. The hubs in the logical topology of the network tended to connect to nodes with lower degree and this had very significant prevalence across the entire graph. While the hubs in the physical topology connected to hubs and low degree nodes to low degree nodes. The clustering coefficient of the logical topology was valued at 0.02140085. With that, the nodes of the physical network were five times less likely to connect in triplets with a clustering coefficient of 0.004462356. The clustering coefficient (i.e. transitivity) in the dual representation of the physical network's graph was 0.8128601, again in full alignment with my expectations.


Figure 26: Link communities of the logical topology of North Lanarkshire.

## *9.1.6 Renfrewshire*

The logical topology of the network was consisted 57.218 nodes connected to each other by 170.183 links and the single-link clustering algorithm found 10.451 communities. The largest community was composed of 268 nodes. The level at which the dendrogram of the logical graph was partitioned to maximise the number of communities was 0.497. At the median height, the dendrogram had a value of 0.6995 and the partition density at this point had a value of 0.02490. At the mean height of

0.6700755 for the network's link dendrogram, the partition density had a value of 0.002201509 and at the maximal height of 1, the partition density was 0.03897092.

The greedy algorithm divided the Renfrewshire's nodes into 543 communities at the optimised modularity value of 0.97 for the logical graph. While the same algorithm divided the dual representation of the physical graph in 394 communities at the optimised modularity value of 0.99. The Giant Component of the physical topology of Renfrewshire's infrastructure was reconstructed by a graph with 134.975 links that connected 121.753 building units of various uses. Its dual representation had 134.975 nodes and 376.996 links.

Renfrewshire's logical graph had the clustering coefficient of 0.03269303 and as expected, the physical topology of the same network had a clustering coefficient of 0.005091117. In the dual syntax the physical topology had transitivity value of 0.8274154. The logical network was disassortative with a coefficient value of -0.3949496, while the degree-degree correlations of nodes with similar degree in the physical network was very strong. This was expressed by the high assortativity value of 0.5516727, which was significantly higher (0.8862654) for the dual syntax of its physical graph.

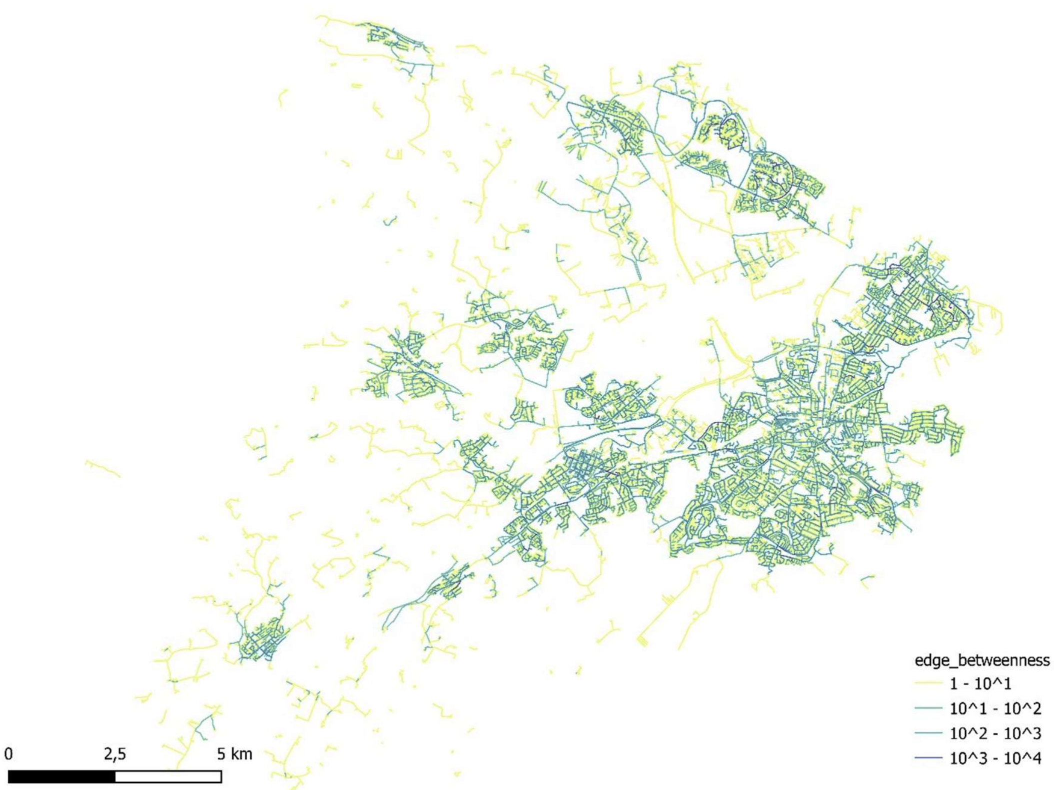



Figure 27: Link communities of the logical topology of Renfrewshire.

## 9.1.7 South Lanarkshire

South Lanarkshire's physical topology was connected in a GC of 243.867 nodes, representing both building units and intersections, and 268.514 links, representing the street segments. The dual representation of this network was constructed of 268.514 nodes connected in GC with 695.790 links. The logical network was assembled from 113.975 nodes connected by 299.402 links and grouped into 7.858 communities. The largest community had 608 nodes. At the level of 0.749, the dendrogram was partitioned at a cut that produced 7.858 link communities. At the level of 1, the partition density had a value of 0.1204997. At the median hight of the network's link dendrogram cut of 0.7788500, the partition density value was 0.008777024 and at the mean height of 0.7509495, the partition density was 0.007889889.  The standard deviation for the height of the median and the mean values of the partition density were 0.1554476 and 0.003086961 respectively. For the optimal modularity value of 1, the dual graph of the physical network was divided in 546 communities.

The logical network's clustering coefficient was 0.01767537 and the one of the physical topology was 0.004930175, while the transitivity of its dual graph was 0.7863492. The assortativity degree of the logical network was -0.3760491, which indicates a relatively disassortative network. The physical topology was assortative with a high value of  0.5229201 and its dual representation had assortativity degree of 0.8734513.

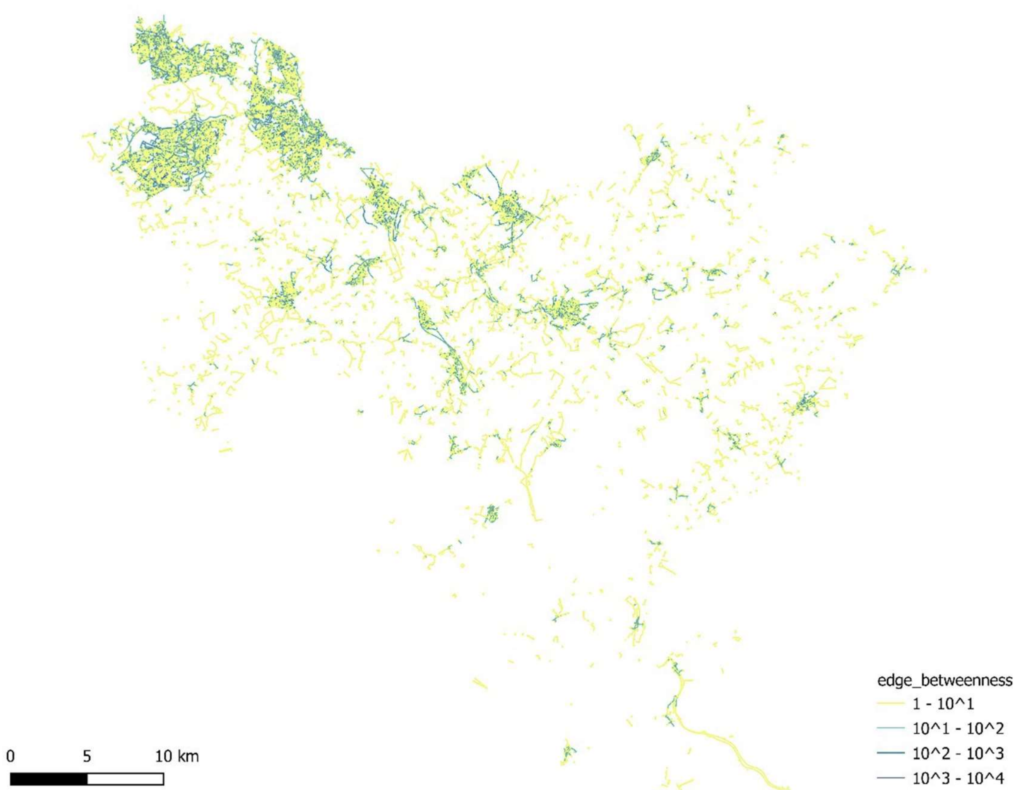



**Figure 28: Link communities of the logical topology of South Lanarkshire.**

## *9.1.8 West Dunbartonshire*

The last network for which the physical and logical topologies were analysed, was the network in West Dunbartonshire. The logical network of this shire had 26.731 nodes connected to each other by 72.590 links, in 3.772 communities of which the largest one had 227 nodes. The physical topology was mapped in a graph of 57.961 links, connecting 55.306 nodes into a Giant Component. The dual representation of the physical topology had 57.961 nodes connected with each other by 96.647 links.

At the level of 0.57 of the link dendrogram of the logical topology, the number of communities that were detected in the logical network was maximised. At the level of 1 of the same dendrogram, the partition density was 0.04482061. At the mean height of 0.6756 it was 0.02755 and at the median height of 0.7, the partition density was measured at 0.03109. The greedy multi-level modularity algorithm divided the nodes into 369 communities at the optimised value for modularity of 0.97. The optimised modularity of the dual representation of the physical network at the value of 0.99 grouped the nodes in 195 communities.

The clustering coefficient of the logical topology was 0.04877561 and the one of the physical topology was 0.006145133. The transitivity of the physical network's dual graph was 0.4808543. Both topologies of the graphs were disassortative with a coefficient of -0.370036 for the logical and -0.1522869 for the infrastructure network. Contrary to this, the dual representation of the graph for the infrastructure network was assortative, with the value for the nodes' degrees correlation coefficient (i.e. the assortativity degree) of 0.4287698.

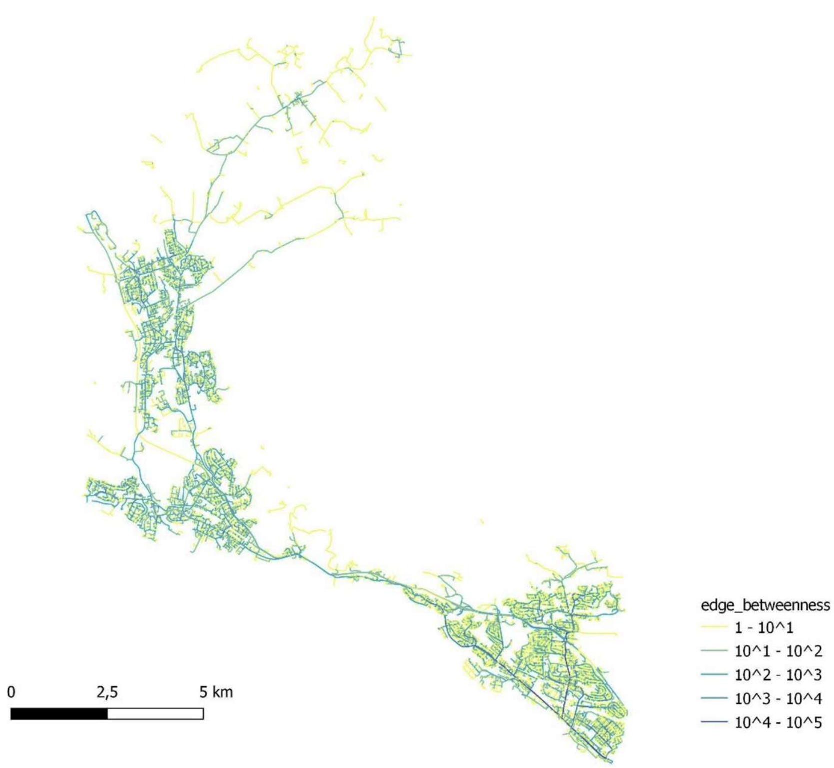



Figure 29: Link communities of the logical topology of West Dunbartonshire.

