# **Visualization of Traditional Chinese Medicine Formulas**

## Abstract

**Background:** Traditional Chinese medicine (TCM) formulas are combinations of Chinese herbal medicines. Knowledge of classic medicine formulas is the basis of TCM diagnosis and treatment and is the core of TCM inheritance. The large number and the flexibility of medicine formulas make memorization difficult, and understanding their composition rules is even more difficult. The multifaceted and multidimensional properties of herbal medicines are important for formula understanding, but these are usually separated from the formula information. Furthermore, these data are presented as texts and cannot be analyzed jointly and interactively.

**Objective:** This work aims to devise a visualization method for TCM formulas that shows the composition of medicine formulas and the multidimensional properties of herbal medicines involved, and supports the comparison of medicine formulas.

**Methods:** A TCM formulas visualization method with multiple linked views is proposed and implemented as a web-based tool after a close collaboration between visualization and TCM experts. The composition of medicine formulas is visualized in a formula view with a similarity-based layout supporting the comparison of compositing medicines; a shared medicine view complements the formula view by showing all overlaps of pair-wise formulas; a dimensionality reduction plot of medicines enables the visualization of multidimensional medicine properties. Medicines are color-encoded with a perceptual-guided color map that encodes multidimensional TCM attributes and the similarity measure at the same time. The color map is calculated by a data-driven interpolation scheme. With simple interactions, users could flexibly select medicines or formulas of interest, and the corresponding elements in other views are highlighted through brushing and linking.

**Results:** Our method is applied to two typical categories of medicine formulas, namely, tonic formulas and heat-clearing formulas, which contain 20 and 26 formulas composed of 58 and 73 herbal medicines, respectively. Each herbal medicine has a 23-dimensional characterizing attribute. TCM experts explored the two datasets with our web-based tool and quickly gained insight into formulas and medicines of interest as well as the overall features of the formula groups that are difficult to identify with the traditional text-based method. Moreover, feedback from the experts indicates the usefulness of our method.

**Conclusions:** Our TCM formulas visualization method is able to visualize and compare complex medicine formulas and the multidimensional attributes of herbal medicines involved with an interactive web-based tool. Insights are gained into two typical medicine formula categories by TCM experts using our method. Overall, the new method is a promising first step to new TCM formulas education and analysis methodologies.

**Keywords:** visualization; Chinese medicine formulas; interactive data analysis;

## Introduction

Understanding and applying classic medicine formulas is the basis of TCM diagnosis and treatment and is the core of TCM inheritance. Syndrome differentiation and treatment is a core method in TCM. In clinical practice, prescriptions are based on classic medicine formulas and corresponding medicines may be adjusted according to the symptoms of patients. A typical prescription may contain several medicine formulas, but it is a challenge to identify the involved formulas and understand their effects.

Learning and teaching Chinese medicine formulas is difficult. The traditional education method is to recite classical medicine formulas with their composition rules [39, 8]. However, formula information is presented in texts (Table 1) or static figures and pictures [30], and the composition rules could not be intuitively understood. Data mining and some visual presentations are adopted in the existing computerized analysis of TCM medicine formulas [9, 7, 44] . However, these methods are query-based and do not allow users to interactively explore medicine formulas, and the relatively simple visualization cannot provide an overview of a group of medicine formulas or an in-depth comparison of formulas.

In this paper, we propose a visualization method for TCM medicine formulas to address the issues of existing methods. As a result of a close collaboration between visualization and TCM experts, our method provides compact and clear visualization of medicine formulas and multidimensional attribute data of herbal medicines. While our method is data-driven, we pay close attention to the perception aspect of the visualization. Specifically, a layout algorithm is designed to improve the comparability and reduce visual clutter in the formula visualization; medicines are color mapped with a 2D color map generated with radial basis function interpolation in a perceptual-uniform color space with TCM-concept inspired colors, while their properties drive their placement in the dimensionality reduction plot. Our method is realized as a web-based interactive tool that comprises three linked views: a medicine formula view, a dimensionality-reduction-based medicine view, and a matrix view of shared medicines of formulas. With brushing-and-linking, users could flexibly select medicines or formulas of interest in one view, highlighting corresponding components in other views.

The usefulness of our method is demonstrated by two use cases of typical medicine formula groups. Two TCM experts analyzed these medicine formulas with our method in a free exploration session. They considered that the new method could effectively reveal the compositional principle of medicine formulas in an intuitive way, and could assist the learning of TCM formula composition theories.

Table 1: Part of the original text-based medicine formula information summarized from the textbook Chinese Medicine Formulas (Tenth Edition) [19][.](#_bookmark32)

|  |  |
| --- | --- |
| Formula | Medicines |
| 八珍汤 | 人参、熟地黄、当归、川芎、白术、茯苓、白芍、炙甘草、生姜、大枣 |
| 参苓白术散 | 人参、白术、茯苓、莲子、薏苡仁、山药、桔梗、大枣、甘草、砂仁、白扁豆 |
| 生脉散 | 人参、麦冬、五味子 |
| 四君子汤 | 人参、甘草、白术、茯苓 |
| 大补阴丸 | 熟地黄、龟甲、黄柏、知母 |
| 四物汤 | 熟地黄、白芍、川芎、当归 |
| 地黄饮子 | 熟地黄、山茱萸、肉苁蓉、巴戟天、麦冬、远志、生姜、附子、茯苓、大枣、五味子、石斛、石蒲、肉桂、薄荷 |

## Methods

### Data Descriptions

Classifications of Chinese herbal medicines are multi-faceted and multi-leveled [2].  Siqi (四气), Wuwei (五味), and Guijing (归经) have been an important part of TCM research [1]. Guijing regards the orientation of Chinese herbal medicines, which is to closely connect the functions of medicines with the organs and meridians (脏腑经络) of the human body.

Chinese herbal medicines can be divided into four properties (Siqi): cold (寒), hot (热), warm (温), and cool (凉) according to their functions on the human body. The cold medicines generally have the role of heat-clearing and detoxifying (清热解毒), which are used to treat febrile diseases. The warm medicines typically have the role of warming Yang(阳) to expel coldness (温阳散寒), which are used to treat cold diseases. In addition, medicines with gentle properties are namely Ping (平). Wuwei means five flavors: pungent (辛), sweet (甘), sour (酸), bitter(苦), and salty (咸). The pungent medicines generally taste spicy (such as ginger) or cool (such as mint and borneol), and are used to sweat and relieve Qi(气). Sweet medicines typically taste sweet, for example, licorices and wheat. They have the effect of easing and nourishing. They are also used to reconcile medical properties. Sour medicines, such as black plum and gallnut, have an astringent effect and are used to stop sweating and diarrhea. The bitter medicines have the functions of clearing heat, drying dampness, and catharsis. Finally, salty medicines have the functions of diarrhea and defecation, softening, and firmness. It is believed that these factors are associated with body heat production processes or metabolic activities, and may also play a role in the digestive system, nervous system, cardiovascular system, and so on [30].

Another important concept is Jun-Chen-Zuo-Shi (君臣佐使). Jun- Chen-Zuo-Shi is the principle of the compatibility of TCM formulas. Junyao (君药), or namely, principal medicines as used hereafter, play a major role against the main disease or syndrome. It is the primary medicine in the formulas. Texts in blue in Table 1 indicate Junyao in the corresponding formulas.

In this work, the medicine formulas data are extracted from the key medicine formulas of the textbook Chinese Medicine Formulas (Tenth Edition) [19] as shown in Table 1. The multidimensionial medicine attribute data are retrieved from the SymMap database [42]. Here, the Siqi and Wuwei properties are combined and represented as 23-dimensional vectors.

### Requirement Analysis and Method Overview

Our work aims to devise a method that supports the joint visualization of medicine formulas and the attributes of corresponding medicines. The visual design should support the comparison of formulas and facilitate the classification of medicines based on their SiqiWuwei properties. Visualization and TCM experts work closely together to analyze the requirements of the visual analysis method for medicine formulas. The requirements are summarized as follows.

R1: clear visualization of medicine formulas.

R2: comparing different medicine formulas with ease.

R3: principal medicines should be highlighted.

R4: associating medicine formulas and attributes of the corresponding medicines.

R5: visual elements should be effectively perceived.

R6: interactions should be easy.

R7: visual designs should reflect general concepts of TCM.

Our method is the result of an iterative development process with quick prototypes. Prototypes were realized based on the requirements and proposed to the TCM expert (SP, who is one of the authors) and improvements were made given the feedback of the TCM expert.

The workflow of our method is shown in Figure 1: the medicine formulas information and the multidimensional medicine attribute data are prepared as the input; medicine attribute data are projected to the low dimensional space (2D) and pair-wise distances are calculated; medicine formulas data is arranged by our similarity-based layout algorithm and visualized as an icicle plot; shared medicines of each pair of formulas are calculated and visualized as a matrix; and next, colors are designed for medicines using our perceptual-guided data-driven color encoding method.

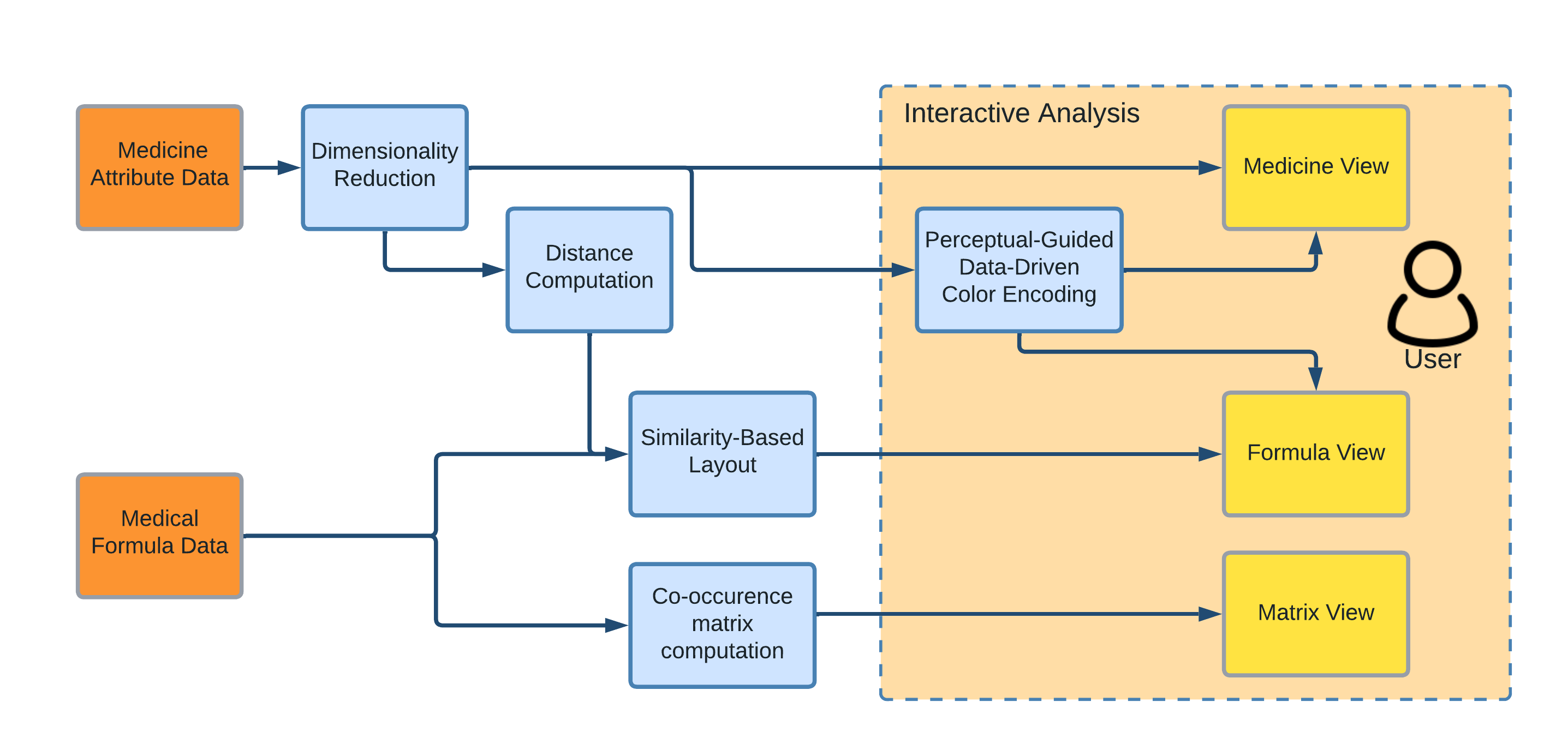


Fig. 1: The workflow of our method.

