Mapping the conceptual system of an urban theory and its evolution: a text analysis of space syntax conference papers over 20 years

Kimon Krenz, Sophia Psarra and **Vinicius M. Netto**
E-mails: k.krenz@ucl.ac.uk, s.psarra@ucl.ac.uk, vmnetto@fe.up.pt
ORCIDs: 0000-0001-6077-7282, 0000-0002-6674-7735, 0000-0003-4175-5701

Revised version received 25 May 2023

Abstract: This work presents a new methodology for analysing the conceptual system of a scientific theory over time. Through the application of a quantitative text analysis to a large corpus of conference proceedings from 1997 to 2017, this study offers insight into the development and relationships of concepts within the domain of space syntax as a representative field within urban morphology. The evolution of this specific field is tracked for the first time with the aim of inspiring comparable analyses in other areas of urban morphology and urban studies.

Keywords: space syntax, conceptual system, quantitative text analysis, concept identification, conference proceedings

Contact: Kimon Krenz, Faculty of the Built Environment, University College London UCL, 22 Gordon Street, London WC1H 0QB (corresponding author).

Since its inception in the 1970s, space syntax has consolidated into a comprehensive theoretical and methodological framework for investigating the relationship between society and space. Over time, the field has grown into a significant and varied area of research, comprising a large international and diverse research community which utilises this framework to investigate a variety of spatio-temporal urban conditions. Space syntax brought to light a theory of a whole new dimension of society–space relations. An important contribution of this might arguably be a mathematical description of spatial relationships through graphs (Hillier, 1996a; Hillier and Hanson, 1984), and the identification of how these relationships influence pedestrian movement and co-presence, economic activity and their externalities (Hillier, 1996b; Hillier *et al*., 1993). In space syntax theory, urban space is not seen as the background to human activity, but as an active participant in shaping the potential of human action and interaction. In other words, the social and the physical are in a reciprocal relationship (Hillier and Hanson, 1984; Hillier and Vaughan, 2007). Hence space syntax is significantly different from existing investigations which – despite the spatial turn in geography and related disciplines (Harvey, 1989; Soja, 1989) – have been mainly approached through socio-economic and political factors, by almost excluding the spatial. Notable exceptions are Henri Lefebvre's theories of the production of space (1991) and Anthony Giddens' work (1984) among others.

This spatial lens in space syntax has provided researchers and practitioners in the field of urban research with a variety of theoretical and analytical concepts to describe urban phenomena for which they would otherwise lack the language and tools to consider. It should be noted that space syntax is not solely an urban theory, but a theory about the relationship between built space and society that includes aspects of both architectural and urban theories. In this paper, the focus is on the urban theory component of space syntax. A prominent idea in this theory is Hillier's (1996b) concept of 'spatial configuration' and its relationship to movement and land uses which has informed numerous explorations of the location of economic activities in cities (Porta *et al*., 2012; Alalouch *et al*., 2019). Other examples are the concepts of spatial segregation and integration casting light on the role of transport systems in contested urban spaces (Rokem and Vaughan, 2019; Shtern and Rokem, 2021), or the spatial clustering of ethnic communities (Vaughan and Penn, 2006; Legeby, 2010). The concepts of centrality and choice have been used to analyse the impact of spatial measures on the housing market (Xiao *et al.*, 2016; Law, 2017) and crime in urban settings (Hillier and Shu, 2000; Hillier and Sahbaz, 2009).

Such concepts and theories have cross-fertilized, intersected with and informed various disciplines relevant to the field of urban research. For example, space syntax has been extensively employed in archaeology to investigate how elusive social relations and processes are materialized through the configuration of cities, public spaces and domestic architecture (Dawson, 2002; Fredrick and Vennarucci, 2021). The role of topological structures and geometrical features in navigation and cognition as driving forces in spatial behaviour is now largely recognized in cognitive studies and the neuroscience (Javadi *et al*., 2017; Coutrot *et al*., 2022). The degree to which spatial structures influence walkable urban neighbourhoods has increasingly been recognized as an influential factor in the field of public health (Koohsari *et al*., 2016; McCormack *et al*., 2021). On the other hand, space syntax has not been developed in a vacuum; instead, numerous preceding theorizations of society and space have informed the development of concepts in the field. Influences from sociology and anthropology can be identified in Durkheim's (2014) concept of solidarities and Turner's (1969) communitas respectively. Other influences from sociology in space syntax is the notion of co-presence as an elementary condition of social awareness, as in Goffman (1963), and a view of society as encounter systems fundamental to social reproduction, as in Giddens (1984).

Overall, space syntax is a concept-rich theory that, due to its long existence, constitutes an ideal example for investigating the development of ideas in a research field (see Yamu *et al.* (2021) for a recent review of frequently-used concepts in space syntax theory). For instance, up to the present, this field has produced over 1,500 conference papers by more than 1,100 international researchers, making these the most representative body of scholarly publications. This growing pool of texts provides large temporal data on the usage of concepts that can capture the emergence, consolidation, and continuing production in this research area. This data is used to investigate the following questions: can the evolution of this urban theory be traced through the use of specific concepts and what can such an analysis reveal about the development of ideas in this field?

Methodologically, this study positions itself in the field of scientometrics which aims to map the dynamics of scientific fields and understand the evolution of knowledge. We combine and adapt established scientometric (Chen and Song, 2019) and bibliometric (Donthu *et al*., 2021) methods such as science mapping (co-word analysis) with network analysis (network metrics, clustering and visualisation techniques). However, our work goes beyond the sole evaluation of scholarly impact through the measurement of standardized metrics, and offers a mapping of the conceptual system of an urban theory.

Previous research utilizing scientometric approaches has employed quantitative text analyses to scientific publications in order to extract thematic focuses through the articles' keywords, and has gained insights into collaborations through the creation of networks based on citations between these articles (Bergeaud *et al.*, 2017; Raimbault 2019). This paper differs from this previous work in that, instead of looking at keywords chosen by authors, concepts are extracted from articles through experts and quantitative methods, and focuses on the use of these concepts in texts, both individually and in conjunction. The intention is to trace points of influence on space syntax by other approaches to cities and urban analysis. Finally, by introducing quantitative text analysis to urban research, this work aims to contribute to explorations that address the development of urban research over time.

The methodological approach deployed in this analysis involves three steps. First, concepts in the field are identified through a combined qualitative and quantitative text analysis that processes large amounts of texts and identifies terminologies. These terminologies encompass theoretical concepts (such as 'solidarity' or 'visibility'), methodological concepts (such as 'integration' or 'choice') and technical concepts (such as 'angular analysis') (this is linguistic identification). Secondly, the frequencies of the occurrence of these concepts and terms are examined statistically over time (tracing concepts over time). Thirdly, networks of relationships among these notions are explored to see whether there are patterns between established concepts that stabilize over time, or recently emerge (identifying conceptual systems). Networks of concepts and terms are mapped through their co-occurrence in text and their structure examined. This process is then used to understand patterns and trends in the evolution of knowledge in space syntax. Such analysis can shed light on the possibility of the relationship between a stable structure of ideas ('canonical') with concepts occurring less persistently in the body of papers.

Text-analysis: methods

Linguistic identification

A concept can be defined as an abstract representation or definition of an entity or phenomenon, generally in the form of a generic idea which has been generalized from particular instances (Saitta and Zucker 2013); it is through the formulation of concepts that humans create generalized theoretical understanding. The identification of concepts through quantitative methods is a challenging endeavour. What particularly constitutes a concept depends highly on the theoretical framework into which a concept is embedded (Blumer, 1931). This makes the formulation of a positive definition, that is, based on the existence of actual properties or components, difficult to achieve. While computers outperform humans in tasks such as processing large amounts of text, they are still weak with more complex tasks such as the extraction of meaning from texts, a fundamental aspect in the identification of concepts. For this research, a mixed approach of quantitative and qualitative methods is proposed to identify theoretical, methodological, and technical concepts.

For the quantitative part, a 'bag-of-words' model (Harris, 1954) is employed to establish a list of words ranked according to their frequency across the database of conference articles. This model is commonly used in natural language processing and information retrieval, and builds on the core assumption that a word's frequency relates to its importance within a text (Zipf, 1932). This implies that some words occur significantly more often than others, allowing the identification and interpretation of their role within a language. In our research, word frequencies are computed for single words (unigrams) and a combination of two words (bigrams) with the use of the R package 'quanteda' for quantitative analysis of textual data (Benoit *et al*., 2018; R Development Core Team, 2018). The general aim is to identify concepts and terminologies that are used frequently, rather than those used only once.