## 9.2 DISCUSSION ON THE FINDINGS

The results presented in this thesis suggest that there is a great potential for some methods that emerged in network science to be applied to spatial analyses. These methods specifically applied in complex networks studies are not inherently spatial, although the phenomenon they study may be. For the field of urban studies and modelling complexity of human behaviour in cellular space, as shown in this thesis, we can enrich these models by incorporating the primal and dual space syntax together with the spatial constraints as nodes or links attributes. Modelling a fragment of the many different complexities that emerge at the large regional scale, showed the benefits of using complex networks models for holistic sustainability and equity analyses. Namely, we can take away many conclusions from the results presented in this Chapter.

First, the results from this study showed that hubs in complex urban systems are not necessarily as important as nodes with high community centrality. For the network of the city of Glasgow the linear regression model showed even negative correlation between the community-based centrality and the node's degree (Table 5). This meant that the higher the node's degree, the smaller the node's importance within the community. If we were to confirm the importance of hubs as observed in social networks, we would have seen near-perfect correlation between the node's degree and the node's community-based centrality. This can be used as an argument to conduct more in depth analyses of urban projects that have the goal to promote equity and improve the quality of the built environment by introducing infrastructure synergies. Such analysis can address the failure of new infrastructure projects that tend to usually focus on connecting places to hubs to prioritise connecting places to communal central nodes.

Furthermore, in the analysed graphs we can see that in all of the eight regression models (Table 5) between community centrality and node betweenness centrality, strong lack of correlation between these two measures. The trivial correlation of betweenness centrality, which measures the number of shortest paths in the graphs going through the graph's nodes, shows that these paths do not pass through the nodes which are of central importance to the communities, i.e. the ones with high community-based centrality. In the case of Glasgow City, these measures were even negatively correlated.

Just as a reminder, as presented in 5.1.1, *betweenness centrality* of a node measures the number of paths passing through the given node that connect pairs of nodes and this is important because increasing this metric reduces the distances in the network. The nodes with very high betweenness centrality are also the hubs in networks. The correlation coefficients of the regression models presented in Table 5 showed no correlation between the hubs in the communities and their community centrality values. Thus the results in all eight regression models confirmed the research hypothesis that structurally important hubs of the physical network do not share the same importance in the relational topology of its functional counterpart.

Second, the regression models of the community centrality to the Burt's constraint in the graphs were negatively correlated. The theoretical underpinning of structural holes and weak ties, is similar. The weak tie hypothesis says that there is a chance greater than a probability that person B and C are connected, if person A is connected to both person B and C. The weak ties in social theory are the ones which contribute to the flow of more novel information. Subsequently, communities of nodes that have many weak ties tend to be overwhelmed with novel information supressing the role of strong ties in these communities. Thus, structural holes measure this 'structural gap' between the individuals which are not connected but have a complementary source of information (i.e. person A). Essentially, the hypothesis of structural holes explains the position of nodes in the network and the way they benefit from their position in the network as a potential source of complementary information. Both measures stem from social network research, and are not directly relevant for the networked infrastructure. However, the inverse measure of structural holes is the one that holds specific interest for this thesis.

Specifically, the *constraint* measure calculates the reverse of the structural holes (Everett and Borgatti, 2020). The constraint measures the lack of a gap between two nodes that hold complementary resources in a network. This measure is more reliable than structural holes since it takes the hierarchy in the network into consideration. Therefore the constraint and not the structural holes as a measure is more important for this study. When networked infrastructure are improved to encourage infrastructure synergies and equity in social accessibility, the goal of the network design is to reduce the gaps in the structure by increasing the number of nodes that have low constraint. I.e. finding the nodes with largest constraint and reducing their constraint by introducing new links to nodes that hold complementary resources to the ones whose constraint is being reduced.

The value of the constraint for a node is higher if the central node (i.e. the largest hub or the ego) has less (or more but redundant) nodes. For small egocentric (i.e. highly centralised) networks, the value of the constraint can be larger than one and for very large egocentric networks it can never reach the value of one. This measure captures the accessibility of an individual node in the network to many nodes that are not redundant. To maximise flow capital, a node in the network would typically connect to as many nodes as possible that don't have a connection with each other. The bigger the constraint value of a node is, the fewer the structural holes are, and the lower the 'node's capital' is (i.e. the node is not benefiting from novel sources) (op cit.). Thus, I expected that the 'boundary' [143] nodes which do not necessarily act as hubs in the community (and have more outwards than inwards community links) to be the ones that also have low constraint score. The results confirmed this expectation.

The nodes that had very high community-based centrality were also the ones whose constraint score approached the value of 0 and they were the ones that enjoyed the highest benefits in their community from novel connections. This was prevalent in all eight regression analyses which summary results are presented in the Appendix. All models displayed long tailed power-law distributions which meant that the constraint follows exponential decline with the increase of the node's community-based centrality and after a certain value for the community centrality the value for the constraint remains the same. There was also significant negative correlation (see Table 5) between the node's community based centrality measure and the Burt's constraint of their hubs. The strength of the Pearson's coefficient was the highest for East Renfrewshire and West Dunbartonshire, and the lowest (i.e. approaching a negligible value) for the communities in Glasgow. Inverclyde and Renfrewshire showed very high negative correlation while the remaining three council areas had moderate negative correlation.

*Closeness centrality* was the last measure interesting for this study. The results of closeness centrality in the eight council areas calculated how easy it is for nodes to be reached from the node being analysed and the opposite, how easy it is from other nodes to reach the node in question. Thus I expected that nodes with high community centrality will also be the ones with very high closeness value, however the results for all eight regression models showed no correlation between the two. Instead, the regression line showed perfectly horizontal trend for any of the communities analysed across all eight networks, regardless of their size or community centrality values.

There are two other results that resulted from this study worth elaborating on. The *community membership matrices* (see pp 126-129) present the count of the top 20 nodes that appear most frequently in overlaps in the 20 largest link communities. These matrices show the size of the overlap the communities have, which essentially is measured by their number of shared nodes (i.e. same coloured grids). The shared nodes are the ones that increase the network's modularity. The community membership matrix for East Dunbartonshire was comprised of 20 nodes which did not appear in any of the 20 communities more than one time. Same was true for East Renfrewshire, Renfrewshire and Inverclyde. The communities of South Lanarkshire were the ones that shared eight

[143] i.e. nodes that are not central to the communities, and with that are not the community hubs.

of their top 20 nodes, followed by the ones in North Lanarkshire, Glasgow and West Dunbartonshire, which shared seven, four and one node respectively. These results show that the existing modularity (or the lack there of) in the respective networks is very low to non-existent.

This result can partially be biased from the choice criteria of how relational (i.e. functional) links were defined in this study. In the Introduction section of this chapter I already presented that in all eight council areas approximately half of the households were more than 400 meters away from any destination point analysed. In other words, not all of the household units and destination locations are part of the communities discovered with the hierarchical link clustering algorithm. Many of them did not fulfil the cut-off criteria of 400 meters to be part of the OD accessibility matrices. Nevertheless, the opportunity to increase the network's modularity by introducing new nodes that will have high community centrality in the network will significantly increase the community membership participation of nodes and reduce the height of the optimal cut of the link communities' dendrogram.

The second measure worth mentioning was the *community connectedness.* This measure calculates how connected a community is towards other communities in the network. I.e. whether the community is bridging several other communities or it is inwardly connected within itself and relatively isolated from the rest of the network. This measure is the reverse of the modularity measure and nodes that belong to more highly connected communities scored higher on their community centrality. In the Appendix we can see that only in few communities the connectedness spiked with over 100 connections and the remining communities were divided; half of them were predominantly inwardly connected and the other half were more outwardly connected.

Knowing the community index makes this community discovery process very transparent. Since the community indices are known, we can find exactly which communities are inwardly and focus our future design efforts on improving their connectivity with the remining communities.

# Chapter 10: Conclusion

## 10.1 INTRODUCTION

Link communities of urban systems represent the functional units that support people to go about their daily lives. From the literature review presented in Chapters 3 and 4 I have concluded that spatially these functional unions are constrained by the urban form, the land use and the networked nature of infrastructure. Using metropolitan Glasgow (Figure 10) as the study region, this research modelled the social accessibility of over half a million, both single and multi-family residential buildings, to various categories of points of interests, important buildings, commercial locations and outdoor recreational facilities. The result of the model were dendrograms of the overlapping link communities of the social infrastructure. As explained in the earlier paragraph in Chapter 9, to measure accessibility as a social indicator I was only interested in the physical network-based distance between these locations.

In summary of the research presented in this thesis, link communities have a great potential for spatial analyses, offering new scale to study the built environment. They overcome some of the general shortcomings like resolution limitations (Kevin and Krizek, 2003; Strominger et al., 2016), associated with the more traditional ad-hoc area-based administrative units of analyses.

Research reviewed in Chapters 5 and 6 has shown that systems which are more modular are also the ones that have highly overlapping *communities,* lower network diameters and are more resilient. These overlaps emerge in systems which have highly co-dependent functional components connected with each other by *functional links*. The literature review revealed scarce evidence of studying complex urban systems from this aspect.

The goal of this modelling project was to study the overall connectivity of the Glasgow's metropolitan region, focused on equity of social accessibility of usual places of residence to various destinations of interest. This was motivated by two factors. First there was extensive study on the region's socio-economic inequality tributed to the spatial inequality of the urban form. This was a direct corollary from the not very prosperous urban planning decisions made in Glasgow's post-industrial planning period. The standing research advised on future extensive research to consider the historic and socio-economic context of the built environment.

Second, large body of knowledge that studies the sustainability of the urban form has demonstrated that multiple urban forms can be sustainable. However, most of the research is fragmented, focusing either on studying the relationship between morphology and travel behaviour or morphology and social-dynamics or urban metabolism. These schools of thought share a common conclusion that morphology is ultimately determined by the ever-changing interactions between the residents, the economy and the institutions. The goal of this thesis was to identify and quantify the 'ever-changing interaction' and characterise their structure, as well as to offer an overarching modelling framework that will supersede the fragmentation of the modelling approaches listed above.

Modelling with an overarching framework, as shown throughout this thesis, improves our understanding of the spatially bounded physical, economic, social and geological processes that shape our complex world. Through analytical lenses, this overarching method ultimately offers structural approach to improving these processes' sustainability.

## 10.2 CONCLUSION FROM THE RESEARCH FINDINGS

First, the results from this study showed that hubs in complex urban systems are not necessarily as important as nodes with high community centrality. The summary results from the regression models presented in Table 5 and the Appendix showed no correlation between the hubs in the communities and their community centrality values, confirming the research hypothesis.

Second, none of the eight logical topologies in the council areas showed high clustering coefficient and shared very few nodes across their communities. These networks had typical degree distributions as also reported in many of the studies reviewed in Chapters 4 and 5. In alignment with the results in this thesis, Boeing's (2017, p159) work presented very low degree of the American street network (app. 2.8 streets per). While the values of the total network length was extremely high. In these networks, named by Kaiser and Hilgetag (2004) as *linear scale-free networks,* hubs were not usually observed and therefore they are almost linear.

The degree distribution of the physical graphs and the line-graphs generated in my thesis (see Appendix), as expected were built out of nodes with relatively low degrees. Contrary to the degree distribution of these graphs (which was characterised by very small power-law[144]), the degree distribution of the eight logical graphs did follow a well-defined power-law distribution. The presence of hubs and the scale-free property in the logical topology of the networks studied in this thesis was in alignment with previous findings that concluded this type of distribution to be typically observed in social and economic networks (Snijders, 2001; König et al., 2007; König and Battiston, 2009). The most connected hub, i.e. the node with the highest degree of app. 1300, was found in the network of East Dunbartonshire, the second largest was the one of North Lanarkshire with the largest hub had a degree of 550. The remaining logical topologies had degree distribution where the degree of the hub with the highest rate of connections ranged between the values of 350 and 450. In summary of the

[144] As expected based on Kaiser and Hilgetag (2004) findings, see discussion following in the next paragraphs.

results, all eight networks did behave in accordance to the properties listed in p102 (this thesis) that are a typical representation of spatially bounded networks.

