Association Analysis for Visual Exploration of Multivariate Scientific Data Sets

Xiaotong Liu, and Han-Wei Shen, *Member, IEEE*

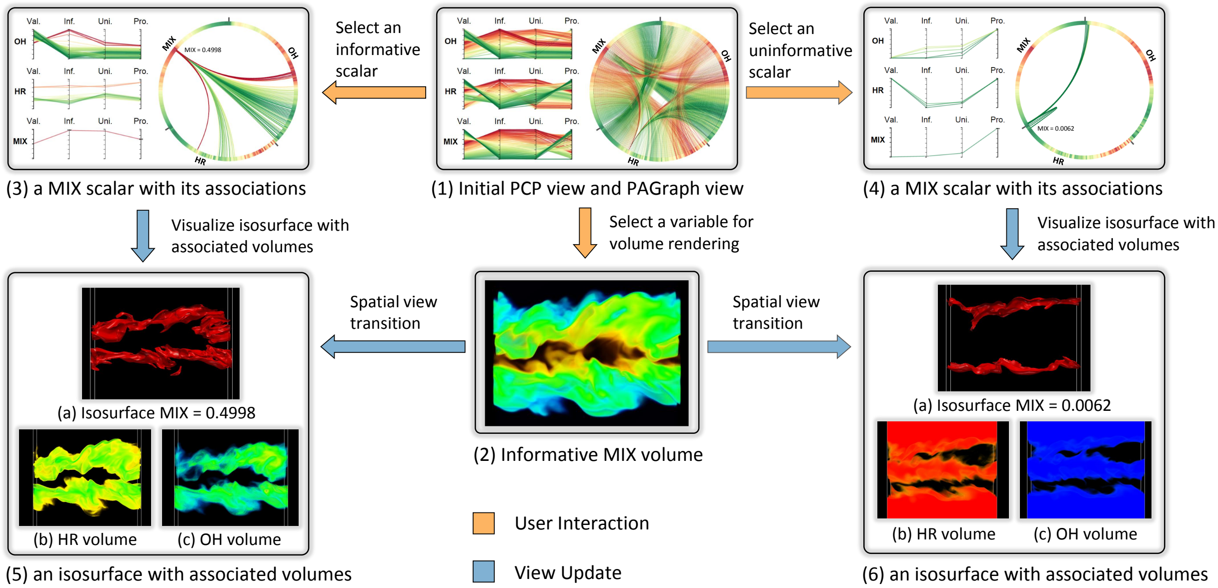


Fig. 1. An illustration of our association-guided exploration framework using the Turbulent Combustion data set. Starting from the initial PCP/PAGraph view (1), users can select one variable to see the informative volume rendering in the spatial view (2), or select specific scalars to see their associations in the PCP/PAGraph view (3, 4), together with the isosurface visualization and associated volumes in the spatial view (5, 6).

Abstract—The heterogeneity and complexity of multivariate characteristics poses a unique challenge to visual exploration of multivariate scientific data sets, as it requires investigating the usually hidden associations between different variables and specific scalar values to understand the data's multi-faceted properties. In this paper, we present a novel association analysis method that guides visual exploration of scalar-level associations in the multivariate context. We model the directional interactions between scalars of different variables as information flows based on association rules. We introduce the concepts of informativeness and uniqueness to describe how information flows between scalars of different variables and how they are associated with each other in the multivariate domain. Based on scalar-level associations represented by a probabilistic association graph, we propose the Multi-Scalar Informativeness-Uniqueness (MSIU) algorithm to evaluate the informativeness and uniqueness of scalars. We present an exploration framework with multiple interactive views to explore the scalars of interest with confident associations in the multivariate spatial domain, and provide guidelines for visual exploration using our framework. We demonstrate the effectiveness and usefulness of our approach through case studies using three representative multivariate scientific data sets.

Index Terms—Multivariate data, association analysis, visual exploration, multiple views.

1 Introduction

Scientific simulations often create multiple variables describing different physical properties within the same spatial domain. Usually, certain scalar values from a subset of variables carry a greater importance to the understanding of the underlying phenomena than others [15]. For instance, in climate simulation, a hurricane typically forms as the *warm*, *moist* air over the ocean rises in the *low* air pressure area, which creates *strong* winds and *heavy* rainfall. However, without sufficient prior knowledge, it is generally difficult for users to select informative scalar values in a multivariate data space and understand how they are associated with scalar values of the other variables [14]. This makes visual exploration of multivariate scientific data sets a difficult task using only the existing visualization techniques such as direct volume rendering [17] and isosurface visualization [19]. More importantly, the heterogeneity and complexity of multivariate characteristics poses a unique challenge to this non-trivial problem, as it requires investigating the usually hidden associations between different variables and specific scalar values to understand the multi-faceted properties of the data sets.

Considering that each scalar value of a variable, or scalar in short, has a certain amount of information, encoded by the corresponding spatial data points, scalar-level interaction can be viewed as the amount of shared information that flows from one scalar to another scalar of a different variable. In this sense, the roles that scalars play in

• *Xiaotong Liu and Han-Wei Shen are with The Ohio State University. E-mail: {liu.1952, shen.94}@osu.edu*

Manuscript received 31 Mar. 2015; accepted 1 Aug. 2015; date of publication xx Aug. 2015; date of current version 25 Oct. 2015. For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

the entire data set are not equally informative. An informative scalar should reveal information about the associated scalars of other variables and how they interact [6]. The directional aspect of information flows should be taken into account to determine how informative one scalar is compared with the other in an association. Commonly-used correlation measures such as correlation coefficients and mutual information mostly focus on studying the symmetric and average relationship between variables, and thus are unable to reflect the directional scalar-level relationships. Consequently, even for two correlated variables, it is still unclear how a specific scalar in one variable is associated with another scalar of the other variable. In addition, the enormous multivariate space complicates the search of potentially interesting scalars. Due to the large number of possible associations with scalars of many different variables, it is non-trivial to determine the relative informativeness that a scalar has in the entire multivariate domain.

In this paper, we present a novel association analysis method that identifies informative scalars in multivariate data sets. To understand how the scalars interact with each other, we model the directional interactions between scalars of different variables as information flows based on *association rules*. Association rules are originally proposed by Rakesh Agrawal et al. [2] to analyze relationships of sale products from supermarket basket transactions, and confident rules can infer the sales of other products by observing the sales of a given product. In our method, we treat each scalar as a sale product and a co-occurrence of scalars between different variables as a transaction. We define the amount of information that flows from one scalar to another as the confidence of an association rule, and estimate it using conditional probability. We model all possible associations between scalars of different variables using a probabilistic association graph called PAGraph, and determine the informativeness of scalars in the PAGraph based on information propagation. The PAGraph is analogous to a social network in terms of two important factors — influence/informativeness and passivity/uniqueness: (1) in a social network, an individual of high influence can attract his audience and lead the interactions, while an individual of high passivity is often inert in response to the interactions; (2) in a PAGraph, an informative scalar can reliably infer the existence of other scalars of different variables, while a unique scalar can hardly be predicted by other scalars of different variables. To quantify informativeness and uniqueness from scalar-level associations in the PAGraph, we propose the Multi-Scalar Informativeness-Uniqueness (MSIU) algorithm based on an information propagation model in social networks [23]. Integrating our association analysis method with interactive visualization techniques, we present an exploration framework with multiple interactive views to explore the scalars of interest with confident associations in the multivariate spatial domain, and provide guidelines for visual exploration using our framework. We demonstrate the effectiveness and usefulness of our approach in exploratory analysis using three representative multivariate data sets.

2 Related Work

In this section, we briefly review and discuss the previous research that are closely related to our work.

