Title
Flows, Bits, Relationships: Construction of Deep Spatial Understanding

Authors
Sandoval Olascoaga, Carlos
Faichney, John

Publication Date
2015-10-01

Peer reviewed

FLOWS, BITS, RELATIONSHIPS: CONSTRUCTION OF DEEP SPATIAL UNDERSTANDING

Carlos Sandoval Olascoaga
Massachusetts Institute of Technology
John Victor-Faichney
Aditazz Inc.

ABSTRACT

The number of variables acting upon urban landscapes is numerous and interconnected, closely resembling complex systems in constant dynamic transformation (Jacobs 1961). If we are to produce urban systems capable of contributing to the robustness and resiliency of cities, we ought to understand and represent the comprehensive network of actors that construct contemporary urban landscapes. On one hand, the natural sciences approach the analysis of complex systems by primarily focusing on the development of models capable of describing their stochastic formation, remaining agnostic to the contextual properties of their individual components and oftentimes discretizing the otherwise continuous relationships among parts (Hidalgo 2015, A Tale of Two Literatures).

In contrast, this research proposes an open-source web mapping and analysis tool for designers based on an alternative visual representation and the subsequent contextual construction of recursive relationships among complex layers of spatial information. The tool structures and organizes the various spatial datasets through an interconnected graph database, allowing N-dimensional connectivity across the datasets. The tool is based on a back-end system that constructs a graph-based representation of the data, and a front-end interactive web interface that enables the user to construct sets of rules to query the database and build relationships across datasets. These relationships can be examined in up to 4 dimensions through time series representations of the data. The network representation model allows both the analysis of network metrics and simulation through network models, helping us identify areas of spatial intensity within the complex systems represented through the networks. We argue that through an alternative data and visual representation it is possible to build deep relationships across datasets that contribute to a more appropriate understanding of the complex causal relationships that construct urban landscapes.

We implemented the proposed methodology in Hong Kong. By assembling and correlating a wide variety of seemingly disparate urban, sociological, ecological and tabular data, we explored the capacities of a network data representation to encounter new and serendipitous relationships between these disparate spatial constituents, presenting a more agile, deep, and emergent approach to data visualization and urban representation. The spatial relationships developed become a part of the spatial system,

creating a recursive feedback loop, where the new layers are added to a crowdsourced online repository of urban knowledge. The integration of data across levels through an alternative data representation has revealed novel, emergent relationships between different datasets and scales of information, allowing us to understand spatial constituents and its implications in the local and larger system, contributing to a growing open source repository of urban knowledge.

1 INTRODUCTION

1.1 ENTROPY AND INFORMATION

Physical systems embody information. In statistical physics, information can be thought as a measurement of the amount/volume of communication required to transmit a message about the state of a system (Shannon 1948); in this sense, systems with higher order tend to embody a higher degree of information. When the correlation among the parts of a system is high, the information embodied by the system will be highly ordered, but at the same time high redundancy within the information also embodies less information. For example, we can think about three different images of 100x100 pixels: the first image is composed by black pixels only, the second image is the picture of a human face, and the third picture is composed of completely random pixels. Although the first image has a high degree of correlation among its parts, it is also highly redundant, embodying less information. The second image embodies the most information: we can find certain correlations among its parts that help us understand the system as a whole. Finally, the third image embodies less information due to the random relationships among its parts.

In contrast, entropy could be though as a measurement of the multiple states of a system (Hidalgo 2015, How Information Grows). This could be understood as the amount of disorder existing within the possible physical states of a system: where there is a wide range of possible system states and less constraints on the system, entropy will be larger. In contrast, information describes just a particular state of a physical system.

Physical systems, including humans have the capacity to process, or compute this information, attempting to find paths that lead to physical states of a higher order or information. More importantly, humans have the capacity to find meaning in information; we compute information and establish meanings through prior and contextual knowledge, we construct relationships through ordered states of systems. Cities represent dynamic physical entities, where the flows of information are continuously transmitted and accumulated within its constituent physical systems. Their growth is only possible through the capacity of its individual constituents to compute information, both processing the information of their environment, and finding the correlations that allow them to organize themselves in states of higher order. We can therefore think of the individual components of the larger physical system not only as carriers of information, but also as systems capable of transmitting their respective information to other systems. The dynamic exchange between physical systems allows the growth of information through the interactions among systems by continuously and collectively computing with the information given.