The database used in this analysis consists of all papers published in the space syntax symposia proceedings from 1997 to 2017. Publications are converted into a TXT document type and cleaned from page numbers, author, and affiliation information, as well as the list of references. The database is further cleaned through automated methods (for example regular expressions and quanteda-based stop-word removals). The number of publications in the database is 1,089 with a total of 46,601 unique unigrams. Word frequencies are computed for all types within the corpus and compiled a list of the 6,000 most frequent unigrams and bigrams (see Figure 1 for the frequency distribution of the 40 most frequent unigrams and bigrams). These two lists were shown to three independent expert researchers active in space syntax with the request to mark/code all words that are or could be characterized as 'concepts' within the field. It is acknowledged that starting from the 6,000 most frequent unigrams and bigrams can form a limitation for the identification of very new concepts, as their frequency is inevitably too low to be in the top 3,000 tokens. However, the aim is not to identify all potential concepts and their scientific significance, but rather to arrive at a list of concepts whose development is sufficiently frequent to be traced over time. Simultaneously, the number of papers per conference increases with time, which potentially allows for newly-developed concepts to be picked up by the analysis regardless of whether they are widely used in the field.

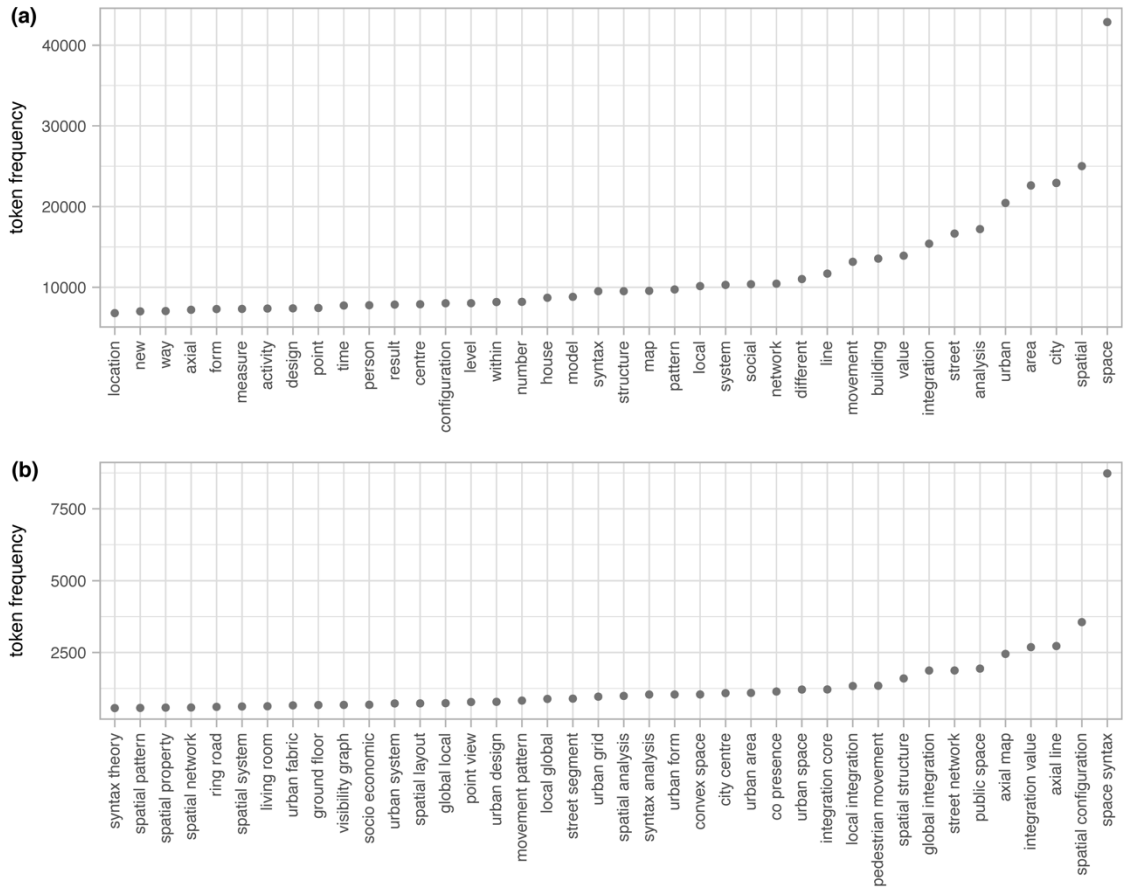

Figure 1. The 40 most used (a) unigrams and (b) bigrams and their frequency across all conferences.

Tracing concepts over time

Once the list of concepts by the three experts was identified, their trajectories were traced over time. This was done by computing frequencies for each of the concepts per conference year. This provided a quantification of their respective usage and their classification into the categories of consistent, declining or increasing concepts over time, and the identification of emergent concepts. If a concept is not present in early conferences, but occurs at some point and is increasingly used over time, it is identified as an emerging concept. If a concept is frequently used in early years, but less so in later conferences, is categorized as a declining concept, and likewise if there is no significant change in frequency, a concept is classified as a consistent concept. A concept under sustained usage over time is classified as a consistent concept. This classification should be seen as a continuous number of increasing or decreasing degrees. Furthermore, the computed frequency is standardized to allow comparisons of trajectories across years and differing frequencies (henceforth 'scaled word frequency'). Firstly, frequencies are divided by the total number of tokens per conference year in order to be able to compare frequencies between different years. Secondly, frequencies are standardized by subtracting the mean and dividing by the standard deviation of the entire vector of word frequencies. Finally, the data trend for each word is calculated and their slope is compared across all concepts. Based on this comparison, concepts are grouped into the aforementioned categories.

Identifying conceptual systems

In addition to the concepts' trajectory, it is of particular interest to identify the relationship among different concepts. For this, the co-occurrence for each word in relation to all other words is computed and the relationship between different concepts is examined. This is done by counting how often each

concept occurs in each paper and the entire corpus; the resulting vectors of occurrences are then correlated with each other and form a word co-occurrence matrix. This matrix can be visualized as a network of relationships. Within this network, edges are undirected and weighted by the respective correlation value. Formal descriptions of network properties and morphology provide insights into how and to which extent concepts relate to each other.

Results

Concepts in the field of space syntax

Using the previously-introduced list of most frequent words, the three researchers independently identified 816 terms in total, of which 205 are unigrams and 611 are bigrams. For reasons of intercoder reliability, that is, the extent to which two or more independent researchers/coders agree on the coding, only those terms are selected which have been identified by at least two researchers. Only 284 (222 bigrams and 65 unigrams) of these terms are classified as 'concepts' by at least two researchers. This means that three-quarters of the most-frequent concepts in space syntax are predominantly bigrams – two-word combinations. These bigram concepts are, for example, words such as (in unrelated order) 'foreground network', 'generic function' or 'isovist occlusivity'. Examples of identified unigrams are 'interface', 'permeability', or 'choice'. **Error! Reference source not found.** shows the 20 most frequent of these 284 identified terms as well as the number of papers in which each occurs.

Table 1. (a) 20 most frequent uni- and bigrams identified by researchers in the field, among a total of 284 terms; (b) number of newly emerged concepts per conference year since 1997.

(a)

#	unigram	frequency	No of papers	bigram	frequency	No of papers
1	integration	15401	933	space syntax	8723	980
2	local	10147	856	spatial configuration	3559	661
3	syntax	9510	1001	axial line	2724	434
4	configuration	8018	923	integration value	2619	528
5	axial	7206	668	axial map	2452	455
6	global	6690	747	global integration	1863	349
7	choice	4926	628	street network	1852	319
8	accessibility	4531	607	spatial structure	1598	458
9	depth	4470	596	local integration	1312	258
10	core	3882	587	integration core	1207	227
11	visibility	3665	418	co presence	1040	215
12	path	3511	530	convex space	1039	217
13	syntactic	3275	537	urban grid	962	249
14	connectivity	3224	518	street segment	881	184
15	node	3147	394	movement pattern	803	267
16	isovist	2775	216	visibility graph	674	158
17	centrality	2628	338	visual field	532	158
18	visitor	2291	328	step depth	530	100
19	configurational	2279	495	syntactic analysis	524	219
20	angular	2169	265	natural movement	494	191

(b)

Concepts	193	23	27	9	22	4	2	0	3	1	0
Year	1997	1999	2001	2003	2005	2007	2009	2012	2013	2015	2017

Concept trajectories

The established series of keywords has been used to perform a temporal analysis of word frequencies. Concepts are subsequently grouped according to their commonalities in terms of slope values in order to classify a concept into the identified categories (increasing, declining or consistent). Figure 2 shows a visualization of plots of the ten most increasing, declining, and consistent concepts. The share of increasingly-frequent concepts (trend line slope > 0.10) is 80 of 284, twice as large as the number of declining concepts at 27 (trend line slope < -0.10), while the majority shows rather steady development across the observed time period at 180.

