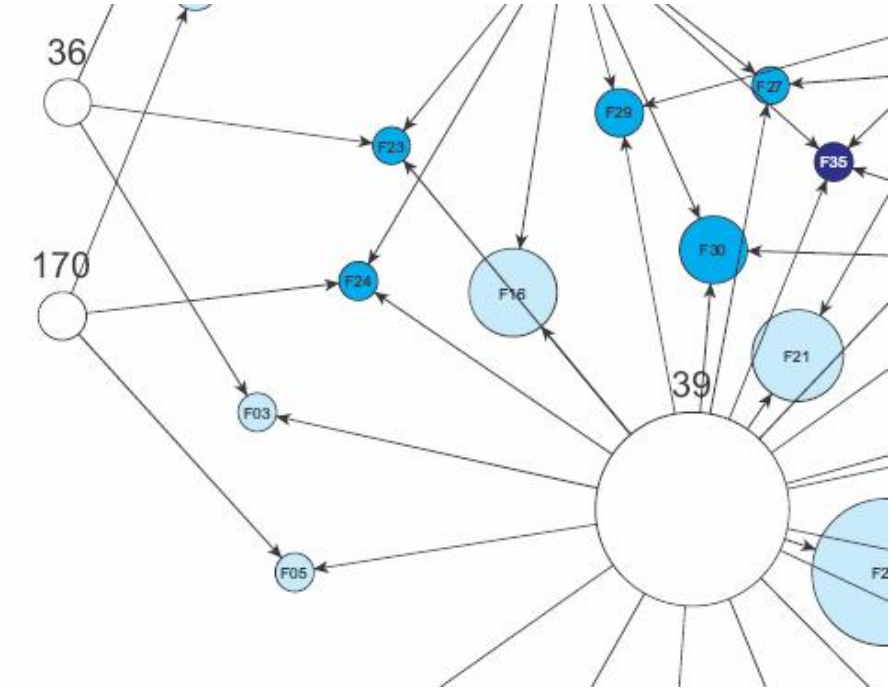
A Framework for Visualizing Association Mining Results



Gurdal Ertek & Ayhan Demiriz

Sabancı Üniversitesi

sakarya üniversitesi
http://www.sakarya.edu.tr

Introduction

Data Mining

- Developing deeper understanding of the data
- Discovering hidden patterns
- Coming up with actionable insights
- Identifying relations between variables, inputs and outputs
- Predicting future patterns

Introduction

- **Association mining**
  - Very popular data mining method
  - Produces interpretable and actionable results
- **Association rules**
  - "If
    the customer purchases Item A,
    then with probability C
    she will buy Item B."

Antecedent

Consequent

Introduction

$$C = \frac{frequency(A \cap B)}{frequency(A)}$$

Confidence: **Conditional probability of having Item B given Item A.**

$$S = \frac{frequency(A \cap B)}{T}$$

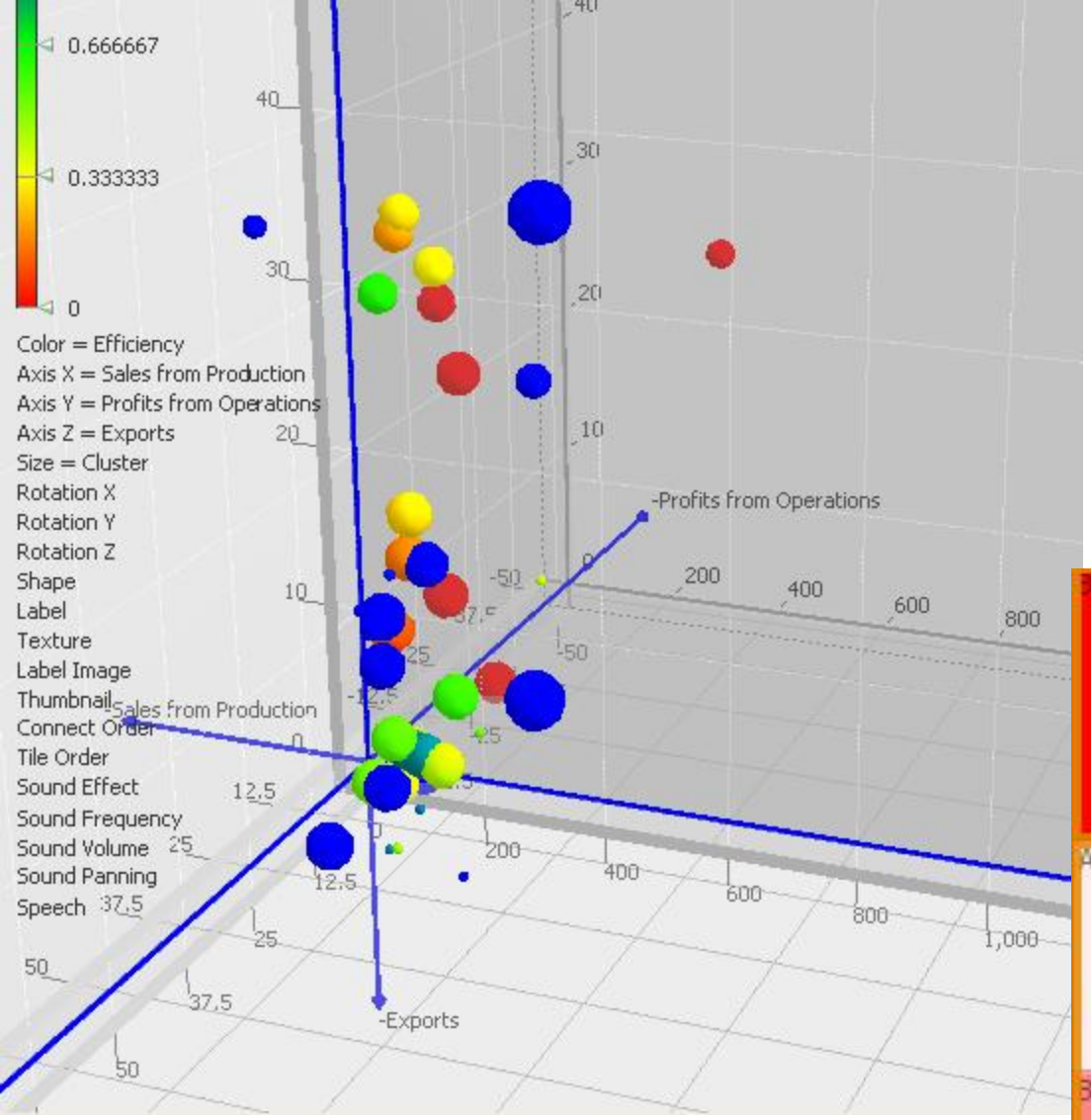
Support: **Fraction of transactions having both Item A and Item B.**