We can augment our understanding of the growth and development of the physical system we inhabit through a process of discovery and computation of the paths that lead to higher degrees of information, allowing us to reason about the states of the system that enable this growth.

1.1.1 Information as Networks / Complex Systems

The real-world functions through the interaction and cooperation between the agents that constitute spatial networks. The collection of the many systems that make up space has been referred to as complex systems. In this sense, complicated networks of information describing the interactions among their different agents compose complex systems. Despite big differences in their constitution, the evolution and emergence of the different spatial and non-spatial networks can be described through a number of common rules and mechanisms. They are organized by common principles, making their constitution similar from each other. The understanding of the networks that constitute complex spatial systems can allow us to describe and understand the complex system itself, making its description relevant to city planning and urban design. The discovery of interactions between networks is fundamental for the understanding of the larger complex system. These relationships are not always evident; however, the exploration of their causal formation and relationships offers insights to their interactions, allowing the reasoning of the high-level workings of the system.

1.2 HUMAN REASONING

It has been proposed that human reasoning, inference and problem-solving can be constructed by activating previous mental states through multiple relationships with current contextual conditions (Minsky 1979). Memory is a key mental structure that allows human reasoning and the construction of knowledge, where the methods of representation, storage, and retrieval allow the usage of past memory data for problem solving.

Memories are created after ideas, memorable events, or successful problem solving has been experienced. When these structures of thought are created, the numerous and distinct parts and agents of the brain that were active at the time become connected to them. These mental structures can then be later activated to induce a partial mental state that resembles the original mental state that created the memory in the first place. Partial mental states can then be used to solve new problems through generalization, inference, or abstraction (Minsky 1988).

Mental structures of memory are interconnected in a complex network; learning occurs through the interaction of these networks with contextual events and problem solving. Therefore, reasoning could be thought of as a useful way to combine different fragments of knowledge and previous experiences, enabled by memory. The activation and combination of these relationships involves the original causal reconstruction and its relationship to new contextual events.

We propose that similar reasoning methods can be abstracted and implemented in the understanding and representation of the complex systems that exist in contemporary cities: building deep causal relationships among continuous visual information and spatial structures.

1.3 MAPPING TECHNOLOGIES

Digital mapping tools are primarily based on GIS technologies. As a result of the development of GIS, there is a disconnect between the technology and design thinking; the incorporation of the technology in design is difficult and superficial in best cases, mostly acting as a cartographic and presentation tool, without a direct agency and feedback loop in the design process.

These mapping tools construct relationships and representations by matching spatial datasets independently and widely, resulting in new discretized and independent layers of information (Chrisman 2006); the relationships found by GIS tools are mainly topological, and while it is possible to observe spatial effects of an interaction of the datasets, this representation lacks the appropriate depth and multi-dimensional characteristics of spatial relationships. Landscapes are a complex combination of land, nature, and society. The construction of landscapes results from the interaction of space and society. Landscapes can be understood through social engagement with the world, rather than a detachment of its multiple discretized layers. Spaces therefore become lived through the act of dwelling (Ingold 1993). The construction of landscapes is a continual process in which components are woven together into a set of relations: it is evident that digital mapping technologies fall short in the construction of representations capable of depicting the real-world.

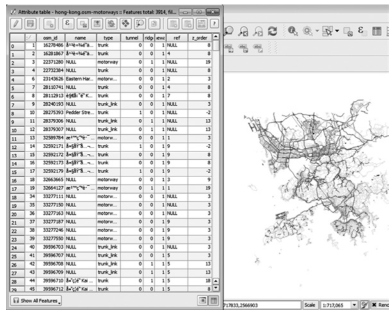

Figure 1
GIS Interface: disjoint visualization and analysis (Hong Kong).

1.3.1 Origins of Current Mapping Technologies

Resulting from the need to develop a method to represent the diverse conditions of the environment, during the 1950's and 1960's the Scottish landscape architect Ian McHarg introduced a representation method based on the use of transparent map overlays (McHarg 1995). McHarg's method aimed to represent the continuous nature of the world, utilizing visual processing to make sense of the imperfect and stratified characteristics of the landscape.

