

# FP-Viz: Visual Frequent Pattern Mining

Daniel A. Keim\*  
University of Konstanz  
Germany

Jörn Schneidewind†  
University of Konstanz  
Germany

Mike Sips‡  
University of Konstanz  
Germany

## Abstract

Frequent pattern mining plays an essential role in many data analysis tasks including association-, correlation-, and causality analysis and has broad applications. Examples are market basket analysis and web click stream analysis. Although a number of efficient methods for mining frequent patterns were proposed in the past, there exist only a small number of visual exploration tools for discovering frequent patterns. In this paper we present a novel visualization technique for exploring frequent itemsets interactively, based on a radial visual layout approach.

**Keywords:** Visual Data Exploration, Frequent Pattern Mining, Visualization technique

## 1 Introduction

Frequent pattern mining plays an essential role in many data analysis tasks including association-, correlation-, and causality analysis. One of the most common application examples is market basket analysis, where analysts are interested in determining what products customers purchase together which in turn might be used to optimize the shelf layout of a supermarket. Another application example is web click stream analysis, where a web administrator might be interested in detecting killer pages. Numerous techniques for mining frequent pattern have been proposed in the past, like the Apriori algorithm or the FP-growth[1] approach. But there exist only a few visualization techniques for frequent pattern analysis which allow the user to get a general idea of patterns contained in the data, and to interactively explore these patterns by changing important parameters, like the minimum support threshold in order to discover the effects immediately or to select only special items and discover only frequent itemsets which contain these items.

To support all these demands and to present an intuitive representation of detected frequent patterns, we introduce the FP-Viz technique. The basic idea of this approach is to provide a radial hierarchical layout, similar to the Sunburst [2] and Interring technique[3], to visually represent the frequent patterns in a high level view, and to allow the user to get details on demand by providing drill down and selection capabilities.

## 2 Visual Frequent Pattern Mining

The basic idea of our proposed method is to use an extended version of the FP-growth method to construct an initial frequent pattern tree (FP tree) from the input dataset  $D$

\*keim@inf.uni-konstanz.de

†schneide@inf.uni-konstanz.de

‡sips@inf.uni-konstanz.de

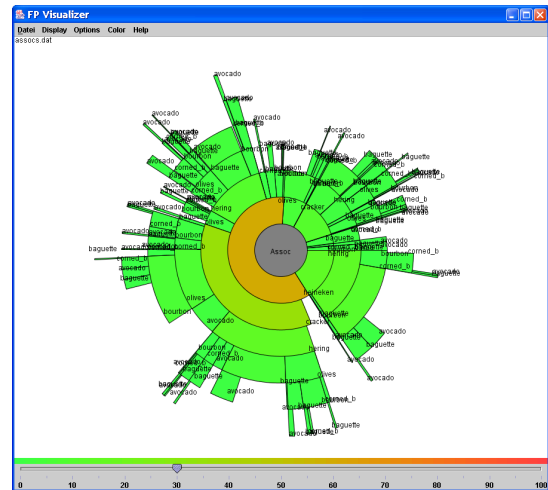


Figure 1: FP-Viz Framework: Radial Hierarchical Frequent Pattern Visualization. Shown is the FP tree visualization for the SAS Assocs example. Color shows item support. The slider allows the adjustment of the *minsup* threshold.

with respect to an predefined minimal support parameter *minsup*,  $0 \leq \text{minsup} \leq 100$ , where an itemset contained in the data is frequent, if its support is greater than or equal to the *minsup* threshold. In the next step, the FP tree is visualized and the user has the option to get a high level view of the data (itemsets) contained in the FP tree and to interactively explore details in the data like the support of items and itemsets. In order to take the special structure of the resulting FP tree into account, a radial hierarchical visualization technique is used to visualize the frequent pattern tree. Figure 1 shows the basic idea of our technique.

Since the FP tree represents a hierarchy of items and a path in the tree represents itemsets, where the hierarchy level represents the frequency (importance) of the corresponding item, we employed a radial visualization method which can represent such information. The root of the tree is visualized by a circle in the middle of the visualization. Starting from the root, each subtree, i.e. each pattern contained in the tree, is visualized by a circular ordering of circle segments, where each circle segment represents a single node in the tree. Segments which have the same distance from the center of the FP-Ring, have the same tree level in the underlying hierarchical structure. The order in which the circle segments of each level are placed, depends on the frequencies of the corresponding items. Color is used to visualize the support of each item or to distinguish the different items, depending on users' demands. The choice of the *minsup* value is critical in many applications, if it is too high, the FP tree is empty, if it is too low, the number of items in the FP tree is too high. Therefore FP-Viz provides the possibility to refine the *minsup* threshold interactively, and to see the changes instantly. Additionally the user may select single items to explore only frequent itemsets containing these items.