Stemming from the discourse in Chapter 3, the low value of the network's degree can be attributed to the restrictive and 'grid-like' controlled planning schemes imposed to Western cities (Major, 2018). Therefore, the average streets per node in cities that grew more organically (Alexander, 1965) could be very different. Since I have observed average degree in the physical graphs I can conclude that planned street networks have a scale — i.e. an average degree — contrary to street networks of ancient cities that evolved through organic movements. According to the extensive street design analyses performed by Major's work (2018) these organically created street networks are scale-free.

Kaiser and Hilgetag (2004) further discovered that although networks with a geospatial dimension are scale-free, they are not small-worlds. Their average shortest path (same as the shortest step-length) is two times larger than the randomly generated model counterpart. Consequently, these networks' diameters are very large. The clustering coefficient (neighbours of a node that are neighbours themselves) of these networks highly depended on the physical limits of growth. High clustering coefficient is a signature of high network's modularity (a network with community structure). When growth was spatially bounded, the density of the network remained constant. The clustering coefficient of the eight graphs in this thesis was relatively low and they had large network diameters.

From these findings I concluded that the networked infrastructure of metropolitan Glasgow holds great potential to improve its overall social accessibility, sustainability and resilience by increasing its clustering coefficient. The resilience of the complex system could be improved by increasing both the number of nodes with high community centrality and links with high betweenness.

## 10.3 FUTURE RESEARCH AVENUES

There are two main avenues for future research. First I didn't explore in depth the cut-sets of the overlapping nodes. They can be potentially important when deciding where and with what type of urban design solutions we address networks' sustainability and improve their social access. For example, structural analyses of overlaps spatially can identify communities of links which can benefit greatly of increased number of links with high edge betweenness, or increased number of nodes with large value for community-based centrality. Burt's constraint is relevant measure in the networked infrastructure analysed, and we should test several design solutions for new overlaps that reduce the value of this measure. Since, as already discussed in the Findings, the lower this constraint is, the more the nodes in the community enjoys from novel connections.

Second, Borgatti's work (2005) highlights the importance of defining precisely the nature of the type of 'flow' we are trying to assume based on the way it moves through the network. This is in fact determined by the nature of the flow itself. Namely, he distinguishes among three types *didactic diffusion, replication-based flows* and *non-deterministic traffic flows*. His contribution relevant to this research is to be mindful about the type of flows under study. The process' typology determines the relevance of the node's centrality measures, which means that centrality is in fact a node-level product of implicit flows modelling, therefore is an artifact of the input criteria that predefine certain participation of a node in the network under study.

Thus, movement of people (commute and other forms of travel—i.e. traffic) belongs under the non-deterministic traffic flows, while the movement of used goods is categorised as *trails* (not passing the same link it went through again). Money are moving by *walks* (and is modelled by Markov process). Information is best modelled as *diffusion by replication* and in the same time exists at multiple locations. The last one that he studied and is of interest for urban networked infrastructure is the movement of a package that travels by the *shortest path* in the geodesic network-based distance from its origin towards its destination. The flows like energy, water, food and waste (incl. solid and water) were not studied in his work. Mapping in future the logical part of the network with these processes before we build the logical topology graph will give insights on capacity constraints as well, which will make the model in better alignment with reality.

Since the betweenness centrality of nodes was one of the response variables in the statistical models presented in this thesis, future research should explore flows clearly defining them, bearing in mind Borgatti's (op cit.) classifications. The research presented in this thesis generalised the notion of a flow, defined only by the accessibility (proximity) of locations. In future research the methodology and results presented here can be extended to include specific flow types and capacity constraints accordingly. Thus, according to the classifications of the flow(s) studied by Borgatti, for the resources conveyed by the networked infrastructure can already be assumed that in centralised networks they move like money— since utilities directly depend on purchasing ability. However, in the case of decentralised networks, these flows (i.e. everyone can self-generate any type of resource) would probably resemble a non-deterministic process. This is left to be explored in future research.

# Appendix

| Residuals | Min | 1Q | Median | 3Q | Max |
|---|---|---|---|---|---|
| East Dunbartonshire | -122.56 | -12.54 | -3.73 | 5.73 | 1429.36 |
| East Renfrewshire | -5.50 | -3.68 | -2.66 | -1.55 | 452.38 |
| Glasgow City | -11.01 | -4.89 | -2.81 | -0.58 | 494.18 |
| Inverclyde | -29.939 | -5.538 | -1.281 | 3.990 | 304.281 |
| North Lanarkshire | -7.88 | -4.59 | -2.95 | -1.27 | 528.44 |
| Renfrewshire | -29.45 | -4.91 | -0.85 | 4.02 | 326.50 |
| South Lanarkshire | -3.87 | -3.36 | -2.57 | -1.59 | 432.13 |
| West Dunbartonshire | -19.272 | -4.095 | -1.343 | 2.130 | 297.364 |

Table 7: Residuals of the lm between node 'link-community based' centrality and node degree

| Residuals | Min | 1Q | Median | 3Q | Max |
|---|---|---|---|---|---|
| East Dunbartonshire | -1556729 | -128903 | -47797 | 40726 | 50766142 |
| East Renfrewshire | -3364022 | -460414 | -86403 | 37872 | 120306156 |
| Glasgow City | -6.506e+25 | -6.414e+25 | -6.248e+25 | -6.172e+25 | 9.121e+28 |
| Inverclyde | -2510714 | -256051 | -78823 | 138776 | 53721635 |
| North Lanarkshire | 1.638e+29 | -1.451e+29 | -1.413e+29 | -1.329e+29 | 1.512e+32 |
| Renfrewshire | -3.834e+19 | -2.472e+18 | -1.075e+18 | 1.084e+18 | 1.506e+21 |
| South Lanarkshire | -5.325e+07 | -2.184e+07 | -1.765e+07 | -7.536e+06 | 3.207e+10 |
| West Dunbartonshire | -247638 | -43224 | -13070 | 20421 | 9592527 |

Table 8: Residuals of the lm between node 'link-community based' centrality and node betweenness centrality

| Residuals | Min | 1Q | Median | 3Q | Max |
|---|---|---|---|---|---|
| East Dunbartonshire | -7.385e-08 | 1.730e-10 | 6.230e-10 | 1.068e-09 | 3.004e-09 |
| East Renfrewshire | -1.559e-09 | -1.179e-09 | 5.969e-10 | 6.252e-10 | 6.505e-10 |
| Glasgow City | -1.022e-08 | -1.520e-11 | 6.380e-11 | 1.259e-10 | 2.846e-10 |
| Inverclyde | -3.737e-09 | -1.751e-09 | 1.100e-09 | 1.338e-09 | 1.515e-09 |
| North Lanarkshire | -1.479e-10 | -9.316e-11 | 3.844e-11 | 4.610e-11 | 6.825e-11 |
| Renfrewshire | -2.284e-10 | -1.324e-10 | -7.744e-11 | 1.333e-10 | 1.417e-10 |
| South Lanarkshire | -2.644e-11 | -1.203e-11 | 2.734e-12 | 9.693e-12 | 1.599e-11 |
| West Dunbartonshire | -7.995e-10 | -1.543e-10 | -7.826e-11 | 1.770e-10 | 3.393e-10 |

Table 9: Residuals of the lm between node 'link-community based' centrality and closeness centrality

| Residuals | Min | 1Q | Median | 3Q | Max |
|---|---|---|---|---|---|
| East Dunbartonshire | -0.26671 | -0.04205 | -0.01043 | 0.02362 | 0.85620 |
| East Renfrewshire | -0.66823 | -0.11817 | -0.01901 | 0.07234 | 1.75599 |
| Glasgow City | -0.34260 | -0.08005 | -0.01369 | 0.05057 | 0.77798 |
| Inverclyde | -0.48803 | -0.10705 | -0.03055 | 0.01924 | 1.31837 |
| North Lanarkshire | -0.44835 | -0.09186 | -0.00471 | 0.06293 | 0.67416 |
| Renfrewshire | -0.54830 | -0.10293 | 0.05765 | 0.00643 | 2.83991 |
| South Lanarkshire | -0.60759 | -0.11264 | -0.00711 | -0.10766 | -0.51236 |
| West Dunbartonshire | -0.71542 | -0.12469 | -0.01977 | 0.14866 | 1.53882 |

Table 10: Residuals of the lm between node 'link-community based' centrality and Burt's constraint

| Coefficients | | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|---|
| East Dunbartonshire | (Intercept) | -36.2084 | 0.4006 | -90.4 | <2e-16 *** |
| | Community cent. | 11.0716 | 0.0847 | 130.7 | <2e-16 *** |
| East Renfrewshire | (Intercept) | 5.54784 | 0.13914 | 39.872 | <2e-16 *** |
| | Community cent. | 0.07177 | 0.06047 | 1.187 | 0.235 |
| Glasgow City | (Intercept) | 12.01402 | 0.11415 | 105.24 | <2e-16 *** |
| | Community cent. | -1.65337 | 0.03679 | -44.94 | <2e-16 *** |
| Inverclyde | (Intercept) | -1.99001 | 0.13523 | -14.71 | <2e-16 *** |
| | Community cent. | 4.94048 | 0.05294 | 93.32 | <2e-16 *** |
| North Lanarkshire | (Intercept) | 8.87637 | 0.11720 | 75.73 | <2e-16 *** |
| | Community cent. | -0.77398 | 0.03843 | -20.14 | <2e-16 *** |
| Renfrewshire | (Intercept) | -2.02297 | 0.06555 | -30.86 | <2e-16 *** |
| | Community cent. | 5.87008 | 0.03147 | 186.53 | <2e-16 *** |
| South Lanarkshire | (Intercept) | 4.57494 | 0.09085 | 50.359 | <2e-16 *** |
| | Community cent. | 0.29094 | 0.03420 | 8.507 | <2e-16 *** |
| West Dunbartonshire | (Intercept) | -1.13022 | 0.10943 | -10.33 | <2e-16 *** |
| | Community cent. | 4.01386 | 0.05067 | 79.22 | <2e-16 *** |

Table 11: Coefficients-- Estimate for the lm between node 'link-community based' centrality and node degree

| Coefficients | | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|---|
| East Dunbartonshire | (Intercept) | -361718 | 8293 | -43.62 | <2e-16 *** |
| | Community cent. | 101132 | 1754 | 57.67 | <2e-16 *** |
| East Renfrewshire | (Intercept) | -322090 | 22560 | -14.28 | <2e-16 *** |
| | Community cent. | 307151 | 9804 | 31.33 | <2e-16 *** |
| Glasgow City | (Intercept) | 6.506e+25 | 1.095e+25 | 5.942 | 2.83e-09 *** |
| | Community cent. | -8.944e+23 | 3.529e+24 | -0.253 | 0.8 |
| Inverclyde | (Intercept) | -138776 | 13938 | -9.957 | <2e-16 *** |
| | Community cent. | 152949 | 5456 | 28.032 | <2e-16 *** |
| North Lanarkshire | (Intercept) | 1.283e+29 | 2.122e+28 | 6.045 | 1.5e-09 *** |
| | Community cent. | 4.606e+27 | 6.958e+27 | 0.662 | 0.508 |
| Renfrewshire | (Intercept) | -1.084e+18 | 1.471e+17 | -7.372 | 1.71e-13 *** |
| | Community cent. | 1.778e+18 | 7.059e+16 | 25.188 | < 2e-16 *** |
| South Lanarkshire | (Intercept) | 1368298 | 3266304 | 0.419 | 0.675 |
| | Community cent. | 6167915 | 1229676 | 5.016 | 5.29e-07 *** |
| West Dunbartonshire | (Intercept) | -20421 | 2336 | -8.742 | <2e-16 *** |
| | Community cent. | 27401 | 1082 | 25.336 | <2e-16 *** |