While the principles of spatial analysis and mapping can be traced back to McHarg's ecological method, the emergence of contemporary GIS technology originated in

the 1960's and the 1970's at Harvard's Laboratory for Computer Graphics (Chrisman 2006). However, the methods implemented in these early versions largely differed with McHarg's: instead of concentrating on an integrative and synthetic approach to landscape representation, the efforts at the Laboratory of Computer Graphics focused on a topological analysis of spatial layers. The divergence in methodology situated the primarily analytical technologies developed closer to the cartographical field than to design thinking. Historical accounts of the development of GIS mark this methodology shift from McHarg's as a result of the computational tools and conditions of the era, rendering it too complicated to establish an interface and a representation resembling McHarg's methods, transforming GIS into a purely analytical device (Chrisman 2006).
The GIS processing and representation model fails to both properly describe the continuous nature of the world and become a useful tool for designers beyond cartographic mapping purposes. In order to develop a method that it is not only useful for the analysis of landscape conditions, but also for its direct design implementation, there is a need for the construction of a continuous model, capable of visually conjoining the contextual landscape, the larger set of relationships, and the calculations required to synthesize spatial representations.

2 NETWORKS REVIEW

In 1961, Jane Jacobs started to think of cities as being described as "organized complexity", where the numerous spatial constituents are highly interconnected and dependent on each other. In this sense, the isolation of any of the multiple spatial variables evident in current methods of mapping and representation results in the reduction of the urban systems to a not-exhaustive representation of the world.
In this section, we propose the implementation of an alternative data representation, focusing on the construction of causal relationships through the creation of networks among spatial layers. We propose the use of a novel analytical and representation tool that structures and organizes the various spatial datasets through an interconnected graph/network database.

2.1 NETWORK REPRESENTATION

Real-world data can be represented through graphs. At the same time, analytical models have been developed to simulate and compare models with real-world networks, allowing us to abstract, simulate and identify specific conditions of the network in a controlled model. The understanding of the components and properties of the representation provides a useful roadmap for their use in developing their appropriate spatial representation.

In mathematics, graphs are used to represent networks. A graph represents entities that can be connected with each other through pair-like links or edges. The objects represented in a graph are called nodes, and they represent the fundamental unit of graphs (Barabasi 2015). Nodes can have a number of properties and are indivisible.

The relationships between the nodes of a graph are specified by edges, or connections among nodes. In more complex graph representations, edges or relationships can be directed of undirected, where the relationship is symmetric or asymmetric. Similarly edges can be weighted according to discrete relationships.

2.1.1 Network Properties

The organization of a network can be described based on its topology, or arrangement of nodes and edges. At the same time, a number of network properties that analyze the topological conditions of the network can be described. These properties can direct our reasoning and attention towards the spaces of spatial intensity within a larger landscape.

The network degree is the number of edges that are incident on a node. Nodes with a higher degree are more connected to other nodes, and can become network hubs. Network paths are a sequence of nodes in which each node is adjacent to the next one. The shortest path between two nodes is defined as the number of edges along the shortest path connecting them. Finally, the clustering coefficient is a description of the connectedness of the neighbors of a node (Barabasi 2015).

2.1.2 Relationship Among Networks

Spatial relationships in the world are the result of interactions among a variety of independent sets of conditions and spatial and non-spatial actors. When describing networks through graph relationships, it is possible to construct models of isolated networks, or models that can represent relationships among different sets or spatial layers.

Mono-partite networks are the simplest network representation. All the nodes of mono-partite networks are part of the same set of information. Road networks can be an example of mono-partite networks, where each node is road starting point or intersection with other road.

Multi-partite networks are networks whose nodes can be divided into multiple independent sets or sublayers of information (Bosak 1990). In this representation, nodes from different networks can be freely related to each other, constructing multi-dimensional relationships among its nodes. For example, if we represent a social network we could map both the relationships between users, and the social posts each user has, combining two distinct sets of nodes.

3 METHODS

Discussed here are the essential structural components of the implementation of a 4-dimensional visual mapping and analysis engine. We present the practicalities of constructing and extending the database, querying the knowledge base, and representing the data, and the relative merits of each implementation in comparison to traditional mapping tools.

The study is presented in the context of a research cluster at the Smartgeometry 2014 Conference in Hong Kong and further developed at the Computer Science and Artificial Intelligence Lab (CSAIL) at MIT. Supported by the research cluster, the authors implemented the ideas leading to the development of the tool and generated a pipeline for acquiring, constructing, and visualizing information in the urban context of Hong Kong. We initially utilized a number of spatial and non-spatial datasets, from traditional GIS information, to social media data and crowdsourced data.

