



An Ecology of Conflicts

Using Network Analytics to Explore the Data of Building Design

Daniel Cardoso Llach^(✉) and Javier Argota Sánchez-Vaquerizo

Computational Design Laboratory, School of Architecture,
Carnegie Mellon University, Pittsburgh, PA, USA
dcardoso@cmu.edu, reivajar@gmail.com

Abstract. The scale and socio-technical complexity of contemporary architectural production poses challenges to researchers and practitioners interested in their description and analysis. This paper discusses the novel use of network analysis techniques to study a dataset comprising thousands of design conflicts reported during design coordination of a large project by a group of architects using BIM software. We discuss in detail three approaches to the use of network analysis techniques on these data, showing their potential to offer topological insights about the phenomenon of contemporary architectural design and construction, which complement other forms of architectural analysis.

Keywords: Architecture · Network analysis · Design ecology · BIM · Data visualization

1 Introduction

Network science techniques allow for the relational analysis, management, and representation of data [3]. These have been used in multiple fields including, for example, medicine and pharmacology studies [4], security and defense [5], humanities research [6, 7], and organizational management [8] among others. In the design fields, network analysis techniques have been applied, for example, to spatial analysis for architectural [9], and urban [10, 11] studies, and—more recently—to the analysis of the dynamics of product adoption [12]. Novel in our approach is the use of network analysis techniques on trace data about design conflicts reported during the collaborative process of building design and construction coordination. The data was collected semi-automatically using a custom software tool installed in the coordination logs of a group of architects using BIM software during several months [1]. Each tuple in the dataset includes information about a design conflict—a problem arising in the process of coordinating different design and construction trades, such as architecture, mechanical systems, structure, concrete, etc. The dataset includes information about each conflict’s location, the organizations involved, an index number, among others.

Design conflicts are central to the processes of design and construction enabled by Building Information Modeling (BIM). A design conflict, such as a clash between a structural column and a ventilation duct, typically arises from the clash between two

different models provided by different organizations such as, for example, the architecture and the mechanical engineering contractors. In the project documented here, the design conflicts are identified in the BIM model, documented through screenshots (see Fig. 1), and registered in a spreadsheet sometimes called “issue log.” These live documents and the conflicts they describe are central to the everyday coordination of this project—a large hospital complex in the Middle East. For detailed descriptions about this project, its coordination, and the collection of these data—as well as some earlier data visualizations—see [1, 2].

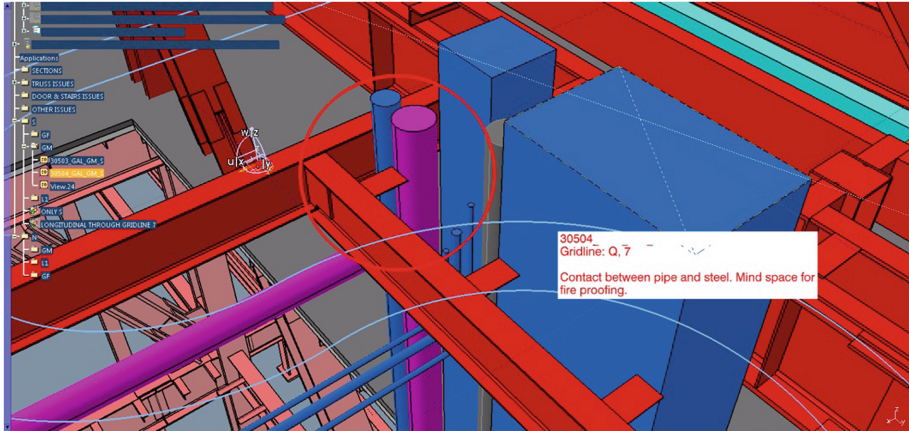


Fig. 1. A screenshot of a design conflict caused by a clash between the steel and the MEP model. The annotations indicate the conflict’s location and description.

In this paper we focus on how network analyses of design conflict data can make visible relationships between entities across dissimilar categories including concepts, spaces, materials, people, and organizations, offering what Drieger has termed ‘topological insights’ [14] on the evolution of a design process. The analyses presented demonstrate that the combination of static and dynamic networks, and text analysis, can enrich our understanding of contemporary architectural production. By documenting concrete approaches to applying network science techniques to study design conflict data, this paper offers architecture and design researchers conceptual and analytical tools to approach the increasingly complex socio-technical practices of building design.

The first analysis aggregates the conflicts into a basic static network (Fig. 2). Focusing on high-level features of each conflict (its location and the organizations involved) this analysis offers a spatial and organizational portrait of the design process through a directed-force graph. The second analysis explores the temporal dimension of the dataset, creating a dynamic network able to capture shifting trends in the data, such as the changing importance of locations and organizations over time. The third analysis uses text mining techniques on coordinators’ descriptions of each conflict, revealing how concepts (e.g. ‘beam,’ ‘steel,’ or ‘clash’) cluster along different spatial, organizational or temporal dimensions. Finally, we present a low-level semantic analysis based on Cube Analysis [15], which offers valuable insight into workflow and managerial aspects of this specific design ecology.

The following sections offer details and illustrations of three network analyses performed, and a discussion of their potentials and limitations.

