Literature Review 2

Computer Graphics I

Zubin Bhuyan (Student id: 01744486)

Paper 1: Chen W, Huang Z, Wu F, Zhu M, Guan H, Maciejewski R. (2017). **VAUD: A Visual Analysis Approach for Exploring Spatio-Temporal Urban Data**. I*EEE Transactions on Visualization and Computer Graphics*, Issue- 99. Doi: 10.1109/TVCG.2017.2758362

Paper 2: Rachel Shadoan, Chris Weaver. 2013. **Visual Analysis of Higher-Order Conjunctive Relationships in Multidimensional Data Using a Hypergraph Query System**. *IEEE Transactions on Visualization and Computer Graphics* Vol.19, Issue 12 (December 2013), Pages- 2070-2079. DOI=http://dx.doi.org/10.1109/TVCG.2013.220

The challenge of analyzing multi-dimensional data increases with the dimensionality; this is especially true for identifying relationships which exist among the attributes. Additionally, the representation framework must allow for the representation of all the patterns and associations which exist across dimensions. The work presented in [Paper 2] makes use of hyper graphs to construct queries and explore n-ary conjunctive multi dimensional relationships. In their representation, a hyper-graph is composed of nodes and hyper-edges. Nodes represent subset of values, and hyper-edges represent conjunctive relationships.

The query languages proposed by Shadoan, et al. allows the explicit construction of queries which can be applied to categorical and categorically reducible dimensions of a dataset. They study the types of relations expressible by applying cross-filtering on attribute relationships, and also provide a modular approach for construction of visual graph query. They provide an example using the "Electronic Enlightenment" (EE) dataset. The EE dataset is a collection of more than 58000 letters and information of about 7000 senders and recipients, city of origin, language and other relevant information. This database has a number of many-to-many relations which makes it ideal candidate to test their *interactive visual tool*.

Table 1. Relationship Degree

Degree	Example	Value	Description
Unary (1)	Author Name	Voltaire	Voltaire wrote letters
Binary (2)	Author Name, Language	Voltaire English	Voltaire wrote in English
Higher-order (3+)	Author Name, Language, Source Country, Destination	Voltaire English France England	Voltaire wrote in English from France to England

Table 2. Relationship Connectivity

Connectivity	Example	Description
One-to-one	Person to Birth Date	A person is born on one date
One-to-many	Person to Language	A person may speak many languages
Many-to-many	Language to Country	Many languages may be spoken in a given country, and many countries speak the same language.

Logical association between attributes of a relation is called the *logical basis* of the relation. The number of attributes linked to a relationship is called its *degree*. Table 1 illustrates three types of degrees. Similarly, *connectivity* is the type of mapping that exist among attributes of a relation (as shown in Table 2). Defining such metrics enable the analysis of large multidimensional datasets by studying the subsets of the data.

