

# TNS EX-A-MINE™

## TextMiner

Strategies against  
Information Overload





- Text Mining is designed to study textual information such as responses to open-ended questions, interviews, titles, journal articles, etc.
- It may be used for automatic categorization of text using a dictionary approach as well as for manual coding.
- Numerous exploratory data analysis and graphical tools can be used to explore the relationships between the content of documents and information stored in categorical or numeric variables such as the gender or the age of the respondent, brand use, etc.
- Relationships among words or categories can be identified using **hierarchical clustering** and **multidimensional scaling** analysis. **Correspondence analysis** and **heatmap plots** are used to explore relationships between keywords and different groups of individuals or brands.



### ■ Call centre contacts

- Which are the main issues of customers who contact the call centre?
- Which complaints and problem constellations occur in combination?

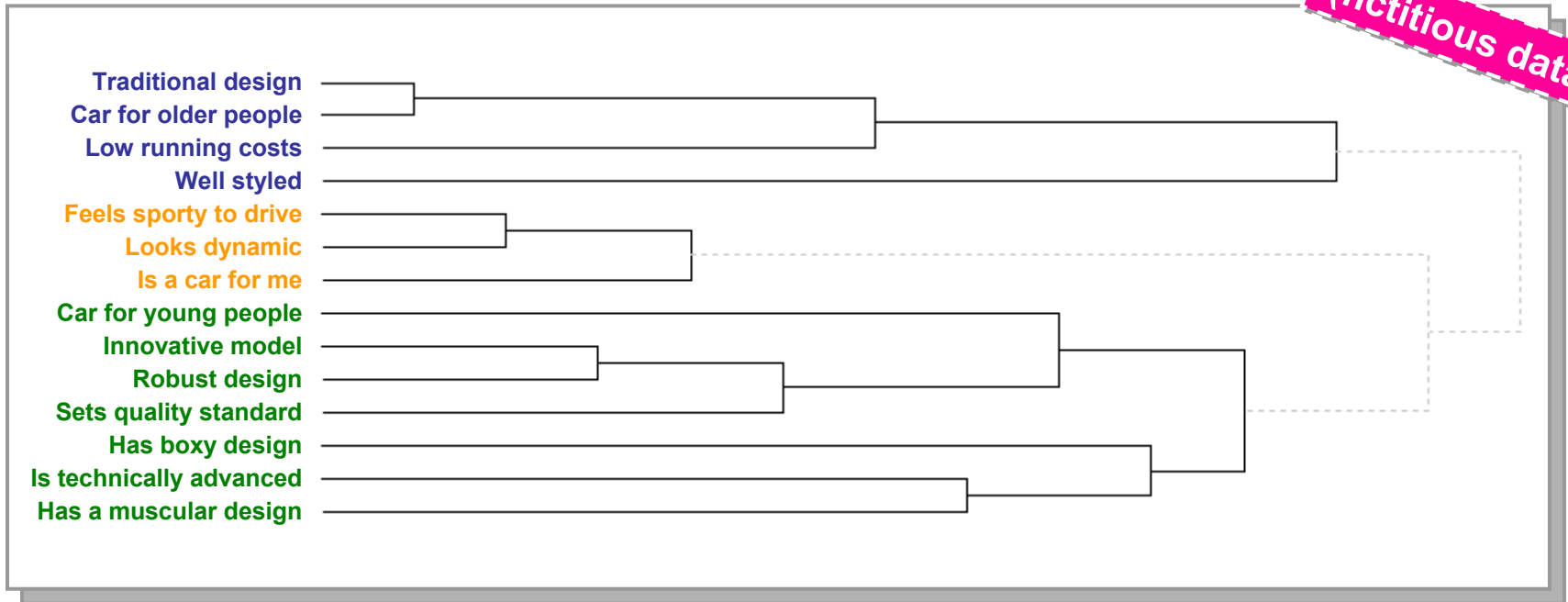
### ■ Open-ended questions in surveys

- Which issues do the respondents evaluate positively, and which ones negatively?
- Which concepts are commonly used in combination and which can be summarised to superordinate categories?

### ■ Analysis of interviews und documents

- To which categories can texts / documents / interviews / Emails be summarised automatically?