2 Methods

2.1 Static Network Analysis: Tracing Actors and Features

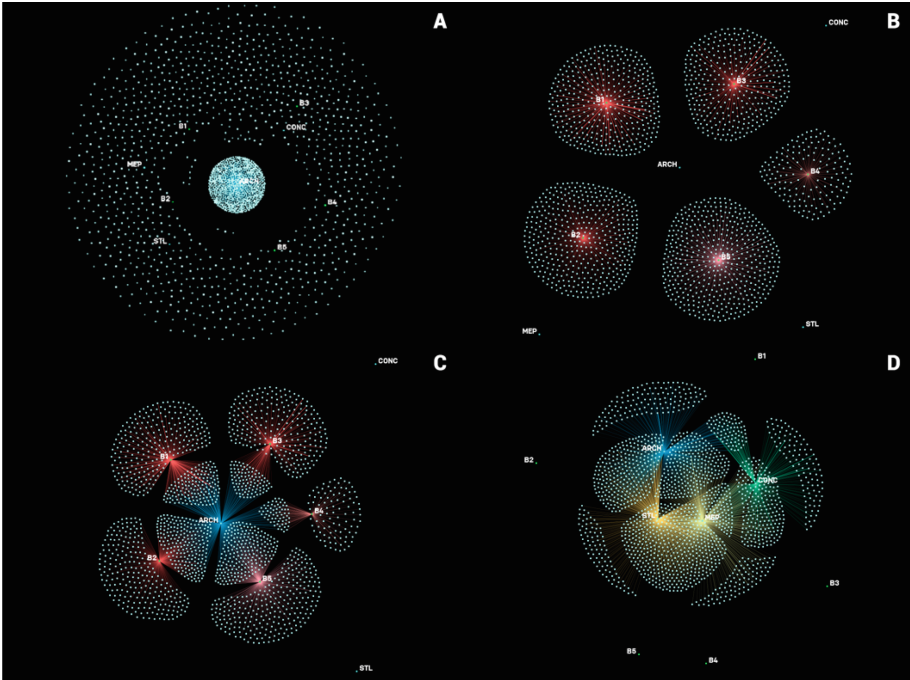


Fig. 2. Different clustering outputs depending on the activated network links: conflicts x trade with only ARCH trade links (A), conflict x location with independent clusters for each building (B), conflict x trade x location with only ARCH related trade links (C), and conflict x trade network showing different clusters depending on the related trades (D).

The first analysis is realized by constructing a simple static network based on high-level features in the design conflict data. Basic static networks are the simplest representations in network science. A common example of static networks is a network based on friendships ties in a class. The network is defined by the links between pairs of individuals. However, the network’s analytical potential is realized in its representation of overlapping and coincident connections (incidents) across the whole group. There are two approaches to static network analysis. In the first one, networks comprise nodes belonging to the same category. Their connections are built implicitly from incidents of their variables. In the second approach (our choice for this analysis), networks comprise nodes belonging to different categories [16, 17]. For example, between nodes

representing the instances and those representing the variables' values. In our dataset, design conflicts are the high-level features configuring the nodes, which are organized in relating to values such as location and organization. These values are key to the management of the architectural project. The result of this analysis is a “meta-network” [19] composed by two distinct networks: one relating conflicts and their locations, which provides information about the spatial dimension of design coordination, and another relating conflicts and the organizations involved in their resolution, which provides information about managerial aspects of the design process.

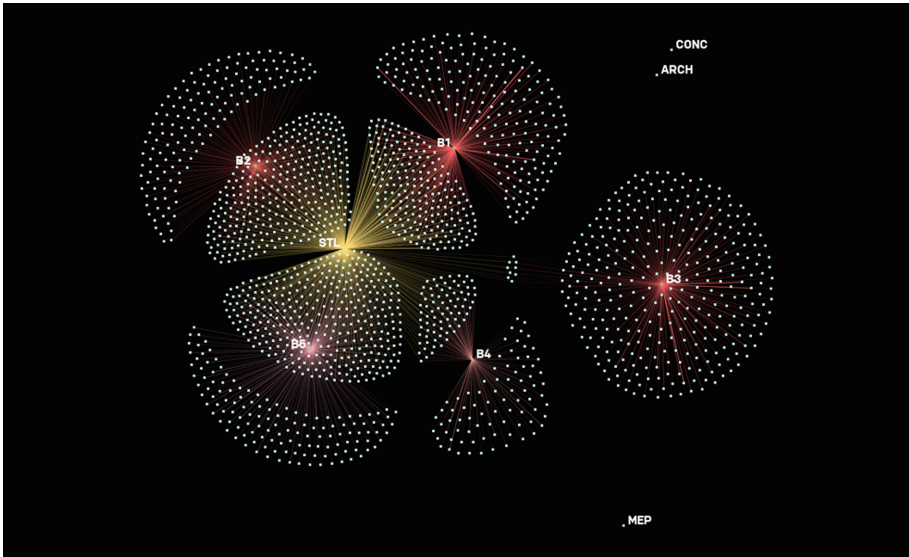


Fig. 3. Clustering of conflicts based on Steel related issues (STL) grouped by trade and location networks. ‘B’ labels indicate a specific zone and the number indicates the floor.