Multivariate data analysis is a well researched topic in the literature of scientific visualization. Wong and Bergeron [33], and Fuchs and Hauser [7] gave extensive reviews of multivariate data analysis and visualization. A fundamental research problem in this field is to study the relationships of variables in the multidimensional data space. Sauber et al. [24] introduced local correlation coefficients to visualize relationships in multivariate data sets. Gosink et al. [9] took normalized dot product between two gradient fields from two variables to derive correlation fields, which were then used to analyze the variable interactions with a third variable. Janicke et al. [14] adapted local statistical complexity from finite state cellular automata to identify informative regions in multivariate data. Nagaraj et al. [21] proposed a gradient-based measure that reveals the relationships of variables for the purpose of comparative visualization. Wang et al. [29] employed transfer entropy to study the causal relationships between variables. Turkay et al. [28] proposed a multi-dimensional data exploration model by linking the items space and dimensions space. Tatu et al. [27] developed an interactive visualization system for exploring interesting subspaces of high-dimensional data. Guo et al. [11] integrated parallel coordinates plot (PCP) and multidimensional scaling (MDS) plots for manipulating multidimensional transfer functions. Watanabe et al. [30] applied biclustering techniques to extract feature subspaces from multivariate data. While most existing methods offer insights into the average relationships between variables, little study has focused on the specific relationships between different scalar values in different variables. Recently, Biswas et al. [6] employed the surprise and predictability metrics to measure the specific mutual information between a scalar value and a variable. These two metrics mostly focus on one-way interactions between two variables, and the extension to more than two variables is non-trivial. In this paper, we present a bottom-up approach to analyze scalar-level relationships in different variables for identifying the informative and unique scalars in multivariate data sets. Our approach copes with more than two variables by using two new metrics, informativeness and uniqueness, to reflect two-way interactions between one scalar value and other scalars of different variables.

Interaction and interface design are critical for visual exploration of complex data sets. Introduced by Becker and Cleveland [5], brushing became one of the most popular interactions for exploring data of interest. Martin and Ward [20] extended brushing to high-dimensional data space. Janicke et al. [13] proposed to brush an attribute space derived from the high-dimensional data space. Coordinated and multiple views are an effective design for visual exploration through different presentations, where the changes in one representation will update the other views immediately [4]. Many examples are given by Roberts [22] in a survey of coordinated and multiple views in exploratory visualization. To represent relationships of variables, node-link diagrams and PCP are commonly used, as in Wang et al. [29], Yang et al. [34], and Biswas et al. [6]. In our work, we link a radial graph visualization with multiple PCPs to support brushing scalars of interest and simultaneously showing the confident associations. These two interactive views are also linked with volume rendering and isosurface visualization to provide a multi-view interface for visual exploration of multivariate scientific data sets.

Social network analysis studies social relationships from the standpoint of network theory. An important task is to identify influential people based on the information flows and propagation in the networks. Watts and Dodds [31] found that influence depends on how willing the audience are to adapt their behaviors and beliefs. Agarwal et al. [1] discovered that the most influential social media users were not necessarily the most active. Weng et al. [32] quantified the influence of Twitter users based on topical similarities and link structures. Goyal et al. [10] proposed to learn influence from past information propagation. Romero et al. [23] devised a model to evaluate the influence and passivity of people in a social network based on information propagation, where the influence and passivity mutually depend on each other. Liu et al. [18] attempted to analyze the influence and passivity conditioned on specific themes. Inspired by Romero et al. [23], we define informativeness and uniqueness in multivariate data sets in analogy to influence and passivity in social networks, and develop an association analysis algorithm to measure informativeness and uniqueness of scalars in multivariate data sets.

3 System Overview

The main objective of this work is to guide visual exploration of multivariate data sets based on scalar-level associations. Essentially, we analyze scalar-level associations to determine the informativeness and uniqueness of scalars, which describe the information flows between scalars of different variables and reveal how they interact with each other in the multivariate domain. Based on scalar-level associations, we design multiple interactive views to develop a framework for exploring scalar-level associations in multivariate data sets. Figure 2 illustrates our analytical workflow. From the input multivariate data set, we model all possible associations using a probabilistic association graph (Section 4.1 and 4.2). The informativeness and uniqueness

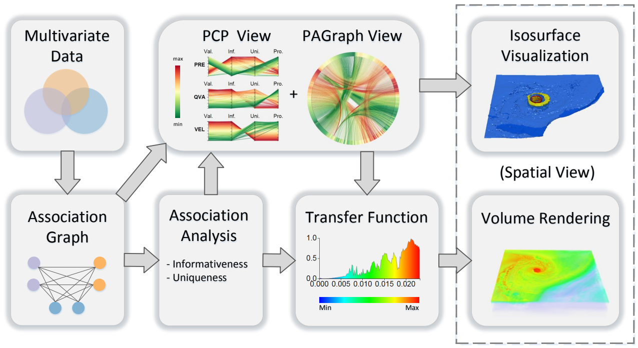


Fig. 2. Overview of the analytical workflow. From the input multivariate data set, we model scalar-level associations using a probabilistic association graph, and apply the proposed association analysis method to quantify the informativeness and uniqueness of scalars. Multiple interactive views and spatial visualizations are created to explore scalar-level associations in the multivariate data set.

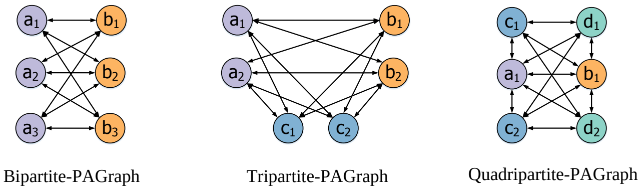


Fig. 3. An illustration of PAGraphs for multivariate data sets (distinct colors are used for nodes from different variables). Left: bipartite-PAGraph for two variables. Middle: tripartite-PAGraph for three variables. Right: quadripartite-PAGraph for four variables.

of scalars are quantified by the proposed association analysis method (Section 4.3). To explore the scalars with associations of interest, the association graph with informativeness and uniqueness is used to create the PCP view and the PAGraph view, which are linked with isosurface visualization and direct volume rendering to reveal the spatial relationships of the selected scalars (Section 5.1).

4 Association Analysis in Multivariate Data Sets

In this section, we first model the directional interactions between scalars of different variables as information flows based on association rules, and describe how to represent scalar-level associations using a probabilistic multipartite graph. Then we introduce the concepts of informativeness and uniqueness to describe how information flows between scalars of different variables and how they are associated with each other in the multivariate domain. Finally we describe the Multi-Scalar Informativeness-Uniqueness (MSIU) algorithm to evaluate the informativeness and uniqueness of scalars.

4.1 Modeling Directional Interactions as Information Flows

Considering that each scalar has a certain amount of information, encoded by the corresponding spatial data points, scalar-level interaction can be viewed as the amount of shared information that flows from one scalar to another scalar of a different variable. To quantify the information flows between scalars of different variables, we make use of *association rules*, which have been widely applied to discover interesting relationships among sale products in large transaction data sets [2]. An association rule is defined as an implication of the form $x \rightarrow y$. An example from the supermarket domain could be $Bread \rightarrow PeanutButter$, suggesting that a strong association exists between the sale of bread and peanut butter because many customers who buy bread also buy peanut butter in one transaction. For a given association rule $x \rightarrow y$, the strength of the rule can be measured in terms of its *confidence*, which can be interpreted as an estimate of the conditional probability $p(y|x)$. The confidence of an association rule can be used to infer the *dependency* that y has on x. The higher the confidence, the more likely it is for x to infer the existence of y. In other words, y is more likely to depend on x.

To make the analogy from the supermarket domain to the multivariate domain, let us treat each scalar as a sale product, and a co-occurrence of the scalars between different variables as one transaction. Considering two scalars x_i and x_j from two different variables X_i and X_j, two possible association rules $x_i \rightarrow x_j$ and $x_j \rightarrow x_i$ can be learned based on their occurrence in the multivariate space. The joint probability $p(x_i, x_j)$ reflects the amount of information shared between the two scalars, and the conditional probability $p(x_j|x_i)$ reflects the amount of shared information that flows from x_i to x_j. Naturally, $p(x_j|x_i)$ quantifies the confidence of an association rule $x_i \rightarrow x_j$ in the multivariate scenario. Unlike most traditional correlation analysis that only studies the average symmetric relationships between variables, such an association-rule-based approach allows us to investigate pointwise directional relationships between scalars of different variables.