The overall trends of each of the ten concepts with most increasing frequencies (Figure 2(a)) are highly comparable, with relative variation in their curve development. Differing from these common developments, the time of emergence of concepts is one of the key differences. However, while some terms (for example 'nach value' or 'angular distance') steadily increase over time, others (for example 'betweenness' or 'background network') increase, showing higher variation in their curve development. The ten concepts with most decreasing frequencies (Figure 2(b)) exhibit much higher variation in their curve compared to the ten most increasing concepts. Most concepts have been highly used in early conferences, after which their use declined rapidly within the first five conferences. Concepts such as 'space occupancy' and 'global movement', for example, were highly used in the first conference (1997), declining, and almost disappearing entirely shortly thereafter. Others, such as 'axial graph', followed the same trajectory, but experienced a sharp increase in 2007, just to disappear shortly afterwards. Overall, one can observe that a series of concepts have significantly lost importance over time. Trajectories of decline are not always steady processes and short temporary recurrences are not rare, yet their overall decline is clearly visible. A view on the ten concepts with slope values < 0.1 and > -0.1 unveils a third kind of trajectory, namely concepts that feature no substantial increase or decrease over time (Figure 2(c)). Such concepts have mostly been present in all conference years and can be identified as the concepts that define the 'space syntax canon'.

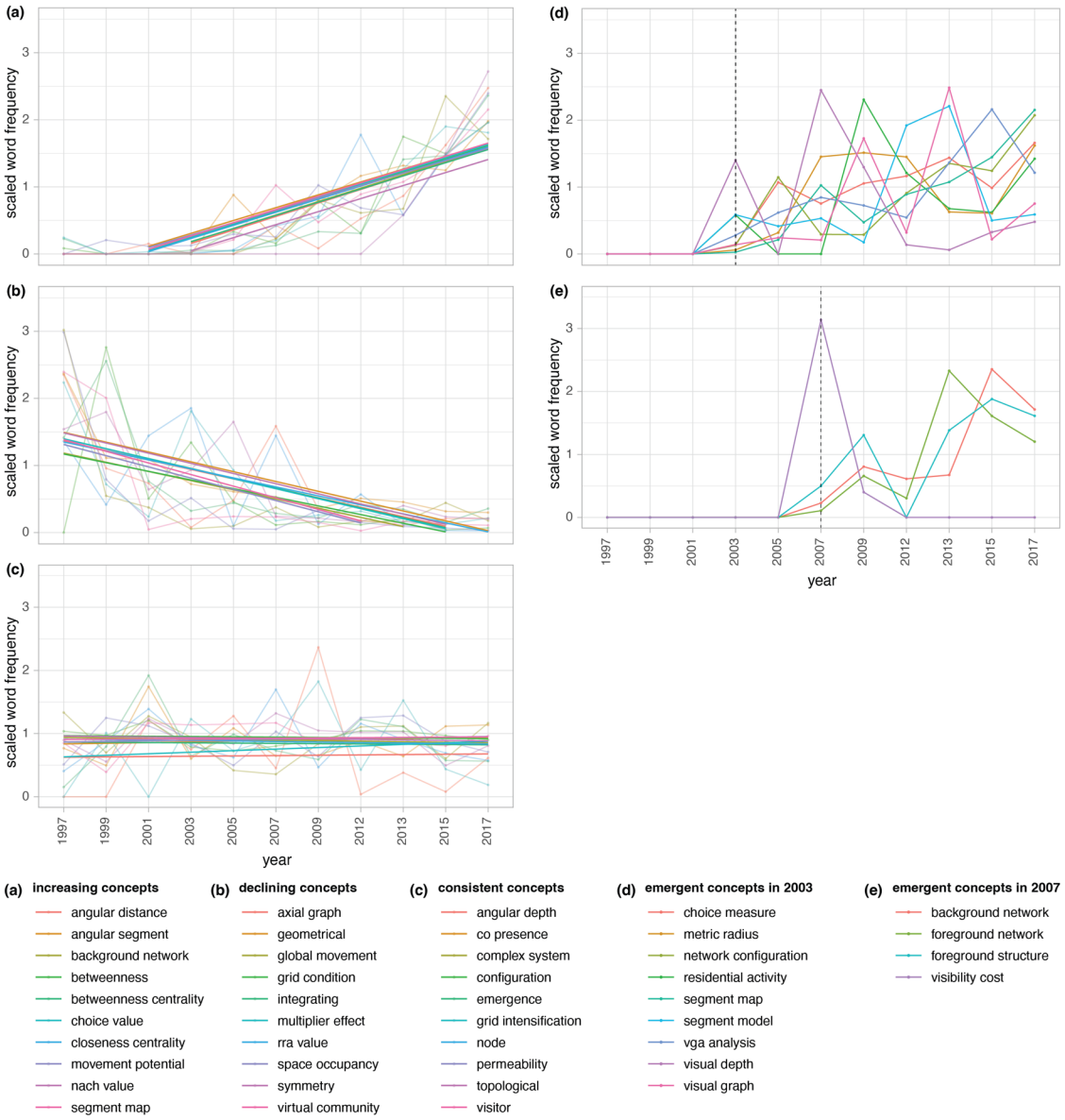

Figure 2. The ten most (a) increasing, (b) declining, and (c) consistent concepts and terms, as well as concepts that emerged in (d) 2003 and (e) 2007, and their scaled word frequencies per conference year (1997 to 2017).

Furthermore, by focusing on years in which concepts occur for the very first time, occurrences of high linguistic and conceptual novelty as well as those lacking such can be identified. For this, the occurrence of new concepts per conference year was counted (Table 1(b)). By definition, the first symposium in 1997 was the one with the highest number of 'newly' introduced concepts. Due to the difficulty in determining whether all these concepts were truly 'new' at the time, this analysis focuses on the second to eleventh symposia. Here, a clear trend of decreasing conceptual novelty is apparent. In 1999, 23 new concepts were introduced, yet five conferences later in 2009 only three new concepts were introduced. The conferences in the years 1999, 2001 and 2005 had the highest number of newly-introduced concepts. Such a shift in conceptual production might be related to an increasing endeavour to investigate the application of concepts in research. At the same time, these results need to be considered with care, as concepts that have very recently emerged might not have been utilized sufficiently frequently to feature in the initial list of 6,000 most frequent words. New concepts introduced in more recent years might not have found sufficient frequency to feature above the threshold considered in the proposed filter. However, this implies that the proposed method would not be able to pick up any novel concepts in recent years, which is not the case. A closer examination of the moment of emergence and development of newly-introduced concepts can shed light on this situation. Figure 2

also shows plots of scaled word frequencies of all words that have been newly introduced in the conference years 2003 (Figure 2(d)) and 2007 (Figure 2(e)), exhibiting conceptual emergence. In 2003, a series of 9 concepts were introduced with different relative frequencies. Here, what differs substantially is not only the relative frequency at the moment of introduction but also its development in the following years. Whereas, for example, the concept 'choice measure' had relatively low levels of use in the year of its introduction, it became consistently higher over the following years, with little variation in its occurrence. In contrast, the concept of 'visual depth', shows an 'M'-shaped development. First, this concept was highly used in the year it was first introduced, then not used at all in 2005, followed by high levels of usage in 2007 and 2009; finally, it disappeared entirely after 2009. Such a spark of usage in one year, followed by a decline, also characterizes concepts that emerged in 2007 (Figure 2(e)). Here the concept of 'visibility cost' was introduced for the first time; its usage decreased significantly in 2009 and was never used since. An example of concepts whose pattern of occurrence indicates a parallel development is 'foreground network' and 'background network'; their curve shows the co-dependence and interrelationship between the two concepts. Investigating trajectories of concepts over time has provided insights into the widespread significance of some concepts, the emergence of new concepts and the moment in which this occurs, and the presence of concepts that show identical or similar trajectories. However, this investigation does not capture the actual relationships between two or more concepts. Network analyses can be used to further explore these relations.

