

Spreadsheet Interface Which Simplifies the Visualization of the Network Structures

Satoshi Koike, Kazuo Misue, and Jiro Tanaka

Department of Computer Science, University of Tsukuba
koike@iplab.cs.tsukuba.ac.jp, {misue, jiro}@cs.tsukuba.ac.jp

Abstract. We developed a system that extracts and visualizes the network structures from relational data on spreadsheets. Visualization through network diagrams obtained from relational data is indispensable to understanding their structure. However, time-consuming tasks are involved in visualizing these networks. The purpose of this research was to support the extraction and visualization of network structures from relational data. We developed an interface that coordinates spreadsheets with visualization tools. It simplifies the visualization of networks from relational data and enables their structures to be easily understood.

Keywords: Network Visualization, Spreadsheet, Graph Drawing, Wizard.

1 Introduction

There are numerous relational data throughout the world. Examples include data on personal connections and purchasing history. These relational data are stored in databases and are used in various scenarios. Retail stores, e.g., convenience stores, better understand customer attitudes from their purchasing history.

Relational data often compose network structures. We can find useful information by understanding the structures. However, it is difficult to understand the structure of networks from relational data when the network is represented as lists of characters and values. The visual representation of networks is essential to understanding the structure of networks from relational data.

Various methods of visualization and many visualization tools that enable network diagrams to be drawn have been developed. However, as databases and visualization tools are not sufficiently coordinated, we have to do time-consuming tasks such as extracting and translating data to visualize networks.

We took the approach of coordinating spreadsheets and visualization tools to solve this problem. Spreadsheets are tools for manipulating tabulated data, and they are also used to edit relational data. If it is easily to extract and visualize networks from relational data on spreadsheets, we can easily analyze these with network diagrams.

2 Networks Extracted from Database

There is a connection between members written on the same record in a database. Various networks are composed of these relations. We focused on two kinds of networks in this research, “networks based on relations” and “networks based on similarities”.

2.1 Database

A database, D , is formulated as $D = \{r_1, r_2, \dots, r_m\}$ in this paper. r_i is the i th record. The p th field in record r is formulated as $f_p(r)$.

2.2 Network Based on Relations

There are connections between fields’ members in the same record. A network is composed when the members of two fields in the database are identified as nodes. This network is based on the relation between two fields.

The network based on the relation between field f_p and field f_q can be formulated as weighted graph G_r (shown in (3)). Set V of nodes, set E of edges with weights, and function w_r that gives the weight of an edge comprise G_r (shown in (4), (5), and (6)).

$$G_r = (V, E, w_r) \tag{3}$$

$$V = F_p \cup F_q, F_p = \{f_p(r) | r \in D\} \tag{4}$$

$$E = \{(a, b) | f_p(r) = a \wedge f_q(r) = b \text{ for } \exists r \in D\} \tag{5}$$

$$w_r : E \rightarrow R \tag{6}$$

Function w_r can be selected from various methods of calculation. There are some examples in the following.

Set $R_{p,q}(a, b)$ of records where a and b are placed in the same record, r , are shown in (7). The element count of $R_{p,q}(a, b)$ gives an indication of the strength of the relation between a and b . Considering this value as the weight of an edge, w_r is as shown in (8).

$$R_{p,q}(a, b) = \{r \in D | f_p(r) = a \wedge f_q(r) = b\} \tag{7}$$

$$w_r((a, b)) = |R_{p,q}(a, b)| \tag{8}$$

The value of numerical field $f_w(r)$ can be assumed to be the value of the strength of the relation between $f_p(r)$ and $f_q(r)$. The method of calculation to add the value of $f_w(r)$ is selected from various choices (e.g., sum or average). Considering this value as the weight of the edge, w_r is shown in (9) and (10).

$$w_r((a, b)) = \sum_{r \in R_{p,q}(a,b)} f_w(r) \tag{9}$$

$$w_r((a, b)) = \frac{\sum_{r \in R_{p,q}(a,b)} f_w(r)}{|R_{p,q}(a, b)|} \tag{10}$$

It is necessary to determine fields f_p, f_q as vertices and field f_w as the weight of an edge to extract the network based on the relation. In addition, it is necessary to determine function w_r .

2.3 Network Based on Similarities

The members in a database are connected with the other field’s members written on the same record. The similarity of the members can be calculated by using its relations to the other field’s members.

When focusing the relation between field f_p ’s members and field f_q ’s members, the similarities of f_p ’s members are calculated. Here, f_q is called a reference field for the similarities. Two or more fields can be used as reference fields at the same time.

