EUROGRAPHICS - IEEE TCVG Symposium on Visualization (2005)
K. W. Brodlie, D. J. Duke, K. I. Joy(Editors)

VisBiz: A Business Process Visualization Case Study

Ming C. Hao, Daniel A. Keim*, Umeshwar Dayal, Joern Schneidewind*
(ming.hao, umeshwar.dayal)@hp.com, (keim, schneide)@informatik.uni-konstanz.de
Hewlett Packard Research Laboratories, Palo Alto, CA

Abstract

Business process management involves many parameters and relationships and is modeled as complex business process workflows. A common way to analyze the process data is by using flowcharts. Visual analysis of a large-scale chart, however, is too complex. In this case study, we employ a novel visualization technique, called VisBiz. VisBiz reduces data complexity by automatically analyzing operational data and abstracting the most critical parameters that influence business process. The basic idea is to select the most relevant parameters and layout them on a "triple-attributes" circular graph based on their relationships and user domain knowledge. VizBiz transforms the attributes to nodes and the process flows to lines. VisBiz derives a new process flow matrix to link the process of multiple circular graphs as the analyst introduces more parameters for further analysis. The results of the real-world credit card fraud study show the significant advantages of this technique in finding fraud distribution patterns and root causes of frauds.

1. Motivation

Business process data is inherently large and complex, most often too complex to be directly visualized. Usually the business processes consist of many steps and alternatives, and every data instance may take a different path through the process.

There exist a number of business-oriented visualization techniques [A to P], such as the SeeSoft line representation technique [Eic92] used for visualizing program changes, ILOG Jviews [Ilo04] used for analyzing workflow processes, and E_BizInsights [EB00] used for web path analysis, and parallel coordinates [ID90] used for visualizing correlations. All these methods aim at reducing the time to turn business data into information, which in turn reduces the business decision-making time.

2. VisBiz Approach

In this case study, we apply a new interactive visualization technique, VisBiz, to reduce data complexity by abstracting the most critical parameters (attributes) that influence business processes. The basic idea of this technique is to visually mine relationships between important operation parameters and to map the parameters into visualizations.

Our study contains the following visual analysis steps:

- Determine the critical parameters from business *correlation matrix* and user domain knowledge.
- Layout on a triple-attribute symmetric circular graph to represent a set of critical attributes (nodes).
- Drill down to a two-attribute visualization to observe clearly a subset of business process (transactions).
- Use *a process flow matrix* based on transaction process paths to automatically link multiple circular graphs together to show more than three parameters.
- Use animation to observe business process flows over time.

VizBiz has been designed to work for large volumes of complex business process data. The automatic analysis determines the important relationships, and the visualization graphs show the relationships. VizBiz provides a new layout that gives more weight to important data values. We start the visualization with the three most critical parameters on one circular graph. Then we generate additional graphs to analyze more than three parameters, as needed.

2.1. Business Process Visualization

VisBiz transforms business attributes to nodes, with the lines between nodes representing a process path of a transaction record on a triple-attribute circular graph. A circular graph is used because there is more space on the edge of a circle and is more intuitive. The three attributes are used for partitioning the left side, center, and right side of a circular graph.

VisBiz uses colored lines for specific business metrics. The width of the line represents the number of lines with the same process flow.

*Presently with University of Constance, Germany.

The color of a line represents the average value of the metrics. The order of the nodes is arranged from the bottom to the top of the edge of a circular graph based on the value of the metrics.

2.2. Attribute Circular Graphs

The triple-attribute graph can be drilled down to a two-attribute graph when a user selects one of the nodes from an attribute axis (e.g., Region 6, from attribute 1 in Figure 1A), which shows a subset of business process instances. The graphs (Figure 1B for Region 6 and 1C for Region 2), contain two halves of a circular graph. The left and right nodes represent the other two attributes of the business processes (e.g., Fraud Amount and Fraud Count).

2.3. Multiple Attribute Circular Graphs

Using a *process flow matrix*, VisBiz allows analysts to focus on a node (attribute) and show its related nodes and process paths across multiple circular graphs. The unrelated nodes and paths are removed or faded out. The analyst can easily analyze the process flow by tracing the lines starting from the original node.