Conceptual systems

To understand how concepts and terms are related to each other, a network system of related concepts is constructed. This network is based on significant correlations between the co-occurrences of concepts and allows the visualization of the conceptual system of the space syntax field. This is done by counting the occurrence of each of the 284 concepts for the entire corpus (1,089 papers), resulting in a co-occurrence matrix (284 x 1,089). These matrix columns are correlated with each other in all possible combinations (that is, each concept against every other concept) and the resulting correlation matrix forms the basis of the network creation. Only concepts that feature a significant p-value ($p < 0.05$) and a Pearson's correlation coefficient above a threshold of +0.1 are then connected to each other. Figure 3 shows the result of this process: a network system of all identified space syntax terms and concepts. The conceptual system offers relevant insights into clusters of concepts. Clusters in such networks are identified quantitatively using a random walks community detection algorithm (Walktrap: Pons and Latapy, 2005).

We highlight four fundamental clusters in the network. These four clusters are distinctive in terms of their network relationships and in relation to their temporal development. Cluster 1, is a dense network of predominantly emerging concepts; cluster 2 is a dense network of predominantly declining concepts; cluster 3 is a sparser network with heterogeneous usages of concepts, and cluster 4 is a less-densely connected network of declining theoretical concepts. More significantly, the first two clusters are densely related to each other (there is a high number of network edges between concepts of both clusters) indicating the replacement of one category of methodological concepts with another, while also maintaining their interrelationship. In contrast, clusters 3 and 4 are more isolated. This indicates that theoretical and methodological concepts are less frequently used together in texts.

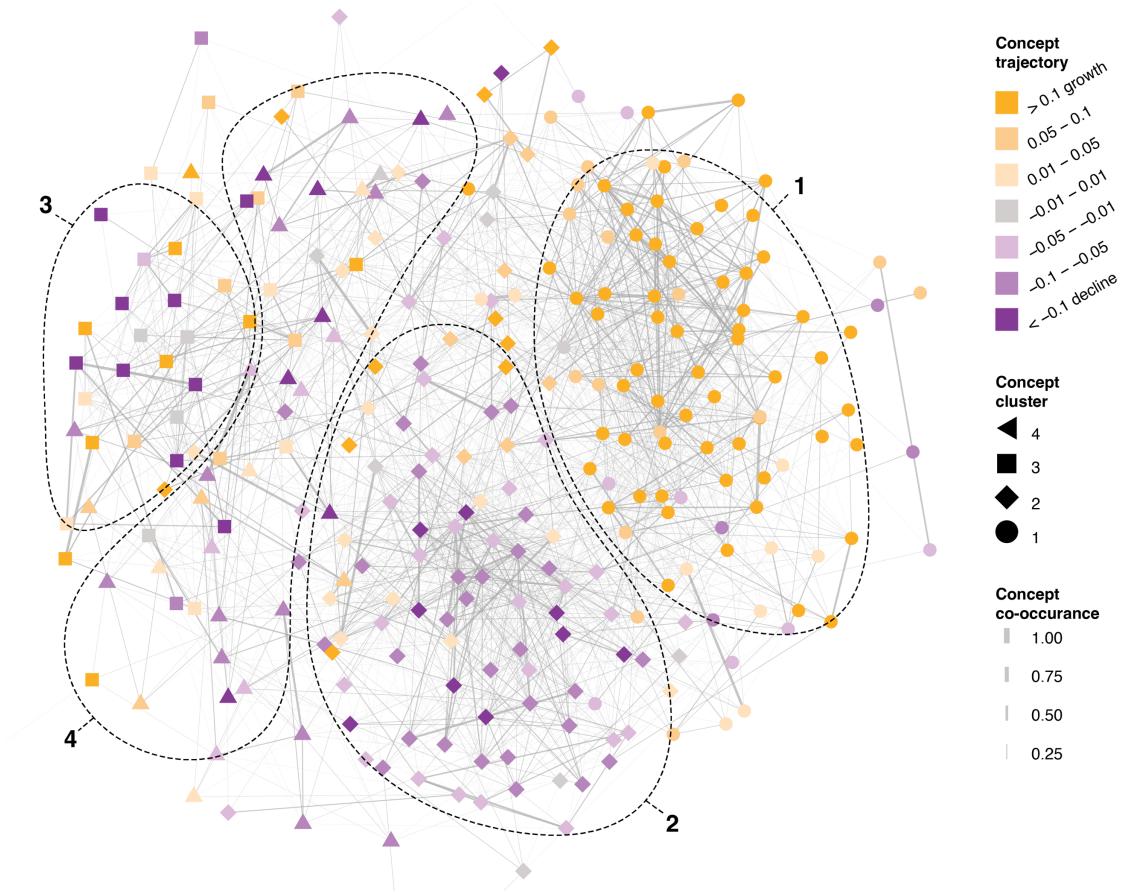

Figure 3. Network of concept relationships. Edge thickness indicates the strength of the relationship based on the correlation coefficient, while purple and yellow colours highlight decreasing and increasing usage over time. Symbol nodes and dashed lines highlight distinctive conceptual sub-clusters.

A closer examination of these clusters shows the nature of these conceptual subsystems, which are distinctive networks within the conceptual network. The network node size in Figures 4-7 is weighted according to the network degree, or connectivity of a node; this gives an indication of how many other concepts a particular concept is related to (that is, feature a correlation of R2>0.1).