### Dimensionality Reduction and Distance Computation

Attributes used in our method are Siqi and Wuwei in an M-dimensional space (M=23). The Siqi and Wuwei attributes of an herbal medicine can be written as a vector **P** of binary valued elements:



The M-dimensional space is then dimensionality reduced to 2D, and an herbal medicine can be represented with a 2D vector **p**:

.

UMAP [21] is used for its structure preservation ability and computational efficiency.

The distance between medicines is the basis of our subsequent similarity-based layout computation and visualization. We define the distance *d(u, v)* between two medicines *u* and *v* as the L2-norm, i.e., Euclidean distance, between their corresponding 2D vectors **pu** and **pv**, respectively:

.

The distance between **Pu** and **Pv** in the original M-dimensional space is also considered. However, our experiment shows that discriminating medicines based on the distance with **P** is more difficult than with the projected vectors **p**, and makes the resulting visualization more difficult for comparison and comes with more visual clutter.

### Formulas Visualization

#### Domain Expert Evaluation of Set Visualization Methods

Typically, a dozen formulas and even more medicines are involved in a category of formulas. From a set visualization perspective, both the number of sets and set elements are large; therefore, it requires a suitable visualization that scales well and is easy to understand.

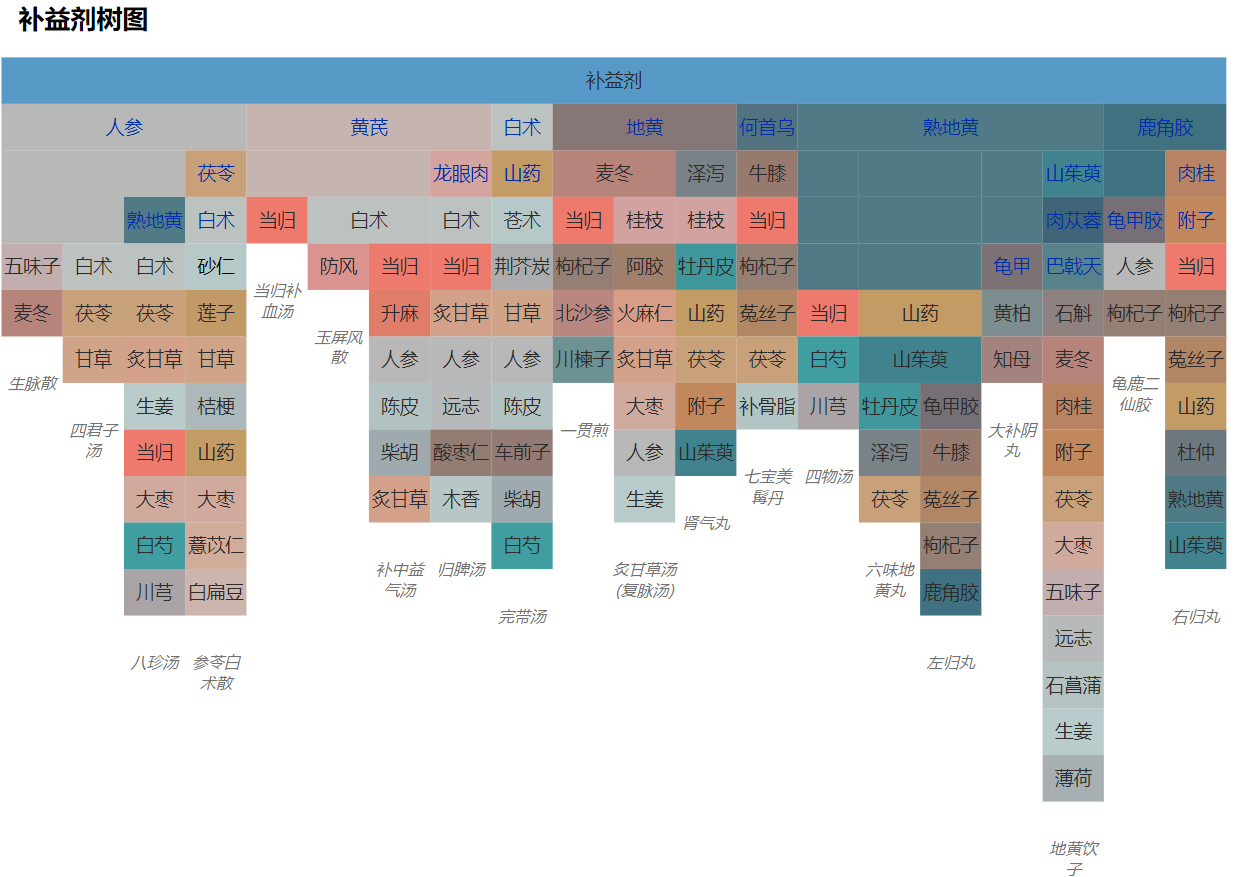
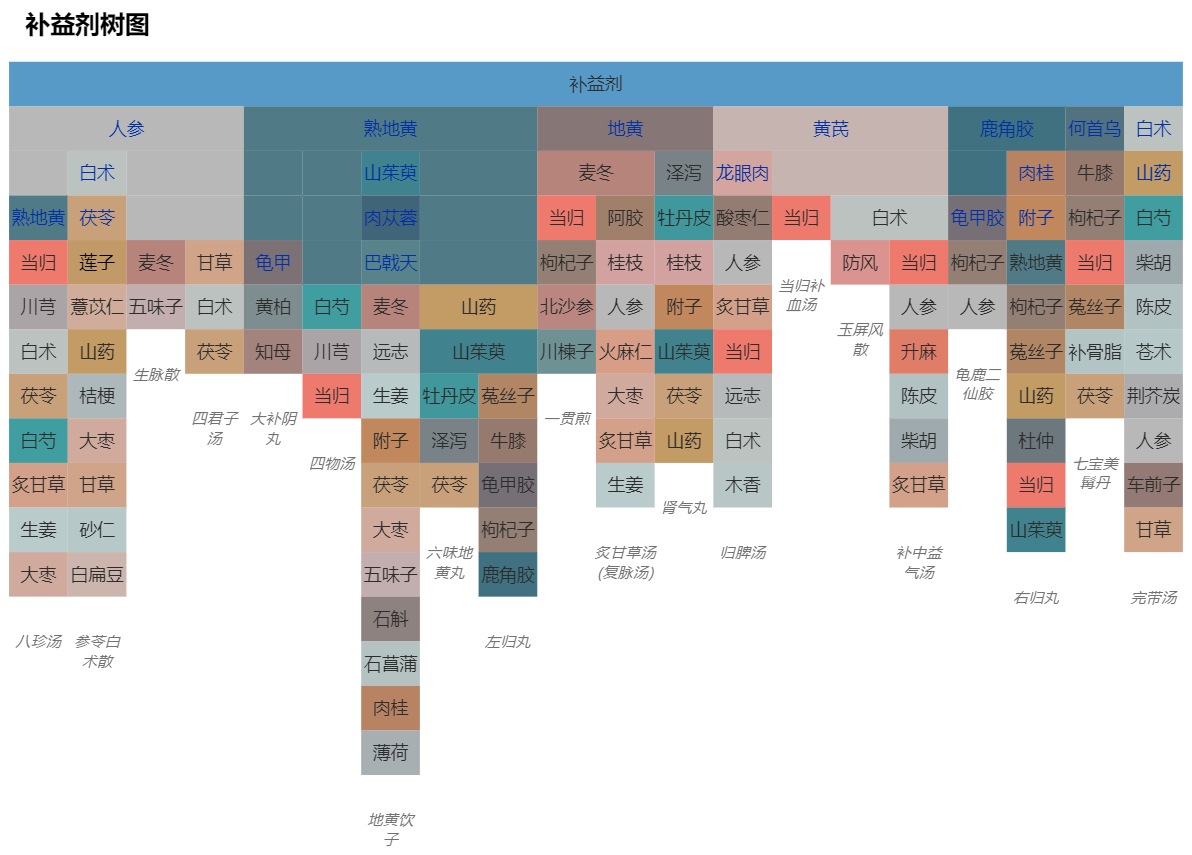
We performed an evaluation of popular sets visualization techniques for the design of a proper set visualization method with the TCM expert (SP). Figures of a Euler diagram, a node-link diagram, and matrix-based methods included in a set visualization survey paper [2] were shown to the TCM expert. The expert was asked to rank the feasibility of these methods for medicine formulas visualization based on the scalability, the ease of understanding, and the support of comparison. The matrix-based method is ranked first by the TCM expert, followed by the node-link diagram, the Euler diagram, and the overlay. The expert found that the matrix-based method is easy to understand and scales to a large number of sets and set elements. However, she was concerned about the compactness of this method as the matrix is typically large but sparse for medicine formulas data. Also, she found it difficult to trace elements of a set and compare sets if the matrix is large. For the node-link diagram, the expert found it easier to understand than the matrix, but the crossings of links make tracing difficult for many sets. The expert preferred the intuitiveness of the Euler diagram and the overlay. However, they did not scale well for the number of sets. Especially, overlapping sets, typically the case of medicine formulas data as formulas in a category, share the same medicines quickly makes these methods unusable. Therefore, the expert considered the Euler diagram and the overlay infeasible for our case.

Based on this informal evaluation, we decided to devise a sparse matrix-based method based on the evaluation to show formulas and corresponding medicines to meet requirements R1 and R2. To support the analysis of overlapping medicines within formulas, a co-occurrence matrix view is used to complement the formulas view.

#### Icicle Plot of Medicine Formulas

Our formula-medicine matrix (set-element matrix) treats formulas (sets) as columns and medicines (elements) as rows. The formula-medicine matrix can be shown with a sparse representation as a collection of formula columns of their corresponding medicine rows. This representation is similar to an icicle plot for hierarchy visualization. It has the potential to support the comparison of similar medicine formulas If properly laid out. Furthermore, the icicle plot allows for encoding medicines in a hierarchy to separate principal medicines and other medicines.

Each record in the medicine formula data contains the name of the formula, names of medicines, and tags for principal medicines (Table [1).](#_bookmark1) We set the content of elements of the icicle plot to names of medicines, and use each column to show a medicine formula as shown in Figures [2](#_bookmark9) and [3.](#_bookmark10)



1. (b)

Fig. 2: Icicle plots with (a) the original order of medicine formulas data and (b) our similarity-based layout.