VisBiz differs from existing visualization techniques, statistical reporting, and charting, and adds the following benefits:

- Simplifies the visualization of complex business processes by partitioning the process into multiple views and provides identification of key patterns and relationships.
- Visualizes the business process executions and their changes over time.
- Instantly captures and identifies business process flow, exceptions, and bottlenecks in executions, thereby enabling analysts to proactively improve their business processes.

3. Case Study- A Credit Card Fraud

We have conducted experiments with VisBiz for fraud analysis at HP Research Laboratories. For fraud detection, fraud analysts would like to discover new patterns and relationships in the transaction data. Examples of analysis that they typically need include the following:

1. What is the fraud geographical distribution?
2. What is the fraud growing rate in the last three years?

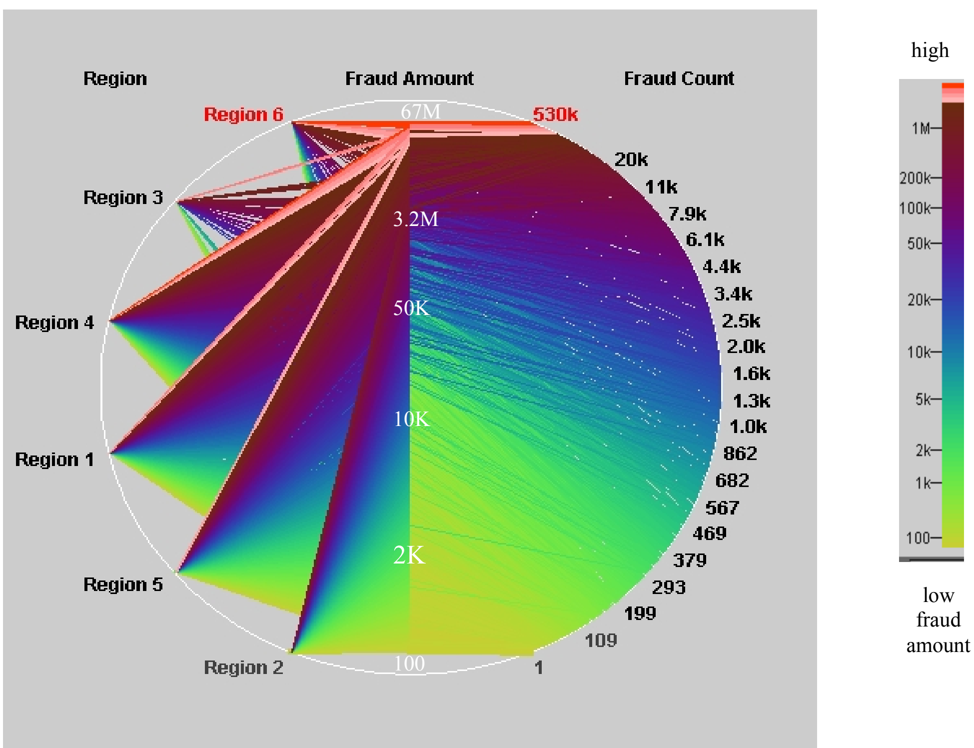


Figure 1A: Credit Card Fraud Geographical Distribution
(Three Attributes: Region, Fraud Amount, and Fraud Count)