# Output Example: Hierarchical Clustering

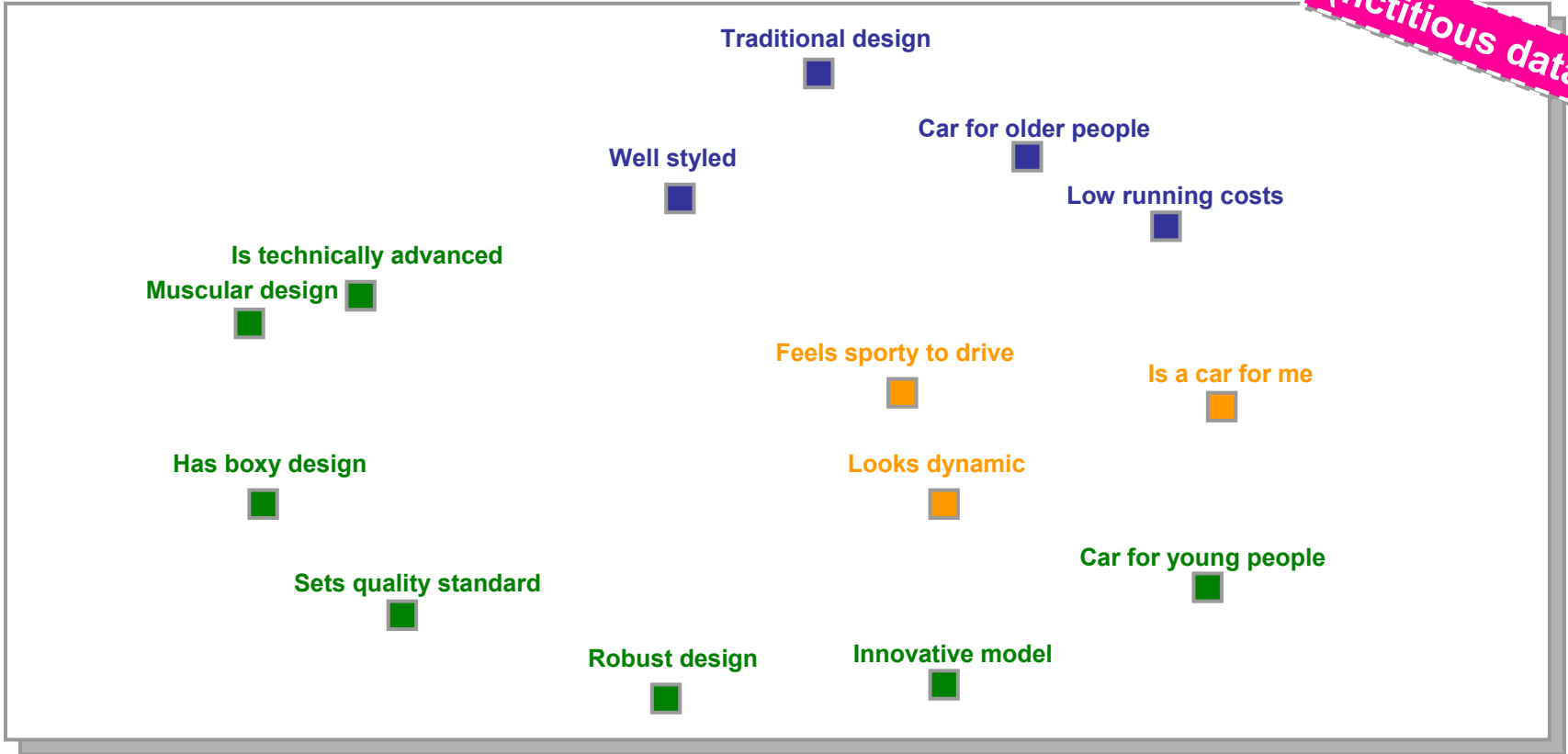


- The **dendrogram** shows keyword co-occurrences or word proximity.

# Output Example: Multidimensional Scaling



(fictitious data)



■ **Multi-dimensional scaling** is used to examine proximity of keywords.