3.1 GRAPH DATABASES

Traditional relational databases store information in tables, organizing data in rows and columns. In this sense, relational databases store multi-dimensional properties of different objects on a k*d matrix or table, where the number of objects in the database corresponds to k rows and their number of attributes corresponds to d columns. Every row has a unique key that can be used to combine queries between multiple data matrices or tables. Relationships between different tables can be added through logical forms of one-to-many or many-to-many. For example, a relationship of parents to sons represents a many-to-many relationship, where each son has two parents and every parent has multiple sons. While relational databases can retrieve a large number of records within a single table (wide queries), the retrieval of deep relationships among objects is cumbersome, inefficient, and requires a predefined structure among tables, leaving no space for exploratory queries and information retrieval. The retrieval of information in traditional databases is based on similarity of attributes.

Differently, graph databases represent data through nodes and node attributes. The structure of the graph database is highly flexible and allows the establishment of a multidimensional set of relationships between distinct datasets, allowing the creation of multi-partite networks. Based on the properties of the nodes or any additional criteria, relationships can be constructed arbitrarily between nodes. The edges can also be assigned properties related to the specifics of the relationships. Information retrieval is deep: by querying an object, it is possible to obtain its adjacent nodes without the need complex and cumbersome logical expressions characteristic of relational databases. We argue that this representation of information allows the understanding of the interrelated causal relationships characteristic of urban space, and an exploratory approach to spatial understanding and representation.

3.1.1 Back-end Implementation and Deep Queries

The development of an open-source, continuously growing repository of urban knowledge of Hong Kong was a major goal of the research. It became evident that for the project to live beyond its initial development, its data structure would have to be accessible and scalable through the interaction with the larger online community. We chose Neo4J as our graph database. Neo4J is currently the most popular graph database, providing a large online support community and wrappers in a number of programming languages around its proprietary query language Cypher. The objects are stored in multi-partite graphs, and they are structured as nodes, properties, and edges. The interaction with the database is possible through Cypher, which is a declarative query language. Declarative languages tend to operate based on problem descriptions and language specific implementations compared to a workflow through a set of instructions and execution states (Steinfeld and Sandoval Olascoaga 2014). Through Cypher it is possible to create and update the data-structure and query the database. Queries constructed through the database could be considered deep, as they return the set of nodes that are topologically adjacent to the node being retrieved. According to the way the relationships are structured, the queries have the capacity to construct deep causal relationships between diverse datasets. The data nodes are queried by matching a node or set of nodes though names or property patterns; it is also possible to further constrain the structure of the relationship patterns looked for by further specifying the relationships queried. The flexibility and semantic power allowed by Cypher and the database provide an excellent framework for the exploratory nature of place-making in an urban context: through a direct interaction with the user, it is possible to both visualize and compute information, discovering the relationships leading to an understanding of the network of spatial actors, while adding new relationships and layers of information to a growing urban repository.

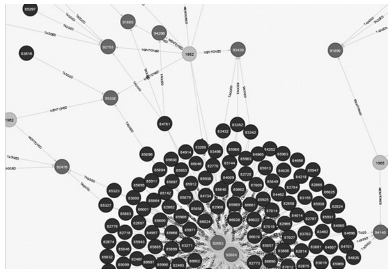

Figure 2
Sample query result of the database.

3.1.2 Relationship Construction

The advantage of using a different data representation is evident with the construction of relationships among the nodes of the dataset, and more importantly among other datasets. Relationships are constructed through a set of rules established by the user. The implementation of the visualization enabled by the graph database, synthesizes multiple layers of information in a 4-dimensional space. Based on the visual representations, the users choose to construct relationships by exploring the intersection of dataset properties, topological relationships, and/or a set of simulations and generalizations enabled by predefined machine learning techniques.

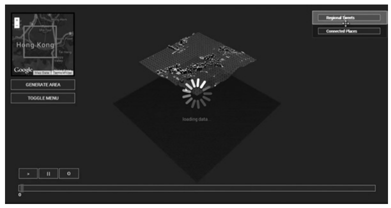

Figure 3
User interface for visual query construction.

The users interact with the database through a visual user interface, allowing them to construct queries through buttons and selections of the web-interface, as seen in *(Figure 1)*. The relationships constructed by the users can create new layers of information for the model, add new properties to the nodes, or create new edges between nodes and datasets, translating the user interaction into a continuous and crowd-sourced growth of the urban-knowledge platform.