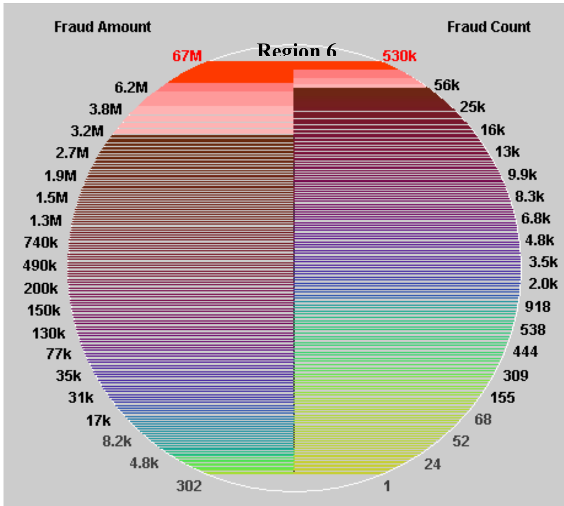


Figure 1B: Region 6 Fraud Credit Card Distribution

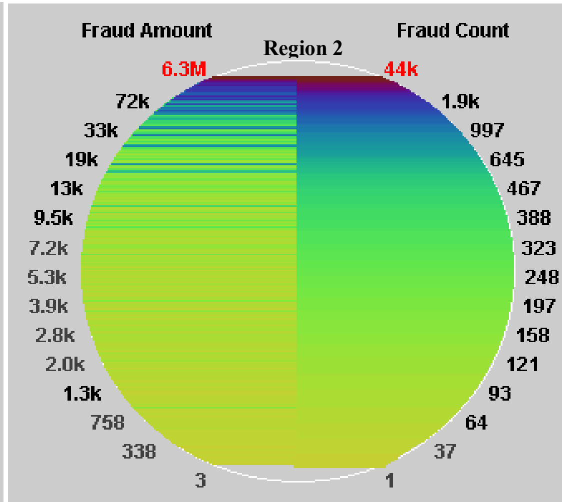


Figure 1C: Region 2 Fraud Credit Card Distribution

3. Which are the three countries with the highest fraud amount over time?
4. Which types of credit cards have the most fraud? Issued by which country?
5. How do we find a fraud? Where does the fraud come from? (from an account or a stolen credit card?)

3.1. Fraud Distribution Pattern Analysis

To address the first question in Figure 1A, we select the three most related attributes: Region, Fraud Amount, and Fraud Count. The attribute 1 nodes comprise the Regions 1-6 and reside on the left side of the circular graph. The attribute 2 nodes are the Fraud Amount and reside in the middle axis of the circular graph. The attribute 3 nodes are the Fraud Count and reside on the right side of the circular graph.

The linked lines represent the connections between the nodes. The color represents the fraud amount. For fast identification, nodes are ordered by Fraud Amount from bottom to top (highest) on the circular graph.

The color represents the average value of the fraud amounts. The width of the lines represents the number of transactions. In addition, the analysts can quickly discover exceptions: red lines to represent high fraud amount or counts above the top 1%.

Region 6 has the highest fraud amount (more red and burgundy) and resides on the top left of the circular graph. Region 2 has the lowest fraud amount (more yellow and green) and resides on right bottom of the circular graph.

Fraud amount and count (Figure 1A) has a high correlation as shown by the colored lines connected from the fraud amount nodes to the fraud count nodes, such as the high fraud amount nodes (burgundy) connected to the high fraud count nodes (burgundy).

The analyst selects a node to show the relationships with the other two attributes (i.e., Fraud Amount and Fraud Count) as illustrated in Figure 1B and Figure 1C. Figure 1B shows high fraud amount and fraud count (more lines are burgundy) and more exceptions (more red lines). Figure 1C shows low fraud amount and fraud count (more green and yellow).

By comparing the flows in Figures 1B and 1C, Region 6 has less transactions but a higher fraud amount and fraud count (fewer lines and more blue, burgundy, and red) than Region 2.

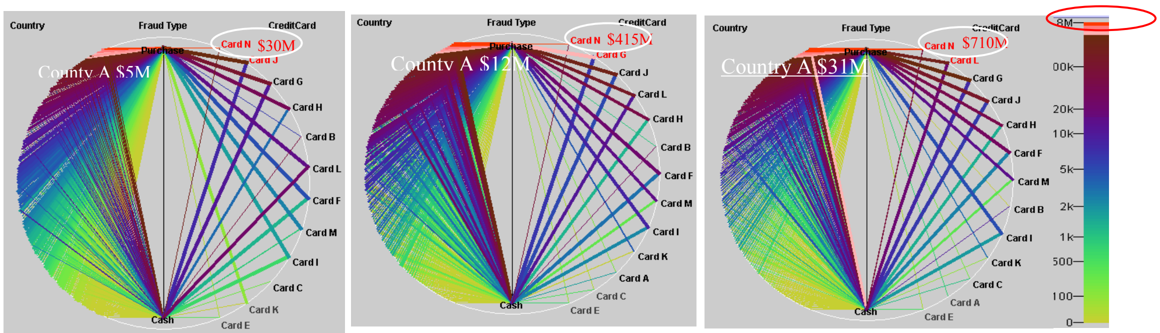


Figure 2A: Year of 2000 **Figure 2B**: Year of 2001 **Figure 2C**: Year of 2002 Country Fraud Amount