# Output Example: Crosstabs

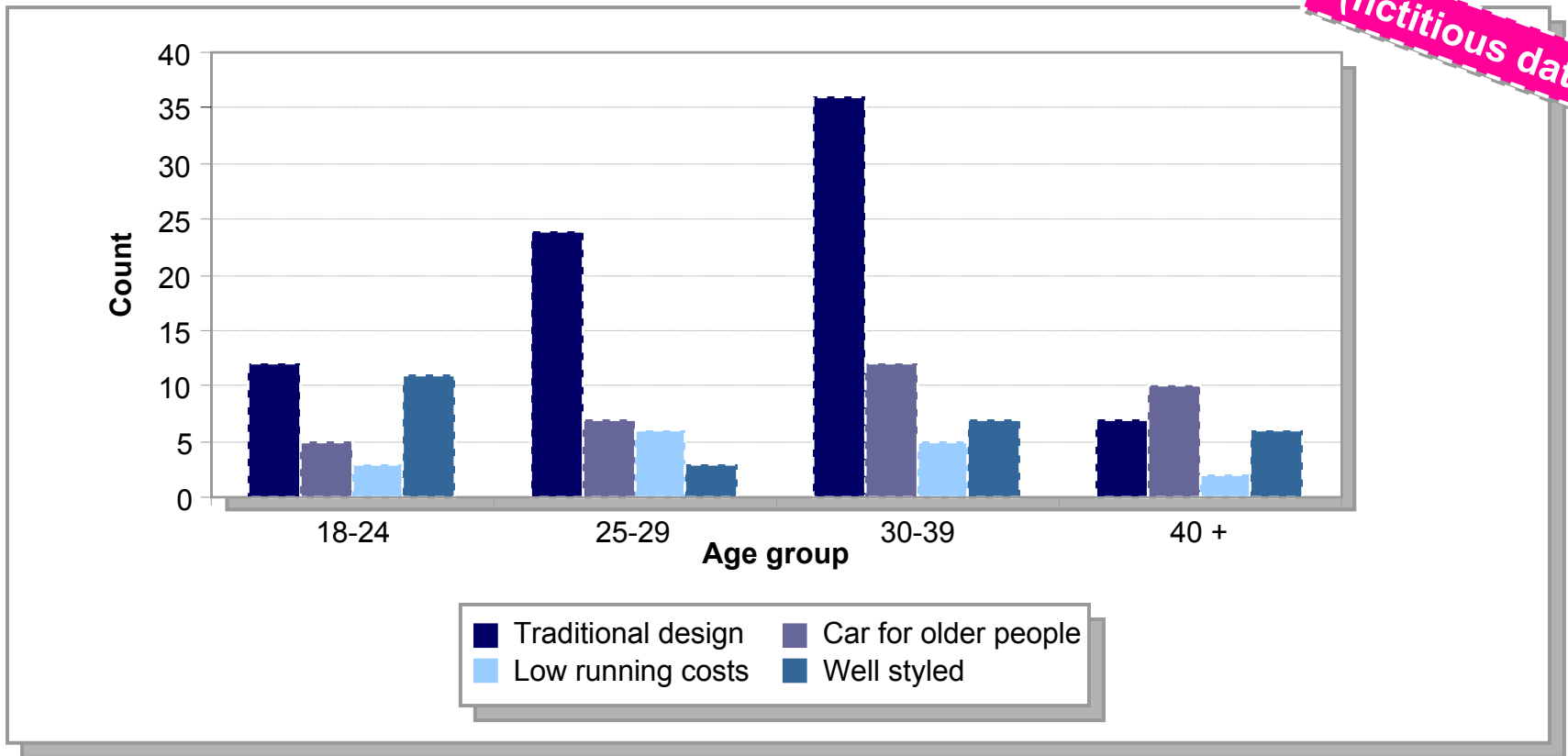


**(fictitious data)**

	Age groups			
	18-24	25-29	30-39	40+
Traditional design	12	24	36	7
Car for older people	5	7	12	10
Low running costs	3	6	5	2
Well styled	11	3	7	6
Feels sporty to drive		3	1	1
Looks dynamic	1		3	1
Is a car for me	11	14	13	5
Car for young people	5	6	5	4
Innovative model	3	4	7	5
Robust design	12	19	26	2
Sets quality standard	2	3	7	1
Has boxy design	1		8	
Is technically advanced	3	2	4	2
Has a muscular design	2	7	7	6

- **Crosstabulation** examines the relationship between words or content categories and the values of the selected independent variables (in this case age groups, one could also use car brands, gender etc.).