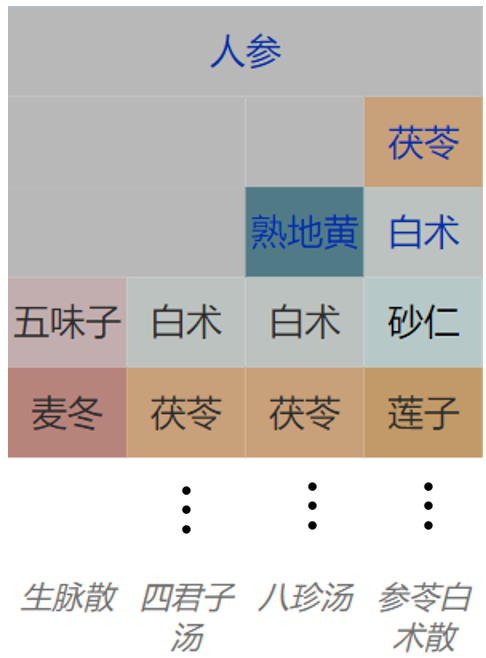


Fig. 3: The design of the icicle plot of medicine formulas. Each column of the icicle plot contains a medicine formula, which comprises principal medicines (texts in blue) and other medicines (texts in black). The name of formula is placed under its column.

In our design, principal medicines are highlighted and treated differently than other medicines to meet requirement R3. As shown in Figure 3, principal medicines are placed on the top levels of the hierarchy and colored blue. Formulas with common principal medicines are grouped together. Rows are padded so that the top of all none- principal medicines are aligned for comparison (R2). For example, rows are padded for Ginseng (人参) as seen in Figure 3. The name of a medicine formula is placed under its corresponding column in italic font face with a fixed vertical spacing as shown in Figure 2. This design is simple yet effective: the height of each column is used as an additional cue to the horizontal position for quick alignment of a formula and its name.

Since the set-based formula information has to be converted into columns of the icicle plot, an ordering is needed for medicines in a formula. However, medicines in the original data have no specific ordering: the resulting icicle plot of medicine formulas of tonic formulas with the initial ordering of medicines has been shown in Figure 2 (a). The plot is cluttered and comparing elements of medicine formulas is difficult as frequent context switch has to be made while searching for a same medicine. Therefore, we propose a similarity-based layout method to facilitate easier comparison and clearer visualization of medicine formulas than using the original ordering.

#### Similarity-Based Layout Computation

Our goal is to design a layout based on the corresponding medicines' similarity so that those with similar attributes are grouped together to enable effective comparison and reduce visual clutter in the icicle plot. Our method is an efficient greedy algorithm with two steps: first, the arrangement of principal medicines, and then we arrange the remaining medicines.

To facilitate the explanation, we introduce the similarity sequence  for a set of medicines . The element *si* of *S* reads:



where ** is the distance between *s* and *h* with Equation 3, and *t* is a random number between 1 and *n*.

**Arrangement of Principal Medicines**

In this step, columns of the icicle plot are sorted based on the similarity of principal medicines. If a medicine is the only principal medicine in a certain medicine formula, it is assigned as the top-level principal medicine. We denote the set of all such medicines as *Hs*. The first element *s1 = ht* is randomly selected from the set *Hs*, and the rest of the sequence is set by finding the medicine *hj* with the shortest distance to the previously ordered element *si*. The sorted top-level principal medicines are placed on the first row of the icicle plot.

We now process formulas with more than one principal medicine. For a top-level principal medicine *hi*, a set  denotes all medicine formulas that have *hi* as a principal medicine, and the *j*-th formula in  is denoted as . If its principal medicine contains elements of *Hs*, add that formula to set ; if none of the principal medicines in a formula is contained in *Hs*, *hj* is selected randomly as the top-level medicine and added to *Hs*. An example is Wandaitang (完带汤) in Figure 2.

For each , formulas with single principal medicine are sorted from left to right by the number of remaining medicines; the number of principal medicines sorts formulas with multiple principal medicines. Principal medicines that are not top-leveled are sorted according to the distance and laid out as subsequent children nodes (as rows). The padding ensures that all non-principal medicines start on the same row.

For example, Figure 3 visualizes set  of Figure 2, with Ginseng (人参) as the top-level principal medicine, and Bazhentang (八珍汤) and Shenlingbaizhusan (参苓白术散) have more than one principal medicines (columns 2 and 3, respectively). Therefore, the principal medicine rows are padded to three rows as Shenlingbaizhusan has a maximum of three principal medicines.

**Arrangement of Remaining Medicines**

Next, the remaining medicines are arranged. Define the medicine in the *j*-th column, and *k*-th row in the *i*-th multiple medicine formulas set  in the icicle plot as *hijk*. Each formula column is then the previously introduced set. We construct the position sequence of  as from left to right. The leftmost column is sorted by the distance-based ordering using Equation 3. Starting from the second column from the left, medicines are sorted by local similarity. We align the same medicines in adjacent columns even if they are not from the same . Other medicines are sorted according to their distances to medicines on its left column within the set :



If  contains more elements than , construct the hierarchy with similarity ordering as in Equation 4.

Figure 2 (b) shows the icicle plot of tonic formulas with the new similarity layout. Compared to the original layout (Figure 2 (a)), the alignment of medicines is improved and the same medicines in adjacent columns are aligned vertically. For example, note how Baizhu (白术), Fuling (茯苓), and Ginseng (人参) are aligned as non-principal medicines in Figure 3 (b), whereas in Figure 2 (a), such alignments are nonexistent.

### Visualization of Shared Medicines in Formulas

The icicle plot visualization allows effective visualization of the comprising medicines of formulas and the comparison of neighboring formulas. However, comparing formulas that are far apart, e.g., having different principal medicines, in the layout is also necessary to provide a deeper understanding of formulas. Therefore, a co-occurrence matrix view of formulas is included to complement the icicle plot. The benefit of using a matrix view is that all formulas' complete pair-wise intersection information can be effectively represented and easily identified.

As shown in Figure 4, the matrix has formulas as rows and columns, and an element of the matrix is the number of shared medicines of the corresponding formulas of its row and column. With a sequential color map, this view allows the user to quickly examine each formula's overlapping information against all others by focusing on a row or a column. Also, the color encoding effectively draws the attention of the user to formulas with the highest number of shared medicines: in this case, Zuoguiwan (左归丸) and Youguiwan (右归丸) as highlighted in red in Figure 4.

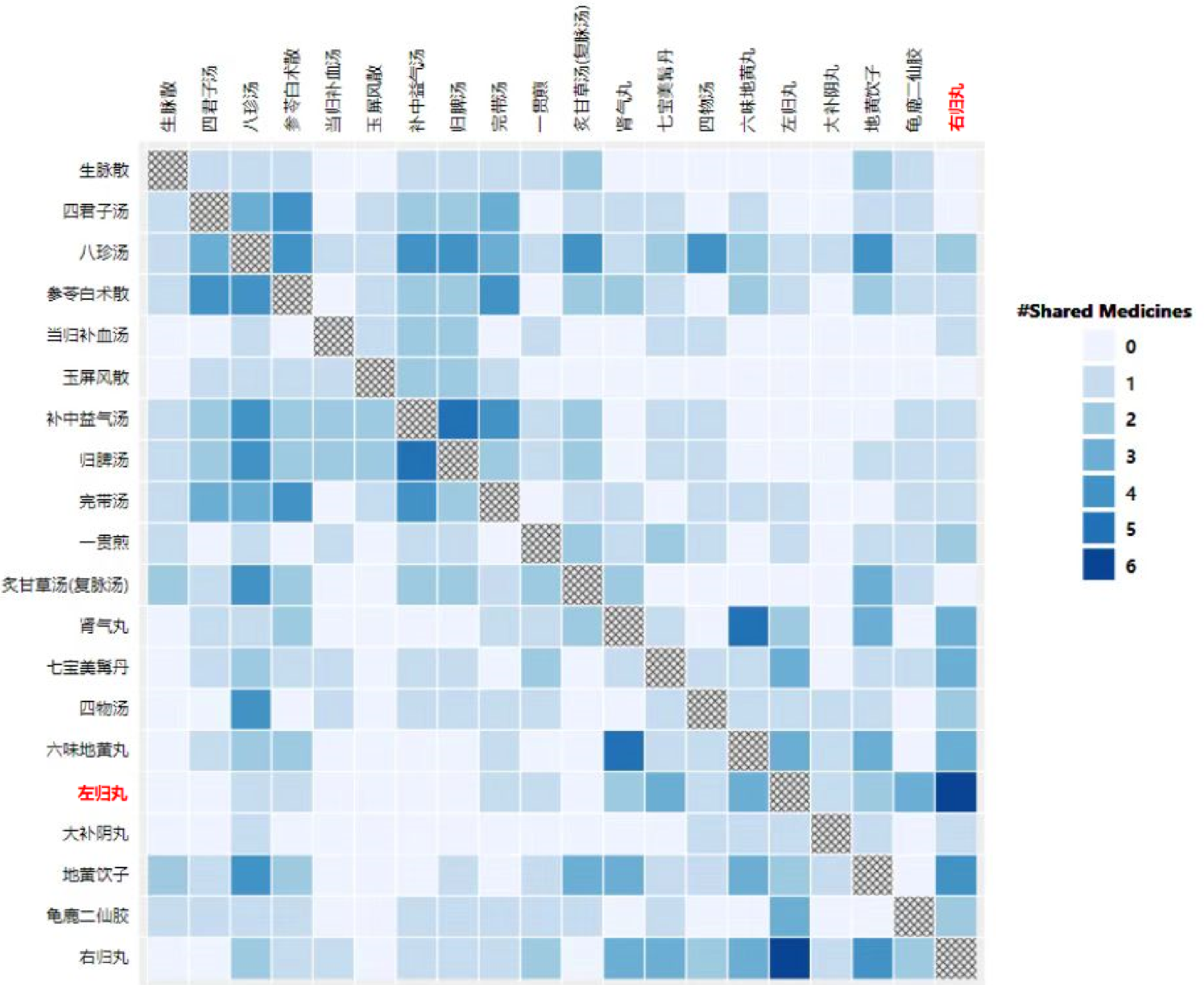


Fig. 4: The shared medicine matrix view of formulas.