Introduction

- **Our study**
  - **Single-dimensional**
  - **Single-level**
  - **Boolean association rules**
  - **In the context of market basket analysis**
- **Literature**
  - **Multitude of interestingness measures (ex: support)**
  - **Efficient algorithms to compute these measures**
  - **Few studies focus on the *interpretation* of the association mining results**

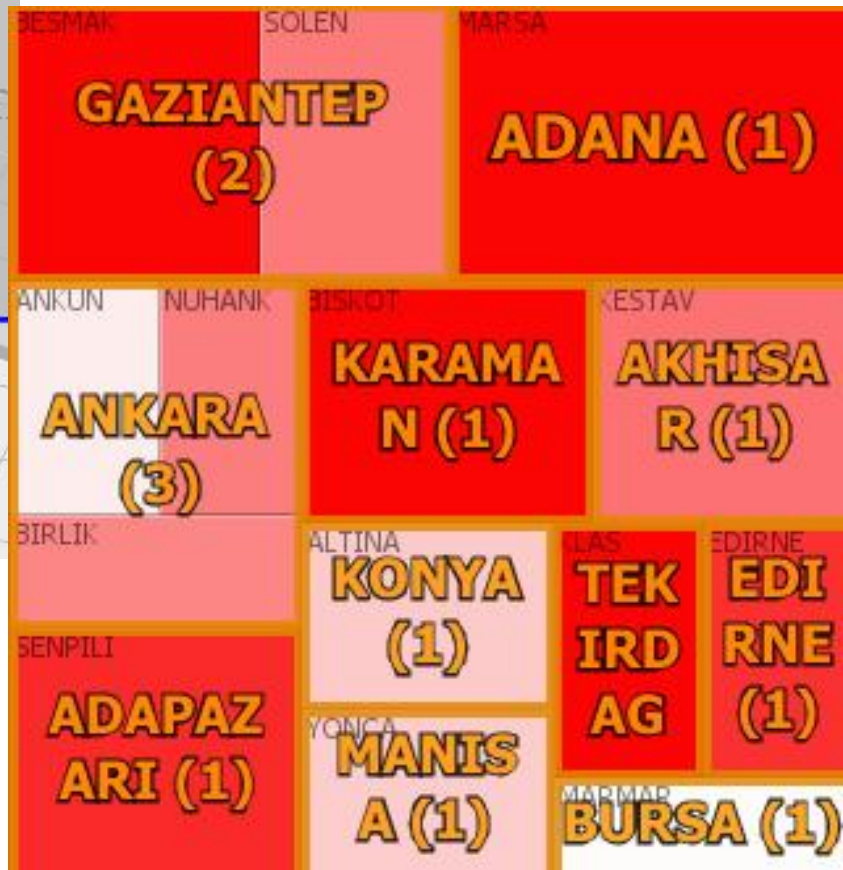
Information Visualization

- Growing field of computer science
- Visually representing multi-dimensional data for knowledge discovery
- User-friendly and successful software tools
  - Miner3D
  - Spotfire
  - Advizor
  - DBMiner
  - IBM Intelligent Miner Visualization
  - Omniscope
  - yEd



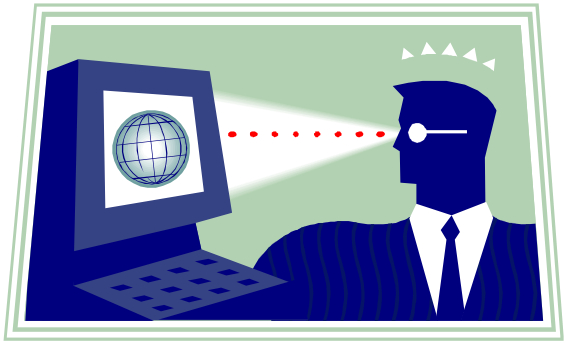
Miner3D
www.miner3d.com

Omniscope
www.visokio.com



Information Visualization

- **The analyst does not have to understand complex algorithms.**



- **Almost no training required.**

- **There are no limits to the types of insights that can be discovered.**

Our Study

- **Motivation**
  - **Visualizing the results of association mining can help end-users significantly in knowledge discovery**
- **Contribution**
  - **Framework that merges association mining with information visualization**
  - **Flexible and human-centered way of discovering insights**

Framework

- **Graph-based framework**
- **Visualizing and interpreting the results of association mining algorithms as directed graphs.**
- **The items, the itemsets, and the association rules are all represented as nodes.**
- **Arcs represent the links between the items and the itemsets/associations.**

Framework

- **Exploit already existing**
  - **Graph Drawing Algorithms**
  - **Software *(yEd Graph Editor)***

