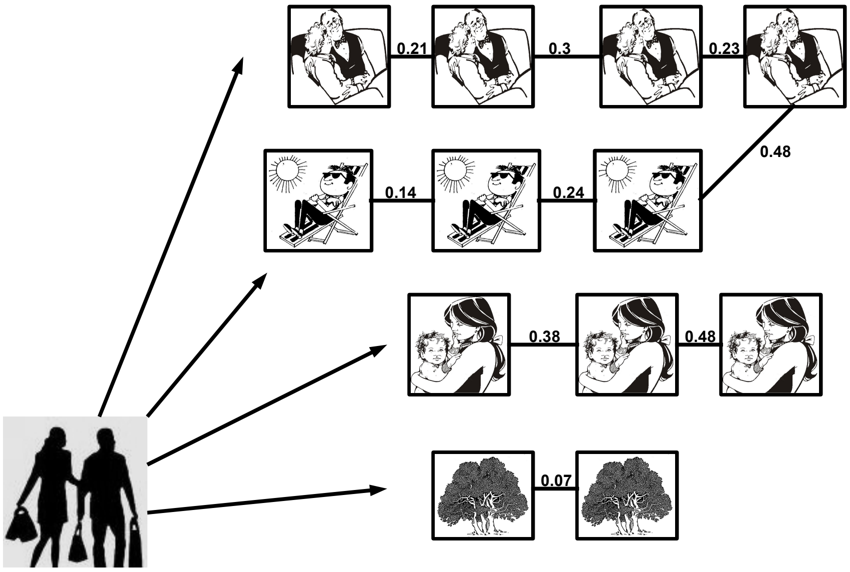
Dynamische Clusteranalyse für DM-Verkaufsdaten

Diplomarbeit von Selma Mukhtar

Thesis of Selma Mukhtar

April 2009



Supervisors: Prof. Dr. Dorothea Wagner
and Dipl.-Math.-tech. Robert Görke

Institute of Theoretical Informatics
Universität Karlsruhe (TH)

Declaration

Hiermit erkläre ich, dass ich die vorliegende Diplomarbeit selbstständig verfasst und die vorgestellten Ergebnisse ohne die Hilfe Dritter erarbeitet habe. Ich habe auf keine anderen als die angegebenen Quellen und Hilfsmittel zurückgegriffen.

Selma Mukhtar
Karlsruhe, den 24. April 2009

Acknowledgment

First of all, I would like to thank my supervisors Prof. Dr. Dorothea Wagner and Robert Görke for their great support throughout my thesis. Especially, I want to thank Prof. Wagner for guiding my work into the right direction and for providing me with essential hints. I am grateful that she was open about the cooperation with *dm-drogerie markt* and that she gave me the opportunity and financial support to visit a training course which offered me valuable and interesting insights in data mining techniques. From the first day on, I felt integrated into her institute and enjoyed the pleasant working atmosphere. My deepest thanks go to Robert Görke from whom I have learnt so many things about scientific working and writing. He always took the time to discuss my results with me. I also thank him for initiating the cooperation with *dm-drogerie markt*.

Further I want to thank Andreas Gessner, Thomas Orth, Saskia Koch and particularly Michael Martin of *dm-drogerie markt*. They supported me with their expertise and expert opinions. Additionally they provided me with all data, information and equipment necessary for my work. Furthermore I would like to thank them for organizing a store internship and for letting me participate in several training courses. They gave me so many interesting insights into the world of retail.

Finally, I would like to thank Martin Holzer for giving me useful hints and ideas and Todd Slater for proofreading this thesis.

This thesis is dedicated to my parents,
Dorothee Stark and Mubarak Mukhtar.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

In the last decade large interbranch loyalty programs became a major trend among German companies and their customers. Being the biggest German loyalty program, *PAYBACK* started out in 2000 and by 2008 about 60% of all German private households owned a *PAYBACK*-card (Gmb08). Other German loyalty programs, like *HappyDigits* launched in 2001 or *DeutschlandCard* launched in 2008, show similar track records. Loyalty programs do not only try to commit customers to their companies by rewarding loyal buying patterns but they also collect a huge amount of data about their customers. When issuing the program's card the customer is normally required to provide some personal and demographic data. Furthermore, every time the customer presents his loyalty card at a participating company's cash register to receive some bonus, data about his purchase is stored. So receipts can be assigned to single customers and their buying behaviors can be tracked, which opens up new possibilities for highly effective and selective marketing concepts and for very close customer relationships.
This costly collection of customer data needs to be analyzed systematically to be of any value for the companies. The more information a company can gain about its customers and their purchasing behaviors the better it can respond to their needs and the closer to the market it actually is.
A possible data analysis obviously suggests itself: The extraction of customer profiles. One can try many ways to accomplish this, but we chose graph clustering.

Our Contribution

The content of this work is a case study. We want to find out if graph clustering is a data mining technique that is suitable to extract meaningful customer profiles in *PAYBACK*-card data provided by the German retail chain *dm-drogerie markt*, in the following abbreviated as *dm*. In our approach, we model the data as a network of customers and apply graph clustering methods to find groups of customers with similar buying behaviors. We begin by designing, implementing and analyzing different potential graph models that reflect similarity in shopping behavior. Having found the most appropriate graph model that yields reasonable clusterings, we try to identify stable customer profiles. Such profiles consist of customer clusters that verifiably occur in several graph clusterings of different points in time. We apply several techniques to identify and confirm the stability of those customer profiles, e.g., we require their clusters to have a similar customer base and similar buying patterns. Finally, we corroborate our results by trying to detect these stable customer profiles in the *PAYBACK*-card data of three totally different stores and by comparing those customer profiles on store-level. By verifiably identifying the same stable customer profiles in three different stores and by considering the period of an entire year, we can present very substantiated *dm*-customer profiles and a reliable and precise technique.

Previous and Related Work

Data mining is a huge and established field with many applications and techniques. But in this work we concentrate on graph clustering and so we refrain from discussing standard data mining techniques and refer the reader to (HK06) and (Fer03) and to references therein for an introduction to general data mining concepts and an overview of state-of-the-art methods.

Graph clustering techniques, similar to ours, were used for other purposes. Hopcroft et al. (HKKS04) try to extract natural communities in the citation graph of the NEC CiteSeer[1] scientific literature database. In the *citation graph*, vertices represent papers in the CiteSeer database and a directed edge from paper A to paper B represents the citation of paper B in paper A. A natural community is defined as follows: Several graph clusterings of a NEC CiteSeer database snapshot are generated by randomly removing a small set of papers before each clustering run. Then the single clusters of each clustering are compared to the single clusters of the first generated clustering with the matching papers they contain. A cluster of the first generated clustering is called a natural community if this cluster has a sufficiently high (paper set)-*match* value (see Section 3.3.2) to a sufficient amount of other clusters in other clusterings. Natural communities are supposed to represent 'the true hidden structure of the data'. Because if such a cluster appears verifiably in a certain amount of these generated clusterings, it can not be a coincidence.

In (Gla08) a *time-expanded* graph model is designed. Based on this model, graphs of the e-mail network of the computing faculty of the *Universität Karlsruhe (TH)* are generated, clustered and evaluated. A *time-expanded* graph connects vertices of graphs at different points in time via inter-time edges, if those vertices "resemble" each other. The *time-expanded* graph is described in Section 3.3.1. The idea behind this is that in a clustered *time-expanded* graph one can track the temporal evolution of single clusters. Furthermore, clusters can be smoothened over time and certain outliers can be diminished.

[1]CiteSeer is a scientific literature digital library.

Outline

The outline of this work is as follows. In the next section, we describe the used source data. In Chapter 2, the graph theoretical concepts used throughout this work and necessary preliminaries are introduced. Furthermore, different considered graph models that are analyzed in the experimental study are presented. Different quality measurements for clusterings as well as different comparison measurements for single clusters, and the used clustering method are introduced in Chapter 3. Chapter 4 contains the experiments, in which we test our methods on the different graph models of Chapter 2 and extract and confirm stable customer profiles in the data of different stores. In Chapter 5 we describe a universal way to import our graph clustering results into many other applications. Finally, Chapter 6 summarizes this work and gives an outlook and a brief note on two promising applications.

1.2 Data Source

We used the sales data of the German retail chain *dm* from mid 2004 to mid 2008. The data was gathered via the customer loyalty program *PAYBACK*. This data is based on a massive customer base and carefully collected, thus it is highly reliable; however, we should keep two flaws of the data in mind: only about 50 percent of *dm* customers have a *PAYBACK* card and it is possible to lend it or use it infrequently. Being aware of this, we can now base our study on *PAYBACK*. If a customer presents his *PAYBACK* card during the payment process at the cash register, customer ID, date, store ID and bought items are recorded. The database contains customer master data, article master data, store master data and receipt data. The customer master data includes information about the customers like age, gender, postal code and favorite store. In the article master data, each of the [redacted] *dm* articles is classified into one of [redacted] assortments[2], into one of [redacted] sub-assortments[1] and into much more, but we will only operate on article- or sub-assortment-level. Additionally the articles are divided into [redacted] different brands. The store master data contains postal code, location, assortment size and sales area of every store. In the receipt data each record describes when and where each customer bought particular items and how many of them. This is by far the largest set of data.
Because the data amount of all 1012 German *dm* stores is huge, we made some temporal and local restrictions. That is we restricted the receipt data to single months or quarters of a year and to single stores. Just limiting the data to a single month of 2006 means that we have to deal with approximately [redacted] receipts. In 2008 a single month even yielded over [redacted] receipts. Further restricting the data to a single store (and a

[2] Warensortiment (assortment)
[1] Warenbereich (sub-assortment)

Store	Store Label	City	Postal Code	Location Label	Sales Area Label
22	S-SILLEN-BUCH	Stuttgart	70619	[redacted]	400 to 499 sqm
242	M-OEZ	München	80993	[redacted]	500 to 599 sqm
518	KA-KÄPPELE	Karlsruhe	76131	[redacted]	over 600 sqm

Table 1.1: The considered stores.

single month) leaves only up to [redacted] receipts.
To remain representative, we selected three different stores in Karlsruhe, Stuttgart and Munich. As shown in Table 1.1, they all differ in location and sales area.

Chapter 2

Graph Modeling

2.1 Very Brief Introduction to Graph Theory

Throughout this thesis, we will use the notation of (BE05). We refer the reader to (Die00) for a more detailed introduction to graph theory.
Let $G = (V, E, \omega)$ be an undirected, connected, and simple *graph* with $n := |V|$ *vertices* and $m := |E|$ *edges*. The *edge weight* is defined as $\omega : E \rightarrow \mathbb{R}^+$. We denote the sum of all edge weights with $\mathcal{W}$. For a vertex v, we define the *vertex weight* $\omega(v)$ as the sum of the weights of its incident edges. The *degree* $deg(v)$ of a vertex is the number of edges incident to the vertex. The *adjacency matrix* $\mathcal{A}$ of graph G is a symmetric $n \times n$ matrix where entry a_{ij} is the edge weight from vertex i to vertex j or zero if $\{v_i, v_j\} \notin E$, and the diagonal entry a_{ii} is zero.
Let $C = \{C_1, .., C_k\}$ be a partition of V. We call C a *clustering* of G and the elements C_i *clusters*. Let $E(C_i) := \{\{v, w\} \in E : v, w \in C_i\}$. Then $E(C) := \bigcup_{i=1}^{k} E(C_i)$ is the set of *intra-cluster edges* and $E \backslash E(C)$ the set of *inter-cluster edges*. $m(C)$ denotes the number of intra-cluster edges and $\overline{m(C)}$ the number of inter-cluster edges.
Figure 2.1 depicts an example of a clustered graph G_1 with $n = 6$, $m = 7$ and $\mathcal{W} = 23$. Vertex b has the vertex weight 8 and a degree of 3. The adjacency matrix of G_1 is shown in (2.1). The clustering $C = \{C_1, C_2\}$ of Figure 2.1 consists of two clusters C_1 and C_2 with $C_1 = \{\text{a}, \text{b}, \text{c}\}$ and $C_2 = \{\text{d}, \text{e}, \text{f}\}$. The number of intra-cluster edges $m(C)$ is 6 while the number of inter-cluster edges $\overline{m(C)}$ is 1.

$$\begin{pmatrix} 0 & 2 & 3 & 0 & 0 & 0 \\ 2 & 0 & 5 & 1 & 0 & 0 \\ 3 & 5 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 6 & 2 \\ 0 & 0 & 0 & 6 & 0 & 4 \\ 0 & 0 & 0 & 2 & 4 & 0 \end{pmatrix} \tag{2.1}$$

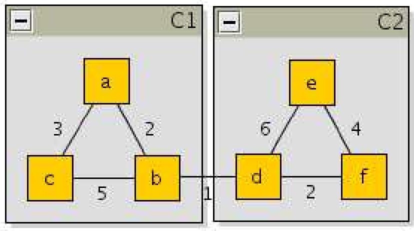


Figure 2.1: A clustering of graph G_1 consisting of two clusters C_1 and C_2.

2.2 Similarity Coefficients

Similarity Coefficients are designed to quantify the likeness between two vectors or two sets of objects by specifying one numeric value. This value increases as the similarity of the vectors increases. We need similarity coefficients for the edge weights of our graphs. Below, we describe a basic representation of our data, some characteristics of similarity coefficients and five commonly used similarity coefficients. We assume that the single objects or vector components are elements of $\mathbb{N}_0$.

2.2.1 The Basic Representation

The calculation of similarities between customers requires the purchases to be represented as collections of items, analog to (JF87). The frequency with which a customer has bought a certain item can be taken into account in the form of a weight. A matrix representation of the collection of purchases, as in Table 2.1, is appropriate. The items are associated with the rows and the purchases of the customers are associated with the columns of the matrix.
Table 2.2 shows a concrete example. This example points out two important aspects about weights. First, if we compare the weights within a column to each other, we can see which items are important for the representation of the customers‘ buying behavior. For example, $Customer_1$ buys mainly cat products and no baby products. Second, comparing the weights within a row might show some similarities between customers. $Customer_1$ and $Customer_3$ both buy mainly cat products and $Customer_2$ and $Customer_4$ are more into baby products.
However, in some cases the weight of an item is not very meaningful. $Customer_5$ buys cat and baby products and almost every customer buys plastic bags. So, $Customer_5$ is not easy to describe and plastic bags are not very helpful to discriminate between customers. Finally, the last items are bought very infrequently compared to the cat food, but they are probably more interesting and distinguishing items than cat food. That is why we are more interested in whether or not a customer has bought a certain item than in how frequently he has bought it.
In this context, two geometric representations of our data are feasible. In Figure 2.2

	Purchases / Customers			
Items	C_1	C_2	...	C_m
A_1	W_{11}	W_{12}		
A_2	W_{21}	W_{22}		
.				
.				
A_n				W_{nm}

Table 2.1: A matrix representation of the purchases of the customers.

	Purchases / Customers				
Items	C_1	C_2	C_3	C_4	C_5
dm-BrandB1 KatzeSeelachs	3	0	1	0	1
BrandA1 Milchbrei	0	7	0	5	2
Plastiktuete	4	3	1	2	0
dm-BrandB1 KatzeKalb+Geflügel	10	0	7	0	1
BrandA1 Frühstück-Brei	0	1	0	1	0
BrandC1 Gardinenwaschmittel	0	1	1	0	1
BrandD Scheuermilch	0	1	1	0	1

Table 2.2: An example of a matrix representation.

customers are points in space and items are represented by coordinates. For instance, the vector of $Customer_1$ lies on the axis of $Item_1$, at (3,0) in space. In other words, $Customer_1$ bought $Item_1$ three times and never bought $Item_2$. Another possible geometric representation would have customers as coordinates and items as points in space.

2.2.2 Characteristics of Similarity Coefficients

Jones, et al. (JF87), proposed some attributes, with which similarity coefficients can be compared to each other. They are illustrated in Figures 2.3 to 2.6. Let sim($\vec{A}$,$\vec{B}$) be the similarity value of the vectors $\vec{A}$ and $\vec{B}$:

Angle monotonicity - Let $\vec{B}$ and $\vec{C}$ be two vectors of the same length. If the angle between $\vec{A}$ and $\vec{B}$ is smaller than the angle between $\vec{A}$ and $\vec{C}$, than sim($\vec{A}$,$\vec{B}$) must be higher than sim($\vec{A}$,$\vec{C}$).

Radial monotonicity - Let $\vec{B}$ and $\vec{C}$ be two vectors of the same direction. If $\vec{B}$ is longer than $\vec{C}$, than sim($\vec{A}$,$\vec{B}$) must be higher than sim($\vec{A}$,$\vec{C}$).

Component-wise monotonicity - Let us increment any component of $\vec{B}$ and call it $\vec{B}_2$. sim($\vec{A}$,$\vec{B}_2$) must be higher than sim($\vec{A}$,$\vec{B}$).

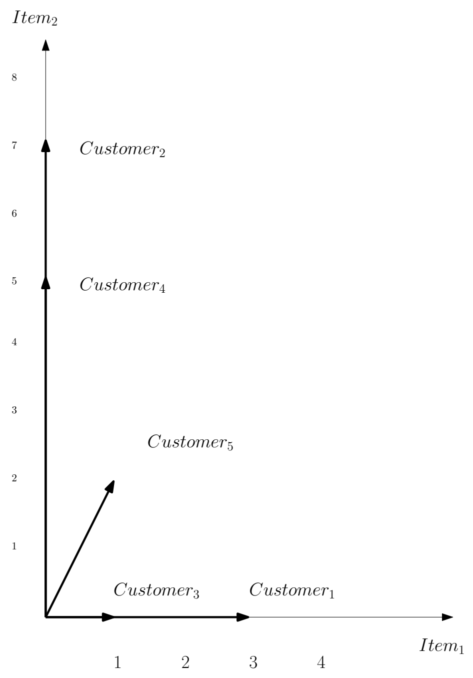


Figure 2.2: A geometric representation.

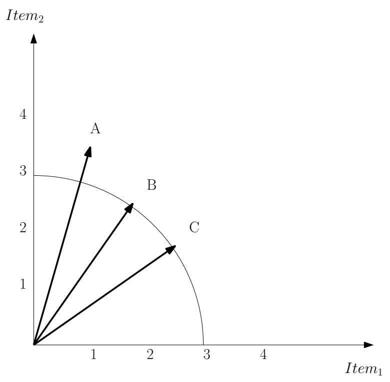


Figure 2.3: An illustration of the *angle monotonicity* characteristic (JF87). It means that sim($\vec{A}$,$\vec{B}$) > sim($\vec{A}$,$\vec{C}$).

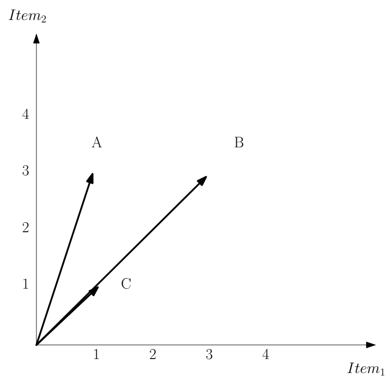


Figure 2.4: An illustration of the *radial monotonicity* characteristic (JF87). It means that sim($\vec{A}$,$\vec{B}$) > sim($\vec{A}$,$\vec{C}$).

Unbounded single-component influence - The increment of one component can result in an infinitely high similarity value.

Boundedness of similarity values - The similarity value has an upper limit, mostly 1.

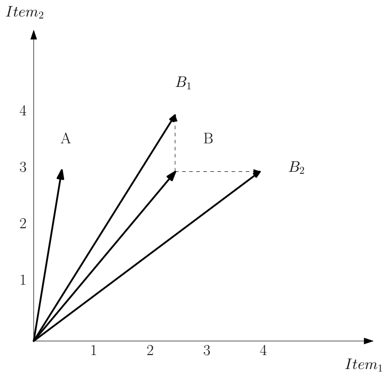


Figure 2.5: An illustration of the *component-wise monotonicity* characteristic (JF87). It means that sim($\vec{A}$,$\vec{B}_2$) > sim($\vec{A}$,$\vec{B}$).

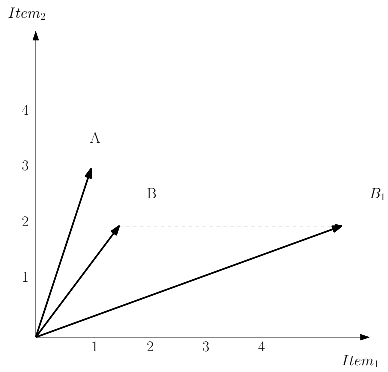


Figure 2.6: An illustration of the *unbounded single-component influence* (JF87). It means that sim($\vec{A}$,$\vec{B}_1$) > sim($\vec{A}$,$\vec{B}$).

2.2.3 Simple Matching Coefficient

The *simple matching coefficient* or *inner product measure* is the simplest of all similarity coefficients. Equation 2.2 shows the *simple matching coefficient* between the customers C_1 and C_2. In the boolean case, as shown in Equation 2.3 (vR79), the *simple matching coefficient* is composed of the cardinality of the intersection between the set of items bought by $Customer_1$ and the set of items bought by $Customer_2$.

$$\sum_{i=1}^{n} W_{iC_1} \cdot W_{iC_2} \tag{2.2}$$

$$|X \cap Y| \tag{2.3}$$

Figure 2.7 illustrates the contours of equal similarity for the reference vector $\vec{R}(0.5, 1.0)$. For two different vectors $\vec{A}$ and $\vec{B}$ that point to the same contour, sim($\vec{R}, \vec{A}$) is equal to sim($\vec{R}, \vec{B}$). A disadvantage of this coefficient is, that its similarity values do not have an upper limit. The lack of a perfect similarity value makes ranking very difficult. Single-components should not have a dominating influence either. As mentioned in Section 2.2.1, frequently bought items like plastic bags are often less interesting than rarely bought ones. In our case, the more is not the better. Furthermore, our focus lies on the combination of items a customer has bought. So the angle between two vectors should have a much bigger influence on the similarity value than the length of a vector or the increment of a single component.

But despite its many disadvantages, the *simple matching coefficient* is (normalized) often part of other similarity coefficients.

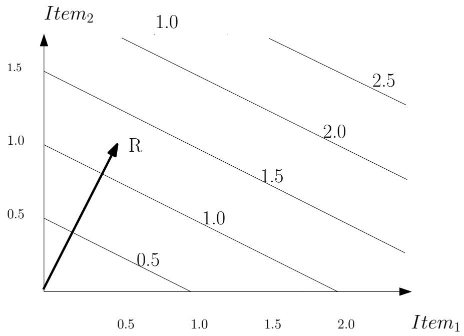


Figure 2.7: Contours of equal similarity of the simple matching coefficient. (JF87)

2.2.4 Cosine Coefficient

The *cosine coefficient*, as shown in Equation 2.4, consists of the *simple matching coefficient*, which serves as the comparison operation, divided by the product of the Euclidean lengths of both vectors. Notice that the denominator of this formula normalizes both vectors to the Euclidean lengths of one. Because of this normalization, only the directions and not the lengths of both vectors have an impact on their similarity value. Thus, the *cosine coefficient* is angle monotone and not radially monotone.
An outline of the regions of equal similarity is given in Figure 2.8. In 2-space, those regions are wedge-shaped.
Another advantage of the *cosine coefficient* is the boundedness of similarity values. As illustrated by Figure 2.8, the maximal similarity value of one lies in the wedge of the contemplated vector.

$$\frac{\sum_{i=1}^{n} W_{iC_1} \cdot W_{iC_2}}{\sqrt{\sum_{k=1}^{n}(W_{kC_1})^2} \cdot \sqrt{\sum_{j=1}^{n}(W_{jC_2})^2}} \tag{2.4}$$

$$\frac{|X \cap Y|}{\sqrt{|X| \times |Y|}} \tag{2.5}$$

2.2.5 Overlap Coefficient

The formulas of the *overlap coefficient* are shown in Equations 2.6 and 2.7. The numerator of this coefficient is a different comparison operation. Unlike the other discussed similarity coefficients, it does not contain the *simple matching coefficient*. The denominator acts as a normalization operation.

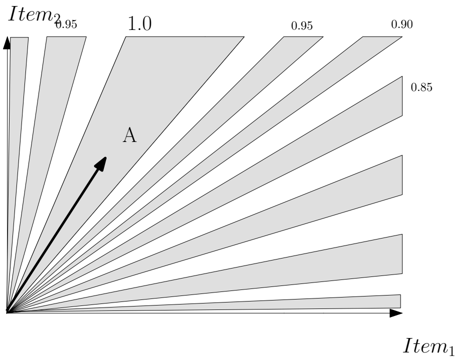


Figure 2.8: Regions of approximately equal similarity of the cosine coefficient (JF87). (Of course cosine is continuous.)

$$\frac{\sum_{i=1}^{n} min(W_{iC_1}, W_{iC_2})}{\min(\sum_{k=1}^{n} W_{kC_1}, \sum_{j=1}^{n} W_{jC_2})} \tag{2.6}$$

$$\frac{|X \cap Y|}{\min(|X|, |Y|)} \tag{2.7}$$

The *overlap coefficient* has a very unique behavior. Figure 2.9 illustrates the regions and contours of equal similarity. The two black boxes represent regions of maximal similarity. For every vector $\vec{B}$, that falls into one of the black boxes, sim($\vec{A}$,$\vec{B}$) is equal to one. So, if all components of Vector $\vec{B}$ are smaller (/bigger) than all components of Vector $\vec{A}$, then sim($\vec{A}$,$\vec{B}$) is maximal.
Figure 2.10 demonstrates this major disadvantage of the *overlap coefficient*. Vectors $\vec{B}$ and $\vec{C}$ both fall into the upper black box of vector $\vec{A}$. Their similarity values are equal, although the angle between $\vec{A}$ and $\vec{B}$ is much smaller than the angle between $\vec{A}$ and $\vec{C}$ and the first components of $\vec{A}$ and $\vec{C}$ differ a lot more than the first components of $\vec{A}$ and $\vec{B}$. Table 2.3 summarize all characteristics of this coefficient.

2.2.6 Jaccard Coefficient

The formula of the *jaccard coefficient* for binary vectors is shown in Equation 2.8. It is composed of the cardinality of the intersection between the set of items bought by $Customer_1$ and the set of items bought by $Customer_2$ divided by the cardinality of the union of both sets. In (Fer03) the weighted version of the *jaccard coefficient* is defined like in Equation 2.9. The disadvantage of this version is that the denominator can become 0 even if all vector components are nonnegative. This can yield discontinuities in the similarity of vectors. In 2-space there are places where tiny changes of a vector can yield

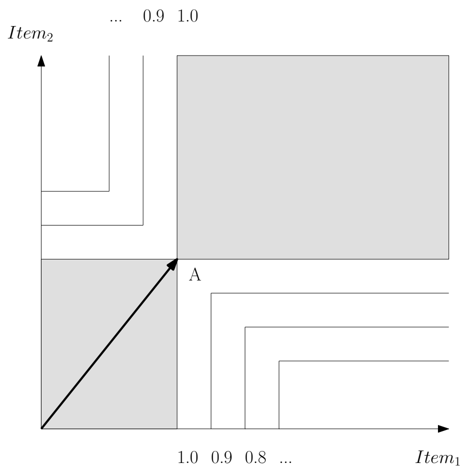


Figure 2.9: Regions and contours of equal similarity of the overlap coefficient. (JF87)

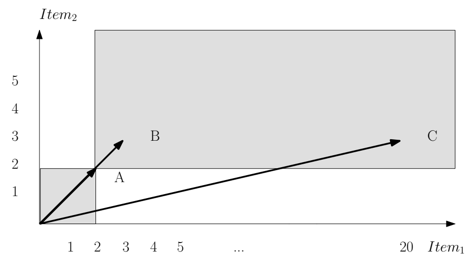


Figure 2.10: Overlap coefficient example.

arbitrarily huge changes in the similarity to another vector. If (a,b) is the reference vector, these discontinuities are given for all points on the following straight line:

$$y = \frac{1-a}{b-1}x + \frac{a+b}{b-1} \ .$$

For instance, if the reference vector is (4,2), this straight line is $y = -3x + 6$. The vector (0,6.2) has a similarity of -62 to the reference vector, while the vector (0,6.1) has a similarity of -122 to the reference vector.
In (Tan57) Tanimoto introduces the *tanimoto coefficient*, which is shown in Equation 2.10. The formula differs from that of Equation 2.9 in that its denominator always remains positive by not taking the sums of vector components but the sums of squared vector components. This formula extends the *cosine coefficient*. For binary vectors it is equal to the *jaccard coefficient*, so it can be seen as a legitimate weighted version of the *jaccard coefficient*. The *tanimoto coefficient* is very popular in biological taxonomy and information retrieval (HK06). Its characteristics are similar to that of the *cosine coefficient* as can be seen in Table 2.3.

$$\frac{|X \cap Y|}{|X \cup Y|} \qquad (\textit{Boolean Jaccard Coefficient}) \tag{2.8}$$

$$\frac{\sum_{i=1}^{n} W_{iC_1} \cdot W_{iC_2}}{\sum_{i=1}^{n} W_{iC_1} + \sum_{i=1}^{n} W_{iC_2} - \sum_{i=1}^{n} W_{iC_1} \cdot W_{iC_2}} \qquad (\textit{Weighted Jaccard Coefficient}) \tag{2.9}$$

$$\frac{\sum_{i=1}^{n} W_{iC_1} \cdot W_{iC_2}}{\sum_{i=1}^{n} W_{iC_1}^2 + \sum_{i=1}^{n} W_{iC_2}^2 - \sum_{i=1}^{n} W_{iC_1} \cdot W_{iC_2}} \qquad (\textit{Tanimoto Coefficient}) \tag{2.10}$$

2.2.7 Summary

An overview of the similarity coefficient characteristics is shown in Table 2.3. We need similarity values to have an upper limit, because we want to rank similarity values. Further, we do not like to have a big influence of single-components, because we want to value frequently and infrequently bought items alike. The same argument speaks against radial monotonicity and component-wise monotonicity. This means that the *simple matching coefficient* is not very suitable for our purposes. But a similarity coefficient should be angle monotone, because this means that the smaller the angle between two vectors, the more similar is each component of one vector to the corresponding component of the other vector. The *overlap coefficient* is not angle monotone.
Comparing the introduced characteristics of the *cosine coefficient* to that of Tanimoto's

		Similarity Coefficients		
Characteristics	Simple	Cosine	Overlap	Jaccard/Tanimoto
Angle monotonicity	+	+ (regardless of vector length)	-	+
Radial monotonicity	+	- (radial-constant)	-	-
Component-wise monotonicity	+	- (only if angle changed)	-	-
Unbounded single-component influence	+	-	-	-
Boundedness of similarity values	-	+ ([0,1])	+ ([0,1])	+ ([0,1])

Table 2.3: A comparison of similarity coefficient characteristics.

version of the *jaccard coefficient* results in a tie. But an unwanted feature of the *cosine coefficient* is that in the area where two vectors have a very small angle (close to 0°) a change of the angle yields an extremely low change in the similarity value. But particularly in this area, changes are very interesting to observe. Thus, we decide to use Tanimoto's version of the *jaccard coefficient* as our main similarity coefficient.

2.3 Thresholds

When generating a graph the user can adjust among others the three parameters *minimum edge weight*, *absolute edge weight* and *single weight*. The *minimum edge weight* is a threshold that allows only customers with a minimum purchasing similarity (edge weight) to at least one other customer to be represented as vertices in the graph. Thus, the lower the *minimum edge weight* the more vertices are in the graph, until all customers of the underlying data are represented in the graph (*minimum edge weight*=0). The purpose of this parameter is that we only want to take customers with a minimum purchasing similarity into consideration, because otherwise they can not help in defining any customer cluster. Analog to this, the *absolute edge weight* is an absolute threshold that allows only customers to be represented in the graph that bought a minimum number of items which at least one other customer bought. The following example illustrates the difference to the *minimum edge weight*: We set the *minimum edge weight* to 0.9, which is an extremely high value,

and we set the *absolute edge weight* to 2, which is an extremely low value. Two customers are described by the customer vectors[1] (1,0,0,0,0,0) and (1,0,0,0,0,0). Their (Jaccard) edge weight is 1.0, because they have the exact same vectors. If we only consider the *minimum edge weight* of 0.9 during the graph generation, both customers are in the graph. But if we only consider the *absolute edge weight* of 2, both customers are not part of the graph, because they both bought just one item.
The *single weight* reflects the minimum amount of items a customer is required to have bought. The customer is only represented in the graph if he bought a certain amount of items. Applying a high *single weight* yields only major customers and excludes customers that only bought small amounts. To illustrate this, we use a second example. We set the *single weight* to 3, the *absolute edge weight* to 1 and the *minimum edge weight* to 0.01. Two customers are described by (1,0,1,0,1,0) and (0,1,0,1,0,1). If we just consider the *single weight* during graph generation, both customers are part of the graph, because they both bought 3 items. But if we just consider the *absolute edge weight* or the *minimum edge weight*, these customers are not in the graph.
In our tool, the user can adjust all three parameters. But we removed the *single weight* during our experiments, because it turned out to be irrelevant. Only high thresholds of *single weight* (≥ 10) had an impact, but we deem such values too restrictive in our model as we do not only want to consider heavy buyers. We fixed the *absolute edge weight* to 5, because it turned out to be dominated by the the *minimum edge weight*. In our experiments the *minimum edge weight* ranged from about 0.2 to 0.03. In our view of the relevance of the data, it is the most interesting of the three parameters, because it does not expect absolute values of the customers.

2.4 Graph Types

A variety of questions can be posed to the given source data, introduced in Section 1.2. The appropriate model helps to find answers to these questions. In this work, we make use of the graph model. Remember that in the graph model entities are represented by vertices, relations are represented by edges, and natural communities or groups of similar entities are expressed by clusters.
So, for instance we might be interested in the customers‘ main reasons for shopping at *dm*. In other words, we want to know which items are essential and should stay in the assortment to keep the customers. To reveal these essential items, we need to find groups of similar shopping carts. In this case, a graph model, whose vertices represent receipts and whose edges represent common items, would be appropriate. Clusters in this graph type are the sought-after groups of similar shopping carts.
Another example is a graph whose vertices stand for items and whose edges stand for customers (or just receipts). The clusters of this graph represent groups of similar items. This information is useful for the shop layout. It would make sense to place items that are

[1] A summary of the customer's receipts.

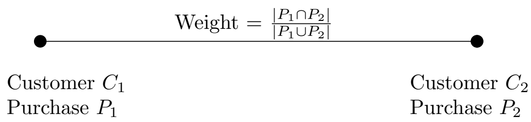


Figure 2.11: The structure of *KAK*-Graphs.

bought by a similar buyership in different areas of the shop to prolong the customers stay in the shop.
But what graph type suits best, if we want to reveal customer profiles and behaviors?
We now introduce three different graph types that will be analyzed in this work. Note that those graph types were part of a long progress and that one graph type evolved from another graph type. But now we want to already provide the formalisms.

2.4.1 *KAK*-Graphs

As mentioned before, we intend to find customer profiles. That is, we want to discover groups of customers with similar buying patterns. So, an obvious graph model would consist of customers as vertices and common purchases as edges. The basic structure of this graph type is illustrated in Figure 2.11. We call this graph type *KAK*-Graph, because of the customer–item–customer (in German: *K*unde–*A*rtikel–*K*unde) relation it represents. The similarity value of two customers, represented by the corresponding edge weight, is calculated using the Jaccard similarity coefficient. As described in Section 2.2.6, the Jaccard similarity of two customers is determined by dividing the number of the intersection of their bought items by the number of the union of their bought items.
The example of Table 2.2 and Figure 2.2 can easily be transformed into a *KAK*-Graph, as demonstrated by Figure 2.12. Note that we only want to take customers with a minimum purchasing similarity into consideration, because otherwise they can not belong to any customer cluster. So in this example, we determine a minimum edge weight of 0.1. This means that only customers with at least one similarity over 10 percent are kept. In Figure 2.12, only the solid edges and the vertices incident to those edges remain in the graph. A clustering algorithm splits a graph into dense groups of vertices that are sparsely connected (see Section 3.2). Such an algorithm would most likely group Customer C_1 and Customer C_3 in one cluster and Customers C_2, C_4 and C_5 in another cluster, as shown in Figure 2.13. As mentioned in Section 2.2.1, Customer C_1 and Customer C_3 mainly buy cat products and Customer C_2 and Customer C_4 mainly buy baby products. So in this case, transforming our source data into a *KAK*-Graph and applying a clustering algorithm reveals a *Cat* customer profile and a *Baby* customer profile.

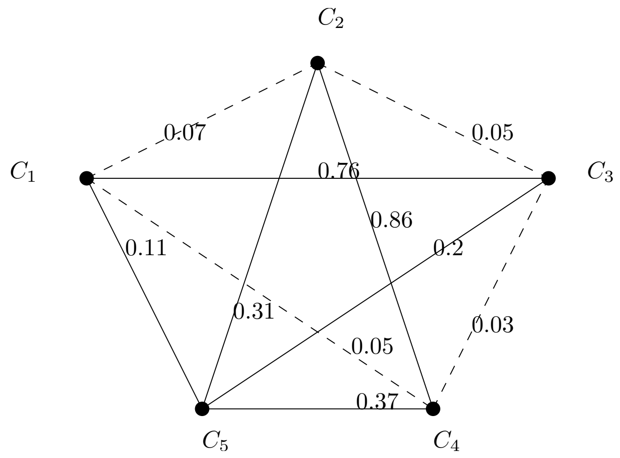


Figure 2.12: Example of a KAK-Graph.

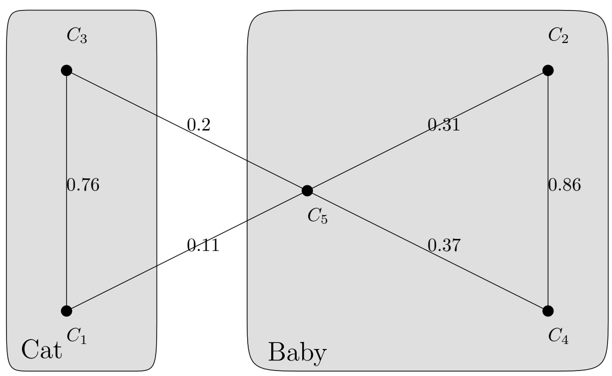


Figure 2.13: Example of a clustered KAK-Graph.

	Purchases / Customers				
Marke/Warenbereich	C_1	C_2	C_3	C_4	C_5
dm-BrandB1 Katze	1	0	1	0	1
BrandA1 Babynahrung	0	1	0	1	1
? Sontiges/Te	1	1	1	1	0
BrandC1 WPR	0	1	1	0	1
BrandD WPR	0	1	1	0	1

Table 2.4: Binary matrix representation.