Table 12: Coefficients-- Estimate for the lm between node 'link-community based' centrality and node betweenness centrality

| Coefficients | | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|---|
| East Dunbartonshire | (Intercept) | 7.286e-08 | 8.097e-11 | 899.83 | <2e-16 *** |
| | Community cent. | 5.507e-10 | 1.712e-11 | 32.17 | <2e-16 *** |
| East Renfrewshire | (Intercept) | 2.420e-09 | 7.993e-12 | 302.783 | <2e-16 *** |
| | Community cent. | 1.965e-11 | 3.474e-12 | 5.658 | 1.54e-08 *** |
| Glasgow City | (Intercept) | 1.010e-08 | 4.873e-12 | 2072.32 | <2e-16 *** |
| | Community cent. | 3.146e-11 | 1.570e-12 | 20.04 | <2e-16 *** |
| Inverclyde | (Intercept) | 4.097e-09 | 1.546e-11 | 265.04 | <2e-16 *** |
| | Community cent. | 1.118e-10 | 6.051e-12 | 18.47 | <2e-16 *** |
| North Lanarkshire | (Intercept) | 1.567e-10 | 4.538e-13 | 345.20 | <2e-16 *** |
| | Community cent. | 8.119e-12 | 1.488e-13 | 54.57 | <2e-16 *** |
| Renfrewshire | (Intercept) | 4.420e-10 | 7.077e-13 | 624.57 | <2e-16 *** |
| | Community cent. | 5.960e-12 | 3.397e-13 | 17.54 | <2e-16 *** |
| South Lanarkshire | (Intercept) | 8.902e-11 | 6.614e-14 | 1345.89 | <2e-16 *** |
| | Community cent. | 1.725e-12 | 2.490e-14 | 69.28 | <2e-16 *** |
| West Dunbartonshire | (Intercept) | 1.525e-09 | 1.711e-12 | 891.09 | <2e-16 *** |
| | Community cent. | 5.012e-11 | 7.923e-13 | 63.27 | <2e-16 *** |

Table 13 : Coefficients-- Estimate for the lm between node ‘link-community based’ centrality and closeness centrality

| Coefficients | | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|---|
| East Dunbartonshire | (Intercept) | 0.2944948 | 0.0011181 | 263.4 | <2e-16 *** |
| | Community cent. | -0.0256917 | 0.0002364 | -108.7 | <2e-16 *** |
| East Renfrewshire | (Intercept) | 0.6890654 | 0.0016654 | 413.8 | <2e-16 *** |
| | Community cent. | -0.1395492 | 0.0007238 | -192.8 | <2e-16 *** |
| Glasgow City | (Intercept) | 0.3806404 | 0.0010312 | 369.1 | <2e-16 *** |
| | Community cent. | -0.0336214 | 0.0003323 | -101.2 | <2e-16 *** |
| Inverclyde | (Intercept) | 0.5041577 | 0.0019423 | 259.6 | <2e-16 *** |
| | Community cent. | -0.0874244 | 0.0007604 | -115.0 | <2e-16 *** |
| North Lanarkshire | (Intercept) | 0.5187337 | 0.0012070 | 429.8 | <2e-16 *** |
| | Community cent. | -0.0678930 | 0.0003957 | -171.6 | <2e-16 *** |
| Renfrewshire | (Intercept) | 0.5576506 | 0.0011556 | 482.5 | <2e-16 *** |
| | Community cent. | -0.1048832 | 0.0005548 | -189.1 | <2e-16 *** |
| South Lanarkshire | (Intercept) | 0.7364242 | 0.0013805 | 533.4 | <2e-16 *** |
| | Community cent. | -0.1237857 | 0.0005197 | -238.2 | <2e-16 *** |
| West Dunbartonshire | (Intercept) | 0.7291229 | 0.0018408 | 396.1 | <2e-16 *** |
| | Community cent. | -0.1575414 | 0.0008523 | -184.8 | <2e-16 *** |

Table 14: Coefficients-- Estimate for the lm between node ‘link-community based’ centrality and Burt’s constraint

# Bibliography

[Github user] crocovert (2021). *Networks plugin for QGIS 3*. Originally written by Patrick Palmier. Version 2.6.5. URL: https://github.com/crocovert/networks/ [last accessed: 2021/05/17]

[lecture notes]. *Networks.* Available from: http://grail.cba.csuohio.edu/~sanchita/network.ppt. 15 (1), pp.1-13.

Ahlfedlt, G., and Pietrostefani, E., (2017). *Demystifying Compact Urban Growth: Evidence From 300 Studies From Across the World.* Coalition for Urban Transitions, London and Washington, DC. Available at: http://newclimateeconomy.net/content/cities- working-papers.

Ahn, Y. Y., Bagrow, J. P., & Lehmann, S. (2010a). Link communities reveal multiscale complexity in networks. *Nature*, 466(7307), 761–764. https://doi.org/10.1038/nature09182

Ahn, Y. Y., Bagrow, J. P., & Lehmann, S. (2010b). Link communities reveal multiscale complexity in networks_Supplementary Information. *Nature*, 466(7307), 761–764. https://doi.org/10.1038/nature09182

Ahvenniemi, H. et al. (2017). What are the differences between sustainable and smart cities? *Cities*, 60 234–245.

Albert, R. and Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74 (1), 47–97. Available from http://dx.doi.org/10.1103/RevModPhys.74.47%5Cnpapers2://publication/doi/10.1103/RevModPhys.74.47%5Cnhttp://stacks.iop.org/1478-3975/1/i=3/a=006?key=crossref.7e041937ef77358d60afe44f40925c5b%5Cnhttp://link.aps.org/doi/10.1103/RevModPhys.74.47.

Albert, R., Jeong, H. and Barabási, A.-L. (2000). Error and attack tolerance of complex networks. *Nature*, 406 (27), 378–381. Available from http://www.nature.com/doifinder/10.1038/35054111.

Alberti, M. (2016). *Cities that think like planets: Complexity, Resilience, and innovation in hybrid ecosystems*. Seattle and London: University of Washington Press.

Alexander, C. (1965). *A City is Not a Tree*. Architectural Forum, 122, 1, pp 58-62.

Allan, J. (2005). Virtual Water: A Strategic Resource Global Solutions to Regional Deficits. *Ground Water*. 36. 545 - 546.

Allan, J.A. (1996). Policy responses to the closure of water resources: regional and global issues. In: Howsam, P. and Carter, R.C. (eds). *Water policy: allocation and management in practice.* Proceedings of international conference, Cranfield University, UK, 23-24 September, 1996: 3-12.

Allen, P.M. (2001). Models, knowledge creation and their limits. *Paper presented at the NEXSUS Workshop on The Limits to Knowledge*, Central European University, Budapest, June 23–24.

Alonso, W. (1964). *Location and land use.* Harvard University Press, Boston.

Amaral, L. A. N., Scala, A., Barthélemy, M. and Stanley, H. E. (2000). *Proceedings of the National Academy of Sciences of the United States of America,* 97, 11149.

An, L. et al. (2014). Agent-based modeling in coupled human and natural systems (CHANS): Lessons from a comparative analysis. *Annals of the Association of American Geographers*, 104 (4), 723–745.

An, L., Linderman, A., Qi, J.,Shortridge, A. and Liu, J. (2005). Exploring complexity in a human environment system: An agent-based spatial model for multidisciplinary and multiscale integration. *Annals of the Association of American Geographers,* 95 (1), 54-79.

Anderson, P.W. (1972). *More Is Different*. Science, 177 (4042), 393–396.

Andersson, C., Frenken, K. and Hellervik, A. (2006). A complex network approach to urban growth. *Environment and Planning A*, 38 (10), 1941–1964.

Arcaute, E., Molinero, C., Hatna, E., Murcio, R., Vargas-Ruiz, C., Masucci, A. P., & Batty, M. (2016). Cities and regions in Britain through hierarchical percolation. *Royal Society Open Science*, 3(4). https://doi.org/10.1098/rsos.150691

Bak, P. (1996). *How Nature Works.* New York: Springer.

Bak, P. and Chen, K. (1989).The physics of fractals. *Physica D,* 38, 5-12.

Bak, P., Chen, K. and Creutz, M. (1989). Self-organized criticality in the 'game of life'. *Nature,* 342, 780-782.

Bak, P., Tang, C. and Wiesenfeld, K. (1987). Self-Organized Criticality: An Explanation of 1/f Noise. *Physical Review Letters*, 59 (4), 381–384.

Barabási, A.-L. (2016). *Network science.* 3rd ed. Cambridge: Cambridge University Press.

Barabási, A.L., Albert, R. and Jeong, H. (1999). Mean-field theory for scale-free random networks. *Physica A: Statistical Mechanics and its Applications*, 272 (1), 173–187.

Barabási, A.L., Albert, R. and Jeong, H. (2000). Scale-free characteristics of random networks: The topology of the world-wide web. *Physica A*: *Statistical Mechanics and its Applications*, 281 (1), 69–77.

Baran, P. (1964). *On distributed communications: Introduction to distributed communications networks*. 51. Santa Monica, California.

Barenblatt, G. (1996). *Scaling, Self-similarity, and Intermediate Asymptotics: Dimensional Analysis and Intermediate Asymptotics* (Cambridge Texts in Applied Mathematics). Cambridge: Cambridge University Press. doi:10.1017/CBO9781107050242

Barrat, A. and Weigt, M. (2000). On the properties of small-world network models. *Eur. Phys. J. B*, 94 85–94.

Barredo, J. I., & Gómez-Delgado, M. (2008). Towards a set of IPCC SRES urban land use sce- narios: Modelling urban land use in the Madrid region. In *Modelling Environmental Dynamics.* Paegelow, M. and Camacho Olmedo, M. T. (eds). Berlin Heidelberg: Springer. 363-385.

Barreira-González, P., Gómez-Delgado, M. and Aguilera-Benavente, F. (2015). From raster to vector cellular automata models: A new approach to simulate urban growth with the help of graph theory. *Computers, Environment and Urban Systems*, 54 . Elsevier Ltd119–131. Available from http://dx.doi.org/10.1016/j.compenvurbsys.2015.07.004.

Barthélemy, M. (2011). Spatial networks. *Physics Reports*, 499 (1–3). Elsevier B.V.1–101. Available from http://dx.doi.org/10.1016/j.physrep.2010.11.002.

Barthélemy, M. (2016). *The Structure and Dynamics of Cities: Urban Data Analysis and Theoretical Modeling*. Cambridge: Cambridge University Press. doi: 10.1017/9781316271377.

Barthélemy, M. et al. (2013). Self-organization versus top-down planning in the evolution of a city. *Scientific Reports*, 3 (1), 1–12.

Batty M. and Longley P. (1989). Urban growth and form: Scaling, fractal geometry, and diffusion-limited aggregation. *Environ. Plann. A* 21: 1447–1472.

Batty, M. (1983). Linear Urban Models. *Papers of the Regional Science Association,* Regional science Research Institute in cooperation with the Dept. of Regional Science, Wharton School, University of Pennsylvania, 53(1), 1-25.

Batty, M. (2005a). *Cities and complexity: understanding cities with ben, agent-based models, and fractals*. Cambridge, Mass., London: MIT. 565

Batty, M. (2005b). Urban Modelling. In: Kitchin, R., and Thrift, N., eds. *International Encyclopedia of Human Geography*., 1–18.

Batty, M. (2006). Rank clocks. *Nature*, 444 (7119), 592–596.

Batty, M. (2008). The Size and Shape of Cities. *Science*, 319 (5864), 769–771

Batty, M. (2009). Accessibility: In Search of a Unified Theory. *Environment and Planning B: Planning and Design,* 36 (2), 191–194.

Batty, M. (2013). *The New Science of Cities*. Cambridge, MA: MIT Press.

Batty, M. (2016). Classifying urban models. *Environment and Planning B: Planning and Design*, 43 (2), 251–256.

Batty, M. (2019). Urban analytics defined. *Environment and Planning B: Urban Analytics and City Science*, 46 (3), 403–405. Available from http://journals.sagepub.com/doi/10.1177/2399808319839494.

Batty, M. (2022). Integrating space syntax with spatial interaction. *Urban Informatics*, 1(4), 1–23. https://doi.org/10.1007/s44212-022-00004-2

Batty, M. and Longley, P. A. (1994). *Fractal Cities: A Geometry of Form and Function.* London: Academic Press.

Batty, M. and Torrens, P.M. (2005). Modelling and prediction in a complex world. *Futures*, 37 (7 SPEC.ISS.), 745–766.

Bavaud, F., Kordi, M. and Kaiser, C. (2018). Flow autocorrelation: a dyadic approach. *Annals of Regional Science*, 61 (1), 95-111. Springer Berlin Heidelberg95–111. Available from https://doi.org/10.1007/s00168-018-0860-y.

Benenson, I. and Torrens, P. M. (2004). *Geosimulation: Automata-based Modeling of Urban Phenomena.* First edn. Chichester: Wiley-Blackwell. 287

Berry, B.J.L. (1964). Cities as systems within systems of cities. *Papers of the Regional Science Association*, 13, 146–163. https://doi.org/10.1007/BF01942566.