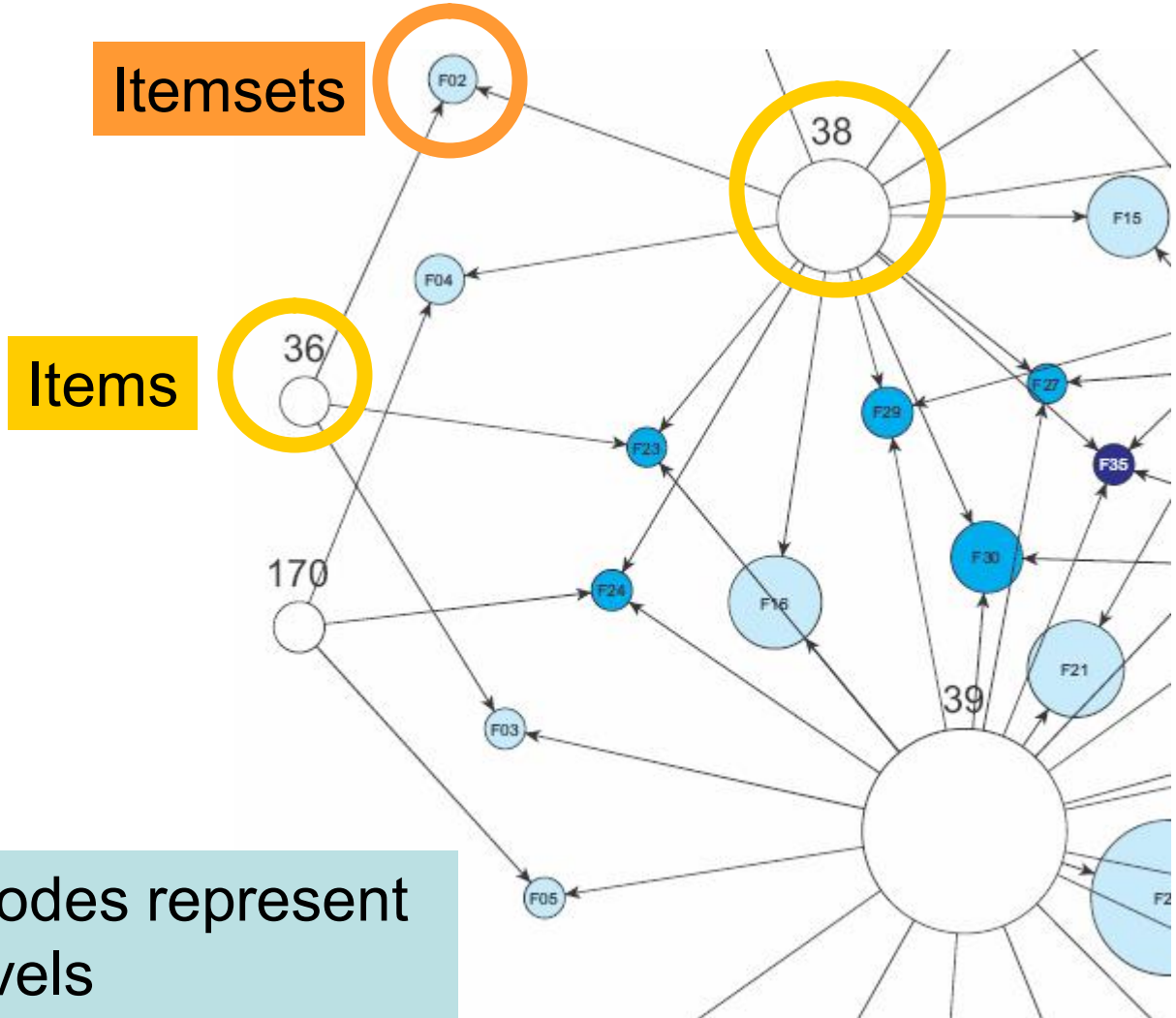
  in the information visualization literature

- **For visualization of results which are generated by already existing**
  - **Association Mining Algorithms *(Apriori)***
  - **Software *(Borgelt's implementation)***