2.4.2 *KMK*-Graphs

Another possibility to find customer profiles is provided by the so-called *KMK*-Graph. This graph type represents a customer–sub-assortment*brand–customer (in German: *K*unde–Warenbereich**M*arke–*K*unde) relation. The structure of the *KMK*-Graph is a little different to that of the *KAK*-Graph. Vertices still stand for customers. However, in this graph type edges stand for purchases in the same sub-assortment and of the same brand. So, all items in the same sub-assortment and of the same brand are merged. As mentioned in Section 2.2.1, one of our objectives is to valorize infrequently bought items. For this purpose, we now use binary customer vectors. In these vectors, '1' means that the customer has bought at least one item in this sub-assortment and of this brand, and '0' means the customer has not bought any item of this sub-assortment*brand combination. The geometric representation of a *KMK*-Graph is illustrated by Figure 2.14. Note that the article master data of *dm* contain [redacted] sub-assortment*brand combinations and so the number of components of each customer vector is equal to this number.
The matrix representation of the example shown in Table 2.2 now shrinks to the matrix representation in Table 2.4. Customer C_1 bought thirteen items in the *Cat* sub-assortment and of the brand *dm-BrandB1* . In the new representation, these thirteen items just count as one. So, we see that frequently bought items are combined in one sub-assortment*brand combination while very infrequently bought items like *BrandC1 Gardinenwaschmittel* still count the same. This is due to the fact that in the *Jaccard* calculation (2.10) the high influence of many items inside the *Cat* cluster does no longer overshadow the infrequently bought items.
In the resulting *KMK*-Graph (2.15), Equation 2.8 is used to calculate the edge weights. If we determine a minimum edge weight of 0.3 and apply a clustering algorithm on our *KMK*-Graph, we again receive two clusters, as shown in Figure 2.16. But this time, customers C_2, C_3 and C_5 are grouped together. Unlike the other customers, those three have bought items in the sub-assortment *WPR*[3]. In other words, we reveal a new *WPR* customer profile using the *KMK*-Graph type.

[3]This sub-assortment contains cleaning supplies.

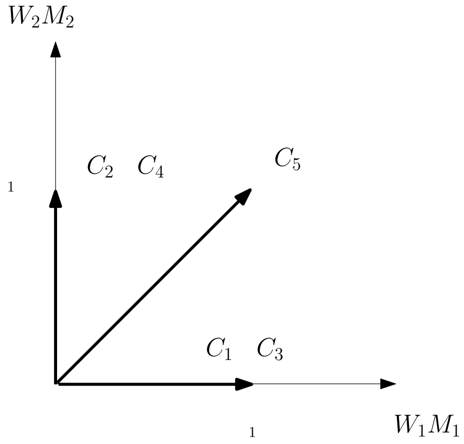


Figure 2.14: Geometric representation of a *KMK*-Graph.

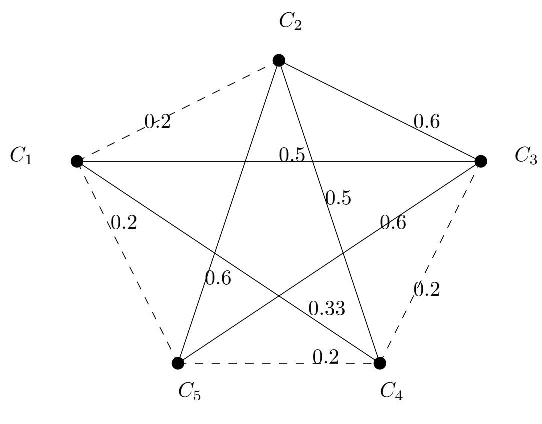


Figure 2.15: Example of a *KMK*-Graph.

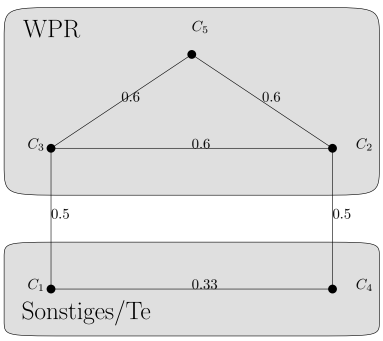


Figure 2.16: Example of a clustered *KMK*-Graph.

2.4.3 *KhK*-Graphs

In the *KhK*-Graph type, a customer is characterized by its most frequently bought brand of every sub-assortment. *KhK* stands for the customer–most-frequent-brand*sub-assortment–customer (in German: *K*unde–*h*äufigste-Marke*Warenbereich–*K*unde) relation. So, the more favorite brands two customers have in common, the higher their similarity value.
In this case, two different matrix representations are feasible. The first one consists of string values. Every row of this matrix represents one sub-assortment. For every sub-assortment a customer gets the string value of his favorite brand assigned. Compared to the *KMK*-Graph, in this representation a customer vector consists of many fewer components. The second feasible representation has as many components as the representation of the *KMK*-Graph and again, we use binary values. But here, '1' stands for the customers' favorite brand. This means that every customer has at most one '1' per sub-assortment. The advantage of the second representation is that it enables the calculation of similarity values. Note that it is easy to switch between both representations. Depending on whether we require less dimensions or we need to calculate similarity values, we can pick the appropriate representation. To distinguish between the two binary representations of the *KMK*- and the *KhK*-Graph, we call the latter the *categorical representation*. Compared to the *KMK*-Graph, we lose the information about whether or not a customer has bought other (not favorite) brands, but on the other hand we gain the information about the favorite brands of the customer.
To demonstrate the features of the *KhK*-Graph, we use a second example. In Table 2.5, the items are already ordered by sub-assortments. Here, in the first sub-assortment Customer C_2 bought the brand *BrandE1* most frequently. So, in the string representation for the field of this sub-assortment the value 'BrandE1' is assigned to him, as shown in 2.11. The corresponding categorical vector of customer C_2 is shown in 2.12. Table 2.6 shows the complete categorical matrix representation. The corresponding *KhK*-Graph is shown in Figure 2.17. If we once again determine a minimum edge weight of 0.3 and apply a clustering algorithm on this *KhK*-Graph, we receive two *Cleaning Supplies* clusters of different brands (*BrandI1* and *BrandJ1*)(Figure 2.18).
Now, let us compare this clustering result with the result we would get using the *KMK*-Graph type. Table 2.7 contains the binary matrix representation of this example and Figures 2.19 and 2.20 show the corresponding (clustered) *KMK*-Graph. With the same minimum edge weight of 0.3, we receive two totally different and less informative clusters. The first one is a mixture of cat products, cleaning supplies and plastic bags and the customers of the second cluster just have plastic bags in commen. In other words, in this example the resulting clusters and their expressiveness heavily depend on the choice of the graph type. This statement is even emphasized by the resulting clusters of the corresponding *KAK*-Graph (A.1). With the same edge weight we receive two *plastic bag* clusters (A.2) and with a lower edge weight, we get one *Dog* and one *Cat* cluster (A.3). But both clustering results are completely different to the clustering results of the *KMK*- and *KhK*-Graphs.

Items	Purchases / Customers C_1	C_2	C_3	C_4	C_5
dm-BrandB1 KatzeLamm	5	0	2	0	5
BrandE1 Mit Kaninchen	4	4	0	0	0
BrandF1 Geflügelcocktail	0	3	4	0	0
dm-BrandB1 Katze Trockenfutter	1	0	1	0	1
BrandG1 Wunderknochen	0	2	0	1	0
BrandH1 Rind	0	0	0	4	3
BrandI1 Allzweck-Reiniger	2	1	0	0	1
BrandJ1 Aprilfrisch	1	0	1	1	0
Plastiktuete	2	3	4	2	4

Table 2.5: Matrix representation of the *KhK*-Example.

Marke/Warenbereich	Purchases / Customers C_1	C_2	C_3	C_4	C_5
dm-BrandB1 Katze	1	0	0	0	1
BrandE1 Katze	0	1	0	0	0
BrandF1 Katze	0	0	1	0	0
BrandG1 Hund	0	1	0	0	0
BrandH1 Hund	0	0	0	1	1
BrandI1 WPR	1	1	0	0	1
BrandJ1 WPR	0	0	1	1	0
? Sontiges/Te	1	1	1	1	1

Table 2.6: Categorical matrix representation of the *KhK*-Example.

$$\begin{pmatrix} SA1 & 'BrandE1' \\ SA2 & 'BrandG1' \\ SA3 & 'BrandI1' \\ SA4 & '?' \end{pmatrix} \tag{2.11}$$

$$\begin{pmatrix} B_1^1 & SA1 & 0 \\ B_1^2 & SA1 & 1 \\ B_1^3 & SA1 & 0 \\ \hline B_2^1 & SA2 & 1 \\ B_2^2 & SA2 & 0 \\ \hline B_3^1 & SA3 & 1 \\ B_3^2 & SA3 & 0 \\ \hline B_4^1 & SA4 & 1 \end{pmatrix} \tag{2.12}$$

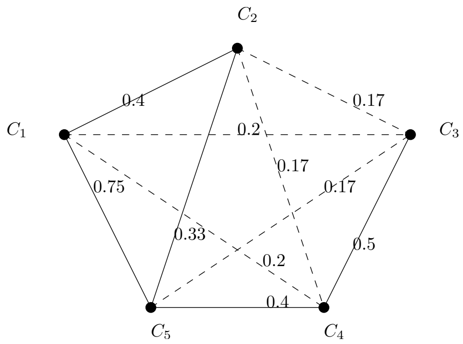


Figure 2.17: *KhK*-Graph of the *KhK*-Example.

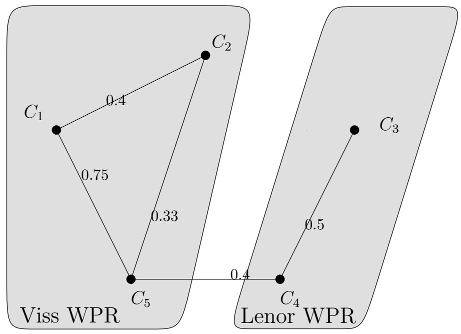


Figure 2.18: Clustered *KhK*-Graph of the *KhK*-Example.

	Purchases / Customers				
Marke/Warenbereich	C_1	C_2	C_3	C_4	C_5
dm-BrandB1 Katze	1	0	1	0	1
BrandE1 Katze	1	1	0	0	0
BrandF1 Katze	0	1	1	0	0
BrandG1 Hund	0	1	0	1	0
BrandH1 Hund	0	0	0	1	1
BrandI1 WPR	1	1	0	0	1
BrandJ1 WPR	1	0	1	1	0
? Sontiges/Te	1	1	1	1	1

Table 2.7: Binary matrix representation of the *KhK*-Example.

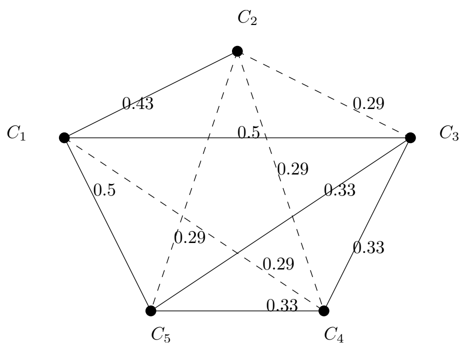


Figure 2.19: *KMK*-Graph of the *KhK*-Example.

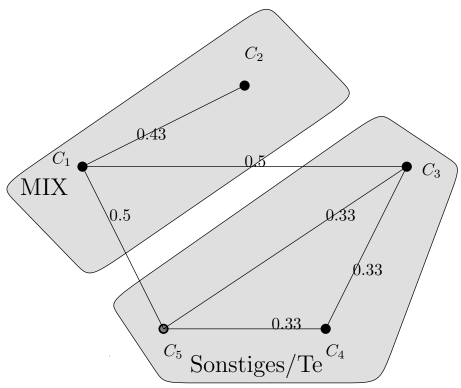


Figure 2.20: Clustered *KMK*-Graph of the *KhK*-Example.

Chapter 3

Graph Clustering

3.1 Quality Measurements for Clusterings

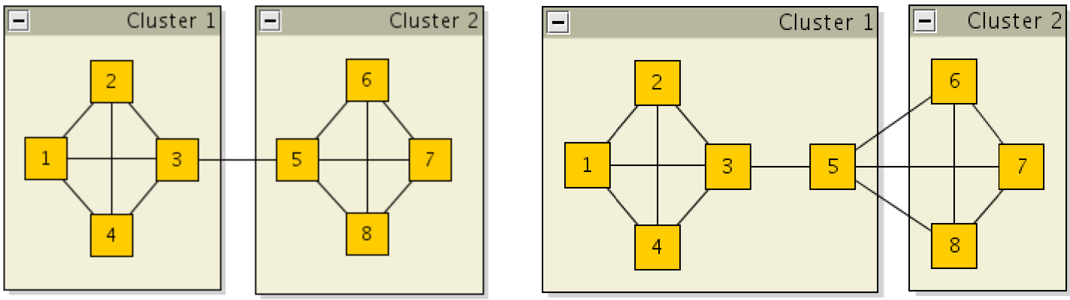
In a set of customers, we want to find customer groups that are as well-defined as possible. This means, we search for customer clusters, whose customers are a lot more similar to other customers of the same cluster than to customers of other clusters. In other words, we look for a customer clustering of high quality. To measure the quality of a clustering, we use structural indices, as described in (BGW03), (BGW07) and (GGW07).

3.1.1 Coverage

Coverage(C) is a simple quality measurement for graph clusterings. It measures the quality of a clustering C by comparing the number of intra-cluster edges to the total number of all edges (Equation 3.1). Equation 3.2 describes the weighted version of *coverage*(C). The range of *coverage*(C) is [0,1]. Basically, the higher the quality of a clustering C, the higher *coverage*(C). Figure 3.1.1 illustrates this principle. 3.1(a) and 3.1(b) are two possible clusterings of one graph. Intuitively one would group the vertices like in 3.1(a). If every edge weight is equal to one, $coverage(C) = \frac{12}{13}$ and $coverage(C') = \frac{10}{13}$. So, clustering 3.1(a) is of higher quality.
But, note that *coverage*(C) reaches its maximum, if C consists of only one cluster containing all vertices of the graph. Therefore, additional constraints on the clustering like a minimum number of clusters or a maximum cluster size are necessary.

$$coverage(C) \quad := \quad \frac{m(C)}{m} = \frac{m(C)}{m(C) + \overline{m(C)}} \tag{3.1}$$

$$coverage_\omega(C) \quad := \quad \frac{\omega(m(C))}{\omega(m)} = \frac{\omega(m(C))}{\omega(m(C)) + \omega(\overline{m(C)})} \tag{3.2}$$



(a) Clustering C (b) Clustering C'

Figure 3.1: Two possible clusterings of one graph.

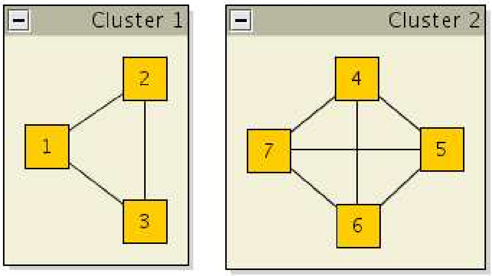


Figure 3.2: Clustering with two completely connected cliques.

3.1.2 Performance

The *performance*(C) of a graph clustering C is the fraction of 'correctly interpreted pairs of vertices' with respect to all possible pairs of vertices (Equation 3.3). A 'correctly interpreted pair of vertices' is either a pair of vertices connected by an intra-cluster edge or two vertices of two different clusters that are not connected by an inter-cluster edge.
Performance(C) reaches its maximum for graph clusterings, whose clusters are completely connected cliques without inter-cluster edges. Figure 3.2 depicts an example for such a clustering.

$$performance(C) \quad := \quad \frac{m(C) + \sum_{\{v,w\} \notin E, v \in C_i, w \in C_j, i \neq j} 1}{\frac{1}{2} n(n-1)} \tag{3.3}$$

3.1.3 Modularity

The *modularity*(C) of a clustering C is defined as *coverage*(C) minus the expected value of *coverage*(C). The expected value of *coverage*(C) is the *coverage* value of a graph clustering that results from a graph in which all expected vertex degrees remain the same but edges are placed at random. In (New04), *modularity* is described in detail. Equation 3.4 shows the formula of *modularity*(C). The weighted version of *modularity*(C) is shown in Equation 3.5. Remember from Section 2.1 that $\omega(v)$ is defined as the vertex weight of a vertex v. It is the sum of the weights of its incident edges. We denote the sum of all edge weights with $\mathcal{W}$. Weighted $modularity_\omega(C)$ is a true generalization of *modularity*(C), because weighted $modularity_{\omega=1}(C)$ with every weight ω set to 1 yields *modularity*(C).
Note the following trade-off: To maximize *coverage*(C), the number of intra-cluster edges of C should be high and the minimization of the expected value of *coverage*(C) is attained by having many clusters with small total degrees.
The range of *modularity*(C) is $[-\frac{1}{2}, 1)$. If *coverage*(C) is higher than expected, *modularity*(C) is positive and otherwise negative. *Modularity*(C) reaches its maximum for clusterings consisting of many, equally sized cliques with no inter-cluster edges. The clustering of Figure 3.2 already has a relatively high *modularity*(C) of 0.44.

$$Modularity(C) := Coverage(C) - \sum_{\mathcal{C} \in C} \left(\frac{\sum_{v \in \mathcal{C}} deg(v)}{2m} \right)^2 \tag{3.4}$$

$$modularity_\omega(C) := Coverage_\omega(C) - \sum_{\mathcal{C} \in C} \left(\frac{\sum_{v \in \mathcal{C}} \omega(v)}{2\mathcal{W}} \right)^2 \tag{3.5}$$

3.2 Clustering Methods

In this section, we introduce the *Greedy Modularity Clustering* algorithm, described in (GGW07). Despite the large variety of graph clustering algorithms, we only use the *Greedy Modularity Clustering* algorithm, because it is a well-known and good algorithm that yields sufficiently reliable clustering results. Besides we already have many degrees of freedom, so we better limit ourselves to one well working algorithm.

3.2.1 Greedy Modularity Clustering (Newman)

The *Greedy Modularity Clustering* algorithm (Algorithm 1) yields clusterings with high *modularity*(C) values. Note that the detection of a clustering with maximum modularity is $\mathcal{NP}$-complete (BDG+08). Hence, the *Greedy Modularity Clustering* algorithm uses a greedy heuristic. Tests indicate that this algorithm yields results close to the optimum

(BDG+08).
It starts with the singleton clustering and merges in every step those two clusters whose merging increases the *modularity*(C) value the most. The iteration stops, after all vertices are merged together into one big cluster. At the end, the intermediate clustering C with the highest *modularity*(C) is returned.
Using appropriate data-structures and an efficient implementation, Algorithm 1 runs in $\mathcal{O}(n^2 \log n)$ (GGW07).

Algorithmus 1 : Greedy Modularity Clustering

Input : Graph $G = (V, E)$
Output : Clustering C of graph G

$C \leftarrow$ singleton clustering of graph G;
dendrogram $\leftarrow$ store Tuple $(C, mod(C))$;
while $|C| \neq 1$ **do**
  maxvalue $= -\infty$;
  forall $\mathbf{C}_i, \mathbf{C}_j \in C$ *with* $i \neq j$ **do**
    $C' = (C \setminus \{\mathbf{C}_i, \mathbf{C}_j\}) \cup \{\mathbf{C}_i \cup \mathbf{C}_j\}$;
    $value \leftarrow mod(C') - mod(C)$;
    if value > maxvalue **then**
      candidate$_1 \leftarrow \mathbf{C}_i$;
      candidate$_2 \leftarrow \mathbf{C}_j$;
      maxvalue $\leftarrow$ value;
  $C \leftarrow (C \setminus \{\text{candidate}_1, \text{candidate}_2\} \cup \{\text{candidate}_1 \cup \text{candidate}_2\}$;
  dendrogram $\leftarrow$ store Tuple $(C, mod(C))$
Select clustering C from dendrogram with maximum modularity $mod(C)$

3.3 Time-Dependent Clusterings

A dynamic graph evolves over time. It consists of a chronological sequence of single graph states. Time-dependent clustering tries to find groups within a given dynamic graph. Note that here all data is available at the beginning of our procedure. This is an offline problem and so the considered temporal evolution completely happened in the past.
In our case, time-dependent clustering might be useful to smoothen customer clusters over time and diminish outliers, to observe the temporal evolution of customer groups, and to identify stable customer clusters. A stable cluster is a single cluster that occurs identifiably in several graph states, and thus is no incidental result. A stable customer cluster is a substantiated and very solid customer profile.
In this work, we go into two different time-dependent clustering approaches. The first one connects the single graph states to create one *time-expanded* graph and then clusters this graph. And the second one tries to identify similar clusters in different single, clustered

graph states.

3.3.1 *Time-Expanded* Graph

The *time-expanded* Graph is described in (GGWW06). Let $\mathcal{G} := \left\{(V, E, \omega) \,|\, E \subseteq \binom{V}{2}, \omega : E \to \mathbb{R}^+\right\}$ be the set of all undirected weighted graphs and $T := (t_1, t_2, .., t_d)$ be a sequence of discrete points in time. Given a graph sequence $s : T \to \mathcal{G}$, the *Time-Expanded* graph $G = (\mathcal{V}, \mathcal{E}, \widetilde{\omega})$. The vertice set $\mathcal{V}$ is a union of all vertice sets V of the single graphs in s (graph states). The edge set $\mathcal{E}$ is a union of the intra-graph edge set $\mathcal{E}_{graph}$ and the inter-time edge set $\mathcal{E}_{time}$. $\mathcal{E}_{graph}$ is defined as the set of all edges of the single graph states. This means that an intra-graph edge is incident to two different vertices of the same graph state. $\mathcal{E}_{time}$ is defined as the set of all edges that are incident to the same vertices at two different graph states. As mentioned in (Gla08), there are several possibilities to determine the number $k < d$ of neighboring graph states that can be connected via inter-time edges. We only use *time-expanded* graphs, in which just vertices of directly neighboring graph states are connected ($k = 1$). The weight of an intra-graph edge $\{(v, t), (w, t)\} \in \mathcal{E}_{graph}$ is the weight of the edge incident to v and w in the corresponding single graph state. There are again several methods to calculate the weight of the inter-time edges. We implemented three of them:

Alpha - All inter-time edge weights are equal to the same fix value α.

Simple Jaccard - The inter-time edge weights are calculated using the boolean Jaccard Similarity Coefficient of Equation 2.8. The weight of the inter-time edge $\{(v, t), (v, t')\} \in \mathcal{E}_{time}$ is the cardinality of the intersection between the set of neighbors of (v, t) and the set of neighbors of (v, t') divided by the cardinality of the union of both sets.

Complex Jaccard - Let v_t be the column vector of (v, t) in the adjacency matrix $\mathcal{A}(t)$ and $\mathrm{v}_{t'}$ be the column vector of (v, t') in the adjacency matrix $\mathcal{A}(t')$. The weight of edge $\{(v, t), (v, t')\} \in \mathcal{E}_{time}$ is calculated, using the Jaccard Similarity Coefficient of Equation 2.10 on the vectors v_t and $\mathrm{v}_{t'}$.

Figure 3.3 shows an example of a sequence of three single graph states with $T = (t_1, t_2, t_3)$. The corresponding *time-expanded* graph with k=1 and *simple jaccard* inter-time edge weights is shown in Figure 3.4. Applying a clustering algorithm on this *Time-Expanded* graph yields three stable customer clusters, as illustrated by Figure 3.5. The analog (clustered) *time-expanded* graph with *alpha* inter-time edge weights can be seen in Figure B.1 (B.2). We see that in every point in time certain customers come and go and certain edges change, but the number of clusters remains the same. Furthermore, the clustered *time-expanded* graph illustrates how the single clusters evolve over time.

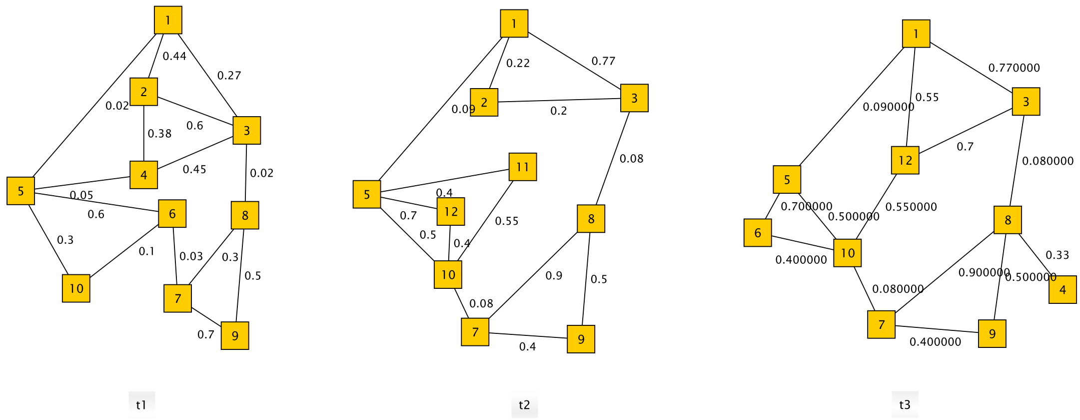


Figure 3.3: A graph sequence.

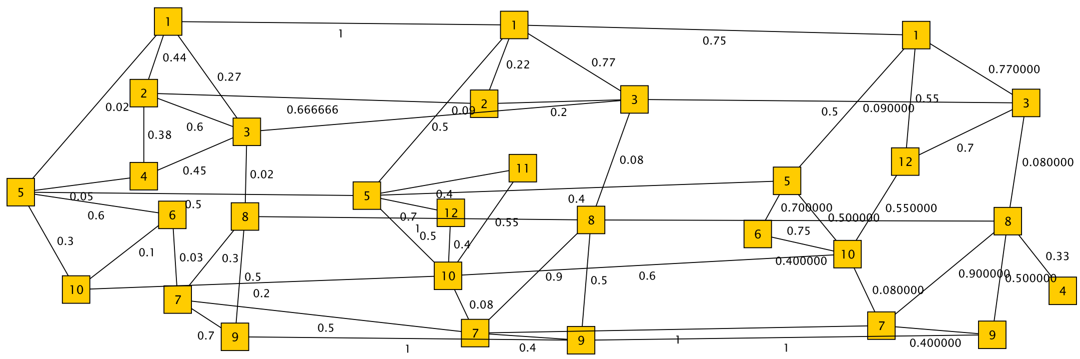


Figure 3.4: *Time-expanded* graph from Figure 3.3 with simple jaccard.

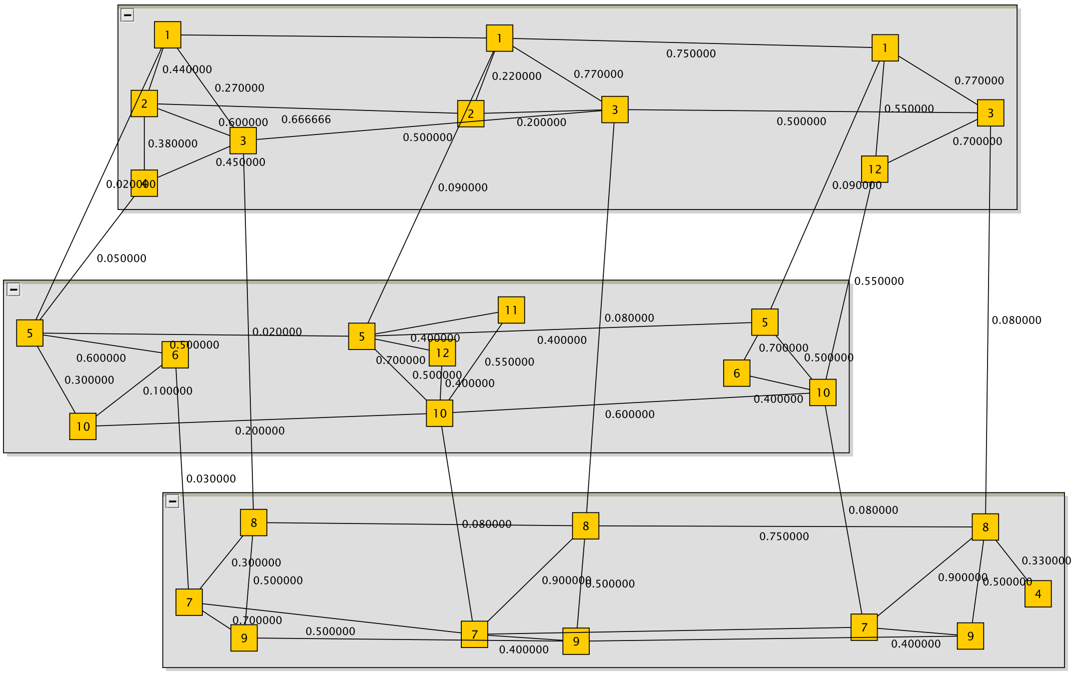


Figure 3.5: Clustered *time-expanded* graph from Figure 3.4 with simple jaccard.

3.3.2 Comparison Measurements for Clusters

Comparison measurements for clusters try to identify similar clusters in different graph clusterings. So first, every single state of a graph sequence is clustered separately. Then, we try to find similar clusters in the resulting clusterings using a comparison measurement.
In our case, detecting similar customer clusters in different clustered graph states, for example in clustered graphs of different quarters, directly leads us to stable customer clusters, and so, to potential customer profiles.
In this section, we present the two comparison measurements *bestmatch* and *bestCD* and a quality characteristic to further compare detected similar clusters.

Bestmatch

The *bestmatch* comparison measurement for clusters is proposed in (HKKS04). In this work, the authors use *bestmatch* to track evolving communities in the NEC CiteSeer database. For a given cluster C_i in one clustering $\mathcal{C}$, *bestmatch* determines the cluster C_j in another clustering $\mathcal{C}'$ that contains most of the vertices of cluster C_i. In our case, vertices are customers. So, two clusters match, if they contain similar customers.
To determine the *bestmatch* cluster of a given cluster C_i, we first have to calculate the *match* values between cluster C_i and every cluster C_j of clustering $\mathcal{C}'$. Therefore, the

number of equal vertices in both clusters is divided by the total number of vertices in the bigger cluster, as shown in Equation 3.6. *bestmatch* then calculates the cluster with the maximum *match* value to cluster C_i (3.7).
Figure 3.6 illustrates this for the example of Section 3.3.1. The *bestmatch* clusters are determined from left to right, because this direction corresponds to the temporal evolution[1]. Each big edge connects a cluster with its *bestmatch* cluster. The edge weights of the big edges, are the corresponding *match* values. Note that we receive exactly the same three stable clusters as we received clustering the *time-expanded* graph in the previous section.

$$match(C_i, C_j) = \min\left(\frac{|C_i \cap C_j|}{|C_i|}, \frac{|C_i \cap C_j|}{|C_j|}\right) \tag{3.6}$$

$$bestmatch(C_i, \mathcal{C}') = \max_{C_j \in \mathcal{C}'}\left(match\left(C_i, C_j\right)\right) \tag{3.7}$$

Centroids and *BestCD*

In our context, a *centroid* is the average customer vector of a certain cluster. Let us go back to our *KhK*-example of Section 2.4.3 and have a look at Table 2.6 and Figure 2.18. The *WPR BrandJ1* cluster consists of the customers C_3 and C_4. So, the *centroid* vector of this cluster is $(0,0,\frac{1}{2},0,\frac{1}{2},0,1,1)$. According to this, the *centroid* vector of the *WPR BrandI1* cluster with the customers C_1, C_2 and C_5 is $(\frac{2}{3},\frac{1}{3},0,\frac{1}{3},\frac{1}{3},1,0,1)$. The seventh component of the *WPR BrandJ1* *centroid* vector is 1. This means that 100% of the *WPR BrandJ1* cluster's customers bought at least one item of this sub-assortment*brand combination and *BrandJ1* is their favorite *WPR* brand. Analog to this, the sixth component of the *WPR BrandI1* *centroid* vector means that this sub-assortment*brand combination is beside the *Sonstiges/Te ?*[2] combination the most popular combination of this cluster's customers.
The *centroid* distance between two centroid vectors $\vec{c}$ and $\vec{z}$ is calculated using the Euclidean distance between both vectors, shown in Equation 3.8. In our example, the *centroid* distance of the *WPR BrandI1* and the *WPR BrandJ1* cluster is 1.55. Note that, due to binary customer vectors, the range of *centroid* vector components of *KMK*- and *KhK*-graphs is [0,1]. Let d be the number of *centroid* vector components[3]. The range of the *centroid* distance between *centroids* of *KMK*- or *KhK*-graphs is $[0,\sqrt{d}]$. The maximum *centroid* distance of $\sqrt{d}$ is reached, if both *centroids* have a maximum difference of 1 in every component. A *centroid* distance below 1 is considered to be good, because then there can be no maximum difference in any component. And even if there is no difference of 0 in any component, the difference in each component must be below $\frac{1}{\sqrt{d}}$.
Like the *bestmatch* comparison measurement, we use the *bestCD* comparison measurement

[1] We want to know: What happens to customers of one graph state in the next graph state? We are less interested in: Where do customers of one graph state come from?

[2] This sub-assortment*brand combination contains uninteresting items like plastic bags.

[3] In *KhK*-graphs d is equal to the number of sub-assortment*brand combinations, which is [redacted].

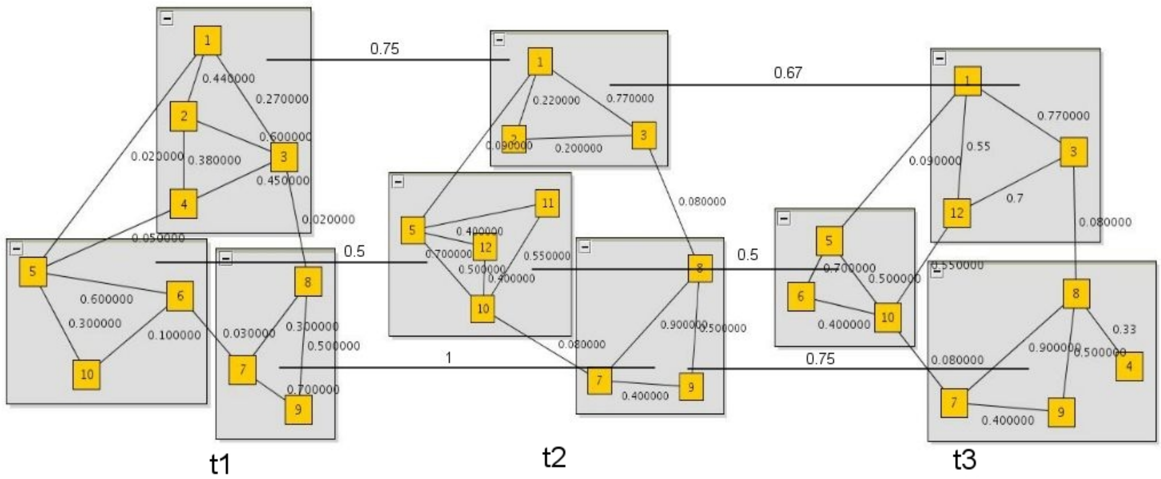


Figure 3.6: Best Match Clusters.

to compare a certain cluster of one clustering to the clusters of another clustering. More precisely, *bestCD* determines the cluster C_j of a clustering $\mathcal{C}'$ with the minimum *centroid* distance to cluster C_i (of another clustering $\mathcal{C}$). Let $\vec{c}$ and $\vec{z}$ be the centroid vectors of clusters C_i and C_j with $C_j \in \mathcal{C}'$. Equation 3.9 shows, how the *bestCD* cluster C_j of cluster C_i is determined.

$$cd(\vec{c}, \vec{z}) = \|\vec{c} - \vec{z}\|_2 = \sqrt{\sum_{i=1}^{n} (\vec{c_i} - \vec{z_i})^2} \tag{3.8}$$

$$bestCD(C_i, \mathcal{C}') = \min_{C_j \in \mathcal{C}'} \left(cd\left(\vec{c}, \vec{z}\right)\right) \tag{3.9}$$

Avgdist

Avgdist measures the average distance of a cluster's customer vectors to their *centroid* vector. In Equation 3.10 the formula of *avgdist* is expressed without a square root, because we want to uprate outliers. Even without the root the formula remains monotone. *Avgdist* indicates the similarity and closeness of the customer vectors to their *centroid* vector. If two clusters are considered similar regarding their low *cd* value, *avgdist* can be an additional instrument to further confirm or reject their similarity. The similarity is weak, if the *avgdist* value of one cluster C_i is much bigger than the *avgdist* value of another cluster C_j. Because then the customer vectors of cluster C_j are on average much closer to their *centroid* value than the customer vectors of cluster C_i to their *centroid* vector. So, the *centroid* vectors are just incidentally similar, but the actual customer vectors of both clusters are not.
Note that *avgdist* is no sufficient comparison measurements for clusters. The *avgdist* values of two clusters can be equal or quite similar although the clusters are not similar.

$$avgdist(C) = \frac{\sum_{j=1}^{n} \left(\sum_{i=1}^{m} \left(\vec{k}_{i,j} - \vec{c_i}\right)^2\right)}{n} \tag{3.10}$$

Chapter 4

Experiments

The main focus of this work is on the detection of meaningful customer profiles. We try to find groups of customers, that are characterized by a certain, easily recognizable purchase behavior. Therefore, we successively analyze the three different graph types introduced in Section 2.4. We want to find appropriate parameters yielding customer groups that are consistent with experts' opinions. Note that those graph types are part of a long progress and that one graph type evolves from the insights gained from another graph type.
Our basic procedure is illustrated in Figure 4.1. It consists of two parts. In the first part, we try to find a promising graph model and promising settings. In the second part, we try to filter out stable clusters[1], that can then be imported into the data mining component of *ToolA* [2] via *PMML* (Chapter 5). Part one is divided into a design phase, an implementation phase and an experiment phase. The selection of a graph model happens in the design phase. Note that we commit ourselves to the *Jaccard* similarity coefficient and the *Greedy Modularity Clustering* algorithm, described and justified in Sections 2.2.6 and 3.2.1. The next phase consists of implementing these concepts. The implementation is done in Java. Based on the particular graph model, selected in the design phase, we implement and adjust a framework that provides the possibility to generate and cluster series of graphs, each with different settings. The experiment phase can be repeated several times for the same graph type. During each repetition, the settings are modified or techniques like sampling or elimination are applied.
If the experiments do not give satisfying results, we identify the reasons for this, return to the design phase and choose a different graph model. In case we find a promising graph type and promising settings, we continue with part two of our procedure. In this part, based on the clusterings constructed in part one, we try to identify stable customer clusters using *Bestmatch* and *BestCD* calculations (3.3.2). If these calculations do not return good enough values and we find out why, we return to the experiment phase of part one, modify certain parameters and again calculate the *match* and *cd* values.

[1] Bigger clusters that are no incidental result, but occur in several graph states. For further explanation have a look at Section 3.3.

[2] A business intelligence, enterprise reporting, and OLAP software suite employed by *dm*.

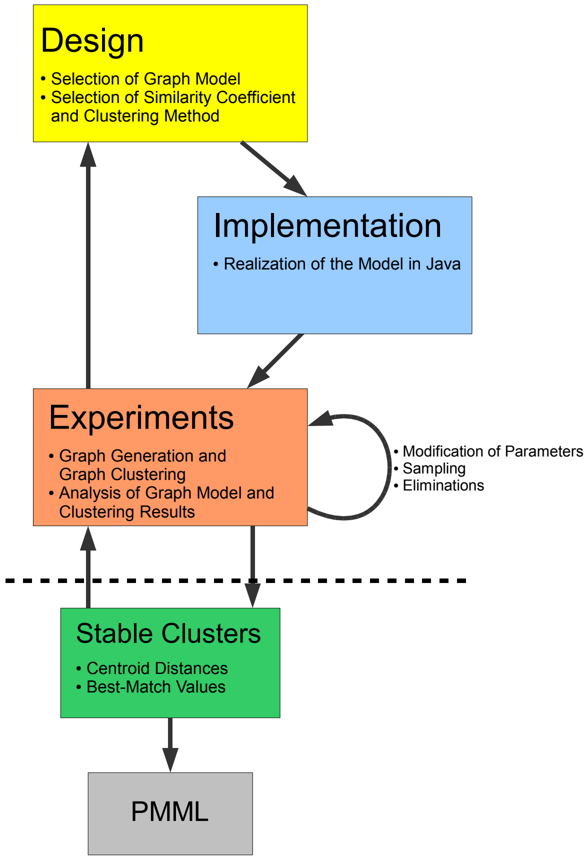


Figure 4.1: The basic procedure of our experiments.