3.2 FRONT-END WEB INTERFACE

The front-end is an AngularJS app that leverages Angular's unique bi-directional data binding to synchronize the data displayed and edited by the UI with data on the server. The primary component of the UI is a 4D urban scale visualization initially constructed by compiling data from multiple datasets.

The dataset from the graph database is fed through a series of Javascript classes responsible for generating custom visualizations. The visualization takes place on a HTML canvas using ThreeJS. A series of interface widgets and client-side data structures allow the modification and construction of database queries, as well as the presentation of data through different ranges of the time-series.

4 DIMENSIONAL WEB ENVIRONMENT

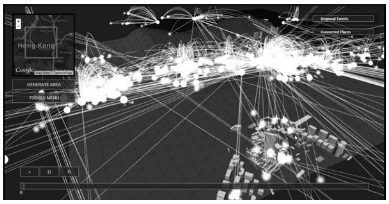

Figure 4
Initial time sequence.

Landscapes are not only land, nature or space; rather they ought to be considered stories that unfold overtime. The construction of landscapes is the result of a continuous process of an interaction between people and their surroundings: a synthesis of time, space, and experiences (Ingold 1993). A way to construct causal relationships is to map the transformation of urban spaces through time: in order to expand the exploratory capacities of the platform, we built a time-based visualization and analysis feature for the datasets and relationships. It is common for this web spatial datasets to have a timestamp associated with them, adding the capacity to visualize the dynamic and ever-changing spatial relationships. Through the interaction with a time slider, the interface can present the time-based iterative addition of datasets and relationships to the visual representation in an aggregated or isolated manner.

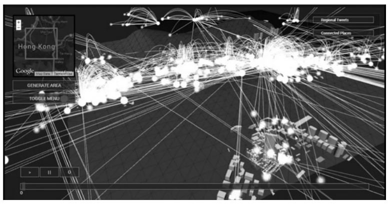

Figure 5
Final time sequence, aggregated data.

Constructing Queries and New Layers through Iterative Rules

Once a query has been constructed based on a given rule (for example, through unsupervised clustering, relate all the tweets that are categorized as food), it can be added as an additional relationship or information layer to the model. In this sense, the new relationships between datasets become part of the multi-dimensional graph, and can be added to the synthetic visualization that providing a holistic view of space. When a new query is constructed, the new layers of information that have been found through visual calculation will now be obtained as an adjacent node or relationship that has been added to the graph.

Semantic Analysis

Our use of non-traditional and dynamic datasets allows us to inquire into the socio-spatial methods through which we construct meaning, and therefore space. Semantic analysis enables us to relate traditional and non-traditional datasets through the analysis and differentiation between the syntactic structures of text. Probabilistic methods allow the approximation and generalization of concepts in large sets of socio-spatial textual datasets that can be easily obtained throughout the web, enabling the creation of relationships among them. The tool implements supervised and unsupervised learning algorithms to perform basic semantic analysis, like sentiment analysis, parent-child classification and trending topics in the datasets, tracing their social and spatial implications in urban landscapes.

We used topic modeling to discover the abstract "topics" that occur in a collection of social media posts. Topic modeling allows the examination of a set of documents and the discovery of its topics and their balance, based on the statistics of the words in each document. We used Latent Dirichlet Allocation (LDA), an unsupervised learning algorithm (Blei et al. 2003). The algorithm allows the dataset to be divided into a number of subgroups of statistically common topics.

Additionally, the platform uses a sentiment analysis algorithm, to determine the overall contextual polarity of a given textual dataset. The polarity result judges or evaluates the affective state, or the intended emotional communication. We implemented a Naive Bayes Classifier (Russell and Norvig 2003), a probabilistic classifier that was previously trained with a corpus of over 2,000 document rating classifications. Through an exploratory analysis of the polarity or sentiments of the datasets throughout the urban landscape, users are able to construct sentiment thresholds that enable them to relate datasets with similar semantic characteristics.

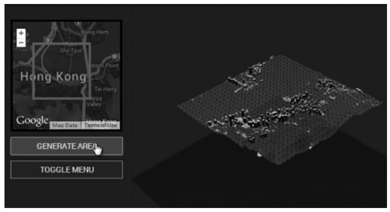

Figure 6
Base map construction.