### Perceptual-Guided Data-Driven Color Encoding

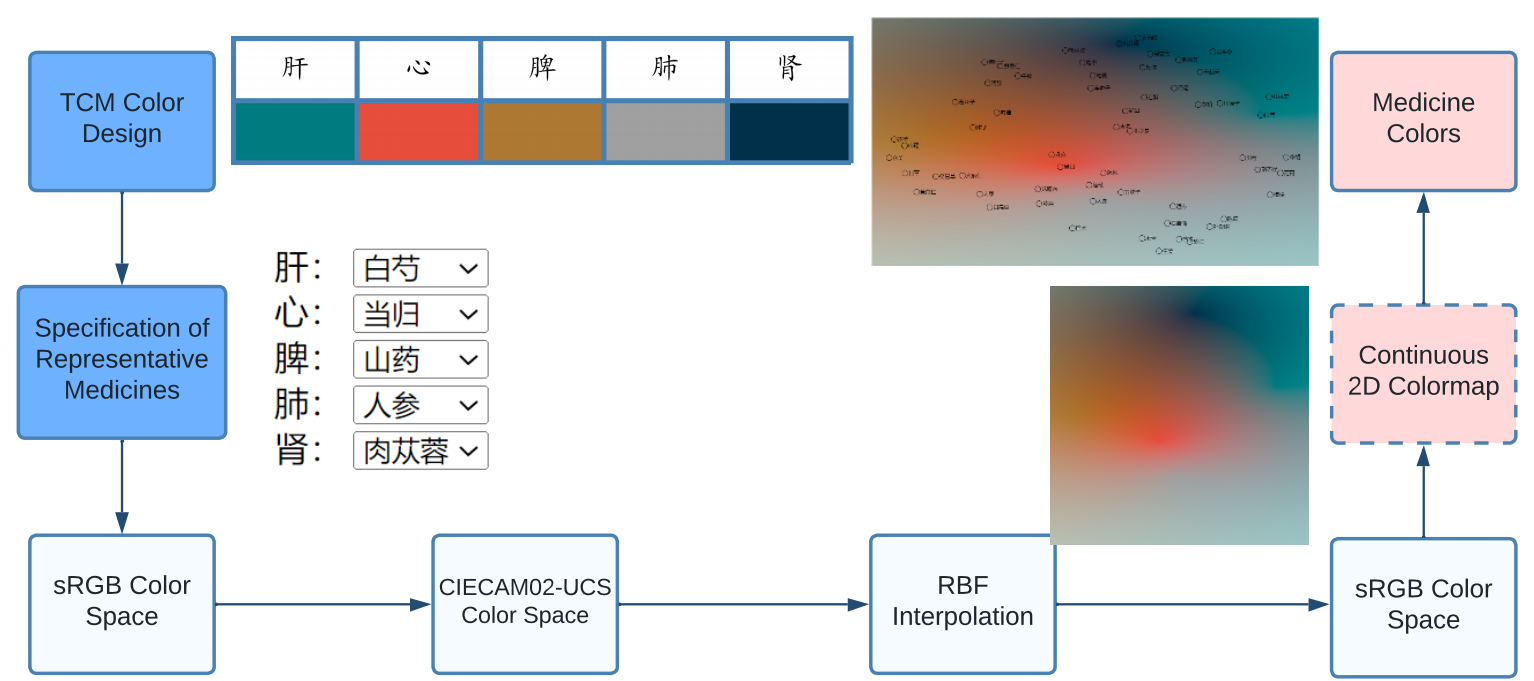
Colors are assigned based on multidimensional attributes in the medicine and formula views. Color encoding is guided by perception that assists the quick understanding of attributes of medicines and distances between medicines. The workflow of our color encoding method is illustrated in Figure 5. The method is based on the 2D dimensionality reduced space derived from the multidimensional medicine attribute data and requires the knowledge of users to identify representative medicines within it. For a group of medicine formulas, medical experts can identify several representative medicines based on their TCM attributes with our TCM-concept-inspired colors (R7). These colors are transformed into a perceptual uniform color space and interpolated with radial basis functions in there to get the medicine colors and/or the continuous 2D color map that spans the entire dimensionality reduced attribute space.

Fig. 5: The pipeline of our color encoding method.

#### TCM-Concept Inspired Representative Color Design

Colors of representative medicine are carefully chosen to show TCM concepts. These TCM concepts include Wuxing (五行), Wuse (五色), and Wuzang (五脏) as summarized in Figure 6. The associated colors are handpicked to show the connection to Wuse with perceptual and aesthetic considerations—the luminance of colors should not vary too much, and saturated colors should be avoided. Initially, we experimented with isoluminant colors that are beneficial for metric comprehension. However, the TCM expert considers the resulting colors are not distinct enough in the medicine view. As a trade-off, the representative colors are chosen to have a relatively small range of luminance (14 ≤Y ≤ 62).

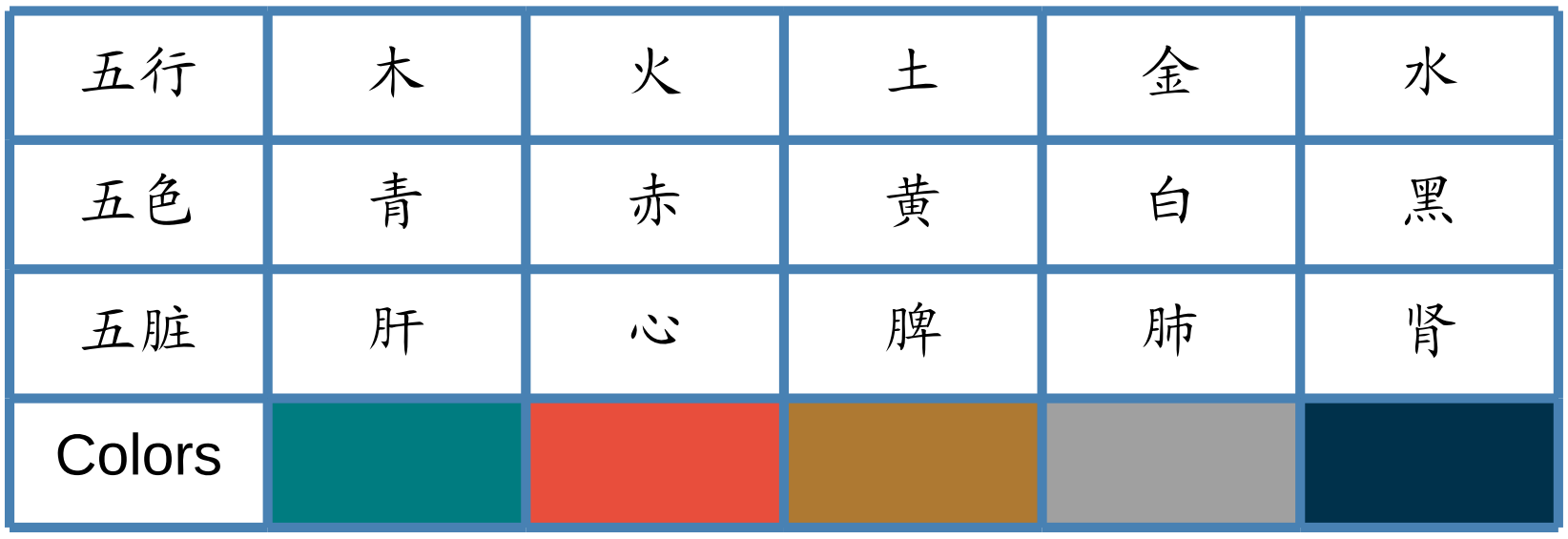


Fig. 6: Colors designed for medicine based on TCM concepts.

#### Perceptual-Uniform Color Space

For perceptual uniformity, we use the CIECAM02-UCS color space[20] to calculate colors of the remaining medicine with color interpolation. This color space is spanned by three parameters  that is calculated by first transforming XYZ color stimuli to the channels of the CIECAM02 color appearance model followed by a transformation from CIECAM02 to CIECAM02-UCS. As shown in Figure 5, we transform the colors of representative medicine from sRGB to CIECAM02-UCS through CIEXYZ. Then, the RBF interpolation is performed for *J*,and  channels, respectively. Next, the interpolated colors are converted back to sRGB for display.

#### Radial Basis Function Color Interpolation

Radial basis function (RBF) interpolation enables the interpolation of unstructured data, e.g., a few scattered points or point clouds—making them a natural choice for our method. The RBF interpolation function s(**x**) at location **x** can be written as:

,

where *K* is the number of known data points (*K* = 5 according to Figure 6), **x***k* are the locations of known data points,  is the radial basis function of a distance r between **x** and **x***k* (here, the Euclidean distance), and *wk* are unknown weights. If **x** = **x***k* ,, interpolated values should be data values y at that location: s(**x**) = f (**x**) = y.

We experimented with several radial basis functions, e.g., Gaussian, cubic, thin-plate functions, and choose the linear radial basis function:

.

The choice is made due to two reasons: first, the measure of Euclidean distance which matches the distance of medicines, and second, it also generates least duplicate colors among typical radial basis functions.