  in the data mining literature

Framework

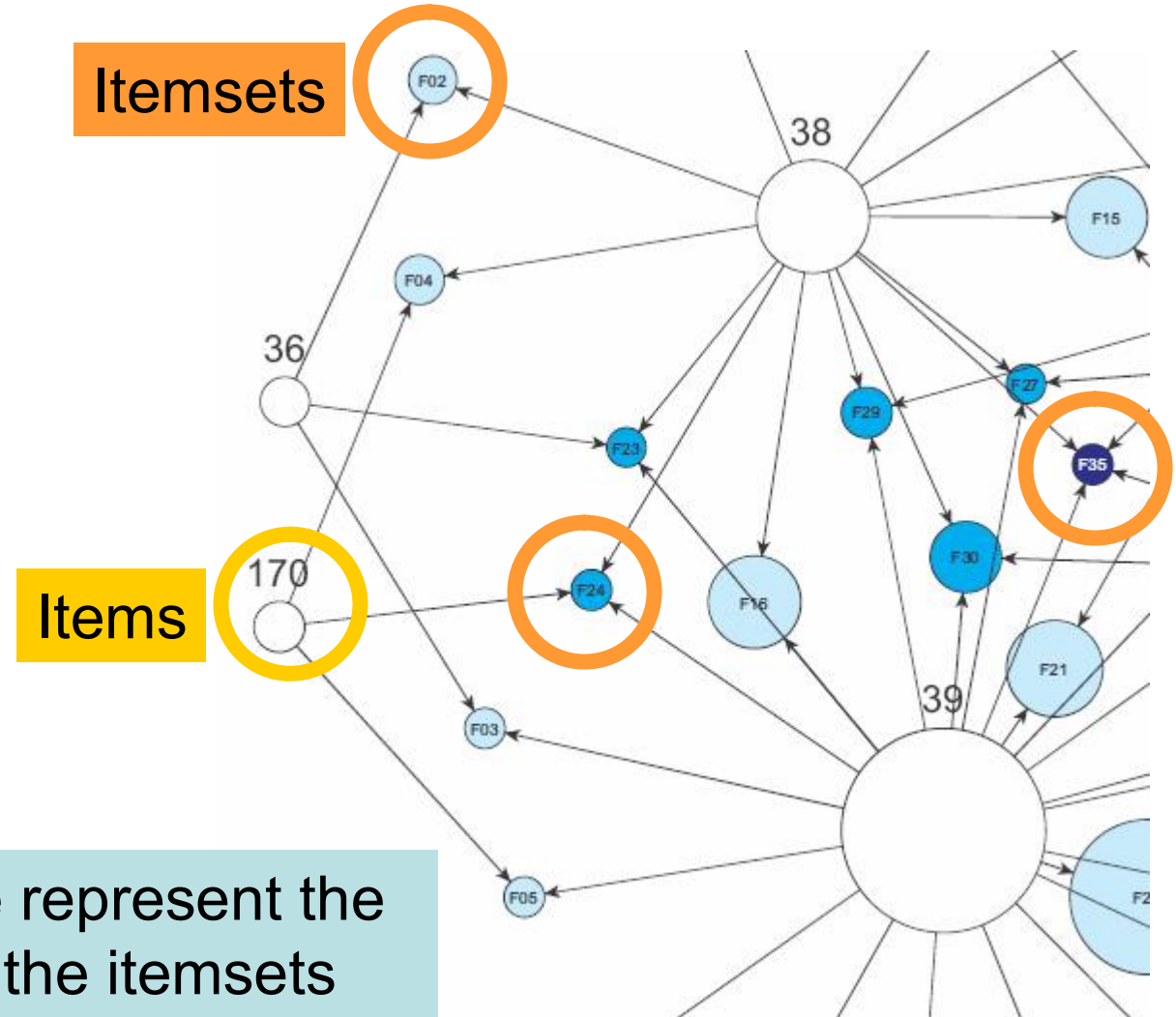
- **Visualizing Frequent Itemsets**



Areas of the nodes represent the support levels

Framework

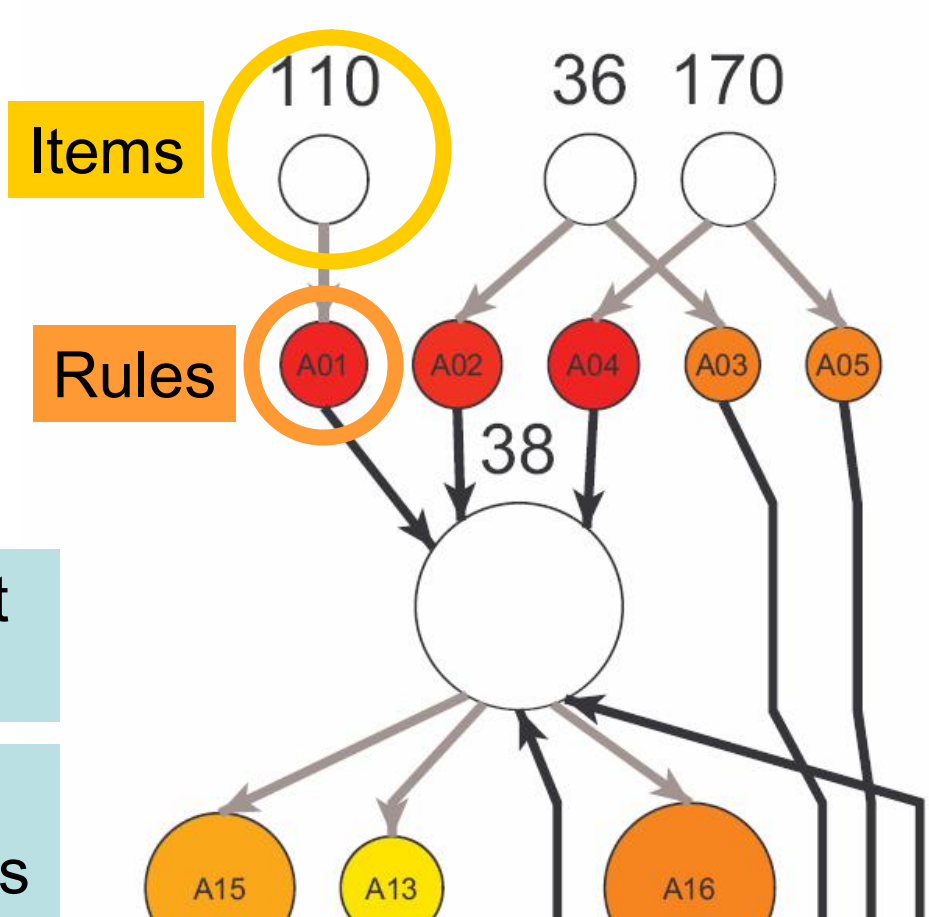
- **Visualizing Frequent Itemsets**



Shades of blue represent the cardinalities of the itemsets

Framework

- **Visualizing Association Rules**



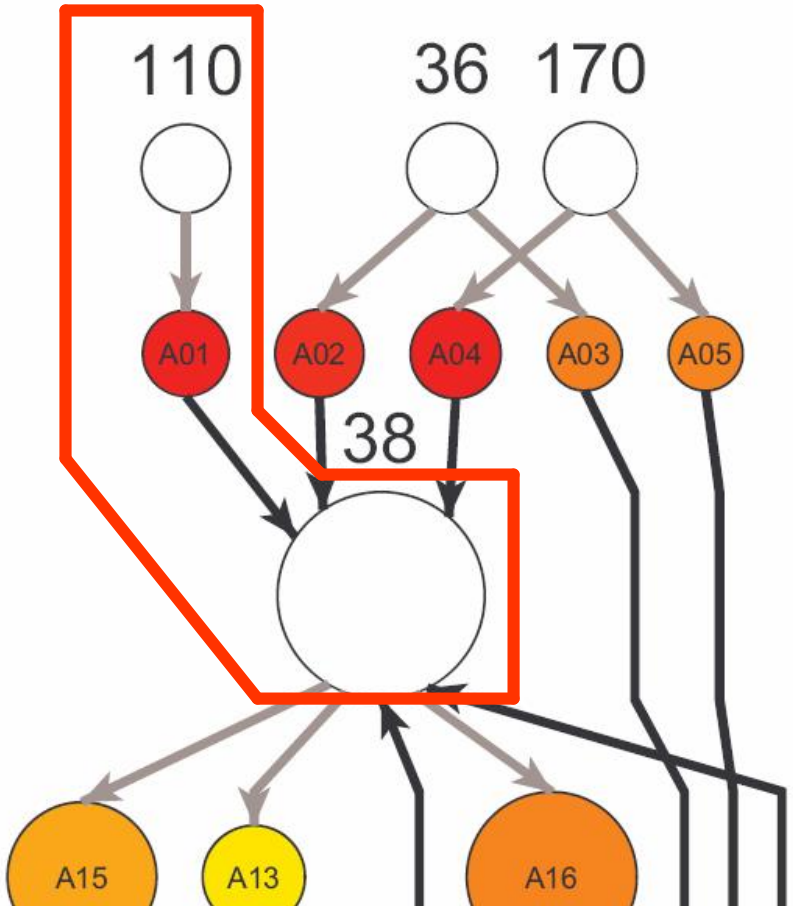
Areas of the nodes represent the support levels

Node colors show the confidence values of the rules

Framework

- **Visualizing Association Rules**

Rule A01: **If** Item110 **Then** Item38



Incoming arcs of the rule nodes are shown in **grey** and outgoing arcs are shown in **black**

Steps in Implementing the Framework

1. Collect market transactions data
2. Run an efficient implementation of the Apriori algorithm to generate
   - Frequent itemsets
   - Association rules
3. Translate the results of the Apriori algorithm into graph specifications
4. Create the graph objects based on the calculated specifications
5. Run the available graph layout algorithms and try to visually discover interesting and actionable insights