# Output Example: Bar Charts

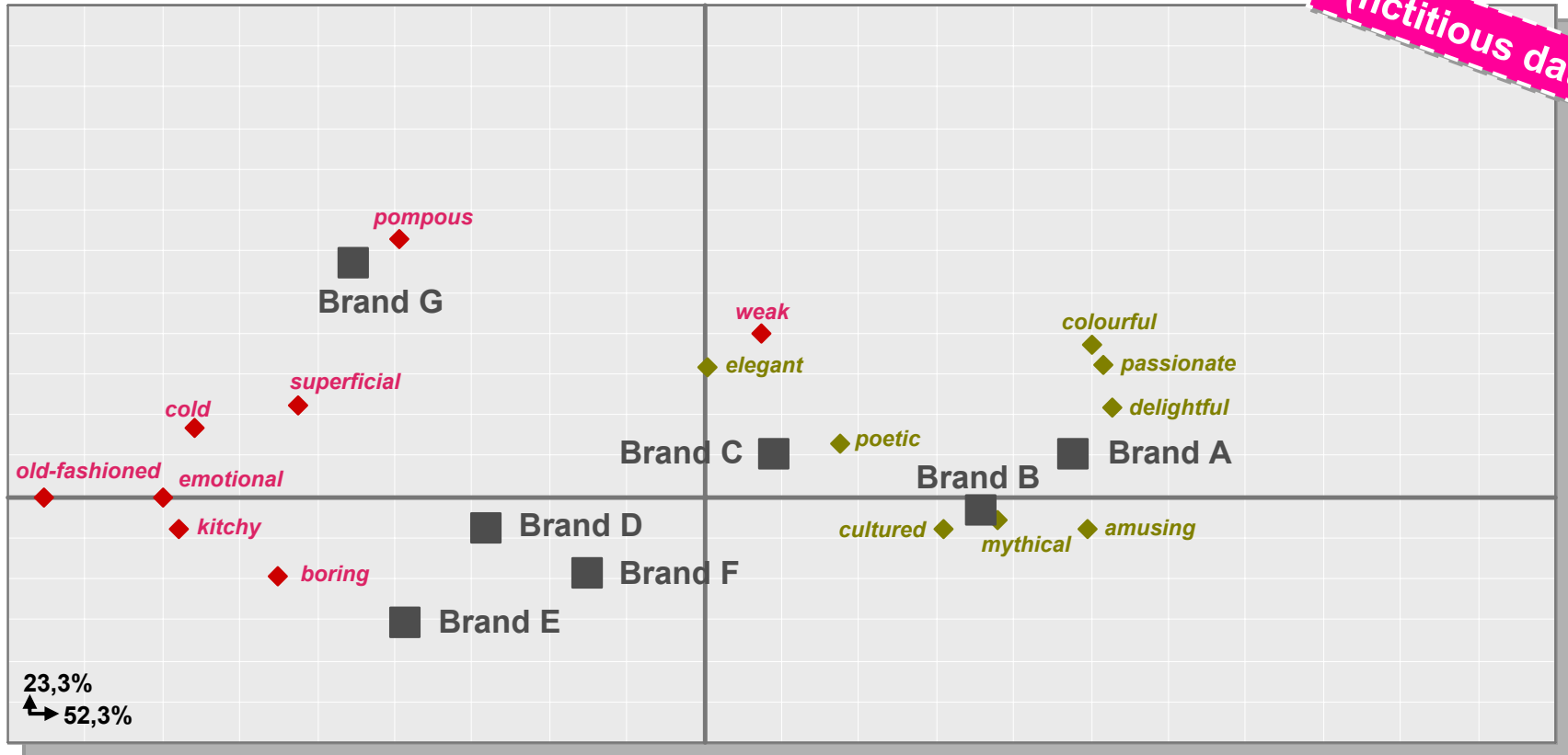


- From the Crosstabs, **bar charts** can be used to examine the relationship between keywords and categorical variables.