4.2 Probabilistic Association Graph

Given a multivariate data set, the scalars and the associated multi-scalar interactions can be represented by a directed graph, where every node $i \in V$ is a scalar x_i and every directed edge $e(i,j) \in E$ is an association rule $x_i \rightarrow x_j$ with the confidence $p(x_j|x_i)$. Essentially, the graph represents information propagation between scalars of different variables, as edges reflect the information flows. We call this the probabilistic association graph $PAGraph(V,E)$. For a system of M variables, since every edge connects two nodes from different variables and no edge connects two nodes from the same variable, a PAGraph is basically a *multipartite graph* whose nodes can be divided into M disjoint sets. Figure 3 illustrates examples of PAGraphs with $M = 2,3,4$.

Since a continuous scalar field contains an infinite number of values, each value domain is discretized into N bins to represent the entire field. We take the value at the center of a bin to represent the scalars that lie within the range of the bin. As a result, the number of nodes $|V|$ is upper-bounded by MN, and the number of edges $|E|$ is upper-bounded by $M(M-1)N^2$. Since many scalar values of different variables do not co-occur in the multidimensional space, $|E|$ is often much smaller than $M(M-1)N^2$. To compute the conditional probability $p(x_j|x_i)$ for a pair of scalars of two different variables, we use 1D histogram to approximate the probability distribution $p(x_i)$ for every variable, and 2D joint histogram to approximate the joint probability distribution $p(x_i,x_j)$ for every pair of variables, and hence we have $p(x_j|x_i) = p(x_i,x_j)/p(x_i)$.

4.3 Informativeness and Uniqueness of Scalars

With the PAGraph, it is now possible to identify informative scalars over the multivariate domain, which can infer the existence of other scalars of different variables. From the standpoint of information flows, this is similar to identifying influential individuals in a social network based on their information propagating activities. In this section, we first discuss the analogy between social networks and PAGraphs in terms of two important factors: influence/informativeness and passivity/uniqueness. Then we describe how to determine the informativeness and uniqueness of scalars in the multivariate domain using our PAGraph.

4.3.1 From Social Networks to Association Graphs

In the field of social network analysis, *influence*, which makes people adapt their behavior, attitudes or beliefs, has been an important factor that directs the dynamics of social media interactions [16]. High popularity does not necessarily infer high influence and vice-versa [23]. The further one's messages are propagated in the network by their connected people, the more influence they may have on others. Equally important is *passivity*, as many people are passive information consumers and do not forward the information to the network [23]. Passivity reflects the barrier to the propagation of messages that is often difficult to overcome. In a brief way, the influence and passivity measures of individuals in social networks can be interpreted as follows:

- **Influence**: an individual of high influence can attract his audience and lead the social interactions.
- **Passivity**: an individual of high passivity is often inert in response to the social interactions.

We observe two factors similar to influence and passivity in the multivariate domain. First, inter-dependencies exist between scalars of different variables, and the scalars that often occur with others actively engage in multi-scalar interactions, and thus have more information in terms of their predictability of the dependent scalars. It is also true that highly frequent scalars are not necessarily the most informative, as they can correspond to background noise or uninteresting regions. Second, certain scalars may not be easily inferred from other scalars of different variables, and thus are considered unique in the multi-scalar associations. In analogy to influence and passivity in social networks, we define the *informativeness* and *uniqueness* of scalars in the multivariate domain:

- **Informativeness**: a scalar of high informativeness can reliably infer the existence of other scalars of different variables.
- **Uniqueness**: a scalar of high uniqueness can hardly be inferred by other scalars of different variables.

To determine informativeness and uniqueness in multivariate data sets, we propose a multivariate association analysis method named the *Multi-Scalar Informativeness-Uniqueness (MSIU)* model, based on the *Influence-Passivity (IP)* model [23]. Next we introduce the basics of the IP model and its applications in social network analysis, and then describe our MSIU model in the multivariate scenario.

4.3.2 Basics of the Influence-Passivity Model

The IP model was first proposed by Romero et al. [23] to determine the influence and passivity of people in a social network based on the structural properties of the network as well as the diffusion behaviors among people. It utilizes the pairwise association information between people to calculate the relative influence and passivity each person has on the whole network. The IP model lies on the basis of the following relations in a social network:

- A person's influence relies on the number of people he influences as well as their passivity.
- A person's influence relies on how much he attracts the attention of his audience compared to other people.
- A person's passivity relies on the influence of those who he is exposed to but not influenced by.
- A person's passivity relies on how much he rejects other people's influence compared to everyone else.

Given a social network $G(V,E)$ with individuals V and edges E, the weight w_{ij} on an edge $e(i,j) \in E$ represents the ratio of influence that person i does exert on person j to the total influence that i attempts to exert on j. Romero et al. [23] used the ratio of retweets as w_{ij}, which is the number of i's tweets retweeted by j divided by the total number of i's tweets. For every edge $e(i,j) \in E$, the *acceptance rate* is defined as:

$$a_{ij} = \frac{w_{ij}}{\sum_{k:(k,j)\in E} w_{kj}}. \quad (1)$$

This metric reflects the dedication j has to i in terms of information propagation. It measures the amount of influence that j accepts from i scaled by the total influence accepted by j from all people in the network. For example, Figure 4 shows that j accepts influence from four neighbors including i, and the acceptance rate a_{ij} reflects how much i attracts the attention of j competing with other j's neighbors

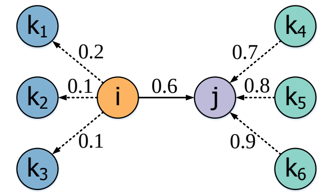


Fig. 4. An illustration of information propagation in a social network or a PAGraph. The direction of an edge shows the direction of information propagation.

who may also influence j. On the other hand, the *rejection rate* is defined as:

$$r_{ij} = \frac{1 - w_{ij}}{\sum_{k:(i,k)\in E} (1 - w_{ik})}. \quad (2)$$

This metric measures the amount of influence that j rejects from i scaled by the total influence from i rejected by all people in the network. This can also be illustrated using the example in Figure 4: i attempts to influence four neighbors including j, and the rejection rate r_{ij} assesses how much j rejects i's influence compared to other i's neighbors who may also reject i's influence.

4.3.3 The Multi-Scalar Informativeness-Uniqueness Model

Inspired by the IP model in [23] for social network analysis, we propose the MSIU model to evaluate the informativeness and uniqueness of scalars in the multivariate domain. Table 1 shows the correspondence between the IP model in [23] and our MSIU model. In our MSIU model, we take a *PAGraph* as the input graph, which corresponds to a social network in the IP model. Our MSIU model considers *scalars* as nodes, and edges are constructed based on *association rules*, which is analogous to the concept of Twitter users being nodes and retweeting relationships being edges in the IP model. The edge weight in our MSIU model is the *confidence* of the association rule, which corresponds to the ratio of retweets in the IP model. While popularity in the IP model is reflected by the number of followers, it is the probability of occurrence for a scalar in our MSIU model. The relations between informativeness and uniqueness can also be interpreted in the multivariate domain:

- A scalar's informativeness depends on the number of scalars it infers as well as their uniqueness.
- A scalar's informativeness depends on how reliably it infers the existence of other scalars compared to everyone else.
- A scalar's uniqueness depends on the informativeness of those who it is exposed to but not informed by.
- A scalar's uniqueness depends on how much it rejects other scalars' information compared to everyone else.

Given a pair of scalars in two different variables, the amount of information that flows from one scalar x_i to another x_j is x_i's confidence of inferring the existence of x_j, which is measured by $p(x_j|x_i)$. Considering all possible associated scalars, the acceptance rate of x_j with respect to x_i is defined as:

$$a_{ij} = \frac{p(x_j|x_i)}{\sum_{k:(k,j)\in E} p(x_j|x_k)}. \quad (3)$$

The acceptance rate in our MSIU model measures the amount of information that x_j accepts from x_i scaled by the total information accepted by x_j from all its associated scalars in the PAGraph. Take Figure 4 as an example: x_j is informed by four other scalars, with confidence $p(x_j|x_i) = 0.6$, $p(x_j|x_{k_4}) = 0.7$, $p(x_j|x_{k_5}) = 0.8$ and $p(x_j|x_{k_6}) = 0.9$,

Table 1. The correspondence between the IP and MSIU models.