Bettencourt, L.M.A. (2013). The origins of scaling in cities. *Science*, 340 (6139), 1438–1441.

Bettencourt, L.M.A. and Lobo, J. (2016). Urban scaling in Europe. *Journal of the Royal Society Interface*, 13 (116), .

Bettencourt, L.M.A. et al. (2007). Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences*, 104 (17), 7301–7306. Available from http://www.pnas.org/cgi/doi/10.1073/pnas.0610172104.

Bettencourt, L.M.A., Lobo, J. and Youn, H. (2013). The hypothesis of urban scaling: formalization, implications and challenges (March 2014), . Available from http://arxiv.org/abs/1301.5919.

Bianconi, G., Kryven, I. and Ziff, R. M. (2019). Percolation on branching simplicial and cell complexes and its relation to interdependent percolation. *Phys. Rev. E*, 100:062311

Blondel, V.D., Guillaume, J. L., Lambiotte, R. and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment 2008* (10), P10008.

Boccaletti, S. et al. (2014). The structure and dynamics of multilayer networks. *Physics Reports*, 544 (1), . Elsevier B.V.1–122. Available from http://arxiv.org/abs/1407.0742%0Ahttp://dx.doi.org/10.1016/j.physrep.2014.07.001.

Boeing, G. (2018). A Multi-Scale Analysis of 27,000 Urban Street Networks: Every US City, Town, Urbanized Area, and ZillowNeighborhood*. *Environment and Planning B: Urban Analytics and City Science*,18.

Boeing, Geoffrey D. (2017). *Methods and Measures for Analyzing Complex Street Networks and Urban Form* []. Thesis. University of California, Berkeley. Available from https://osf.io/preprints/socarxiv/93h82/.

Boerema, A. et al. (2017). Are ecosystem services adequately quantified? *Journal of Applied Ecology*, 54 (2), 358–370.

Bogart, D. (2009). Inter-modal network externalities and transport development: Evidence from roads, canals, and ports during the English industrial revolution. *Networks and Spatial Economics*, 9 (3), 309–338.

Bolstad, P. (2012). *GIS fundamentals: a first text on geographic information systems*. White Bear Lake, Minn, Eider Press.

Borgatti, S. P. (2005). Centrality and network flow. *Social Networks*, 27(1), 55–71. https://doi.org/10.1016/j.socnet.2004.11.008

Borgatti, S. P., Everett, M. G., & Shirey, P. R. (1990). LS sets, lambda sets and other cohesive subsets. *Social Networks,* 12(4), 337–357. https://doi.org/10.1016/0378-8733(90)90014-Z

Bosch, M., Chenal, J. and Joost, S. (2019). Addressing urban sprawl from the complexity sciences. MPRA Paper, 1–19.

Bourne, J. W. and Simmon, L. S. (1978). *Systems of Cities: Readings on Structure, Growth, and Policy* (0 ed.). New York: Oxford University Press.

Brandes, U. et al. (2008). On Modularity Clustering. *IEEE Transactions on Knowledge and Data Engineering*, 20 (2), 172–188.

Braveman, P. (2014). What is Health Equity: And How Does a Life-Course Approach Take Us Further Toward It?. *Matern Child Health J* 18, 366–372 https://doi.org/10.1007/s10995-013-1226-9

Breheny, M. (1997). Urban compaction: feasible and acceptable? *Cities*, 14 (4), 209–217.

Brelsford, C. et al. (2017). Heterogeneity and scale of sustainable development in cities. *Proceedings of the National Academy of Sciences*, 114 (34), 8963-8968. Available from http://www.pnas.org/lookup/doi/10.1073/pnas.1606033114.

Buhl, J. et al. (2004). Efficiency and robustness in ant networks of galleries. *European Physical Journal B,* 42 (1), 123–129.

Campbell, J. E. and Shin, M. (2012). *Essentials of Geographic Information Systems.* Sailor Academy. Available from: https://saylordotorg.github.io/text_essentials-of-geographic-information-systems/index.html Accessed on 03.06.2019

Castells Manuel (2000). *The rise of the network society*. 2nd ed. The information age, Economy, society and culture v. 1. Oxford: Blackwell.

Castells, M. (2010). Globalisation, networking, urbanisation: Reflections on the spatial dynamics of the information age. *Urban Studies*, 47 (13), 2737–2745.

Cecchini, A. and Rizzi, P. (2001). Is urban gaming simulation useful? *Simulation and Gaming*, 32 (4), 507–521.

Chen, F. et al. (2015). Data mining for the internet of things: Literature review and challenges []. *International Journal of Distributed Sensor Networks*.

Choay, F. (1996). *The rule and the model: on the theory of architecture and urbanism*. Cambridge, Mass ; London: MIT Press.

Cleveland State University-College of Business-DBA-CSI. (unknown). *Network Topology: Physical and Logical.*

Clifton, J., Díaz-Fuentes, D. and Fernández-Gutiérrez, M. (2016). Public Infrastructure Services in the European Union: Challenges for Territorial Cohesion. *Regional Studies*, 50 (2), . Taylor & Francis358–373.

Collins, J. P., Kinzig, A., Grimm, N. B., Fagan, W. F., Hope, D., Wu, J., Borer, E. T., Collins, J. P., Kinzig, A., Grimm, N. B., Fagan, W. R., Hope, D., Wu, J. and Borer, E. T. (2000). A New Urban Ecology: Modeling human communities as integral parts of ecosystems poses special problems for the development and testing of ecological theory. *American Scientist*, 88 (May), 416–425.

Costa, L. da F., Rodrigues, F. A., Travieso, G. and Villas Boas, P. R. (2005). Characterization of complex networks: A survey of measurements. Available from http://arxiv.org/abs/cond-mat/0505185%0Ahttp://dx.doi.org/10.1080/00018730601170527.

Couclelis, H. (1985). Cellular worlds: A framework for modeling micro–macro dynamics. *Environment and Planning A*, 17, 585-596.

Couclelis, H. (2005). Where has the future gone? Rethinking the role of integrated land-use models in spatial planning. *Environment and Planning A,* 37 (8), 1353–1371.

Coutard, O., Hanley, R. E. and Zimmerman, R. (2005). *Sustaining Urban Networks: the Social Diffusion of Large Technical Systems.* Hanley, R.E., Graham, S., and Marvin, S., eds. *The Networked Cities Series*.

First Edn. Abingdon: Routledge. Available from http://www.tandfebooks.com/isbn/9780203357118%5Cnhttp://samples.sainsburysebooks.co.uk/9781134345700_sample_502037.pdf.

Cressie, N. (1993). *Statistics for Spatial Data.* New York: Wiley.

Crucitti, P., Latora, V. and Porta, S. (2006). Centrality measures in spatial networks of urban streets. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 73 (3), 1–5.

Curits, A. and Fotheringham, A. S. (1995). Large Scale Information Surfaces: An Analysis of City Name Recalls in the United State, *Geoforum,* 26, 65-78.

De Nadai, M. et al. (2016). The Death and Life of Great Italian Cities: A Mobile Phone Data Perspective. In: *The 26th International ACM Conference on World Wide Web (WWW)*.

Dearden, J., Jones, M.W. and Wilson, A. (2015). DynaMoVis: visualization of dynamic models for urban modeling. *Visual Computer*, 31 (6–8), . Springer Berlin Heidelberg1079–1088.

Del Genio, C.I., Gross, T. and Bassler, K.E. (2011). All scale-free networks are sparse. *Physical Review Letters*, 107 (17), 1–4.

Dendrinos, D.S. and Rosser, J.B.J. (1992). Fundamental issues in nonlinear urban population dynamic models. *Regional Science*, 26 135–145.

Densham, Paul J., (1999). Spatial decision support systems. In: Maguire, D.J., Goodchild M. F. and Rhind, D. W. (eds.) *Geographical Information systems: principles and applications*. London: Longman, 403-412.

Derrible, S. (2017). Urban infrastructure is not a tree: Integrating and decentralizing urban infrastructure systems. *Environment and Planning B: Urban Analytics and City Science*, 44 (3), 553–569.

Dessemontet, P., Kaufmann, V. and Jemelin, C. (2010). Switzerland as a single metropolitan area? A study of its commuting network. *Urban Studies*, 47 (13), 2785–2802.

Ducruet, C. and Beauguitte, L. (2014). Spatial Science and Network Science: Review and Outcomes of a Complex Relationship. *Networks and Spatial Economics*, 14 (3–4), 297–316.

Ducruet, C. and Lugo, I. (2011). Structure and dynamics of transportation networks: Models, methods and applications. *The SAGE Handbook of Transport Studies*,(May 2015), 347–364.

Dupuy, G., Van Schaick, J. and Klaasen, I.T. (2008). *Urban Networks: Network Urbanism.*

Ellison, G., Glaeser E. L., and Kerr, W. (2007). What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns. *National Bureau of Economic Research*, Working Paper, 13068.

Emmanuel, R. and Krüger, E. (2012). Urban heat island and its impact on climate change resilience in a shrinking city: The case of Glasgow, *UK. Building and Environment*, 53 . Elsevier Ltd137–149. Available from http://dx.doi.org/10.1016/j.buildenv.2012.01.020.

Emmanuel, R. and Loconsole, A. (2015). Green infrastructure as an adaptation approach to tackling urban overheating in the Glasgow Clyde Valley Region, UK. *Landscape and Urban Planning*, 138 . Elsevier B.V.71–86. Available from http://dx.doi.org/10.1016/j.landurbplan.2015.02.012.

Engelen, G., Lavalle, C., Barredo, J. I., Van der Meulen, M., and White, R. (2007). The MOLAND modelling framework for urban and regional land-use dynamics. In: *Modelling Land-Use Change*. Koomen, E. and Stillwell, J. (Eds.). Netherlands: Springer. 297–320.

Epstein, J.M. (1999). Agent-based computational model and generative social science, *Complexity,* 4, 41–60.

Erdös, P. and Rényi, A. (1959). On random graphs. *Publicationes Mathematicae*, 6 290–297.

Essam, J.W. (2002). Percolation theory. In: *Reports on Progress in Physics*. Elsevier, 833–912.

Euler, L. (1736). Solutio problematis and geometriam situs pertinentis. *Comment. Acad. Sci. U. Petrop* 8, 128–40.

European Union (EU). (2011). *Territorial Agenda of the European Union 2020: Towards and Inclusive, Smart and Sustainable Europe of Diverse Regions*. Available at: http://www.eu2011.hu.

Evans, R. et al. (no date). Urban Morphology. Urban Design, (93), 16–33.

Evans, T. S. and Lambiotte, R. (2010). Line graphs of weighted networks for overlapping communities. *European Physical Journal B*, 77 (2), 265–272. https://doi.org/10.1140/epjb/e2010-00261-8.

Everett, M. G. and Borgatti, S. P. (2020). Unpacking Burt's constraint measure. *Social Networks*, 62 (February), 50–57. https://doi.org/10.1016/j.socnet.2020.02.001

Fairburn, J., Walker, G., and Smith, G. (2004). *Investigating environmental justice in Scotland: links between measures of environmental quality and social deprivation SNIFFER report UE4(03)01*. Edinburgh: SNIFFER (Scotland and Northern Ireland Forum For Environmental Research).

Farkas, I., Abel, D., Palla, G., and Vicsek, T. (2007). Weighted network modules. *New Journal of Physics*, 9, 180-180.

Farley, K.A. (2007). Grasslands to tree plantations: forest transition in the Andes of Ecuador. *Annals of the Association of American Geographers,* 97 (4), 755–771.

Fiksel, J. (2006). Sustainability and resilience : toward a systems approach. *Sustainability: Science, Practice and Policy*, 2 (2), 14–21.

Florida, R. (2002) *The rise of the creative class: and how it's transforming work, leisure, community and everyday life*. New York, NY, Basic Books.

Ford Jr., L. R. (1956). *Network Flow Theory*. In Defense Documentation Center for Scientific and Technical Information.

Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, 486 (3–5), 75–174.

Fortunato, S., & Barthélemy, M. (2007). Resolution limit in community detection. *Proceedings of the National Academy of Sciences of the United States of America*, 104(1), 36–41. https://doi.org/10.1073/pnas.0605965104

Fotheringham, A. S., Brunsdon, C. and Charlton,M. (2005). *Quantitative Geography: Perspectives on Spatial Data Analysis.* Sage Publications. London.