# Output Example: Correspondence Analysis

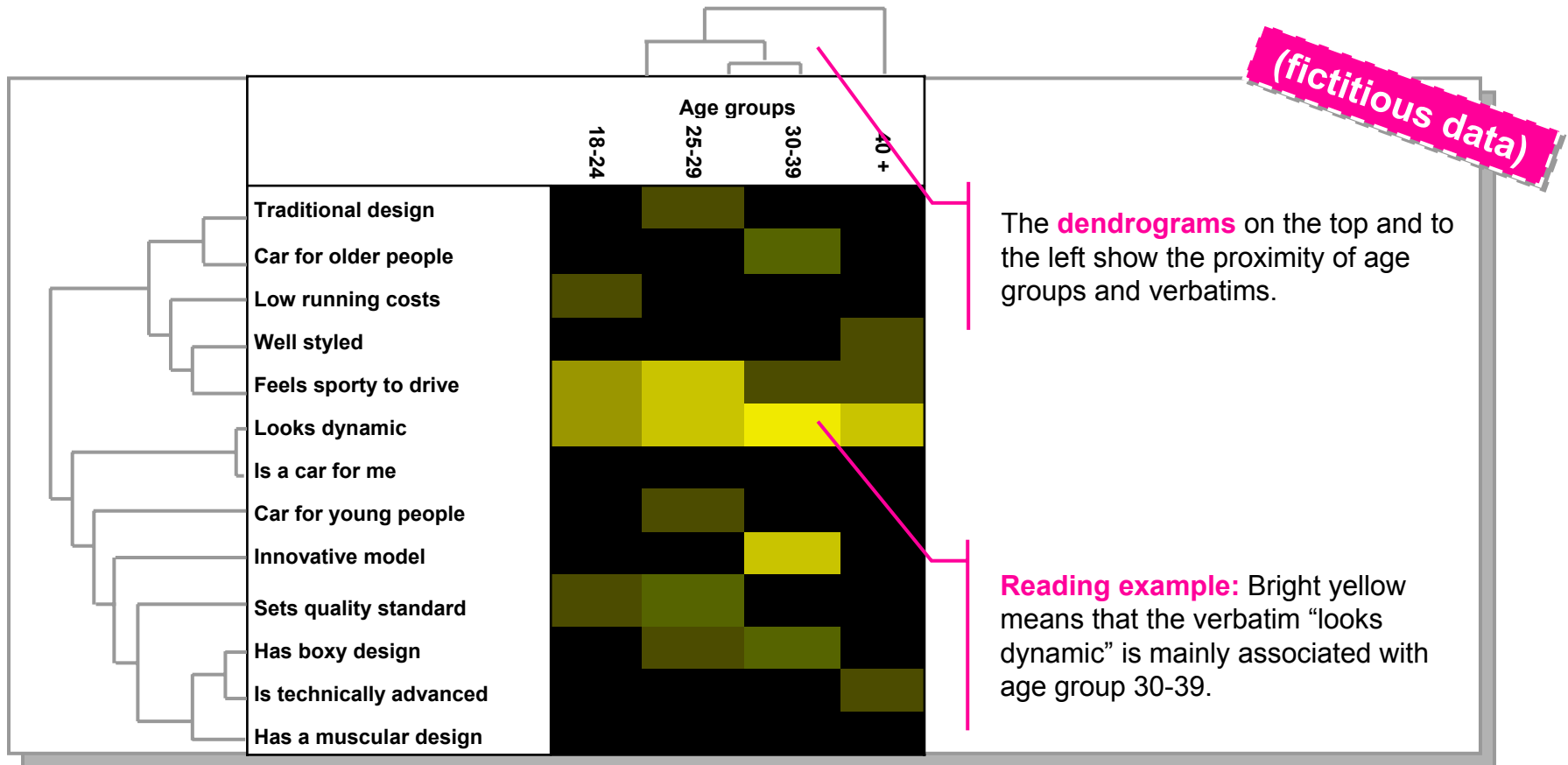


(fictitious data)



■ **Correspondence analysis** can be used to graphically represent the relationship between objects (brands, products) and verbatims.