Case Study

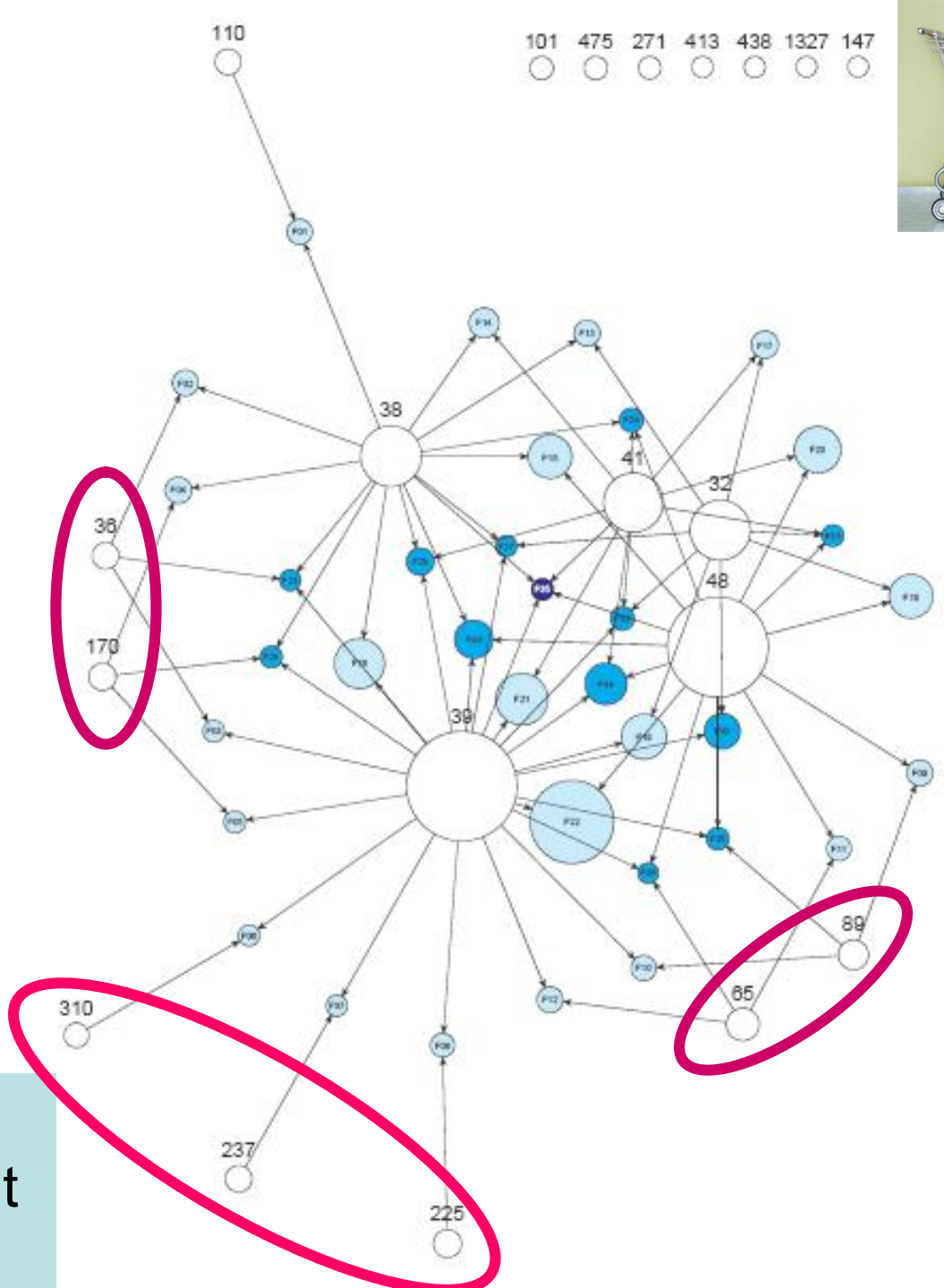
Analysis of Supermarket Sales Data

The Data

Field Name	Desciption
TRANSACTION_ID	Transaction ID
PRODUCT_NO	Product Number

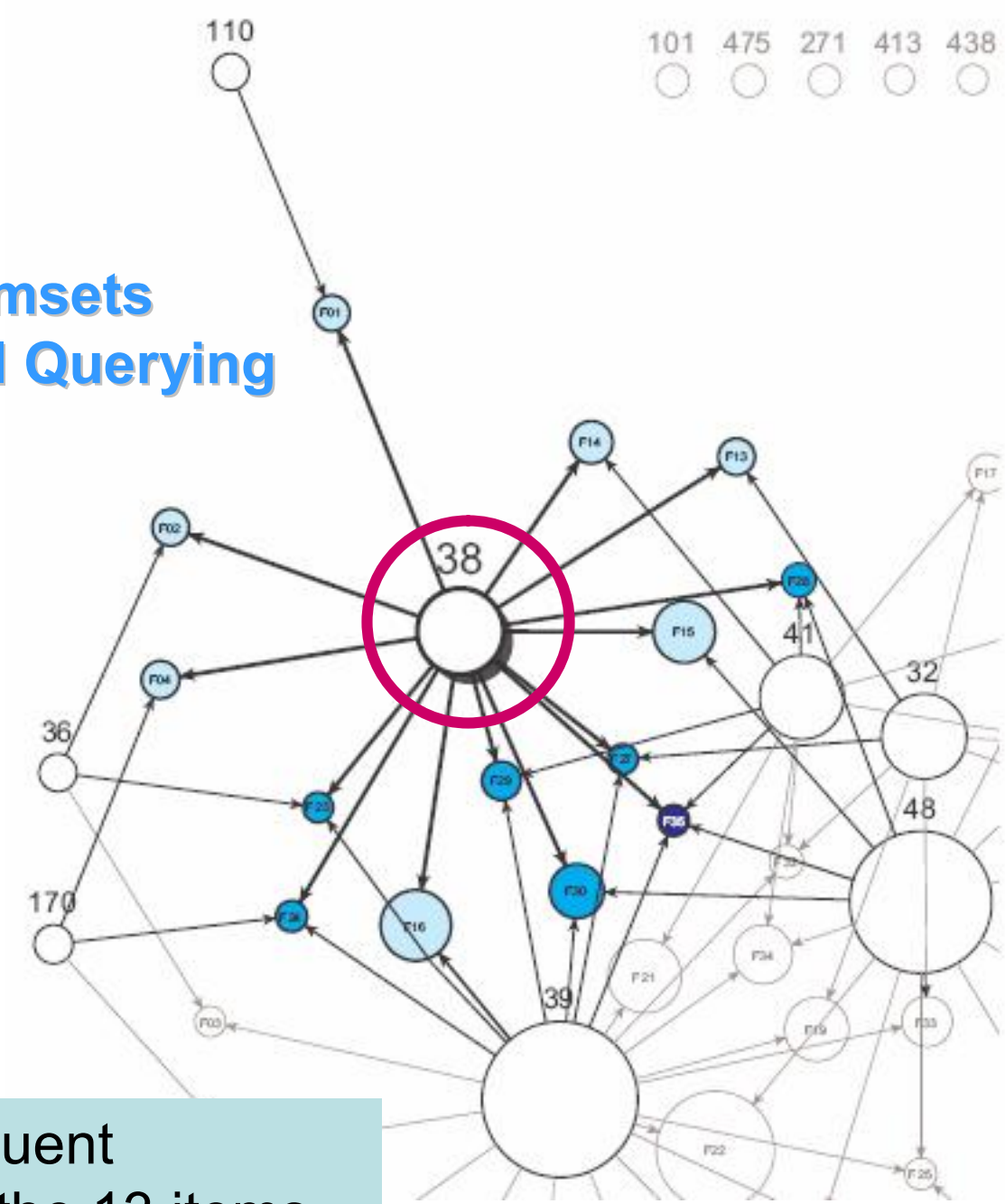
- Belgian supermarket
- ~90,000 transactions
- ~16,000 unique items