A network is based on the similarity is composed when we consider that all the different members of a field as nodes. The network based on the similarities of all the different members of f_p can be formulated as a weighted complete graph, G_s (shown in (11)). Set V of nodes, set E of edges with weights, and function w_r that gives the weight i.e., similarity of an edge comprise G_s (shown in (12), (13), and (14)).

$$G_s = (V, E, w_s) \tag{11}$$

$$V = F_p, F_p = \{f_p(r) | r \in D\} \tag{12}$$

$$E = F_p \times F_p \tag{13}$$

$$w_s : E \rightarrow R \tag{14}$$

Function w_s can also be selected from various methods of calculation. The following has some examples (shown in (15), (16), and (17)).

$$w_s((a_i, a_j)) = \frac{1}{\sum_{b \in F_q} |w_r((a_i, b)) - w_r((a_j, b))|} \tag{15}$$

$$w_s((a_i, a_j)) = \frac{1}{\sqrt{\sum_{b \in F_q} (w_r((a_i, b)) - w_r((a_j, b)))^2}} \tag{16}$$

$$w_s((a_i, a_j)) = \frac{\sum_{b \in F_q} (w_r((a_i, b)) - v(a_i, b))(w_r((a_j, b)) - v(a_j, b))}{\sqrt{\sum_{b \in F_q} (w_r((a_i, b)) - v(a_i, b))^2} \sqrt{\sum_{b \in F_q} (w_r((a_j, b)) - v(a_j, b))^2}}, \quad (17)$$

$$v(a, b) = \frac{\sum_{b \in F_q} w_r((a, b))}{|F_q|}$$

It is necessary to determine fields f_p as vertices, fields f_q as the target for calculating the similarities, and field f_w as the value of the strength of the relation to extract the network based on similarities. It is also necessary to determine function w_r . In addition, it is necessary to determine function w_s .

3 Extracting and Visualizing Networks

The steps involved in visualizing networks are as follows.

Step 1. Select the type of network

It is necessary to determine what kind of network is to be extracted. Two kinds of networks (“networks based on relation G_r ” and “networks based on similarity G_s ”) were presented in Section 2.

Step 2. Select the fields

It is necessary to determine the extraction fields or method of calculation to extract the networks selected in Step 1. As described in Section 2, it is necessary to determine f_p, f_q, f_w, w_r to extract G_r and it is necessary to determine f_p, f_q, f_w, w_r, w_s to extract G_s .

Step 3. Select the style of visualization

It is necessary to determine the style of visualization of the network. As there are various methods of visualization, it is necessary to choose one according to the kind of the network and the purpose of the visualization. (AnchoredMap Layout [1] and Radial Layout [2] can be chosen for the system proposed in this paper.)

Step 4. Extract the network

The selected network is extracted after Step 3 and the network data are translated to data that correspond to the visualization tool. (The system proposed in this paper outputs GraphML [3] and CSV from the data on the spreadsheet. GraphML is extended XML to represent the graph. GraphML is used in many tools as an input data format. CSV is used in several Web applications as an input data format.)

Step 5. Visualize with tool

The network is visualized by inputting data to the tool.

There are numerous choices in extracting the network; therefore, it is necessary to select from these. Moreover, the extracted network needs to be translated to data that

correspond to the visualization tool. These tasks are time consuming and difficult for people who do not have skills to translate data.

4 Wizard for Extracting/Visualizing Network

A wizard is often used to support complex tasks. It is an interface that guides users through complex tasks, using step-by-step dialogs. The wizard is used for setting up applications.

It is further used for visualizing information. An interface called “Graph Wizard” is implemented on a spreadsheet to draft coordinate graphs such as bar and line graphs. Although it is necessary to select from various choices such as data fields and types of graphs, it is possible to easily draw these using the Graph Wizard.

Visualizing the network is similar to drawing these coordinate graphs. A spreadsheet is also used to present and edit relational data. It is also necessary to select from numerous choices such as data fields and methods of visualizing the network. The wizard can also be applied to visualizing the network for this reason.

We propose the wizard as an interface to support the extraction and visualization of networks from relational data. The user can extract and visualize networks from relational data with this wizard by answering its questions.

4.1 System Architecture

Figure 1 outlines the system architecture.

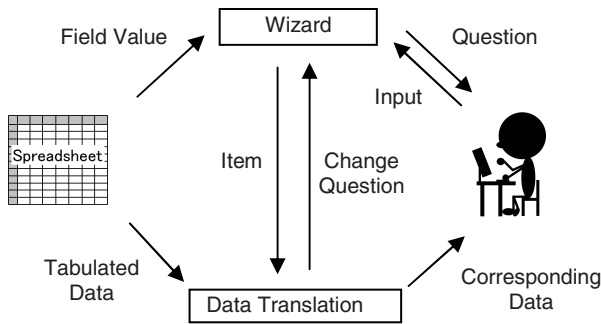


Fig. 1. System Architecture

The wizard component promotes input by users by asking them questions. If it needs the field value for the question, it obtains the value from the sheets. The data translation component changes the questions asked by the wizard. Based on the inputs by the wizard, it translates the tabulated data into the data that correspond to the visualization tool.

4.2 Implementation

The items that are necessary to extract and visualize networks can be classified into three steps.

- Step 1: Type of network
- Step 2: Field
- Step 3: Style of visualization

The wizard presents a question to the user to set these items through dialog-input forms. He or she extracts and visualizes the network by inputting items into the dialog forms.

A dialog input form is divided into three parts (Figs. 2, 3, 4, and 5) according to the item that needs to be extracted to visualize the network.