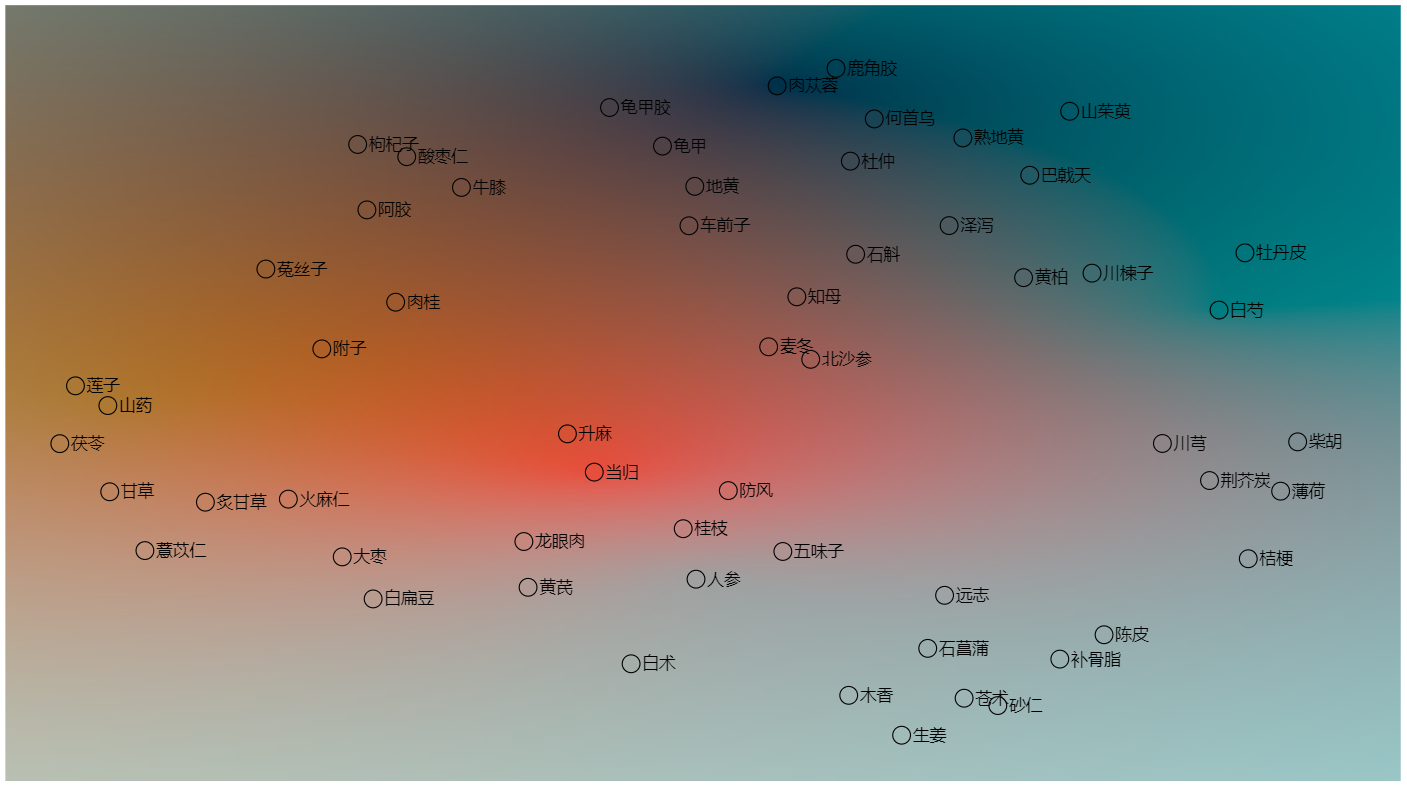
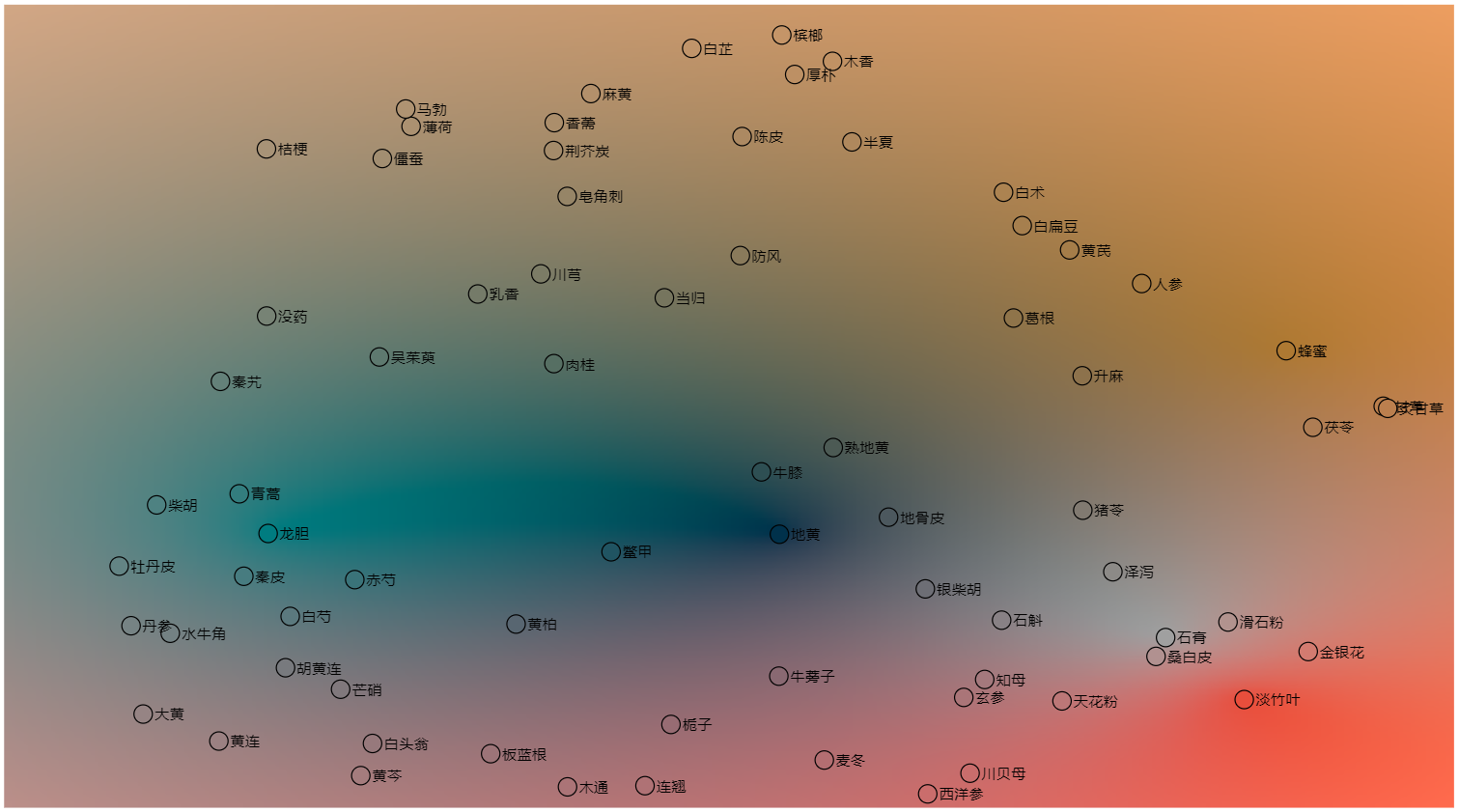
#### Color Assignment

Continuous 2D color maps of two groups of medicine formulas generated by RBF interpolation over the entire 2D domain are shown in Figures 7 (a and b). Smooth transitioning between attributes of medicines can be seen with the 2D color maps, while color differences indicate distances between medicines. Therefore, 2D continuous color maps are a useful tool for examining the distribution of medicines in the multidimensional space of a certain medical formula.

To assign colors to medicines, the 2D location of each medicine in the dimensionality reduced space is used for the interpolation of colors using Equation 6. Medicine colors overlaid on the continuous color map are shown for the two formula groups as shown in Figure 7 (c and d). For efficiency, only the colors of points of medicines shown in medicine formulas need to be calculated if the overall trend in the 2D domain is not the focus.

1. (b)

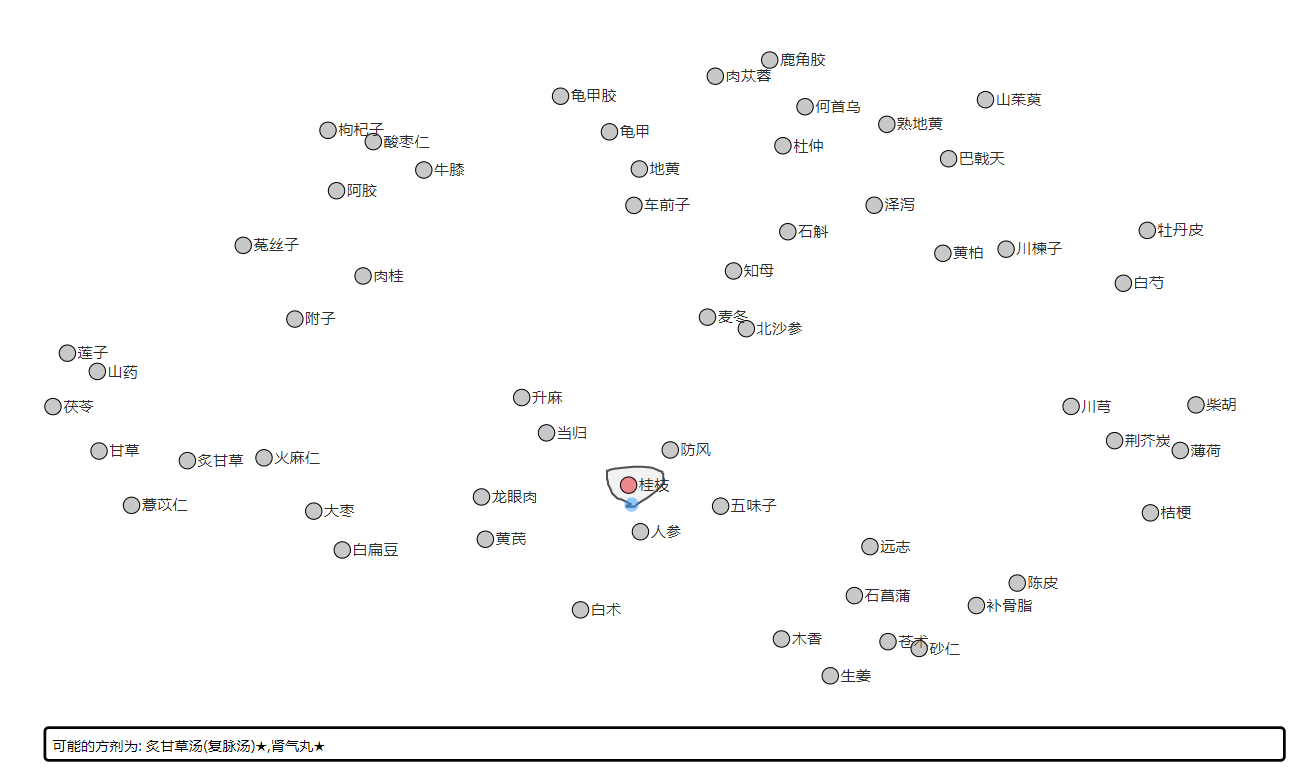
(c) (d)

Fig. 7: Color encoding with our method for tonic formulas (the left column) and heat-clearing formulas (the right column). Continuous 2D colormaps are shown in (a) and (b), respectively. Medicine colors are calculated based on their positions in the 2D domain (c and d).