Frequent Itemsets



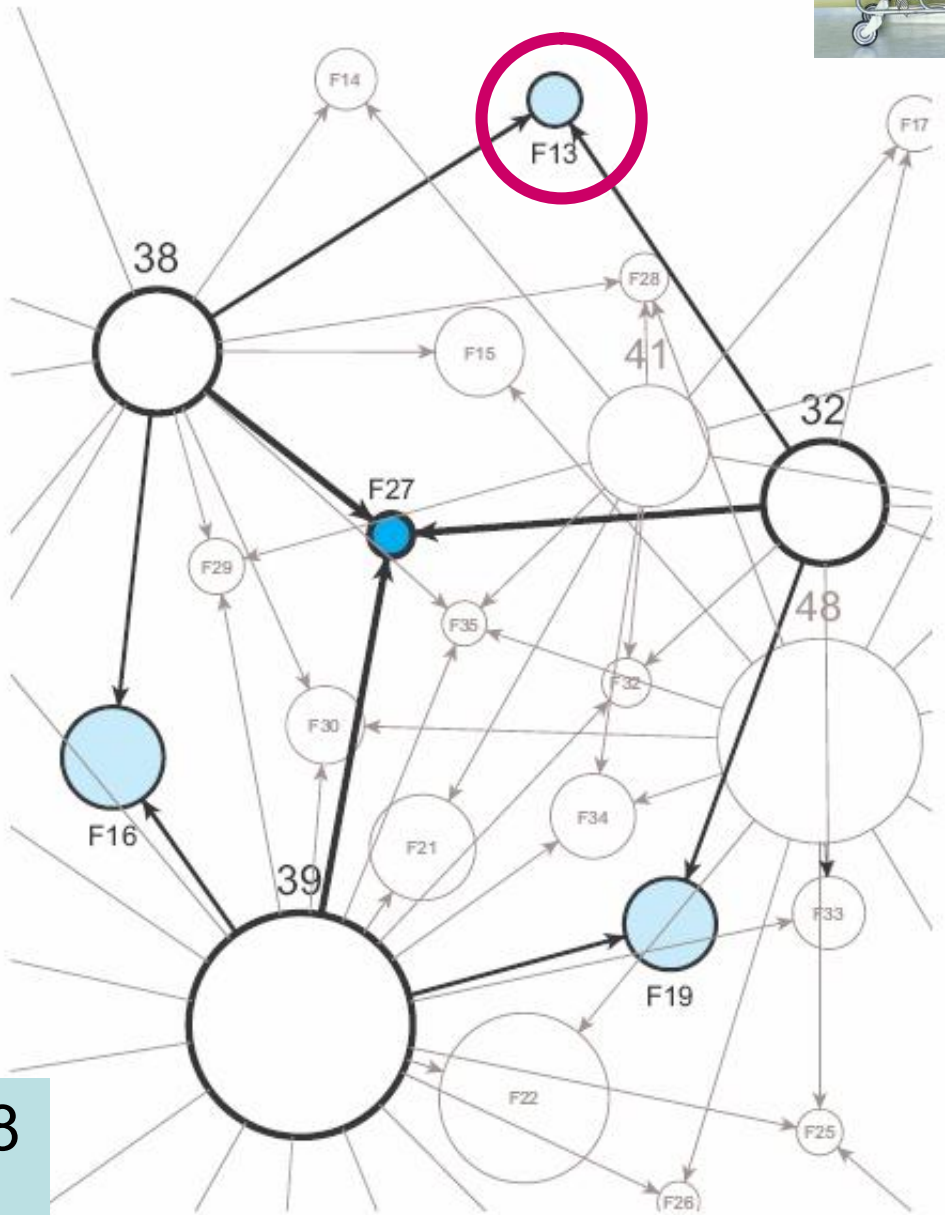
Support > 2%
Classic Organic Layout
Catalog Design

Frequent Itemsets
Interactive Visual Querying



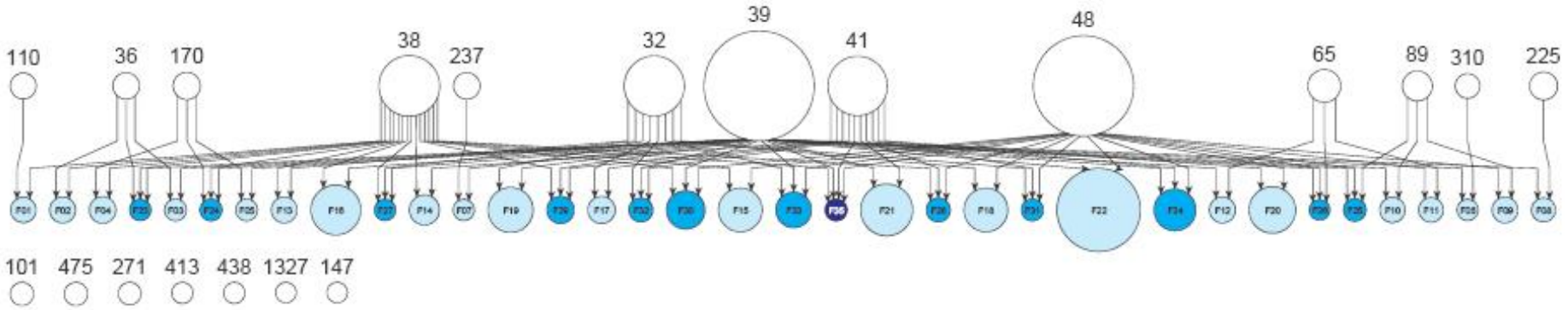
Item 38 forms frequent itemsets with 7 of the 13 items

Frequent Itemsets Interactive Visual Querying



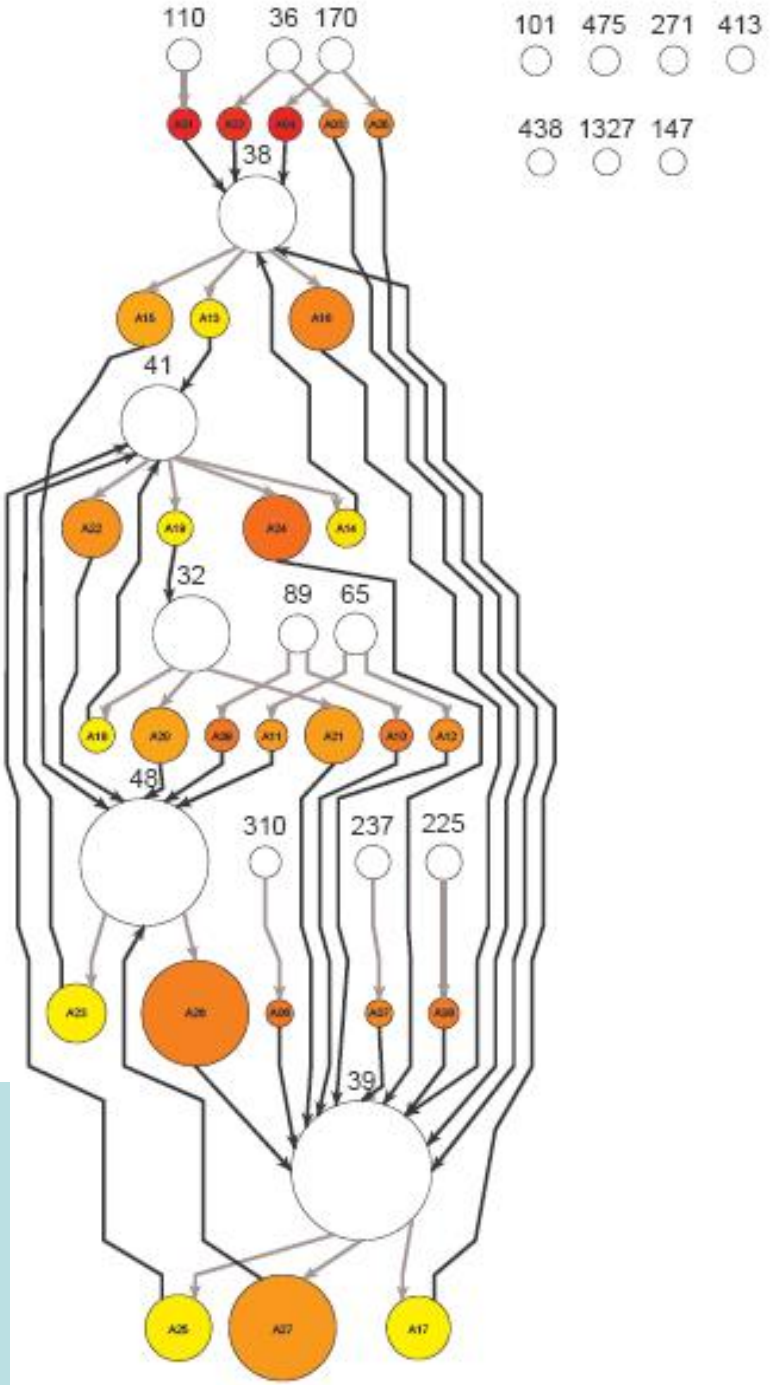
Association between items 38 and 32 is significantly low