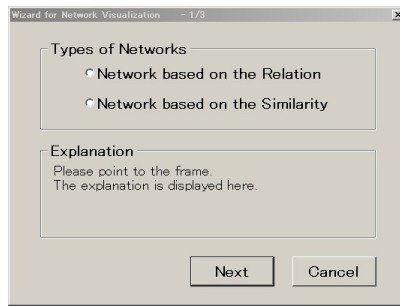


Fig. 2. Dialog-input form for Step 1

Figure 2 shows a dialog form for two types of networks. A text explanation is displayed in a frame in the middle of the form. The user selects a network (“relation” or “similarity”) using the buttons on the form. The system changes to the next dialog form depending on the input selected.

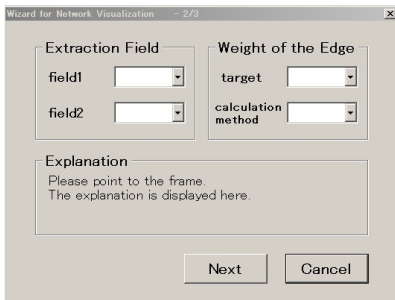


Fig. 3. Dialog-input form for Step 2 (relation)

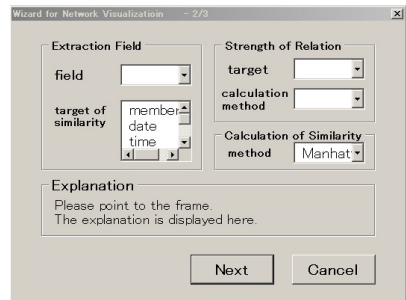


Fig. 4. Dialog-input form for Step 2 (similarity)

Figure 3 shows the dialog form to select a field to extract the network based on relations. The user selects a field extracted as nodes from the left frame. The value of

the box corresponds to the field on the sheet. He or she selects the field and method of calculating the edge weight from the right frame. The value of the “target” box corresponds to the numerical field on the sheet.

Figure 4 shows the dialog form to select a field to extract the network based on similarities. The user selects a field extracted as nodes and the target field of similarity from the left frame. The user selects the field and method of calculating the relational strength from the upper right frame. The user selects the method of calculating similarity from the lower right frame.

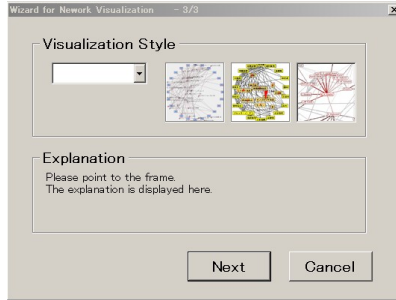


Fig. 5. Dialog-input form for Step 3

Figure 5 shows the dialog form for the style of visualization. The user selects a method from the frame.

The corresponding data are output after he or she has selected from the form.

5 Discussion

We compared the tasks of manually extracting and visualizing the network based on relations and similarities with those done by the wizard.

Table 1. Comparison of tasks

| manual procedure | wizard |
|---|--|
| 0. Prepare the translation program 1. Extract the network by manual 2. Output the intermediary file 3. Translate the intermediary file to the corresponding data 4. Visualize by the tool | 1. Extract the network and output the corresponding data by answering the question from the wizard 2. Visualize by the tool |

It is necessary to output an intermediary file to output corresponding data in the manual procedure. In addition, users have to prepare the translation program that translates the intermediary file to corresponding data. With the wizard, on the other

hand, they can output the corresponding data simply by answering questions asked by the wizard, which enables networks to be visualized.

6 Related Works

There has been some research that has taken the approach of coordinating tabulated data and network diagrams.

Henry and Felete developed a tool called “MatrixExplorer” to assist in the understanding of network structures [4]. This tool displays tabulated data and network diagrams simultaneously and visually represents their correspondences.

Mackinlay developed A Presentation Tool (APT) that designs effective representations of the relational data [5]. This tool designs representations including bar and scatter charts according to the features of the relational data.

However, we took a new approach that uses a wizard to coordinate tabulated data and network diagrams.

7 Conclusions

We proposed a wizard as an interface to extract and visualize the structures of networks. We implemented it to extract and visualize networks from the relational data available on spreadsheets. It is much easier to visualize networks with this wizard and we intend to apply it to surveying other networks.

References

1. Misue, K.: Drawing Bipartite Graphs as Anchored Maps. In: Proceedings of Asia-Pacific Symposium on Information Visualization, pp. 169–177 (2006)
2. Heer, J., Card, S.K., Landay, J.A.: Prefuse: a toolkit for Interactive Information Visualization. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 421–430 (2005)
3. Brandes, U., Pich, C.: GraphML Transformation. In: Pach, J. (ed.) GD 2004. LNCS, vol. 3383, pp. 89–99. Springer, Heidelberg (2005)
4. Henry, N., Felete, J.D.: MatrixExplorer: a Dual-Representation System to Explorer Social Networks. *IEEE Transaction on Visualization and Computer Graphics* 12(5), 677–684 (2006)
5. Mackinlay, J.: Automating the Design of Graphical Presentations of Relational Information. *ACM Transactions on Graphics* 5(2), 110–141 (1986)