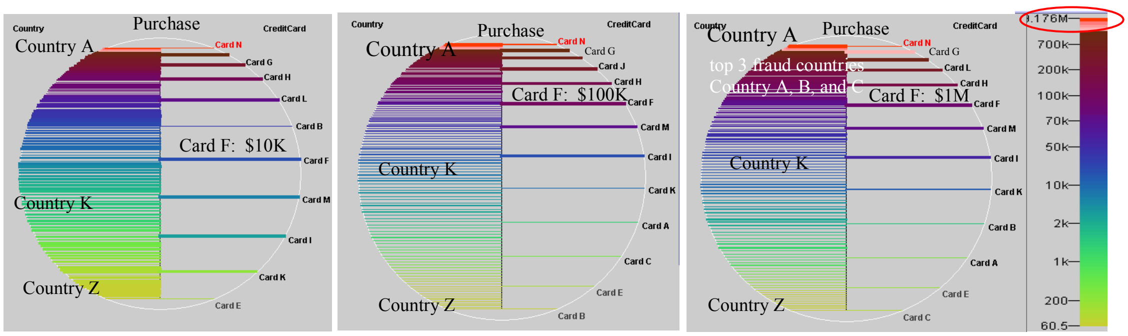


Figure 2D: 2000 Purchase Fraud **Figure 2E**: 2001 Purchase Fraud **Figure 2F**: 2002 Purchase Fraud Credit Card Fraud Amount

3.2. Fraud Growth Rate Analysis

To answer the second and third questions, VisBiz animates the graphs over the historical evolution of the sales transactions. We select three different attributes: Country, Fraud Type and Credit Card. We map them on a circular graph over time as shown in Figures 2A-2F. The colored lines are ordered by fraud amount. The countries with the highest fraud amount are in red at the top of the left side of the circular graphs.

VisBiz shows fraud growth patterns over three years. By comparing the flows displayed in figures 2A, 2B, and 2C, the analyst can quickly determine that the year 2002 has the most fraud transactions and the highest fraud amount (more lines, more blue and burgundy). The fraud rate has grown threefold from 2000 to 2002.

Figures 2D-2F is generated when the analyst selects the Purchase nodes in Figures 2A-2C. The business flows displayed in Figures 2D, 2E, and 2F also show that the fraud amount has grown significantly from 2000 to 2002. For example, credit card F has a total fraud amount increasing from $10K to $1M during the period 2000 to 2002. From Figure 2F, the top three countries (A, B, and C, color red) can be identified to have the highest fraud amount which exceeds $15M. Card N, G, and L (colored red) have the highest fraud amount (over $9M) in 2002 (indicated from the color scales).

3.3. Credit Card Usage Analysis

Figure 3 answers the fourth question on credit card usage from a re-arranged triple-attribute (Fraud Type, Credit Card, and Country) circular graph: Card N has the highest fraud amount (colored red). Card F has the most purchase transactions and is used by