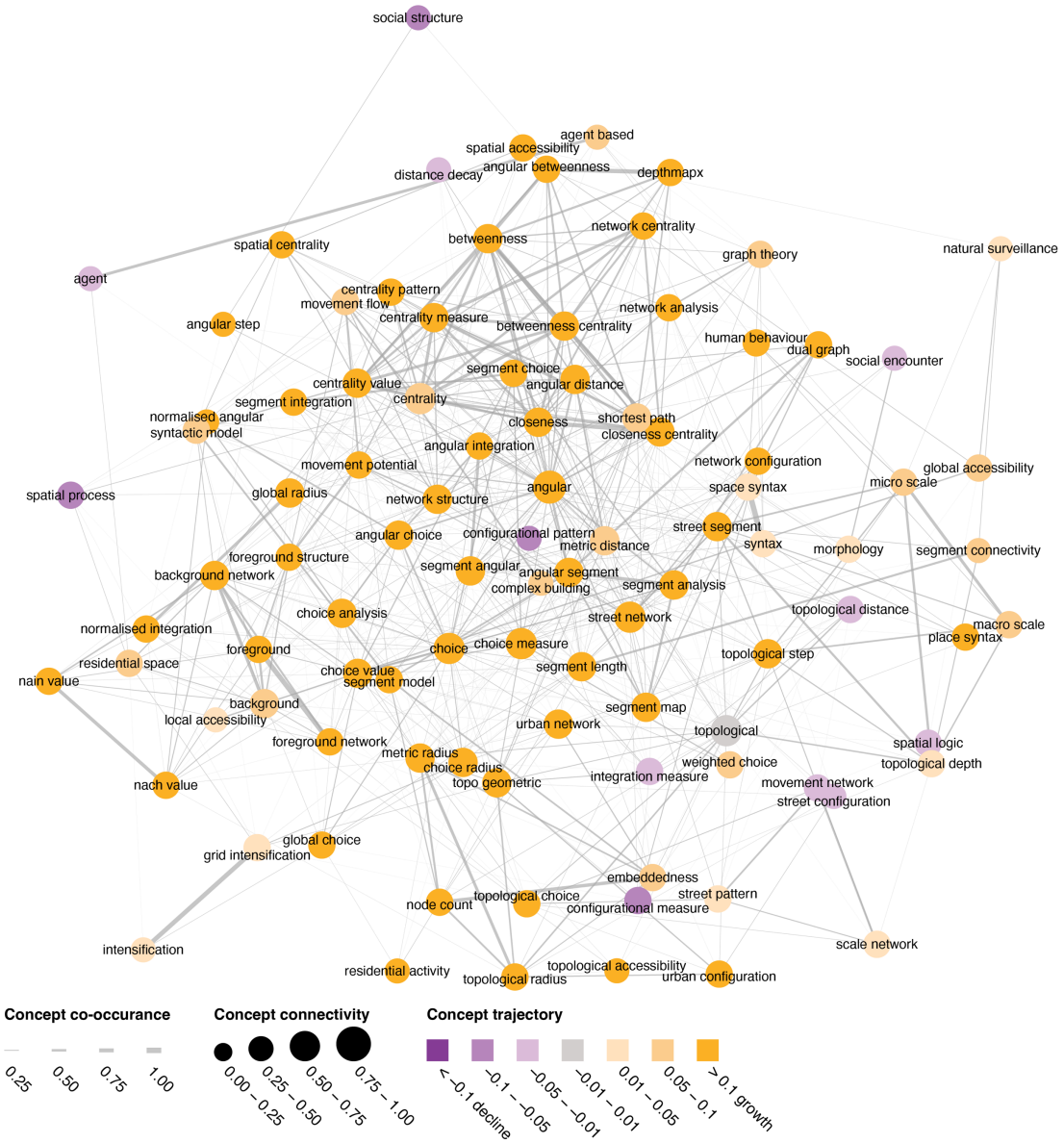

Figure 4. Cluster community 1. In this and Figures 5–7, edge thickness indicates the strength of the relationship based on the correlation coefficient, node size indicates the connectivity of concepts based on the network degree, while purple and yellow colours highlight decreasing and increasing usage over time.

Figure 4 shows the network of cluster 1. The concept with the highest degree (the number of connections a concept has to other concepts) is ‘angular’, followed by ‘betweenness centrality’ and ‘angular choice’. All concepts are of a rather methodological nature, with an emphasis on spatial network terminologies such as angularity, segments, centrality, shortest path, and choice/betweenness. Since all of the concepts exhibit an increasing trend, this cluster appears as one of the most important research directions in the field of space syntax. There are further sub-clusters within this conceptual network. For example, ‘betweenness centrality’, ‘closeness centrality’, ‘network centrality’ form a distinctive sub-network, as well as technical concepts like ‘nach value’, ‘nain value’, and ‘normalized angular’, which indicate specific sets of words that are simultaneously used by authors employing any of these concepts.

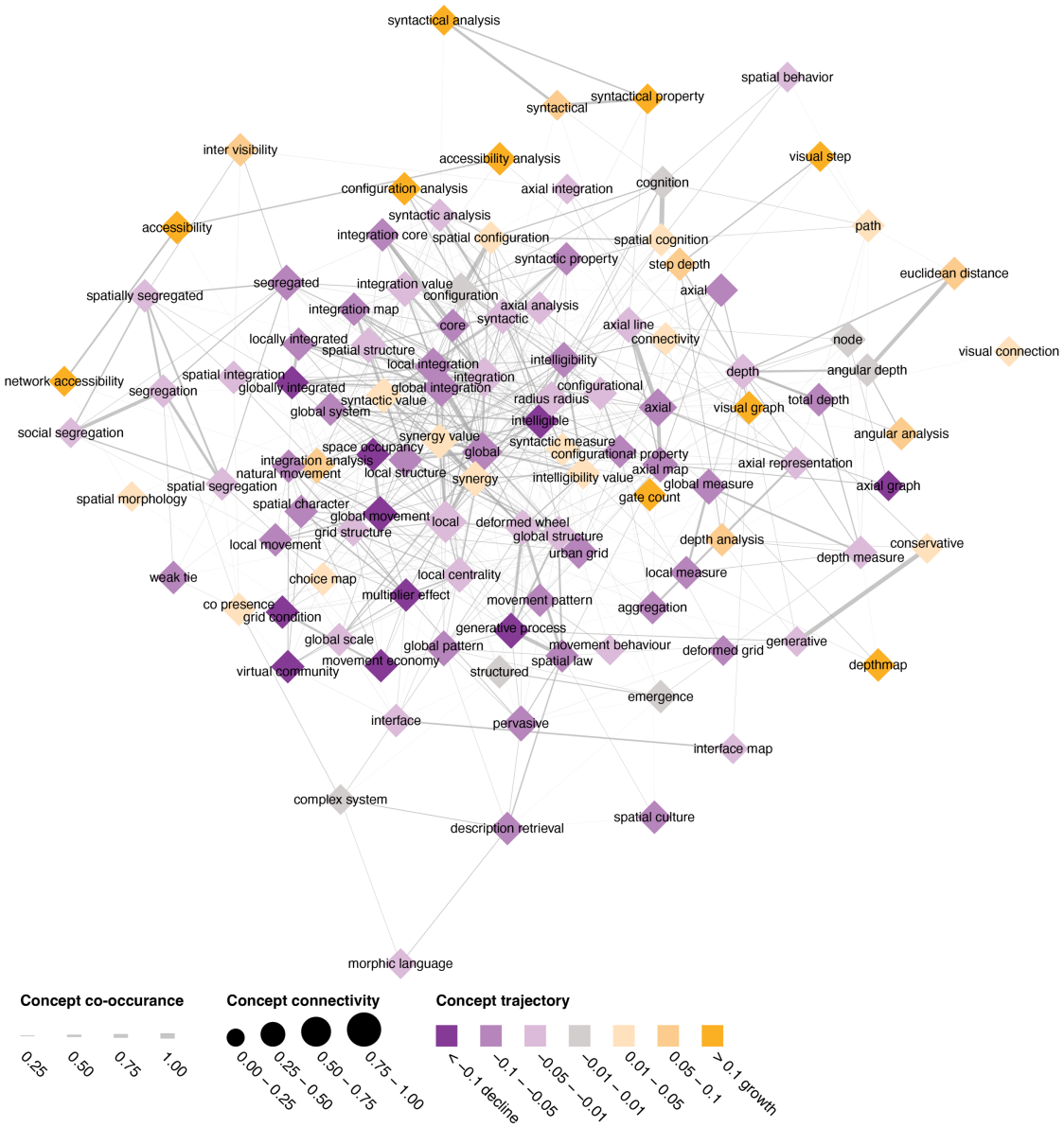

Figure 5. Cluster community 2.

The second conceptual subsystem (cluster 2) features a different network morphology compared to cluster 1 (Figure 5). All concepts exhibit declining trajectories and point to a strand of theoretical and methodological terminology that loses its importance in the field. The concepts with the highest connectivity in cluster 2 are 'integration', 'axial' and 'global', to which two tree-like subgroups connect. The first subgroup, connected to 'integration', defines the work on segregation with concepts such as 'social segregation' and 'spatial segregation'; the second subgroup, connected to 'global', is formed of concepts such as 'intelligibility' and 'synergy'. Overall, cluster 2 has more theoretical concepts related to properties of a configurational nature, such as 'global', or cognitive nature, such as 'intelligibility', as opposed to cluster 1, which features a higher number of methodological concepts. Furthermore, cluster 2 relates to the earlier methodological work in space syntax, when axial line maps formed the main method of analysis. Compared to cluster 1, this early methodology has declined over time, as the introduction of angular analysis (A. Turner, 2000) and the subsequent method for angular segment analysis breaking axial lines into segments resulted in an increasing number of studies employing segment maps. This seems to coincide with the growing usage of concepts from network theory.