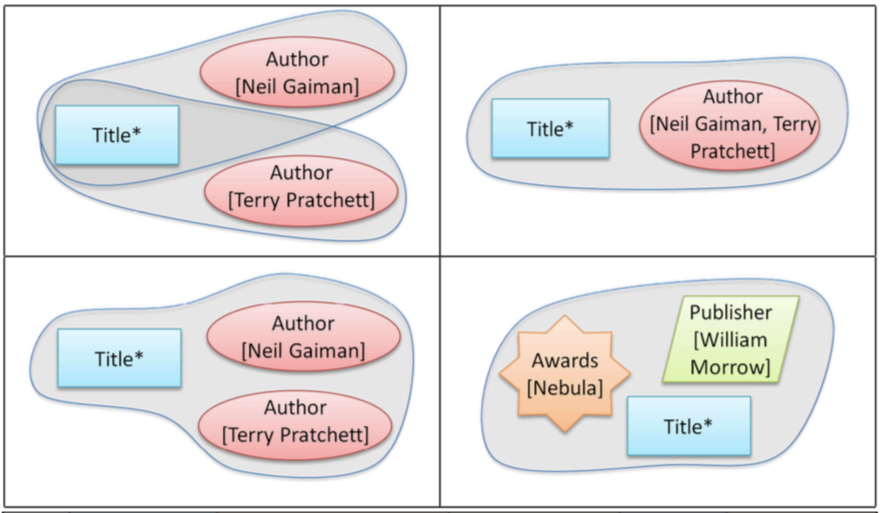
HYPERGRAPH QUERYING LANGUAGE AND MATCHING QUERIES

The implementation of the querying and representation in [Paper 2] is done in a graph theoretic approach. Hyperedges and nodes form a hypergraph, which is a query graph. More formally, a query graph $q = (N, H, T, V, f)$, where N is the set of nodes, H is the set of hyper edges, T is the set of attribute types, V is the attributes (values), and f is the mapping $T \rightarrow V$.

If *d* is a surjective mapping $d: T \rightarrow f(t)$, a query graph *q* can be compared to *d* based on the following rules:

- Given that $n = (t, U)$ is a node, $n \in N$, n *matches* d $\Leftrightarrow U \cap d(t)$
- For a given hyper graph $h \in H$, h *matches* d $\Leftrightarrow \forall\, n \in h$, d *matches* n
- Query q *matches* d $\Leftrightarrow \forall h \in H$, d *matches* h

There are two phases of the query matching algorithm. First, the mapping on the data is set up and this is performed only once. Then the hyper graph is evaluated by linearly scanning the table. This phase is performed every time a query is performed.



ID	Title	Author	Publisher	Date	Awards
1	American Gods	Neil Gaiman	William Morrow	09/02/2003	Hugo, Nebula
2	Neverwhere	Neil Gaiman	William Morrow	09/02/2003	
3	Good Omens	Neil Gaiman, Terry Pratchett	Guild	5/1/1990	
4	Neuromancer	William Gibson	Ace Books	7/1/1984	Hugo, Nebula
5	Going Postal	Terry Pratchett	HarperTorch	9/27/2005	Nebula
6	Xenocide	Orson Scott Card	Tor Books	8/15/1992	Hugo

Fig. 1 Examples of hypergraph queries on a small dataset of books.

Examples of hyper-graph queries is given in Fig. 1. The top right and top left queries are equivelt and expresses the question, “What books have been written by Neil Garmin OR Terry Pratchett?” The bottom left query expresses “Which books were written by Neil Gaiman AND Terry Pratchett?” The bottom right query expresses “Which books published by William Morrow won Nebula awards?”

TIME COMPLEXITY

The worst case time-complexity is $O(m(2^n-1))$, where n is the number of nodes and m is the table dimension in the query. This case is for a fully connected graph which has hyper edges between every possible subset of nodes. The authors mention that the average time is much better, but does not give specific details. They further mention that, the worst case scenario is highly unlikely to occur.

EXPRESSIVITY OF HYPERGRAPH QUERIES

The attribute relationship graphs used in the hypergraph query system facilitate the exploration of n-ary, many-to-many, inter/intra-dimensional relatoonships. The task of higher-order query development is simplified by this process as it compresses the query expression into one single view. The authors plan to extend the expressivity by allowing designation of equivalent nodes and the ability to define nested queries.

When compared with classifiers which deal with a single type of task, graph based representations have clear advantage as it is able to querying over complex knowledge bases and answer more complex questions. A similar work in this direction is done in [Paper 1]. Chen, et. al. in their implementation allows multi-source data, which is more heatergenous in nature. Their approach is very much different than that of [Paper 2] (even though both work may be categorizes as visual query systems).

VAUD QUERY MODEL

Visual Analysis Approach for Exploring Spatio-Temporal Urban Data (VAUD) is a query model which can perform cross-domain correlations and deductions from multiple sources. It also has a visual analytics framework. Their data used in their implementation is citywide urban data.

- Data Description:
  Anonymized data collected in a city of 14 million citizens is used to demonstrate their systems capabilities. The dataset contains information such as road networks, coordinates and information about points of interests and real estates such as type of property, price, year of construction, etc. Additionally, it has over 300 billion locations records of mobile phone usages, social media and microblogging information, taxi GPS trajectories and taxi driver details.