	The IP model in [23]	**Our MSIU model**
Input graph	Social network	PAGraph
Nodes	Twitter users	Scalars
Edges	Retweets	Association rules
Edge weight	Ratio of retweets	Rule's confidence
Popularity	Number of followers	Probability
Output	Influence	Informativeness
	Passivity	Uniqueness

Algorithm 1 Multi-Scalar Informativeness-Uniqueness (MSIU)

Input: a multivariate data set V.
Output: informativeness I_i and uniqueness U_i for every scalar $x_i \in V$.

Compute probability distribution $p(x_i)$ for each variable, and joint probability distribution $p(x_i, x_j)$ for every pair of variables.
Construct a $PAGraph(V, E)$ with E being edges based on association rules of V, and w_{ij} for each $e(i, j)$ being $p(x_i, x_j)$.
Compute the acceptance rate a_{ij} and rejection rate r_{ij} for each $x_i \in V$ according to equations (3) and (4) respectively.
Initialize informativeness $I_i = 1$ and uniqueness $U_i = 1$ for each $x_i \in V$.
repeat
  for each $x_i \in V$ **do**
    $I_i \leftarrow \sum_{j:(i,j)\in E} a_{ij} U_j$
  end for
  for each $x_i \in V$ **do**
    $U_i \leftarrow \sum_{j:(j,i)\in E} r_{ji} I_j$
  end for
  for each $x_i \in V$ **do**
    $U_i \leftarrow \frac{U_i}{\sum_{k\in V} U_k}$
    $I_i \leftarrow \frac{I_i}{\sum_{k\in V} I_k}$
  end for
until I_i and U_i converge

and hence we have $a_{ij} = 0.6/(0.6+0.7+0.8+0.9) = 0.2$. The acceptance rate a_{ij} reflects how much x_i can infer the existence of x_j competing with other x_j's neighbors who may also inform x_j. In this example, although the information that flows from x_i to x_j is 60%, the relative information that x_j accepts from x_i is very small (20%) when considering all the associated scalars of x_j. Such a relative measure is desired in our multivariate scenario as it takes all possible associations into account for quantifying the relative information that flows from one scalar to another.

Given a pair of scalars in two different variables, the amount of information from x_i that is rejected by x_j is the chance of x_j's absence when x_i is observed, which is measured by $p(\bar{x_j}|x_i) = 1 - p(x_j|x_i)$. Considering all possible associated scalars, the rejection rate of x_j with respect to x_i is defined as:

$$r_{ij} = \frac{1 - p(x_j|x_i)}{\sum_{k:(i,k)\in E}(1 - p(x_k|x_i))} = \frac{p(\bar{x_j}|x_i)}{\sum_{k:(i,k)\in E}(p(\bar{x_k}|x_i)}. \quad (4)$$

The rejection rate in our MSIU model measures the amount of information that x_j rejects from x_i scaled by the total information from x_i rejected by all its associated scalars in the PAGraph. In the example of Figure 4, we see that x_i attempts to inform four other scalars, with rejected information $p(\bar{x_j}|x_i) = 1-0.6 = 0.4$, $p(\bar{x_{k_1}}|x_i) = 1-0.2 = 0.8$, $p(\bar{x_{k_2}}|x_i) = 1-0.1 = 0.9$ and $p(\bar{x_{k_3}}|x_i) = 1-0.1 = 0.9$, and hence we have $r_{ij} = 0.4/(0.4+0.8+0.9+0.9) = 0.13$. The rejection rate r_{ij} assesses how much x_j rejects x_i's information compared to other x_i's neighbors who may also reject x_i's information. In this example, the amount of information that is rejected by x_j from x_i is relatively small (13%) after considering all possible associated scalars of x_i.

Both acceptance rates and rejection rates in our MSIU model are used to obtain the relative informativeness and uniqueness that one scalar has in the entire PAGraph, based on the information flows modeled by the association rules. We employ the iterative scheme in the IP model [23] to compute the informativeness and uniqueness in our MSIU model:

$$I_i \leftarrow \sum_{j:(i,j)\in E} a_{ij} U_j \quad (5)$$

$$U_i \leftarrow \sum_{j:(j,i)\in E} r_{ji} I_j \quad (6)$$

In the initial state, the informativeness and uniqueness values are set to 1 for every scalar, and are iteratively updated as the information flows between pairs of scalars until convergence. At the end of each iteration, I_i and P_i are normalized to be within $[0, 1]$ respectively. Romero et al. [23] have shown that this iterative process converges in tens of iterations for even large graphs that have one million edges.

Essentially, the informativeness and uniqueness are quantified based on how information is propagated between scalars of different variables. From equation (5), we see that a scalar x_i is more informative if the following conditions hold: (1) x_i is associated with many other scalars of different variables; (2) the information of x_i is highly accepted by the associated scalars (a_{ij} is high); and/or (3) the associated scalars are unique (U_j is high), which means that x_i is able to infer the existence of scalars that are generally hard to be predicted. Similarly, a scalar x_i is more unique if (1) x_i is associated with many other scalars of different variables; (2) x_i mostly rejects the information of the associated scalars (r_{ji} is high); and/or (3) the rejected scalars are informative (I_j is high), which means that even the generally informative scalars are not able to infer the existence of x_i.

The complete process of determining the informativeness and uniqueness that each scalar has in the multivariate domain is referred to as the Multi-Scalar-Informativeness-Uniqueness (MSIU) algorithm (Algorithm 1). We now analyze the time complexity of the MSIU algorithm. For a data set of M variables $\times$ K data points, 1D histograms and 2D joint histograms can be computed in $O(MK + M(M-1)K)$ time. Assuming the number of bins in each scalar range is N, the PAGraph can be constructed in $O(M(M-1)N^2)$. Computing the acceptance rates and rejection rates takes $O(|V| + |E|)$ time. The iterative computation of informativeness and uniqueness values takes $O(c(|V| + |E|))$, where c is the total number of iterations. According to Romero et al. [23], c is a small number since the iterative process converges very quickly. Since $|V|$ is upper-bounded by MN, and the number of edges $|E|$ is upper-bounded by $M(M-1)N^2$ ($|E|$ is often much smaller than $M(M-1)N^2$ due to the sparsity of multidimensional data), overall the MSIU algorithm takes $O(MK + M(M-1)K + c(MN + M(M-1)N^2))$ time.

5 Association-Guided Exploration Framework

We combine our association analysis method with interactive visualization techniques to develop a framework for exploring scalar-level associations in multivariate data sets. In this section, we describe the design considerations and choices of our system interface, and provide guidelines for visual exploration using our framework. A supplementary video that accompanies this paper demonstrates our user interface and the exploration work flow.

5.1 User Interface

The user interface consists of three components: the PAGraph view, the multi-faceted PCP view, and the spatial view.

5.1.1 PAGraph View

The PAGraph view is to visualize the PAGraphs of confident associations between scalars of different variables. Because a PAGraph typically has over hundreds of nodes and thousands of edges, conventional

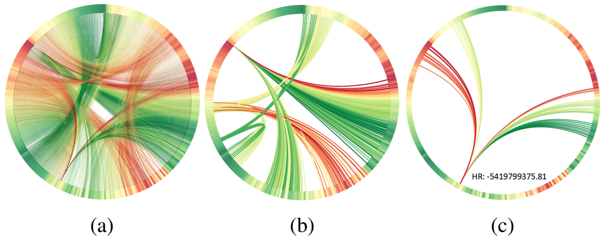


Fig. 5. Visualization of PAGraphs with different number of associations. (a) Top 10% confident associations; (b) Associations of a few scalars; (c) Associations of one particular scalar.