The implementation of simple classification algorithms allows the users to develop semantic relationships of seemingly hidden properties among the datasets. Through the iterative process of constructing new layers of information, continuous new visual representations of the city are developed that enable us to compute and find the serendipitous organization of information.

5 IMPLEMENTATION

The system was initially built and implemented in Hong Kong, a city characteristic for its dense interconnected public spaces. While surveying conventional datasets of the city, it became evident that these were insufficient to represent the social complexity and convey a proper analysis of its dense urban fabric. In order to explore methods of representation capable of providing a more holistic view of the city, we explored non-traditional datasets that could be added to the alternative data representation being proposed.

5.1 BASE MAP CONSTRUCTION

The web interface allows the construction of a 3D base map; instead of pre-loading an entire region from SHP files, the user selects the area of interest based on an interactive web map. A base Google map is presented on a smaller window, where the user selects a bounding box to query. The geographical vertices are used to query the Open Street Maps API and obtain a 3D urban context; the JSON obtained from the API is parsed by the application and used to create three.JS 3D mesh objects. The Google maps API is used to construct the site terrain; data points with elevation are returned by the API and used to construct an interpolated three.JS 3D mesh.

5.2 GIS

Conventional GIS files were obtained from the area. The SHP files are processed with a geoDjango application, employing the geo-spatial library GDAL to project them into a consistent geographical projection and translated into JSON files that can be served and queried by the web application. The datasets can be overlaid in a synthetic manner on top of the 3D base-map previously constructed.

5.3 DATA SCRAPPING

The use of dynamic information describing the social interactions and networks present in the area was a central aspect in the construction of a diversified network representation. A set of python scripts was developed to scrape a number of social media APIs. Over 500,000 geo-located data points were obtained from Twitter, Foursquare, Flickr, Instagram, and Facebook. The scripts also parsed the responses obtained from the APIs into readable dictionaries that made it easier to obtain the relevant information and stored in the server as consistent JSON files. While the data APIs used in the project do not publicly release either all their data or live data, a number of similar scripts were developed to obtain on demand data that is generated on an almost real-time manner.

Social media has increasingly acquired social and political power. At the time of the development of the project in Hong Kong, student protests were beginning to pick up steam in the area: through the use of social media, we attempted to discover areas of social and political intensity throughout the city. Our analysis sought to investigate the relationship of the protest and counter-protest communities to specific geographic boundaries, attempting to construct relationships between these conditions and other spatial and non-spatial characteristics of the areas.

5.4 RESULTS

5.4.1 Query Construction

The construction of a UI to effectively construct queries proved to be a challenge. On one hand, the visual representation motivates the exploration of relationships, on the other hand, the queries submitted to the database need to be specific to the objects and their properties within structured language syntax.

In the current implementation of the tool, we use two different methods to construct queries. First we built a visual interface that allows the selection of datasets and properties through radio buttons, attempting to match datasets through the selections. Second, it is possible to query adjacent nodes to a given node or dataset, and it is possible to refine the search by returning the shortest paths among nodes, returning the nodes that are more closely connected to the queried object. In order to easily construct queries from unstructured language to structured database syntax, there is a need to build a natural language processing engine into the system.

The users have been particularly interested in creating semantic and spatial relationships between social media datasets. Through the construction of time-based queries, they have inquired into the circadian rhythms of urban spaces. Similarly, users have been interested in visualizing spaces that share similar conversation topics, or conversation moods, relating additional spatial datasets to make assumptions of spatial characteristics contributing to such social events.

5.4.2 Eastern Semantic Analysis

The original implementation of the semantic analysis algorithms, in particular the sentiment analysis algorithm relied heavily on western language. For the initial implementation, we were only relying on English datasets; however, it became evident that the English datasets were highly contextualized to Eastern expressions and meanings. In order to obtain a more nuanced description of the meanings and relationships behind textual datasets, a database of Eastern phrases was crowdsourced and added to the original training set. When running the queries, it is possible to explore their construction through both the western and eastern-based semantic analysis.

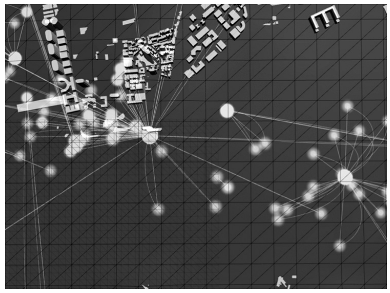

Figure 7
Tweets related to specific urban events.