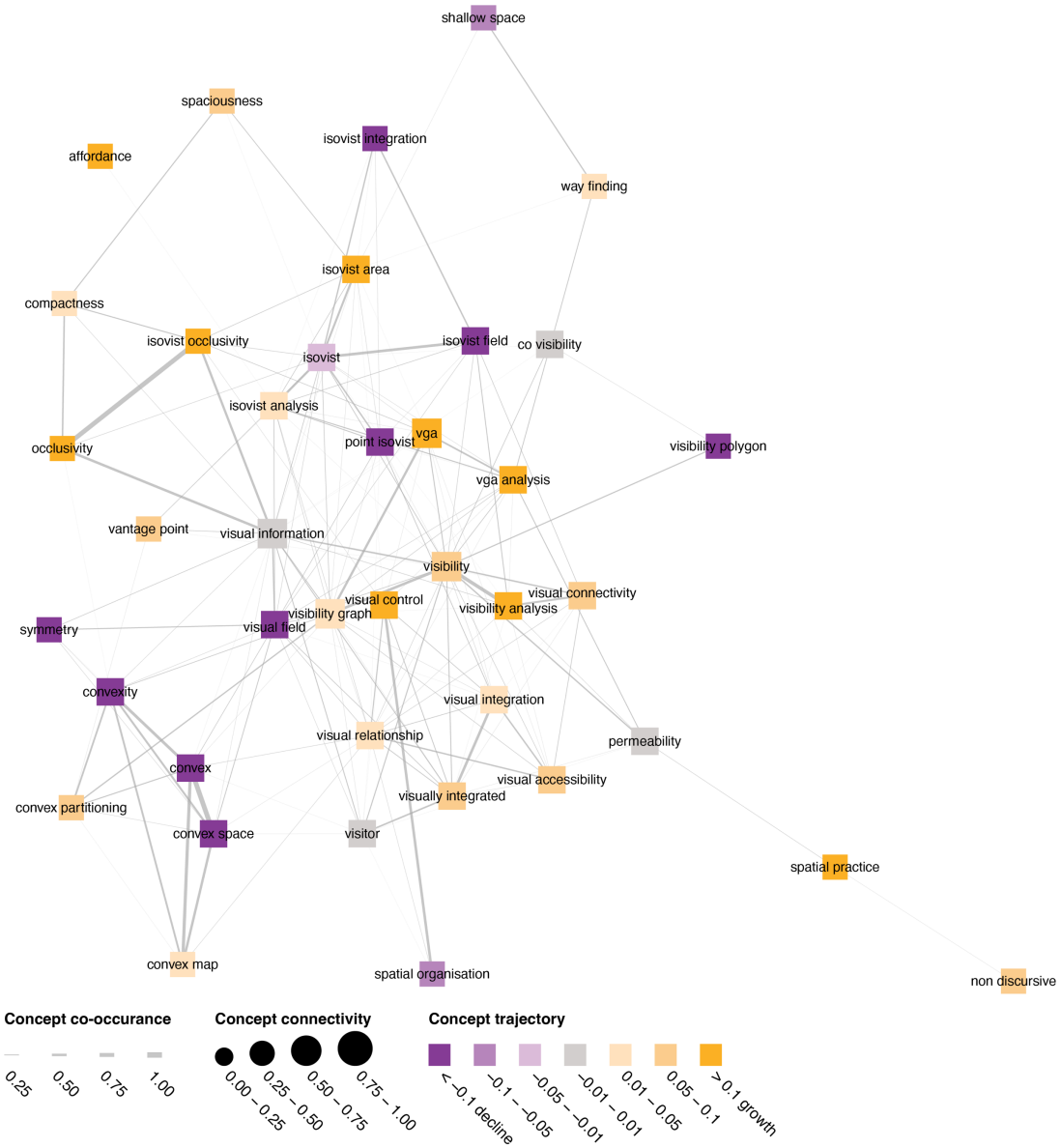

Figure 6. Cluster community 3.

The third conceptual subsystem (cluster 3) is not characterized by a clear trend in terms of the usage of concepts (Figure 6). The concepts with the highest degree are ‘isovist’, ‘visibility graph’ and ‘visual integration’. These concepts are mostly used in the analysis of buildings and small-scale spaces, pointing to a particular strand of research, which features a more heterogeneous development compared to clusters 1 and 2.

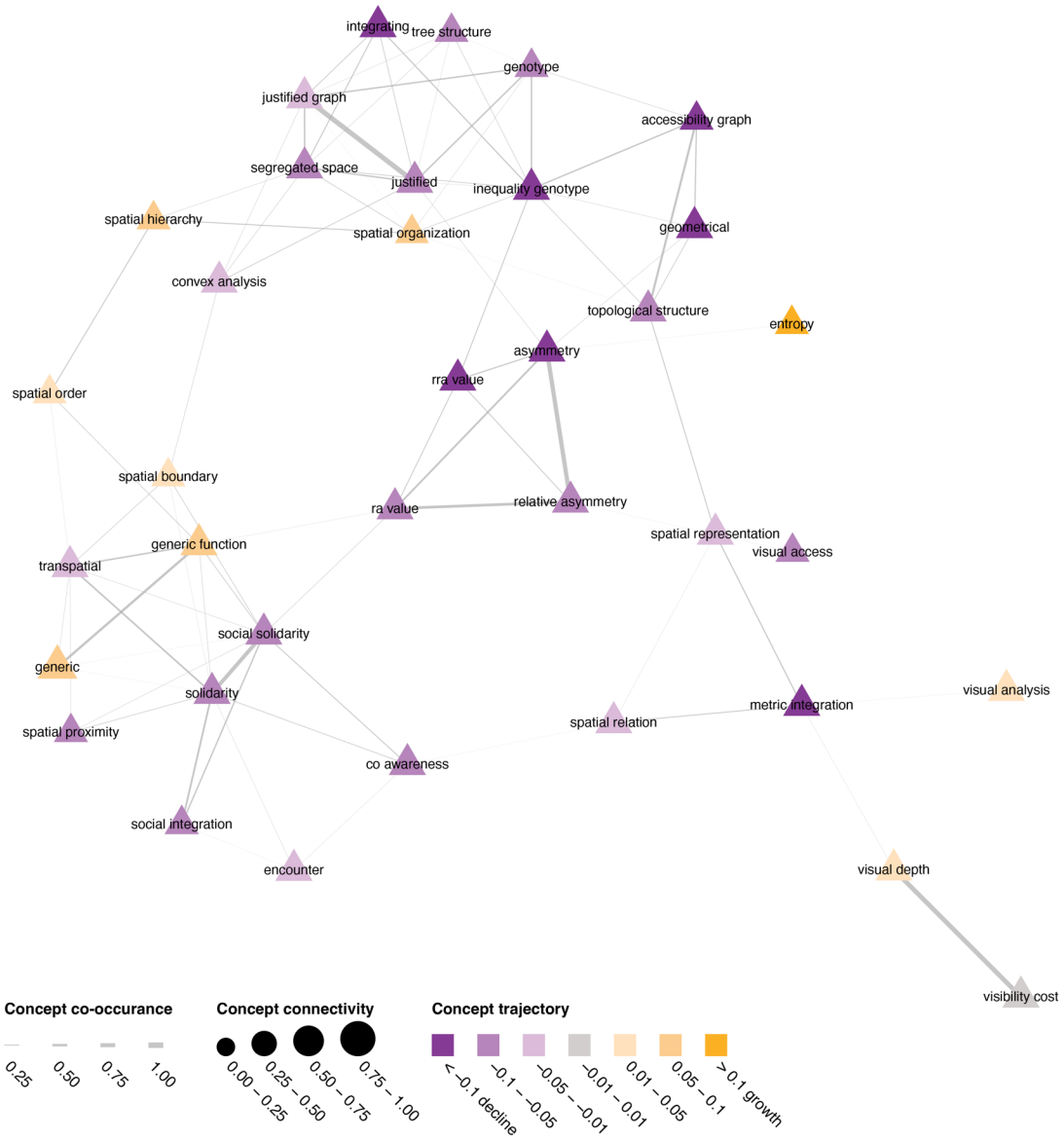

Figure 7. Cluster community 4.

Finally, a fourth cluster shows the relative decline in the use of early theoretical concepts (Figure 7). The network analysis of space syntax concepts reveals that a constellation of methodological concepts and technical terms has not only come to be proportionally dominant over theoretical terms, but sociological concepts like ‘solidarity’, ‘encounter’ and ‘transpatial’ have significantly declined in usage.

Discussion

Our analyses showed a number of trends within the space syntax field. As the field evolved, new concepts emerged (such as ‘metric radius’, ‘topological radius’, ‘angular integration’), suggesting intensified efforts for analytical advancement and novelty. There is a clear split between early theoretical and methodological spatial concepts (‘integration’, ‘axial line’) and newer methodological ones (‘choice’, ‘segment’). This change shows that the field has sought methodological connections with network science and other approaches that use network analysis. This trend is indicated by the growing use of concepts from network theory, such as ‘betweenness centrality’, the growing use of GIS tools and road-centred maps (OpenStreetMap data), which are open access and technically comparable

to segment maps. Accordingly, early terms from the major original work in the field, *The social logic of space* (Hillier and Hanson, 1984), such as 'relative asymmetry' and 'axial map', are used to lesser degrees since 1997.

Similarly, a declining use of topological distance and increasing use of metric and angular distance can be identified. Segment analysis based on angular distance emerges as the central methodology, along with the concepts of 'closeness' and 'betweenness centrality'. Concepts like 'occlusivity' (M. L. Benedikt, 1979) and 'visibility analysis' form a cluster of their own, as Visibility Graph Analysis and Isovist analysis become more commonly used, due to increasing computational power (M. Benedikt and McElhinney, 2019). Here, the influence of computational advances on the use of language and research focus, thereby shaping the field itself, is visible. The results suggest that the use of techniques and technical development is the main focus in this field, with an emphasis on advancing knowledge through analytical application. However, less attention appears to be paid to theoretical exploration.