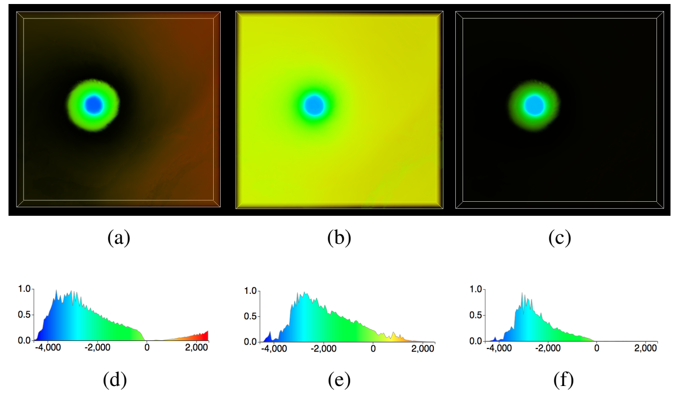


Fig. 6. Volume rendering of the PRE variable in the Hurricane Isabel data set based on average associations with the scalars of the HR and OH variables. Informative PRE volume (a) with its transfer function (d); Unique PRE volume (b) with its transfer function (e); Informative and unique volume (c) with its transfer function (f).

node-link diagrams can suffer from visual cluttering due to node overlap and edge crossing. An alternative visualization is the adjacency matrix, which shows how nodes are connected together through the intersection of rows and columns. However, considering the sparsity of multidimensional data as well as the multi-partite property of PAGraph, the PAGraph is often sparse and many display areas will be wasted in the corresponding adjacency matrix visualization. In searching the design space for visualizing directed graphs, we choose a radial graph visualization (Figure 5): the scalars are arranged clockwise around a circle in an ascending order of the values, while the entire circle is divided into multiple segments for multiple variables (the scalars that do not have associations are omitted). Associations between scalars of different variables are drawn as curves connecting the associated scalars together. Each scalar node on the circle is colored by the its informativeness in a green-yellow-red order (see Figure 2 for the color legend); and the color of an arc is consistent with one of the associated scalars to facilitate visual tracking. The boundaries between the scalars of different variables along the circle are marked by small line segments (see Figure 7.1).

In an initial PAGraph view (Figure 5a), the top confident associations are visible (10% by default), and users can adjust the edge density of the PAGraph to be shown in a control panel. Users can hover on a scalar node in the outermost circle of the PAGraph to view its confident associations, and a tooltip will pop up for showing the corresponding scalar value in focus (Figure 5c).

5.1.2 Multi-Faceted PCP View

With our analysis method, a scalar value (val.) now has three additional facets: informativeness (inf.), uniqueness (uni.) and probability (pro.). A multi-faceted view of scalars can reveal the relationships between the facets and prompt selecting scalars of interest [15]. To this end, we employ parallel coordinates plot (PCP) [12], with the following special design considerations:

Rank-based Axes. The density of polylines on a PCP axis is typically not uniform, due to the non-uniform distribution of data values. As a result of such overplotting, users may not be able to select a small portion of scalars of interest in dense areas through brushing [5]. We transform the values on each axis into a *rank space*, where the position of a polyline on an axis shows the *relative rank* of the facet for the corresponding scalar. Since ranks are uniformly distributed on an axis, users can brush a small region for selecting a few scalars of interest (Figure 7.2). By default the polylines are colored by their informativeness in a green-yellow-red order (see Figure 2 for the color legend).

Juxtaposed Plots. To show the multi-ranks of scalars of multiple variables simultaneously, we employ the juxtaposition design [8], using multiple side-by-side PCPs — each PCP corresponds to multi-faceted scalars of one variable (Figure 7.1). Such juxtaposed plots allow users to focus on selecting an interesting value range in one variable, while still maintaining the information of the multi-ranks in other variables for switching the selection.

Cross-linked Views. In addition to standard brushing and linking across axes [5], our design also supports brushing and linking across different variables — when users select scalars in the PCP of one variable through brushing, the top associated scalars of other variables are extracted from the PAGraph (10% by default), and are simultaneously highlighted in the other PCPs. Furthermore, these associations are also highlighted in the PAGraph view (Figure 7.2). The values of the selected scalars are shown in a separate window so that users can understand and refine their selection.

5.1.3 Spatial View

While the PAGraph view and the PCP view together serves as the interactive interface for users to select scalars of interest and view their associations in detail, the spatial view provides users with the most direct and intuitive view of the data through volume rendering (Figure 7.3b and 7.3c) and isosurface visualization (Figure 7.3a). To design an automatic insightful transfer function that saves users' effort from the tedious and time-consuming manual specification, while also revealing the associations between scalars of different variables, we propose the following two schemes:

Transfer Function with Average Associations. The informativeness and uniqueness of a scalar together reveals the association information about scalars of other variables in an average sense. Therefore, the transfer function that reflects average scalar-level associations takes one of the following options: $opacity(informativeness(x))$, $opacity(uniqueness(x))$, or $opacity(informativeness(x) \times uniqueness(x))$. The corresponding volumes determined by these transfer functions are referred to as *informative volume*, *unique volume*, and *informative and unique volume* respectively, as illustrated in Figure 6.

Transfer Function with Specific Associations. When users select a scalar of interest in the interface, in addition to showing the corresponding isosurface visualization (Figure 7.3a), the *associated volumes* highlight the scalars of other variables that are associated with the selected scalar. Thus, for a given scalar x, the transfer function of its associated volume Y is set to $opacity(p(y|x))$. Figure 7.3b and 7.3c present an example of two associated volumes with respect to a selected scalar.

Empirically, we use $opacity(f(x)) = [f(x)]^2$ to emphasize the relatively high informative/unique scalars or the relatively confident associations. The color transfer function $color(x)$ maps the scalar values in a blue-green-red order, as shown in Figure 2.

5.2 Exploration Guidelines

For a given multivariate data set, we assume that users already have some pre-selected variables in mind, as is often the case for domain scientists in real-world applications [6]. The task to address in this work is to explore scalars of interest in the given variables based on scalar-level associations. We follow the visualization mantra

"Overview first, zoom and filter, then details-on-demand" by Ben Shneiderman [25] to guide the exploration process:

Overview first. The initial PCP and PAGraph views provide an overview of the multiple facets of scalars and their associations (Figure 8.1). Users can select a variable from the PCP labels to initiate the spatial view with the informative/unique volume of the variable (Figure 8.2).

Zoom and filter. To understand how scalars of other variables are associated with a particular scalar, users can select a scalar by brushing the PCP view and see the corresponding associations in the PAGraph view. For instance, users can first brush the Inf./Uni. axis to selected a few informative/unique scalars, and then brush the Pro. axis to narrow down the focus to a particular scalar of interest (illustrated in Figure 8.3 as well as in the supplementary video).

Details-on-demand. To gain insights into the spatial relationship of the selected scalar and its associated scalars of other variables, the spatial view shows the isosurface visualization of the selected scalar and its associated volumes of the other variables (Figure 8.4).

6 RESULTS

We demonstrate the effectiveness of our methods through experiments on three representative multivariate data sets: Hurricane Isabel, Ionization Front Instability, and Turbulent Combustion. The experiments were conducted on a desktop machine with an Intel core i7-2600 CPU, 16 GB of RAM and an NVIDIA Geforce GTX 560 GPU with 2GB texture memory. The interactive exploration process in each experiment is demonstrated in the supplementary video that accompanies this paper.

6.1 Case Study 1: Hurricane Isabel Data Set

The Hurricane Isabel data set is an atmospheric simulation created by the Weather Research and Forecast model, courtesy of the National Center for Atmospheric Research and the U.S. National Science Foundation. The resolution of the grid is $250 \times 250 \times 50$ at each time step. To investigate the relationships between pressure, humidity and wind speed in the hurricane simulation, we selected the Pressure (PRE), Water Vapor Mixing Ratio (QVA) and Wind Velocity Magnitude (VEL) variables at time step 18 for our experiment on this data set.

Figure 7.1 presents the initial PCP view and PAGraph view of the selected variables. We started the exploration by selecting the PRE variable and examine the resulting informative/unique volumes, as shown in Figure 6. We can see that the hurricane eye and eyewall structures are highlighted by the informative and unique PRE volume in Figure 6c, which has low air pressure as illustrated in Figure 6f.