For example, the visualization of the resulting static network, a force-directed graph [18], creates clusters of design conflicts in relation to the variables analyzed. Force-directed graphs minimize intersections while clustering related nodes. This rendering can outline automatically clusters of data with similar features. For example, Fig. 2B renders conflicts in each part of the building in a single image. Figure 2C, by contrast, renders the relative importance of ARCH (Architecture) at each building area—thus spatializing coordination trends. Similarly, Fig. 3 offers a focused picture of the relative importance of conflicts in a specific zone of the project.

Displaying all the possible links between conflicts, locations, and trades can offer a high-level insight about coordination, but the resulting analysis can be confusing. For example, by combining 50 interconnected clusters we produced a network displaying the relative importance of different types of conflicts during the process of design coordination. However, while a high-level picture of the state of coordination is useful, conflict features are lost, and important aspects of hierarchy and centrality are glossed over—we therefore omit it. Instead, using only high-level features makes possible a clearer and more actionable analysis.

2.2 Dynamic Network Analysis: Tracing Change

Our second analysis uses dynamic networks to explore how the design process evolves over time. Dynamic network analysis adds to conventional network metrics the capacity to observe the evolution of the network’s density over time—defined as the ratio of existing links over the maximum possible number. This is a good indicator of the relative importance of each network within the whole “meta-network.” Further, the link count over time is an indicator of a network’s overall activity. Taking design conflicts’ timestamps indicating the moment when the conflict was reported by a BIM coordinator, this analysis makes visible the evolving relationships between types of conflict, locations, and trades, over their lifetime [22].

Table 1. Dynamic metrics for trades related networks

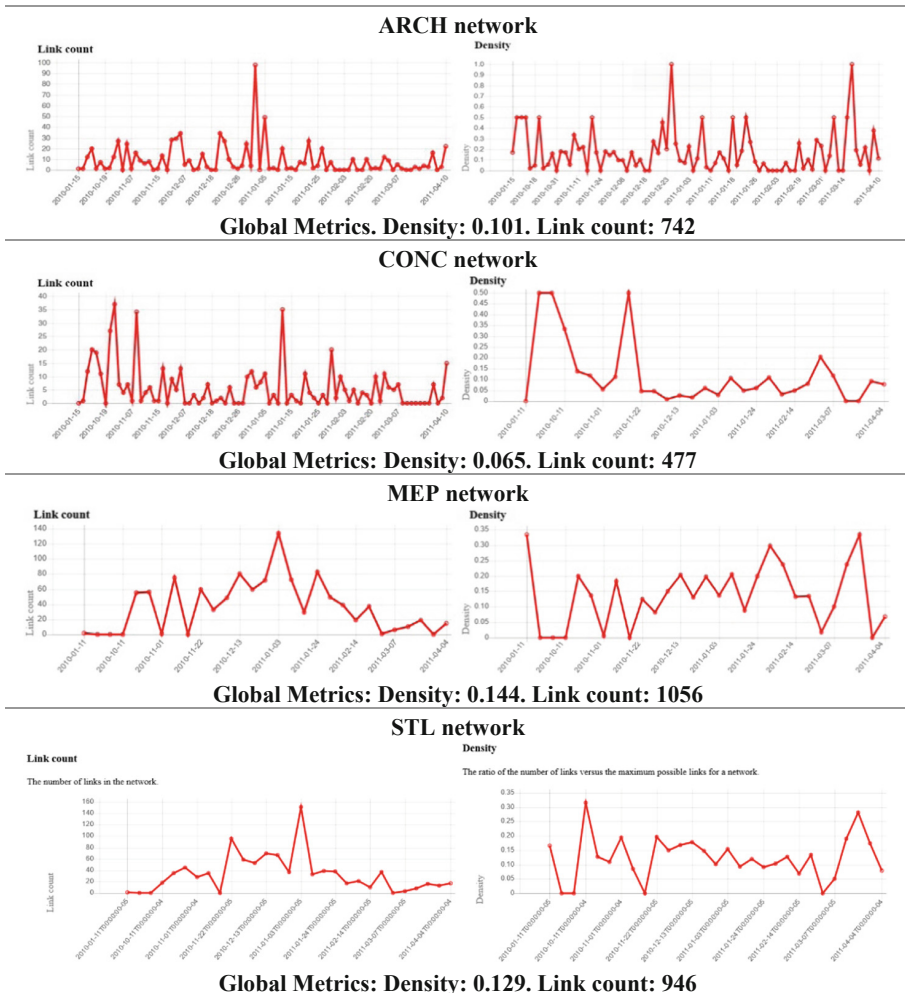
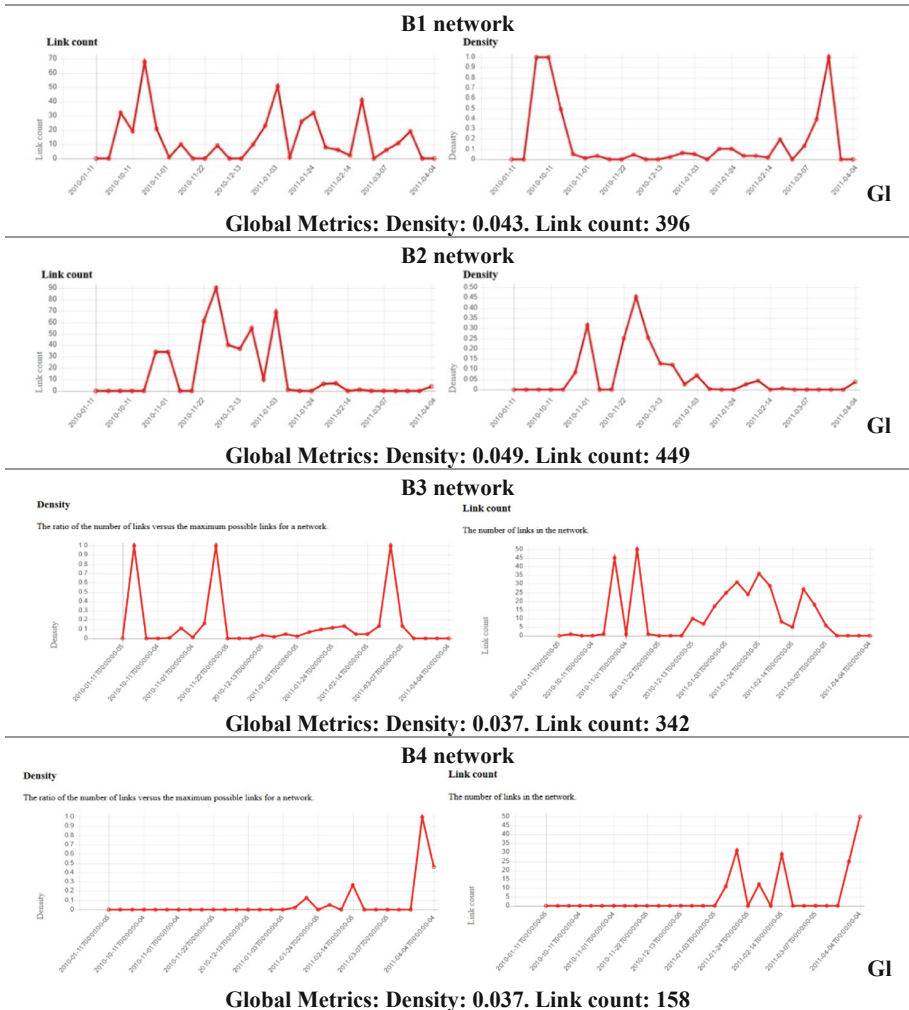