DATA REPRESENTATION

A primary aim of this work is to effectively manage data objects and their connections. Data objects such as people, taxis, etc., and the relations that exist among them can be inferred using muktiple sources of data. All objects are allowed to have four attributes: *identification*, *spatial*, *temporal* and *descriptive*. Some of the relateonships which exist among the objects are *pre-built*, and some are extracted in run-time.

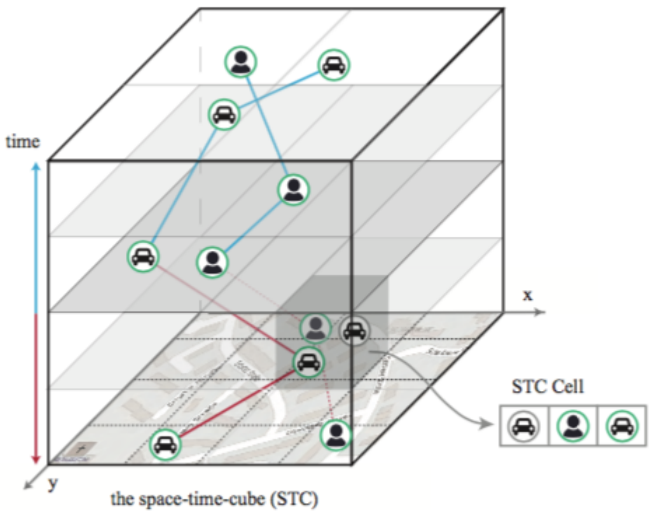


Fig. 2 Spatio-temporal cube representing two cars and a person

The time-period of the data is sliced into smaller periods. The idea of **space-time cubes** (STC) is used to represent spatio-temporal aspects of the objects. STC is a 3D grid where X-Y plane represents geographic area and the Z axis represents time. Fig. 2 is an illustration of a scenario which of two cars and a person being represented by a STC.

QUERYING

Data fusion from multiple sources and cross-domain queries calls for structures and clear query specifications. This is why all queries are actually a sequence of simpler atomic expressions. An atomic query has 3 parts: query condition, operation and result. (Fig. 3(a)) Further, an atomic query combination can be a *union*, *intersection* or *complement*. An extraction indicates that some part of an objected has to be extracted. It is composed of a query result, extraction operation and objects' components. These components which have been extracted can be further be used as atomic queries for new conditions. Complex tasks can be described using compound queries and extractions. Another important point to mention here is that mode of queries can be source, destination and multi-source mode.

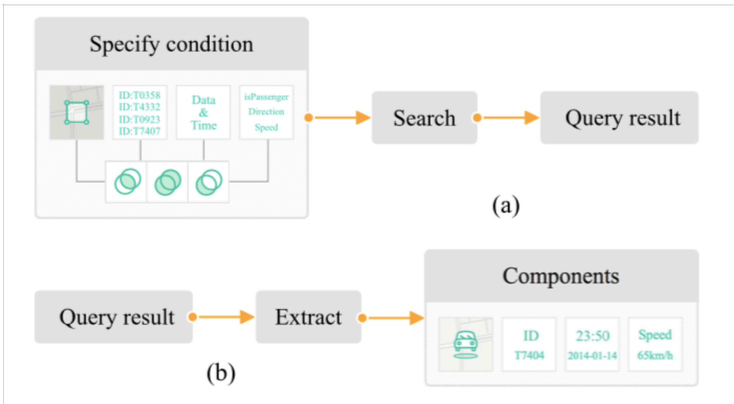


Fig. 3 (a) Atomic query, (b) Extraction

The VAUD systems scene view gives spatial and attributes information and the query view allows an interactive process for query generation and visual reasoning. VAUD is an excellent tool to analyze cross-domain data sources and perform reasoning over it to mine interesting patterns and results. The usability of the query generation system can be further enhanced by performing preliminary graph-theoretic analysis to show interesting directions of data exploration.

===============