To understand how the PRE scalars interact with the scalars of the other variables QVA and VEL, we brushed the initial PCP view in Figure 7.1 to select one informative scalar PRE = -2674.17 in Figure 7.2. The linked PAGraph view in Figure 7.2 highlights the confident associations starting from a single node (PRE = -2674.17). From the linked PCP of VEL in Figure 7.2, we observe that a majority of the associated VEL scalars have high values. Meanwhile, we found that the selected PRE scalar is associated with high/medium QVA values that represent a decent amount of water vapor. When it comes to the spatial view in Figure 7.3, we found that the isosurface PRE = -2674.17 in Figure 7.3a corresponds to the hurricane eyewall. The associated QVA volume in Figure 7.3b presents the long bands of rain clouds that spiral inward to the hurricane eye, and the associated VEL volume in Figure 7.3c highlights strong winds near the hurricane eyewall. In this sense, this informative PRE scalar with its associated QVA and VEL scalars presents a joint feature from multiple variables — the hurricane eyewall with strong wind and spiral rainbands.

As a follow-up comparison study, we select an uninformative scalar PRE = 141.61 in Figure 7.4. We see that this scalar mostly interacts with low VEL values and low QVA values. The green PCP polylines in Figure 7.4 imply that these associated scalar values are also uninformative. The spatial view in Figure 7.5 further explains such associations: the isosurface PRE = 141.61 spreads out the entire spatial domain except around the hurricane eye (Figure 7.5a), which is associated with mostly low water vapor on the continent (the dark blue regions in Figure 7.5b) and weak wind faraway from the hurricane eye (the dark blue regions in Figure 7.5c).

6.2 Case Study 2: Ionization Front Instability Data Set

The Ionization Front Instability data set is a simulation created by Mike Norman and Daniel Whalen for investigating the effects of instabilities where radiation ionization fronts scatter around primordial gas in the formation of galaxies. Ionization front is a transition region where interstellar gas changes from a mostly neutral state to a mostly ionized state, which involves the abundances of many chemical species such as hydrogen and helium. The resolution of the grid is $300 \times 124 \times 124$ at each time step. To investigate the dependency of the ionization process with respect to temperature, we selected the variables of H, H^+ (HP), He (HE), and Temperature (TEM) and time step 100 in our experiment.

For this data set, the initial PCP view and PAGraph view are presented in Figure 8.1. We initiated our exploration with the informative and unique TEM volume in Figure 8.2, determined by the transfer function $opacity(informativeness(x) \times uniqueness(x))$. We found that the TEM scalars around 10000K are highly informative and unique. We then select an informative and unique scalar TEM = 10588.16 in the PCP view in Figure 8.1. From the updated PCP view and PAGraph view in Figure 8.3, we can see that many scalars are associated with this temperature value. To examine how the selected chemical species react at this temperature 10588.16K, we resort to the spatial view of the associated volumes in Figure 8.4b, 8.4c and 8.4d. It is obvious that the regions overlapping the isosurface TEM = 10588.16 in Figure 8.4a are mostly highlighted, particularly in the associated H and HP volumes. According to the studies of hydrogen ionization [26], 10000K is roughly the temperature where a majority of neutral hydrogen become ionized hydrogen. Furthermore, we also explored uninformative TEM scalars from the PCP view in Figure 8.1, which have very high temperature values. No confident associations (top 10%) were found between those uninformative scalars and the H/HP/HE scalars (the resulting PCP and PAGraph views are not shown due to the lack of space).

6.3 Case Study 3: Turbulent Combustion Data Set

The Turbulent Combustion data set is a combustion simulation produced by Dr. Jacqueline Chen at Sandia Laboratories through US Department of Energy's SciDAC Institute for Ultrascale Visualization. The resolution of the grid for each time step is $240 \times 360 \times 60$. The Mixture Fraction (MIX) variable in this data set, ranging from 0 (pure oxidizer) to 1 (pure fuel), signifies the proportion of fuel and oxidizer and generally reflects the characteristics of the flame: when the chemical reaction rate is greater than the turbulent mixing rate, a complete burning is indicated (the scalar MIX = 0.42 corresponds to a complete burning as reported in [3]); weak or extinguished burning occurs when the turbulent mixing rate exceeds the chemical reaction rate. As the characteristics of local combustion also depends on the Heat Release Rate (HR) and Mass Fraction of the Hydroxyl Radical (OH) from the neighboring flame elements, we are interested in exploring scalar-level associations of the MIX, HR and OH variables. We have selected time step 65 to conduct our experiment.

The initial PCP view and PAgraph view are presented in Figure 1.1. From the PCP view, we observe that the most informative MIX scalars lie in the middle range of MIX. Since flames typically burn in a balanced mixture of fuel and oxidizer (either pure oxidizer or pure fuel will result in extinction of reaction), we expect these informative MIX scalars correspond to the burning flame regions, which strongly interact with scalars of the HR and OH variables. The informative MIX volume in Figure 1.2 illustrates the spatial locations of such flame structures.

To investigate the associations between the MIX scalars and the scalars of the other variables HR and OH, two specific MIX scalars that have contrast informativeness have been selected in Figure 1.3 and 1.4 respectively. The linked PAGraph view in Figure 1.3 highlights a lot of associations with MIX = 0.4998, which we expect to represent a burning flame. From the linked PCP view of OH in Figure 1.3, we

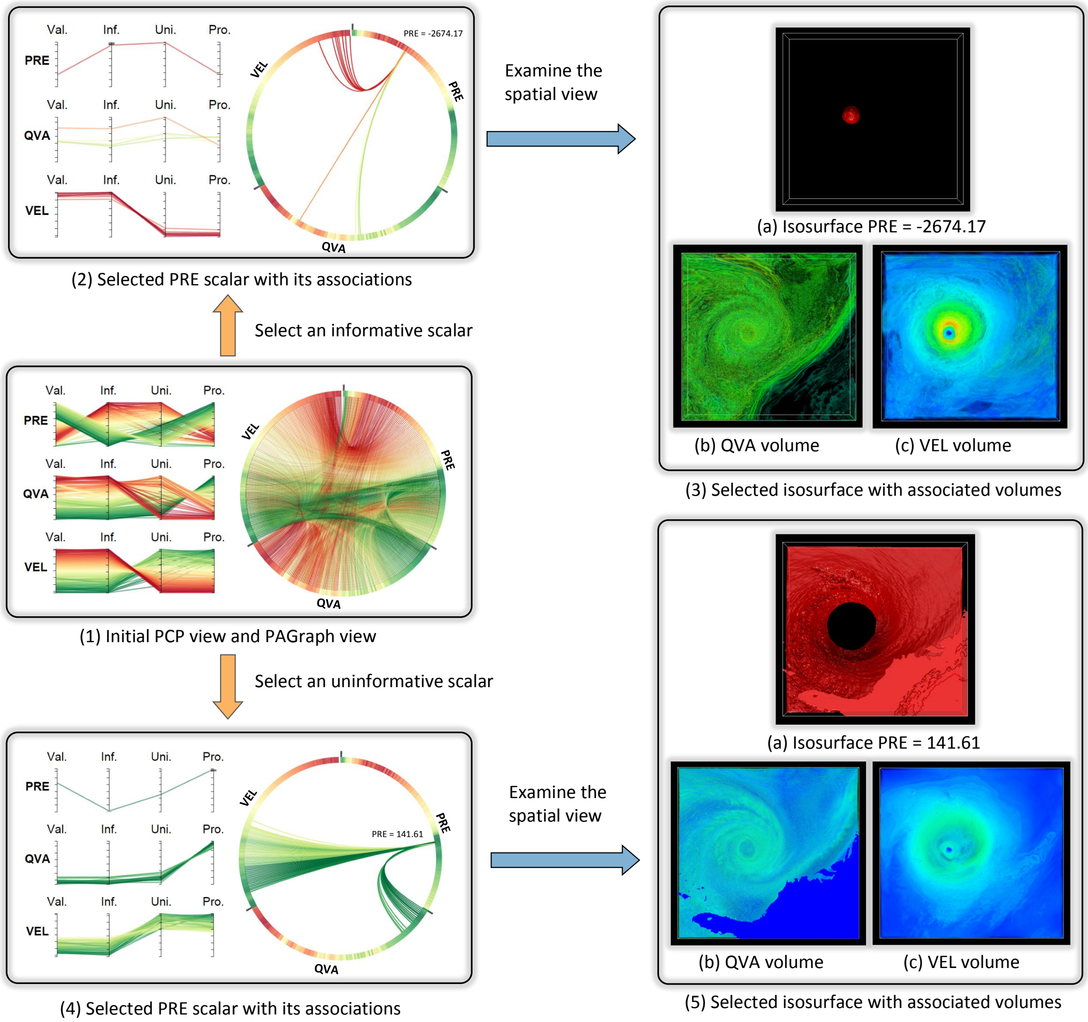


Fig. 7. Experiments on the Hurricane Isabel data set. See Section 6.1 for details about the exploration process.