Table 2. Dynamic metrics for location-based network.



A resolution of weekly increments eases distortions caused by unusual reporting frequency on specific days. A comparison between the density of the network and its total number of links offers a perspective on the relative weight of each location and each trade at each time frame. For example, in the ARCH (Architecture) network (see Table 1) we can see ARCH related design conflicts peaking in the middle of the observation period. The coincidence of peaks across the network indicates a managerial

focus on this type of design conflict. The STL (Steel) network, in contrast, exhibits a different trend. The peak of conflict activity does not align with the peak density of steel issues in the overall network—which occupies the beginning and end of the observed period. Finally, the total link counts and density indicate that MEP (Mechanical, Engineering, and Plumbing) conflicts were the most active, and offered the most challenges to the design coordination team—a result consistent with direct ethnographic observation of this design process [13].

Comparing density and link count in trade networks (Table 1) and location networks (Table 2) offers a glimpse of their distinctive dynamic signatures. Trade networks (networks organized by the trade organization in charge of conflicts) show a higher variability in their size and density values, while location networks (networks organized by the location of conflicts) are more homogenous. The temporal patterns also differ. Salient in the trade network is the mismatch between the peaks highlighting higher-incidence of issues across the whole set. This hints at specific managerial challenges during coordination. However, when checking the temporal trends of the location of the issues, there are alignments between both temporal targets. These delineate the hierarchy and division of the work within the project—and a managerial focus on coordination progress by building, and not by work trades or packages. This outlines a specific managerial trait of this project.

2.3 Text Analysis: Tracing Concepts

The third and last analysis explores the potential of text analysis techniques to explore design coordination data. We discuss three approaches to text mining, each facilitating a different type of analysis: stems-count, text mapping, and cube analysis.

Stems-Count

Stems-count is a basic text-mining strategy that identifies the most frequent stems—roots of words—throughout the design conflict descriptions. The use of stems instead of tokens allows us to identify terms without the noise of word derivation. Using Python Notebook we tokenize each description and rid it of “stopwords” such as prepositions, and articles, obtaining 374 different roots with a typical power law distribution of natural language [20]. For legibility purposes, only the 24 stems with a frequency over 50 were selected:

```
listOfKeywords = ['beam', 'cell', 'clash', 'column', 'concret',
'corner', 'discrep', 'door', 'duct', 'edg', 'fireproof', 'fur',
'issu', 'level', 'mep', 'miss', 'model', 'partit', 'pipe',
'slab', 'stair', 'steel', 'support', 'wall']
```