Eventually, we back up our results. To this end, we try to extract the detected stable clusters from the data of two other stores using the same means. The final stable clusters are then transformed into *PMML* and later imported into the data mining component of *ToolA* .
The huge amount of *dm* source data (1.2) forces us to considerable reductions during the algorithm engineering process. When the process is established, we can work with larger mounds of data.
First, we restrict our data to store *242*. This store is located in a shopping center in downtown Munich and is, due to a sales area of 586 sqm, one of the bigger *dm* stores. We choose to start with this store, because it has one of the hightest sales of all German *dm* stores.

4.1 *KAK*-Graphs – [Item-Level]

First, we analyze the *KAK*-graph model, introduced in Section 2.4.1. Remember that our main objective is to find customer groups that reflect the different purchase behaviors of the *dm* customer base. So we start with the model that contains the most detailed information about the purchases. In the *KAK*-graph model, similarities are calculated using the purchased items of the customers. The items are not merged like in the *KMK*- or *KhK*-graphs.

4.1.1 I.Attempt: *Time-Expanded* Graph

For our first experiment, we implemented the *time-expanded* graph. Figure 4.2 shows our tool for generating graphs developed by us and reveals that we allowed several degrees of freedom concerning the graph generation, like the graph model or the minimum edge weight. Two of them regard the *time-expanded* graph generation: We can adjust the number of graph states and we can select one out of three methods to calculate the weights of the inter-time edges. But we only allow directly sequenced graph states to be bridged by inter-time edges, i.e., no inter-time edge spans three or more states.
Our first generated *KAK*-graph is a *time-expanded* graph with four different graph states of the months November 2004, December 2004, January 2005 and February 2005. *Simple Jaccard* is used to calculate the weights of the inter-time edges and the minimum edge weight of the intra-graph edges is set to 0.1. Figure 4.3 shows this graph after the *Greedy Modularity Clustering* algorithm was applied. Note that the colors of the vertices represent the customers' age ranges (Table 4.1). No color means that the customer's age is not specified. As can be seen in Table 4.2, the clustering has a high quality.
To be able to identify the purchase behaviors of the single customer clusters, we determine for each cluster the most frequent sub-assortment and the most frequent brand of the collectively bought items. In the case of our clustered *time-expanded* *KAK*-graph, the dominant sub-assortments are *Cat Food* and *Baby Food* and the dominant brands are *dm-BrandB1* and *BrandL1* . Dominant means that 23 out of the 41 clusters (56%) are *Baby Food/BrandL1* [3] and *Cat Food/dm-BrandB1* clusters, including the five biggest clusters. Interestingly, a high brand loyalty among *Cat Food* and *Baby Food* customers becomes apparent. For instance, in the five biggest clusters the respective percentage of those brands that are most frequently bought by their customers ranges from 70% to 91%.
Figure 4.4 shows the 12 clusters that are connected via inter-time edges. Only two clusters span more than one graph state. This is partly due to the fact that only few vertices are connected via inter-time edges. This means that the customers of the connected clusters change a lot. Furthermore, many inter-time edge weights are zero. So, the neighborhood of customers that do appear in more than one graph state changes a lot, too. But, although

[3] A cluster is called after the sub-assortment and the brand that contain items collectively bought by most of the cluster's customers.

Figure 4.2: The user interface. We allowed two degrees of freedom regarding the *time-expanded* graph generation. Here, the *simple jaccard* is chosen to calculate the inter-time edge weights and four graph states are selected.

the clientele of the clusters changes over the months, we still have one big stable *Cat Food* cluster and one smaller stable *Baby Food* cluster that appear in all four graph states. Moreover, there exists one small *Dog Food* cluster that appears in only two graph states.
In summary, this first attempt already yields two stable customer clusters and a high brand loyalty among *Cat Food* and *Baby Food* customers. Unfortunately, these two *Cat Food* and *Baby Food* clusters are extremely dominant. Potentially, they overshadow other interesting clusters.
Because the *KAK*-graph of this attempt consists of single months, customers that do not shop every month in the same *dm* store do not recur in every graph state. This might be a reason for the few inter-time edges and for the dominance of cat food and baby food buyers. For instance, a customer that just buys toothpaste every few months, would not have any influence on the graph's clusters. Therefore, in our next attempt we try to increase the influence of such customers by considering four quarters of a whole year instead of just single months.

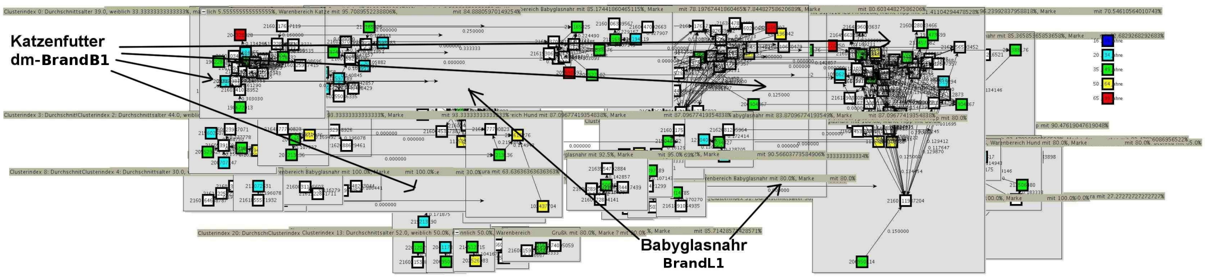


Figure 4.3: Clustered *time-expanded KAK*-graph of the months November 2004, December 2004, January 2005 and February 2005.

Vertex Color	Age Range [years]
blue	16 - 19
turquoise	20 - 34
green	35 - 49
yellow	50 - 64
red	65 - 99

Table 4.1: The colors of the vertices represent the customers' age ranges. No color means that the customer's age is not specified.

Quality Measurement	cov_w	per	$sig_{cov,sub}$
	0.98	0.93	0.83

Table 4.2: Quality measurements of clustered *Time-Expanded KAK*-graph of Figure 4.3.

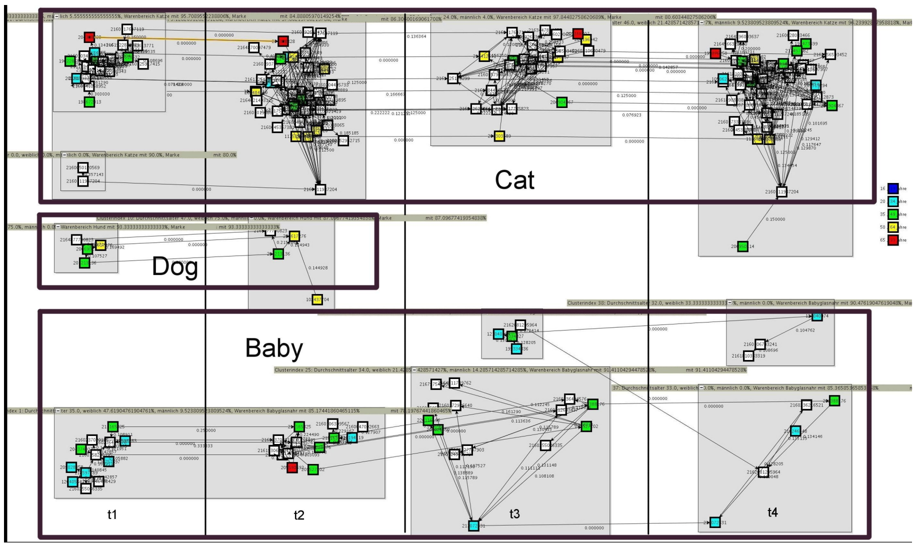


Figure 4.4: Cluster evolution of the *time-expanded KAK*-graph of Figure 4.3.

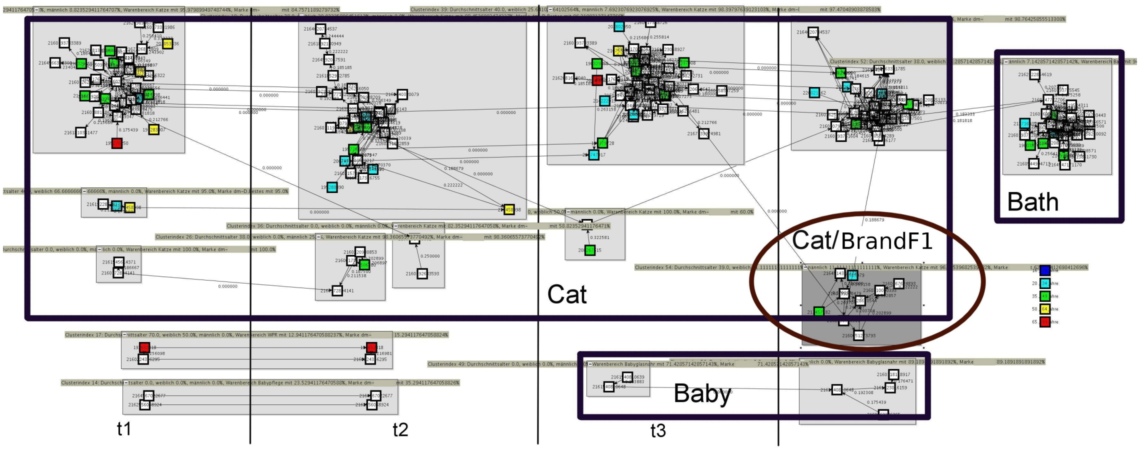


Figure 4.5: Clustered *time-expanded KAK*-graph consisting of the four quarters of 2005.

4.1.2 II.Attempt: Quarters

The extremely dominant *Cat Food* and *Baby Food* clusters of the first experiment appeared most likely because of the high frequency in which these items are bought. Food is needed daily and mostly bought by the week. So, in the next attempt, we examined the four quarters of an entire year instead of just single months. Consequently, items that are bought every few months, like for example sanitary products, should carry more weight.
We generated a *time-expanded KAK*-Graph consisting of the four quarters of 2005. *Simple Jaccard* is used for the inter-time edge weights and the minimum intra-graph edge weight is set to 0.175. The clustering of this graph has a high quality again (Table 4.3) and a very clear structure.
Once again, the most common clusters are either *Baby Food/BrandL1* or *Cat Food/dm-BrandB1* clusters (47%). But this time, the *Baby Food* clusters are very small. Figure 4.5 shows the bigger clusters and clusters that are connected via inter-time edges. We see that there is just one big *Cat Food/dm-BrandB1* cluster connected via inter-time edges in every graph state. We call this cluster stable, despite the small splitting-offs. Moreover, there is one mid-size *Cat Food/BrandF1* cluster and one big *Bath/BrandM1* cluster in the winter quarter.
In other words, this attempt yields only one stable *Cat Food* cluster and many small *Baby Food* clusters. We assume that there exists more than two different customer groups in the *dm* customer base. Other customer clusters might exist in addition to or underneath the two *Cat Food* and *Baby Food* clusters. So it is likely that either more fine-grained clusters are hidden by the *Cat Food* and *Baby Food* clusters or that those two clusters extremely dominate other coexisting clusters.

Quality Measurement	cov_w	*per*	$sig_{cov,sub}$
Quarters	0.99	0.94	0.86

Table 4.3: Quality measurements of clustered *time-expanded* *KAK*-graph of Figure 4.5.

Quality Measurement	cov_w	*per*	$sig_{cov,sub}$
Eliminations	0.98	0.96	0.90

Table 4.4: Quality measurements of clustered *time-expanded* *KAK*-graph with eliminations. See Figure 4.6.

4.1.3 III.Attempt: Eliminations

The first two experiments yielded a very dominant *Cat Food/dm-BrandB1* cluster that possibly overshadowed other interesting clusters. So the next attempt is the elimination of similarities based on items of this sub-assortment*brand combination. This means that the *Jaccard* similarity value of two customers is calculated using all their bought items, except the *Cat Food/dm-BrandB1* items.
We generated a *time-expanded* *KAK*-Graph of January, February and March 2005. Again, *Simple Jaccard* is used for the inter-time edge weights. The intra-graph edge weights are calculated without considering *Cat Food/dm-BrandB1* items and their minimum weight is set to 0.1. The clustering of this graph has a very high quality (Table 4.4). For instance, a *Coverage* value of 98% means that there are only 2% inter-cluster edges. The graph consists of quite strongly connected components. Note that the *Greedy Modularity Clustering* (3.2.1) algorithm tends to transform these components into clusters with only little regard to the actual edge weights. So more fine-grained clusters inside these components can easily be missed.
As shown in Figure 4.6, the resulting clusters are very small. But, we can still see the *Baby Food/BrandL1* amassment. 10 out of 24 clusters (42%) are *Baby Food/BrandL1* clusters and two *Baby Food/BrandL1* clusters appear in two different graph states. A *Cat Food/BrandF1* cluster and a *Dog Food/dm-BrandB1* cluster are the only two other clusters that are connected via inter-time edges and appear in more than one graph state. So, in this experiment, we received *Baby Food/BrandL1* , *Cat Food* and *dm-BrandB1* clusters. These are all clusters that we have already received in Attempt I and II.

4.1.4 Conclusion *KAK*-Graphs

In this section, we tried to find interesting customer clusters in *KAK*-Graphs. For that purpose, we examined three different *time-expanded* clustered *KAK*-Graphs. The first one consisting of four chronologically ordered single months, the second consisting of the four quarters of 2005 and the third one with elimination of *Cat Food/dm-BrandB1* items.

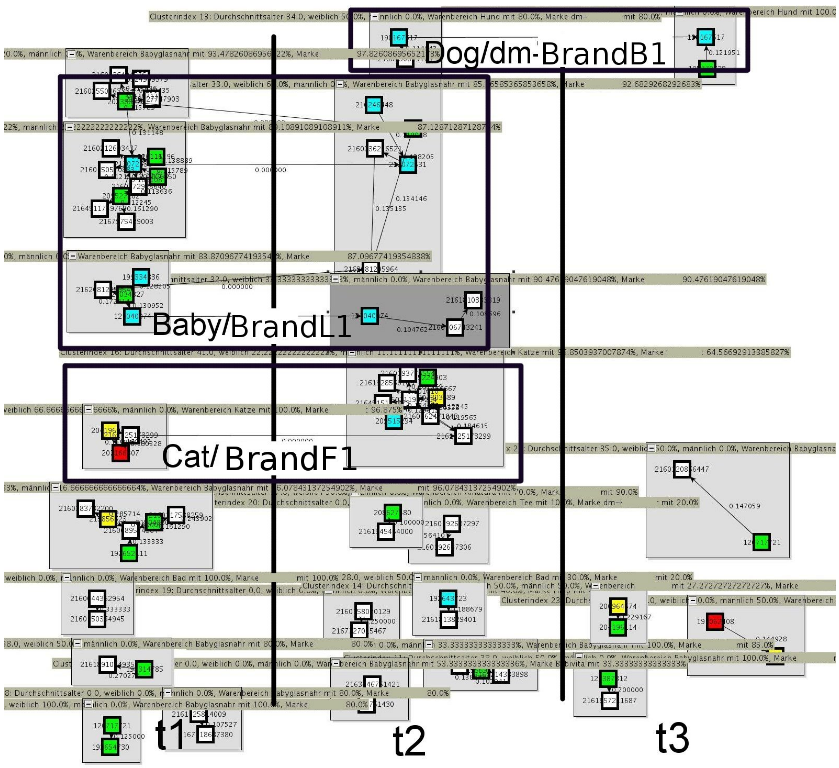


Figure 4.6: Clustered *time-expanded KAK*-graph with *Cat Food/dm-BrandB1* eliminations.

Basically, in all three cases we received stable and very persistent *Baby Food* and *Pet Food* clusters. This is due to the specific characteristics of such sub-assortments. Customers usually buy several flavors of the same product at once, because they like their babys or pets to have a variety in their daily food. And they also buy a big amount of these items, because these items are consumed daily. These specific characteristics explain the high edge weights between pet food buyers and between baby food buyers. After having eliminated *Cat Food/dm-BrandB1* and *Baby Food/BrandL1* items, we could continue with further eliminations. But we expect further eliminations to result in clusters of items that belong to sub-assortments with the above mentioned specific characteristics. An example for another sub-assortment with such specific characteristics is the *Bath* sub-assortment. Products like shower gel, liquid hand soap or bath salt are also needed daily and their buyers like to try different flavors of the same product.
Note that we are not just interested in frequently bought items, but in frequently and infrequently bought items alike. And, we are definitely not just interested in items of sub-assortments with some specific characteristics. So with the next graph model we try to achieve a more equal consideration of all items and sub-assortments.

4.2 *KMK*-Graphs – [Sub-assortment*Brand-Level]

As a possible remedy to the issues found in the context of the *KAK*-graph, we now evaluate the *KMK*-graph model. As described in Section 2.4.2, the *KMK*-graph model represents the customer–sub-assortment*brand–customer relation. Similarities between customers are no longer calculated using the items a customer bought. They are calculated using the sub-assortment*brand combinations of the items a customer bought.
The advantage of using sub-assortment*brand combinations is that sub-assortments and brands do not change a lot over the years while single items have a high fluctuation. Oftentimes new items are included into the assortment or unpopular items are no longer carried.
The *KAK*-graph favors frequently bought items. We hope to remove this disadvantage with the design of the *KMK*-graph. In this graph type binary customer vectors are used and items in the same sub-assortment and of the same brand are merged. So, now frequently and infrequently bought items should have the same influence on the similarity calculation.

4.2.1 I.Attempt: One Quarter

To get a general idea of the clusters that result from this graph type, we first generate a simple (not *time-expanded*) *KMK*-graph consisting of three months. We use a minimum edge weight of 0.2. Figure 4.7 shows the resulting clustered graph. Many of the clusters

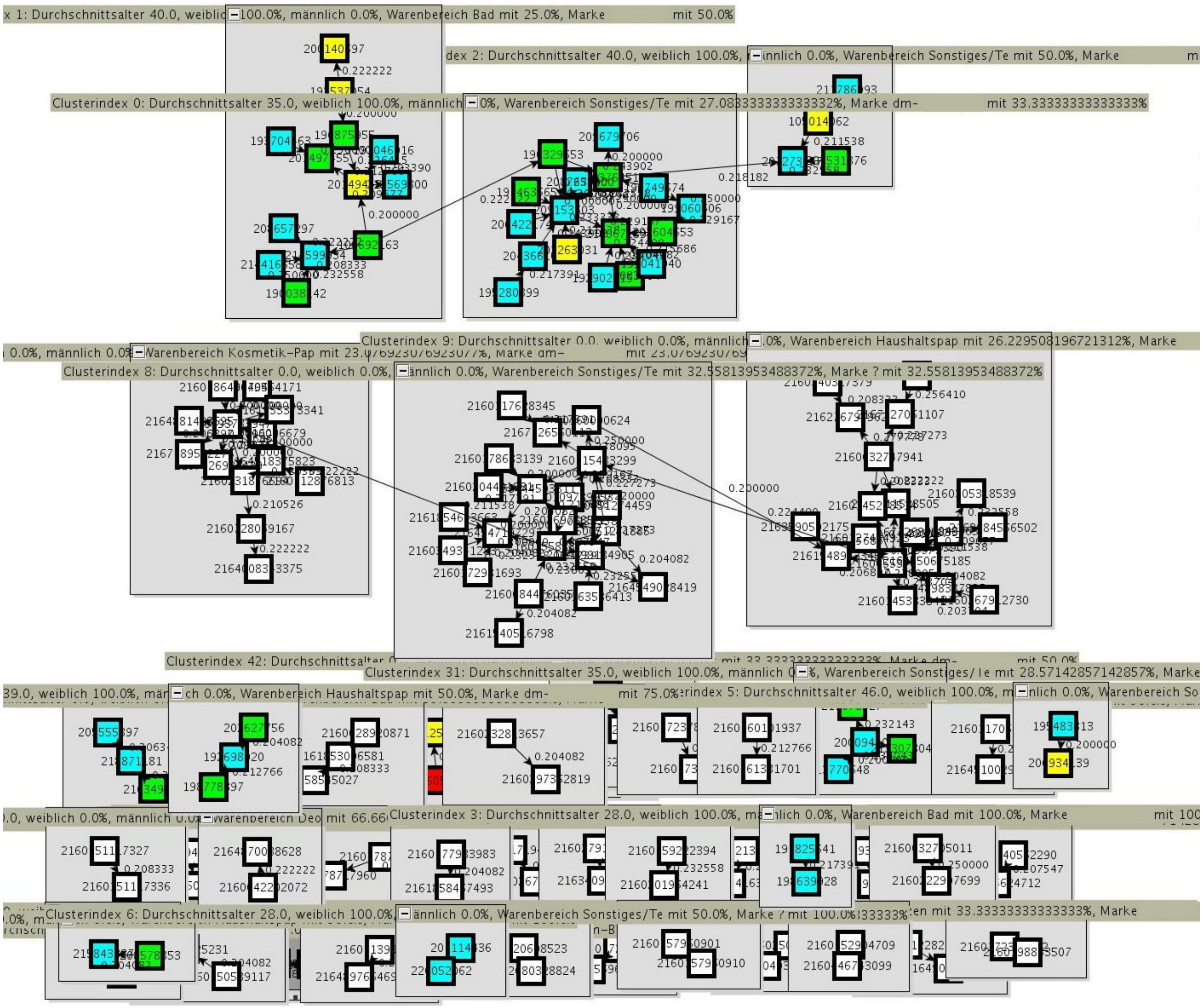


Figure 4.7: Clustered *KMK*-graph of three months.

are characterized[4] by sub-assortments like *Sonstiges/Te*, *Kosmetik-Pap*, *Hygienepapie* or *Haushaltspap*. These sub-assortments contain items like plastic bags, toilet paper, cotton swabs, tissues et cetera. Furthermore, most clusters (62%) are characterized by *dm*-owned brands like *dm-BrandN1* , *dm-BrandO1* or *dm-BrandK1* . These items are bought by most of the *dm* customers. They are not very suitable to distinguish between certain customer groups.

[4]A cluster is characterized by the sub-assortment that contains items collectively bought by most of the cluster's customers.

Quality Measurement	cov_w	per	$sig_{cov,sub}$
I.Attempt	0.97	0.96	0.87
II. Attempt	0.99	0.95	0.90

Table 4.5: Quality measurements of the clustered *KMK*-graphs of Figures 4.7 and 4.8.

graph state	average age	female [%]	male [%]
t_1	37	87	12
t_2	38	89	10
t_3	37	87	12
t_4	38	86	13

Table 4.6: Age and gender distribution of the colored stable cluster of the *time-expanded KMK*-graph of Figure 4.8.

4.2.2 II.Attempt: *Time-Expanded* Graph of Whole Year

Then, we generate a *time-expanded KMK*-graph consisting of the four quarters of 2005, similar to the *time-expanded KAK*-graph in Section 4.1.2. The high quality of the clustering of this graph can be seen from Table 4.5. This clustering consists of 754 clusters. Figure 4.8 shows the eight biggest clusters, that are connected via inter-time edges. Remember that the colors of the vertices represent the customers' age ranges. No color means that the customer's age is not specified. We can easily recognize two stable clusters that appear in each of the four graph states. All customers of the first stable cluster have specified their age, while the customers of the second stable cluster have not. Table 4.6 shows that the average ages and the percentages of the female and male customers of the first colored stable cluster are almost equal in every single graph state. This indicates the inner homogeneity of the two stable clusters. Seven of the eight shown clusters are characterized by the sub-assortment *Sonstiges/Te* and the eighth one by the sub-assortment *Haushaltspap*. Four clusters feature the brand *dm-BrandN1* . As already mentioned, these sub-assortments contain items like tissues, plastic bags and toilet paper, that are bought by most of the *dm* customers. Besides, the average age of all *dm* customers in 2005 was 42 years. In this year, 86% of the *dm* customers that specified their gender were female and 13% male. Comparing these facts with the age and gender distribution of the first colored stable cluster (Table 4.6) shows that they are very similar.

In summary, the two stable clusters are characterized by sub-assortments that contain items bought by most of all customers. And additionally, the age and gender distribution of the first colored stable cluster is quite close to the overall age and gender distribution of the *dm* customer base. So this experiment yields two average customer profiles.

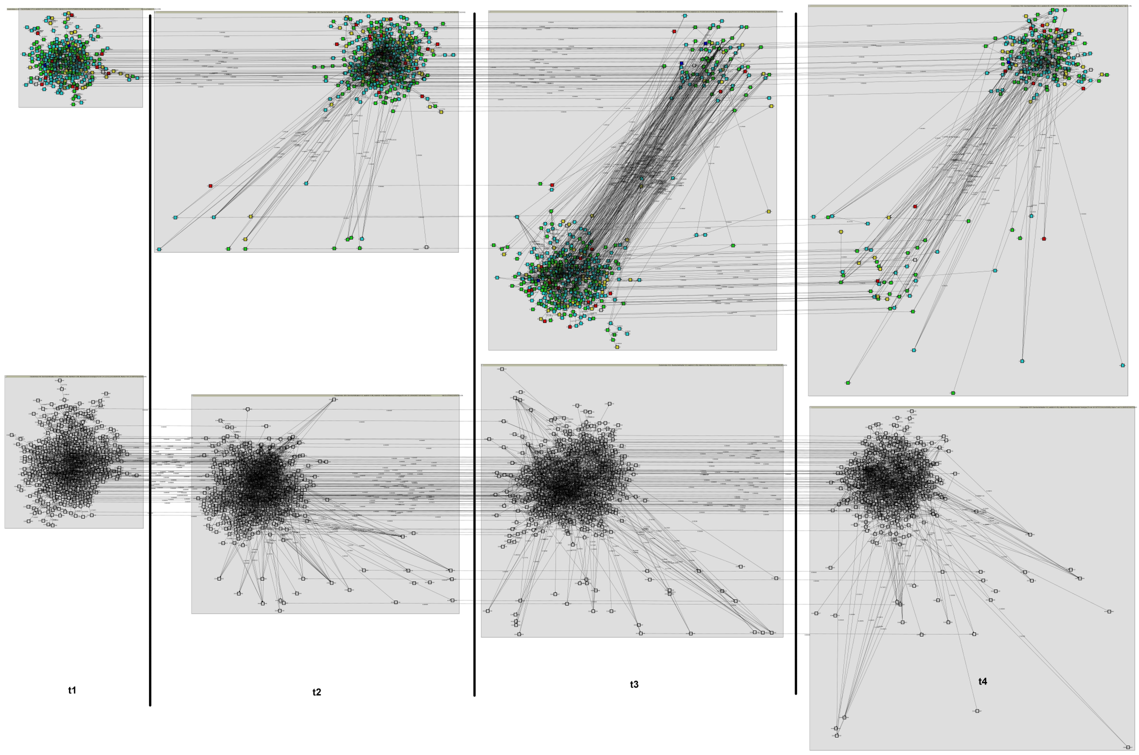


Figure 4.8: Clustered *time-expanded* *KMK*-graph of 2005.

4.2.3 Conclusion *KMK*-Graphs

In this section, we analyzed two different clustered *KMK*-graphs. The first one consisting of three sequential months and the second, *time-expanded* one consisting of the four quarters of 2005. Both clusterings mostly contained clusters characterized by sub-assortments and brands that are not appropriate to discriminate between *dm* customers. Items of these sub-assortments and brands are needed by almost everybody. They are bought additionally to the individual needs and are usually no motive to go to a certain store or to go shopping at all. Furthermore, one of the two stable clusters of the *time-expanded* clustered *KMK*-graph had a female and male customer proportion similar to the general gender distribution of the *dm* customer base. The same applied to the average age of this stable clusters' customers.
So, the *KMK*-graph just partly met our expectations by not favoring very frequently bought items. In exchange of the *Baby* and *Pet Food* customer groups resulting from clustered *KAK*-graphs, we received even less interesting average customer groups from clustered *KMK*-graphs.

4.3 *KhK*-Graphs – [Sub-assortment*FavoriteBrand-Level]

Finally, we analyze the *KhK*-graph type. Remember from Section 2.4.3 that the *KhK*-graph model represents customer–most-frequent-brand*sub-assortment–customer relations. A customer is characterized by his most frequently bought brand of each sub-assortment.
In the last two sections, we learned about the *KAK*-graph that it favors frequently bought items and about the *KMK*-graph that its clusters reflect the average *dm* customer. The resulting clusters of both graph types did not turn out to be acceptable and diverse *dm* customer profiles. So, by merging a customer's bought items, similar to how it is done in the *KMK*-graph, frequently bought items should no longer dominate infrequently bought ones. And, by just considering the customers' favorite brands, we hope to sharpen the expressiveness of the resulting customer groups by disregarding items that are bought by everybody.
Another advantage of the *KhK*-graph is the possibility of a compressed matrix representation by just memorizing the string value of the favorite brand for each sub-assortment. So, a customer vector just consists of as many dimensions as there are sub-assortments. In the case of the *KMK*-graph the number of dimensions is given by the number of brand*sub-assortment combinations and in the case of the *KAK*-graph it is given by the number of customers (or at least by the number of items). Currently, *dm* has about [redacted] customers[5] with *PAYBACK* cards, [redacted] different items and [redacted] brand*sub-assortment combinations, but only [redacted] sub-assortments. We found out, that the number of dimensions is crucial for the successful *PMML*-import into the data mining component of *ToolA* .

4.3.1 I.Attempt: Sampling

First, we generate a *KhK*-graph containing the first quarter of 2005. The minimum edge weight is set to 0.2. Figure 4.9 shows the resulting clusters after applying the *Greedy Modularity Clustering* algorithm to this graph. There are three very big clusters with more than 420 customers, one big cluster containing 129 customers and 15 very small clusters with 2 to 10 customers. So, the clusters differ a lot in size. The black beams are edge bundles and stand for the many inter-cluster edges. This clustering has a *Coverage* value of 0.58 (Table 4.7), which means that 42% of all edges are inter-cluster edges. The inferior quality of this clustering is even confirmed by its low *Modularity* value.
One possible reason for this inferior clustering could be the high minimum edge weight of 0.2. The clustering contains 1581 customer vertices which account for only 4% of all customers in this quarter. These customers are selected, because they all have a minimum purchasing similarity value of 20% to at least one other customer, which is quite restrictive.

[5] Customer IDs in our data source.

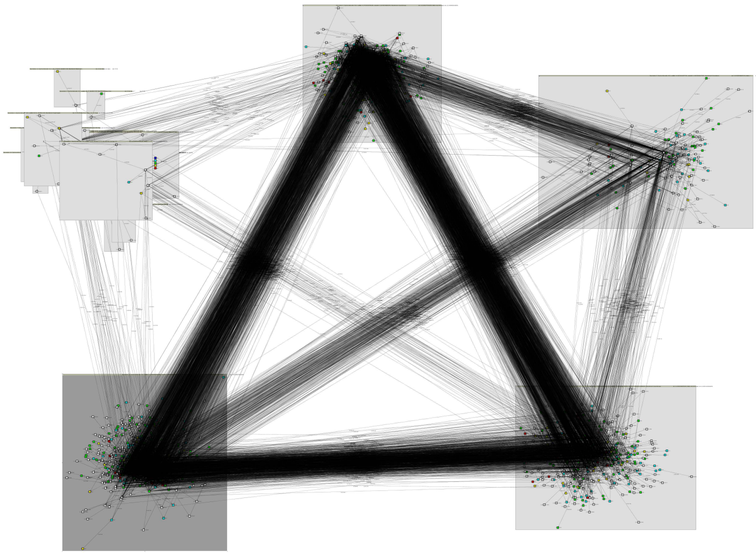


Figure 4.9: Clustered *KhK*-graph containing the first quarter of 2005.

These selected customers are probably not very representative of the *dm* customer base. A solution to this problem could be *Sampling*. That is, we first sample the customer base and then reduce the minimum edge weight in a way, that the resulting clustering again contains a similar number of customers and a number which our methods in their present form can handle.

So, next we generate a customer sample by selecting those customers of the first quarter of 2008, whose IDs are divisible by 4. Then, we genereate a *KhK*-graph, that contains only sampled customers. With a reduced minimum edge weight of 0.05, we receive a clustering containing 1196 customer vertices. As we can see in Figure 4.10, the resulting clustering looks like the clustering without *Sampling* (Figure 4.9). Unexpectedly, its quality is even worse.

The high amount of inter-cluster edges between the four biggest clusters in both clusterings indicates a certain similarity between those clusters and cluster *centroids*[6] respectively. A closer examination of the cluster *centroids* reveals that most of the clusters are extremely dominated by some popular *dm*-owned brands like *dm-BrandK1* , *dm-BrandN1* , *dm-BrandP1* , *dm-BrandQ1* , *dm-BrandD* et cetera[7]. In other words, these clusterings reflect more or less one big customer profile, namely the *dm* customer. All *dm* customers have the purchase of popular *dm* items in common.

Consequently, we hope to reveal more interesting customer groups underneath this one big *dm* customer group by eliminating the specific popular *dm*-owned brands, considering them as non-discriminative.

[6] *Centroids* are explained in Section 3.3.2.

[7] Table E.1 contains these popular *dm*-owned brands.

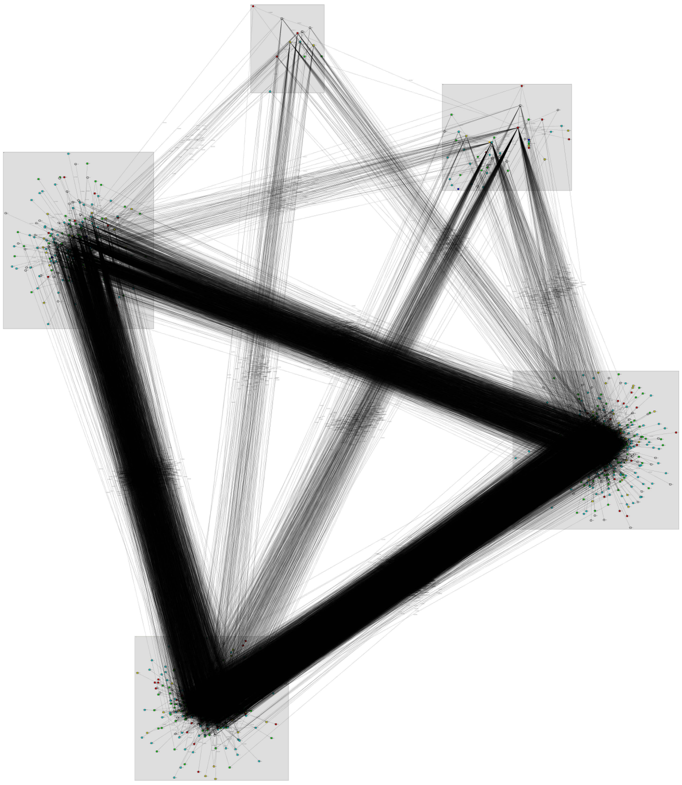


Figure 4.10: Clustered *KhK*-graph containing the first quarter of 2008 with sampling.

Quality Measurement	cov_w	per	$sig_{cov,sub}$
05Q1	0.58	0.73	0.26
08Q1 with Sampling	0.54	0.67	0.14

Table 4.7: Quality measurements of clustered *KhK*-graphs of Figures 4.9 and 4.10.

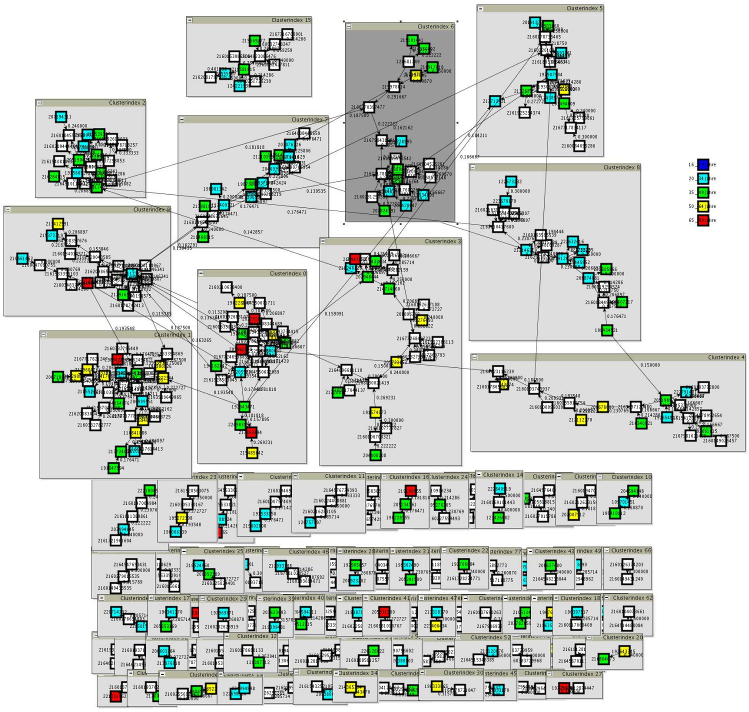


Figure 4.11: Clustered *KhK*-graph containing the first quarter of 2006 with eliminations.

4.3.2 II.Attempt: Eliminations

The last experiment resulted in one big *dm* customer group due to strong similarities between *dm* customers based on popular *dm*-owned brands, which almost everybody seems to buy regularly. So, next we eliminate similarities based on sub-assortment*brand combinations containing popular *dm*-owned brands. This means that a customers favorite brand of every sub-assortment is determined without considering popular *dm*-owned brands. Then the *Jaccard* similarity value of two customers is calculated using their favorite, not (popular) *dm*-owned brands of every sub-assortment.

Figure 4.11 depicts a clustered *KhK*-graph of the first quarter of 2006 with such eliminations. This clustering is a tremendous improvement compared to the clusterings of the first attempt. Table 4.8 highlights the high quality of this clustering. Furthermore, we received several midsize clusters instead of just three to four huge clusters.

So our assumption has been confirmed: The elimination of specific popular *dm*-owned brands yields better clustering results. The dominance of the *dm*-owned brands shows that they are certainly well received by the customer base, but they are also non-discriminatory and they obliterate possible customer profiles.

Since we have found a promising graph type and promising settings, the next step of our basic procedure (Figure 4.1) is to find stable clusters, which can serve as representative *dm* customer profiles.

Quality Measurement	cov_w	per	$sig_{cov,sub}$
06Q1 with Eliminations	0.95	0.97	0.91

Table 4.8: Quality measurements of clustered *KhK*-graph of Figure 4.11.