Frequent Itemsets



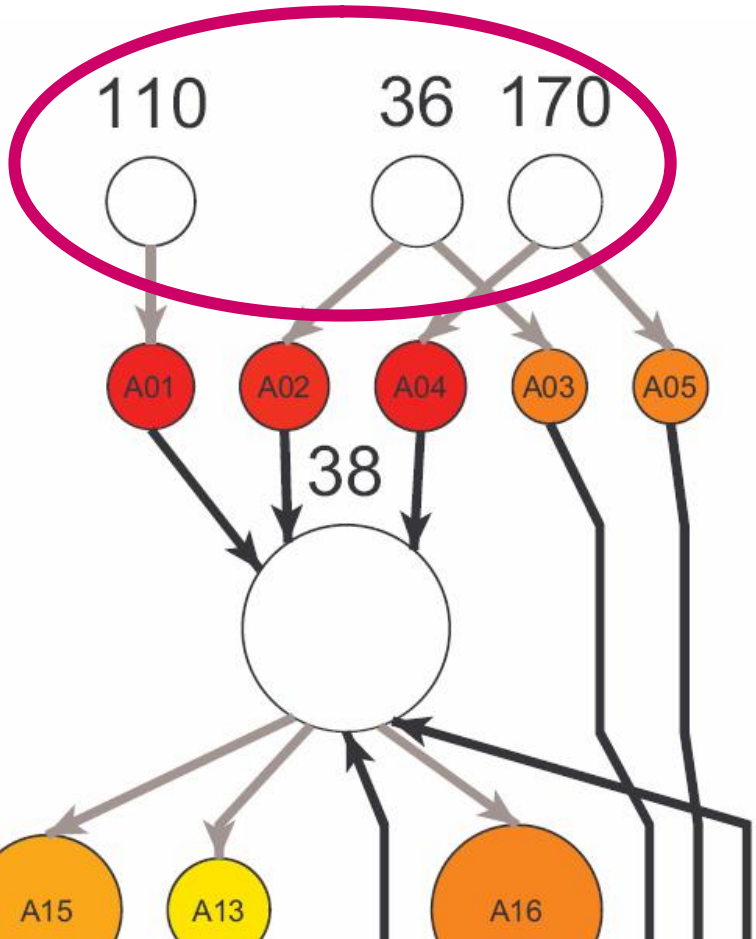
Interactive Hierarchical Layout
Planning supermarket shelf layouts

Association Rules



Support > 2%
Confidence > 20%
Classic Hierarchical Layout
Cross-selling strategies

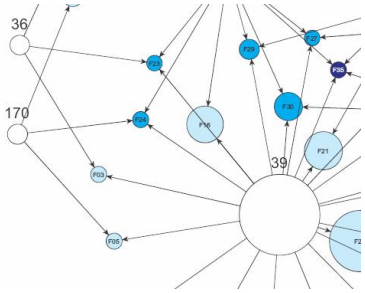
Association Rules

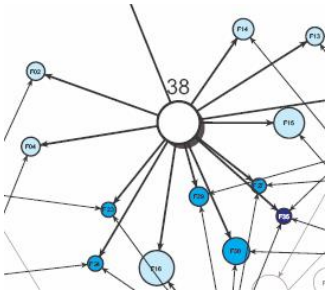


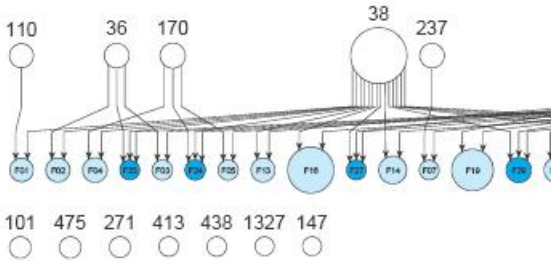
- Initiate a promotional campaign for Items 110, 36, or 170
- Place these items next to Item 38

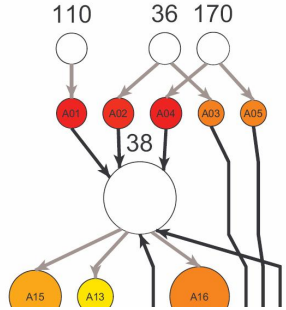
Association Rules

Other “sales drivers”


36
170
39


38


110
36
170
38
237
101
475
271
413
438
1327
147


110
36
170
A01
A02
A04
A03
A05
38
A15
A13
A16

THANK YOU