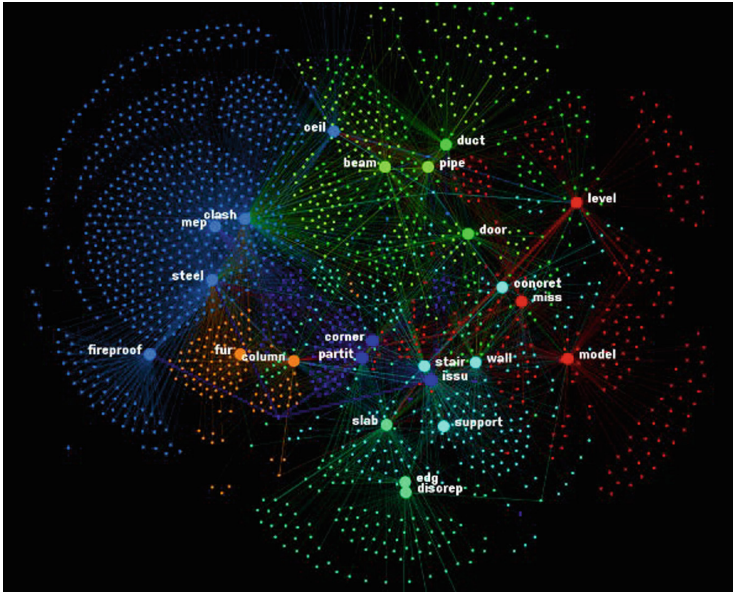


Fig. 4. Static network conflict (small nodes) x Stem (large nodes). Links colored by Louvain grouping for improving visualization. Louvain value: 0.4916405.

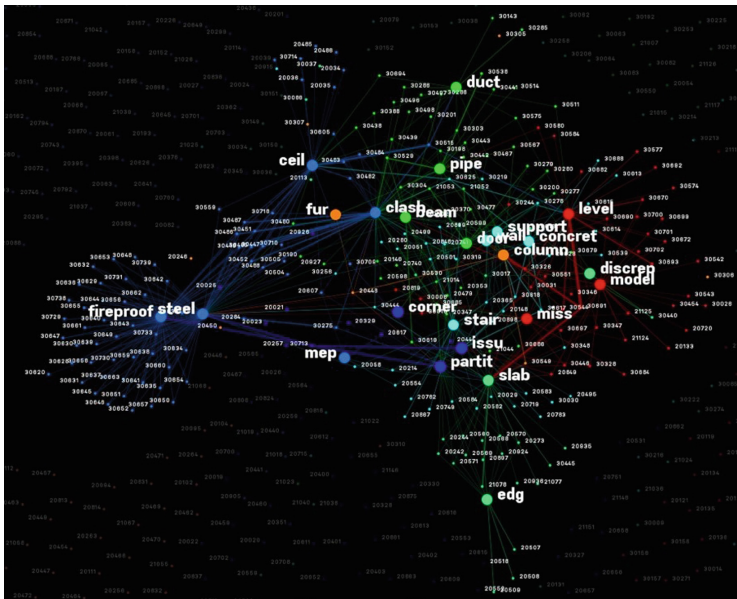


Fig. 5. Static network conflict (small nodes labelled with the ID) x Stem (large nodes). Links colored by Louvain grouping [19]. Louvain value: 0.4916405. Showing only links with an Issue_length (Issue string length) over 49 as simplification method for the network.

Following the process described in the previous section, we created a static network connecting two different classes of nodes: conflicts and stems. The resulting network was too complex to offer useful insights (Fig. 4), so we simplified it by focusing our analysis on the subset of 224 design conflicts with the longest descriptions (Fig. 5). This introduced a bias which may be avoided in future analyses by focusing instead on conflicts of a single trade, building zone, or timeframe. The resulting network shows a central core of highly correlated and common issues and conflicts comprising stems such as ‘steel’, ‘mep’, ‘clash’, ‘fireproof’, and their close periphery ‘corner’, ‘partit’, ‘ceil’, and ‘issu.’

Stem-count analysis offers visual insight into the concept-scape of this design process, mapping the relative importance of certain problems (such as clashes between steel and mechanical systems) in the process of designing this specific building. An important limitation of this analysis is the fact that the data can vary significantly from coordinator to coordinator, and from issue to issue. This reflects their different habits of record-keeping and areas of professional specialization [1]. A second limitation to this analysis is that it does not account for the relevance of each term in the network from a communicative point of view. This is a subject we explore in our next analysis.

Text Mapping

In addition to occurrence and frequency, text mapping techniques can help make visible relations across concepts and their meanings. Based on cognitive and communicative models [21], they examine a text and produce semantic networks based on these relations.

To transform the tabular original data into the text-only data, we used a Python Notebook. We then used ORA-Netmapper to map the text into two static networks: one based on the cross-classification of concepts depending on their context, meaning, and ontological category, and a semantic network based on the relationship between concepts. For the first, we mapped each relevant term to one of the following categories: “agent,” “belief,” “location,” “organization,” “resource,” “task,” or “unknown” (see Fig. 6), often used in text-mapping analyses, following the MetaOntology algorithm already implemented on ORA-Netmapper [15]. Ignoring the most common classification (“unknown”, in grey), the two categories more populated are “resource” (in turquoise) and “task” (in blue). It is important to note that this pre-defined set of categories is a limiting factor in our study, and may explain the over-classification of data as “unknown.” For the highly idiosyncratic data studied in this paper such generic settings don’t seem to fit. Getting rid of the “unknown” category, however, unveils a nuanced landscape of interrelated meanings across nodes classified as “Resource” and “Task” (see Fig. 7). Unsurprisingly, labels for trade names (“arch”, “conc”, “mep”, and “stl”) were highly ranked. This analysis reveals topics that would be overlooked by a frequency and occurrences analysis. For example, seldom mentioned terms such as “land-water-use” and “tools_and_appliance” seem relevant because of their position in the semantic network. Their position suggests that addressing the design issues underlying those terms would have significant impact on the overall design coordination of the project.