see that a majority of the associated OH scalars have high values. This is consistent with the fact that the highly convoluted flame regions often have rapid radical diffusion rates [3]. Similarly, we found that the scalar MIX = 0.4998 is associated with many low (negative) HR values that represent high heat release rates. In contrast, an uninformative scalar MIX = 0.0062 in Figure 1.4, which we expect to be a weakly burning region due to the lack of fuel, has only a few associations. From the linked PCP views of OH and HR in Figure 1.4, we found that the associated scalars are low OH values and high HR values. This indicates that weakly or extinguished burning regions often have low radical diffusion rates and low heat conduction rates.

The spatial views in Figure 1.5 and 1.6 confirmed our findings: the associated HR/OH volume mostly overlaps with the flame structure represented by the isosurface MIX = 0.4998, while the isosurface MIX = 0.0062 presents two outer layers in the spatial domain that is far away from the strong flame region, and the associated HR/OH volumes highlight the regions that are either outer layers (where oxidizer is too high) or inner layers (where fuel is too high).

7 Discussion and Future Work

The MSIU model brings a novel perspective to visual exploration of multivariate data sets. Previous correlations methods include correlation coefficients, mutual information, gradient-based measures, etc. Compared to these approaches, our method has two major differences: (1) the MSIU model takes into account the associations of scalars, which are the more specific relationships between variables. These scalar-level associations reveal complex interactions between scalars of different variables, and can be used to classify the scalars in terms of their informativeness and uniqueness; (2) our approach can easily cope with more than two variables, as the MSIU model constructs a multipartite association graph that captures the relationships of scalars in the given variables. In contrast, existing correlation approaches mostly focus on the relationships of two variables, and the extension to more than two variables is often non-trivial.

The association graph with the informativeness and uniqueness learned from the MSIU model is used for volume rendering that encodes the relationships between scalars in one variable and scalars of the other variables. This unique property differentiates our method from many existing multi-dimensional transfer function design approaches, as they usually map a multidimensional vector to a single color and opacity for describing all the variables, sometimes with additional derived variables. Consequently, it can be difficult to understand the different contributions of individual variables regarding the multivariate interactions. Our association-aware transfer functions, encoding the average or specific associations, are automatically created for each variable respectively. Therefore, the scalars in the spatial view have their intrinsic physical properties based on the semantics of the underlying variable.

Our method works well for multivariate data sets where scalars of different variables co-occur, as shown in the three case studies. As stated in Section 5.2, we assume that users already have some preselected variables in mind prior to the exploration in real world scenarios [6], and our method has been applied to several variables of

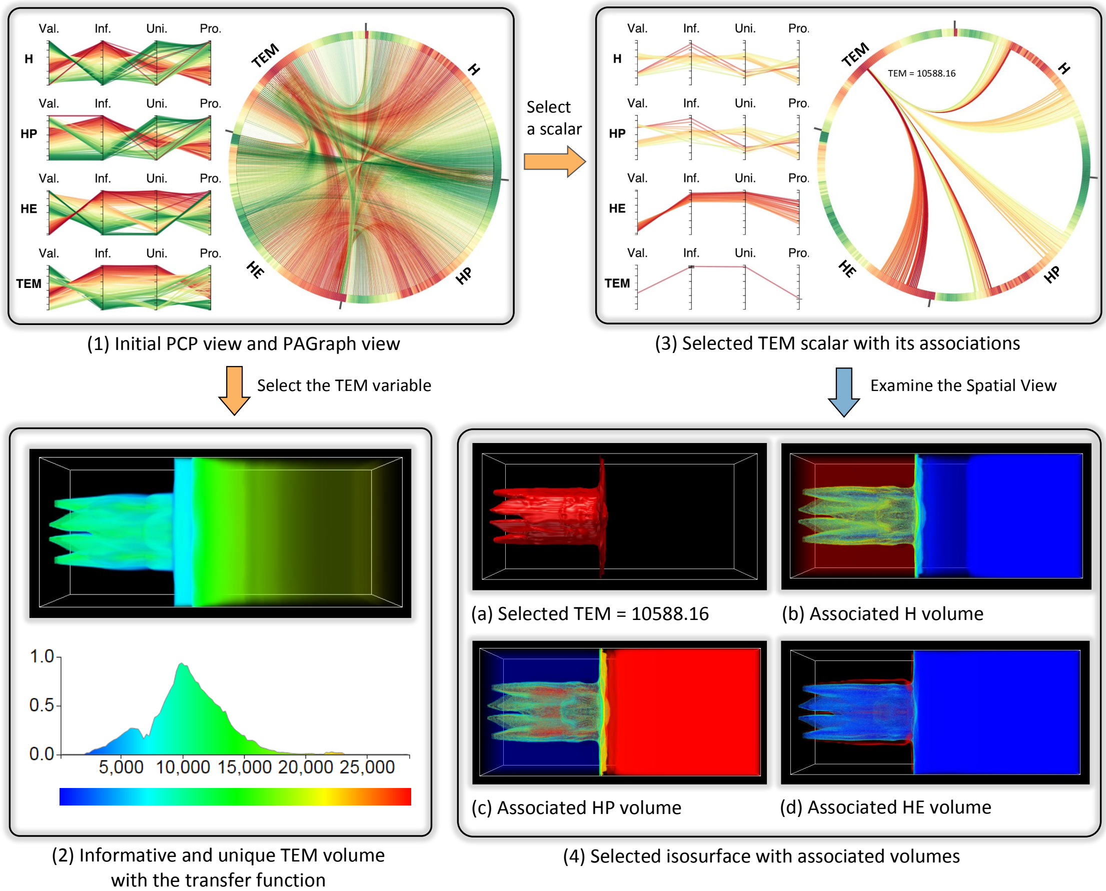
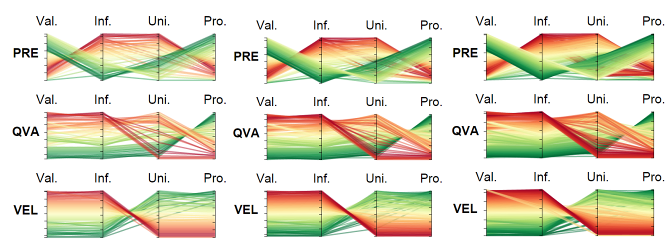


Fig. 8. Experiments on the Ionization Front Instability data set. See Section 6.2 for details about the exploration process.

Table 2. Average runtime (in seconds) for the MSIU algorithm.

	64 bins	128 bins	256 bins
Hurricane Isabel	1.72	2.62	11.95
Ionization Front	4.33	6.99	14.73
Combustion	2.43	3.40	10.61



(a) 64 bins (b) 128 bins (c) 256 bins

Fig. 9. The ranks of informativeness, uniqueness and probability with different number of bins for the Hurricane Isabel data set.

interest in each case study respectively. Although one can consider all the variables simultaneously, it can be very difficult to interpret the relationships between a scalar and scalars of a large number of variables. In contrast, pair-wise association or a small number of multivariate associations are more intuitive and easier to understand. The trade-off between multivariate complexity and interpretability needs to be better understood in the future.