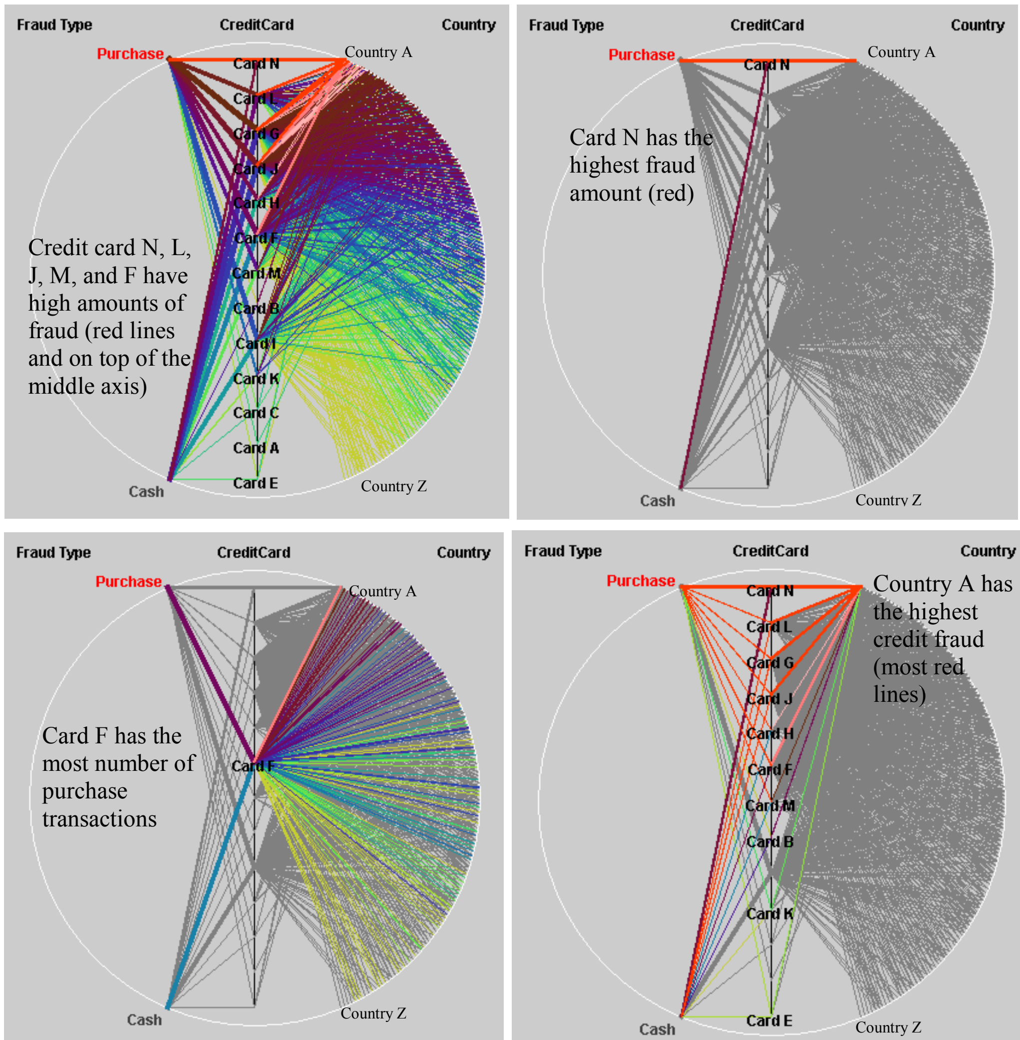


Figure 3: Credit Card Usage Analysis

many countries (more lines). Country A has the most fraud (red lines at the top of the circle. Cards N, L, G, and J have the highest fraud amounts and the most fraud from purchase transactions.

3.4. Fraud Root Cause Detection

To answer the fifth question about where does the fraud come from and the cause of a fraud, we take the following steps:

3.4.1. Find Important Attributes

First, we select three critical parameters (attributes) with the highest correlations from the *Business Correlation Matrix*, namely Fraud amount, Fraud count and Sales Volume, as shown in Figure 4A.

Second, we generate a circular graph as shown in Figure 4C. In Figure 4C, these three parameters are highly correlated as seen by the nearly parallel lines, except for some outliers crossing from low Sales Volumes to high Fraud Count.

Columns	Pearson's correlation coefficient
Fraud Amount / Fraud Count	.97
Sales Volume / Number of Total Tran...	.9
Sales Volume / Fraud Count	.72
Sales Volume / Fraud Amount	.72
Number of Total Tran... / Fraud Count	.7

Figure 4A: Business Correlation Matrix

Attribute1 Fraud Amount	Attribute2 Fraud Count	Attribute3 Sales Volume	Attribute4 Country	Attribute5 Region	Attribute6 Issuer	Attribute7 Purchase Quarter	Attribute8 Credit Card	Attribute 9 Status
47M	530K	48B	A	6	1	2002-3Q	N	ACCOUNT
28M (row2)	430K	1	A	6	1	2002-3Q	G	ACCOUNT
210K (row3)	120K	110M	A	4	5	2002-2Q	F	STOLEN
57K	27K	3.3M	K	4	22	2002-1Q	I	ACCOUNT
…	…	…	…	…	…	…	…	…

Figure 4B: Process Flow Matrix (each attribute is a node on the linked circular graphs)

Then, we construct the 2nd and 3rd circular graphs based on the fraud domain knowledge, as shown in Figure 4D and Figure 4E. Figure 4D shows which country and region the fraud comes from and who is the fraud credit card issuer. Credit cards issued by Issuer 1 from countries in Region 6 have the most fraud (red line on top of the circle). Figure 4E shows Credit Card N, and Status ACCOUNT have the most fraud in 2002-3Q (on top of the circular graph, more red and burgundy lines).