# Output Example: Heatmaps

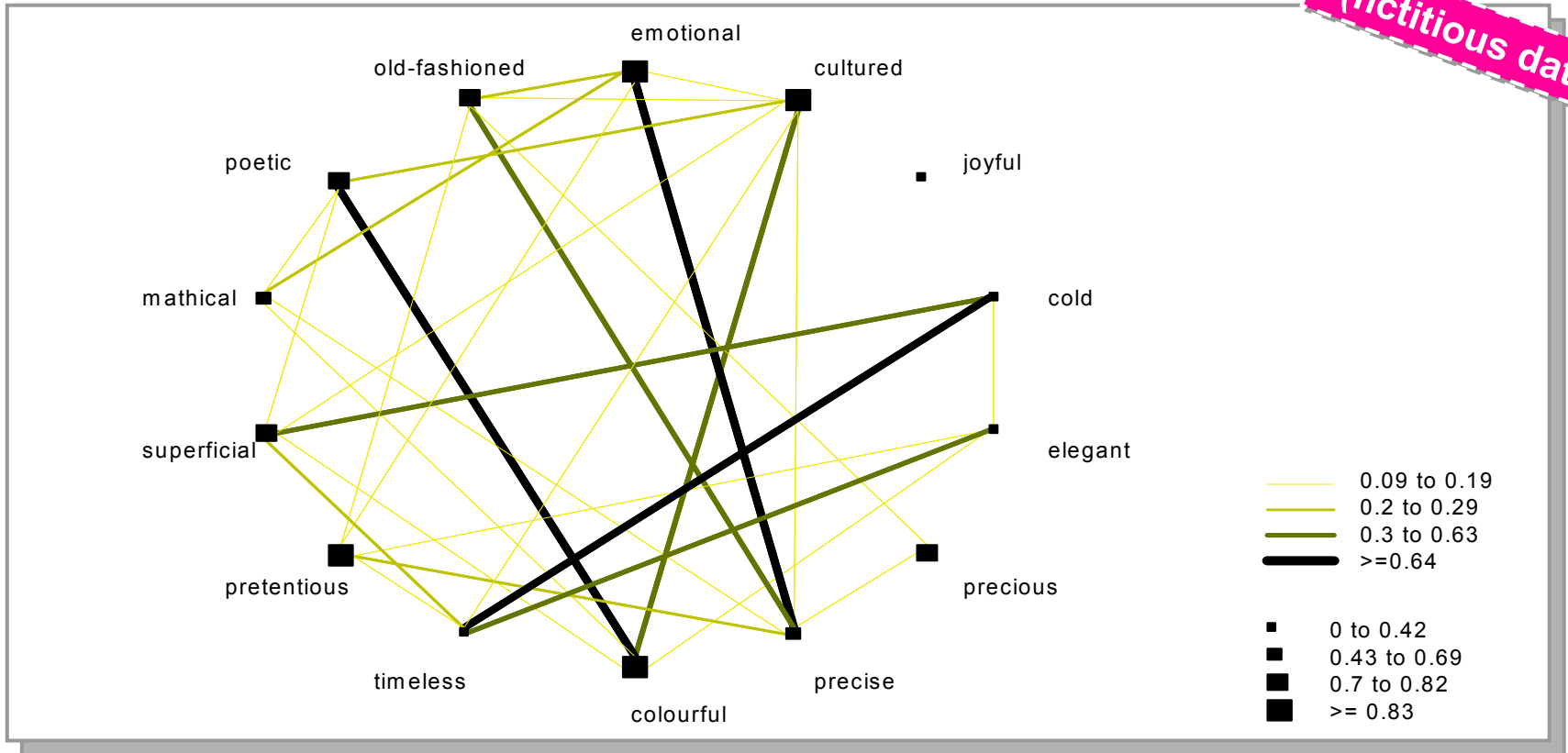


- **Heatmaps** provide another way to graphically explore relationships in large tables. In a heatmap plot **relative frequencies** are represented by different color brightness or tones.

# Output Example: Web Graphs



(fictitious data)



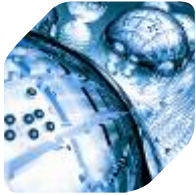
■ **Web graphs** show which verbatims frequently occur conjointly. Thick lines represent strong associations, thin lines represent weak connections.

# TNS EX·A·MINE™ at a Glance



- **Competence Centre** for Data Mining and Data Fusion of TNS Infratest

- Support of all TNS Market Sectors in the areas of **Data Mining**, **Data Fusion** and **Holistic Segmentation** for analytical CRM and Database Marketing



- **Highly specialised (Senior) Consultants** and (Senior) Data Analysts whose areas of expertise include Economics, Statistics and IT



- Long **intersectoral and international project experience** with complex analytic designs and study conceptions

- Toolbox with **state-of-the-art methods** in the areas of classic multivariate statistics and Data Mining



- Close relations with research facilities and software industry; ongoing adjustment of methods and tools to the **state-of-the-art**

# The TNS EX·A·MINE™ Algorithms-Toolbox

- **Multivariate statistics**
  - Logistic, Categorical, Linear Regression, EM Algorithm
  - Multivariate Adaptive Regression Splines (MARS)
  - Ridge Regression, Robust Regression
  - Cluster Analysis, Latent Class Analysis
- **Decision Trees / Decision Rules, Automatic Learning**
  - C&RT, C5.0, QUEST, CHAID, Association rules
  - MART – Multiple Additive Regression Trees, Random Forest
  - Nearest Neighbours / Instance based learning EX·A·MINE Profiler
- **Artificial Neural Networks**
  - Cascade Correlation Learning Architecture, MLP, SOM
- **Hybrid Methods**
  - Automatic OLAP Navigation and Search
  - Genetic Algorithms for variable selection
  - Neuro Fuzzy Algorithms, interactive visualisation of data



Holistic  
Customer  
Understanding  
[EX▪A▪MINE  
Services]

**Dr. Robert Hartl**  
Tel. +49 89 5600 – 1320  
[robert.hartl@tns-infratest.com](mailto:robert.hartl@tns-infratest.com)