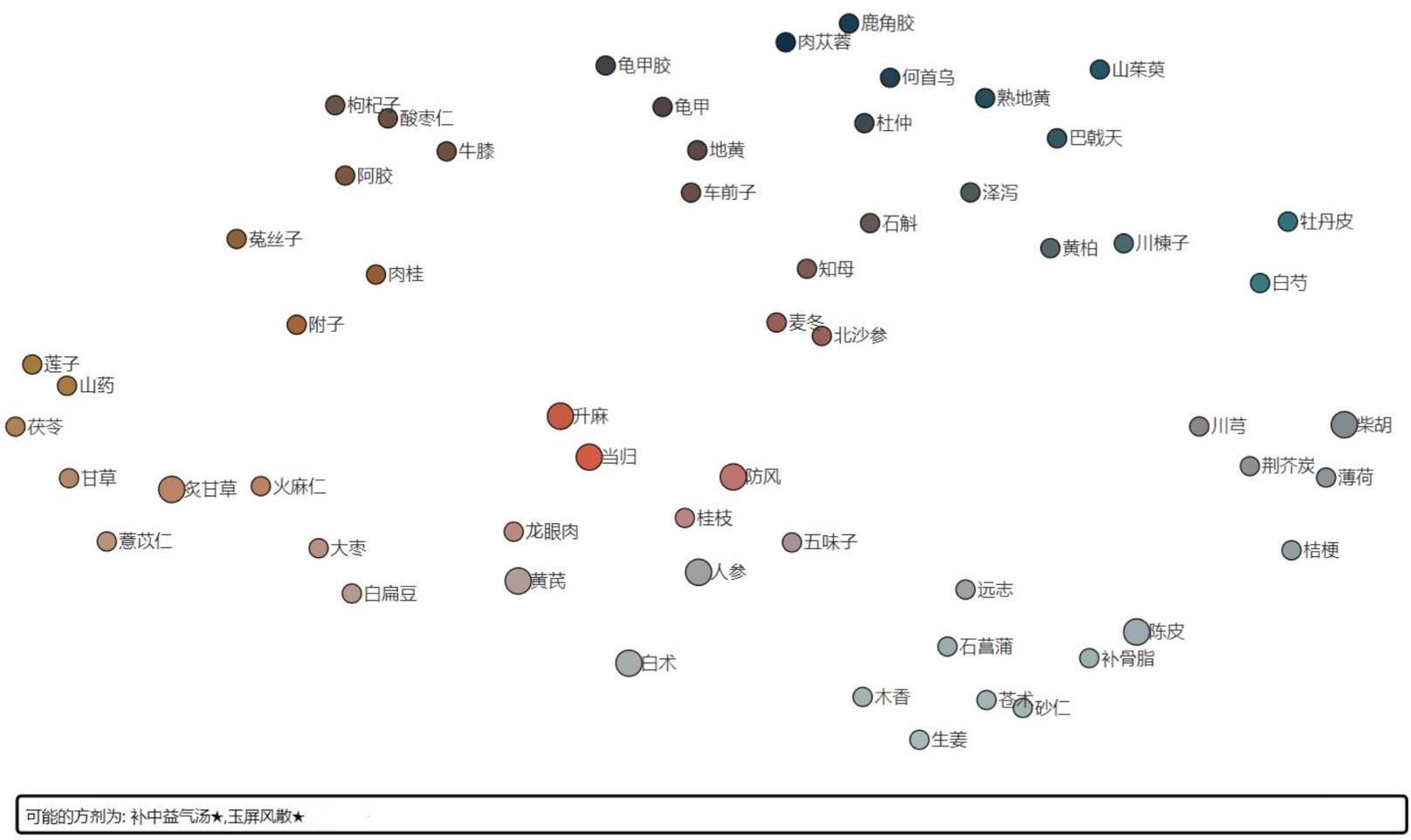
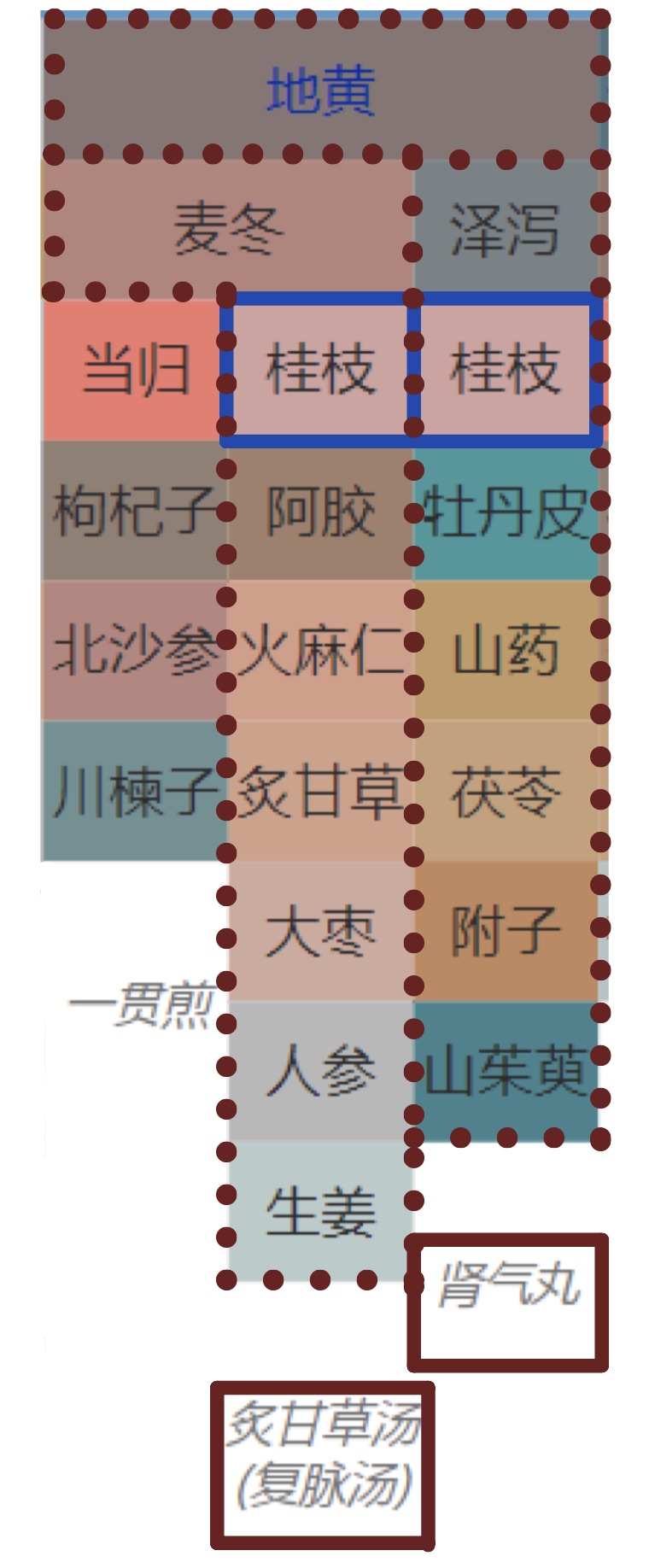
### User Interactions

Our visualization method supports interactive exploration within the formula view, the matrix view, and the medicine view. Brushing-and-linking enables the connections between these three views (R4). In the formula view, the names of all formulas are shown whenever the mouse hovers over a medicine as shown in Figure 8(a). The matrix view highlights corresponding formulas when the mouse hovers over an element. In the medicine view, a lasso tool allows users to flexibly select medicines of interest. All potential formulas are shown as text under the scatterplot of the medicine view (Figure 8(c)). Representative medicines can be assigned and updated through the selection boxes on top of the medicine view (Figure 9). These user interactions are easy to use and intuitive to users who are not familiar with interactive visualization. Therefore, requirement R6 is satisfied.

Brushing-and-linking enables visual connections between the formula view and the medicine view interactively. All medicines are highlighted in the medicine view with enlarged size (Figure 8(b)) if any formula is selected in the formula view (Figure 8(a)). Conversely, whenever any medicine is selected in the medicine view (Figure 8(c)), the formula view is updated as shown in Figure 8(d). Here, all selected formulas are highlighted with blue solid lines, and formulas containing the selected medicines are highlighted with red dashed lines. As a result, brushing-and-linking helps enhance the understanding of users of the composition of medicine in formulas (R5).



1. (b)

(c) (d)

Fig. 8: User interactions in our method: (a) mouse hovering in the formula view and (b) corresponding updates in the medicine view; (c) lasso selection in the medicine view and (d) corresponding changes in the formula view.

### Implementation

Our method is implemented as a web-based interactive visual analysis tool as shown in Figure 9. Data processing procedures were realized in python aided by the 'umap' package for dimensionality reduction, the 'scipy' package for RBF interpolation, and the 'colour' package for color space transformations, Visualization and user interactions were realized in JavaScript aided by the 'D3' package, and the communication between python and JavaScript components is achieved using the 'eel' package.

## Results

### Overview

The evaluation of our method is performed with the analysis of two representative use cases-- tonic and heat-clearing formulas---by two TCM experts (SP and XH). They were asked to analyze the formulas using the web-based tool with think-aloud protocol analysis and provide feedback after the session. Both experts were systematically trained in TCM, and have obtained clinical degrees and certificates in TCM. One has obtained a doctoral degree in TCM (SP), while the other has been working in clinical for over nine years (XH). Both experts have more than fourteen years of expertise in TCM.

After introducing our new method, the participants were asked to explore the medicine formulas data with our visualization tool while the observer observed and talked to the participants. Afterward, they were asked to provide further feedback on the method.

Visualizations of the two use cases presented to the TCM experts as in the web-based tool are shown in Figure 9. There, the three linked views: the dimensionality reduction-based medicine view, the medicine formulas view, and the shared medicine matrix view, are laid out from left to right.



1. Tonic formulas.



1. Heat-clearing formulas.

Fig. 9: Visualizations of two typical groups of medicine formulas with our method: (a) tonic formulas and (b) heat-clearing formulas.

### Statistics of Datasets for Evaluation

The tonic formulas (Figure 9(a)) contain 20 formulas and 58 medicines with 17 principal medicines and a median of 1 principal medicine per formula. The median of medicines per formula is 7.5, with a minimum of 2 and a maximum of 15. The average number of shared medicines of a pair of formulas is 1.09, the median being 1, the maximum of 6, and the minimum of 0.

The heat-clearing formulas (Figure 9(b)) contain 26 formulas and 73 medicines with 25 principal medicines with a median of 1 principal medicine per formula. The median of medicines per formula is 6.5, with a minimum of 2 and a maximum of 17. The average number of shared medicines between a pair of formulas is 0.9822, the median being 1, the maximum of 7, and the minimum of 0.

### Use Cases

Expert PS started the analysis by looking at the overall distributions of medicines and used her knowledge to assign representative medicines for each medicine category listed in Figure 6. The resulting continuous 2D colormaps show that the center of the attribute space of tonic formulas is red (Figure 7(a)), while heat-clearing formulas have the center of its space as green and black (Figure 7(b)). These indicate the different properties of tonic and heat-clearing formulas and are in line with the related TCM concepts.

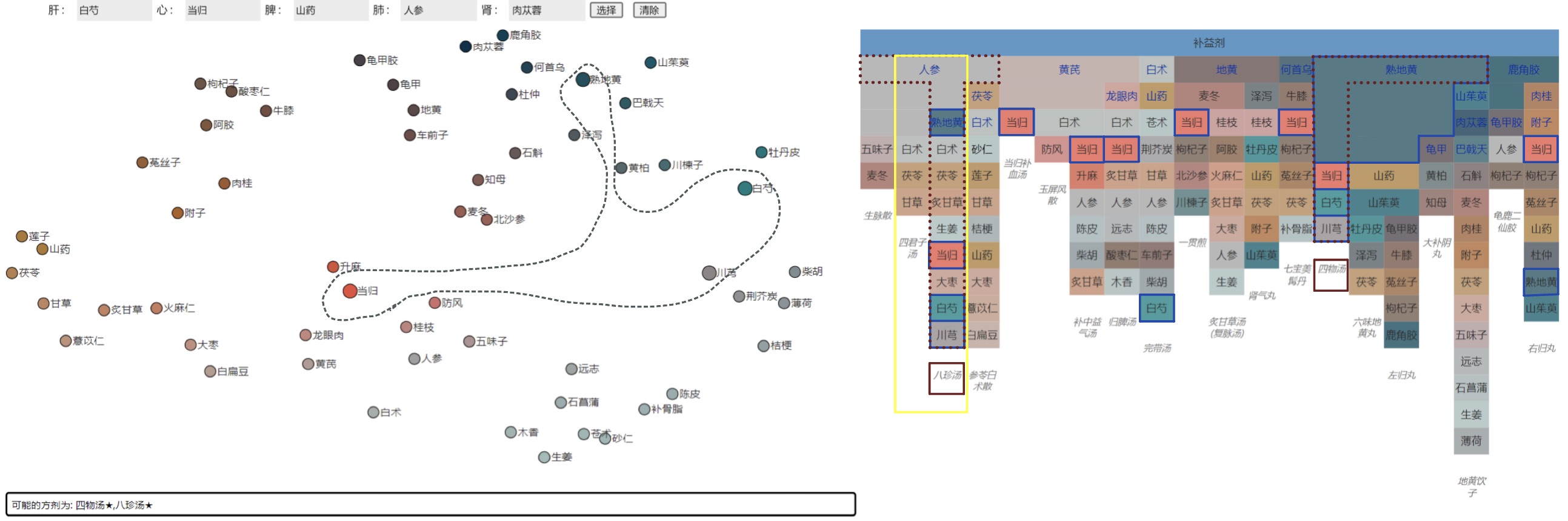


Fig. 10: The analysis of tonic prescriptions with our method. A lasso selects four medicines of interest in the medicine view (left), and corresponding formulas are highlighted in the formula view (right).

In the icicle plot of tonic formulas (Figure 10(right)), it is easily seen that two adjacent columns are similar: the Bazhentang (八珍汤) contains the Sijunzitang (四君子汤) as highlighted in the yellow box. The TCM expert then analyzed the difference between these two formulas. She used the lasso tool in the medicine view to select four other medicines in Bazhentang as shown in Figure 10(left). The text below the scatterplot shows that formulas containing these medicines are Bazhentang and Siwutang (四物汤). These two formulas were selected with red dashed lines and the selected medicines are highlighted with blue solid lines in the formula view (Figure 10(right)). A close examination shows that the lasso selected medicines form Siwutang. Moreover, it can be seen that Bazhentang is the combination of Sijunzitang and Siwutang.