Fotheringham, A.S. (1986). Modelling hierarchical destination choice (US migrants). *Environment and Planning A*, 18 (3), 401–418.

Frankhauser, P. (1998). Fractal geometry of urban patterns and their morphogenesis. Discrete

Freeman, L.C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, 1 (3), 215–239.

Fujita, M. and Ogawa, H. (1982). Multiple equilibria and structural transition of non-monocentric urban configurations. *Regional Science and Urban Economics,* 12(2), 161-196.

Gabaix, X. (1999). Zipf's Law for Cities: an explanation. *The Quarterly Journal of Economics*, 114 (August), 739–767.

Garnett, P. (2018). Total Systemic Failure. *Science of the Total Environment journal*, 626, 684–688.

Gayle, D. (2016). Daily commute of two hours – reality for 3.7m UK workers. *The Guardian.* Available at: https://www.theguardian.com/money/2016/nov/18/daily-commute-of-two-hours-reality-for-37m-uk-workers [Accessed 19 October 2018].

Geary, R. C. (1954). The contiguity ratio and statistical mapping. *The Incorporated Statistician*, 5, 115-145.

Geurs, K.T. and van Wee, B. (2004). Accessibility evaluation of land-use and transport strategies: review and research directions. *Journal of Transport geography*, 12(2), pp.127-140.

Ghisellini, P., Cialani, C. and Ulgiati, S., (2016). A review on circular economy: the expected transition to a balanced interplay of environmental and economic systems. *J. Clean. Prod.* 114, 11–32.

Giles-Corti, B. and Donovan, R. J. (2002). The relative influence of individual, social and physical environment determinants of physical activity. *Soc Sci Med*, 54(12):1793-812. doi: 10.1016/s0277-9536(01)00150-2. PMID: 12113436.

Giljum, S. et al. (2015). Global Patterns of Material Flows and their Socio-Economic and Environmental Implications: A MFA Study on All Countries World-Wide from 1980 to 2009. *Resources*, 3, 319–339.

Girvan, M. and Newman, M.E.J. (2001). Community structure in social and biological networks, 1–8. Available from http://arxiv.org/abs/cond-mat/0112110 http://dx.doi.org/10.1073/pnas.122653799.

Glaeser, E.L. (2008). The Economic Approach to Cities. *Ssrn*,1–40.

Glaeser, E.L. and Gottlieb, J.D. (2009). The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States. *Journal of Economic Literature*, 47 (4), 983–1028.

Glasgow City Council (2017). *Glasgow City Development Plan*. 137. Available from https://www.glasgow.gov.uk/index.aspx?articleid=16186.

Glass, R. (1998). *The Social Background of a Plan: A Study of Middleshbrough.* First edn. Routledge. 364.

Goodstadt, V. (2006). Strategic planning in the Glasgow metropolitan region. In: *Strategic planning for regional development in the UK: a review of principles and practices*. Routledge, 316–337. Available from https://doi.org/10.1016/j.solener.2019.02.027%0Ahttps://www.golder.com/insights/block-caving-a-viable-alternative/%0A

Graham, S. and Marvin, S. (2001). *Splintering Urbanism: Networked Infrastructures, Technological Mobilities and the Urban Condition,* (1st ed.), London: Routledge.

GRASS Development Team. (2020). *Geographic Resources Analysis Support System (GRASS) Software*, Version 7.8. Open Source Geospatial Foundation. https://grass.osgeo.org

Guimerà, R. et al. (2003). Self-similar community structure in a network of human interactions. *Physical Review E*, 68 (6), 1–4.

Gunderson, L. H., and C. S. Holling, editors. (2002). *Panarchy: understanding transformations in human and natural systems*. Island Press, Washington, D.C., USA.

Haining, R. (1990). *Spatial data analysis in the social and environmental sciences.* Cambridge: Cambridge University Press.

Handy, S. (1996). Methodologies for exploring the link between urban form and travel behaviour. *Transportation Research,* 1(2), p151-165.

Hansen, M. C., Yalcin, H. and Hayes, J. P. (1999). Unveiling the ISCAS-85 benchmarks: a case study in reverse engineering. *IEEE Design and Test of Computers,* vol. 16. no. 3. pp.72-80.
doi: 10.1109/54.785838

Hansen, W.G. (1959). How accessibility shapes land use. *Journal of American Institute of Planners,* 25 (1), 73–76.

Hanzl, M., Geerse, A., Guschl, L. and Dewan, R. (2021). Urban metabolism and land use optimization: In quest for modus operandi for urban resilience. In: Santos, P. P., Chmutina, K., Von Mendig J. and Raju, E. (eds), *Understanding Disaster Risk: A Multidimensional Approach,* Elsevier.

Hardy, M.A. and Bryman, A. (2009). *Handbook of data analysis*. First ed. Los Angeles; London: SAGE. 711.

Hartwell, L.H. et al. (1999). From molecular to modular cell biology. *Nature*, 402 (S6761), C47–C52. Available from http://www.nature.com/articles/35011540.

Heidegger, M. (1977). *The question concerning technology, and other essays.* Harper torchbooks. New York; London: Harper and Row. 182.

Hellervik, A., Nilsson, L. and Andersson, C. (2019). Preferential centrality – A new measure unifying urban activity, attraction and accessibility. *Environment and Planning B: Urban Analytics and City Science*, 46 (7), 1331–1346. https://doi.org/10.1177/2399808318812888

Heppenstall, A., Malleson, N. and Crooks, A. (2016). "Space, the Final Frontier": How Good are Agent-Based Models at Simulating Individuals and Space in Cities? *Systems*, 4 (1), 9.

Hillier, B. (2007). *Space is the machine*. Space Syntax. London.

Hillier, B., and Hanson, J. (1984). *The Social Logic of Space*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511597237

Hillier, B., Leaman, A., Stansall, P., and Bedford, M. (1976). Space Syntax. *Environment and Planning B: Planning and Design*, 3(2), 147–185. doi:10.1068/b030147.

Hirtle, S. C. and Jonides, J. (1985). Evidence of Hierarchies in Cognitive Maps, *Memory and Cognition*, 13, 208-217.

Hobbs, R. J., Higgs, E. and Harris, J. A. (2009). Novel Ecosystems: Implications from Conservation and Restoration. *Trends in Ecology and Evolution*. 24(11). 599-605.

Holme, P. et al. (2007). Korean university life in a network perspective: Dynamics of a large affiliation network. *Physica A: Statistical Mechanics and its Applications*, 373 (2), 821–830.

Ihaka, R. and Gentleman, R. (1996). R: A Language for Data Analsyis and Graphics. *Journal of Computational and Graphical Statistics*, 5 (3), 299–314.

Illenberger, J., Nagel, K. and Flötteröd, G. (2013). The role of spatial interaction in social networks. *Networks and Spatial Economics*, 13(3): 255-282.

Isard, W. (1954). Location Theory and Trade Theory: Short-Run Analysis. *Quarterly Journal of Economics,* 68 (2), 305–320. doi:10.2307/1884452. JSTOR 1884452.

Jabareen, Y.R. (2006). Sustainable urban forms: Their typologies, models, and concepts. *Journal of Planning Education and Research*, 26 (1), 38–52.

Jacobs, J. (1961). *The Death and Life of Great American Cities*. New York: Vintage Books.

Janssen, M.A. et al. (2006). Toward a network perspective of the study of resilience in social-ecological systems. Ecology and Society, 11 (1).

Jenks, M. and Jones, C. (2010). Issues and Concepts. In: Jenks, M., and Jones, C., eds. *Dimensions of the Sustainable City*.First. Oxford: Springer, 18. Available from http://link.springer.com/10.1007/978-1-4020-8647-2.

Jia, G. et al. (2012). Community Detection in Social and Biological Networks using Differential Evolution. Learning and Intelligent Optimization, Lecture No 71–85.

Kaiser, M. and Hilgetag, C.C. (2004). Spatial growth of real-world networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 69 (32), 1–6.

Kalinka, A. T., and Tomancak, P. (2011). linkcomm: An R package for the generation, visualization, and analysis of link communities in networks of arbitrary size and type. *Bioinformatics*, 27(14), 2011–2012. https://doi.org/10.1093/bioinformatics/btr311.

Kauffman, S. A., (1990). Requirements for evolvability in complex systems: orderly dynamics and frozen components, *Physica D*, 42, 135-152.

Kauffman, Stuart A. (1993). *The origins of order, self-organization and selection in evolution* (1st ed.). New York and Oxford: Oxford University Press

Kennedy, C., Cuddihy, J. and Engel-Yan, J. (2007). The Changing Metabolism of Cities. *Journal of Industrial Ecology*,11 (7), 43-59.

Kevin J. and Krizek 2003. Residential Relocation and Changes in Urban Travel: Does Neighborhood Scale Urban Form Matter?, *Journal of the American Planning Association*, 69 (3), 265-281.

Kitchin, R. (2017). Urban Science: A Short Primer. 1-13.

König, M. and Battiston, S. (2009). From Graph Theory to Models of Economic Networks. A Tutorial. Lecture Notes in Economics and Mathematical Systems, 613 (Spring), 23–63. Available from http://www.springerlink.com/index/10.1007/978-3-540-68409-1.

König, M.D. (2009). Dynamic R&D Networks: The Efficiency and Evolution of Interfirm Collaboration Networks.

König, M.D. et al. (2007). On Algebraic Graph Theory and the Dynamics of Innovation Networks. Networks and Heterogeneous Media, 3 (2), 201–219. Available from http://arxiv.org/abs/0712.2752.

Krüger, E., Drach, P., Emmanuel, R. & Corbella, O. (2013). Urban heat island and differences in outdoor comfort levels in Glasgow, UK, *Theoretical and Applied Climatology*, vol. 112, no. 1-2, pp. 127-141.

Krugman, Paul. (1991). Increasing Returns and Economic Geography. *Journal of Political Economy,* 99(3), 483-499.

Kuby, M. et al. (2005). A Comparison of Geographic Information Systems , Complex Networks , and Other Models for Analyzing Transportation Network Topologies. 58. Hampton.

Lambin, E.F. and Meyfroidt, P. (2010). Land use transitions: Socio-ecological feedback versus socio-economic change. *Land Use Policy,* 27 (2), 108–118. Available from http://linkinghub.elsevier.com/retrieve/pii/S0264837709001288.

Lambin, E.F. and Meyfroidt, P. (2011). Global land use change, economic globalization, and the looming land scarcity. *Proceedings of the National Academy of Sciences*, 108 (9), 3465–3472. Available from http://www.pnas.org/cgi/doi/10.1073/pnas.1100480108.

Lambooy, J.G. (1998). *Economies of agglomeration and spatial development : Cities in the era of the knowledge economy* []. Utrecht: University of Utrecht 1–29.

Lancichinetti, A., and Fortunato, S. (2009). Community detection algorithms: a comparative analysis, *Physical Review E,* 80 (5), 056117.

Langfelder, P., Zhang, B. and Horvath, S. (2007). Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics,* 2008 24(5), 719-720.

Langton, C. (1986). Studying artificial life with cellular automata. *Physica D*, 22, 120-149.

Langton, C.G. (1990). Computation at the edge of chaos: Phase transitions and emergent computation. *Physica D: Nonlinear Phenomena*, 42 (1–3), 12–37.

Langton, C.G. (1996). Artificial Life. In: Boden, M., ed. *The Philosophy of Artificial Life.* Oxford.

Latora, V. and Marchiori, M. (2001). Efficient behavior of small-world networks. *Physical Review Letters*, 87 (19), 198701-1-198701–4.

Lee, D.B. (1973). Requiem for Large-Scale Models. *Journal of the American Planning Association*, 39 (3), 163–178.

Lee, J., Zhang, Z. Y., Lee, J., Brooks, B. R., and Ahn, Y. Y. (2017). Inverse Resolution Limit of Partition Density and Detecting Overlapping Communities by Link-Surprise. *Scientific Reports*, 7 (1), 1–13. https://doi.org/10.1038/s41598-017-12432-1

Lehmann, S. (2016). Sustainable urbanism: towards a framework for quality and optimal density? *Future Cities and Environment,* 2 (8).