All the computations in the MSIU algorithm are done in the preprocessing stage prior to the exploration process. One parameter in the MSIU algorithm is the choice of bin size N, which is related to the discretization of scalar values and the calculation of probability distributions. Table 2 reports the computational performance of the MSIU algorithm with different number of bins. We also examined the relative ranks of informativeness and uniqueness when using different numbers of bins. It is observed that a majority of the relative ranks of informativeness and uniqueness remained the same for different numbers of bins with minor variations in all three data sets. Due to space constraint, we only report the comparison results for the Hurricane Isabel data set (Figure 9). While we used 128 bins in this work, in the future we would like to experiment with non-uniform discretization as well as multi-level discretization.

8 Conclusion

In this paper, we present a novel association analysis method to guide visual exploration of scalar-level associations in multivariate data sets. We model the directional interactions between scalars as information flows based on association rules. We introduce the concepts of informativeness and uniqueness to describe how information flows between scalars of different variables and how they are associated with each other in the multivariate domain. Based on scalar-level associations represented by a probabilistic association graph, we transform the problem of identifying informative and unique scalars into a network analysis problem of finding influential and passive people in social networks, and propose the Multi-Scalar Informativeness-Uniqueness (MSIU) algorithm. We develop an exploration framework with multiple interactive views to explore the scalars of interest with confident associations in the spatial domain, and provide guidelines for visual exploration using our framework. We demonstrate the effectiveness of our approach through case studies in three application domains.

REFERENCES

[1] N. Agarwal, H. Liu, L. Tang, and P. S. Yu. Identifying the influential bloggers in a community. In *Proceedings of the 2008 international conference on web search and data mining*, pages 207–218. ACM, 2008.

[2] R. Agrawal, T. Imieliński, and A. Swami. Mining association rules between sets of items in large databases. In *ACM SIGMOD*, volume 22, pages 207–216, 1993.

[3] H. Akiba, K.-L. Ma, J. H. Chen, and E. R. Hawkes. Visualizing multivariate volume data from turbulent combustion simulations. *Computing in Science and Engineering*, 9(2):76–83, 2007.

[4] M. Q. W. Baldonado, A. Woodruff, and A. Kuchinsky. Guidelines for using multiple views in information visualization. In *Advanced Visual Interfaces*, pages 110–119, 2000.

[5] R. A. Becker and W. S. Cleveland. Brushing scatterplots. *Technometrics*, 29(2):127–142, 1987.

[6] A. Biswas, S. Dutta, H.-W. Shen, and J. Woodring. An information-aware framework for exploring multivariate data sets. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2683–2692, 2013.

[7] R. Fuchs and H. Hauser. Visualization of multi-variate scientific data. *Computer Graphics Forum*, pages 1670–1690, 2009.

[8] M. Gleicher, D. Albers, R. Walker, I. Jusufi, C. D. Hansen, and J. C. Roberts. Visual comparison for information visualization. *Information Visualization*, 10(4):289–309, 2011.

[9] L. J. Gosink, J. C. Anderson, E. Wes Bethel, and K. I. Joy. Variable interactions in query-driven visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6):1400–1407, 2007.

[10] A. Goyal, F. Bonchi, and L. V. Lakshmanan. Learning influence probabilities in social networks. In *Proceedings of the third ACM international conference on Web search and data mining*, pages 241–250. ACM, 2010.

[11] H. Guo, H. Xiao, and X. Yuan. Scalable multivariate volume visualization and analysis based on dimension projection and parallel coordinates. *Visualization and Computer Graphics, IEEE Transactions on*, 18(9):1397–1410, 2012.

[12] A. Inselberg. The plane with parallel coordinates. *The Visual Computer*, 1(2):69–91, 1985.

[13] H. Janicke, M. Bottinger, and G. Scheuermann. Brushing of attribute clouds for the visualization of multivariate data. *Visualization and Computer Graphics, IEEE Transactions on*, 14(6):1459–1466, 2008.

[14] H. Janicke, A. Wiebel, G. Scheuermann, and W. Kollmann. Multifield visualization using local statistical complexity. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6):1384–1391, 2007.

[15] J. Kehrer and H. Hauser. Visualization and visual analysis of multifaceted scientific data: A survey. *Visualization and Computer Graphics, IEEE Transactions on*, 19(3):495–513, 2013.

[16] R. T. A. Leenders. Modeling social influence through network autocorrelation: constructing the weight matrix. *Social Networks*, 24(1):21–47, 2002.

[17] M. Levoy. Display of surfaces from volume data. *Computer Graphics and Applications, IEEE*, 8(3):29–37, 1988.

[18] X. Liu, S. Parthasarathy, H.-W. Shen, and Y. Hu. Galaxyexplorer: Influence-driven visual exploration of context-specific social media interactions. In *Proceedings of the 24th International Conference on World Wide Web Companion*, pages 215–218, 2015.

[19] W. E. Lorensen and H. E. Cline. Marching cubes: A high resolution 3d surface construction algorithm. In *ACM siggraph computer graphics*, pages 163–169. ACM, 1987.

[20] A. Martin and M. Ward. High dimensional brushing for interactive exploration of multivariate data. In *Visualization, 1995. Visualization '95. Proceedings., IEEE Conference on*, pages 271–, 1995.

[21] S. Nagaraj, V. Natarajan, and R. S. Nanjundiah. A gradient-based comparison measure for visual analysis of multifield data. In *Computer Graphics Forum*, volume 30, pages 1101–1110. Wiley Online Library, 2011.

[22] J. C. Roberts. State of the art: Coordinated & multiple views in exploratory visualization. In *Coordinated and Multiple Views in Exploratory Visualization, 2007. CMV'07. Fifth International Conference on*, pages 61–71. IEEE, 2007.

[23] D. M. Romero, W. Galuba, S. Asur, and B. A. Huberman. Influence and passivity in social media. In *Machine learning and knowledge discovery in databases*, pages 18–33. Springer, 2011.

[24] N. Sauber, H. Theisel, and H.-P. Seidel. Multifield-graphs: An approach to visualizing correlations in multifield scalar data. *IEEE Transactions on Visualization and Computer Graphics*, 12(5), Sept. 2006.

[25] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *IEEE Symposium on Visual Languages*, pages 336–343. IEEE, 1996.

[26] J. T.-l. Su. *An electron force field for simulating large scale excited electron dynamics*. PhD thesis, California Institute of Technology, 2007.

[27] A. Tatu, F. Maas, I. Farber, E. Bertini, T. Schreck, T. Seidl, and D. Keim. Subspace search and visualization to make sense of alternative clusterings in high-dimensional data. In *IEEE Conference on Visual Analytics Science and Technology*, pages 63–72. IEEE, 2012.

[28] C. Turkay, P. Filzmoser, and H. Hauser. Brushing dimensions-a dual visual analysis model for high-dimensional data. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2591–2599, 2011.

[29] C. Wang, H. Yu, R. W. Grout, K.-L. Ma, and J. H. Chen. Analyzing information transfer in time-varying multivariate data. In *Pacific Visualization Symposium*, pages 99–106. IEEE, 2011.

[30] K. Watanabe, H.-Y. Wu, Y. Niibe, S. Takahashi, and I. Fujishiro. Analyzing information transfer in time-varying multivariate data. In *Pacific Visualization Symposium*. IEEE, 2015.

[31] D. J. Watts and P. S. Dodds. Influentials, networks, and public opinion formation. *Journal of consumer research*, 34(4):441–458, 2007.

[32] J. Weng, E.-P. Lim, J. Jiang, and Q. He. Twitterrank: finding topic-sensitive influential twitterers. In *ACM international conference on Web search and data mining*, pages 261–270. ACM, 2010.

[33] P. C. Wong and R. D. Bergeron. 30 years of multidimensional multivariate visualization. In *Scientific Visualization, Overviews, Methodologies, and Techniques*, pages 3–33, Washington, DC, USA, 1997. IEEE Computer Society.

[34] D. Yang, E. A. Rundensteiner, and M. O. Ward. Analysis guided visual exploration of multivariate data. In *Visual Analytics Science and Technology, 2007. VAST 2007. IEEE Symposium on*, pages 83–90. IEEE, 2007.