In the matrix view (Figure 9(a-right)), most formulas have overlapping herbal medicines with Sijunzitang(四君子汤) and Bazhentang(八珍汤), suggesting that tonic formulas are built on the medicine composition of these two formulas.

It is known that the main role of Sijunzitang or Bazhentang is "invigorating Qi and blood". The understanding of Qi and blood in TCM is the basic substance of human body, which can reflect the importance of all supplements to Qi and blood in the matrix view. Meanwhile, Yin and Yang are two interdependent, opposite, complementary, and exchangeable aspects of nature. Qi is Yang (阳, positive), blood is Yin (阴, negative), Qi and blood are dependent. TCM physicians usually prescribe for diseases in which Qi and blood deviate from balance. The expert considered that this visualization is suitable for beginners to pay attention to the "Qi and blood" supplement for tonic formulas.

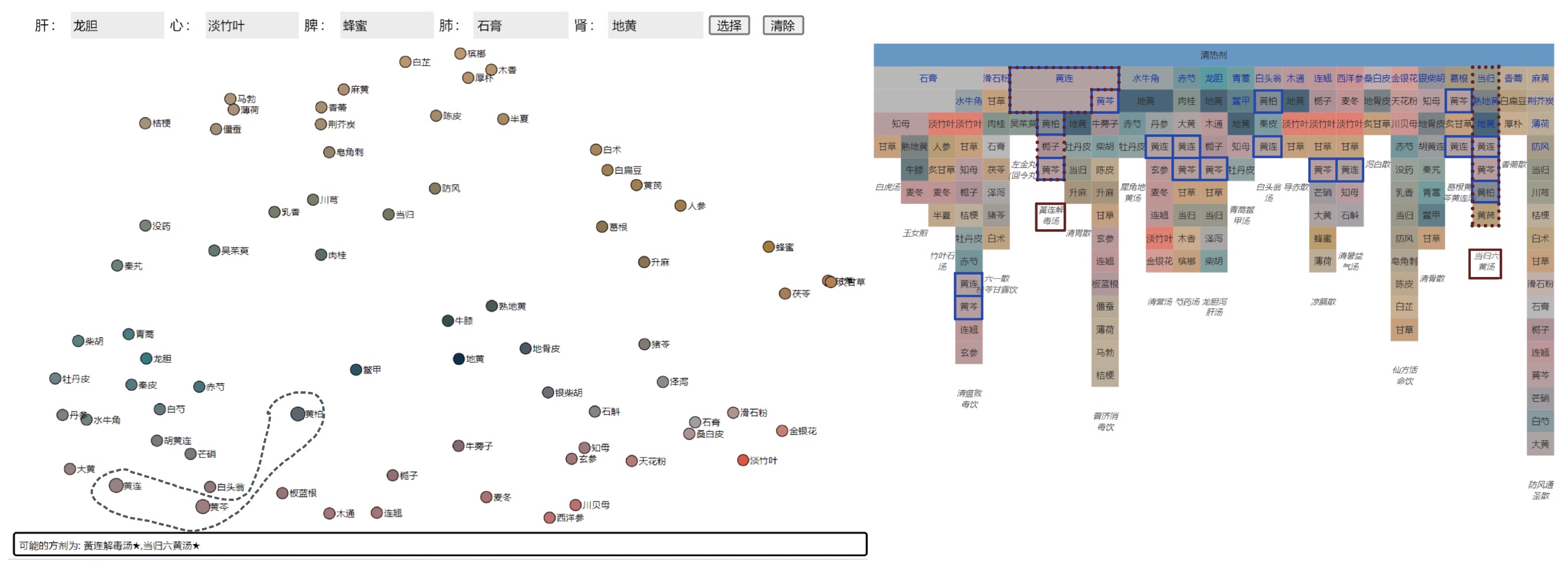


Fig. 11: Interactive analysis of heat-clearing formulas with our method.

The analysis of heat-clearing formulas is shown in Figure 11. TCM expert XH was interested in Sanhuang (三黄): Huanglian (黄连), Huangqin (黄芩), and Huangbo (黄柏), which is a commonly used medicine combination for clearing heat and detoxification in TCM. The three medicines are relatively close in the medicine view (Figure 11(left)), and the expert used a lasso to select them. Both Huanglian-jiedutang (黄连解毒汤) and Danggui-liuhuangtang (当归六黄汤) contain Sanhuang as suggested by the text below. The expert further examined the formula view (Figure 11(right)) where these two formulas were highlighted. According to the medicine attributes, the function of Huanglian-jiedutang is to clear heat and detoxify. While the composition of Danggui-liuhuangtang contains tonic medicines, meaning that in addition to clearing heat and detoxification, it also has the effect of nourishing Yin (滋阴). Unlike the tonic formulas, not many overlaps are seen in the matrix view (Figure 9 (b-right)). Most formulas have overlapping herbal medicines with Qingwenbaiduyin (清瘟败毒饮), which have the function of clearing heat and detoxification. This can be a reminder for beginners to pay attention to the relationship between this formula and other formulas in the heat-clearing formulas.

Both experts made positive comments on the coloring of medicines. For example, Danggui (当归) is a blood tonic medicine and corresponds to red. On the other hand, Shigao (石膏) works on the lungs and is colored white.

### TCM Expert Feedback

Overall, both experts think that our method is able to clearly disassemble complex formulas and assist the memorization of their functionalities. The interactive visual analysis process is new to them and is helpful for enhancing their understanding of formula composition theories by making and testing their own hypothesis. They think that the color encoding of medicines allows TCM students and beginners to understand the effect of medicines more intuitively and facilitates memorization. Beginners have difficulty understanding the similarities and differences between multiple similar formulas. With the lasso tool, beginners can test multiple medicine combinations to better understand the similarities and differences between formulas and, therefore, better understand an actual prescription. Also, they consider brushing-and-linking a beginner-friendly way to understand the relationships between medicines and formulas.

The experts suggest that other than assisting the learning of TCM formulas for beginners, the method can be extended to facilitate the learning of actual treatment plans for TCM physicians. The TCM theory system includes the process of "theory, method, formula, and medicine", and a treatment plan with prescriptions is performed to assess the effectiveness of formulas. The experts suggest supporting multiple lassos as future work to facilitate the building up of a prescription by adding medicines from an initial known set of medicines to learn actual treatment plans.

## Discussion

### Principal Results

The proposed method is able to effectively visualize complex TCM formulas and multidimensional medicine attribute information. The joint analysis of medicine formulas and corresponding medicines is possible with user interactions and brushing-and-linking between multiple views within our web-based tool.

### Limitations

Our method does not directly support the visualization of overlaps of more than two medicine formulas, i.e., intersections of more than two sets. However, such information can be implicitly gained by visual searching in the medicine formula view and interactively selecting medicines of interest that would highlight all formulas containing the shared medicines.

Another limitation is that the dimensional reduction view does not explicitly show the multidimensional properties but the relative distances between medicines. This could be addressed using additional multidimensional visualization techniques such as parallel coordinates.

### Comparison with Prior Work

Visualization methods are used in the research of TCM. Visual recognition and visualization are proposed for TCM pulse information—the pulse information is quantified and visualized to support a more accurate diagnosis [37]. Digital tongue images that are important in TCM are recognized and analyzed with a visualization of tongues [43]. Infrared thermal imaging visualization enables users to see and assess physiological states or pathological conditions intuitively as the temperature of local tissues or the whole body may change due to illness [26]. Visualization based on a 3D human model of Chinese medicine pulses could facilitate the teaching, understanding, and communication of meridians and acupoints [40].

Specific visualization methods for herbal medicine formulas are available. Cold and hot properties are visualized as indicators for herbal medicine formulas in a formula analysis platform [7]. However, this method covers only two properties and does not show the multidimensional attributes of medicines. Knowledge graph visualization is proposed for many medicine formulas through manual and natural language processing [9]. Network visualization is used to show the composition of medicine formulas to assist in constructing medicine formulas databases [44]. However, these methods rely on querying based on text input and do not support interactive visualization and analysis. A visual analysis method of TCM health records has been available recently due to the collaboration between TCM and visualization experts [11]. This method supports the analysis of time-varying TCM health records and comparing medicines in formulas of different patients. However, no visualization method is available for herbal medicine formulas yet.

Set is an important research subject in visualization. Set visualization techniques are reviewed in a survey by Alsallakh et al. [2]. The visualization of set memberships can be categorized into different strategies, including Euler and Venn diagrams [3, 35, 33, 23, 41, 34, 28], node-link diagrams [36, 6, 24], matrix-based methods [29, 22, 18], and aggregation methods [1, 13, 10]. Matrix-based methods support a large number of sets and elements, and also all set relationships. However, the full representation of matrix is often spatially inefficient for a large row or column numbers. In our case, the matrices of sets are sparse; therefore, we use a sparse matrix representation to show the set information, i.e., the formulas information, as an icicle plot.

The icicle plot[14] is a popular hierarchical data visualization technique. Hierarchical data visualization techniques can be classified into explicit techniques, i.e., trees using node-link diagrams, and implicit techniques that no explicit edges are drawn. Implicit hierarchy visualization techniques are summarized in an extensive survey [31], and the main benefit of implicit techniques is the efficient use of space making them more suitable for large hierarchical data than trees. Popular implicit methods include treemaps [12, 32] and icicle charts[14]. Icicle plots utilize the vertical placement of nodes to show the hierarchical information which naturally allows showing the list of medicines within a formula. With our augmented icicle plot in the similarity-based layout, our TCM experts consider it easy to understand and allows for quick comparison of formulas.

Multidimensional data can be effectively visualized using dimensionality reduction techniques[38]. Nonlinear dimensionality reduction methods [16] are more suitable to preserve complex high-dimensional structures than linear methods[5]. Currently, t-SNE[15] and UMAP[21] are the most popular nonlinear dimensionality reduction methods as they could preserve the neighboring information in the high-dimensional space. We choose UMAP in our method as it is more efficient and overcomes several limitations of t-SNE.

Perception is important in visualization, and color perception, among others, is most relevant in our case. With numerous medicines and formulas, colors provide critical classification information that assists the users for the reasoning of formulas composition. A survey of the use of colors in visualization can be found elsewhere [45]. A key concept for the effective use of colors is perceptual uniformity, i.e., the perceived color difference should match the data value difference. Perceptual uniformity is used for color map design[27, 17], and an online tool ColorBrewer is readily available[[1]](#footnote-1), but these methods do not support flexible color specification for 2D color maps. Instead, we propose a color encoding method for drugs based on a 2D color map created by RBF interpolation with user-specified representative drugs assigned with TCM concept-inspired colors. To achieve perceptual uniformity, colors has to be computed in a uniform color space. CIELab is perhaps the most well-known perceptual uniform color space [4]. However, studies show that the uniformity performance of CIELab is not satisfactory [20]. Recently, several color spaces based on the CIECAM02 color appearance model [25] with better uniformity than CIELab are available. In our method, we choose the CIECAM02- UCS color space[20] for its good performance.

### Conclusions

We have introduced a visualization method for TCM formulas. Requirements and design choices of our method are made through a close collaboration between visualization and TCM experts in an iterative, quick-prototyping fashion. Our method supports interactive visualization of medicine formulas with a similarity-based layout complemented by a matrix view of shared medicines by formulas, and multidimensional attribute data of medicines are visualized with a dimensionality-reduction method. Colors of visual elements are assigned with a perceptual-guided data-driven color encoding method that achieves perceptual uniformity and reflects TCM concepts. The web-based tool that implements our method supports interactive analysis and comparison of medicine formulas and corresponding medicines with brushing-and-linking between different views. The evaluation of our method with TCM experts demonstrates the effectiveness of our method for joint TCM formula composition and medicine property analysis. Further feedback from the experts suggests that our method has potential for educating TCM formula composition theories, and modernizing TCM inheritance methods.