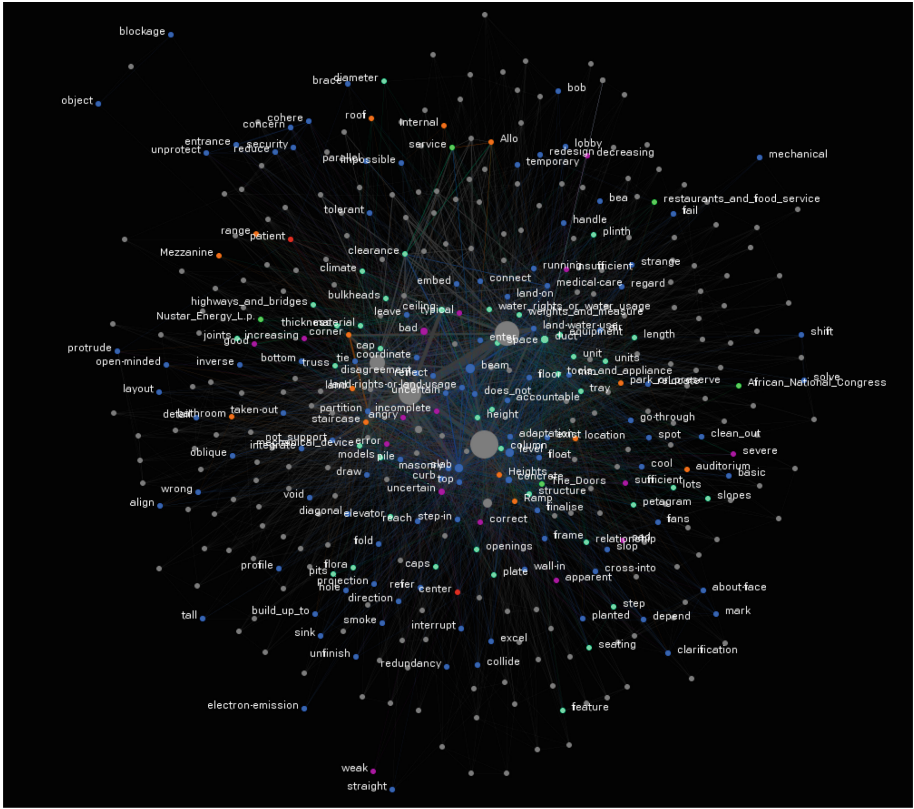


Fig. 6. Static network based on the terms cross-classification from the descriptions on the field Issue. Each color denotes a term category: red = agent, purple = belief, orange = location, green = organization, turquoise = resource, task = blue, grey = unknown. (Color figure online)



Fig. 7. Static network based on the cross-classification of terms from the field issue, only displaying nodes and links labelled as “Resource” (in blue) and “Task” (in green). Link with values over 50. (Color figure online)

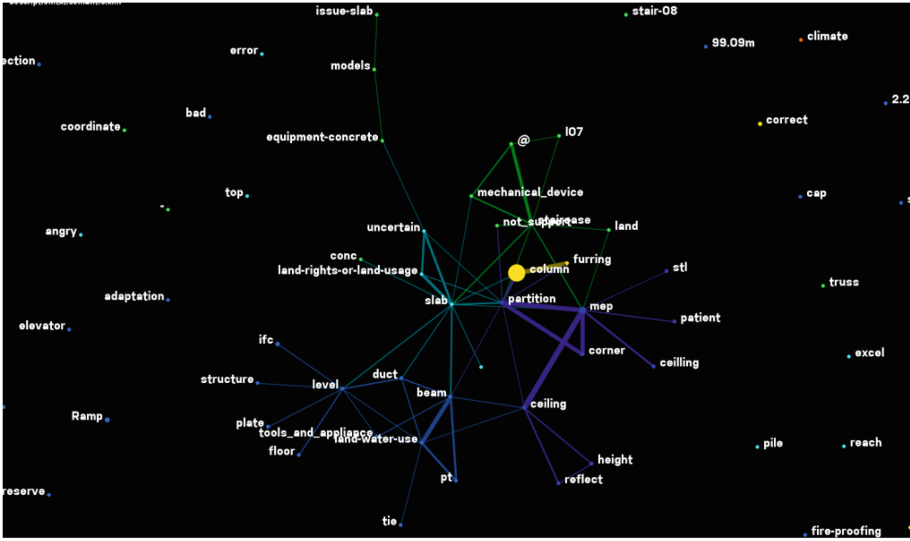


Fig. 8. Regenerated semantic static network with Louvain clustering (Louvain value 0.4557173), removing trade labels.

A second semantic network shows connectivity across concepts. It is a representation focused on knowledge, while cross-classification focuses on the role of each term. Rather than representing the whole network, we again isolate the core using Louvain clustering. Trades labels again occupy key positions, so we remove them in order to focus our analysis on less obvious connections (Fig. 8). This results in concepts such ‘column’, ‘mep’, and ‘level’ occupying a central position in the network. The remaining terms are key concepts identified only in this analysis. Those concepts are ‘false’, ‘ramp’, ‘ifc’, ‘uncertain’, and ‘park_or_preserve’. As they only appear in this analysis, it can be assumed that there are underlying factors fostered by the metrics which drive semantic relationships.