Developed independently of network theory in the 1980s, the notion of integration (Hillier and Hanson, 1984) approximates the measure of closeness centrality (Sabidussi, 1966; Freeman, 1977) in network theory, and has been recently replaced by it. The same applies to the measure of betweenness centrality that substituted the measure of choice. These changes clearly indicate that space syntax research is adapting and aligning its concepts with network theory, communicating with a wider audience from other fields. This is a reasonable development that seeks methodological advancement based on established mathematical concepts. A similar trend refers to the recent borrowing of concepts, such as 'affordance', 'agent-based', 'entropy' and 'spatial practice' from other disciplines. The notion of 'affordance' (slope value of +0.14) was first introduced in 2001 and has been continuously rising since 2007, bridging space, cognition, and performance. After an initial peak in 2003, the concept 'agent-based' (slope value of +0.06) is also consistently increasing since 2012. The increased frequency of the term might be associated with the software depthmapX, one of the initial computational platforms in space syntax. Finally, the concept of 'entropy' (slope value of +0.10) links space syntax research with many disciplines, from information theory and physics to biology, social theory, urban studies and analytics (Shannon, 1948; Bailey, 1990; Gleick, 2011; Batty *et al*., 2014; Hidalgo, 2015; Davies, 2019). Its use within space syntax might be related to those other fields, ranging from assessing diversity in the environment to describing its levels of order.

Previous critiques of space syntax raised concerns about the apparent isolation of the field and risks of a self-referential terminology (Westin, 2014). However, the frequent usage of borrowed concepts suggests increasing influence by other fields. On the other hand, in terms of theoretical notions, there are relatively few new concepts in the field. Concepts related to the notion of morphogenesis, such as 'generative process', 'spatial law', as well as 'morphic language' and 'description retrieval', are declining after peaks in 2001 and 2003 respectively. In contrast, concepts related to cognition ('spatial cognition', 'way-finding', 'visual information', 'visibility', 'visual connection' and 'visual integration') are stable, while concepts related to 'convexity' are declining: possibly indicating a decline of building studies or local-scale analysis of urban spaces. One exception is the increased use of the concept of 'visual control'.

Importantly, a telling decline in the use of early concepts rooted in sociology and anthropology, such as 'virtual community', 'solidarity', 'encounter', 'social structure', and 'segregated space', can be observed. Concepts related to theoretical definitions of movement ('movement economy', 'natural movement', 'movement pattern') are also losing importance. The increasing use of concepts such as 'human behaviour' suggests seeking alignment with behavioural science as opposed to the early influences of Durkheim on Hillier and Hanson (1984). On the other hand, concepts such as 'spatial practice' indicate influences from the social theories of Lefebvre and de Certeau.

Conclusion

It is expected that a well-defined set of theoretical concepts should guide a research field (Kuhn, 1962), including new methodological developments. However, the decrease in usage of theoretical concepts, coupled with the increase of methodological and technical ones, suggests that – as a scientific field –

space syntax is refining its methodological and technical dimensions rather than using them to draw new theoretical directions. On the one hand, the apparent decline in the use of concepts since 2005 may be seen as a function of the growing stability of a theory as a paradigm, in Kuhn's (1962) sense of a theory established around a set of concepts and methodological rules that ensure the coherence of its applications and further theoretical and methodological developments. On the other hand, it might be related to a growing conservatism apparent in the recursive use of its own conceptual terminology and its reproduction as 'normal science' (Kuhn, 1962). Alternatively, having grown its conceptual framework over the first ten-fifteen years since the mid-1970s, and supported by the recent explosion of digital technologies and data, the field has been mainly directing its efforts in building empirical evidence and testing its key concepts and theories. Growing achievements in other fields such as network science (Barabási, 2013) have clearly played a role, influencing space syntax researchers toward technologically-driven explorations.

However, from the outset, space syntax explored new ways of describing the socio-spatial dimensions of buildings and cities not simply by representing and measuring spatial relationships through graph analysis, but also by providing a theory of space, one which should account for "how and why different societies generate different spatial patterns" (Hillier and Hanson, 1984, p. 14) rather than interpreting a variable (different societies or social patterns) by a constant (one particular instance of observable phenomenon). According to Hillier and Hanson (1984, p. 8), as opposed to analysing space, analysing behaviour and looking at their relationship, a theory of space considers that "society already pervades those patterns that need to be analyzed" while space carries social determination in its very form as object. A set of "postulates" (Hillier and Hanson, 1984, pp. 95–7) linking different categories and measures of space with social categories in *The social logic of space* established what the authors described as an "interpretive framework", a layered and structured set of relationships between spatial morphology, spatial measures and socio-spatial interpretation. Therefore, the strength of space syntax has historically lain not simply in its quantifiable and technological aspects, but also in the systemic interrelationship of theoretical and methodological ideas, and spatial and social concepts. The rising emphasis on network-based concepts and methodology at the dispense of theoretical ones indicates that the link between the theoretical and the methodological in space syntax has shifted to the methodological side of the relationship.

Nevertheless, confirming earlier observations (Griffiths and Netto, 2015), the analysis shows a clear 'technological turn' in the field, one in which conceptual ideas increasingly have a 'ghosted' rather than active presence. New data sources and research techniques have been incorporated into theoretical and empirical work in recent quantitative approaches in urban morphology. There is certainly a growing emphasis on spatial analysis and urban analytics driving a "new science of cities" (Batty, 2013), but are there other fields undergoing a similar trend? Exposing points of contact between space syntax and other approaches, this work can contribute to a broader assessment of urban theories within urban research.

In summary, the results from the analyses suggest that: (i) concepts and terms evolve in interdependent trajectories; (ii) ideas show evolutionary developments, with some emerging and gaining growing attention, others showing signs of stability, while others declining. This includes networks of 'canonical' concepts, which lost stability and centrality over time. This work also identified (iii) an overall decline in the use of early concepts rooted in sociology and anthropology; (iv) an apparent trend of decreasing conceptual novelty over time; (v) increasing influence by other fields; and finally (vi) strong signs of a 'technological turn', which describes a shift of the field towards the development of techniques, tools and knowledge based on technological progress. This research contributes to a unique understanding of space syntax research as a collective body of knowledge. It also introduces quantitative text analysis to urban research as a tool for the first time observing and analysing trajectories of knowledge over the years. By using new tools of analysis to approach the evolution of a particular theory, this work aims to stimulate similar explorations in other fields of urban research.

References

Alalouch, C., Al-Hajri, S., Naser, A. and Al Hinai, A. (2019) 'The impact of space syntax spatial attributes on urban land use in Muscat: implications for urban sustainability'. *Sustainable Cities and Society* 46, paper 101417. https://doi.org/10.1016/j.scs.2019.01.002

Bailey, K.D. (1990) *Social entropy theory* (State University of New York Press, Albany, NY).

Barabási, A.-L. (2013) 'Network Science', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 371, paper 20120375. https://doi.org/10.1098/rsta.2012.0375

Batty, M. (2013) *The new science of cities* (MIT Press, Cambridge, MA).

Batty, M., Morphet, R., Masucci, P. and Stanilov, K. (2014) 'Entropy, complexity, and spatial information', *Journal of Geographical Systems* 16, 363–85. https://doi.org/10.1007/s10109-014-0202-2

Benedikt, M.L. (1979) 'To take hold of space: isovists and isovist fields', *Environment and Planning B* 6: 47–65. https://doi.org/10.1068/b060047

Benedikt, M. and McElhinney, S. (2019) 'Isovists and the metrics of architectural space', in Ficca, J. and Kulper, A. (eds) *Proceedings of the 107th ACSA Annual Meeting* (ACSA Press, Pittsburgh, PA) 1–10.

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S. and Matsuo, A. (2018) 'Quanteda: an R package for the quantitative analysis of textual data', *Journal of Open Source Software* 3, paper 774. https://doi.org/10.21105/joss.00774

Bergeaud, A., Potiron, Y. and Raimbault, J. (2017) 'Classifying patents based on their semantic content', *PLOS ONE* 12, paper e0176310. https://doi.org/10.1371/journal.pone.0176310

Blumer, H. (1931) 'Science without concepts', *American Journal of Sociology* 36, 515–33.