### Acknowledgements

We thank Xiaoxuan Hu for participating in the expert evaluation, and providing valuable insights and suggestions for improvements. The funding information will be added if the paper is accepted.

### Conflicts of Interest

None declared.

### Abbreviations

CIECAM02: International commission on illumination (CIE) color appearance model 2002

CIECAM02-UCS: CIECAM02 uniform color space

RBF: Radial basis function

TCM: Traditional Chinese medicine

UMAP: Uniform manifold approximation and projection for dimension reduction

## References

[1] B. Alsallakh, W. Aigner, S. Miksch, and H. Hauser, Radial Sets: Interactive Visual Analysis of Large Overlapping Sets, IEEE Transactions on Visualization and Computer Graphics, vol. 19, no. 12, pp. 2496–2505, 2013.

[2] B. Alsallakh, L. Micallef, W. Aigner, H. Hauser, S. Miksch, and P. Rodgers, The State-of-the-Art of Set Visualization, Computer Graphics Forum, vol. 35, no. 1, pp. 234–260, 2016. [Online]. Available: https://­onlinelibrary.wiley.com/­doi/­abs/­10.1111/­cgf.12722

[3] M. E. Baron, A Note on the Historical Development of Logic Diagrams: Leibniz, Euler and Venn, The Mathematical Gazette, vol. 53, no. 384, pp. 113–125, 1969. [Online]. Available: http://­www.jstor.org/­stable/­3614533

[4] CIE, “Colorimetry, 4th edition,” Commision Internationale de l’Eclairage, Tech. Rep., 2018.

[5] J. P. Cunningham and Z. Ghahramani, Linear dimensionality reduction: Survey, insights, and generalizations, Journal of Machine Learning Research, vol. 16, no. 1, pp. 2859–2900, 2015.

[6] M. Dörk, N. Henry Riche, G. Ramos, and S. Dumais, PivotPaths: Strolling through Faceted Information Spaces, IEEE Transactions on Visualization and Computer Graphics, vol. 18, no. 12, pp. 2709–2718, 2012.

[7] J. Gao, Construction of Visual Analysis Platform for Cold and Heat Properties of Formulae Based on Quantitative Study, Ph.D. dissertation, Beijing University of Chinese Medicine, 2009.

[8] X. Gao, Chinese Pharmacy. Beijing: The Press of Traditional Chinese Medicine, 2004.

[9] W. Guo, Research and Implementation of Knowledge Mapping of Traditional Chinese Medicine Prescription, Lanzhou University, 2019.

[10] H. Hofmann, A. P. J. M. Siebes, and A. F. X. Wilhelm, Visualizing Association Rules with Interactive Mosaic Plots, in Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2000, pp. 227–235. [Online]. Available: https://­doi.org/­10.1145/­347090.347133

[11] X. Hu, S. Peng, H. Hou, N. Yang, Y. Lv, and l. Zhou, Visual Analysis of Traditional Chinese Medicine Health Records, Journal of Computer-Aided Design and Computer Graphics, vol. 33, no. 12, p. 10, 2021.

[12] B. Johnson and B. Shneiderman, Tree-maps: a space-filling approach to the visualization of hierarchical information structures, in Proceeding Visualization '91, 1991, pp. 284–291.

[13] R. Kosara, F. Bendix, and H. Hauser, Parallel Sets: interactive exploration and visual analysis of categorical data, IEEE Transactions on Visualization and Computer Graphics, vol. 12, no. 4, pp. 558–568, 2006.

[14] J. B. Kruskal and J. M. Landwehr, Icicle plots: Better displays for hierarchical clustering, The American Statistician, vol. 37, no. 2, pp. 162–168, 1983. [Online]. Available: http://­www.jstor.org/­stable/­2685881

[15] E. P. L. van der Maaten and H. van den Herik, Dimensionality reduction: A comparative review, Tilburg University, Tech. Rep., 2009.

[16] J. A. Lee and M. Verleysen, Nonlinear Dimensionality Reduction. New York: Springer-Verlag, 2007.

[17] H. Levkowitz and G. Herman, Color scales for image data, IEEE Computer Graphics and Applications, vol. 12, no. 1, pp. 72–80, Jan 1992.

[18] A. Lex, N. Gehlenborg, H. Strobelt, R. Vuillemot, and H. Pfister, UpSet: Visualization of Intersecting Sets, IEEE Transactions on Visualization and Computer Graphics, vol. 20, no. 12, pp. 1983–1992, 2014.

[19] J. Li, Chinese Herbal Formulas. Beijing: The Press of Traditional Chinese Medicine, 2016.

[20] M. R. Luo, G. Cui, and C. Li, Uniform colour spaces based on ciecam02 colour appearance model, Color Research & Application, vol. 31, no. 4, pp. 320–330, 2006. [Online]. Available: https://­onlinelibrary.wiley.com/­doi/­abs/­10.1002/­col.20227

[21] L. McInnes, J. Healy, and J. Melville, Umap: Uniform manifold approximation and projection for dimension reduction, 2018.

[22] L. Micallef, P. Dragicevic, and J. Fekete, Assessing the Effect of Visualizations on Bayesian Reasoning through Crowdsourcing, IEEE Transactions on Visualization and Computer Graphics, vol. 18, no. 12, pp. 2536–2545, 2012.

[23] L. Micallef and P. Rodgers, eulerAPE: Drawing Area-Proportional 3-Venn Diagrams Using Ellipses, PLOS ONE, vol. 9, no. 7, pp. 1–18, 2014. [Online]. Available: https://­doi.org/­10.1371/­journal.pone.0101717

[24] K. Misue, Drawing Bipartite Graphs as Anchored Maps, in Proceedings of the 2006 Asia-Pacific Symposium on Information Visualisation - Volume 60, 2006, pp. 169–177.

[25] N. Moroney, M. Fairchild, R. Hunt, l. Changjun, M. Luo, and T. Newman, The ciecam02 color appearance model, vol. 10, 2002, pp. 23–27.

[26] A. Ovechkin, S.-M. Lee, and K.-S. Kim, Thermovisual evaluation of acupuncture points, Acupuncture & electro-therapeutics research, vol. 26, no. 1-2, pp. 11–23, 2001.

[27] P. K. Robertson and J. F. O'Callaghan, The generation of color sequences for univariate and bivariate mapping, IEEE Computer Graphics and Applications, vol. 6, no. 2, pp. 24–32, 1986.

[28] P. Rodgers, J. Flower, G. Stapleton, and J. Howse, Drawing Area-Proportional Venn-3 Diagrams with Convex Polygons, in Diagrammatic Representation and Inference, A. K. Goel, M. Jamnik, and N. H. Narayanan, Eds. Springer Berlin Heidelberg, 2010, pp. 54–68.

[29] R. Sadana, T. Major, A. Dove, and J. Stasko, OnSet: A Visualization Technique for Large-scale Binary Set Data, IEEE Transactions on Visualization and Computer Graphics, vol. 20, no. 12, pp. 1993–2002, 2014.

[30] School of Chinese Medicine, Hong Kong Baptist University, Chinese medicine formulae images database, 2017, accessed July 19, 2022. https://library.hkbu.edu.hk/­electronic/­libdbs/­cmfid/­index.html.

[31] H.-J. Schulz, S. Hadlak, and H. Schumann, The design space of implicit hierarchy visualization: A survey, IEEE Transactions on Visualization and Computer Graphics, vol. 17, no. 4, pp. 393–411, 2011.

[32] B. Shneiderman, Tree visualization with tree-maps: 2-d space-filling approach, ACM Trans. Graph., vol. 11, no. 1, pp. 92–99, 1992. [Online]. Available: https://­doi.org/­10.1145/­102377.115768

[33] P. Simonetto, D. Auber, and D. Archambault, Fully Automatic Visualisation of Overlapping Sets, Computer Graphics Forum, vol. 28, no. 3, pp. 967–974, 2009. [Online]. Available: https://­onlinelibrary.wiley.com/doi/­abs/­10.1111/­j.1467-8659.2009.01452.x

[34] G. Stapleton, P. Rodgers, J. Howse, and L. Zhang, Inductively Generating Euler Diagrams, IEEE Transactions on Visualization and Computer Graphics, vol. 17, no. 1, pp. 88–100, 2011.

[35] G. Stapleton, J. Flower, P. Rodgers, and J. Howse, Automatically drawing Euler diagrams with circles, Journal of Visual Languages Computing, vol. 23, no. 3, pp. 163–193, 2012. [Online]. Available: https://www.sciencedirect.com/­science/­article/­pii/­S1045926X12000134

[36] J. Stasko, C. Görg, and Z. Liu, Jigsaw: Supporting Investigative Analysis through Interactive Visualization, Information Visualization, vol. 7, no. 2, pp. 118–132, 2008. [Online]. Available: https://doi.org/10.1057/palgrave.ivs.9500180

[37] A. C. Y. Tang, Review of traditional chinese medicine pulse diagnosis quantification, in Complementary Therapies for the Contemporary Healthcare, M. Saad and R. de Medeiros, Eds. Rijeka: IntechOpen, 2012, ch. 4. [Online]. Available: https://­doi.org/­10.5772/­50442

[38] L. van der Maaten and G. Hinton, Visualizing data using t-sne, Journal of Machine Learning Research, vol. 9, no. 86, pp. 2579–2605, 2008. [Online]. Available: http://­jmlr.org/­papers/­v9/­vandermaaten08a.html

[39] J. Wang, Basic Theory of Traditional Chinese Medicine. Beijing: China Press of Traditional Chinese Medicine, 2016.

[40] M. Wei, Z. Chen, G. Chen, X. Huang, Y. Jin, K. Lao, Z. Li, S. Li, F. Zhong, H. Liang et al., A portable three-channel data collector for chinese medicine pulses, Sensors and Actuators A: Physical, vol. 323, p. 112669, 2021.

[41] L. Wilkinson, Exact and Approximate Area-Proportional Circular Venn and Euler Diagrams, IEEE Transactions on Visualization and Computer Graphics, vol. 18, no. 2, pp. 321–331, 2012.

[42] Y. Wu, F. Zhang, K. Yang, S. Fang, D. Bu, H. Li, L. Sun, H. Hu, K. Gao, W. Wang et al., Symmap: an integrative database of traditional chinese medicine enhanced by symptom mapping, Nucleic acids research, vol. 47, no. D1, pp. D1110–D1117, 2019.

[43] J. Xie, C. Jing, Z. Zhang, J. Xu, Y. Duan, and D. Xu, Digital tongue image analyses for health assessment, Medical Review, vol. 1, no. 2, pp. 172–198, 2021. [Online]. Available: https://­doi.org/­10.1515/­mr-2021-0018

[44] Z. Yan, G. Bo, and C. Meng, Design and implementation of the analysis system of TCM prescription, China Journal of Traditional Chinese Medicine and Pharmacy, vol. 29, no. 5, p. 4, 2014.

[45] L. Zhou and C. D. Hansen, A survey of colormaps in visualization, IEEE Transactions on Visualization and Computer Graphics, vol. 22, no. 8, pp. 2051–2069, 2016.

1. https://colorbrewer2.org [↑](#footnote-ref-1)