4.3.3 Stable Clusters

Now we try to find stable clusters in *KhK*-graphs, in which popular *dm*-owned brands are eliminated. Therefore, we compare either the clusters of sequential quarters of the same year, or the clusters of the same quarter of sequential years to each other. We use the comparison measurements *Bestmatch* and *BestCD*, introduced in Section 3.3.2. A high *match* value of two clusters of different quarters means that those two clusters have a lot of customers in common and are therefore similar to each other. A low *cd* value of two clusters means that the *centroids* of those clusters are very similar and so the customers of both clusters have similar average buying patterns. To find one stable cluster, means that we have to find the *Bestmatch* and/or *BestCD* clusters for a certain cluster in some other quarters with corresponding high *match* and/or *cd* values.
First, we analyze the clusters of the fourth quarters of 2004 (*04Q4*) and 2005 (*05Q4*). For this purpose, we generate the two corresponding clustered *KhK*-graphs with a minimum edge weight of 0.05 each and extract those clusters containing more than 9 customers. This threshold is an arbitrary choice from us, but the clusters should not be too small because *match* and *cd* values need a certain cluster size to be meaningful. For instance, a *match* value of 50% of two clusters of the size 2 has a different relevance than the same *match* value of two clusters of the size 50. It could be a coincidence, if one customer appears in two different clusters. But it is no longer a coincidence if 25 customers appear in two different clusters. The clustered *KhK*-graph of *04Q4* contains 16 big clusters. Their *match* values to corresponding *Bestmatch* clusters range from 2.4% to 8.4% and their *cd* values to corresponding *BestCD* clusters range from 0.6 to 1.7. Table 4.9 reveals a correlation between *Bestmatch* and *BestCD* clusters for five of the 16 big *04Q4* clusters. The *Bestmatch* clusters of these five *04Q4* clusters are equal to their *BestCD* clusters, which strengthens our findings.
Note, that a maximum *match* value (to the *Bestmatch* clusters) of 8.4% is not very high, since it means that at most 8.4% of the customers belonging to a *04Q4* cluster are part of the corresponding *05Q4* *Bestmatch* cluster. But, how many customers of *04Q4* are still customers in *05Q4*? The answer is 25%. Although this relativizes the 8,4%, it still means that only a maximum of one-third of all possible customers match. A possible explanation for these poor *match* values could be the big temporal gap between both clusterings. For this reason, we have a look at the *match* value of the clusters of two sequential quarters. In the case of the big *06Q3* and *06Q4* clusters, the maximum *match* values are even worse, scilicet 7.4%. So, the bad *match* values are not due to the temporal gap.
To get a general idea of these results, we calculate the *Bestmatch* and *BestCD* clusters of the four quarters of 2006 (*06Q1*–*06Q4*) and the *Bestmatch* and *BestCD* clusters of the

fourth quarters of 2004 to 2007 (*04Q4–07Q4*). To visualize this, we generate *Bestmatch*- and *BestCD*-graphs. The vertices of these graphs represent the clusters, the colors of the vertices represent the graph state (year and quarter) and the vertex size represents the size of the cluster. An edge connects a cluster with its *Bestmatch* (*BestCD*) cluster of a neighboring state and the edge weight stands for the *match* (*cd*) value of both clusters. The basic structure of the *Bestmatch*-graph is illustrated by Figure 4.12. Figure 4.13 shows the *Bestmatch*-graph of *06Q1–06Q4*. Note that only the mentioned clusters of *06Q3* and *06Q4*, illustrated by the yellow and grey clusters, have a maximum *match* value of 7.4%. The maximum *match* value of all 2006 quarters is 10.9 %. As can be seen in Figure D.1, the maximum *match* value of *04Q4–07Q4* is just 8.6%.
We wonder about these low *match* values. So next, we try to identify the reasons for this. To this end, we only consider customers that appear in all four graph states. Note that those customers are only the regulars and very loyal *PAYBACK* card users. But they are the only customers, we can make reliable statements about. Thus, we determine the customer intersection of *06Q1*, *06Q2*, *06Q3* and *06Q4*. Then, we generate the corresponding *KhK*-graphs that contain only customers of the determined intersection. Their minimum edge weights are set to 0.05. Finally, we cluster the four graphs and generate the *Bestmatch*-graph (4.14). By just considering the regular customers, we achieve a maximum *match* value of 15.8%. This value is better than before, but still curiously low.
So, we still try to identify the reasons for these relatively low *match* values. Therefore, we reduce the minimum edge weights of the used *KhK*-graphs. Note that first we keep only those customers in the *KhK*-graphs that are contained in the determined customer intersection and then we remove the customers that do not have a minimum purchasing similarity to at least one other customer. So, the lower the minimum edge weight, the more customers of the customer intersection remain in the graphs. Think of a scenario in which a certain customer has irregular buying patterns. He might buy a lot in *06Q1* and *06Q3* and just a few items in *06Q2* and *06Q4*. So he potentially has a higher similarity (edge weight) to other customers in *06Q1* and *06Q3* than he has to customers in *06Q2* and *06Q4*. If in *06Q1* and *06Q3* his highest edge weight to at least one other customer is above the minimum edge weight and in *06Q2* and *06Q4* it is below it, then reducing the minimum edge weight would increase the *match* value by this customer[8].
Figure 4.15 shows the *Bestmatch*-graph equal to the last *Bestmatch*-graph of Figure 4.14, except that the minimum edge weights of the corresponding *KhK*-graphs are set to 0.03. The maximum *match* value of this graph is 39.7% between cluster 12 of *Q2* and cluster 35 of *Q3*. This means that almost 40% of the customers of cluster 12[9] are contained in cluster 35, too.
It would be interesting to find out, if a further reduction of the minimum edge weight would further increase the *match* values and what the highest reachable *match* value is. But as already mentioned the reduction of the minimum edge weight yields much bigger

[8]Of course the *match* value only increases, if the customer does buy similar items in all four quarters and thus would belong to the same customer profile in all four quarters, which seems likely.

[9]Generally speaking, it is almost 40% of the customers of cluster 12 or cluster 35, depending on which cluster is the biggest. See Section 3.3.2.

graphs and during the algorithm engineering process the amount of data we can work with is limited.
To be sure that the single clusterings of the used *KhK*-graphs are still of good quality, we have a look at the quality measurements of one of the clustered graphs. Table 4.10 compares the quality measurements of the clustered *KhK*-graphs of *06Q1* with mentioned eliminations and customer intersection. The quality measurements of the clustered *KhK*-graph with minimum edge weight of 0.05 are excellent, while the quality measurements of the clustered *KhK*-graph with minimum edge weight of 0.03 are still satisfying. But note that quality measurements depend on the size of the graph. The graph with minimum edge weight of 0.03 contains 4151 vertices and the graph with minimum edge weight of 0.05 contains only 321 vertices. In other words, what we have here is a trade-off between the quality measurements, i.e., the clarity, of the single clusterings and the *match* values of the *Bestmatch* clusters. A low minimum edge weight yields high *match* values, but also lower clustering qualities, as more data means more outliers and more garbage.
Now, that we have stable clusters with good *match* values based on clusterings of satisfying quality, we continue with the analysis of these stable clusters. For that purpose we compare the *Bestmatch*-graph of Figure 4.15 to its corresponding *BestCD*-graph. The cut-out of this graph in Figure 4.16 shows only those stable clusters that are contained in the *Bestmatch*-graph, too. The *cd* values range from 0.22 to 0.67, which is very good. So, with the *Bestmatch*- and the *BestCD*-technique we have extracted four stable clusters. Table 4.11 summarizes the *match* and *cd* values of the found stable clusters. Apparently, the *match* values anti-correlate with the *cd* values in a way that high *match* values come along with low *cd* values.
Next, we analyze the age and gender distribution of our four stable clusters. Table 4.12 reveals that the stable clusters differ significantly in their average ages. The average customer ages of the first stable cluster range from 45.7 to 51.1, in the second stable cluster they range from 39.6 to 41.3, in the third one they range from 35.7 to 36.8 and in the last one they range from 42.7 to 46.5. Unfortunately, the gender distribution is less interesting, except for the fact, that the last two stable clusters have a little higher proportion of men than the first two stable clusters. But at large, the gender distribution is quite similar in all clusters.
Finally, we have a closer look at the purchase behaviors of the stable clusters' customers. Therefore, we compare the highest and most interesting *centroid* component values of the single clusters for each stable cluster. Remember from Section 3.3.2 that the *centroid* component value of a certain sub-assortment*brand combination reflects the proportion of the cluster's customers whose favorite brand belongs to this sub-assortment*brand combination. At least this proportion of customers must have bought at least one item in this sub-assortment*brand combination and in the case of the *KhK*-graph, the highest value of one sub-assortment reflects the cluster's customers favorite brand of this sub-assortment. Tables 4.13, 4.14, 4.15 and 4.16 show the highest and most interesting *centroid* component values for our four stable clusters. An underlined *centroid* component value means that this value is the highest one of this cluster. Mind that regarding the *centroid* component values, the single clusters of the stable clusters have partly the same first ranked sub-

Cluster ID	*Bestmatch* Cluster	*match* value	*BestCD* cluster	*cd* value
3	1	0.083	1	0.603
4	13	0.035	13	0.934
8	11	0.084	11	0.908
16	23	0.076	23	0.818
17	21	0.044	21	1.188

Table 4.9: Clusters of *04Q4* and their corresponding *bestmatch* and *bestCD* clusters in *05Q4*.

Quality Measurement	cov_w	per	$sig_{cov,sub}$
Minimum Edge Weight of 0.05	0.95	0.96	0.89
Minimum Edge Weight of 0.03	0.67	0.86	0.46

Table 4.10: Quality measurements of clustered *KhK*-graphs of *06Q1* with eliminations and customer intersection.

assortment*brand combination[10], which is a strong and surprisingly positive observation. For instance, in the stable cluster of Table 4.13 three of four single clusters have the highest *centroid* component value for the sub-assortment*brand combination *Kerzen/BrandR1* . That applies to two of three single clusters of the stable cluster in Table 4.14 for the sub-assortment*brand combination *Sonne*/dm-*BrandS1* and it applies to two of four single clusters of the customer profiles in Table 4.16 for the sub-assortment*brand combination *Hand*/dm-*green-BrandT1* , too. This is even more obvious for the stable cluster in Table 4.15. There, even the order of the top four sub-assortment*brand combinations is almost identical. We name this stable cluster the *Baby/Young Families* cluster due to the affinity to baby products. Almost 50% of the customers in this cluster buy diapers. The clusters' relatively young average age of 36.5 is consistent to that. The stable cluster of Table 4.16 gets the label *Green Concerns* due to the affinity to natural products. In this stable cluster, green brands like *dm-green-BrandT1* and *green-BrandU1* appear more often. Due to the relatively high average age and the tendency to conservative products, we name the stable cluster of Table 4.13 the *Traditional/Premium* cluster. Because of its tendency to workaday products, the last stable cluster of Table 4.14 gets the label *Mainstream*.
In summary, by eliminating popular *dm*-owned brands, by considering only customers that appear in all four graph states and by reducing the minimum edge weight of the single *KhK*-graphs from 0.05 to 0.03, we receive, regarding *Bestmatch* and *BestCD*, four stable clusters that are based on clusterings of satisfying quality. We label them *Traditional/Premium*, *Mainstream*, *Baby/Young Families* and *Green Concerns*. Their stability is even confirmed by an interesting age distribution and by in each case individual buying behaviors.

[10]This first ranked sub-assortment*brand combination of a cluster has the highest *centroid* component value of all sub-assortment*brand combinations. It is bought by most of the cluster's customers.

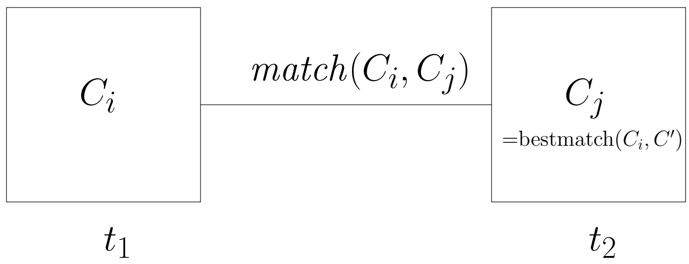


Figure 4.12: Structure of *bestmatch*-graphs. A node represents a cluster. A cluster is connected with its *bestmatch*-cluster and an edge weight represents the *match* value of both clusters.

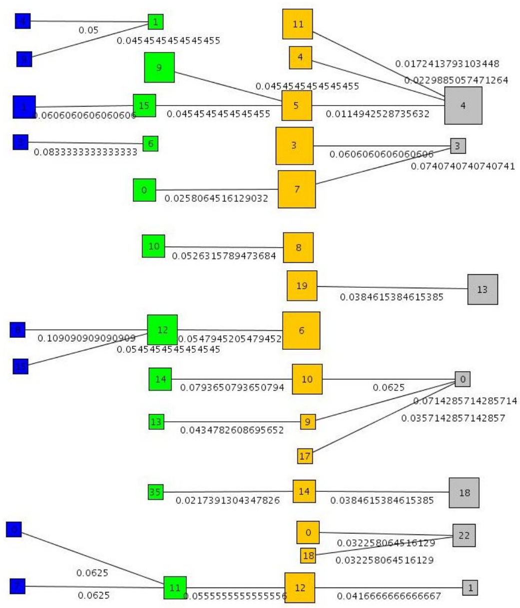


Figure 4.13: *Bestmatch*-graph of 2006 (Q1-Q4). The vertex colors represent the single quarters of 2006. Blue represents the quarter *06Q1*, green represents the quarter *06Q2*, yellow represents the quarter *06Q3* and grey represents the quarter *06Q4*.

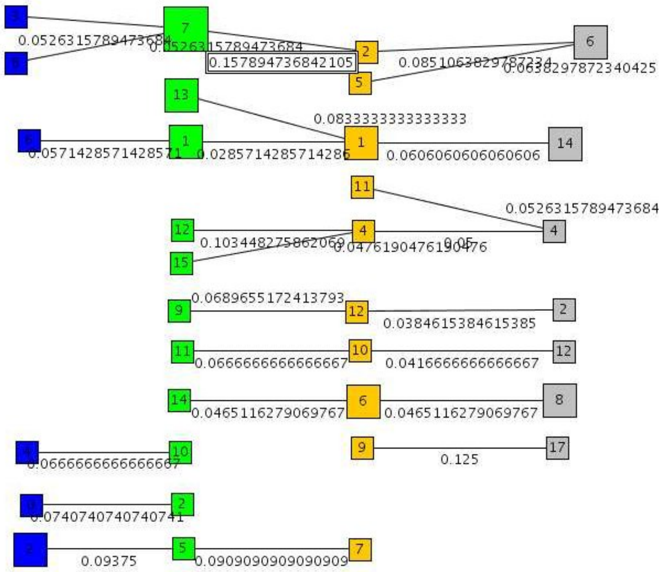


Figure 4.14: *Bestmatch*-graph of 2006 (Q1-Q4) with customer intersection and a minimum edge weight of 0.05.

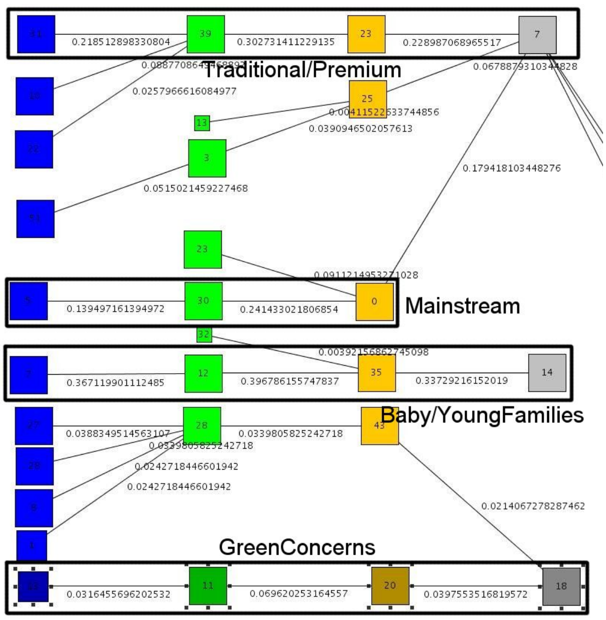


Figure 4.15: *Bestmatch*-graph of 2006 (Q1-Q4) with customer intersection and a minimum edge weight of 0.03.

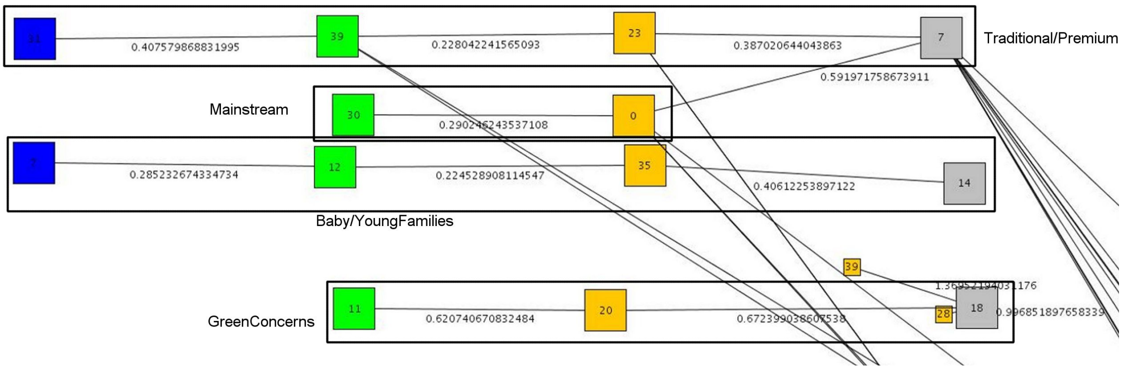


Figure 4.16: *BestCD*-graph of 2006 (Q1-Q4) with customer intersection and a minimum edge weight of 0.03. Only those clusters are included that are also in Figure 4.15.

Cluster ID_1	Cluster ID_2	*match* [%]	*cd*
31	39	21.9%	0.41
39	23	30.3%	0.23
23	7	22.9%	0.39
5	30	14.0%	-
30	0	24.1%	0.29
(0)	(7)	(17.9%)	(0.59)
7	12	36.7%	0.29
12	35	39.7%	0.23
35	14	33.7%	0.41
(13)	(11)	(3.1%)	-
11	20	7.0%	0.62
20	18	4.0%	0.67

Table 4.11: *Match* and *cd* values of stable clusters as in Figures 4.15 and 4.16.

ClusterID	Year/Quarter	avg(Age)	Women [%]	Men [%]
31	06Q1	47.4	87.3% (351)	12.7% (51)
39	06Q2	51.6	87.5% (448)	12.5% (64)
23	06Q3	51.1	89.1% (384)	10.9% (47)
7	0604	45.7	88.7% (713)	11.3% (91)
5	06Q1	40.9	88.9% (296)	11.1% (37)
30	06Q2	41.3	90.2% (468)	9.8% (51)
0	06Q3	39.6	87.7% (500)	12.3% (70)
7	06Q1	35.7	86.5% (244)	13.5% (38)
12	06Q2	36.7	82.6% (300)	17.4% (63)
35	06Q3	36.6	83.5% (289)	16.5% (57)
14	06Q4	36.8	82.8% (322)	17.2% (67)
(13)	(06Q1)	(50.7)	(71.4% (15))	(28.6% (6))
11	06Q2	46.0	86.8% (59)	13.2% (9)
20	06Q3	46.5	80.0% (48)	20.0% (12)
18	06Q4	42.7	84.0% (121)	16.0% (23)

Table 4.12: Age and gender distribution of the customer profiles from Figures 4.15 and 4.16. The single customer profiles are separated by double lines. Note that the customer profiles differ significantly in their average ages, but the average ages of the single clusters of a customer profile vary just a little.

ClusterID/ YearQuarter	*avgdist*	Sub-assortment/Brand	*Centroid* Value [%]
31/06Q1	10.35	Kerzen (45)/BrandR1 (2075)	12.1%
		Filter&Folien (56)/BrandX1 (1243)	10.4%
		Haushaltspap (6)/BrandY1 (1265)	15.3%
		Haarpflege (5)/BrandZ1 (33)	17.0%
		Damenhygiene (3)/BrandA2 (2191)	14.8%
		Kosmetik-Pap (4)/BrandB2 (1241)	15.1%
39/06Q2	9.83	Kerzen (45)/BrandR1 (2075)	<u>28.4</u>%
		Filter&Folien (56)/BrandX1 (1243)	15.3%
		Haushaltspap (6)/BrandY1 (1265)	19.6%
		Haarpflege (5)/BrandZ1 (33)	16.0%
		Damenhygiene (3)/BrandA2 (2191)	15.3%
		Kosmetik-Pap (4)/BrandB2 (1241)	12.7%
23/06Q3	9.65	Kerzen (45)/BrandR1 (2075)	<u>25.3%</u>
		Filter&Folien (56)/BrandX1 (1243)	11.8%
		Haushaltspap (6)/BrandY1 (1265)	14.6%
		Haarpflege (5)/BrandZ1 (33)	15.7%
		Damenhygiene (3)/BrandA2 (2191)	19.2%
		Kosmetik-Pap (4)/BrandB2 (1241)	12.0%
7/06Q4	9.85	Kerzen (45)/BrandR1 (2075)	<u>19.0%</u>
		Filter&Folien (56)/BrandX1 (1243)	8.2%
		Haushaltspap (6)/BrandY1 (1265)	12.7%
		Haarpflege (5)/BrandZ1 (33)	10.1%
		Damenhygiene (3)/BrandA2 (2191)	12.2%
		Kosmetik-Pap (4)/BrandB2 (1241)	8.9%

Table 4.13: Interesting *centroid* component values of the *Traditional/Premium* customer profile from Figures 4.15 and 4.16. Am underlined *centroid* value is the highest one of its cluster.

ClusterID/ YearQuarter	*avgdist*	Sub-assortment/Brand	*Centroid* Value [%]
5/06Q1	10.15	Bonbon (40)/BrandC2 (1772) Nassrasur (66)/*dm*-BrandD2 (218) Sonne (48)/*dm*-BrandS1 (210)	25.7% 13.4% 9.3%
30/06Q2	9.56	Bonbon (40)/BrandC2 (1772) Nassrasur (66)/*dm*-BrandD2 (218) Sonne (48)/*dm*-BrandS1 (210)	15.0% 14.8% 60.0%
0/06Q3	9.60	Bonbon (40)/BrandC2 (1772) Nassrasur (66)/*dm*-BrandD2 (218) Sonne (48)/*dm*-BrandS1 (210)	20.7% 16.0% 43.3%

Table 4.14: Interesting *centroid* component values of the *Mainstream* customer profile from Figures 4.15 and 4.16.

ClusterID/ YearQuarter	*avgdist*	Sub-assortment/Brand	*Centroid* Value [%]
7/06Q1	9.04	Windeln (27)/BrandE2 (899) Babynahrung (33)/BrandA1 (1846) Babyglasnahr (37)/BrandA1 (1846) Babyglasnahr (37)/BrandL1 (1847)	52.8% (1.) 34.3% (2.) 32.1% (3.) 25.2% (4.)
12/06Q2	9.57	Windeln (27)/BrandE2 (899) Babynahrung (33)/BrandA1 (1846) Babyglasnahr (37)/BrandA1 (1846) Babyglasnahr (37)/BrandL1 (1847)	49.4% (1.) 28.8% (2.) 28.6% (3.) 20.3% (4.)
35/06Q3	9.19	Windeln (27)/BrandE2 (899) Babynahrung (33)/BrandA1 (1846) Babyglasnahr (37)/BrandA1 (1846) Babyglasnahr (37)/BrandL1 (1847)	47.5% (1.) 29.5% (3.) 31.8% (2.) 22.0% (4.)
14/06Q4	9.63	Windeln (27)/BrandE2 (899) Babynahrung (33)/BrandA1 (1846) Babyglasnahr (37)/BrandA1 (1846) Babyglasnahr (37)/BrandL1 (1847)	34.6% (1.) 25.3% (2.) 23.5% (3.) 21.0% (4.)

Table 4.15: Interesting *centroid* component values of the *Baby/Young Families* customer profile from Figures 4.15 and 4.16.

ClusterID/ YearQuarter	*avgdist*	Sub-assortment/Brand	*Centroid* Value [%]
(13/06Q1)	9.50	Bad (54)/*dm*-green-BrandT1 (201)	-
		Hand (39)/*dm*-green-BrandT1 (201)	17.5%
		Haarpflege (5)/*dm*-green-BrandT1 (201)	-
		Bodylotion (61)/*dm*-green-BrandT1 (201)	-
		WPR (15)/BrandF2 (238)	-
11/06Q2	9.59	Bad (54)/*dm*-green-BrandT1 (201)	20.3%
		Hand (39)/*dm*-green-BrandT1 (201)	<u>30.4%</u>
		Haarpflege (5)/*dm*-green-BrandT1 (201)	17.7%
		Bodylotion (61)/*dm*-green-BrandT1 (201)	15.2%
		WPR (15)/BrandF2 (238)	12.0%
20/06Q3	9.45	Bad (54)/*dm*-green-BrandT1 (201)	24.2%
		Hand (39)/*dm*-green-BrandT1 (201)	21.9%
		Haarpflege (5)/*dm*-green-BrandT1 (201)	11.7%
		Bodylotion (61)/*dm*-green-BrandT1 (201)	16.4%
		WPR (15)/BrandF2 (238)	-
18/06Q4	9.54	Bad (54)/*dm*-green-BrandT1 (201)	8.0%
		Hand (39)/*dm*-green-BrandT1 (201)	<u>16.2%</u>
		Haarpflege (5)/*dm*-green-BrandT1 (201)	15.0%
		Bodylotion (61)/*dm*-green-BrandT1 (201)	10.4%
		WPR (15)/BrandF2 (238)	8.9%

Table 4.16: Interesting *centroid* component values of the *Green Concerns* customer profile from Figures 4.15 and 4.16.

4.3.4 Conclusion *KhK*-Graphs

In this section we tried to find meaningful customer profiles in clustered *KhK*-graphs. First, we generated and clustered a *KhK*-graph of *05Q1* with a minimum edge weight of 0.2. The resulting clustering of this graph was of inferior quality. We tried to find the reasons for this. So first we tried to improve the representativeness of the graphs' customers. To this end, we generated a customer sample and reduced the minimum edge weight to 0.05. Unexpectedly, applying the sampling technique did not yield a higher quality of the resulting clustering. A closer examination of the *centroids* of the three to four biggest clusters revealed that they were all extremely dominated by popular *dm*-owned brands and thus represented one big *dm* customer profile. So, next we eliminated similarities based on sub-assortment*brand combinations containing popular *dm*-owned brands. The quality of the resulting clustering was excellent, so we continued to search for stable clusters in clustered *KhK*-graphs with such eliminations. Figure 4.17 summarizes the procedure of the extraction of stable clusters. The *06Q1–06Q4* *KhK*-graphs based on the customer intersection of 2006 with a minimum edge weight of 0.03 yielded four clusters that were stable with respect to *Bestmatch* and *BestCD* calculations. We labeled them *Traditional/Premium*, *Mainstream*, *Baby/Young Families* and *Green Concerns* based on the sub-assortment*brand combinations of their highest *centroid* component values. To some extent, their unique age distribution confirmed the stability of the clusters as well as their labels.
In the next sections, we try to confirm our results by extracting similar customer profiles out of analoge *KhK*-graphs based on the source data of two other *dm* stores.

4.4 Second Store

In this section, we analyze the source data of the *dm* store 518. This store differs from the previous store in location and sales area. While the previous store has a sales area of 586 sqm and is located in a Munich shopping center, this store has a much bigger sales area of 679 sqm and is located in Karlsruhe in a small industrial area close to a residential area. Nearby are several supermarkets, a home improvement store and a pet food store, all with big parking lots. So people drive to this area, to run several errands at once.

4.4.1 *KhK*-Graphs – [Sub-assortment*FavoriteBrand-Level]

First, we cluster the *KhK*-graph of *06Q1* (Figure 4.18). Analog to the corresponding *KhK*-graph of the previous store, we receive three huge clusters containing 1068, 783 and 238 customers and eleven small clusters containing 2 to 14 customers. Accordingly, the clustering's quality is unsatisfying (Table 4.17). The table of Section F.1 shows the sub-assortment*brand combinations with the hightest *centroid* component values of the three biggest clusters (4, 6 and 9). We see that the huge clusters of this *KhK*-graph are again

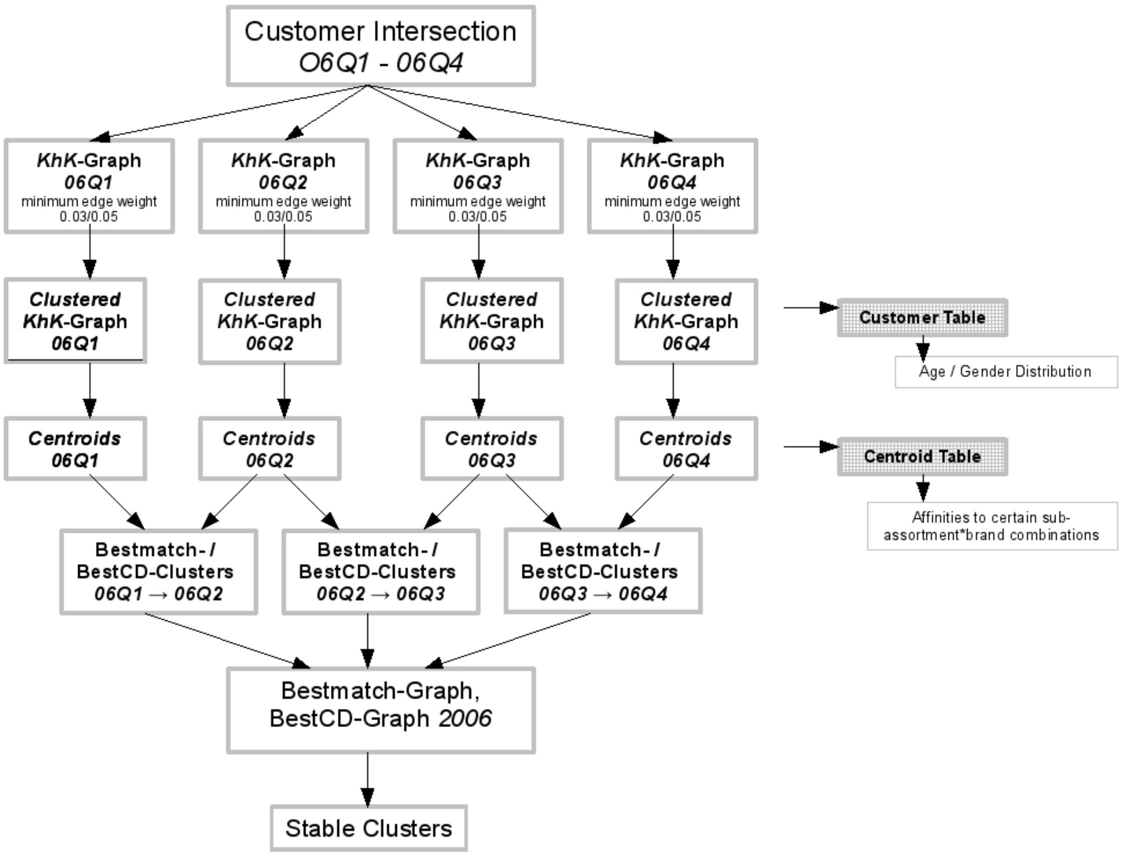


Figure 4.17: Stable cluster extraction.

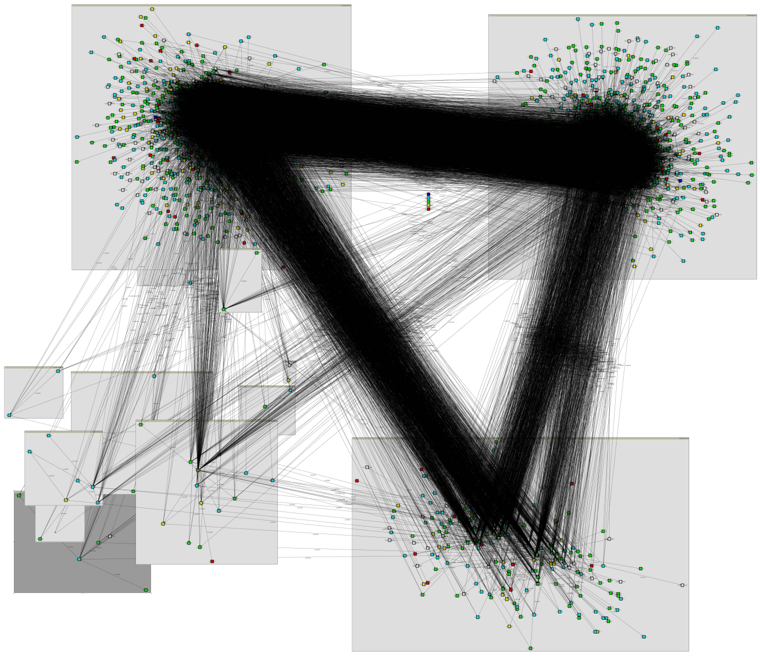


Figure 4.18: Clustered *KhK*-graph of *06Q1*, store 518.

Quality Measurement	cov_w	per	$sig_{cov,sub}$	# Cluster	# Nodes	# Edges
No Eliminations	0.68	0.61	0.22	14	2153	44721
Eliminations	0.68	0.91	0.54	31	440	901

Table 4.17: Quality measurements of clustered *KhK*-graphs of *06Q1* with and without eliminations, store 518.

dominated by specific popular *dm*-owned brands. So, we again eliminate those brands. The resulting clustered *KhK*-graph is shown in Figure 4.19. Instead of three huge clusters we now have several mid-sized clusters. The clustering is of much higher quality than before the elimination (Table 4.17). In other words, the elimination technique takes a positive effect on *KhK*-graphs of store 518, too.

4.4.2 Stable Clusters

Next, we generate the *Bestmatch*- and *BestCD*-graphs of *KhK*-graphs with *dm*-owned brand eliminations. As can be seen in Figures F.1 and F.2 the highest *match* value is 20.3% and the *cd* values range from 0.63 to 1.86. Remember that the corresponding highest *match* value of all 2006 quarters of the previous store was 10.9%. So, here we have double the highest *match* value. It is possible that a store inside a shopping center has a higher proportion of occasional customers than a store in the described industrial area. This could explain the difference between both highest *match* values.

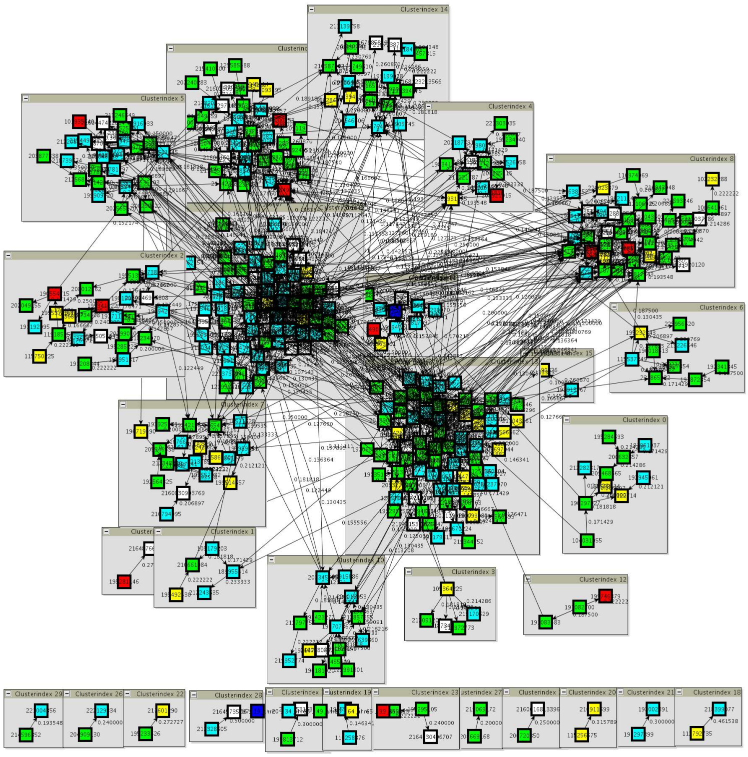


Figure 4.19: Clustered *KhK*-graph of *06Q1* with eliminations, store 518.

But 20.3% is still relatively low. Analog to the previous store, we determine the customer intersection of *06Q1*, *06Q2*, *06Q3* and *06Q4*. This means that only regular customers remain in the *KhK*-graphs. We again set the minimum edge weights of these graphs to 0.05 and cluster them. Then we generate the *Bestmatch*-graph (Figure F.3). Now, the highest *match* value is 24.1%, which is just slightly better.
We again assume, that the reason for this is the too high minimum edge weight of 0.05. Remember that we first keep the customers of the customer intersection in the four *KhK*-graphs and then we remove those customers that do not have a minimum purchasing similarity to at least one other customer. So, we reduce the minimum edge weight to 0.03. The corresponding *Bestmatch*- and *BestCD*-graphs are shown in Figures 4.20 and 4.21. We see that now the highest *match* value is 48.4%. The *cd* values improve, too. They now range from 0.26 to 1.5. Table 4.18 shows, that in contrast to store 242 there is no longer an anti-correlation between the *match* values and the *cd* values[11]. Analog to the previous store there is a trade-off between higher *match* values and a high clustering quality. As can be seen in Table 4.19, reducing the minimum edge weight does not only yield better *match* values, but also a much bigger graph with a poorer clustering quality. Note that though the clustering quality is poorer, it is still sufficiently good.
Next, our expectations are satisfied, because we can confirm the results of store 242. Going over the highest *centroid* component values of big clusters that have high enough *match* and low enough *cd* values to big clusters of other quarters, we discover the same stable customer profiles as in the previous store. But here, only the *Baby/Young Families* profile spans all four quarters, while the *Mainstream* profile, the *Traditional/Premium* profile and the *GreenConcerns* profile span only two quarters. And the *Traditional/Premium* profile is, regarding its highest *centroid* component values, a mix of typical *Traditional/Premium* sub-assortment*brand combinations and typical *Mainstream* sub-assortment*brand combinations.
Some high and interesting *centroid* component values are listed in Tables 4.21, 4.22, 4.23 and 4.24. To facilitate comparisons with the customer profiles of the previous store, we list the exact same sub-assortment*brand combinations. The complete lists of the highest *centroid* component values of the single clusters can be seen in the appendix (Section F.3). Those complete lists show that our customer profiles definitely consist of profile-specific sub-assortment*brand combinations. For instance, the complete *centroid* component value lists of the *Baby/Young Families* clusters are shown in Sections F.3.3 (Cluster 11), F.3.3 (Cluster 13), F.3.3 (Cluster 24) and F.3.3 (Cluster 2). The sub-assortment*brand combination *Baby-Push/BrandG2* has the highest *centroid* component value in Clusters 13 and 24, and it has the second highest *centroid* component value in Clusters 11 and 2. So, these clusters clearly belong to the customer profile *Baby/Young Families*. Another observation is that the *GreenConcerns* customer profile is much clearer now (Table 4.24). For example, the sub-assortment*brand combination *Hand/dm-green-BrandT1* has the highest *centroid* component value of Cluster 7, that is 45.8%.
Table 4.20 contains the age distribution of the four stable clusters. Like in the previous

[11]This table contains only the *match* and *cd* values of the stable clusters.