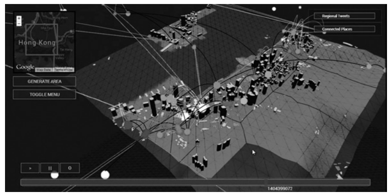

Figure 8
Western sentiment relationships.

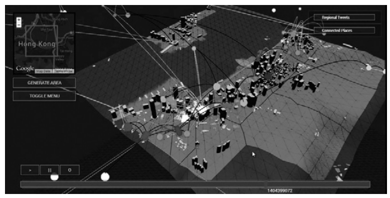

Figure 9
Eastern sentiment relationships.

The algorithm implemented relies on the statistical similarities of the data to a large corpus words. In contrast, linguistic dialects localize expressions and communication, revealing the potential of understanding socio-political and cultural particularities of a place. While we acknowledge the value of data locality, our current analysis methods cannot account for these nuances in the data, but future explorations are being carried to derive methods capable of harnessing the local and temporal nuances of written language.

Upon reflection, the case study yields a set of potential contributions for the use of the tool in an urban mapping and urban design context. These methods are impactful to design pedagogy and practice, both as a way for analyzing and understanding the interwoven relationships between spatial constituents, and as a method to reflect on the implications of design decisions in the larger networked system.

6 CONTRIBUTIONS

The research suggests an alternative data representation that creates an alternative visual representation of information. Through a graph representation of the data, it is possible to explore the use of multi-partite networks to relate datasets in an N-dimensional space, modeling real world relationships holistically. Our tool allows users to visually interact with the data, ask questions, and develop rules through constrains. The tool enables creation of crowd-sourced relationships among the datasets contributing to a growing platform of urban networked urban knowledge. Moreover, the use of a network to represent a complex spatial system is advantageous for easily understanding complex relationships that would be otherwise hard to trace. The network representation model allows both the analysis of network metrics and simulation through network models, helping us identify areas of spatial intensity within the complex systems represented through the networks.

The methodology presented by this study starts to suggest a hybrid analysis approach, combining bottom-up analysis and construction of information layers with a top-down reasoning and goal fulfillment process. The top-down inference approaches landscape by exploring high-level questions and finding deep relationships among global datasets. We explore a question by analyzing the deep causal relationships within different nodes of different datasets. We argue that the construction of a causal representation model is closer to some of the same principles that allow intelligent reasoning and inference, providing the opportunity to develop an understanding of deep relationships among complex systems, facilitated by a personal exploration through a growing dataset. The visual conjunction of information, constraints and causal relationships allows designers to think about the implications of design decisions within a larger complex system.

The interdisciplinary area of inquiry that has been suggested by the work presents numerous potentials within visual representation, analysis systems, and information management for urban representation. Facilitated by the use of a graph database, the integration of data across has revealed novel, emergent relationships between different datasets and scales of information, allowing us to understand spatial constituents and its implications in the local and larger system, providing a growing open source repository of urban knowledge.

7 FUTURE WORK

The current implementation of the interface mostly employs the graph representation as a data storage structure. While steps have been taken to incorporate the use of network properties and analysis to construct additional relationships among the datasets, work is left to fully implement the network analysis algorithms and expose them in the user interface to allow the construction of queries and subsequent relationships among the datasets.

In addition to the machine learning algorithms previously mentioned, we are starting to experiment with the use of unsupervised clustering algorithms to characterize spaces with similar characteristics (both obtained by the data, and constructed by the users). We are hoping to provide visual feedback to users of the different communities of similarities to allow the exploration of urban characteristics by creating inferences through the observed similarities. In addition, the current implementation of the web visualization closely reflects the structure of the database. The development of alternative methods for representing the data relationships is a major area for future development.

Given a set of heterogeneous relationships within large systems, the individual discrete datasets can become decontextualized and lose their heterogeneity. The current algorithmic implementations for analysis, processing, and description of the datasets can simplify the datasets through the development of the model. Our interest in using graphs to model social media and urban data stemmed from the comparatively robust support for creating heterogeneous or ad-hoc relationships between data, something which is far more cumbersome to implement using other data modeling and storage techniques. Through the methods developed for this research, we are hoping to harness our visual apparatus and human reasoning to construct representations capable of describing the world in a more appropriate manner, eventually developing methods capable of complementing the shortcomings of algorithmic data analysis.