Cube Analysis

Lastly, we conducted a cube analysis of the dataset. Cube analysis is a technique that seeks to describe the communicative power of different concepts in a static semantic network. It analyzes the words in a dataset from three perspectives: consensus (frequency), betweenness (a measure of how frequent a particular node (or stakeholder) is a broker of connections among all the other nodes in the network), and total degree (the number of direct connections that a node has in a network [15]). Depending on the combination of those three vectors, concepts can be classified as ordinary words, factoids, buzzwords, emblems, allusions, stereotypes, placeholders, and symbols [15]. The presence of a word in either of these categories suggests it plays a distinct role in the conceptual scaffolding of the design process.

Table 3. The cube analysis identified different typical measurements as “buzzwords.”

Rank	Concept	Consensus	Degree centrality	Betweenness centrality
1	4350 mm	0.0002276	0.00001442	0.005
2	cross-into	0.0002276	0.00001442	0.003
3	1675 mm	0.0002276	0.00002164	0.003
4	2350 mm	0.0002276	0.00001683	0.002
5	1750 mm	0.0002276	0.00001683	0.002
6	dgmsst18	0.0002276	0.00001442	0.002
7	470 mm	0.0002276	0.00001923	0.002
8	oblique	0.0002276	0.00002164	0.002

Somewhat perplexingly, the concepts classified as buzzwords by our analysis were measurements (see Table 3). In network analysis, buzzwords are concepts with a low degree and a low consensus, but a high betweenness. Often these important topics because they appear in many lines of thought, despite not being repeated frequently, and not being necessarily connected with many other concepts. Out of context, measurements are not meaningful concepts, and yet their relevance as buzzwords indicates that they are important elements in the spatial language of the building. In this case, these measurements referred to typical proportion and distances related to the structure, installations, distances between floors, etc. Their presence as buzzwords in our dataset of design conflicts indicates that discrepancies between these measurements in plans and on the real construction was a common source of conflict.

Emblems are concepts with high consensus and betweenness values. This means that they appear frequently, and in many lines of thought—although they may not be connected directly to them apparently in the text (low total degree). The only emblem in our analysis is ‘concrete-mep’ (Table 4). This indicates the prominence of this coupling of trades in the concept-scape of the project. Even when trade tags are not considered as a source field for our analysis, conflicts between the concrete structure and mechanical systems are foregrounded by the analysis. Ethnographic descriptions of the project confirm this relative importance in the project [1].

Table 4. The “concrete-mep” concept was classified as an “emblem” by the cube analysis.

Rank	Concept	Consensus	Degree centrality	Betweenness centrality
1	Concrete-mep	0.001	0.00002164	0.006

Table 5. Stereotypes from the cube analysis on the static semantic network.

Rank	Concept	Consensus	Degree centrality	Betweenness centrality
1	ceiling	0.006	0.00025	0.00003925
2	±50 mm	0.005	0.0006587	0
3	-stair	0.001	0.000101	4.049E-06
4	detail	0.001	0.000101	0
5	stair-23	0.001	0.00007212	0
6	gridline	0.0009105	0.000101	0.00001763
7	clashes	0.0009105	0.00006731	0
8	parallel	0.0009105	0.0000625	0

Stereotypes (Table 5) are concepts with a high degree, high consensus, and low betweenness. It means concepts which are both frequent and highly connected, but not central in the network. Our analysis identifies as stereotypes some commonly used concepts such as “clashes” and “ceiling” (sic). Stereotypes don’t play a key role in the concept-scape. They can be seen as common but unimportant issues. Finally, the misspelling of ‘ceiling’ is an issue which would be addressed later by data cleaning.

On the other hand, symbols (Table 6) are the concepts with the highest values in the three parameters. They offer an initial metric for detecting important concepts in the dataset. The cubic analysis yields some coincidences with past metrics, e.g. “column,” “mep,” “ceiling,” and “level.” Others, such as “false,” “ifc,” “Ramp,” and “uncertain” were already identified using simply the betweenness centrality. There are, however, two new concepts: “beam” and “slab.” Those are key elements which have an important presence in the recorded issues but were missed previously.

Our analysis did not find factoids, allusions, or placeholders in our dataset of design conflicts. Factoids are defined by a high frequency and low levels of betweenness and total degree. Differently, allusions have a high score of total degree but low levels of frequency and betweenness. Finally, placeholders combine a high value of betweenness and total degree with a low frequency. Their absence may be explained by the unconventional style of conflict descriptions, which is utilitarian, often rushed, and thus highly economical in its vocabulary.

Table 6. Symbols in the cube analysis of communicative power on the static semantic network.

Rank	Concept	Consensus	Degree centrality	Betweenness centrality
1	column	0.035	0.002	0.253
2	mep	0.123	0.004	0.108
3	False	0.002	0.0001611	0.109
4	slab	0.045	0.003	0.035
5	ifc	0.008	0.0007933	0.072
6	RAMP	0.002	0.0001563	0.073
7	uncertain	0.022	0.001	0.051
8	ceiling	0.036	0.002	0.03
9	level	0.026	0.002	0.038
10	beam	0.034	0.002	0.029

3 Conclusion

A premise of this paper has been that the socio-technical complexity of today's design and construction industries pose challenges to researchers and practitioners interested in their description and analysis. By documenting the novel use of network analysis techniques on a dataset of design conflicts produced during the design coordination of a large architectural project, we show that computational methods of data collection and analysis offer additional tools to address these challenges, enriching our understanding of contemporary modes of architectural design and construction. Particularly, as this paper has shown, they allow us to collect and navigate larger datasets of digital traces of the design process, and to obtain high-level topological insights about them. As documented above, a combination of static and dynamic networks, and text analysis, can help uncover patterns of conflict in the design process by relating conflicts, concepts, organizations, and spaces. They may also be useful to explore comparatively how different teams—or cultures of architectural production—coordinate their efforts.