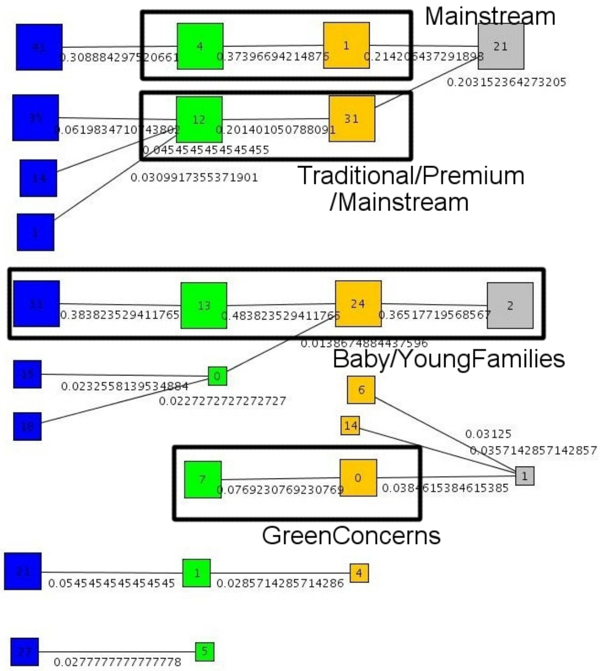


Figure 4.20: Best matches of 2006 (Q1-Q4) with customer intersection and minimum edge weight of 0.03, store 518. Almost the same customer profiles as in store 242 (Figure 4.15).

store, the *Baby/Young Families* single clusters have the youngest average ages. The average ages of this customer profile and those of the *Mainstream* profile are quite similar to the average ages of the corresponding profiles of store 242. That does not apply to the *Traditional* and *GreenConcerns* profiles. Their average ages are much lower. But, note that the average ages of the single clusters inside each customer profile do not differ much. They vary by less than 0.5 years. This additionally confirms their stability. So the age difference between stores 242 and 518 seems to be a demographic phenomenon of some sort.
In summary, applying the same techniques and adjustments of the previous store to *KhK*-graphs of store 518 yields almost the exact same four stable customer profiles. Even though store 242 and 518 differ a lot in size and location.

4.5 Third Store

The *dm* store 22 is analyzed in this section. This store again differs from the two previous stores in location and sales area. It is located in a suburban residential estate of Stuttgart. Due to its sales area of 480 sqm, it is the smallest one of the three stores.

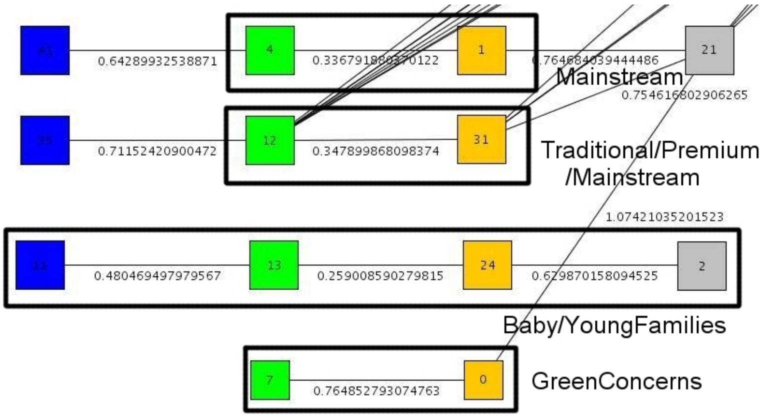


Figure 4.21: *BestCDs* of 2006 (Q1-Q4) with customer intersection and minimum edge weight of 0.03, store 518.

Cluster ID_1	Cluster ID_2	*match* [%]	*cd*
41	4	30.9%	0.64
4	1	37.4%	0.34
1	21	21.4%	0.77
35	12	6.1%	0.71
12	31	20.1%	0.35
31	21	20.3%	0.76
11	13	38.4%	0.48
13	24	48.4%	0.26
24	2	36.5%	0.63
7	0	7.7%	0.77

Table 4.18: *Match* and *cd* values of stable clusters, store 518

Quality Measurement	cov_w	per	$sig_{cov,sub}$	# Cluster	# Nodes	# Edges
Intersection 0.05	0.73	0.85	0.51	30	351	732
Intersection 0.03	0.69	0.77	0.34	60	1914	13761

Table 4.19: Quality measurements of clustered *KhK*-graphs of *06Q1* with eliminations, customer intersection and reduced minimum edge weight, store 518.

ClusterID	Year/Quarter	avg(Age)
(41)	(06Q1)	(41.4)
4	06Q2	41.3
1	06Q3	40.9
(21)	(0604)	(41.4)
(35)	(06Q1)	(40.8)
12	06Q2	42.2
31	06Q3	42.2
11	06Q1	37.0
13	06Q2	37.0
24	06Q3	36.8
2	0604	36.5
7	06Q2	41.0
0	06Q3	41.4

Table 4.20: Age distribution of stable clusters, store 518. Note that the average ages of the single clusters inside each customer profile vary in less than 0.5 years.

ClusterID/ YearQuarter	Sub-assortment/Brand	*Centroid* Value [%]
12/06Q2	Kerzen (45)/BrandR1 (2075)	6.8%
	Filter&Folien (56)/BrandX1 (1243)	8.2%
	Haushaltspap (6)/BrandY1 (1265)	13.6%
	Haarpflege (5)/BrandZ1 (33)	12.4%
	Damenhygiene (3)/BrandA2 (2191)	12.2%
31/06Q3	Kerzen (45)/BrandR1 (2075)	28.4%
	Filter&Folien (56)/BrandX1 (1243)	7.7%
	Haushaltspap (6)/BrandY1 (1265)	13.8%
	Haarpflege (5)/BrandZ1 (33)	14.1%
	Damenhygiene (3)/BrandA2 (2191)	14.5%
	Kosmetik-Pap (4)/BrandB2 (1241)	4.6%

Table 4.21: Interesting *centroid* component values of *Traditional/Premium/Mainstream* stable cluster, store 518.

ClusterID/ YearQuarter	Sub-assortment/Brand	*Centroid* Value [%]
4/06Q2	Bonbon (40)/BrandC2 (1772)	14.8%
	Nassrasur (66)/*dm*-BrandD2 (218)	17.2%
	Sonne (48)/*dm*-BrandS1 (210)	39.9% (2.)
1/06Q3	Bonbon (40)/BrandC2 (1772)	14.7%
	Nassrasur (66)/*dm*-BrandD2 (218)	19.9%
	Sonne (48)/*dm*-BrandS1 (210)	28.1%

Table 4.22: Interesting *centroid* component values of *Mainstream* stable cluster, store 518.

ClusterID/ YearQuarter	Sub-assortment/Brand	*Centroid* Value [%]
11/06Q1	Windeln (27)/BrandE2 (899)	46.4% (1.)
	Babynahrung (33)/BrandA1 (1846)	20.3%
	Babyglasnahr (37)/BrandA1 (1846)	27.4%
	Babyglasnahr (37)/BrandL1 (1847)	13.7%
13/06Q2	Windeln (27)/BrandE2 (899)	37.9%
	Babynahrung (33)/BrandA1 (1846)	17.5%
	Babyglasnahr (37)/BrandA1 (1846)	21.8%
	Babyglasnahr (37)/BrandL1 (1847)	17.4%
24/06Q3	Windeln (27)/BrandE2 (899)	40.1% (2.)
	Babynahrung (33)/BrandA1 (1846)	17.1%
	Babyglasnahr (37)/BrandA1 (1846)	24.8%
	Babyglasnahr (37)/BrandL1 (1847)	13.1%
2/06Q4	Windeln (27)/BrandE2 (899)	29.8%
	Babynahrung (33)/BrandA1 (1846)	16.3%
	Babyglasnahr (37)/BrandA1 (1846)	20.8%
	Babyglasnahr (37)/BrandL1 (1847)	14.8%

Table 4.23: Interesting *centroid* component values of *Baby/Young Families* stable cluster, store 518.

ClusterID/ YearQuarter	Sub-assortment/Brand	*Centroid* Value [%]
7/06Q2	Bad (54)/*dm*-green-BrandT1 (201)	37.3% (2.)
	Hand (39)/*dm*-green-BrandT1 (201)	45.8% (1.)
	Haarpflege (5)/*dm*-green-BrandT1 (201)	25.4%
	Bodylotion (61)/*dm*-green-BrandT1 (201)	16.9%
	WPR (15)/BrandF2 (238)	28.8% (4.)
0/06Q3	Bad (54)/*dm*-green-BrandT1 (201)	29.5% (2.)
	Hand (39)/*dm*-green-BrandT1 (201)	26.9% (4.)
	Haarpflege (5)/*dm*-green-BrandT1 (201)	28.2% (3.)
	Bodylotion (61)/*dm*-green-BrandT1 (201)	10.3%
	WPR (15)/BrandF2 (238)	21.8%

Table 4.24: Interesting *centroid* component values of *GreenConcerns* stable cluster, store 518. The *centroid* component values are much higher and clearer as in the *GreenConcerns* profile of the previous store. See Table 4.16

Quality Measurement	cov_w	per	$sig_{cov,sub}$	# Cluster	# Nodes	# Edges
No Eliminations	0.65	0.66	0.24	9	1520	29356
Eliminations	0.73	0.93	0.61	29	442	823

Table 4.25: Quality measurements of clustered *KhK*-graphs of *06Q1* with and without eliminations, store 22.

4.5.1 *KhK*-Graphs

First, we have a look at unmodified *KhK*-graphs of store 22. The clustered *KhK*-graph of *06Q1* is shown in Figure 4.22. Its structure is reminiscent of the corresponding clustered *KhK*-graphs of the two previous stores. It consists of three huge clusters containing 633, 564 and 288 customers and six small clusters containing 2 to 9 customers. Table 4.25 shows the clustering's poor quality. The table of Section G.1 shows that the huge clusters (1, 2 and 5) are again dominated by popular *dm*-owned brands. So, according to the previous stores we eliminate those brands. As expected the resulting clustered *KhK*-graph (4.23) has a much better structure and the clustering's quality (4.25) is much higher. This means that in this store the elimination of specific *dm*-owned brands improves clusterings of *KhK*-graphs, too.

4.5.2 Stable Clusters

Now, we calculate the *Bestmatch* and *BestCD* clusters of *KhK*-graphs with *dm*-owned brand eliminations. Here, the highest *match* value is 11.0% and the *cd* values range from

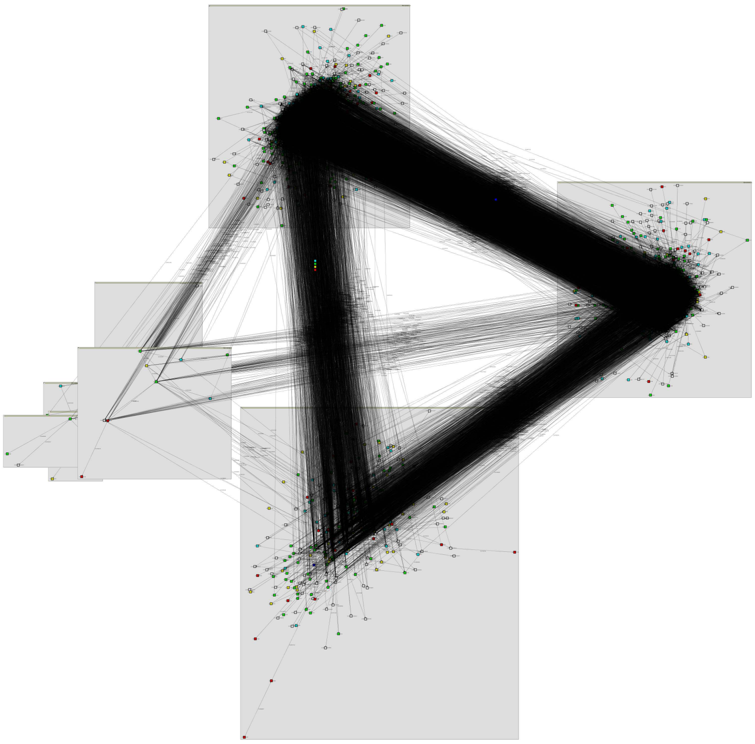


Figure 4.22: Clustered *KhK*-graph of *06Q1*, store 22.

1.27 to 1.58. These values are as unsatisfying as the corresponding values of store 242.
To remain comparable to the last two stores we continue with the generation of the customer intersection of the four quarters of 2006. We again cluster the four *KhK*-graphs whose minimum edge weights are set to 0.05 and whose vertices are contained in the customer intersection. The corresponding *Bestmatch*-graph (Figure G.1) shows, that now the highest *match* value is 29.5%. This is a big improvement compared to the 11%. But we consider this highest *match* value to be an outlier, because the second highest *match* value already is 14%.
Because of the experience gained with the last two stores, we next reduce the minimum edge weight of the four *KhK*-graphs from 0.05 to 0.03 and cluster them. Figures 4.24 and 4.25 depict the corresponding *Bestmatch*- and *BestCD*-graphs. Now, the highest *match* value is 52.7% and the *cd* values range from 0.3 to 1.3. So, the procedure of eliminating specific popular *dm*-owned brands, generating the customer intersection and reducing the minimum edge weight again yields stable customer clusters with relatively high *match* values and relatively low *cd* values. Table 4.26 illustrates the trade-off between higher *match* values and a high clustering quality. Reducing the minimum edge weight to 0.03 impairs the clustering's quality, but just to a still acceptable extent.
Fortunately, we can again confirm the results of stores 242 and 518. Inspecting the highest *centroid* component values of those big single clusters that have high enough *match* and low enough *cd* values to other big single clusters, yields a *Baby/Young Families* customer

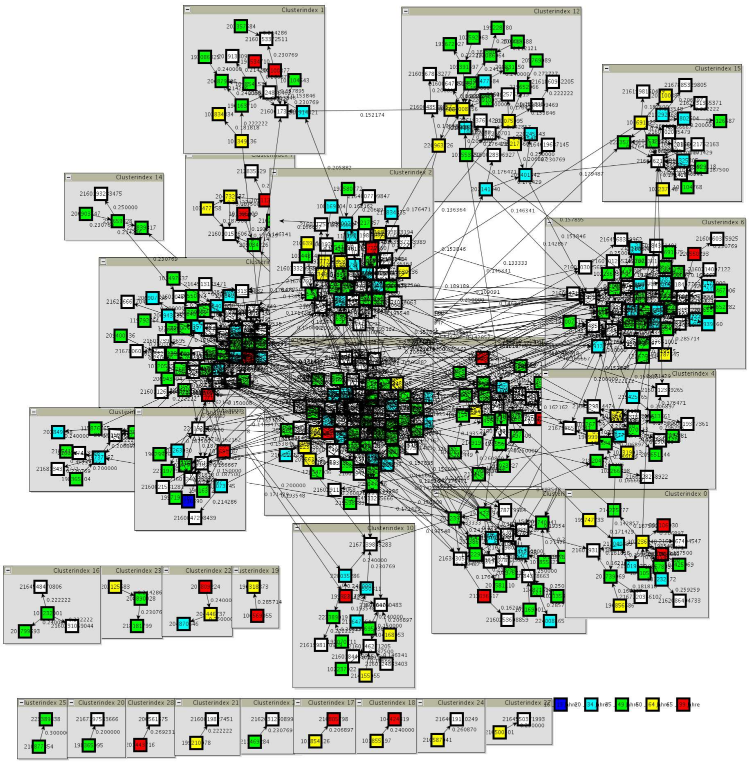


Figure 4.23: Clustered *KhK*-graph of *06Q1* with eliminations, store 22.

Quality Measurement	cov_w	per	$sig_{cov,sub}$	# Cluster	# Nodes	# Edges
Intersection 0.05	0.72	0.94	0.61	31	403	n/a
Intersection 0.03	0.67	0.77	0.35	42	2224	16477

Table 4.26: Quality measurements of clustered *KhK*-graphs of *06Q1* with eliminations, customer intersection and reduced minimum edge weight, store 22.

profile, that spans three quarters, a *Mainstream* customer profile, that spans two quarters, and a *Traditional/Premium* customer profile, that even spans four quarters, see Figures 4.24 and 4.25. We can not assign the *GreenConcerns* customer profile.
For a better comparison to the previous stores, we list the *centroid* component values of the same sub-assortment*brand combinations in Tables 4.28, 4.29 and 4.30, except that for the *Kerzen* sub-assortment of the *Traditional/Premium* customer profile we list the brand *BrandH2* . In this store, the brand *BrandH2* has 2 to 8 percentage points higher *centroid* component values than the brand *BrandR1* (see Tables of Section G.3.3). In the previous stores, the brand *BrandR1* is the most popular brand of the sub-assortment *Kerzen* in the *Traditional/Premium* customer profile. Analog to store 242, the combination *Sonne/dm-BrandS1* has the highest *centroid* component value in both single clusters of the *Mainstream* customer profile. Another observation is that like in store 518 the *Baby-Push/BrandG2* combination is among the top two *centroid* component values in the *Baby/Young Families* customer profile.
The age distribution of the three stable customer profiles is shown in Table 4.27. The age distribution of this store is similar to that of store 242, except that the average ages of this store are between 3 to 5 years higher. It seems plausible, that the customers' average ages are higher in a suburban residential estate than in an industrial area or in a downtown shopping center. Once again[12], the *Traditional/Premium* customer profile has the highest average ages (49.7 – 56.6), the average ages of the *Mainstream* customer profile are in the middle (43.3 – 44.6) and the average ages of the *Baby/Young Families* customer profile are the lowest (39.6 – 40.0). Note that the average ages of the *Baby/Young Families* customer profile vary in less than 0.4 years. This further confirms the profile's stability.
In other words, using the same procedure as in stores 242 and 518 we can confirm three of the four stable customer profiles. Although, all three stores have different sizes and locations.

4.6 Comparison on Store-Level

In the last three sections, we identified three to four stable customer profiles in three totally different stores. They were all stable due to their high *match* and low *cd* values and partially due to their age distribution. If we compare the customer profiles on store-

[12]Like in store 242.

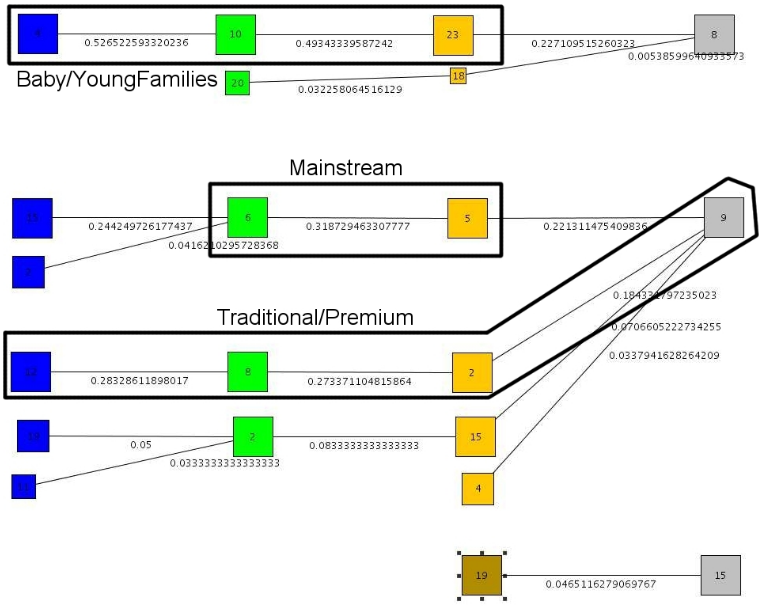


Figure 4.24: Best matches of 2006 (Q1-Q4) with customer intersection and minimum edge weight of 0.03, store 22.

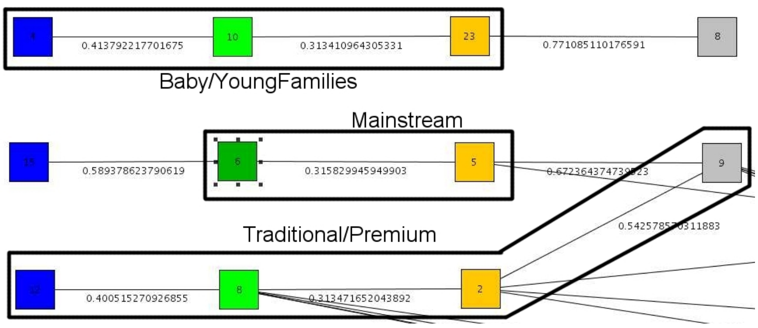


Figure 4.25: BestCDs of 2006 (Q1-Q4) with customer intersection and minimum edge weight of 0.03, store 22.

ClusterID	Year/Quarter	avg(Age)
4	06Q1	39.6
10	06Q2	40.0
23	06Q3	40.0
(15)	(06Q1)	(48.6)
6	06Q2	44.6
5	06Q3	43.3
12	06Q1	49.7
8	06Q2	54.4
2	06Q3	52.7
9	0604	56.6

Table 4.27: Age distribution of stable clusters, store 22.

level, we can see that their sub-assortment*brand combinations with the highest *centroid* component values mostly match and that partially their age distributions match, too. But we still wonder, if those profiles are really similar to each other. On store-level, calculating the *Bestmatch* clusters would not make any sense, because the stores have totally disjoint customer sets. However, calculating the *BestCD* clusters is practicable. Figure 4.26 shows the single clusters of the customer profiles *Traditional/Premium*, *Mainstream* and *Baby/Young Families* of store 242 and their corresponding *BestCD* clusters in store 22. Fortunately, in all four quarters each considered single cluster of store 242 and its *BestCD* cluster of store 22 both belong to the same customer profile. Furthermore, the *cd* values range from 0.35 to 0.79, which means that their *centroid* vectors are relatively similar, too. Note that the *cd* values of the *Traditional/Premium* clusters range from 0.37 to 0.44 and the *cd* values of the *Mainstream* clusters range from 0.35 to 0.42 while the *cd* values of the *Baby/Young Families* clusters range from 0.67 to 0.79. The higher *cd* values of the *Baby/Young Families* clusters can be explained by the slightly different assortments of stores 242 and 22. For instance, store 22 sells products of the sub-assortment*brand combination *Baby-Push/BrandG2* and store 242 does not. In *06*Q1 of store 22, this sub-assortment*brand combination even has the highest *centroid* component value of cluster 4 (48%). This huge difference of both *centroids* in the *centroid* component value of this sub-assortment*brand combination accounts for the higher *cd* value range of the *Baby/Young Families* clusters. So, if we compare customer profiles of different stores to each other, we should keep in mind that their assortments can be slightly different. It is better to compare either just stores with very similar assortments to each other or exclude the assortments' differences from the comparison.
Analog to the comparison of stores 242 and 22, Figure 4.27 shows the clusters of customer profiles and their *BestCD* clusters for the stores 518 and 22. Here, each cluster and its *BestCD* cluster both belong to the same customer profile, too. Their *cd* values range from 0.57 to 0.9.

ClusterID/ YearQuarter	Sub-assortment/Brand	*Centroid* Value [%]
12/06Q1	Kerzen (45)/BrandH2 (2076)	9.3%
	Filter&Folien (56)/BrandX1 (1243)	16.9% (1.)
	Haushaltspap (6)/BrandY1 (1265)	14.7%
	Haarpflege (5)/BrandZ1 (33)	13.9%
	Damenhygiene (3)/BrandA2 (2191)	8.5%
	Kosmetik-Pap (4)/BrandB2 (1241)	14.8% (4.)
8/06Q2	Kerzen (45)/BrandH2 (2076)	5.7%
	Filter&Folien (56)/BrandX1 (1243)	17.3% (4.)
	Haushaltspap (6)/BrandY1 (1265)	19.3% (3.)
	Haarpflege (5)/BrandZ1 (33)	16.9%
	Damenhygiene (3)/BrandA2 (2191)	14.9%
	Kosmetik-Pap (4)/BrandB2 (1241)	12.2%
2/06Q3	Kerzen (45)/BrandH2 (2076)	7.4
	Filter&Folien (56)/BrandX1 (1243)	15.9%
	Haushaltspap (6)/BrandY1 (1265)	18.0%
	Haarpflege (5)/BrandZ1 (33)	13.2%
	Damenhygiene (3)/BrandA2 (2191)	11.5%
	Kosmetik-Pap (4)/BrandB2 (1241)	8.6%
9/06Q4	Kerzen (45)/BrandH2 (2076)	18.0% (3.)
	Filter&Folien (56)/BrandX1 (1243)	14.0% (4.)
	Haushaltspap (6)/BrandY1 (1265)	11.4%
	Haarpflege (5)/BrandZ1 (33)	7.8%
	Damenhygiene (3)/BrandA2 (2191)	-
	Kosmetik-Pap (4)/BrandB2 (1241)	6.6%

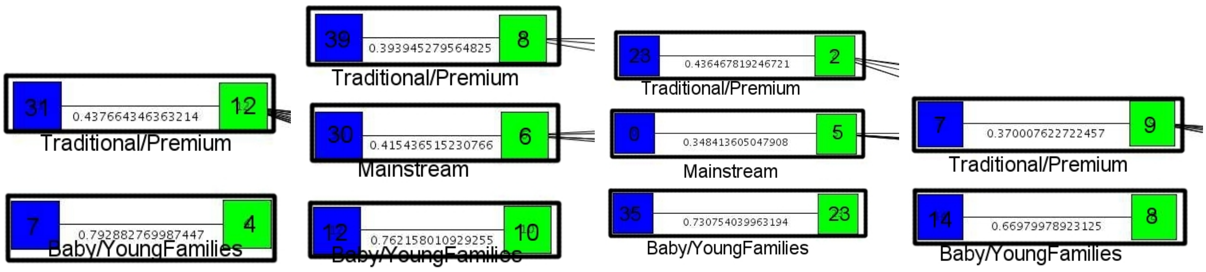
Table 4.28: Interesting *centroid* component values of *Traditional/Premium* stable cluster, store 22.

ClusterID/ YearQuarter	Sub-assortment/Brand	*Centroid* Value [%]
6/06Q2	Bonbon (40)/BrandC2 (1772)	13.8%
	Nassrasur (66)/*dm*-BrandD2 (218)	20.0%
	Sonne (48)/*dm*-BrandS1 (210)	41.2% (1.)
5/06Q3	Bonbon (40)/BrandC2 (1772)	18.3%
	Nassrasur (66)/*dm*-BrandD2 (218)	22.3% (4.)
	Sonne (48)/*dm*-BrandS1 (210)	39.6% (1.)

Table 4.29: Interesting *centroid* component values of *Mainstream* stable cluster, store 22.

ClusterID/ YearQuarter	Sub-assortment/Brand	*Centroid* Value [%]
4/06Q1	Windeln (27)/BrandE2 (899)	33.8% (3.)
	Babynahrung (33)/BrandA1 (1846)	18.1%
	Babyglasnahr (37)/BrandA1 (1846)	20.0%
	Babyglasnahr (37)/BrandL1 (1847)	16.1%
10/06Q2	Windeln (27)/BrandE2 (899)	37.3%
	Babynahrung (33)/BrandA1 (1846)	16.2%
	Babyglasnahr (37)/BrandA1 (1846)	24.5% (3.)
	Babyglasnahr (37)/BrandL1 (1847)	14.8%
23/06Q3	Windeln (27)/BrandE2 (899)	32.3% (2.)
	Babynahrung (33)/BrandA1 (1846)	14.6%
	Babyglasnahr (37)/BrandA1 (1846)	17.1%
	Babyglasnahr (37)/BrandL1 (1847)	14.8%

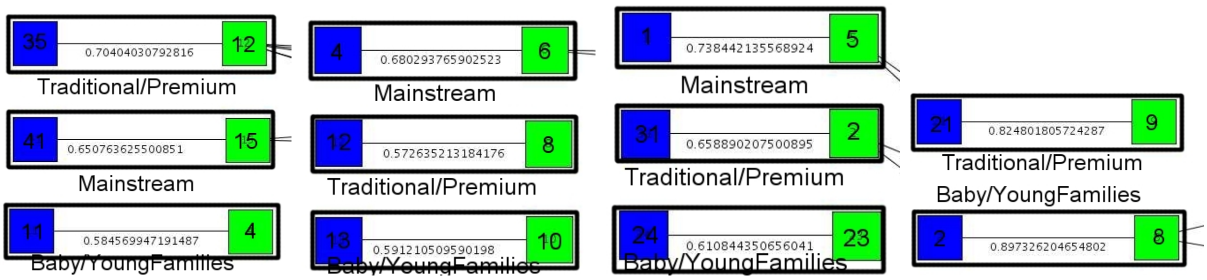
Table 4.30: Interesting *centroid* component values of *Baby/Young Families* stable cluster, store 22.



(a) *06Q1* (b) *06Q2* (c) *06Q3* (d) *06Q4*

Figure 4.26: Centroid distances of clusters from stores 242 and 22. The blue vertices represent single clusters of store 242 and the green vertices represent single clusters of store 22.

Again we find our results corroborated: All three customer profiles are not only stable due to *Bestmatch* and *BestCD* calculations and age distributions of different quarters in the same store, but also due to *BestCD* calculations of the same quarters in different stores. So the comparison of our extracted customer profiles on store-level further confirms the stabilities of those profiles.



(a) *06Q1* (b) *06Q2* (c) *06Q3* (d) *06Q4*

Figure 4.27: Centroid distances of clusters from stores 518 and 22. The blue vertices represent single clusters of store 518 and the green vertices represent single clusters of store 22.

Chapter 5

PMML

The Predictive Model Markup Language (*PMML*) is an XML-based markup language. It is used to describe data mining models like for example clustering models. *PMML* is developed and maintained by the Data Mining Group (*DMG*) (Gro). It offers a vendor-independent way to interchange data mining models for applications. So a user can develop a data mining model with one vendor's application, and use another vendor's application to further work with the model without having to deal with compatibility issues.
We use *PMML* to import our graph clustering results into the data mining component of the *ToolA* [1] software suite employed by *dm*. With this tool *dm* experts can easily classify customers based on their purchasing history (customer vector) in one of our clustering model's customer profiles.
A *PMML* document consists of three basic components. The *header* component contains document information like for example the model's copyright. The *data dictionary* component defines all the attributes (data fields) that are used by the model. It contains information about the attributes' data types, their value ranges and their optypes[2]. The *model* component defines the data mining model, which in our case is a clustering model.

```
<?xml version="1.0"?>
 <PMML version="3.2" xmlns="http://www.dmg.org/PMML-3_2">

   <Header copyright="dmg.org"/>
   <DataDictionary> ... </DataDictionary>
   <ClusteringModel ..> ... </ClusteringModel>

 </PMML>
```

A clustering model can either be center-based or distribution-based. It contains the following sub-components:

[1] A business intelligence, enterprise reporting, and OLAP software.
[2] Continuous, categorical, or ordinal.

mining schema which lists the attributes used in the model,

comparison measure which can either be a distance or a similarity measure,

clustering fields which are the *centroid* vector components and have to be consistent with the attribute names defined in the *data dictionary*,

center fields which contains normalizations of the *centroid* vector fields and

clusters which are the *centroid* vectors of the model's clusters.

Note that the normalization of the *centroid* vector fields is mostly used to map categorical input fields to numeric values. Thus, categorical fields are split into multiple dummy fields.
The following PMML example contains a clustering model that classifies people due to their education and gender. It is part of an example presented at a data mining workshop (MG08) hosted by *ToolA* .

```
<?xml version="1.0"?>
 <PMML version="3.2" xmlns="http://www.dmg.org/PMML-3_2">

   <Header copyright="Copyright (c) 2008 by \microstrategy, Inc."/>

   <DataDictionary numberOfFields="2">
     <DataField dataType="string" optype="categorical" name="Education">
       <Value value="Graduate"/>
       <Value value="High School"/>
       <Value value="Other"/>
       <Value value="Undergraduate"/>
     </DataField>
     <DataField dataType="string" optype="categorical" name="Gender">
       <Value value="Female"/>
       <Value value="Male"/>
     </DataField>
   </DataDictionary>

   <ClusteringModel numberOfClusters="2" modelClass="centerBased"
    modelName="CustomerProfiles">

     <MiningSchema>
       <MiningField name="Education"/>
       <MiningField name="Gender"/>
     </MiningSchema>

     <ComparisonMeasure kind="distance">
       <squaredEuclidean/>
     </ComparisonMeasure>

     <ClusteringField field="Education" compareFunction="absDiff"/>
     <ClusteringField field="Gender" compareFunction="absDiff"/>

     <CenterFields>
       <DerivedField dataType="double" optype="continuous" name="Education=Graduate">
         <NormDiscrete field="Education" value="Graduate"/>
       </DerivedField>
       <DerivedField dataType="double" optype="continuous" name="Education=High School">
         <NormDiscrete field="Education" value="High School"/>
       </DerivedField>
       <DerivedField dataType="double" optype="continuous" name="Education=Other">
         <NormDiscrete field="Education" value="Other"/>
       </DerivedField>
       <DerivedField dataType="double" optype="continuous" name="Education=Undergraduate">
         <NormDiscrete field="Education" value="Undergraduate"/>
       </DerivedField>
       <DerivedField dataType="double" optype="continuous" name="Gender=Female">
         <NormDiscrete field="Gender" value="Female"/>
       </DerivedField>
       <DerivedField dataType="double" optype="continuous" name="Gender=Male">
         <NormDiscrete field="Gender" value="Male"/>
       </DerivedField>
     </CenterFields>

     <Cluster name="Cluster 1">
       <Array n="6" type="real">0.219563 0.412071 0.0863684 0.281998 0.44641 0.55359</Array>
```

```
    </Cluster>
    <Cluster name="Cluster 2">
      <Array n="6" type="real">0.216797 0.401367 0.100586 0.28125 0.429688 0.570312</Array>
    </Cluster>

  </ClusteringModel>

</PMML>
```

The *data dictionary* component describes the two categorical attributes *education* and *gender*. For instance, the categorical attribute *gender* has the data type string and can take the values *Female* or *Male*. This clustering model is center-based and uses all the attributes defined in the *data dictionary*. The *comparison measure* of this model is the Euclidean distance. Each person is described by a binary vector with six vector components. The cluster *centroids* are calculated by the component-wise average of the people vectors. Note that in this example everybody takes a value for every attribute. So the sum of all *centroid* vector components that belong to one attribute is 1.
Next, we describe the *PMML* document's basic layout of our clustering results. As already mentioned, the *data dictionary* component of every *PMML* document defines the attributes used by the model. In the case of the clustered *KhK*-graph the attributes are the 121 sub-assortments and the attributes' values are the brands of the sub-assortments. So, for instance the *data field* of the attribute *Babynahrung* has the following structure:

```
<DataDictionary numberOfFields="121" >
  ..
  <DataField dataType="string" optype="categorical" name="Babynahrung" >
    <Value value="?" />
    <Value value="BrandA1" />
    <Value value="BrandU1" />
    <Value value="BrandV1" />
    <Value value="BrandW1" />
    <Value value="dm-BrandQ1" />
    <Value value="BrandL1" />
    ...
  </DataField>
  ..
</DataDictionary>
```

The *mining schema* sub-component lists all sub-assortments:

```
  <MiningSchema >
    ...
    <MiningField name="Baby-Pull"/>
    <MiningField name="Baby-Push"/>
    <MiningField name="Babyglasnahr"/>
    <MiningField name="Babynahrung"/>
    ...
  </MiningSchema>
```

The *clustering fields* are the *centroid* vector components and contain the sub-assortments, too:

```
<ClusteringField field="Baby-Pull" compareFunction="absDiff"/>
<ClusteringField field="Baby-Push" compareFunction="absDiff"/>
<ClusteringField field="Babyglasnahr" compareFunction="absDiff"/>
<ClusteringField field="Babynahrung" compareFunction="absDiff"/>
```

Like in the previous example, our clustering model is center-based and the model's *comparison measure* is the Euclidean distance. Due to the mapping of categorical input fields to numeric values, the categorical sub-assortment attributes are split into [redacted] sub-assortments*brand combination fields. So, there are [redacted] *center fields* like the following:

```
<CenterFields>
  ...
  <DerivedField dataType="double" optype="continuous" name="c82" >
    <NormDiscrete field="Babynahrung" value="?" />
  </DerivedField>
  <DerivedField dataType="double" optype="continuous" name="c84" >
    <NormDiscrete field="Babynahrung" value="BrandA1" />
  </DerivedField>
  <DerivedField dataType="double" optype="continuous" name="c86" >
    <NormDiscrete field="Babynahrung" value="BrandU1" />
  </DerivedField>
  <DerivedField dataType="double" optype="continuous" name="c89" >
    <NormDiscrete field="Babynahrung" value="BrandV1" />
  </DerivedField>
  <DerivedField dataType="double" optype="continuous" name="c90" >
    <NormDiscrete field="Babynahrung" value="BrandW1" />
  </DerivedField>
  <DerivedField dataType="double" optype="continuous" name="c91" >
    <NormDiscrete field="Babynahrung" value="dm-BrandQ1" />
  </DerivedField>
  <DerivedField dataType="double" optype="continuous" name="c95" >
    <NormDiscrete field="Babynahrung" value="BrandL1" />
  </DerivedField>
  ...
</CenterFields>
```

Analog to Section 3.3.2, the single customer vector is binary, consists of [redacted] components and has at most one '1' per sub-assortment (that stands for the customer's favorite brand). The *centroid* vector of a certain cluster is the average customer vector. Note that unlike in the previous example, the sum of all *centroid* vector components that belong to one attribute does not have to be 1, because not every customer buys items of all sub-assortments. The sub-component *clusters*, that contains those *centroid* vectors has the following structure:

```
  <Cluster name="Cluster 0">
    <Array n="# sub-assortment*brand combinations" type="real">0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.025641 0.0 0.025641 0.0 ... 0.0</Array>
  </Cluster>
  <Cluster name="Cluster 1">
    <Array n="# sub-assortment*brand combinations" type="real">0.0 0.0 0.0
0.033333 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 ... 0.0</Array>
  </Cluster>
  ...
```

Chapter 6

Conclusion

The main purpose of this work was a case study of extracting substantiated customer profiles out of loyalty card data of *dm* customers by means of graph clustering techniques. Roughly speaking, we modeled the data as a network of customers and employed graph clustering methods to find groups of customers with similar purchasing behaviors. We first designed, implemented and analyzed different promising graph models that represent similar shopping behaviors with different strengths and weaknesses. We began with the *KAK*-graph model which represents the customer–item–customer relation (modeling the similarity on item-level). We analyzed three different *time-expanded* approaches of this model. We came to the conclusion that the *KAK*-graph model favors sub-assortments like *Pet Food* or *Baby Food* due to their special characteristics: Customers that buy *Pet Food* or *Baby Food* products, frequently buy several flavors of the same product at once, which exaggerates ties between such customers. So, we continued with the *KMK*-graph model, which is not prone to favoring sub-assortments with such characteristics. This model represents the customer–sub-assortment*brand–customer relation. In this graph model, items in the same sub-assortment and of the same brand are merged, and thus an increased edge weight due to several different flavors and many small packages is avoided. For that purpose, binary customer vectors are used. After analyzing different clustered *KMK*-graphs, we found out that their clusters reflect the average and rather general *dm* customer. So, compared to *KAK*-graphs *KMK*-graphs do not favor very frequently bought items, but on the other hand their clusters are characterized by sub-assortments and brands like those containing items such as plastic bags that are not appropriate to discriminate between *dm* customers. Thus, we carried on with the analysis of the *KhK*-graph which represents the customer–sub-assortment*favorite-brand–customer relation. It differs from the *KMK*-graph in that it just considers the customers' favorite brands of each sub-assortment. In this way, we wanted to keep the *KMK*-graph's advantage of not favoring frequently bought items while trying to sharpen the expressiveness of the resulting customer clusters by ignoring items that are bought by everybody. The first attempts of clustered *KhK*-graphs yielded four huge unsatisfying customer clusters of very low significance that were very similar to each other. But after examining these huge similar clusters, we found out that they were extremely dominated by specific popular *dm*-owned brands like e.g. *dm-BrandK1* or *dm-BrandN1* . So the four clusters represented one overall *dm* customer profile. We found a solution by eliminating similarities based on sub-assortment*brand combinations containing these brands and received a clustering of high quality that consisted of several meaningful mid-sized clusters.

