

# CPS - Company Positioning System: Visualizing the Economic Environment

Bettina Hoser\*  
Universität Karlsruhe (TH)

Michael Blume†  
Universität Karlsruhe (TH)

Jan Schröder‡  
Universität Karlsruhe (TH)

Markus Franke§  
Universität Karlsruhe (TH)

## ABSTRACT

In this paper we present the Company Positioning System CPS, which enables the user to navigate interactively through a given economic environment and analyze companies, trends and patterns within that environment. We describe how the data is analyzed, the strengths and weaknesses of CPS and some interesting insights found in the data provided.

**CR Categories:** K.6.1 [Management of Computing and Information Systems]: Project and People Management—Life Cycle; K.8.1 [Personal Computing]: Application Packages—Graphics

**Keywords:** visualization, cluster, eigensystem analysis

## 1 INTRODUCTION

The data provided for the 2005 InfoVis contest consists of 87,659 technology companies. Attributes given were for example: sales, year founded, number of employees etc. In addition it was possible to use the NAICS code as a product group hierarchy. The task was to visualize trends, patterns and correlations found within this data set.

Pattern analysis within data sets can be achieved with different tools. As components for CPS we selected a generalized eigen-system analysis as presented in [2] and a cluster algorithm based on random walks as described in [1]. These components together with traditional query techniques were used to achieve an interactive analysis tool. For visualization of the different areas of interest we used geographical information, color coding and relational highlighting.

## 2 STRENGTHS

### 2.1 Flexibility of data source

CPS is based on an SQL data base of relational data. This allows a very flexible handling of different data settings. While we included for this paper only the attributes provided for the contest, we could easily broaden the scope of CPS when additional data becomes available.

### 2.2 Intuitive Color Coding

Results are color coded to facilitate intuitive perception. Different generic schemes have been implemented for color coding, e.g. sales volume, number of employees, number of products, thus enabling the user to navigate a multidimensional attribute space. As a special feature the sales per product can be used for color coding. Since this data is not explicitly available in the data, we use the assumption of

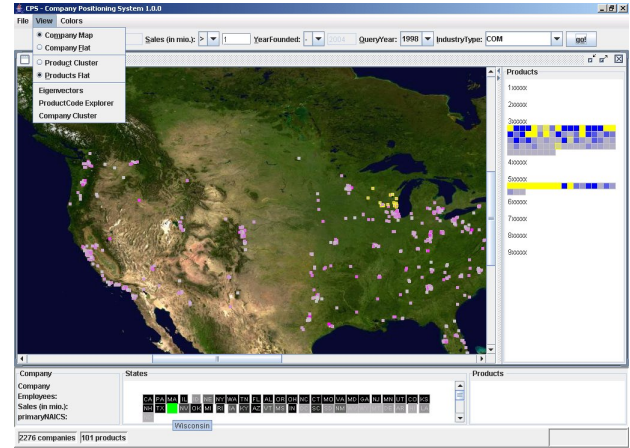


Figure 1: Default screen of CPS.

equally distributed sales volume within a company over its products as a first approximation. The color coding mechanism can easily be adapted to new attribute when they become available.

### 2.3 Company - Products - Geography Relationship

To acquire an overview of the data the default view given in fig.(1) can be used. In this main window the query on the data set can be specified. As a result the geographical location of the queried companies and the products manufactured by those companies is visible. The view can be toggled to a flat array view. In this view the geographical location can still be found by clicking on a state at the bottom of the screen.

These views make it possible to see three kinds of correlations. The user can identify a company and subsequently see which products it sells and where it is located. Also a product can be highlighted and the companies that produce this product and their location is given. As a third possibility a state could be singled out and then the companies within that state and their products are highlighted.

### 2.4 Product or Company Clusters

One of the interesting questions concerning the data set is which products were sold together and which companies produce the same portfolios. The result of these calculations are given in the Product Cluster view of CPS (fig.2). Here the user can select a level of similarity and as a result the cluster centers and the cluster members are color coded within the NAICS code hierarchy.

A result that can be obtained in this view is for example to find products that belong to no cluster at all. This can be achieved by choosing a cluster level close to zero. Products that still show no coloring do not even belong to the overall picture. These products are outliers, since they seem to be produced without any correlation to other products. In the same way companies can be identified which belong to no cluster and thus could be seen as outliers.

\*e-mail: bettina.hoser@em.uni-karlsruhe.de

†e-mail: michael.blume@iw.uni-karlsruhe.de

‡e-mail: jan.schroeder@em.uni-karlsruhe.de

§e-mail: markus.franke@em.uni-karlsruhe.de

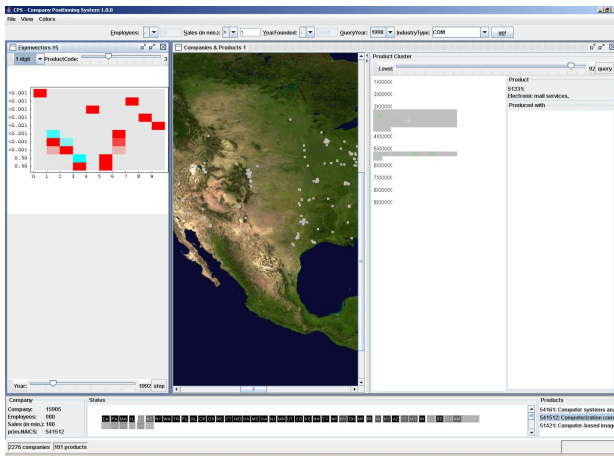


Figure 2: Eigenvector and Product Cluster screen of CPS.