As a study of design coordination, the focus of this project is strictly interpretive and analytical—in fact, the analyses were conducted after the project was finished. However, these methods could also play a role during the design and construction of the building. Real-time topological insights about the coordination process could help design coordination teams identify and solve design problems—and problem “patterns.” This remains to be tested.

While BIM coordination processes offer relatively large and tractable sources of data about building design, it is very important to remember that these do not account for the full complexity of the project, which includes a much richer ecosystem of material practices and social interactions. Similarly, the data itself should not be understood as an inherently truthful nor neutral account of the coordination process, but rather as a collection of situated artifacts—each contingent upon the specific habits of record keeping and professional inclinations of the people and organizations defining them and collecting. Therefore, as proposed in [1], computational methods of data analysis work best in combination with other methods of qualitative observation and reflection. Accordingly, the methods presented in this paper do not aim at replacing or automating design research, but rather at expanding the repertoire of tools available to those interested in developing rich, performative accounts of the socio-technical processes of architectural design.

Acknowledgements. The authors wish to thank Dr. Kathleen M. Carley and the CASOS Group at CMU for their valuable guidance in employing network science techniques.

References

1. Cardoso Llach, D.: Tracing design ecologies: collecting and visualizing ephemeral data as a method in design and technology studies. In: Vertesi, J., Ribes, D. (eds.) *DigitalSTS Handbook*. Princeton University Press (2019)
2. Cardoso Llach, D.: *Builders of the Vision: Software and the Imagination of Design*. Routledge, London, New York (2015)

3. Brandes, U., et al.: What is network science? *Netw. Sci.* **1**, 1–15 (2013)
4. Hopkins, A.L.: Network pharmacology. *Nat. Biotechnol.* **25**, 1110–1111 (2007)
5. Arquilla, J., Ronfeldt, D.: *Networks and Netwars: The Future of Terror, Crime, and Militancy*. RAND, Santa Monica (2001)
6. Weingart, S.: Topic modeling and network analysis. *The Scottbot irregular*, 15 November 2011. <http://www.scottbot.net/HIAL/?p=221>
7. McCallum, A., Corrada-Emmanuel, A., Wang, X.: Topic and role discovery in social networks. In: *Proceedings of the 19th International Joint Conference on Artificial Intelligence, IJCAI 2005*, pp. 786–791. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (2005). <http://dl.acm.org/citation.cfm?id=1642293.1642419>
8. Wu, L., et al.: Mining face-to-face interaction networks using sociometric badges: predicting productivity in an IT configuration task. In: *Proceedings of the International Conference on Information Systems (2008)*
9. Porta, S., Crucitti, P., Latora, V.: The network analysis of urban streets: a primal approach. *Environ. Plann. B: Plann. Des.* **33**, 705–725 (2006)
10. Alexander, C.: A city is not a tree. *Archit. Forum* **122**, 58–62 (1965)
11. He, Y., Luo, J.: Novelty, conventionality, and value of invention. In: Gero, J.S. (ed.) *Design Computing and Cognition '16*, pp. 23–38. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-44989-0_2
12. Sarkar, S., Gero, J.S.: The topology of social influence and the dynamics of design product adoption. In: Gero, J.S. (ed.) *Design Computing and Cognition '16*, pp. 653–665. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-44989-0_35
13. Ammon, S., Hinterwaldner, I. (eds.): *Architecture and the Structured Image in Imagery in the Age of Modelling: Operative Artifacts in the Design Process in Architecture and Engineering*. Springer, Heidelberg (2017)
14. Drieger, P.: Semantic network analysis as a method for visual text analytics. *Procedia Soc. Behav. Sci.* **79**, 4–17 (2013)
15. Carley, K., et al.: *ORA User's Guide 2013*. ISR, SCS, Carnegie Mellon University, Pittsburgh (2013)
16. Borgatti, S.P., Everett, M.G.: Network analysis of 2-mode data. *Soc. Netw.* **19**, 243–269 (1997)
17. Latapy, M., Magnien, C., Del Vecchio, N.: Basic notions for the analysis of large two-mode networks. *Soc. Netw.* **30**(1), 31–48 (2008)
18. Eades, P.: A heuristic for graph drawing. *Congr. Numer.* **42**, 149–160 (1984)
19. Carley, K.: Smart agents and organizations of the future. *Handb. New Media* **12**, 206–220 (2002)
20. Fagan, S., Gençay, R.: An introduction to textual econometrics. In: *Handbook of Empirical Economics and Finance*, pp. 133–153. CRC Press (2010)
21. Johnson-Laird, P.: *Mental Models*. Harvard University Press, Cambridge (1983)
22. Pfitzner, R., et al.: Betweenness preference: quantifying correlations in the topological dynamics of temporal networks. *Phys. Rev. Lett.* **110**, 198701: 1–5 (2013)