3.4.2. Link Process Flows

VisBiz employs the *Process Flow Matrix,* as shown in Figure 4B, to link the nodes and process flows across three circular graphs (Figure 4C, 4D, and 4E) to show an entire process of a business transaction. The *Process Flow Matrix* is constructed as a row and column hash set, one column for each attribute, one row for each transactions record. A node is drawn for each distinct value of an attribute. A transaction record may contain many nodes. VisBiz stores the transaction row number in each node that are used to link to different nodes and lines. Each node may contain many rows. The analyst can selects a node or a line on a graph to display all related transaction paths. Unrelated nodes and lines will not be shown on the graphs. For example, VisBiz uses the transaction row number to link Region 6 to Fraud Amount $28M, Credit Card G and Status ACCOUNT as shown in Figure 4B (row 2). We can also link Fraud Amount 210K to Region 4, Credit Card F, and Status STOLEN, as shown in Figure 4B (row 3).

3.4.3. Identify the Root Cause

Figures 4F-4H is generated when the analyst clicks on a red crossing line in Figure 4C. Only the lines and nodes related to the red line are shown. These automated process link and auto-fade in/out functions enable the analyst to quickly find the root cause of a fraud as follows.

In Figures 4F, 4G, and 4H, we can see that the fraud occurred in 2002-3Q using Credit Card G. The credit card is issued by Issuer 1 in Region 6, Country A with Fraud Amount 28M, Fraud Count 430K, and Sales Volume 1.

Using the above information, the company is able to place strict control on certain countries, such as country A, and certain credit cards, such as Card N (highest fraud, on top of the circle). After knowing the sources of the frauds, the company will be able to take preventative actions.

4. Conclusion

In this case study, we simplify the visualization of complex business processes by partitioning the business process into many linked attribute circular graphs. This visualization provides analysis of key patterns, process link flows, and root cause identifications. We have applied VisBiz to other applications, such as service contract analysis and sales analysis. These initial studies of real-world applications also show significant advantages of this technique in finding distribution patterns, exceptions, and root causes of problems.

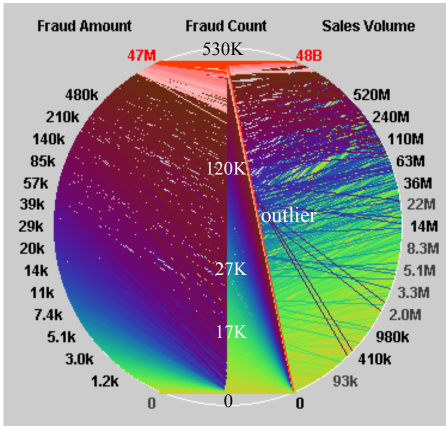


Figure 4C: Discover Exceptions (red crossing lines)

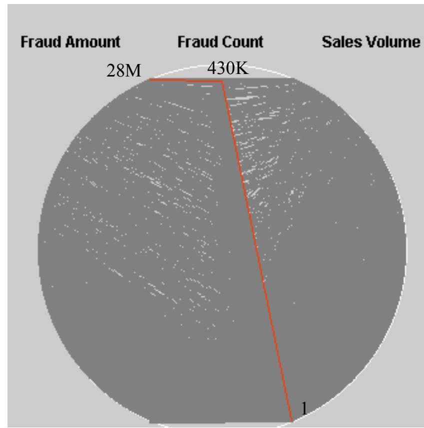
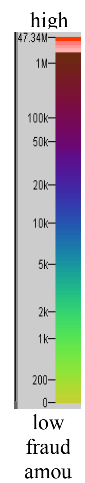