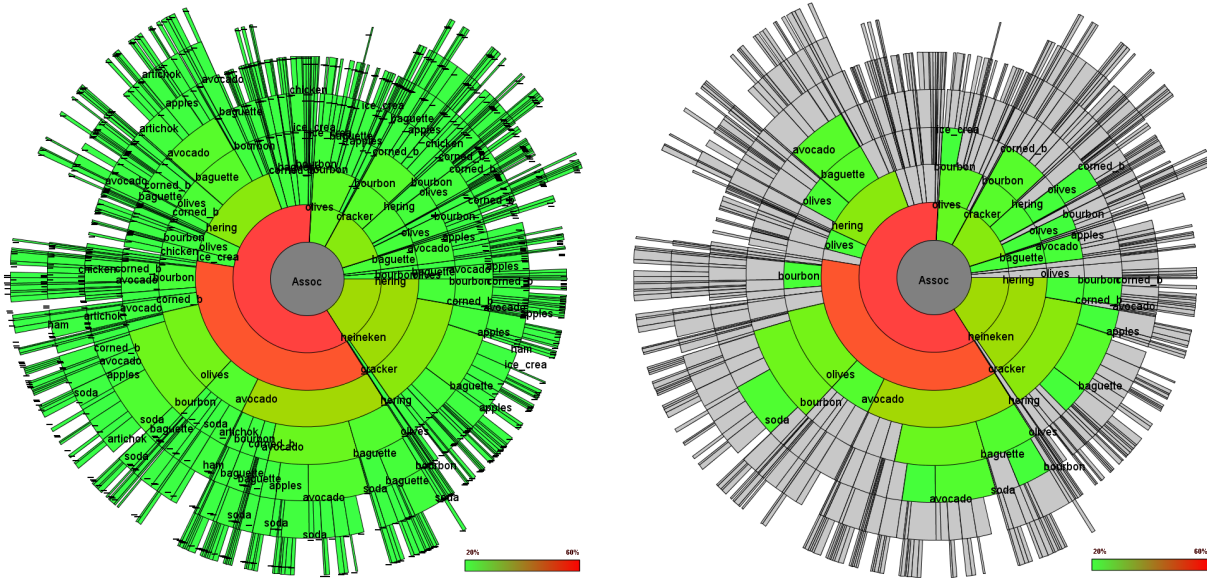


Figure 2: The left figure shows the FP-Viz technique applied to the SAS Assocs data with  $minsup = 30$ . In the right figure the  $minsup$  is then set to 40, infrequent items are faded out. Color shows support. "heineken" (sup=60) and "cracker" (sup=49) are the most frequent items.

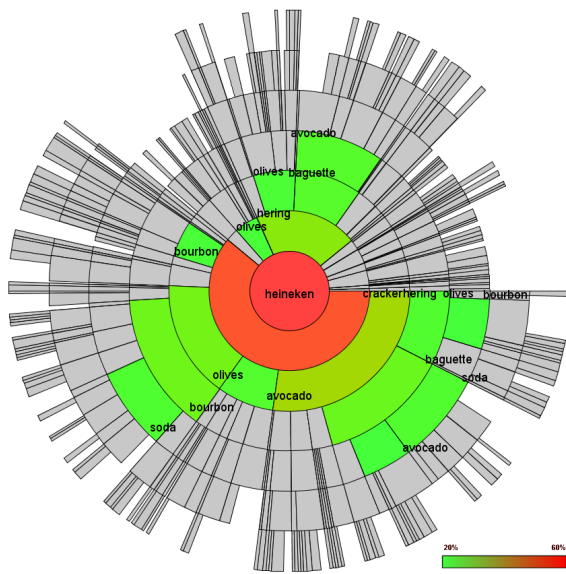


Figure 3: Visualizing the FP tree for item "heineken" from SAS Assocs data set with  $minsup = 40$ . Color shows the support of each FP tree node

### 3 Application Examples

We applied the FP-Viz technique to the SAS Assocs data example. This dataset contains 1000 itemsets containing 7 items each. After the user has defined the  $minsup$  threshold, e.g. by moving an interactive slider, FP-Viz computes a FP tree from the data set that contains only items whose support is  $\geq minsup$ , shown in figure 2 (left). Here the  $minsup$  was set to 30. The figure shows that the first level of the FP tree has 7 nodes, shown by 7 arcs in the innermost circle segment. Item "heineken" has the highest support, it is contained in approximately 60 percent of all transac-

tion. It turns out that "heineken" usually occurs together with "cracker". So the user gets a first impression about interesting itemsets contained in the data and may now adjust the  $minsup$ , shown in figure 2 (right), or select special items. Suppose the user moves the mouse on the arc labeled "heineken" since it may be the most interesting item, than the support for "heineken" is shown. If the user then clicks the mouse, a new FP tree for item "heineken" will be constructed and shown immediately, as shown in figure 3. This new tree contains only items which occur together with "heineken". It turns out that "heineken" occurs usually together with "cracker", "hering" and "olives". Color can either be used to distinguish the different items, by determining a unique color for each item, or to visualize the support of itemsets as shown in figure 2.

### 4 Conclusion and Future work

In this paper we presented the FP-Viz radial visualization tool to visually explore frequent itemsets from large databases. This tool gives a new viewpoint for exploring frequent patterns and allows a high level view on the underlying data. Our tool supports user interaction and is adaptable to a broad variety of applications. Future work will include the extension of the tool and its application to other datasets.

### References

- [1] J. Han, J. Pei, and Y. Yin. Mining frequent patterns without candidate generation. In *Proceedings of the 2000 ACM SIGMOD international conference*, pages 1 – 12, 2000.
- [2] J. T. Stasko and E. Zhang. Focus+context display and navigation techniques for enhancing radial, space-filling hierarchy visualizations. In *Proceedings of the IEEE Symposium on Information Visualization*, page 57, 2000.
- [3] J. Yang, O. M. Ward, and E. A. Rundensteiner. Interring: An interactive tool for visually navigating and manipulating hierarchical structures. In *Proceedings of the IEEE Symposium on Information Visualization*, page 77, 2002.