Chen, C. and Song, M. (2019) 'Visualizing a field of research: a methodology of systematic scientometric reviews', *PLOS ONE* 14, paper e0223994. https://doi.org/10.1371/journal.pone.0223994

Coutrot, A., Manley, E., Goodroe, S., Gahnstrom, C., Filomena, G., Yesiltepe, D., Dalton, R.C., Wiener, J.M., Hölscher, C., Mornberger, M. and Spiers, J.C. (2022) 'Entropy of city street networks linked to future spatial navigation ability', *Nature* 604 (7904), 104–10. https://doi.org/10.1038/s41586-022-04486-7

Davies, P. (2019) *The demon in the machine: how hidden webs of information are solving the mystery of life* (University of Chicago Press, Chicago, IL).

Dawson, P.C. (2002) 'Space syntax analysis of central Inuit snow houses', *Journal of Anthropological Archaeology* 21, 464–80. https://doi.org/10.1016/S0278-4165(02)00009-0

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N. and Lim. W.M. (2021) 'How to conduct a bibliometric analysis: an overview and guidelines', *Journal of Business Research* 133, 285–96. https://doi.org/10.1016/j.jbusres.2021.04.070

Durkheim, E. (2014) *The division of labour in society,* Lukes, S. (ed.), Halls, W.D. (trans.) (Free Press, New York).

Fredrick, D. and Vennarucci, R.G. (2021) 'Putting space syntax to the test: digital embodiment and phenomenology in the Roman house', *Studies in Digital Heritage* 4, 185–224. https://doi.org/10.14434/sdh.v4i2.31521

Freeman, L.C. (1977) 'A set of measures of centrality based on betweenness', *Sociometry* 40, 35–41. https://doi.org/10.2307/3033543

Giddens, A. (1984) *The constitution of society: outline of the theory of structuration* (University of California Press, Berkeley, CA).

Gleick, J. (2011) *The information: a history, a theory, a flood* (Pantheon, New York).

Goffman, E. (1963) *Behavior in public places* (Free Press, New York).

Griffiths, S. and Netto, V.M. (2015) 'Open syntaxes: towards new engagements with social sciences and humanities', *The Journal of Space Syntax* 6(1), i–v.

Harris, Z.S. (1954) 'Distributional structure', *WORD* 10, 146–62. https://doi.org/10.1080/00437956.1954.11659520

Harvey, D. (1989) *The condition of postmodernity: an enquiry into the origins of cultural change* (Blackwell, Oxford).

Hidalgo, C. (2015) *Why information grows: the evolution of order, from atoms to economies* (Basic Books, New York).

Hillier, B. (1996a) *Space is the machine: a configurational theory of architecture* (Cambridge University Press, Cambridge).

Hillier, B. (1996b) 'Cities as movement economies', *Urban Design International* 1, 41–60.

Hillier, B. and Hanson, J. (1984) *The social logic of space* (Cambridge University Press, Cambridge).

Hillier, B., Penn, A., Hanson, J., Grajewski, T. and Xu, J. (1993) 'Natural movement: or, configuration and attraction in urban pedestrian movement', *Environment and Planning B: Planning and Design* 20: 29–66.

Hillier, B. and Sahbaz, O. (2009) 'Crime and urban design: an evidence based approach', in Cooper, R., Evans, G. and Boyko, C. (eds) *Designing sustainable cities* (Blackwell, Oxford) 163–85.

Hillier, B. and Shu, S. (2000) ‘Crime and urban layout: the need for evidence’, in Ballintyne, S., Pease, K. and McLaren, V. (eds) *Secure foundations: key issues in crime prevention, crime reduction and community safety* (Institute for Public Policy Research, London) 224–50.
Hillier, B. and Vaughan, L. (2007) ‘The city as one thing’, *Progress in Planning* 67, 205–30.
Javadi, A.-H., Emo, B, Howard, L.R., Zisch, F.E., Yu, Y., Knight, R., Silva, J.P. and Spiers, H.J. (2017) ‘Hippocampal and prefrontal processing of network topology to simulate the future’, *Nature Communications* 8, 1–11. https://doi.org/10.1038/ncomms14652
Koohsari, M.J., Owen, N., Cerin, E., Giles-Corti, B. and Sugiyama, T. (2016) ‘Walkability and walking for transport: characterizing the built environment using space syntax’, *International Journal of Behavioral Nutrition and Physical Activity* 13, 1–9. https://doi.org/10.1186/s12966-016-0448-9
Kuhn, T.S. (1962) *The structure of scientific revolutions* (University of Chicago Press, Chicago, IL).
Law, S. (2017) ‘Defining street-based local area and measuring its effect on house price using a hedonic price approach: the case study of Metropolitan London’, *Cities* 60, 166–79. https://doi.org/10.1016/j.cities.2016.08.008
Lefebvre, H. (1991) *The production of space* (Blackwell, Oxford).
Legeby, A. (2010) ‘From housing segregation to integration in public space: a space syntax approach applied on the city of Södertälje’, *The Journal of Space Syntax* 1, 92–107.
McCormack, G.R., Koohsari, M.J., Turley, L., Nakaya, T., Shibata, A., Ishii, K., Yasunaga, A. and Oka, K. (2021) ‘Evidence for urban design and public health policy and practice: space syntax metrics and neighborhood walking’, *Health and Place* 67, paper 102277. https://doi.org/10.1016/j.healthplace.2019.102277
Pons, P. and Latapy, M. (2005) ‘Computing communities in large networks using random walks (long version)’, *arXiv* (http://arxiv.org/abs/physics/0512106) accessed 20 June 2023).
Porta, S., Latora, V., Wang, F., Rueda, S., Strano, E., Scellato, A., Cardillo, A., Belli, E., Càrdenas, F., Cormenzana, B. and Latora, L. (2012) ‘Street centrality and the location of economic activities in Barcelona’, *Urban Studies* 49, 1471–88. https://doi.org/10.1177/0042098011422570
R Development Core Team (2018) *R: a language and environment for statistical computing* (R Foundation for Statistical Computing, Vienna) (http://www.r-project.org/) accessed 20 June 2023.
Raimbault, J. (2019) ‘Exploration of an interdisciplinary scientific landscape’, *Scientometrics* 119, 617–41. https://doi.org/10.1007/s11192-019-03090-3
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, 2426–46. https://doi.org/10.1177/0042098018795561
Sabidussi, G. (1966) ‘The centrality index of a graph’, *Psychometrika* 31, 581–603 https://doi.org/10.1007/BF02289527
Saitta, L. and Zucker, J.-D. (2013) *Abstraction in artificial intelligence and complex systems* (Springer, New York).
Shannon, C. (1948) ‘The mathematical theory of communication’, *The Bell System Technical Journal* 27, 379–423.
Shtern, M, and Rokem, J. (2021) ‘Towards urban geopolitics of encounter: spatial mixing in contested Jerusalem’, *Geopolitics*, advance online publication. https://doi.org/10.1080/14650045.2021.1926992
Soja, E. (1989) *Postmodern geographies: the reassertion of space in critical social theory* (Verso, London).
Turner, A. (2000) *Angular analysis: a method for the quantification of space* Working Paper Series (UCL Centre For Advanced Spatial Analaysis, London).
Turner, V. (1969) ‘Liminality and communitas’, in Turner, V. *The ritual process: structure and anti-structure* (Aldine, Chicago) 94–113.
Vaughan, L. and Penn, A. (2006) ‘Jewish immigrant settlement patterns in Manchester and Leeds 1881’, *Urban Studies* 43, 653–71. https://doi.org/10.1080/00420980500535146
Westin, S. (2014) *The paradoxes of planning: a psycho-analytical perspective* (Routledge, London).
Xiao, Y., Orford, S. and Webster, C.J. (2016) ‘Urban configuration, accessibility, and property prices: a case study of Cardiff, Wales’, *Environment and Planning B: Planning and Design* 43, 108–29. https://doi.org/10.1177/0265813515600120
Yamu, C., van Nes, A. and Garau, C. (2021) ‘Bill Hillier’s legacy: space syntax – a synopsis of basic concepts, measures, and empirical application’, *Sustainability* 13, paper 3394. https://doi.org/10.3390/su13063394
Zipf, G.K. (1932) *Selected studies of the principle of relative frequency in language* (Harvard University Press, Cambridge, MA).