Having found a promising graph model that yields reasonable clusterings on one probational quarter, we next tried to identify stable customer profiles, that is, customer clusters that verifiably occur at different points in time, using the data from subsequent quarters in a whole year. The similarity of a customer profile's single clusters was verified by four different comparisons: The single clusters of a customer profile needed to have a similar customer base (high set-*match* values), we required their customers' buying patterns to be similar (low *centroid* distance values), their customers' favorite sub-assortment*brand combinations needed to be alike (similar highest *centroid* vector components) and their customers' age distribution needed to be revealing (small variance of average ages). After calculating the customer intersection of the four quarters of 2006 and reducing the minimum required edge weight (i.e., similarity) for a customer to participate, we finally received the four stable customer profiles *Traditional/Premium*, *Mainstream*, *Baby/Young Families* and *GreenConcerns*.

To this point our results were based on the data of a single store. We confirmed our findings with a final test which broadened the set of data we base our results on. We tried to extract the same customer profiles out of the data of two other, totally different stores and compared those customer profiles on store-level. We attained the same results with the same success by using the same techniques!

On our journey through different models and approaches, we attained a number of secondary results, e.g. the high brand loyalty among *Cat Food* and *Baby Food* customers in the *KAK*-graphs or that some favorite sub-assortment*brand combinations might vary from store to store while still being very closely related, but nevertheless remain specific to a certain profile[1].

We developed several software tools during our process of extracting customer profiles, e.g. for the *time-expanded* graph generation, for the generation of *KMK*- and *KhK*-graphs and for the analysis and comparison of single clusters.

Experts of *dm* corroborated the value of this work. They examined our results and came, among others, to the following conclusion: The biggest benefit lies in the sophisticated and detailed characterization of the customers' purchasing behaviors. Prior approaches just tried to divide customers into different stereotyped categories, which simplified their purchasing behaviors too much, but they did not consider that even a customer who clearly belongs to a certain category buys most likely small amounts of items that are not category-specific. E.g. a customer who has a baby is not only a parent and thus buys not only baby products. Currently, there is a *dm*-internal project in operation that tries to identify *dm* customer segments. Our results are a valuable contribution to this project.

[1]Remember the *Kerzen/BrandR1* versus the *Kerzen/BrandH2* sub-assortment*brand combination in the *Traditional/Premium* customer profile or the *Baby-Push/BrandG2* combination of the *Baby/Young Families* customer profile.

6.1 Outlook

The many insights into the buying patterns of *dm* customers provided by this work revealed some open questions that should be addressed in future work:

1. Are there sub-profiles inside our four stable customer profiles?

   One possible way would be to separately analyze the stable customer profiles by just considering the customer set of one customer profile.

   Another possibility would be to refine the sub-assortment*brand combination. Analyzing sub-sub-assortment* brand combinations might yield interesting sub-profiles.

2. Can further knowledge be gained about the customer profiles by evaluating the *avgdist* comparison measurement for clusters, introduced in Section 3.3.2?

   Avgdist measures the average distance of a cluster's customer vectors to their *centroid* vector. Because this comparison measurement is not sufficient and due to the limited time we did not compare the single clusters of our customer profiles with their *avgdist* values.

3. Are there more systematic ways to determine the sub-assortment*brand combinations that have to be eliminated?

   We assume that those sub-assortment*brand combinations have to be determined manually by *dm* experts who employ background knowledge (and a fair amount of common sense). Because the single sub-assortments differ a lot in size, it would probably make sense to partition the huge sub-assortments.
   An automated procedure to determine the sub-assortment*brand combinations that have to be eliminated is not necessary because the (popular) sub-assortment*brand combinations do not change a lot over the years. It is enough to determine those combinations once and for all.

4. Which customers did we catch and which did we miss?

   Remember that only about 50 percent of *dm* customers have a *PAYBACK* card. Furthermore, we apply several thresholds and some local and temporal restrictions to the data. In what way, if at all, does this bias our results?

6.2 A Brief Note on Two Promising Applications

Although this was no central point of this work, we suggest two different obvious applications for the identified stable customer profiles, and the methods and models developed by us:

1. A customer can be classified into one of the customer profiles. Then the differences of this customer to the profiles's average customers are determined, that is roughly speaking the items which he does not buy but the rest of the profile's customers does. Then, these items are individually promoted for this customer.
   The rationale behind this is that the fact that many customers of this profile buy these items encourages the assumption that this customer might also have an inclination to buy these items, but for some reason did not do so, yet.

2. A product launch scenario:

   (a) Arrange a test run with some selected stores that sell this product on trial.

   (b) Gather all customers that bought this product, calculate their average customer vector and treat this average customer vector as the prototypical buyer of the new product.

   (c) When launching the full release in other stores, call the attention of those customers, whose customer vectors resemble the average customer vector, to the new product.

Appendix A

KAK-Graph of the *KhK*-Example

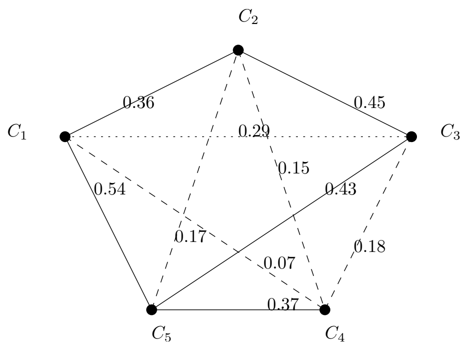


Figure A.1: *KAK*-graph of the *KhK*-example.

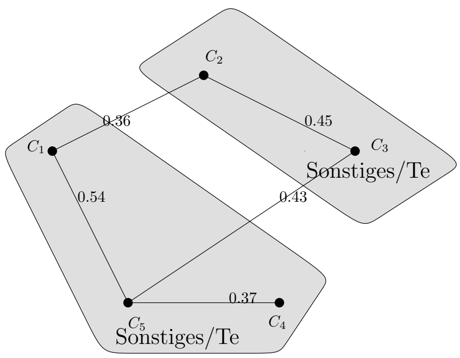


Figure A.2: Clustered *KAK*-graph of the *KhK*-example with minimum edge weight of 0.3.

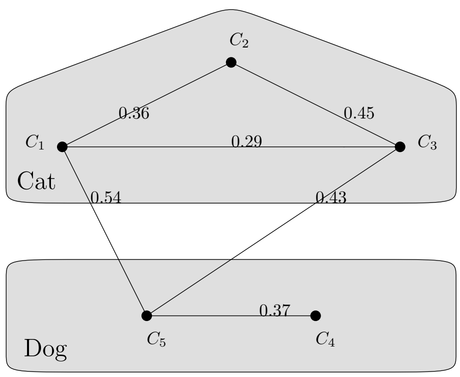


Figure A.3: Clustered *KAK*-graph of the *KhK*-example with minimum edge weight of 0.2.

Appendix B

Time-Expanded Graph

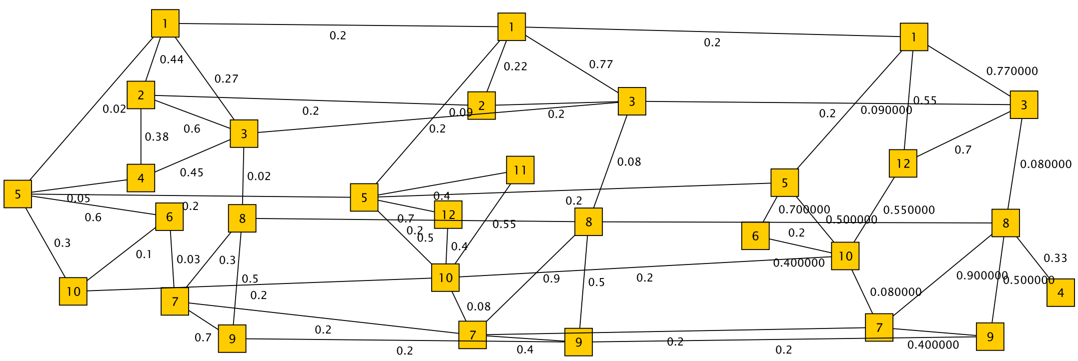


Figure B.1: *Time-expanded* graph with alpha. See Section 3.3.

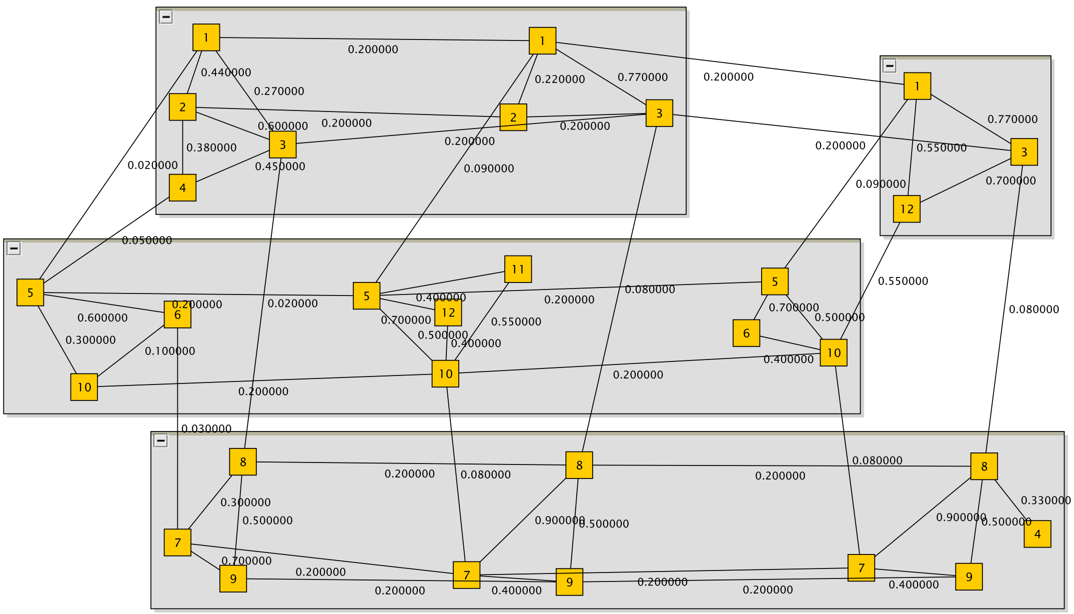


Figure B.2: Clustered *time-expanded* graph, alpha. See Section 3.3.

Appendix C

KhK Graphs of Store 242

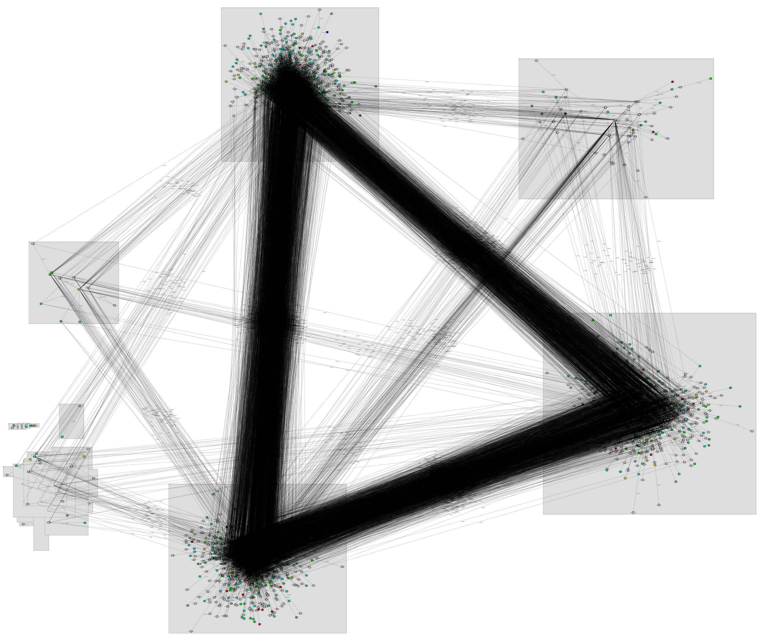


Figure C.1: Clustered *KhK*-graph of Q1 2006.

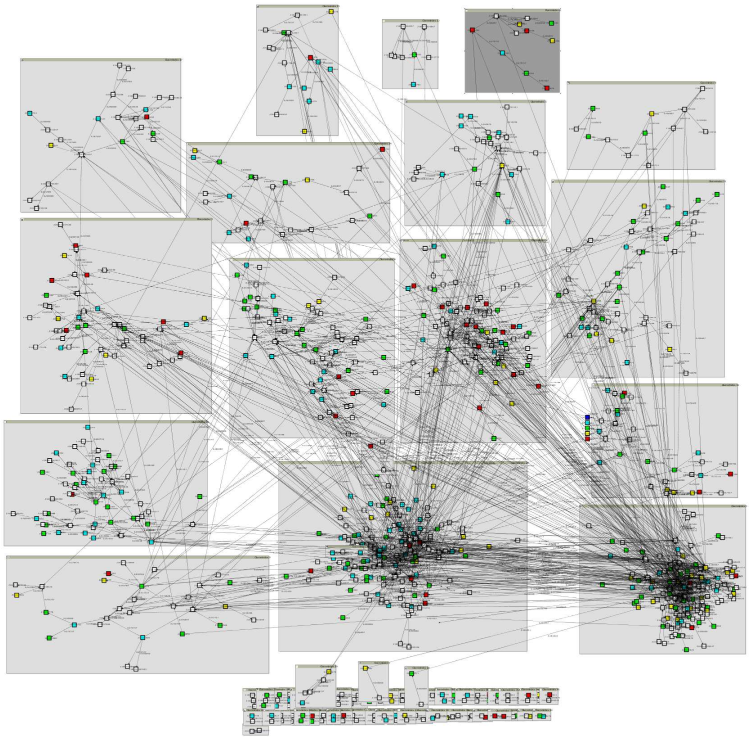


Figure C.2: Clustered *KhK*-graph of *04Q4* with eliminations.

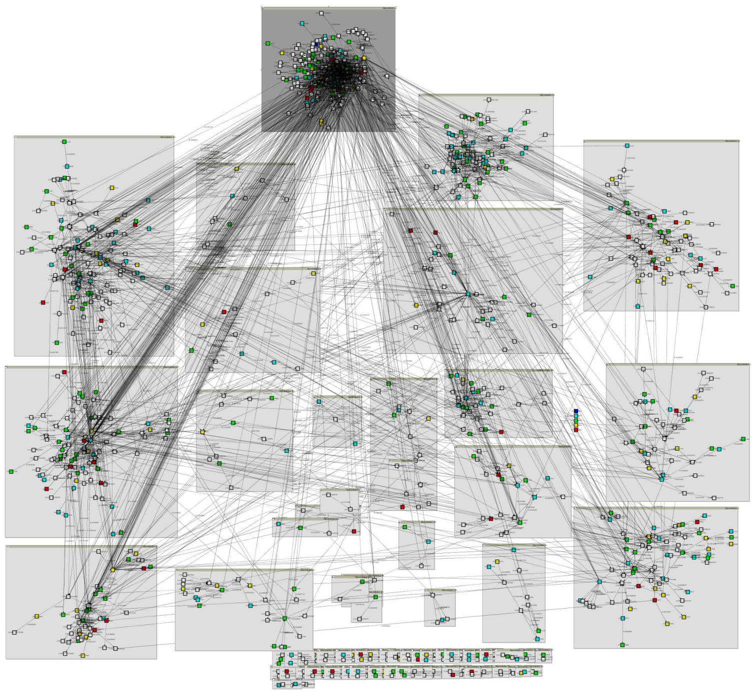


Figure C.3: Clustered *KhK*-graph of*05Q4* with eliminations.

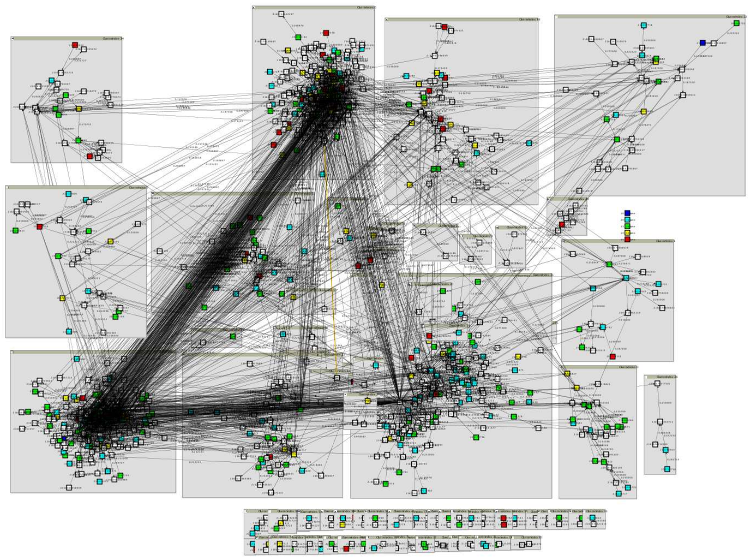


Figure C.4: Clustered *KhK*-graph of *06Q4* with eliminations.

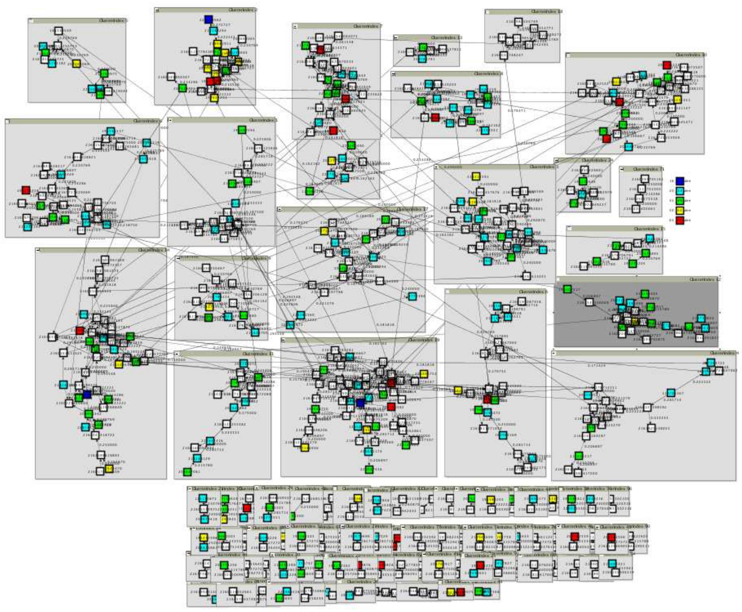


Figure C.5: Clustered *KhK*-graph of *06Q3* with eliminations.

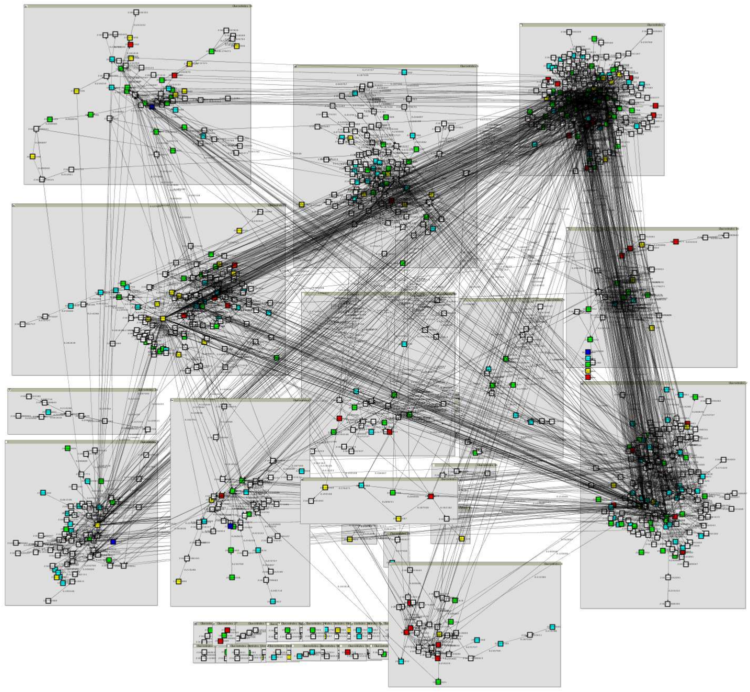


Figure C.6: Clustered *KhK*-graph of *07Q4* with eliminations.

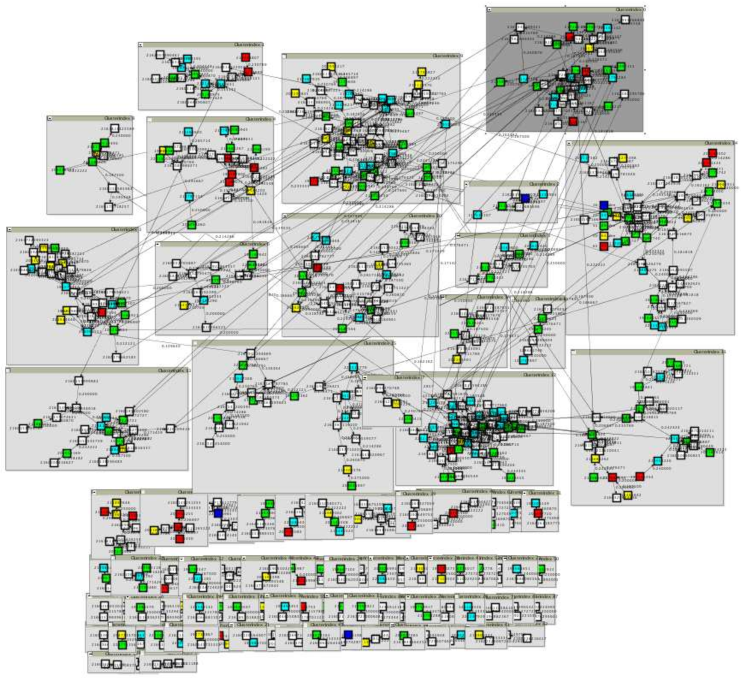


Figure C.7: Clustered *KhK*-graph of *06Q2* with eliminations.

Quality Measurement	cov_w	per	$sig_{cov,sub}$
06Q1	0.62	0.71	0.27
04Q4 with Eliminations	0.73	0.92	0.61
05Q4 with Eliminations	0.71	0.93	0.58
06Q4 with Eliminations	0.69	0.89	0.49
06Q3 with Eliminations	0.91	0.97	0.86
07Q4 with Eliminations	0.68	0.89	0.52
06Q2 with Eliminations	0.92	0.96	0.85

Table C.1: Quality measurements of KhK-graphs in store 242.

Appendix D

Bestmatch- and *BestCD*-Graphs of Store 242

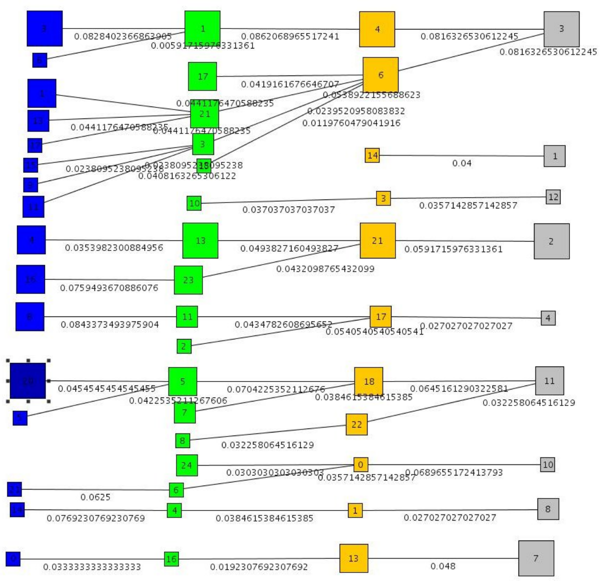


Figure D.1: *Bestmatch*-graph of Q4 (2004-2007).

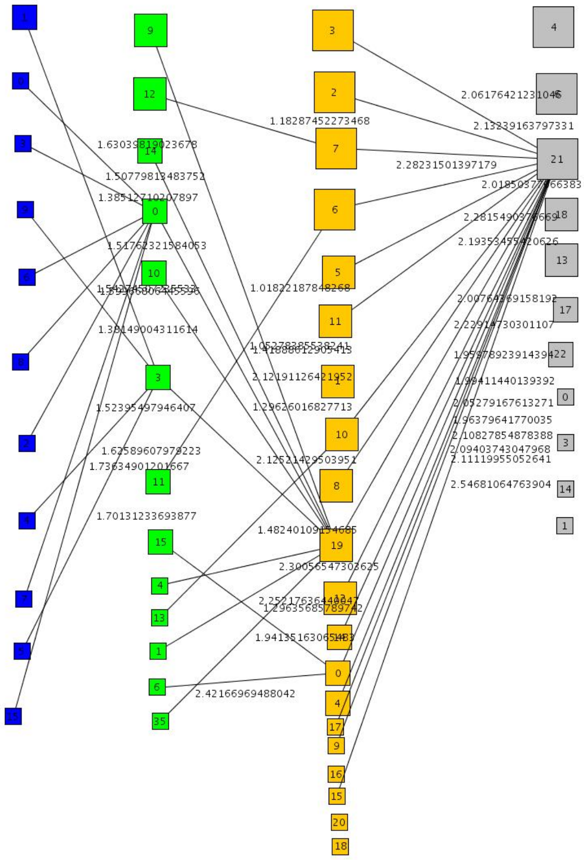


Figure D.2: *BestCD*-graph of 2006.

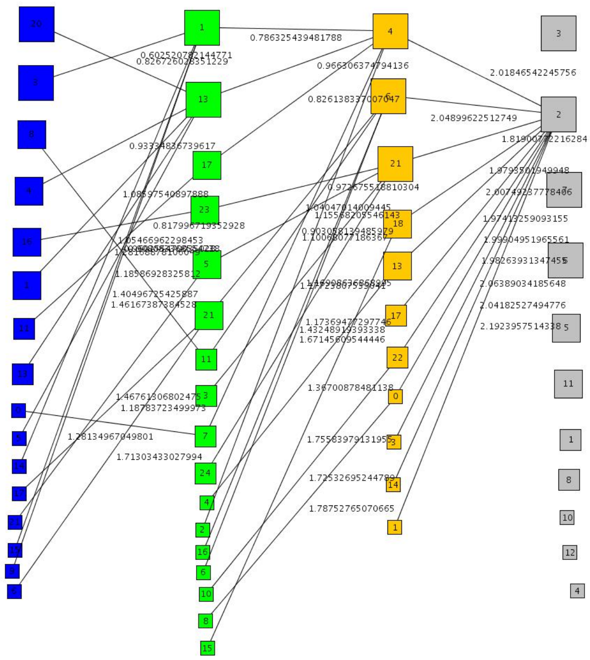


Figure D.3: *BestCD*-graph of Q4 (2004-2007).

Appendix E

Details to Stable Clusters, Store 242

E.1 Eliminated Items in *KhK*-Graphs

Popular *dm*-owned brands:

Sub-Assortment ID	Brand ID	count(ClusterID)	Sub-Assortment	Brand
81	731	24	green-BrandU1	green-BrandU1
4	213	21	Kosmetik-Pap	dm-BrandO1
15	202	20	WPR	dm-BrandD
84	-1	20	Sonstiges/Te	?
19	214	17	Hygienepapie	dm-BrandN1
18	212	15	Haush.artik.	[redacted]
78	381	15	Pharma	[redacted]
56	212	15	Filter&Folien	[redacted]
60	207	15	Gesicht	dm-BrandK1
70	-1	14	Fotoarbeiten	?
54	207	14	Bad	dm-BrandK1
11	207	13	Haarstyling	dm-BrandK1
66	785	13	Nassrasur	[redacted]
6	214	13	Haushaltspap	dm-BrandN1
30	206	13	Babypflege	dm-BrandQ1
3	208	12	Damenhygiene	[redacted]
42	211	12	Mund/Zahn	dm-BrandP1
72	1214	12	Tee	[redacted]
125	200	12	Ebelin Fash	[redacted]
57	94	12	Deo	[redacted]
45	212	11	Kerzen	[redacted]
57	207	11	Deo	dm-BrandK1
5	207	11	Haarpflege	dm-BrandK1

39	207	11	Hand	dm-BrandK1
72	381	11	Tee	[redacted]
40	1787	11	Bonbon	[redacted]
124	200	11	Ebel. Beaut+Manik	[redacted]
58	129	10	Erlebnis dm	[redacted]
57	4	10	Deo	[redacted]
118	1309	10	Make up	[redacted]
75	307	10	Reform	[redacted]
39	4	9	Hand	[redacted]
39	86	9	Hand	[redacted]
3	727	9	Damenhygiene	[redacted]
74	205	9	Filme	[redacted]
42	761	9	Mund/Zahn	[redacted]
42	7	9	Mund/Zahn	[redacted]
30	57	9	Babypflege	[redacted]
73	206	9	Babyzubehr	dm-BrandQ1
128	200	9	Ebelin Hair	[redacted]
129	200	9	EBEL HS ST K	[redacted]
60	82	9	Gesicht	[redacted]
13	219	9	Katze	dm-BrandB1
66	207	9	Nassrasur	dm-BrandK1
60	745	8	Gesicht	[redacted]
78	3	8	Pharma	[redacted]
121	1290	8	Nagelpflege	[redacted]
95	920	8	Accessoires	[redacted]
119	1309	8	Augen Make u	[redacted]
60	201	8	Gesicht	dm-green-BrandT1
40	2248	8	Bonbon	[redacted] [redacted]
61	207	8	Bodylotion	dm-BrandK1
37	1848	8	Babyglasnahr	BrandV1
54	172	8	Bad	[redacted] [redacted]
130	200	8	EBEL HS ST G	[redacted]
33	731	8	Babynahrung	green-BrandU1
26	37	8	Flaschen & S	[redacted]
123	-1	8	Cactus Gruk	?
58	756	8	Erlebnis dm	[redacted]
11	4	8	Haarstyling	[redacted]
51	62	8	Fu	[redacted]
66	175	8	Nassrasur	[redacted] [redacted]
3	768	8	Damenhygiene	[redacted]
27	206	8	Windeln	dm-BrandQ1

E.2 *Centroids* of Stable Clusters

All *centroid* values greater than 5%.

E.2.1 The *Traditional/Premium* Cluster

06Q1, Cluster 31

Cluster	Quarter	*Centroid* Value	Sub-Assortment		Brand	
31	06Q1	0.292863762743281	112	Ostern klein	-1	?
31	06Q1	0.169601482854495	5	Haarpflege	33	BrandZ1
31	06Q1	0.15291936978684	6	Haushaltspap	1265	BrandY1
31	06Q1	0.151065801668211	4	Kosmetik-Pap	1241	BrandB2
31	06Q1	0.148285449490269	3	Damenhygiene	2191	BrandA2
31	06Q1	0.145505097312326	127	Ebelin Body	200	
31	06Q1	0.138090824837813	18	Haush.artik.	1884	
31	06Q1	0.133456904541242	119	Augen Make u	29	
31	06Q1	0.129749768303985	51	Fu	207	dm-BrandK1
31	06Q1	0.126969416126043	61	Bodylotion	4	
31	06Q1	0.1241890639481	60	Gesicht	176	
31	06Q1	0.120481927710843	16	Duftkerzen-M	433	
31	06Q1	0.120481927710843	45	Kerzen	2075	BrandR1
31	06Q1	0.113994439295644	11	Haarstyling	65	
31	06Q1	0.11306765523633	42	Mund/Zahn	187	
31	06Q1	0.103799814643188	56	Filter&Folien	1243	BrandX1
31	06Q1	0.100092678405931	39	Hand	257	
31	06Q1	0.0991658943466173	42	Mund/Zahn	32	
31	06Q1	0.0963855421686747	113	Ostern mitte	-1	?
31	06Q1	0.092678405931418	19	Hygienepapie	1275	
31	06Q1	0.0908248378127896	39	Hand	201	dm-green-BrandT1
31	06Q1	0.0871177015755329	45	Kerzen	2076	BrandH2
31	06Q1	0.0787766450417053	57	Deo	5	
31	06Q1	0.0787766450417053	42	Mund/Zahn	374	
31	06Q1	0.0778498609823911	30	Babypflege	59	
31	06Q1	0.0750695088044486	15	WPR	262	
31	06Q1	0.0732159406858202	53	Posten/Saiso	-1	?
31	06Q1	0.0704355885078777	40	Bonbon	1772	BrandC2
31	06Q1	0.0695088044485635	15	WPR	270	
31	06Q1	0.0695088044485635	3	Damenhygiene	267	
31	06Q1	0.0667284522706209	58	Erlebnis dm	1265	BrandY1
31	06Q1	0.0667284522706209	110	Feinstrick 1	1988	
31	06Q1	0.0630213160333642	5	Haarpflege	66	

31	06Q1	0.06209453197405	36	Entferner	213	dm-BrandO1
31	06Q1	0.06209453197405	54	Bad	257	[redacted]
31	06Q1	0.0611677479147359	56	Filter&Folien	1245	[redacted]
31	06Q1	0.0611677479147359	15	WPR	249	[redacted]
31	06Q1	0.0593141797961075	60	Gesicht	1365	[redacted]
31	06Q1	0.0593141797961075	15	WPR	439	[redacted]
31	06Q1	0.0593141797961075	54	Bad	27	[redacted]
31	06Q1	0.0583873957367933	66	Nassrasur	218	dm-BrandK1 [redacted]
31	06Q1	0.0583873957367933	119	Augen Make u	79	[redacted]
31	06Q1	0.0574606116774791	60	Gesicht	174	[redacted]
31	06Q1	0.0556070435588508	54	Bad	1474	BrandM1
31	06Q1	0.0546802594995366	5	Haarpflege	91	[redacted]
31	06Q1	0.0546802594995366	72	Tee	1213	[redacted]
31	06Q1	0.0546802594995366	11	Haarstyling	756	[redacted]
31	06Q1	0.0537534754402224	19	Hygienepapie	1264	[redacted]
31	06Q1	0.0537534754402224	61	Bodylotion	257	[redacted]
31	06Q1	0.0518999073215941	34	Fotozubehr	2189	[redacted]
31	06Q1	0.0518999073215941	119	Augen Make u	1290	[redacted]
31	06Q1	0.0509731232622799	75	Reform	381	[redacted]
31	06Q1	0.0509731232622799	40	Bonbon	381	[redacted]

06Q2, Cluster 39

Cluster	Quarter	*Centroid* Value	Sub-Assortment		Brand	
39	06Q2	0.283763277693475	45	Kerzen	2075	BrandR1
39	06Q2	0.22154779969651	61	Bodylotion	4	[redacted]
39	06Q2	0.213960546282246	18	Haush.artik.	1884	[redacted]
39	06Q2	0.195751138088012	6	Haushaltspap	1265	BrandY1
39	06Q2	0.160091047040971	5	Haarpflege	33	BrandZ1
39	06Q2	0.152503793626707	56	Filter&Folien	1243	BrandX1
39	06Q2	0.152503793626707	3	Damenhygiene	2191	BrandA2
39	06Q2	0.127465857359636	42	Mund/Zahn	187	[redacted]
39	06Q2	0.126707132018209	4	Kosmetik-Pap	1241	BrandB2
39	06Q2	0.125189681335357	60	Gesicht	176	[redacted] [redacted]
39	06Q2	0.122913505311077	11	Haarstyling	65	[redacted]
39	06Q2	0.10546282245827	42	Mund/Zahn	32	[redacted]
39	06Q2	0.0971168437025797	48	Sonne	210	dm-BrandS1
39	06Q2	0.0971168437025797	39	Hand	257	[redacted]
39	06Q2	0.0955993930197269	15	WPR	262	[redacted]
39	06Q2	0.0940819423368741	127	Ebelin Body	200	[redacted]
39	06Q2	0.090288315629742	51	Fu	207	dm-BrandK1
39	06Q2	0.0880121396054628	30	Babypflege	59	[redacted]

39	06Q2	0.0857359635811836	60	Gesicht	1365	[redacted]
39	06Q2	0.082701062215478	119	Augen Make u	29	[redacted]
39	06Q2	0.0773899848254932	58	Erlebnis dm	1265	BrandY1
39	06Q2	0.0773899848254932	42	Mund/Zahn	1140	[redacted]
39	06Q2	0.0751138088012139	36	Entferner	213	dm-BrandO1
39	06Q2	0.0743550834597876	19	Hygienepapie	1275	[redacted]
39	06Q2	0.0743550834597876	39	Hand	172	[redacted] [redacted]
39	06Q2	0.0705614567526555	57	Deo	5	[redacted]
39	06Q2	0.0705614567526555	54	Bad	782	[redacted]
39	06Q2	0.0690440060698027	40	Bonbon	1772	BrandC2
39	06Q2	0.0660091047040971	15	WPR	439	[redacted]
39	06Q2	0.0652503793626707	51	Fu	69	[redacted]
39	06Q2	0.0652503793626707	56	Filter&Folien	1245	[redacted]
39	06Q2	0.0644916540212443	12	Insekt/Pflan	223	[redacted]
39	06Q2	0.0637329286798179	112	Ostern klein	-1	?
39	06Q2	0.0637329286798179	54	Bad	257	[redacted]
39	06Q2	0.0606980273141123	5	Haarpflege	91	[redacted]
39	06Q2	0.0599393019726859	5	Haarpflege	113	[redacted]
39	06Q2	0.0591805766312595	39	Hand	201	dm-green-BrandT1
39	06Q2	0.0561456752655539	58	Erlebnis dm	32	[redacted]
39	06Q2	0.0553869499241275	3	Damenhygiene	267	[redacted]
39	06Q2	0.0553869499241275	16	Duftkerzen-M	433	[redacted]
39	06Q2	0.0546282245827011	40	Bonbon	1793	[redacted]
39	06Q2	0.0546282245827011	75	Reform	381	[redacted]
39	06Q2	0.0546282245827011	66	Nassrasur	218	dm-BrandK1 [redacted]
39	06Q2	0.0538694992412747	110	Feinstrick 1	1988	[redacted]
39	06Q2	0.0531107738998482	15	WPR	270	[redacted]
39	06Q2	0.0523520485584218	11	Haarstyling	908	[redacted]
39	06Q2	0.0515933232169954	74	Filme	861	[redacted]
39	06Q2	0.0500758725341426	72	Tee	1213	[redacted]
39	06Q2	0.0500758725341426	78	Pharma	897	[redacted]

***06Q3*, Cluster 23**

Cluster	Quarter	*Centroid* Value	Sub-Assortment		Brand	
23	06Q3	0.253041362530414	45	Kerzen	2075	BrandR1
23	06Q3	0.234387672343877	61	Bodylotion	4	[redacted]
23	06Q3	0.192214111922141	3	Damenhygiene	2191	BrandA2
23	06Q3	0.180048661800487	60	Gesicht	176	[redacted] [redacted]
23	06Q3	0.156528791565288	11	Haarstyling	65	[redacted]
23	06Q3	0.156528791565288	5	Haarpflege	33	BrandZ1
23	06Q3	0.145985401459854	6	Haushaltspap	1265	BrandY1