Li, L. et al. (2006). Towards a Theory of Scale-Free Graphs: Definition, Properties and Implications. *Internet Mathematics*, 2 (4), 431–523.

Light, M. T. and Thomas, J. T. (2019) Segregation and Violence Reconsidered: Do Whites Benefit from Residential Segregation? *American sociological review*. [Online] 84 (4), 690–725.

Lin, Y.-R. et al. (2009). Analyzing communities and their evolutions in dynamic social networks. *ACM Transactions on Knowledge Discovery from Data,* 3 (2), 1–31.

Lindley, S., et al., (2011). *Climate change, justice and vulnerability*, Joseph Rowntree Foundation, pp. 1–180. Available from https://www.jrf.org.uk/report/climate-change-justice-and-vulnerability.

Liu, W., et al., (2016). Finding overlapping communities in multiplex networks. *PLoS ONE*, 13 (4), 1–22. Available from http://arxiv.org/abs/1602.03746.

Liu, Z., He, C. and Wu, J., (2016). General spatiotemporal patterns of urbanization: An examination of 16 world cities. *Sustainability* 8(1): 41.

Llewelyn-Davies. (1997). *Sustainable Residential Quality: New Approach to Urban Living,* report for the GOL, DETR and LPAC , London.

Long, H. et al., (2014). Effects of land use transitions due to rapid urbanisation on ecosystem services: Implications for urban planning in the new developing area of China. *Habitat International*, 44 . Elsevier Ltd536–544. Available from http://dx.doi.org/10.1016/j.habitatint.2014.10.011.

Lorrain, D. (2005). Gig@city: The Rise of Technological Networks in Daily Life. in: Coutard et al. (eds.). Coutard, O., Hanley, R. E. and Zimmerman, R. (2005). *Sustaining Urban Networks: the Social Diffusion of Large Technical Systems.*Hanley, R.E., Graham, S., and Marvin, S., eds. *The Networked Cities Series*. First Edn. Abingdon: Routledge. Available from http://www.tandfebooks.com/isbn/9780203357118%5Cnhttp://samples.sainsburysebooks.co.uk/9781134345700_sample_502037.pdf.

Louf, R. and Barthelemy, M., (2013). Modeling the polycentric transition of cities. *Physical Review Letters,* (111) 198702.

Louf, R. and Barthelemy, M., (2014). How congestion shapes cities: from mobility patterns to scaling. *Scientific Reports,* 4 5561.

Lovelace, R. and Ellison, R., (2018). stplanr: A Package for Transport Planning. *The R Journal*, 10(2).

Lowry, I.S. (1965). A short course in model design. *Journal of the American Institute of Planners*, 31, 158–166.

Luck, M. and Wu, J. (2002). A gradient analysis of urban landscape pattern: A case study from the Phoenix metropolitan region, Arizona, USA. *Landscape Ecology*, 17 (4), 327–339.

Maantay, J. and Maroko, A. (2015). 'At-risk' places: Inequities in the distribution of environmental stressors and prescription rates of mental health medications in Glasgow, Scotland, *Environmental Research Letters,* 11 (10) .

Macdonald, F. (2019). The ‘Glasgow effect’ implies cities make us sad. Can the city prove the opposite? *The Guardian*, 16 October. Available from: https://www.theguardian.com/cities/2019/oct/16/urban-living-makes-us-miserable-this-city-is-trying-to-change-that [Accessed 20 September 2020]

Majekodunmi, M., Emmanuel, R. and Jafry, T. (2020). A spatial exploration of deprivation and green infrastructure ecosystem services within Glasgow city. *Urban Forestry and Urban Greening*, 52 (August 2019), . Elsevier126698. Available from https://doi.org/10.1016/j.ufug.2020.126698.

Major, M. D. (2018). *The Syntax of City Space: American Urban Grids.* First Ed. Routledge: New York.

Marshall, S. and Gong, Y. (2009). Urban Pattern Specification. *Solutions*, (November), 1–36.

Masucci, A.P. et al. (2009). Random planar graphs and the London street network. *European Physical Journal B*, 71 (2), 259–271.

Matthews, E. et al. (2000). *The Weigh of Nations: Material Outflows from Industrial Economies*. 126. Washington DC. Available from http://scholar.google.com/scholar?hl=enandbtnG=Searchandq=intitle:THE+WEIGHT+OF+NATIONS+MATERIAL+OUTFLOWS+FROM+INDUSTRIAL+ECONOMIES#5.

McArthur, J. (2018). Comparative infrastructural modalities: Examining spatial strategies for Melbourne, Auckland and Vancouver. *Environment and Planning C: Politics and Space*, 0 (0), 1–21.

McDonough, W. and Braungart, M. (2009). *Cradle to cradle: remaking the way we make things*. London: Vintage. 212.

McGarigal, K. and Marks, B.J. (1995). FRAGSTATS: spatial pattern analysis program for quantifying landscape structure. 134. Available from https://www.fs.usda.gov/treesearch/pubs/3064.

Merçay, A. and Borrie, J. (2006). Physics of Diplomacy ? the Dynamics of Complex Social Phenomena and Their Implications for Multilateral Negotiations. In: Borrie, J., and Randin, V.M., eds. *Thinking Outside the Box in Multilateral Disarmament and Arms Control Negotiations*. United Nations, 129–164.

Molinero, C., Arcaute, E., Smith, D., and Batty, M. (2015). The Fractured Nature of British Politics. https://arxiv.org/abs/1505.00217

Moran, P. A. P. (1950). Notes on continuous statistic phenomena. *Biometrika,* 37, 17-23.

Moudon, A.V. (1997). Urban morphology as an emerging interdisciplinary field. *Urban Morphology*, 1 (1), 3–10.

Naimzada, A.K., Stefani, S. and Torriero, A. (2009). *Lecture notes in economics and mathematical systems.* Fandel, G., and Trockel, W., eds. *Networks, Topology and Dynamics: Theory and Applications to Economics and Social Systems*. Berlin Heidelberg: Springer. Available from http://linkinghub.elsevier.com/retrieve/pii/0304400982900080.

Nakano, S. (2014). Tree Enumeration. In: Kao, M.-Y., ed. *Encyclopedia of Algorithms*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1–4. Available from https://doi.org/10.1007/978-3-642-27848-8_726-1.

Neuman, M. (2005). The compact city fallacy. *Journal of Planning Education and Research*, 25 (1), 11–26.

Neuman, M. (2006). Infiltrating infrastructures: On the nature of networked infrastructure. *Journal of Urban Technology*, 13 (1), 3–31.

Neuman, M. (2022). *Sustainable Infrastructure for Cities and Societies*. London; New York: Routledge.

Neuman, M. and Churchill, S.W. (2015). Measuring sustainability. *Environmental and Ecological Statistics*, 4 (86), 457–482.

Neuman, M. and Hull, A. (2009). The Futures of the City Region, *Regional Studies*, 43 (6), 777-787, DOI: 10.1080/00343400903037511

Neuman, M. and Hull, A. (2011). Introduction: The Futures of the City Region. In: Neuman, M. and Hull, A. *The Futures of the City Region.* Regions and Cities. London; New York: Routledge.

Neuman, M. and Smith, S. (2010). City planning and infrastructure: Once and future partners. *Journal of Planning History*, 9 (1), 21–42.

Newman, M. E. J. (2006). Finding community structure in networks using the eigenvectors of matrices. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 74 (3). https://doi.org/10.1103/PhysRevE.74.036104

Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America*, 103 (23), 8577–8696.

Newman, M. E. J., Forrest, S. and Balthrop, J. (2002). *Physical Review E*, 66, 035101(R).

Newman, M.E.J. (2004). Fast algorithm for detecting community structure in networks. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, 69 (6), 5.

Newman, M.E.J. and Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 69 (2 2), 1–15.

Nourian, P. et al. (2018). Modelling walking and cycling accessibility and mobility: The effect of network configuration and occupancy on spatial dynamics of active mobility. *Smart and Sustainable Built Environment*, 7 (1), 101–116.

Nystuen, J. D. and Dacey, M. F. (1961). A Graph Theory Interpretation of Nodal Regions. *Papers and Proceedings of the Regional Science Association*, 7, 29-42.

O'Sullivan, D.B. (2016). *Graph-based Cellular Automaton Models of Urban Spatial Processes []*. ProQuest LLC.

Oberoi, K.S. et al. (2018). Towards a qualitative spatial model for road traffic in urban environment. *IEEE Conference on Intelligent Transportation Systems*, Proceedings, ITSC, 2018-March (October), 1–6.

OGC (2011). *OpenGIS® Implementation Standard for Geographic information - Simple feature access - Part 1: Common architecture*. Open Geospatial Consortium, Inc,93. Available from http://portal.opengeospatial.org/files/?artifact_id=25355E+Implementation+Standard+for+Geographic+information+-+Simple+feature+access#1.

Olsen, J.R. et al. (2016). Effects of new motorway infrastructure on active travel in the local population: A retrospective repeat cross-sectional study in Glasgow, Scotland. *International Journal of Behavioral Nutrition and Physical Activity*, 13 (1), .

Onnela, J.P. et al. (2007). Structure and tie strengths in mobile communication networks. *PNAS*, 99 (12), . National Academy of Sciences7821–7826. Available from httpwww.pnas.orgcontent104187332 [Accessed 6 September 2018].

OpenStreetMap contributors (2021). Planet dump [Data file from 2020 of database dump$]. Retrieved from https://planet.openstreetmap.org.

Padgham, M, Rudis, B., Lovelace, R., Salmon M. (2017). osmdata. *The Journal of Open Source Software*, 2(14). doi: 10.21105/joss.00305, https://doi.org/10.21105/joss.00305.

Palla, G. et al. (2005). Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435 814–818.

Pan, W. et al. (2013). Urban characteristics attributable to density-driven tie formation. *Nature Communications*, 4 .

Pflieger, G. and Rozenblat, C. (2010). Introduction. Urban networks and network theory: The city as the connector of multiple networks. *Urban Studies*, 47 (13), 2723–2735.

Ponzetto, G.A.M. (2012). *Introduction to Urban Economics Spatial Equilibrium within a City* []. 61. Available from http://www.crei.cat/wp-content/uploads/users/pages/Week01.pdf.

Popham, F. and Boyle, P. J. (2011). Is there a 'Scottish effect' for mortality? Prospective observational study of census linkage studies. *Journal of Public Health,* Sep;33(3):453-8. doi: 10.1093/pubmed/fdr023. Epub 2011 Apr 14. PMID: 21493620.

Portugali, J. (2011). *Complexity, cognition and the city,* Springer. New York.

Pulido Barrera, P., Rosales Carreón, J. and de Boer, H.J. (2018). A multi-level framework for metabolism in urban energy systems from an ecological perspective. *Resources, Conservation and Recycling*, 132 (May 2017) . Elsevier230–238. Available from https://doi.org/10.1016/j.resconrec.2017.05.005.

QGIS.org, (2021). *QGIS Geographic Information System*. QGIS Association. http://www.qgis.org

Ravalde, T. and Keirstead, J. (2017). Comparing performance metrics for multi-resource systems: the case of urban metabolism. *Journal of Cleaner Production*, 163 . ElsevierS241–S253. Available from http://dx.doi.org/10.1016/j.jclepro.2015.10.118.

Ravasz, E. et al. (2002). Hierarchical Organization of Modularity in Metabolic Networks. *Science*, 297 1551–1555.

Reckien, D., Eisenack, K. and Lüdeke, M.K.B. (2011). Land consumption by urban sprawl - a new approach to deduce urban development scenarios from actors' preferences. *Environmental Modelling and Assessment*, 16 (465), 465-477. Available from: https://doi.org/10.1007/s10666-011-9254-6.

Rees, W.E. (1996). Revisiting carrying capacity: Area-based indicators of sustainability. Population and Environment, 17 (3), 195–215. Available from http://link.springer.com/10.1007/BF02208489.

Ripley, B.D. (1981). *Spatial statistics.* Wiley series in probability and mathematical statistics. New York ; Wiley.

Rodriguez-Iturbe, I., Marani, M., Rigon, R. and Rinaldo, A. (1994). Fractal river basins: chance and self-organization. *Water Resources Research*, Cambridge University Press: Cambridge, 12 (30), 3531-3539.