Finally, while the proposed model already suggests an alternative analysis of urban conditions, the capacity to produce strong and deep causal relationships is currently dependent on the datasets available. However, with the increasing ubiquity of data collection, we are expecting to incorporate additional sources capable of complementing our spatial perception and our capacity to construct urban stories through deep relationships.

ACKNOWLEDGEMENTS

Part of the work presented was developed through a research cluster at the Smartgeometry conference 2014 in Hong Kong and it was made possible through the work of Carlos Sandoval Olascoaga (Cluster leader), John Victor-Faichney (Cluster leader), Scott Ewart (Cluster leader), Kyle Steinfeld (Cluster leader), Matthew Shaxted, Plamena Milusheva, Richard Maddock, William Haviland, and Benjamin Coorey. Further development of the work took place at the Computer Science and Artificial Intelligence Laboratory at MIT and was made possible through the support of the Jumex Foundation of Contemporary Arts, and the National Council of Science and Technology.

REFERENCES

Barabasi, A. (2015). Network Science. Cambridge, UK: Cambridge University Press.

Blei, D., A. Ng, and M. Jordan. (2003). Latent Dirichlet Allocation. Journal of Machine Learning Research 3: 993-1022.

Bosak, J. (1990). Decompositions of Graphs. Mathematics and its Applications (Book 47) Springer.

Chrismas, N. (2006). Charting the Unknown: How Computer Mapping at Harvard Became GIS. ESRI Press.

Hidalgo, C. (2015). How Information Grows: The Evolution of Order from Atoms to Economies. New York: Basic Books.

Hidalgo, C. (2015). A Tale of Two Literatures: Social Networks and Network Science. Unpublished draft.

Ingold, T. (1993). The Temporality of Landscape. World Archaelogy, Conceptions of Time and Ancient Society 25(2): 152-174.

Jacobs, J. (1961). The Death and Life of Great American Cities. New York: Random House.

McHarg, I. (1995). Design With Nature, 25th Anniversary Edition. New York: John Wiley & Sons.

Minsky, M. (1988). The Society of Mind. New York: Simon & Schuster.

Minsky, M. (1979). K-Lines: A Theory of Memory. MIT AI Lab, Memo 516.

Russell, S. and P. Norvig. (2003). Artificial Intelligence: A Modern Approach. New Jersey: Prentice Hall.

Shannon, C. (1948). A Mathematical Theory of Communication. Bell System Technical Journal 27(3): 379-423.

Smartgeometry. (2014). Urban Compaction: Smartgeometry Workshop and Conference, Hong Kong, July 14-19. Accessed April 20, 2015, *http://smartgeometry.org/*

Steinfeld, K. and C. Sandoval Olascoaga. (2014). Imperative/Functional/Object Oriented: An Alternative Ontology for Programming Languages in Design. In. ACADIA 2014: Design Agency, Proceedings of the Association for Computer-Aided Design in Architecture 2014 International Conference, Los Angeles, CA, October 23-25.

CARLOS SANDOVAL OLASCOAGA

Carlos Sandoval Olascoaga is an architect who specializes in computational design. Before joining the Design and Computation Group at MIT, he was a research scientist at UC Berkeley, investigating the convergence of design, data science, and social sciences. In the past four years, Carlos has been a Lecturer at UC Berkeley and a Visiting Professor at UNAM, and has taught computational design seminars and workshops in the United States, Italy, China and Mexico. His work has been supported by numerous fellowships, more recently by the IDEA Studio Fellowship at Autodesk, the National Council of Science and Technology, and the Jumex Foundation for Contemporary Arts. He is interested in urban data visualization, computation in design cognition, and technological platforms for design collaboration. Carlos holds a Master of Architecture from UC Berkeley, and a Bachelor of Architecture with Distinction from UNAM in Mexico.

JOHN VICTOR-FAICHNEY

John Victor-Faichney is a designer, educator, and software developer who lives in San Francisco. John is currently a member of the technical staff at Aditazz, a technology firm in San Francisco that implements computationally enabled scenario-based planning for the design of buildings. Prior to this, he was an Associate at Fernau & Hartman Architects. John has been a Lecturer at the University of California, Berkeley teaching computational design and GIS. He has also offered workshops at Diablo Valley College, the 2012 Association for Computer Aided Design in Architecture (ACADIA) conference at the California College of the Arts (CCA), and the 2014 Smartgeometry Conference in Hong Kong.