## 2.5 Shifts in Product Portfolios over time

In order to detect any shifts in the relevance of products over time CPS offers the Eigenvector view (fig.2). In this view the eigenvectors of the product-by-product co-manufactured matrix are presented such that the eigenvector corresponding to the largest eigenvalue (absolute value) is given as the bottom row. The subsequent rows are those eigenvectors belonging to lower absolute eigenvalues. The relevance within the eigensystem is encoded on the y-axis as the covered data variance, which can be calculated from the respective eigenvalue. The most relevant product within a row is color coded in red or cyan, depending on the sign of the eigenvector component. With this view we can detect a shift in the areas of operations of companies over time. This is achieved by calculating the eigensystem of each year. When these eigensystems are then animated, any change in the configuration of the eigenvector components becomes visible, thus revealing a shift.

## 3 WEAKNESSES

As CPS not only incorporates intuitive cross reference techniques but also state of the art cluster and pattern recognition techniques, it is no surprise that the user has to allow for a learning curve. This also holds true for the color coding, the highlighting, the use of menus and the appropriateness of different views for different tasks.

One task that CPS is not build to perform is missing data analysis. The application assumes that the data is complete and valid and allows interpretation.

## 4 EXPLORATION RESULTS

### 4.1 Correlations

We have found interesting correlations in the data. They can be found when using the map or flat view. We can show that there seem to be states more relevant to certain industries than others. For example California has many computer industry companies, while Texas has many oil companies. Within the computer industry field there are two states that feature only one company, but with a very different sales level. The one company in Montana has a high sales level, while the one company in West Virginia has a low sales level.

Another finding might point to tax differences between states. If we highlight for example in the computer industry field the state of California, we see that all companies registered there are also located there. This is not true for Massachusetts. Many companies

registered there are not located there. This might point to the fact that MA has different tax regulations then CA.

## 4.2 Product clusters and unusual products

To see which products were manufactured together and which products are the cluster centers, we change to Product Cluster view. The level at which products are clustered can be adjusted. A level close to 0 helps to check for outliers, while levels close to 100 make it possible to identify products that are within densely knit clusters of co-manufactured products. As findings we can state that at a level between 93 and 95 we find product groups 541 (professional, scientific and technical services) and 334 (computer and electronic product manufacturing) as cluster centers and as strongly connected within clusters. Product codes which fell out of the clustering already at cluster level 1, thus indicating an unusual behavior, can be described as outliers. Since in the overall picture the data points towards the software and computer industries and those connected to it, it is not surprising to find products like 62210 (Psychiatric and Substance Abuse Hospitals) or 327112 (Vitreous China, Fine Earthenware, and Other Pottery Product) as outliers.

## 4.3 Trends: Shifts in areas of operation

As an exploratory result enhancing the results of the product clustering, we have detected with the eigensystem analysis the shift of relevance from manufacturing (NAICS code 3) to services and software (NAICS code 5) around the year 2000. In order to get a more detailed view than the 1st-level of NAICS code, the user can switch to the two digit level. Here we can see, that the product code 33 is connected to the codes 32, 51, 54 and 55. In 1999 the focus switches to code 54 which represents a shift from manufacturing to professional, scientific and technical services. If we even go one step further to the three digit NAICS code view, it becomes clear that the shift from 334 towards services (541) took place, which could represent the internet and software boom.

## 5 CONCLUSION

By using the different views and methods of CPS we have discovered some patterns that would have been very difficult if not impossible to detect with standard tools. The shift of the area of operation in the year 2000 from product group 3 to 5 as well as the clustering of products and companies in the Product Cluster view are good examples of this. The possibility to construct queries on the data and analyze the query or to look at the complete data set seems to have a high potential for further investigation. We believe that CPS can also be used to either dig deeper once the data base is enriched by information on sales by product etc., or to even shift to applications in the field of market analysis or any other multidimensional problem that can be described with a data base of relational data.

## REFERENCES

- [1] Markus Franke and Anke Thede. Clustering of Large Document Sets with Restricted Random Walks on Usage Histories. In Claus Weihs and Wolfgang Gaul, editors, *Classification – the Ubiquitous Challenge: Proceedings of the 28th Annual Conference of the Gesellschaft für Klassifikation e.V.*, Studies In Classification, Data Analysis, and Knowledge Organization, pages 402–409, Berlin Heidelberg New York, 2005. GfKI Deutsche Gesellschaft für Klassifikation, Springer.
- [2] Bettina Hoser. *Analysis of Asymmetric Communication Patterns in Computer Mediated Communication Environments*. Universitätsverlag Karlsruhe, 2005.