Rokem, J. and Vaughan, L. (2019). Geographies of ethnic segregation in Stockholm: The role of mobility and co-presence in shaping the 'diverse' city. *Urban Studies*, 56 (12), 2426–2446. https://doi.org/10.1177/0042098018795561

Romanowicz, A. et al. (2004). European Environment Agency Developments of Land and Ecosystem Accounts: General Overview. *European Environment*, 6 (1), 65–70.

Rosenthal, H. (1960). Unpublished thesis. *Acquaintances and contacts of Franklin Roosevelt,* Massachusetts Institute pf Technology.

Rosenthal, S.S. and Strange, W.C. (2003). Evidence on the Nature and Sources of Agglomeration Economies []. 63.

Roth, C. et al. (2011). Structure of urban movements: Polycentric activity and entangled hierarchical flows. *PLoS ONE*, 6 (1), 2–9.

S.C.U.U. of the W. of E. Bristol. (2012). *Science for Environment Policy In- depth Report: Resource Efficiency Indicators*, Report produced for the European Commission DG Environment. 1-32. Available from http://ec.europa.eu/environment/integration/research/newsalert/pdf/IR4_en.pdf.

Santamouris, M. (2016). *Minimizing energy consumption, energy poverty and global and local climate change in the built environment: innovating to zero: causalities and impacts in a zero concept world.* Elsevier.

Sassen, S. (1990). Economic Restructuring and The American City. *Annual Review of Sociology*, 16 (1990), 465–490. Available from http://www.jstor.org/stable/2083278.

Sassen, S. (2001). *The global city: New York, London, Tokyo*. Princeton, N.J: Princeton University Press.

Schaffler, A. (2018). *Value shifts: The rise of multifunctional infrastructure* []. UC Berkeley. Available from https://escholarship.org/uc/item/1n8904rd.

Schläpfer, M. et al. (2014). The scaling of human interactions with city size. *Journal of the Royal Society Interface*, 11 (98), 1–21.

Schmolke, A., P. Thorbek, D. L. DeAngelis, and V. Grimm. (2010). Ecological models supporting environmental decision-making: A strategy for the future. *Trends in Ecology & Evolution* 25 (8): 479–86.

Schneider, M. (1997). *Spatial Data Types for Database Systems—Finite Resolution Geometry for Geographic Information Systems*. Lecture Notes in Computer Science, vol. 1288. Springer-Verlag, Berlin, Heidelberg, Germany.

Schneider, M. and Behr, T. (2006). Topological relationships between complex spatial objects. *ACM Transactions on Database Systems*, 31 (1), 39–81.

Schweitzer, F. et al. (2009). Economic Networks: The New Challenges. *Science*, 325 (5939), 422–426.

Scott, A. J. (1971). *Combinatorial Programming, Spatial Analysis and Planning.* First edn. London, Methuen and Co Ltd.

Scottish Government (2012). *Vacant and Derelict Land Survey*. Statistical Bulletin Planning Series. Edinburgh: A National Statistics Publication for Scotland.

Scottish Government (2019). Scottish Awards for Quality in Planning 2019 []. 1.

Seto, K.C., Guneralp, B. and Hutyra, L.R. (2012). Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. Proceedings of the National Academy of Sciences, 109 (40), 16083–16088. Available from http://www.pnas.org/cgi/doi/10.1073/pnas.1211658109.

Shafizadeh-Moghadam, H. et al. (2017). Coupling machine learning, tree-based and statistical models with cellular automata to simulate urban growth. *Computers, Environment and Urban Systems*, 64 297–308.

Shen, J., Chen, M. and Liu, X. (2018). Classification of topological relations between spatial objects in two-dimensional space within the dimensionally extended 9-intersection model. *Transactions in GIS*, 22 (2), 514–541.

Sim, A. et al. (2015). Great cities look small. *Journal of the Royal Society Interface*, 12 (109).

Simon, H.A. (1969). *The sciences of the Artificial.* Third. Cambridge, Mass ; London: MIT Press.

Smith, M., Hosking, J., Woodward, A. et al. (2017). Systematic literature review of built environment effects on physical activity and active transport – an update and new findings on health equity. *Int J Behav Nutr Phys Act* 14, 158. https://doi.org/10.1186/s12966-017-0613-9

Snijders, T.A.B. (2001). The Statistical Evaluation of Social Network Dynamics. *Sociological Methodology,* 31 361–395. Available from http://www.jstor.org/stable/3097281?seq=1andcid=pdf-reference#references_tab_contents.

Snijders, T.A.B. (2003). Accounting for degree distributions in empirical analysis of network dynamics []. *Dynamic social network modeling and Analysis: Workshop Summary and Papers.* 1–16. Available from http://scholar.google.com/scholar?hl=enandbtnG=Searchandq=intitle:Accounting+for+Degree+Distributions+in+Empirical+Analysis+of+Network+Dynamics#0.

Solé, R. V. et al. (2008). Robustness of the European power grids under intentional attack. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 77 (2), 1–7.

Song, Y. et al. (2017). The relationships between urban form and Urban commuting: An empirical study in China. *Sustainability*.

Stanford University (2000). *LEDA A Research project.* Accessed on 19 June 2019. Available from: http://graphics.stanford.edu/courses/cs368-00-spring/TA/manuals/LEDA/planar_map.html

Stanley Milgram. (1967). The small world problem. *Psychology Today*, 2(1):60–67. http://snap.stanford.edu/class/cs224w-readings/milgram67smallworld.pdf

Strominger, J., Anthopolos, R. and Miranda, M.L., (2016). Implications of construction method and spatial scale on measures of the built environment. *International journal of health geographics*, 15, 15. https://doi.org/10.1186/s12942-016-0044-x

Sveikauskas, L. (1975). The Productivity of Cities. *The Quarterly Journal of Economics*, 89 (3), 393. Available from https://academic.oup.com/qje/article-lookup/doi/10.2307/1885259.

Talen, E., (1999). Sense of community and neighbourhood form: an assessment of the social doctrine of New Urbanism. *Urban Studies*, 36 (8), 1361–1379. doi:10.1080/0042098993033.

Tan, L.M. et al. (2018). Ecological network analysis on intra-city metabolism of functional urban areas in England and Wales. *Resources, Conservation and Recycling*, 138 (July), 172–182.

te Brömmelstroet, M., Nikolaeva, A., Cadima, C., Verlinghieri, E., Ferreira, A., Mladenovic, M., Milakis, D., Silva, J. de A. e, and Papa, E. (2021). Have a good trip! Expanding our concepts of the quality of travel time with flow theory. *Applied Mobilities*, 00(00), 1–22. https://doi.org/10.1080/23800127.2021.1912947

Tobler, W. R. (1970). Computer movie simulating urban growth in Detroit region. *Economic Geography,* 46, 234-240.

Torrens, P. (2000). How cellular models of urban systems work. *CASA working series papers*. Centre for Advanced Spatial Analysis (University College London).

Turner, A. (2000). Angular Analysis: a Method for the Quantification of Space. *Centre for Advanced Spatial Analysis (CASA),* Working paper, 23.

Turner, M. G. (1988). A spatial simulation model of land use changes in a Piedmont county in Georgia. *Applied Mathematics and Computation*, 27 (1), 39–51. https://doi.org/10.1016/0096-3003(88)90097-5

Turner, M. G. (1989). *Landscape Ecology: The Effect of Pattern on Process*. Annual Review of Ecology and Systematics, 20(1989), 171–197. http://www.jstor.org/stable/2097089

UN DESA. (2018). *2018 Revision of World Urbanisation Prospects*, 3.

van der Veer, J. (2000). Urban Segregation and the Welfare State. Inequality and Exclusion in Western Cities, Sako Musterd and Wim Ostendorf (Eds.). *Journal of Housing and the Built Environment*, 15 (2), 201–204. Available from http://link.springer.com/10.1023/A:1010132332193.

Van Nes, A. and Yamu, C. (2018). Space Syntax: a Method To Measure Urban Space Related To Social, Economic and Cognitive Factors. In: Yamu, C., Poplin, A., Devisch, O., et al., eds. *The Virtual And The Real in Planning and Urban Design: Perspectives, practices and applications*. First. Oxon, New York: Routledge, 136–150.

von Thünen, J. H. (1826). *Der isolirte Staat in Beziehung auf Landwirtschaft und Nationalökonomie*. Wirtschaft and Finan.

Walsh, D. et al. (2010). *It's not 'just deprivation': Why do equally deprived UK cities experience different health outcomes?* Public Health, 124 (9), . Elsevier Ltd487–495. Available from http://dx.doi.org/10.1016/j.puhe.2010.02.006.

Wang, J. and Provan, G. (2008). Generating Application-Specfic Benchmark Models for Complex Systems. *Artificial Intelligence,* 566–571.

Wang, J. and Provan, G. (2009). A comparative analysis of specific spatial network topological models. Lecture Notes of the Institute for Computer Sciences, *Social-Informatics and Telecommunications Engineering*, 5 LNICST (PART 2), 1514–1525.

Wang, J., Zhu, Q. and Mao, Q. (2007). The Three-Dimensional Extension of Space Syntax. *6th International Space Syntax Symposium*, 048.01-048.16. Available from https://pdfs.semanticscholar.org/d97d/d8f927d9a1d9cf5709f57fcc0a36b6bd3579.pdf%0Ahttp://www.spacesyntaxistanbul.itu.edu.tr/papers%5Clongpapers%5C048 - Wang Zhu Mao.pdf.

Wang, P. et al. (2015). Understanding the Spreading Patterns of Mobile Phone Viruses Understanding of Mobile the Spreading Patterns Phone Viruses 324 (5930), 1071–1076.

Wathern, P. (1990). An introductory guide to EIA. In: Wathern, P. (ed.) *Environmental Impact Assessment: Theory and Practice*. First Edn. New York: Routledge. pp. 3-30

Watkins, C., A. (2001). The Definition and Identification of Housing Submarkets. *Environment and Planning A: Economy and Space*, 33 (12), 2235-2253. doi:10.1068/a34162

Weber, M. (1929/1971). *Theory of the Location of Industries*. Chicago: Chicago University Press

West, D.B. (2001). *Introduction to Graph Theory.* Second edi. Prentice Hall.

West, G.B. et al. (1999). The Fourth Dimension of Life: Fractal Geometry and Allometric Scaling of Organisms. *American Association for the Advancement of Science*, 284 (5420), 1677–1679. Available from www.jstor.org/stable/2898653?seq=1andcid=pdf-reference#references_tab_contents.

Wheaton, W. (1974). A comparative static analysis of urban spatial structure. *Journal of Economic Theory,* 9, 223-237.

Wheeler, S. M. (2013). *Planning for Sustainability*. London: Routledge.

White, R. and Engelen, G. (1993). Cellular Automata and Fractal Urban Form: A Cellular Modelling Approach to the Evolution of Urban Land-Use Patterns. *Environment and Planning A: Economy and Space*, 25 (8), 1175–1199. Available from http://journals.sagepub.com/doi/10.1068/a251175.

Williams, K., Burton, E. and Jenks, M. (Eds.) (2000). *Achieving Sustainable Urban Form.* London: E and FN Spon.

Wilson, A.G. (1971). A Family of Spatial Interaction Models, and Associated Developments. *Environment and Planning A: Economy and Space*, 3 (1), 1–32.

Wu, C. et al. (2017). Economic transition, spatial development and urban land use efficiency in the Yangtze River Delta, China. *Habitat International*, 63 . Elsevier Ltd67–78. Available from http://dx.doi.org/10.1016/j.habitatint.2017.03.012.

Yamu, C. and van Nes, A. (2017). An Integrated Modeling Approach Combining Multifractal Urban Planning with a Space Syntax Perspective. *Urban Science*, 1 (37), 1–21.

Zeller, C. (2010). The pharma-biotech complex and interconnected regional innovation arenas. *Urban Studies,* 47 (13), 2867–2894.

Zhang, Y. et al. (2018). CO2 metabolic flow analysis in global trade based on ecological network analysis. *Journal of Cleaner Production*, 170 . Elsevier Ltd, 34–41. Available from https://doi.org/10.1016/j.jclepro.2017.08.244.

Zoomers, A. (2010). Globalisation and the foreignisation of space: Seven processes driving the current global land grab. *Journal of Peasant Studies*, 37 (2), 429–447.