23	06Q3	0.145174371451744	18	Haush.artik.	1884	[redacted]
23	06Q3	0.140308191403082	36	Entferner	213	dm-BrandO1
23	06Q3	0.138686131386861	42	Mund/Zahn	187	[redacted]
23	06Q3	0.131386861313869	42	Mund/Zahn	32	[redacted]
23	06Q3	0.131386861313869	127	Ebelin Body	200	[redacted]
23	06Q3	0.120032441200324	4	Kosmetik-Pap	1241	BrandB2
23	06Q3	0.118410381184104	56	Filter&Folien	1243	BrandX1
23	06Q3	0.111922141119221	51	Fu	207	dm-BrandK1
23	06Q3	0.106244931062449	48	Sonne	210	dm-BrandS1
23	06Q3	0.105433901054339	30	Babypflege	59	[redacted]
23	06Q3	0.0916463909164639	15	WPR	262	[redacted]
23	06Q3	0.0908353609083536	39	Hand	172	[redacted]
23	06Q3	0.089213300892133	42	Mund/Zahn	1140	[redacted]
23	06Q3	0.0738037307380373	57	Deo	5	[redacted]
23	06Q3	0.0721816707218167	54	Bad	782	[redacted]
23	06Q3	0.0705596107055961	39	Hand	257	[redacted]
23	06Q3	0.0681265206812652	15	WPR	249	[redacted]
23	06Q3	0.0681265206812652	19	Hygienepapie	1264	[redacted]
23	06Q3	0.0673154906731549	54	Bad	257	[redacted]
23	06Q3	0.0673154906731549	5	Haarpflege	66	[redacted]
23	06Q3	0.0656934306569343	58	Erlebnis dm	32	[redacted]
23	06Q3	0.0640713706407137	40	Bonbon	381	[redacted]
23	06Q3	0.0640713706407137	40	Bonbon	1772	BrandC2
23	06Q3	0.0632603406326034	75	Reform	306	[redacted]
23	06Q3	0.0624493106244931	56	Filter&Folien	1245	[redacted]
23	06Q3	0.0624493106244931	66	Nassrasur	218	dm-BrandK1 [redacted]
23	06Q3	0.0600162206001622	39	Hand	201	dm-green-BrandT1
23	06Q3	0.0600162206001622	58	Erlebnis dm	1265	BrandY1
23	06Q3	0.0592051905920519	60	Gesicht	1365	[redacted]
23	06Q3	0.0559610705596107	36	Entferner	207	dm-BrandK1
23	06Q3	0.0559610705596107	74	Filme	861	[redacted]
23	06Q3	0.0559610705596107	119	Augen Make u	29	[redacted]
23	06Q3	0.0551500405515004	5	Haarpflege	113	[redacted]
23	06Q3	0.0535279805352798	61	Bodylotion	257	[redacted]
23	06Q3	0.0527169505271695	54	Bad	86	[redacted]
23	06Q3	0.0519059205190592	5	Haarpflege	4	[redacted]
23	06Q3	0.0519059205190592	48	Sonne	4	[redacted]
23	06Q3	0.0519059205190592	5	Haarpflege	91	[redacted]
23	06Q3	0.0510948905109489	19	Hygienepapie	1275	[redacted]

06Q4, Cluster 7

Cluster	Quarter	*Centroid* Value	Sub-Assortment	Brand

7	06Q4	0.190193965517241	45	Kerzen	2075	BrandR1
7	06Q4	0.147629310344828	51	Fu	207	dm-BrandK1
7	06Q4	0.144935344827586	18	Haush.artik.	1884	[redacted]
7	06Q4	0.13739224137931	42	Mund/Zahn	187	[redacted]
7	06Q4	0.126616379310345	6	Haushaltspap	1265	BrandY1
7	06Q4	0.122844827586207	40	Bonbon	1772	BrandC2
7	06Q4	0.12176724137931	3	Damenhygiene	2191	BrandA2
7	06Q4	0.119612068965517	61	Bodylotion	4	[redacted]
7	06Q4	0.113685344827586	116	Modul li	-1	?
7	06Q4	0.11260775862069	127	Ebelin Body	200	[redacted]
7	06Q4	0.108297413793103	11	Haarstyling	65	[redacted]
7	06Q4	0.107219827586207	66	Nassrasur	218	dm-BrandK1 [redacted]
7	06Q4	0.102370689655172	5	Haarpflege	66	[redacted]
7	06Q4	0.101293103448276	5	Haarpflege	33	BrandZ1
7	06Q4	0.100754310344828	36	Entferner	207	dm-BrandK1
7	06Q4	0.0980603448275862	42	Mund/Zahn	32	[redacted]
7	06Q4	0.0980603448275862	60	Gesicht	176	[redacted] [redacted]
7	06Q4	0.0969827586206897	3	Damenhygiene	267	[redacted]
7	06Q4	0.0964439655172414	36	Entferner	213	dm-BrandO1
7	06Q4	0.0953663793103448	16	Duftkerzen-M	212	[redacted]
7	06Q4	0.0889008620689655	4	Kosmetik-Pap	1241	BrandB2
7	06Q4	0.0872844827586207	56	Filter&Folien	1245	[redacted]
7	06Q4	0.0856681034482759	119	Augen Make u	29	[redacted]
7	06Q4	0.0818965517241379	56	Filter&Folien	1243	BrandX1
7	06Q4	0.0802801724137931	57	Deo	5	[redacted]
7	06Q4	0.0802801724137931	119	Augen Make u	1290	[redacted]
7	06Q4	0.0792025862068965	15	WPR	270	[redacted]
7	06Q4	0.0786637931034483	121	Nagelpflege	1309	[redacted]
7	06Q4	0.072198275862069	15	WPR	249	[redacted]
7	06Q4	0.068426724137931	45	Kerzen	2076	BrandH2
7	06Q4	0.068426724137931	39	Hand	257	[redacted]
7	06Q4	0.0678879310344828	120	Lippenpflege	1309	[redacted]
7	06Q4	0.0657327586206897	110	Feinstrick 1	1988	[redacted]
7	06Q4	0.0657327586206897	60	Gesicht	1365	[redacted]
7	06Q4	0.0646551724137931	39	Hand	10	[redacted]
7	06Q4	0.0635775862068965	30	Babypflege	81	[redacted]
7	06Q4	0.0619612068965517	5	Haarpflege	257	[redacted]
7	06Q4	0.0619612068965517	106	Modul klein	-1	?
7	06Q4	0.0614224137931034	5	Haarpflege	91	[redacted]
7	06Q4	0.0608836206896552	119	Augen Make u	156	[redacted]
7	06Q4	0.0598060344827586	58	Erlebnis dm	32	[redacted]
7	06Q4	0.0581896551724138	60	Gesicht	284	[redacted]
7	06Q4	0.0581896551724138	30	Babypflege	59	[redacted]

7	06Q4	0.0549568965517241	54	Bad	257	
7	06Q4	0.0549568965517241	72	Tee	1213	
7	06Q4	0.0538793103448276	54	Bad	782	
7	06Q4	0.0522629310344827	116	Modul li	212	
7	06Q4	0.0517241379310345	39	Hand	201	dm-green-BrandT1
7	06Q4	0.0511853448275862	5	Haarpflege	113	
7	06Q4	0.0495689655172414	60	Gesicht	174	
7	06Q4	0.046875	19	Hygienepapie	1275	
7	06Q4	0.0463362068965517	75	Reform	306	
7	06Q4	0.0457974137931034	15	WPR	439	
7	06Q4	0.0452586206896552	58	Erlebnis dm	1265	BrandY1
7	06Q4	0.0447198275862069	54	Bad	218	dm-BrandK1
7	06Q4	0.0441810344827586	61	Bodylotion	257	
7	06Q4	0.0441810344827586	119	Augen Make u	158	
7	06Q4	0.0436422413793103	15	WPR	262	
7	06Q4	0.0436422413793103	54	Bad	86	
7	06Q4	0.0431034482758621	40	Bonbon	1793	
7	06Q4	0.0431034482758621	106	Modul klein	731	green-BrandU1
7	06Q4	0.0425646551724138	39	Hand	172	
7	06Q4	0.0420258620689655	61	Bodylotion	284	
7	06Q4	0.0420258620689655	5	Haarpflege	4	
7	06Q4	0.0420258620689655	5	Haarpflege	952	
7	06Q4	0.0414870689655173	19	Hygienepapie	1264	
7	06Q4	0.0414870689655173	40	Bonbon	1789	
7	06Q4	0.0414870689655173	42	Mund/Zahn	58	
7	06Q4	0.0414870689655173	57	Deo	257	
7	06Q4	0.040948275862069	119	Augen Make u	79	
7	06Q4	0.040948275862069	57	Deo	173	Hidrofugal
7	06Q4	0.0404094827586207	48	Sonne	210	dm-BrandS1

E.2.2 The *Mainstream* Cluster

06Q1, Cluster 5

Cluster	Quarter	*Centroid* Value	Sub-Assortment		Brand	
5	06Q1	0.257107540173053	40	Bonbon	1772	BrandC2
5	06Q1	0.197775030902349	11	Haarstyling	65	
5	06Q1	0.184177997527812	42	Mund/Zahn	187	
5	06Q1	0.180469715698393	119	Augen Make u	1290	
5	06Q1	0.140914709517923	51	Fu	207	dm-BrandK1
5	06Q1	0.133498145859085	66	Nassrasur	218	dm-BrandK1
5	06Q1	0.133498145859085	36	Entferner	207	dm-BrandK1

5	06Q1	0.126081582200247	3	Damenhygiene	267	[redacted]
5	06Q1	0.11495673671199	39	Hand	201	dm-green-BrandT1
5	06Q1	0.106304079110012	5	Haarpflege	66	[redacted]
5	06Q1	0.102595797280593	127	Ebelin Body	200	[redacted]
5	06Q1	0.0951792336217552	18	Haush.artik.	1884	[redacted]
5	06Q1	0.0951792336217552	121	Nagelpflege	1309	[redacted]
5	06Q1	0.0951792336217552	15	WPR	249	[redacted]
5	06Q1	0.0927070457354759	48	Sonne	210	dm-BrandS1
5	06Q1	0.0902348578491965	54	Bad	201	dm-green-BrandT1
5	06Q1	0.0877626699629172	30	Babypflege	59	[redacted]
5	06Q1	0.0852904820766378	57	Deo	5	[redacted]
5	06Q1	0.0815822002472188	20	Nager	219	dm-BrandB1
5	06Q1	0.0803461063040791	120	Lippenpflege	1309	[redacted]
5	06Q1	0.0778739184177997	112	Ostern klein	-1	?
5	06Q1	0.0766378244746601	120	Lippenpflege	1290	[redacted]
5	06Q1	0.0754017305315204	39	Hand	257	[redacted]
5	06Q1	0.0754017305315204	61	Bodylotion	201	dm-green-BrandT1
5	06Q1	0.0754017305315204	5	Haarpflege	257	[redacted]
5	06Q1	0.0741656365883807	42	Mund/Zahn	32	[redacted]
5	06Q1	0.0716934487021014	60	Gesicht	284	[redacted]
5	06Q1	0.0716934487021014	60	Gesicht	162	[redacted]
5	06Q1	0.069221260815822	54	Bad	86	[redacted]
5	06Q1	0.0679851668726823	5	Haarpflege	967	[redacted]
5	06Q1	0.0679851668726823	72	Tee	1213	[redacted]
5	06Q1	0.0667490729295426	15	WPR	270	[redacted]
5	06Q1	0.0642768850432633	66	Nassrasur	71	[redacted]
5	06Q1	0.0630407911001236	6	Haushaltspap	1265	BrandY1
5	06Q1	0.0618046971569839	118	Make up	29	[redacted]
5	06Q1	0.0605686032138442	5	Haarpflege	952	[redacted]
5	06Q1	0.0605686032138442	60	Gesicht	176	[redacted] [redacted]
5	06Q1	0.0593325092707046	5	Haarpflege	113	[redacted]
5	06Q1	0.0580964153275649	39	Hand	10	[redacted]
5	06Q1	0.0580964153275649	4	Kosmetik-Pap	1241	BrandB2
5	06Q1	0.0580964153275649	75	Reform	306	[redacted]
5	06Q1	0.0580964153275649	36	Entferner	213	dm-BrandO1
5	06Q1	0.0556242274412855	5	Haarpflege	201	dm-green-BrandT1
5	06Q1	0.0556242274412855	58	Erlebnis dm	32	[redacted]
5	06Q1	0.0531520395550062	5	Haarpflege	33	BrandZ1
5	06Q1	0.0531520395550062	54	Bad	1474	BrandM1
5	06Q1	0.0531520395550062	75	Reform	381	[redacted]
5	06Q1	0.0531520395550062	113	Ostern mitte	-1	?
5	06Q1	0.0519159456118665	3	Damenhygiene	2191	BrandA2
5	06Q1	0.0506798516687268	119	Augen Make u	158	[redacted]

5	06Q1	0.0506798516687268	119	Augen Make u	156	[redacted]
5	06Q1	0.0506798516687268	15	WPR	238	BrandF2
5	06Q1	0.0506798516687268	40	Bonbon	1789	[redacted]
5	06Q1	0.0506798516687268	60	Gesicht	1365	[redacted]
5	06Q1	0.0506798516687268	16	Duftkerzen-M	433	[redacted]
5	06Q1	0.0494437577255871	110	Feinstrick 1	1988	[redacted]
5	06Q1	0.0482076637824475	15	WPR	439	[redacted]
5	06Q1	0.0469715698393078	42	Mund/Zahn	58	[redacted]
5	06Q1	0.0469715698393078	119	Augen Make u	29	[redacted]
5	06Q1	0.0469715698393078	45	Kerzen	2076	BrandH2
5	06Q1	0.0457354758961681	119	Augen Make u	79	[redacted]
5	06Q1	0.0457354758961681	40	Bonbon	381	[redacted]
5	06Q1	0.0457354758961681	30	Babypflege	81	[redacted]
5	06Q1	0.0444993819530284	39	Hand	1999	[redacted]
5	06Q1	0.0432632880098888	19	Hygienepapie	1275	[redacted]
5	06Q1	0.0432632880098888	61	Bodylotion	82	[redacted]
5	06Q1	0.0420271940667491	118	Make up	1290	[redacted]
5	06Q1	0.0420271940667491	54	Bad	381	[redacted]
5	06Q1	0.0407911001236094	60	Gesicht	174	[redacted]
5	06Q1	0.0407911001236094	61	Bodylotion	257	[redacted]
5	06Q1	0.0407911001236094	112	Ostern klein	212	[redacted]
5	06Q1	0.0407911001236094	42	Mund/Zahn	85	[redacted]
5	06Q1	0.0407911001236094	61	Bodylotion	284	[redacted]

***06Q2*, Cluster 30**

Cluster	Quarter	*Centroid* Value	Sub-Assortment		Brand	
30	06Q2	0.600973236009732	48	Sonne	210	dm-BrandS1
30	06Q2	0.29683698296837	51	Fu	207	dm-BrandK1
30	06Q2	0.150040551500406	40	Bonbon	1772	BrandC2
30	06Q2	0.148418491484185	66	Nassrasur	218	dm-BrandK1 [redacted]
30	06Q2	0.139497161394972	36	Entferner	207	dm-BrandK1
30	06Q2	0.1338199513382	11	Haarstyling	65	[redacted]
30	06Q2	0.127331711273317	119	Augen Make u	1290	[redacted]
30	06Q2	0.124898621248986	127	Ebelin Body	200	[redacted]
30	06Q2	0.123276561232766	121	Nagelpflege	1309	[redacted]
30	06Q2	0.115166261151663	36	Entferner	213	dm-BrandO1
30	06Q2	0.111111111111111	42	Mund/Zahn	187	[redacted]
30	06Q2	0.10867802108678	5	Haarpflege	66	[redacted]
30	06Q2	0.097323600973236	57	Deo	5	[redacted]
30	06Q2	0.0948905109489051	5	Haarpflege	33	BrandZ1
30	06Q2	0.0932684509326845	3	Damenhygiene	267	[redacted]

30	06Q2	0.0916463909164639	3	Damenhygiene	2191	BrandA2
30	06Q2	0.0900243309002433	42	Mund/Zahn	32	[redacted]
30	06Q2	0.0884022708840227	39	Hand	1999	[redacted]
30	06Q2	0.0884022708840227	120	Lippenpflege	1290	[redacted]
30	06Q2	0.0827250608272506	110	Feinstrick 1	1988	[redacted]
30	06Q2	0.0746147607461476	60	Gesicht	176	[redacted] [redacted]
30	06Q2	0.0738037307380373	5	Haarpflege	257	[redacted]
30	06Q2	0.0738037307380373	39	Hand	257	[redacted]
30	06Q2	0.072992700729927	54	Bad	218	dm-BrandK1 [redacted]
30	06Q2	0.0697485806974858	30	Babypflege	59	[redacted]
30	06Q2	0.0689375506893755	119	Augen Make u	29	[redacted]
30	06Q2	0.0665044606650446	18	Haush.artik.	1884	[redacted]
30	06Q2	0.0616382806163828	58	Erlebnis dm	257	[redacted]
30	06Q2	0.0592051905920519	118	Make up	1290	[redacted]
30	06Q2	0.0575831305758313	39	Hand	201	dm-green-BrandT1
30	06Q2	0.0575831305758313	54	Bad	27	[redacted]
30	06Q2	0.0575831305758313	61	Bodylotion	257	[redacted]
30	06Q2	0.0551500405515004	6	Haushaltspap	1265	BrandY1
30	06Q2	0.0551500405515004	61	Bodylotion	4	[redacted]
30	06Q2	0.0551500405515004	60	Gesicht	1365	[redacted]
30	06Q2	0.0551500405515004	120	Lippenpflege	1309	[redacted]
30	06Q2	0.0543390105433901	60	Gesicht	284	[redacted]
30	06Q2	0.0519059205190592	30	Babypflege	81	[redacted]
30	06Q2	0.0519059205190592	119	Augen Make u	156	[redacted]
30	06Q2	0.0510948905109489	60	Gesicht	162	[redacted]
30	06Q2	0.0494728304947283	4	Kosmetik-Pap	1241	BrandB2
30	06Q2	0.0494728304947283	54	Bad	782	[redacted]
30	06Q2	0.0494728304947283	39	Hand	10	[redacted]
30	06Q2	0.0494728304947283	61	Bodylotion	82	[redacted]
30	06Q2	0.0470397404703974	45	Kerzen	2075	BrandR1
30	06Q2	0.0462287104622871	54	Bad	257	[redacted]
30	06Q2	0.0462287104622871	42	Mund/Zahn	85	[redacted]
30	06Q2	0.0454176804541768	5	Haarpflege	113	[redacted]
30	06Q2	0.0446066504460665	61	Bodylotion	284	[redacted]
30	06Q2	0.0446066504460665	5	Haarpflege	91	[redacted]
30	06Q2	0.0437956204379562	121	Nagelpflege	29	[redacted]
30	06Q2	0.0429845904298459	78	Pharma	897	[redacted]
30	06Q2	0.0429845904298459	11	Haarstyling	756	[redacted]
30	06Q2	0.0429845904298459	5	Haarpflege	952	[redacted]
30	06Q2	0.0421735604217356	15	WPR	439	[redacted]
30	06Q2	0.0413625304136253	54	Bad	86	[redacted]
30	06Q2	0.0413625304136253	15	WPR	270	[redacted]
30	06Q2	0.040551500405515	75	Reform	306	[redacted]

30	06Q2	0.040551500405515	5	Haarpflege	967	[redacted]

***06Q3*, Cluster 0**

Cluster	Quarter	*Centroid* Value	Sub-Assortment		Brand	
0	06Q3	0.433021806853583	48	Sonne	210	dm-BrandS1
0	06Q3	0.310747663551402	51	Fu	207	dm-BrandK1
0	06Q3	0.207165109034268	40	Bonbon	1772	BrandC2
0	06Q3	0.175233644859813	36	Entferner	213	dm-BrandO1
0	06Q3	0.160436137071651	66	Nassrasur	218	dm-BrandK1 [redacted]
0	06Q3	0.154984423676012	127	Ebelin Body	200	[redacted]
0	06Q3	0.143302180685358	42	Mund/Zahn	187	[redacted]
0	06Q3	0.130841121495327	57	Deo	5	[redacted]
0	06Q3	0.129283489096573	121	Nagelpflege	1309	[redacted]
0	06Q3	0.110591900311526	36	Entferner	207	dm-BrandK1
0	06Q3	0.0989096573208723	11	Haarstyling	65	[redacted]
0	06Q3	0.0965732087227414	54	Bad	218	dm-BrandK1 [redacted]
0	06Q3	0.0957943925233645	30	Babypflege	81	[redacted]
0	06Q3	0.0950155763239875	119	Augen Make u	1290	[redacted]
0	06Q3	0.0942367601246106	3	Damenhygiene	267	[redacted]
0	06Q3	0.0919003115264797	15	WPR	270	[redacted]
0	06Q3	0.088785046728972	5	Haarpflege	66	[redacted]
0	06Q3	0.0825545171339564	39	Hand	1999	[redacted]
0	06Q3	0.0817757009345794	39	Hand	201	dm-green-BrandT1
0	06Q3	0.0802180685358255	119	Augen Make u	29	[redacted]
0	06Q3	0.0794392523364486	42	Mund/Zahn	32	[redacted]
0	06Q3	0.0763239875389408	5	Haarpflege	257	[redacted]
0	06Q3	0.0755451713395639	57	Deo	218	dm-BrandK1 [redacted]
0	06Q3	0.0716510903426791	120	Lippenpflege	1309	[redacted]
0	06Q3	0.0716510903426791	15	WPR	439	[redacted]
0	06Q3	0.0716510903426791	53	Posten/Saiso	-1	?
0	06Q3	0.0661993769470405	54	Bad	782	[redacted]
0	06Q3	0.0661993769470405	3	Damenhygiene	2191	BrandA2
0	06Q3	0.0654205607476635	18	Haush.artik.	1884	[redacted]
0	06Q3	0.0630841121495327	30	Babypflege	59	[redacted]
0	06Q3	0.0623052959501558	75	Reform	381	[redacted]
0	06Q3	0.0615264797507788	60	Gesicht	1365	[redacted]
0	06Q3	0.0599688473520249	110	Feinstrick 1	1988	[redacted]
0	06Q3	0.0599688473520249	119	Augen Make u	156	[redacted]
0	06Q3	0.058411214953271	5	Haarpflege	33	BrandZ1
0	06Q3	0.0560747663551402	60	Gesicht	176	[redacted] [redacted]
0	06Q3	0.0560747663551402	5	Haarpflege	952	[redacted]

0	06Q3	0.0552959501557632	54	Bad	82	[redacted]
0	06Q3	0.0552959501557632	66	Nassrasur	71	[redacted]
0	06Q3	0.0552959501557632	5	Haarpflege	91	[redacted]
0	06Q3	0.0529595015576324	20	Nager	219	dm-BrandB1
0	06Q3	0.0529595015576324	75	Reform	1937	[redacted]
0	06Q3	0.0514018691588785	60	Gesicht	284	[redacted]
0	06Q3	0.0490654205607477	5	Haarpflege	967	[redacted]
0	06Q3	0.0482866043613707	5	Haarpflege	1916	[redacted]
0	06Q3	0.0482866043613707	34	Fotozubehr	2189	[redacted]
0	06Q3	0.0467289719626168	5	Haarpflege	4	[redacted]
0	06Q3	0.0467289719626168	42	Mund/Zahn	58	[redacted]
0	06Q3	0.0459501557632399	61	Bodylotion	284	[redacted]
0	06Q3	0.044392523364486	45	Kerzen	2075	BrandR1
0	06Q3	0.044392523364486	119	Augen Make u	79	[redacted]
0	06Q3	0.043613707165109	61	Bodylotion	257	[redacted]
0	06Q3	0.043613707165109	16	Duftkerzen-M	433	[redacted]
0	06Q3	0.0428348909657321	60	Gesicht	162	[redacted]
0	06Q3	0.0428348909657321	48	Sonne	182	[redacted]
0	06Q3	0.0428348909657321	6	Haushaltspap	1265	BrandY1
0	06Q3	0.0428348909657321	11	Haarstyling	756	[redacted]
0	06Q3	0.0420560747663551	54	Bad	20	[redacted]
0	06Q3	0.0420560747663551	70	Fotoarbeiten	2359	[redacted] [redacted]
0	06Q3	0.0412772585669782	5	Haarpflege	113	[redacted]
0	06Q3	0.0412772585669782	30	Babypflege	2067	[redacted]
0	06Q3	0.0404984423676013	118	Make up	1290	[redacted]

E.2.3 The *Baby/Young Families* Cluster

***06Q1*, Cluster 7**

Cluster	Quarter	*Centroid* Value	Sub-Assortment		Brand	
7	06Q1	0.528213166144201	27	Windeln	899	BrandE2
7	06Q1	0.343260188087774	33	Babynahrung	1846	BrandA1
7	06Q1	0.321316614420063	37	Babyglasnahr	1846	BrandA1
7	06Q1	0.252351097178683	37	Babyglasnahr	1847	BrandL1
7	06Q1	0.230407523510972	33	Babynahrung	1847	BrandL1
7	06Q1	0.202194357366771	26	Flaschen & S	206	dm-BrandQ1
7	06Q1	0.19435736677116	30	Babypflege	59	[redacted]
7	06Q1	0.156739811912226	3	Damenhygiene	267	[redacted]
7	06Q1	0.122257053291536	112	Ostern klein	-1	?
7	06Q1	0.108150470219436	42	Mund/Zahn	187	[redacted]
7	06Q1	0.0940438871473354	5	Haarpflege	91	[redacted]

7	06Q1	0.0862068965517241	66	Nassrasur	218	dm-BrandK1 [redacted]
7	06Q1	0.0862068965517241	30	Babypflege	2111	[redacted]
7	06Q1	0.0846394984326019	30	Babypflege	2067	[redacted]
7	06Q1	0.0846394984326019	3	Damenhygiene	2191	BrandA2
7	06Q1	0.0830721003134796	11	Haarstyling	65	[redacted]
7	06Q1	0.0768025078369906	40	Bonbon	1772	BrandC2
7	06Q1	0.0736677115987461	6	Haushaltspap	1265	BrandY1
7	06Q1	0.0721003134796238	42	Mund/Zahn	32	[redacted]
7	06Q1	0.0705329153605016	5	Haarpflege	952	[redacted]
7	06Q1	0.0689655172413793	42	Mund/Zahn	58	[redacted]
7	06Q1	0.0673981191222571	5	Haarpflege	33	BrandZ1
7	06Q1	0.0658307210031348	5	Haarpflege	257	[redacted]
7	06Q1	0.0626959247648903	15	WPR	238	BrandF2
7	06Q1	0.061128526645768	54	Bad	257	[redacted]
7	06Q1	0.061128526645768	53	Posten/Saiso	-1	?
7	06Q1	0.061128526645768	127	Ebelin Body	200	[redacted]
7	06Q1	0.0579937304075235	54	Bad	86	[redacted]
7	06Q1	0.0579937304075235	60	Gesicht	176	[redacted] [redacted]
7	06Q1	0.0564263322884013	119	Augen Make u	29	[redacted]
7	06Q1	0.054858934169279	18	Haush.artik.	1884	[redacted]
7	06Q1	0.054858934169279	60	Gesicht	174	[redacted]
7	06Q1	0.0532915360501567	51	Fu	207	dm-BrandK1
7	06Q1	0.0532915360501567	40	Bonbon	381	[redacted]
7	06Q1	0.0517241379310345	15	WPR	270	[redacted]
7	06Q1	0.0501567398119122	57	Deo	5	[redacted]
7	06Q1	0.0501567398119122	40	Bonbon	1789	[redacted]
7	06Q1	0.04858934169279	119	Augen Make u	156	[redacted]
7	06Q1	0.04858934169279	54	Bad	782	[redacted]
7	06Q1	0.0470219435736677	61	Bodylotion	4	[redacted]
7	06Q1	0.0470219435736677	58	Erlebnis dm	32	[redacted]
7	06Q1	0.0454545454545455	33	Babynahrung	1856	[redacted]
7	06Q1	0.0454545454545455	5	Haarpflege	66	[redacted]
7	06Q1	0.0438871473354232	15	WPR	221	[redacted]
7	06Q1	0.0423197492163009	30	Babypflege	899	BrandE2
7	06Q1	0.0423197492163009	30	Babypflege	268	[redacted]
7	06Q1	0.0407523510971787	15	WPR	262	[redacted]
7	06Q1	0.0407523510971787	4	Kosmetik-Pap	1241	BrandB2
7	06Q1	0.0407523510971787	48	Sonne	210	dm-BrandS1
7	06Q1	0.0407523510971787	61	Bodylotion	257	[redacted]
7	06Q1	0.0407523510971787	75	Reform	1937	[redacted]
7	06Q1	0.0407523510971787	40	Bonbon	1794	[redacted]
7	06Q1	0.0407523510971787	53	Posten/Saiso	133	[redacted]

06Q2, Cluster 12

Cluster	Quarter	*Centroid* Value	Sub-Assortment		Brand	
12	06Q2	0.494437577255871	27	Windeln	899	BrandE2
12	06Q2	0.288009888751545	33	Babynahrung	1846	BrandA1
12	06Q2	0.285537700865266	37	Babyglasnahr	1846	BrandA1
12	06Q2	0.202719406674907	37	Babyglasnahr	1847	BrandL1
12	06Q2	0.19035846724351	33	Babynahrung	1847	BrandL1
12	06Q2	0.179233621755253	26	Flaschen & S	206	dm-BrandQ1
12	06Q2	0.171817058096415	42	Mund/Zahn	187	[redacted]
12	06Q2	0.166872682323857	3	Damenhygiene	267	[redacted]
12	06Q2	0.15327564894932	30	Babypflege	59	[redacted]
12	06Q2	0.143386897404203	48	Sonne	210	dm-BrandS1
12	06Q2	0.102595797280593	11	Haarstyling	65	[redacted]
12	06Q2	0.0976514215080346	112	Ostern klein	-1	?
12	06Q2	0.0964153275648949	48	Sonne	206	dm-BrandQ1
12	06Q2	0.0927070457354759	30	Babypflege	2067	[redacted]
12	06Q2	0.0889987639060569	66	Nassrasur	218	dm-BrandK1 [redacted]
12	06Q2	0.0852904820766378	3	Damenhygiene	2191	BrandA2
12	06Q2	0.0778739184177997	51	Fu	207	dm-BrandK1
12	06Q2	0.0741656365883807	18	Haush.artik.	1884	[redacted]
12	06Q2	0.0741656365883807	40	Bonbon	1772	BrandC2
12	06Q2	0.072929542645241	127	Ebelin Body	200	[redacted]
12	06Q2	0.069221260815822	54	Bad	27	[redacted]
12	06Q2	0.0679851668726823	5	Haarpflege	33	BrandZ1
12	06Q2	0.0679851668726823	5	Haarpflege	91	[redacted]
12	06Q2	0.0667490729295426	30	Babypflege	81	[redacted]
12	06Q2	0.065512978986403	42	Mund/Zahn	32	[redacted]
12	06Q2	0.0630407911001236	6	Haushaltspap	1265	BrandY1
12	06Q2	0.0630407911001236	5	Haarpflege	66	[redacted]
12	06Q2	0.0605686032138442	33	Babynahrung	1856	[redacted]
12	06Q2	0.0593325092707046	37	Babyglasnahr	731	green-BrandU1
12	06Q2	0.0593325092707046	5	Haarpflege	952	[redacted]
12	06Q2	0.0593325092707046	121	Nagelpflege	1309	[redacted]
12	06Q2	0.0593325092707046	75	Reform	1931	[redacted]
12	06Q2	0.0593325092707046	42	Mund/Zahn	58	[redacted]
12	06Q2	0.0580964153275649	58	Erlebnis dm	32	[redacted]
12	06Q2	0.0556242274412855	5	Haarpflege	257	[redacted]
12	06Q2	0.0543881334981459	53	Posten/Saiso	-1	?
12	06Q2	0.0543881334981459	36	Entferner	213	dm-BrandO1
12	06Q2	0.0531520395550062	30	Babypflege	268	[redacted]
12	06Q2	0.0531520395550062	15	WPR	439	[redacted]
12	06Q2	0.0519159456118665	15	WPR	262	[redacted]

12	06Q2	0.0519159456118665	30	Babypflege	2111	[redacted]
12	06Q2	0.0506798516687268	34	Fotozubehr	205	[redacted]
12	06Q2	0.0506798516687268	54	Bad	257	[redacted]
12	06Q2	0.0506798516687268	39	Hand	201	dm-green-BrandT1
12	06Q2	0.0506798516687268	70	Fotoarbeiten	2359	[redacted] [redacted]
12	06Q2	0.0506798516687268	54	Bad	86	[redacted]
12	06Q2	0.0494437577255871	30	Babypflege	899	BrandE2
12	06Q2	0.0482076637824475	26	Flaschen & S	1843	[redacted]
12	06Q2	0.0469715698393078	119	Augen Make u	1290	[redacted]
12	06Q2	0.0469715698393078	75	Reform	2204	[redacted]
12	06Q2	0.0469715698393078	40	Bonbon	381	[redacted]
12	06Q2	0.0457354758961681	61	Bodylotion	4	[redacted]
12	06Q2	0.0457354758961681	119	Augen Make u	156	[redacted]
12	06Q2	0.0444993819530284	74	Filme	861	[redacted]
12	06Q2	0.0444993819530284	58	Erlebnis dm	1265	BrandY1
12	06Q2	0.0432632880098888	39	Hand	61	[redacted]
12	06Q2	0.0432632880098888	66	Nassrasur	71	[redacted]
12	06Q2	0.0432632880098888	15	WPR	270	[redacted]
12	06Q2	0.0432632880098888	5	Haarpflege	4	[redacted]
12	06Q2	0.0432632880098888	60	Gesicht	176	[redacted] [redacted]
12	06Q2	0.0420271940667491	39	Hand	257	[redacted]
12	06Q2	0.0420271940667491	75	Reform	1937	[redacted]
12	06Q2	0.0407911001236094	110	Feinstrick 1	1988	[redacted]
12	06Q2	0.0407911001236094	6	Haushaltspap	1263	[redacted]

06Q3, Cluster 35

Cluster	Quarter	*Centroid* Value	Sub-Assortment		Brand	
35	06Q3	0.474509803921569	27	Windeln	899	BrandE2
35	06Q3	0.317647058823529	37	Babyglasnahr	1846	BrandA1
35	06Q3	0.295424836601307	33	Babynahrung	1846	BrandA1
35	06Q3	0.219607843137255	37	Babyglasnahr	1847	BrandL1
35	06Q3	0.203921568627451	26	Flaschen & S	206	dm-BrandQ1
35	06Q3	0.164705882352941	42	Mund/Zahn	187	[redacted]
35	06Q3	0.162091503267974	33	Babynahrung	1847	BrandL1
35	06Q3	0.149019607843137	3	Damenhygiene	267	[redacted]
35	06Q3	0.12156862745098	30	Babypflege	59	[redacted]
35	06Q3	0.113725490196078	30	Babypflege	2067	[redacted]
35	06Q3	0.112418300653595	48	Sonne	210	dm-BrandS1
35	06Q3	0.105882352941176	66	Nassrasur	218	dm-BrandK1 [redacted]
35	06Q3	0.0941176470588235	51	Fu	207	dm-BrandK1
35	06Q3	0.0849673202614379	18	Haush.artik.	1884	[redacted]

35	06Q3	0.0836601307189542	11	Haarstyling	65	[redacted]
35	06Q3	0.0823529411764706	75	Reform	1937	[redacted]
35	06Q3	0.0823529411764706	60	Gesicht	176	[redacted] [redacted]
35	06Q3	0.0797385620915033	5	Haarpflege	33	BrandZ1
35	06Q3	0.0797385620915033	75	Reform	2204	[redacted]
35	06Q3	0.0784313725490196	42	Mund/Zahn	32	[redacted]
35	06Q3	0.077124183006536	40	Bonbon	1772	BrandC2
35	06Q3	0.073202614379085	127	Ebelin Body	200	[redacted]
35	06Q3	0.0718954248366013	5	Haarpflege	91	[redacted]
35	06Q3	0.0705882352941176	30	Babypflege	2111	[redacted]
35	06Q3	0.0705882352941176	5	Haarpflege	66	[redacted]
35	06Q3	0.069281045751634	30	Babypflege	81	[redacted]
35	06Q3	0.0666666666666667	53	Posten/Saiso	-1	?
35	06Q3	0.0640522875816993	3	Damenhygiene	2191	BrandA2
35	06Q3	0.0640522875816993	54	Bad	27	[redacted]
35	06Q3	0.0627450980392157	42	Mund/Zahn	58	[redacted]
35	06Q3	0.061437908496732	39	Hand	201	dm-green-BrandT1
35	06Q3	0.0588235294117647	54	Bad	782	[redacted]
35	06Q3	0.0588235294117647	5	Haarpflege	4	[redacted]
35	06Q3	0.0562091503267974	6	Haushaltspap	1265	BrandY1
35	06Q3	0.0549019607843137	121	Nagelpflege	1309	[redacted]
35	06Q3	0.0535947712418301	70	Fotoarbeiten	2359	[redacted] [redacted]
35	06Q3	0.0522875816993464	54	Bad	257	[redacted]
35	06Q3	0.0509803921568627	56	Filter&Folien	1245	[redacted]
35	06Q3	0.0509803921568627	5	Haarpflege	257	[redacted]
35	06Q3	0.0496732026143791	119	Augen Make u	29	[redacted]
35	06Q3	0.0496732026143791	36	Entferner	213	dm-BrandO1
35	06Q3	0.0496732026143791	33	Babynahrung	1856	[redacted]
35	06Q3	0.0483660130718954	39	Hand	257	[redacted]
35	06Q3	0.0483660130718954	58	Erlebnis dm	32	[redacted]
35	06Q3	0.0470588235294118	61	Bodylotion	4	[redacted]
35	06Q3	0.0470588235294118	40	Bonbon	381	[redacted]
35	06Q3	0.0470588235294118	37	Babyglasnahr	731	green-BrandU1
35	06Q3	0.0457516339869281	15	WPR	262	[redacted]
35	06Q3	0.0444444444444444	30	Babypflege	899	BrandE2
35	06Q3	0.0444444444444444	30	Babypflege	268	[redacted]
35	06Q3	0.0444444444444444	110	Feinstrick 1	1988	[redacted]
35	06Q3	0.0431372549019608	74	Filme	861	[redacted]
35	06Q3	0.0431372549019608	119	Augen Make u	1290	[redacted]
35	06Q3	0.0431372549019608	54	Bad	218	dm-BrandK1 [redacted]
35	06Q3	0.0405228758169935	66	Nassrasur	254	[redacted]