Figure 4E: Show Overall Purchase Quarters, Credit Cards, and Status Process Flow



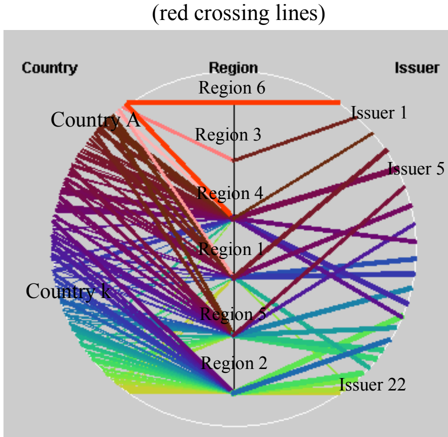


Figure 4D: Show Overall Countries, Regions, and Issuers Process Flow

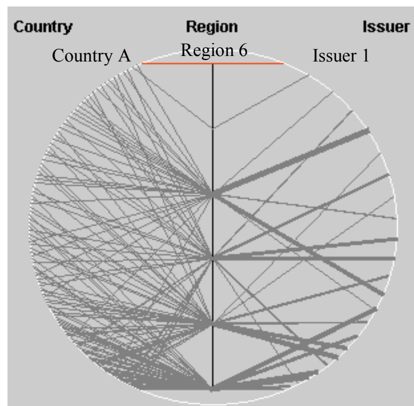


Figure 4F: Show Outlier's Fraud Amount and Fraud Count

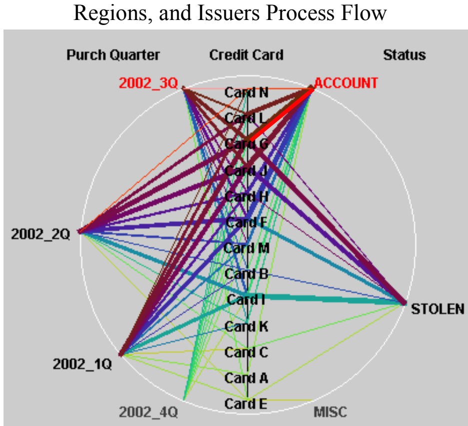


Figure 4H: Show Outlier's Purchase Quarter, Credit Card, and Status

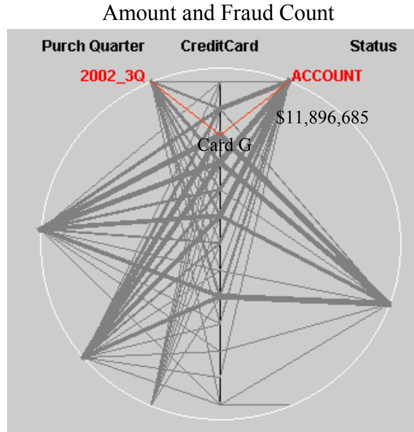


Figure 4G: Show Outlier's Country, Region, and Issuer

Acknowledgements

Many thanks to Kris Halvorsen of HP Laboratories for his encouragement and suggestions and to Manish Bhardwaj from HP consulting and services division for providing comments and data.

References

[AWS92] Ahlberg C., Williamson C., Shneiderman B.: '*Dynamic Queries for Information Exploration: An Implementation and Evaluation*', Proc. ACM CHI Int. Conf. on Human Factors in Computing, Monterey, CA, 1992.

[BEW95] Becker R. A., Eick S. G., Wills G. J.: '*Visualizing Network Data*', IEEE Transactions on Visualizations and Graphics, Vol.1, No.1, 1995.

[BMM*91] Buja A., McDonald J. A., Michalak J., Stuetzle W.: '*Interactive Data Visualization Using Focusing and Linking Linking*', Proc. Visualization '91, San Diego, CA.

[EB00] E_Bizinsights http://www.bizinsights.com

[Eic00] Eick S. G.: '*Visualizing Multi-dimensional Data with ADVISOR/2000',* VisualInsights.

[Eic92] Eick S. G. et al.: '*Seesoft–a tool for visualizing line oriented software statistic*s', IEEE Transactions on Software Engineering, November, 1992.

[GKW02] Grinstein G., Keim D. A., Ward M.: *Information Visualization, Visual Data Mining, and Its Application to Drug Design',* Tutorial, IEEE Visualization 2002, Boston, MA.

[HDH99] Hao, M. C., Dayal U., Hsu M.: '*A Java-based Visual Mining Infrastructure and Application*', Information Visualization, San Francisco, CA. 1999.

[ID90] Inselberg A., Dimsdale B.: `*Parallel Coordinates: A Tool for Visualizing Multi-Dimensional Geometry*', Proc. Visualization ´90, San Francisco, CA.

[Ilo04] ILOG Discovery: Information Visualization. http://www2.ilog.com/preview/Discovery/

[PG88] Pickett R. M., Grinstein G.: '*Iconographic Displays for Visualizing Multidimensional Data*', Proc. IEEE Conf. On Systems, Man and Cybernetics, IEEE Press, Piscataway, NJ, 1988.