06Q4, Cluster 14

Cluster	Quarter	*Centroid* Value	Sub-Assortment		Brand	
14	06Q4	0.345605700712589	27	Windeln	899	BrandE2
14	06Q4	0.252969121140142	33	Babynahrung	1846	BrandA1
14	06Q4	0.235154394299287	37	Babyglasnahr	1846	BrandA1
14	06Q4	0.21021377672209	37	Babyglasnahr	1847	BrandL1
14	06Q4	0.197149643705463	3	Damenhygiene	267	[redacted]
14	06Q4	0.188836104513064	116	Modul li	-1	?
14	06Q4	0.185273159144893	42	Mund/Zahn	187	[redacted]
14	06Q4	0.182897862232779	106	Modul klein	-1	?
14	06Q4	0.166270783847981	26	Flaschen & S	206	dm-BrandQ1
14	06Q4	0.13895486935867	30	Babypflege	2067	[redacted]
14	06Q4	0.13895486935867	30	Babypflege	59	[redacted]
14	06Q4	0.1270783847981	33	Babynahrung	1847	BrandL1
14	06Q4	0.10332541567696	11	Haarstyling	65	[redacted]
14	06Q4	0.0914489311163895	66	Nassrasur	218	dm-BrandK1 [redacted]
14	06Q4	0.0890736342042755	107	Modul mittel	-1	?
14	06Q4	0.0866983372921615	6	Haushaltspap	1265	BrandY1
14	06Q4	0.0855106888361045	18	Haush.artik.	1884	[redacted]
14	06Q4	0.0807600950118765	42	Mund/Zahn	32	[redacted]
14	06Q4	0.0807600950118765	60	Gesicht	176	[redacted] [redacted]
14	06Q4	0.0748218527315915	15	WPR	270	[redacted]
14	06Q4	0.0712589073634204	70	Fotoarbeiten	2359	[redacted] [redacted]
14	06Q4	0.0712589073634204	53	Posten/Saiso	-1	?
14	06Q4	0.0700712589073634	42	Mund/Zahn	58	[redacted]
14	06Q4	0.0688836104513064	40	Bonbon	1772	BrandC2
14	06Q4	0.0665083135391924	40	Bonbon	1789	[redacted]
14	06Q4	0.0653206650831354	5	Haarpflege	33	BrandZ1
14	06Q4	0.0641330166270784	3	Damenhygiene	2191	BrandA2
14	06Q4	0.0641330166270784	5	Haarpflege	91	[redacted]
14	06Q4	0.0629453681710214	40	Bonbon	381	[redacted]
14	06Q4	0.0558194774346793	5	Haarpflege	66	[redacted]
14	06Q4	0.0546318289786223	58	Erlebnis dm	32	[redacted]
14	06Q4	0.0546318289786223	119	Augen Make u	29	[redacted]
14	06Q4	0.0534441805225653	75	Reform	1937	[redacted]
14	06Q4	0.0534441805225653	54	Bad	257	[redacted]
14	06Q4	0.0534441805225653	54	Bad	27	[redacted]
14	06Q4	0.0522565320665083	51	Fu	207	dm-BrandK1
14	06Q4	0.0522565320665083	75	Reform	2204	[redacted]
14	06Q4	0.0510688836104513	39	Hand	257	[redacted]
14	06Q4	0.0510688836104513	16	Duftkerzen-M	212	[redacted]
14	06Q4	0.0498812351543943	119	Augen Make u	1290	[redacted]

14	06Q4	0.0486935866983373	45	Kerzen	2076	BrandH2
14	06Q4	0.0486935866983373	45	Kerzen	2075	BrandR1
14	06Q4	0.0475059382422803	5	Haarpflege	257	[redacted]
14	06Q4	0.0463182897862233	20	Nager	219	dm-BrandB1
14	06Q4	0.0463182897862233	5	Haarpflege	952	[redacted]
14	06Q4	0.0451306413301663	36	Entferner	213	dm-BrandO1
14	06Q4	0.0451306413301663	127	Ebelin Body	200	[redacted]
14	06Q4	0.0451306413301663	15	WPR	249	[redacted]
14	06Q4	0.0439429928741093	57	Deo	5	[redacted]
14	06Q4	0.0439429928741093	30	Babypflege	81	[redacted]
14	06Q4	0.0427553444180522	5	Haarpflege	4	[redacted]
14	06Q4	0.0427553444180522	56	Filter&Folien	1245	[redacted]
14	06Q4	0.0415676959619952	34	Fotozubehr	205	[redacted]
14	06Q4	0.0415676959619952	54	Bad	782	[redacted]
14	06Q4	0.0415676959619952	36	Entferner	207	dm-BrandK1
14	06Q4	0.0403800475059382	108	Modul gross	-1	?
14	06Q4	0.0403800475059382	30	Babypflege	268	[redacted]
14	06Q4	0.0403800475059382	15	WPR	264	BrandJ1
14	06Q4	0.0403800475059382	18	Haush.artik.	1152	[redacted]

E.2.4 The *GreenConcerns* Cluster

06Q2, Cluster 11

Cluster	Quarter	*Centroid* Value	Sub-Assortment		Brand	
11	06Q2	0.30379746835443	39	Hand	201	dm-green-BrandT1
11	06Q2	0.221518987341772	48	Sonne	210	dm-BrandS1
11	06Q2	0.20253164556962	54	Bad	201	dm-green-BrandT1
11	06Q2	0.177215189873418	5	Haarpflege	201	dm-green-BrandT1
11	06Q2	0.158227848101266	42	Mund/Zahn	187	[redacted]
11	06Q2	0.151898734177215	61	Bodylotion	201	dm-green-BrandT1
11	06Q2	0.145569620253165	18	Haush.artik.	1884	[redacted]
11	06Q2	0.139240506329114	66	Nassrasur	218	dm-BrandK1 [redacted]
11	06Q2	0.132911392405063	40	Bonbon	1772	BrandC2
11	06Q2	0.120253164556962	15	WPR	238	BrandF2
11	06Q2	0.113924050632911	42	Mund/Zahn	58	[redacted]
11	06Q2	0.107594936708861	40	Bonbon	381	[redacted]
11	06Q2	0.107594936708861	72	Tee	1213	[redacted]
11	06Q2	0.107594936708861	5	Haarpflege	33	BrandZ1
11	06Q2	0.107594936708861	54	Bad	218	dm-BrandK1 [redacted]
11	06Q2	0.10126582278481	51	Fu	207	dm-BrandK1
11	06Q2	0.0949367088607595	15	WPR	439	[redacted]

11	06Q2	0.0886075949367089	74	Filme	861	[redacted]
11	06Q2	0.0886075949367089	42	Mund/Zahn	1140	[redacted]
11	06Q2	0.0886075949367089	75	Reform	306	[redacted]
11	06Q2	0.0822784810126582	112	Ostern klein	-1	?
11	06Q2	0.0822784810126582	11	Haarstyling	65	[redacted]
11	06Q2	0.0822784810126582	36	Entferner	207	dm-BrandK1
11	06Q2	0.0822784810126582	3	Damenhygiene	267	[redacted]
11	06Q2	0.0822784810126582	58	Erlebnis dm	257	[redacted]
11	06Q2	0.0759493670886076	70	Fotoarbeiten	2359	[redacted] [redacted]
11	06Q2	0.069620253164557	56	Filter&Folien	1243	BrandX1
11	06Q2	0.069620253164557	78	Pharma	897	[redacted]
11	06Q2	0.0632911392405063	58	Erlebnis dm	32	[redacted]
11	06Q2	0.0632911392405063	36	Entferner	213	dm-BrandO1
11	06Q2	0.0632911392405063	53	Posten/Saiso	-1	?
11	06Q2	0.0632911392405063	110	Feinstrick 1	1988	[redacted]
11	06Q2	0.0632911392405063	56	Filter&Folien	1245	[redacted]
11	06Q2	0.0632911392405063	119	Augen Make u	29	[redacted]
11	06Q2	0.0632911392405063	127	Ebelin Body	200	[redacted]
11	06Q2	0.0632911392405063	60	Gesicht	12	[redacted]
11	06Q2	0.0632911392405063	5	Haarpflege	113	[redacted]
11	06Q2	0.0569620253164557	16	Duftkerzen-M	433	[redacted]
11	06Q2	0.0569620253164557	37	Babyglasnahr	731	green-BrandU1
11	06Q2	0.0569620253164557	30	Babypflege	2067	[redacted]
11	06Q2	0.0569620253164557	58	Erlebnis dm	731	green-BrandU1
11	06Q2	0.0569620253164557	42	Mund/Zahn	1132	[redacted]
11	06Q2	0.0569620253164557	9	Hund	219	dm-BrandB1
11	06Q2	0.0569620253164557	39	Hand	61	[redacted]
11	06Q2	0.0569620253164557	39	Hand	172	[redacted] [redacted]
11	06Q2	0.0569620253164557	45	Kerzen	2075	BrandR1
11	06Q2	0.0506329113924051	48	Sonne	4	[redacted]
11	06Q2	0.0506329113924051	60	Gesicht	174	[redacted]
11	06Q2	0.0506329113924051	5	Haarpflege	257	[redacted]
11	06Q2	0.0506329113924051	66	Nassrasur	86	[redacted]
11	06Q2	0.0506329113924051	5	Haarpflege	66	[redacted]
11	06Q2	0.0443037974683544	15	WPR	895	[redacted]
11	06Q2	0.0443037974683544	30	Babypflege	274	[redacted]
11	06Q2	0.0443037974683544	78	Pharma	2136	[redacted]
11	06Q2	0.0443037974683544	20	Nager	219	dm-BrandB1
11	06Q2	0.0443037974683544	119	Augen Make u	1290	[redacted]
11	06Q2	0.0443037974683544	42	Mund/Zahn	891	[redacted]
11	06Q2	0.0443037974683544	12	Insekt/Pflan	212	[redacted]
11	06Q2	0.0443037974683544	60	Gesicht	176	[redacted] [redacted]
11	06Q2	0.0443037974683544	60	Gesicht	162	[redacted]

11	06Q2	0.0443037974683544	57	Deo	257	[redacted]
11	06Q2	0.0443037974683544	30	Babypflege	59	[redacted]
11	06Q2	0.0443037974683544	15	WPR	212	[redacted]
11	06Q2	0.0443037974683544	119	Augen Make u	79	[redacted]
11	06Q2	0.0443037974683544	58	Erlebnis dm	1265	BrandY1

***06Q3*, Cluster 20**

Cluster	Quarter	*Centroid* Value	Sub-Assortment		Brand	
20	06Q3	0.421875	42	Mund/Zahn	187	[redacted]
20	06Q3	0.2421875	54	Bad	201	dm-green-BrandT1
20	06Q3	0.21875	39	Hand	201	dm-green-BrandT1
20	06Q3	0.171875	36	Entferner	213	dm-BrandO1
20	06Q3	0.1640625	61	Bodylotion	201	dm-green-BrandT1
20	06Q3	0.1640625	48	Sonne	210	dm-BrandS1
20	06Q3	0.15625	66	Nassrasur	218	dm-BrandK1 [redacted]
20	06Q3	0.140625	5	Haarpflege	66	[redacted]
20	06Q3	0.140625	57	Deo	173	[redacted]
20	06Q3	0.140625	3	Damenhygiene	267	[redacted]
20	06Q3	0.1171875	121	Nagelpflege	158	[redacted]
20	06Q3	0.1171875	127	Ebelin Body	200	[redacted]
20	06Q3	0.1171875	5	Haarpflege	201	dm-green-BrandT1
20	06Q3	0.109375	119	Augen Make u	49	[redacted]
20	06Q3	0.109375	6	Haushaltspap	1265	BrandY1
20	06Q3	0.09375	30	Babypflege	59	[redacted]
20	06Q3	0.09375	36	Entferner	207	dm-BrandK1
20	06Q3	0.09375	39	Hand	61	[redacted]
20	06Q3	0.09375	74	Filme	861	[redacted]
20	06Q3	0.0859375	40	Bonbon	1789	[redacted]
20	06Q3	0.0859375	15	WPR	262	[redacted]
20	06Q3	0.0703125	121	Nagelpflege	29	[redacted]
20	06Q3	0.0703125	18	Haush.artik.	1884	[redacted]
20	06Q3	0.0703125	39	Hand	257	[redacted]
20	06Q3	0.0703125	45	Kerzen	2075	BrandR1
20	06Q3	0.0703125	56	Filter&Folien	1243	BrandX1
20	06Q3	0.0625	58	Erlebnis dm	32	[redacted]
20	06Q3	0.0625	75	Reform	381	[redacted]
20	06Q3	0.0625	48	Sonne	182	[redacted]
20	06Q3	0.0625	20	Nager	219	dm-BrandB1
20	06Q3	0.0625	18	Haush.artik.	1885	[redacted]
20	06Q3	0.0625	5	Haarpflege	1916	[redacted]
20	06Q3	0.0625	30	Babypflege	2067	[redacted]

20	06Q3	0.0625	51	Fu	69	[redacted]
20	06Q3	0.0546875	53	Posten/Saiso	-1	?
20	06Q3	0.0546875	4	Kosmetik-Pap	1241	BrandB2
20	06Q3	0.0546875	39	Hand	1999	[redacted]
20	06Q3	0.0546875	5	Haarpflege	4	[redacted]
20	06Q3	0.0546875	42	Mund/Zahn	58	[redacted]
20	06Q3	0.0546875	54	Bad	86	[redacted]
20	06Q3	0.0546875	15	WPR	221	[redacted]
20	06Q3	0.0546875	11	Haarstyling	756	[redacted]
20	06Q3	0.0546875	11	Haarstyling	65	[redacted]
20	06Q3	0.0546875	56	Filter&Folien	1245	[redacted]
20	06Q3	0.0546875	40	Bonbon	1775	[redacted]
20	06Q3	0.046875	12	Insekt/Pflan	223	[redacted]
20	06Q3	0.046875	15	WPR	2098	[redacted]
20	06Q3	0.046875	51	Fu	207	dm-BrandK1
20	06Q3	0.046875	15	WPR	249	[redacted]
20	06Q3	0.046875	9	Hund	219	dm-BrandB1
20	06Q3	0.046875	34	Fotozubehr	205	[redacted]
20	06Q3	0.046875	57	Deo	218	dm-BrandK1 [redacted]
20	06Q3	0.046875	60	Gesicht	162	[redacted]
20	06Q3	0.046875	37	Babyglasnahr	1847	BrandL1
20	06Q3	0.046875	19	Hygienepapie	1275	[redacted]
20	06Q3	0.046875	15	WPR	264	BrandJ1
20	06Q3	0.046875	40	Bonbon	381	[redacted]
20	06Q3	0.046875	119	Augen Make u	156	[redacted]
20	06Q3	0.046875	56	Filter&Folien	1244	[redacted]
20	06Q3	0.046875	15	WPR	895	[redacted]
20	06Q3	0.046875	15	WPR	269	[redacted]
20	06Q3	0.046875	34	Fotozubehr	2189	[redacted]
20	06Q3	0.046875	40	Bonbon	1793	[redacted]
20	06Q3	0.046875	15	WPR	212	[redacted]

***06Q4*, Cluster 18**

Cluster	Quarter	*Centroid* Value	Sub-Assortment		Brand	
18	06Q4	0.162079510703364	39	Hand	201	dm-green-BrandT1
18	06Q4	0.149847094801223	42	Mund/Zahn	187	[redacted]
18	06Q4	0.149847094801223	5	Haarpflege	201	dm-green-BrandT1
18	06Q4	0.131498470948012	116	Modul li	-1	?
18	06Q4	0.116207951070336	106	Modul klein	731	green-BrandU1
18	06Q4	0.113149847094801	5	Haarpflege	33	BrandZ1
18	06Q4	0.113149847094801	66	Nassrasur	218	dm-BrandK1 [redacted]

18	06Q4	0.113149847094801	18	Haush.artik.	1884	[redacted]
18	06Q4	0.113149847094801	40	Bonbon	1772	BrandC2
18	06Q4	0.110091743119266	106	Modul klein	-1	?
18	06Q4	0.110091743119266	58	Erlebnis dm	1265	BrandY1
18	06Q4	0.103975535168196	45	Kerzen	2076	BrandH2
18	06Q4	0.103975535168196	61	Bodylotion	201	dm-green-BrandT1
18	06Q4	0.100917431192661	30	Babypflege	59	[redacted]
18	06Q4	0.100917431192661	61	Bodylotion	4	[redacted]
18	06Q4	0.100917431192661	3	Damenhygiene	2191	BrandA2
18	06Q4	0.0978593272171254	36	Entferner	207	dm-BrandK1
18	06Q4	0.0948012232415902	11	Haarstyling	65	[redacted]
18	06Q4	0.0917431192660551	57	Deo	5	[redacted]
18	06Q4	0.0886850152905199	42	Mund/Zahn	32	[redacted]
18	06Q4	0.0886850152905199	45	Kerzen	2075	BrandR1
18	06Q4	0.0886850152905199	15	WPR	238	BrandF2
18	06Q4	0.0886850152905199	60	Gesicht	174	[redacted]
18	06Q4	0.0886850152905199	15	WPR	221	[redacted]
18	06Q4	0.0886850152905199	16	Duftkerzen-M	212	[redacted]
18	06Q4	0.0856269113149847	127	Ebelin Body	200	[redacted]
18	06Q4	0.0795107033639144	54	Bad	201	dm-green-BrandT1
18	06Q4	0.0764525993883792	57	Deo	173	[redacted]
18	06Q4	0.0764525993883792	36	Entferner	213	dm-BrandO1
18	06Q4	0.073394495412844	119	Augen Make u	29	[redacted]
18	06Q4	0.073394495412844	78	Pharma	351	[redacted]
18	06Q4	0.0703363914373089	3	Damenhygiene	267	[redacted]
18	06Q4	0.0703363914373089	5	Haarpflege	91	[redacted]
18	06Q4	0.0703363914373089	42	Mund/Zahn	1140	[redacted]
18	06Q4	0.0703363914373089	42	Mund/Zahn	85	[redacted]
18	06Q4	0.0672782874617737	110	Feinstrick 1	1988	[redacted]
18	06Q4	0.0672782874617737	5	Haarpflege	967	[redacted]
18	06Q4	0.0672782874617737	107	Modul mittel	-1	?
18	06Q4	0.0642201834862385	39	Hand	257	[redacted]
18	06Q4	0.0642201834862385	54	Bad	782	[redacted]
18	06Q4	0.0642201834862385	5	Haarpflege	257	[redacted]
18	06Q4	0.0611620795107034	30	Babypflege	2067	[redacted]
18	06Q4	0.0611620795107034	54	Bad	27	[redacted]
18	06Q4	0.0611620795107034	42	Mund/Zahn	58	[redacted]
18	06Q4	0.0611620795107034	54	Bad	30	[redacted]
18	06Q4	0.0581039755351682	78	Pharma	897	[redacted]
18	06Q4	0.0581039755351682	56	Filter&Folien	1245	[redacted]
18	06Q4	0.055045871559633	15	WPR	269	[redacted]
18	06Q4	0.055045871559633	39	Hand	10	[redacted]
18	06Q4	0.055045871559633	116	Modul li	212	[redacted]

18	06Q4	0.0519877675840979	107	Modul mittel	212	
18	06Q4	0.0519877675840979	72	Tee	2000	

Appendix F

Details to Stable Clusters, Store 518

F.1 *Centroid* values of *KhK* graph *06Q1*

See Section 4.4.1.

Cluster	S-A ID	Brand ID	*Centroid* Value	Quarter	Sub-Assortment	Brand
9	19	214	0.795880149812734	06Q1	Hygienepapie	dm-BrandN1
6	6	214	0.773109243697479	06Q1	Haushaltspap	dm-BrandN1
9	6	214	0.76498127340824	06Q1	Haushaltspap	dm-BrandN1
6	19	214	0.752100840336134	06Q1	Hygienepapie	dm-BrandN1
4	81	731	0.739463601532567	06Q1	green-BrandU1	green-BrandU1
4	19	214	0.735632183908046	06Q1	Hygienepapie	dm-BrandN1
4	6	214	0.715197956577267	06Q1	Haushaltspap	dm-BrandN1
9	4	213	0.687265917602996	06Q1	Kosmetik-Pap	dm-BrandO1
9	56	212	0.626404494382023	06Q1	Filter&Folien	
4	30	206	0.61941251596424	06Q1	Babypflege	dm-BrandQ1
6	4	213	0.600840336134454	06Q1	Kosmetik-Pap	dm-BrandO1
4	56	212	0.582375478927203	06Q1	Filter&Folien	
6	81	731	0.579831932773109	06Q1	green-BrandU1	green-BrandU1
9	81	731	0.577715355805243	06Q1	green-BrandU1	green-BrandU1
6	15	202	0.525210084033613	06Q1	WPR	dm-BrandD
9	39	207	0.524344569288389	06Q1	Hand	dm-BrandK1
9	18	212	0.523408239700375	06Q1	Haush.artik.	
4	4	213	0.492975734355045	06Q1	Kosmetik-Pap	dm-BrandO1
4	15	202	0.491698595146871	06Q1	WPR	dm-BrandD
4	27	206	0.485312899106003	06Q1	Windeln	dm-BrandQ1
6	56	212	0.420168067226891	06Q1	Filter&Folien	
9	15	202	0.401685393258427	06Q1	WPR	dm-BrandD
4	18	212	0.397190293742018	06Q1	Haush.artik.	

F.2 *Bestmatch*- and *BestCD*-graphs of Store 518, with Eliminations

See Section 4.4.2.

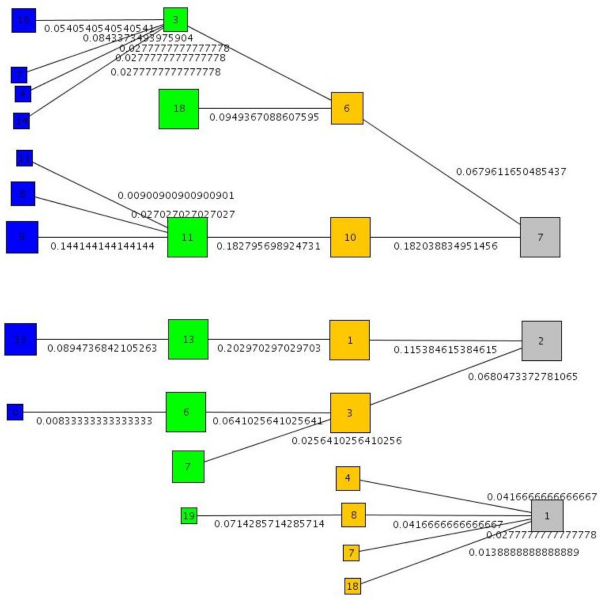


Figure F.1: *Bestmatch*-graph with eliminations, store 518.

F.3 *Centroids* of Stable Clusters

All *centroid* values greater than 5%.

F.3.1 The *Mainstream* Cluster

06Q2, Cluster 4

Cluster	S-A ID	Brand ID	*Centroid* Value	Quarter	Sub-Assortment	Brand
4	103	-1	0.447314049586777	06Q2	Tchibo	?
4	48	210	0.398760330578512	06Q2	Sonne	dm-BrandS1
4	44	-1	0.397727272727273	06Q2	Presse	?
4	51	207	0.22004132231405	06Q2	Fu	dm-BrandK1
4	110	1988	0.210743801652893	06Q2	Feinstrick 1	
4	66	218	0.171487603305785	06Q2	Nassrasur	dm-BrandK1
4	53	-1	0.158057851239669	06Q2	Posten/Saiso	?

4	40	1772	0.147727272727273	06Q2	Bonbon	BrandC2
4	36	207	0.142561983471074	06Q2	Entferner	dm-BrandK1
4	131	1988	0.138429752066116	06Q2	Grobstrick	[redacted]
4	127	200	0.135330578512397	06Q2	Ebelin Body	[redacted]
4	54	218	0.134297520661157	06Q2	Bad	dm-BrandK1 [redacted]
4	42	187	0.133264462809917	06Q2	Mund/Zahn	[redacted]
4	20	219	0.101239669421488	06Q2	Nager	dm-BrandB1
4	60	162	0.0898760330578512	06Q2	Gesicht	[redacted]
4	42	32	0.0888429752066116	06Q2	Mund/Zahn	[redacted]
4	11	65	0.0867768595041322	06Q2	Haarstyling	[redacted]
4	75	381	0.0847107438016529	06Q2	Reform	[redacted]
4	75	306	0.0774793388429752	06Q2	Reform	[redacted]
4	18	1884	0.0764462809917355	06Q2	Haush.artik.	[redacted]
4	36	213	0.0764462809917355	06Q2	Entferner	dm-BrandO1
4	3	2191	0.0702479338842975	06Q2	Damenhygiene	BrandA2
4	5	33	0.0681818181818182	06Q2	Haarpflege	BrandZ1
4	42	58	0.0671487603305785	06Q2	Mund/Zahn	[redacted]
4	61	4	0.0671487603305785	06Q2	Bodylotion	[redacted]
4	119	29	0.0661157024793388	06Q2	Augen Make u	[redacted]
4	111	1988	0.0619834710743802	06Q2	Feinstrick 2	[redacted]
4	119	1290	0.0619834710743802	06Q2	Augen Make u	[redacted]
4	39	201	0.0619834710743802	06Q2	Hand	dm-green-BrandT1
4	15	270	0.0609504132231405	06Q2	WPR	[redacted]
4	121	1309	0.0599173553719008	06Q2	Nagelpflege	[redacted]
4	120	1290	0.0599173553719008	06Q2	Lippenpflege	[redacted]
4	56	1245	0.0599173553719008	06Q2	Filter&Folien	[redacted]
4	57	218	0.0599173553719008	06Q2	Deo	dm-BrandK1 [redacted]
4	6	1265	0.0588842975206612	06Q2	Haushaltspap	BrandY1
4	5	66	0.0578512396694215	06Q2	Haarpflege	[redacted]
4	57	5	0.0568181818181818	06Q2	Deo	[redacted]
4	5	952	0.0557851239669422	06Q2	Haarpflege	[redacted]
4	28	210	0.0557851239669422	06Q2	Sonnenbrille	dm-BrandS1
4	5	257	0.0547520661157025	06Q2	Haarpflege	[redacted]
4	16	433	0.0547520661157025	06Q2	Duftkerzen-M	[redacted]
4	9	219	0.0537190082644628	06Q2	Hund	dm-BrandB1
4	40	381	0.0537190082644628	06Q2	Bonbon	[redacted]
4	3	267	0.0526859504132231	06Q2	Damenhygiene	[redacted]
4	15	262	0.0526859504132231	06Q2	WPR	[redacted]
4	12	1881	0.0516528925619835	06Q2	Insekt/Pflan	[redacted]
4	112	-1	0.0516528925619835	06Q2	Ostern klein	?
4	30	59	0.0516528925619835	06Q2	Babypflege	[redacted]
4	103	490	0.0506198347107438	06Q2	Tchibo	[redacted]
4	54	257	0.0506198347107438	06Q2	Bad	[redacted]

4	42	891	0.0506198347107438	06Q2	Mund/Zahn	[redacted]

***06Q3*, Cluster 1**

Cluster	S-A ID	Brand ID	*Centroid* Value	Quarter	Sub-Assortment	Brand
1	44	-1	0.470588235294118	06Q3	Presse	?
1	103	-1	0.412874583795782	06Q3	Tchibo	?
1	51	207	0.366259711431743	06Q3	Fu	dm-BrandK1
1	48	210	0.280799112097669	06Q3	Sonne	dm-BrandS1
1	66	218	0.198668146503885	06Q3	Nassrasur	dm-BrandK1 [redacted]
1	127	200	0.172031076581576	06Q3	Ebelin Body	[redacted]
1	53	-1	0.16315205327414	06Q3	Posten/Saiso	?
1	40	1772	0.146503884572697	06Q3	Bonbon	BrandC2
1	131	1988	0.13984461709212	06Q3	Grobstrick	[redacted]
1	54	218	0.130965593784684	06Q3	Bad	dm-BrandK1 [redacted]
1	11	65	0.123196448390677	06Q3	Haarstyling	[redacted]
1	36	213	0.113207547169811	06Q3	Entferner	dm-BrandO1
1	119	1290	0.110987791342952	06Q3	Augen Make u	[redacted]
1	42	32	0.105438401775805	06Q3	Mund/Zahn	[redacted]
1	36	207	0.102108768035516	06Q3	Entferner	dm-BrandK1
1	57	218	0.0887902330743618	06Q3	Deo	dm-BrandK1 [redacted]
1	5	33	0.0887902330743618	06Q3	Haarpflege	BrandZ1
1	40	381	0.0832408435072142	06Q3	Bonbon	[redacted]
1	110	1988	0.0810210876803552	06Q3	Feinstrick 1	[redacted]
1	20	219	0.0799112097669256	06Q3	Nager	dm-BrandB1
1	34	-1	0.0788013318534961	06Q3	Fotozubehr	?
1	15	270	0.0765815760266371	06Q3	WPR	[redacted]
1	5	66	0.0754716981132075	06Q3	Haarpflege	[redacted]
1	16	433	0.074361820199778	06Q3	Duftkerzen-M	[redacted]
1	9	219	0.0732519422863485	06Q3	Hund	dm-BrandB1
1	75	381	0.0710321864594895	06Q3	Reform	[redacted]
1	60	162	0.0710321864594895	06Q3	Gesicht	[redacted]
1	42	58	0.0699223085460599	06Q3	Mund/Zahn	[redacted]
1	42	1140	0.0688124306326304	06Q3	Mund/Zahn	[redacted]
1	42	187	0.0688124306326304	06Q3	Mund/Zahn	[redacted]
1	39	1999	0.0677025527192009	06Q3	Hand	[redacted]
1	18	1884	0.0677025527192009	06Q3	Haush.artik.	[redacted]
1	121	1309	0.0654827968923418	06Q3	Nagelpflege	[redacted]
1	70	2359	0.0577136514983352	06Q3	Fotoarbeiten	[redacted] [redacted]
1	15	249	0.0554938956714761	06Q3	WPR	[redacted]
1	12	1881	0.0554938956714761	06Q3	Insekt/Pflan	[redacted]
1	75	306	0.0543840177580466	06Q3	Reform	[redacted]

1	45	2075	0.0532741398446171	06Q3	Kerzen	BrandR1
1	57	5	0.0521642619311876	06Q3	Deo	[redacted]
1	60	176	0.0521642619311876	06Q3	Gesicht	[redacted] [redacted]
1	12	212	0.0521642619311876	06Q3	Insekt/Pflan	[redacted]
1	5	257	0.051054384017758	06Q3	Haarpflege	[redacted]

(*06Q4*, Cluster 21)

Cluster	S-A ID	Brand ID	*Centroid* Value	Quarter	Sub-Assortment	Brand
21	103	-1	0.439252336448598	06Q4	Tchibo	?
21	44	-1	0.42803738317757	06Q4	Presse	?
21	106	-1	0.355140186915888	06Q4	Modul klein	?
21	116	-1	0.353271028037383	06Q4	Modul li	?
21	131	1988	0.254205607476636	06Q4	Grobstrick	[redacted]
21	66	218	0.222429906542056	06Q4	Nassrasur	dm-BrandK1 [redacted]
21	110	1988	0.211214953271028	06Q4	Feinstrick 1	[redacted]
21	40	1772	0.190654205607477	06Q4	Bonbon	BrandC2
21	16	212	0.188785046728972	06Q4	Duftkerzen-M	[redacted]
21	6	1265	0.153271028037383	06Q4	Haushaltspap	BrandY1
21	51	207	0.138317757009346	06Q4	Fu	dm-BrandK1
21	36	207	0.134579439252336	06Q4	Entferner	dm-BrandK1
21	42	187	0.134579439252336	06Q4	Mund/Zahn	[redacted]
21	42	58	0.121495327102804	06Q4	Mund/Zahn	[redacted]
21	45	2075	0.117757009345794	06Q4	Kerzen	BrandR1
21	107	-1	0.11588785046729	06Q4	Modul mittel	?
21	119	1290	0.108411214953271	06Q4	Augen Make u	[redacted]
21	5	33	0.106542056074766	06Q4	Haarpflege	BrandZ1
21	11	65	0.100934579439252	06Q4	Haarstyling	[redacted]
21	56	1245	0.100934579439252	06Q4	Filter&Folien	[redacted]
21	45	2076	0.0990654205607477	06Q4	Kerzen	BrandH2
21	54	218	0.0953271028037383	06Q4	Bad	dm-BrandK1 [redacted]
21	20	219	0.0953271028037383	06Q4	Nager	dm-BrandB1
21	53	-1	0.0953271028037383	06Q4	Posten/Saiso	?
21	127	200	0.0953271028037383	06Q4	Ebelin Body	[redacted]
21	39	201	0.091588785046729	06Q4	Hand	dm-green-BrandT1
21	54	201	0.0897196261682243	06Q4	Bad	dm-green-BrandT1
21	3	2191	0.0878504672897196	06Q4	Damenhygiene	BrandA2
21	40	381	0.085981308411215	06Q4	Bonbon	[redacted]
21	18	1884	0.0841121495327103	06Q4	Haush.artik.	[redacted]
21	116	212	0.0803738317757009	06Q4	Modul li	[redacted]
21	15	270	0.0803738317757009	06Q4	WPR	[redacted]
21	34	-1	0.0785046728971963	06Q4	Fotozubehr	?

21	57	5	0.0785046728971963	06Q4	Deo	
21	40	1789	0.0766355140186916	06Q4	Bonbon	
21	70	2359	0.0766355140186916	06Q4	Fotoarbeiten	
21	15	238	0.0766355140186916	06Q4	WPR	BrandF2
21	119	29	0.0747663551401869	06Q4	Augen Make u	
21	5	257	0.0747663551401869	06Q4	Haarpflege	
21	42	32	0.0747663551401869	06Q4	Mund/Zahn	
21	118	29	0.0728971962616822	06Q4	Make up	
21	3	267	0.0728971962616822	06Q4	Damenhygiene	
21	30	2067	0.0710280373831776	06Q4	Babypflege	
21	15	221	0.0691588785046729	06Q4	WPR	
21	60	284	0.0654205607476635	06Q4	Gesicht	
21	9	219	0.0635514018691589	06Q4	Hund	dm-BrandB1
21	58	1265	0.0635514018691589	06Q4	Erlebnis dm	BrandY1
21	36	213	0.0635514018691589	06Q4	Entferner	dm-BrandO1
21	34	205	0.0635514018691589	06Q4	Fotozubehr	
21	120	1309	0.0635514018691589	06Q4	Lippenpflege	
21	57	218	0.0616822429906542	06Q4	Deo	dm-BrandK1
21	107	212	0.0616822429906542	06Q4	Modul mittel	
21	60	176	0.0616822429906542	06Q4	Gesicht	
21	5	66	0.0616822429906542	06Q4	Haarpflege	
21	5	113	0.0598130841121495	06Q4	Haarpflege	
21	5	201	0.0598130841121495	06Q4	Haarpflege	dm-green-BrandT1
21	60	162	0.0579439252336449	06Q4	Gesicht	
21	58	32	0.0579439252336449	06Q4	Erlebnis dm	
21	5	91	0.0579439252336449	06Q4	Haarpflege	
21	75	381	0.0579439252336449	06Q4	Reform	
21	59	199	0.0579439252336449	06Q4	KLKD-Push	BrandG2
21	39	257	0.0579439252336449	06Q4	Hand	
21	61	4	0.0560747663551402	06Q4	Bodylotion	
21	108	-1	0.0560747663551402	06Q4	Modul gross	?
21	60	174	0.0542056074766355	06Q4	Gesicht	
21	54	1474	0.0542056074766355	06Q4	Bad	BrandM1
21	72	1213	0.0542056074766355	06Q4	Tee	
21	5	1916	0.0542056074766355	06Q4	Haarpflege	
21	75	306	0.0542056074766355	06Q4	Reform	
21	119	156	0.0542056074766355	06Q4	Augen Make u	
21	106	731	0.0523364485981309	06Q4	Modul klein	green-BrandU1
21	39	1999	0.0523364485981309	06Q4	Hand	
21	19	1275	0.0504672897196262	06Q4	Hygienepapie	
21	78	351	0.0504672897196262	06Q4	Pharma	

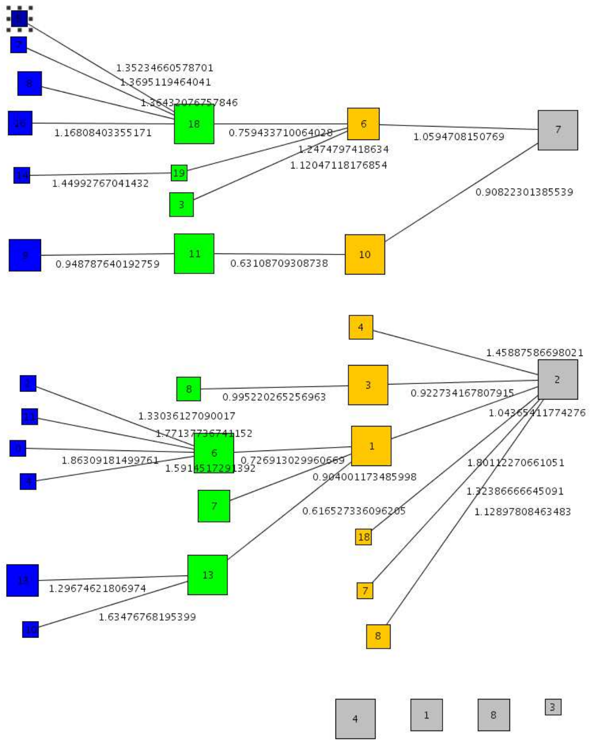


Figure F.2: *BestCD*-graph with eliminations, store 518.

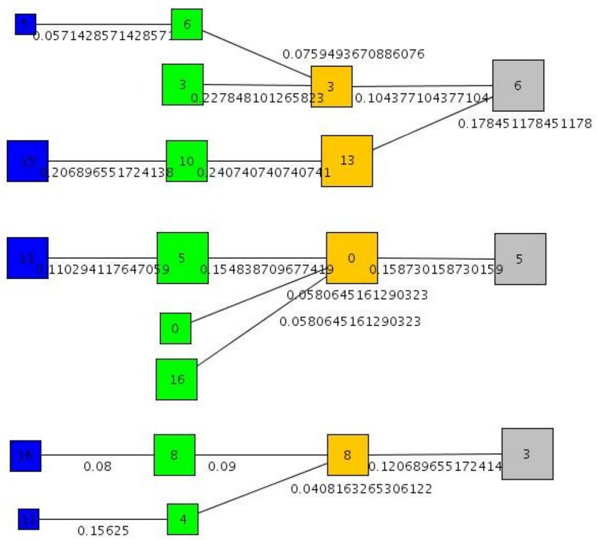


Figure F.3: *BestMatch*-graph with eliminations and customer intersection, store 518.

(*06Q1*, Cluster 41)

Cluster	S-A ID	Brand ID	*Centroid* Value	Quarter	Sub-Assortment	Brand
41	103	-1	0.494609164420485	06Q1	Tchibo	?
41	16	433	0.211590296495957	06Q1	Duftkerzen-M	[redacted]
41	112	-1	0.191374663072776	06Q1	Ostern klein	?
41	51	207	0.172506738544474	06Q1	Fu	dm-BrandK1
41	11	65	0.149595687331536	06Q1	Haarstyling	[redacted]
41	127	200	0.137466307277628	06Q1	Ebelin Body	[redacted]
41	40	1772	0.133423180592992	06Q1	Bonbon	BrandC2
41	53	-1	0.122641509433962	06Q1	Posten/Saiso	?
41	48	210	0.121293800539084	06Q1	Sonne	dm-BrandS1
41	66	218	0.121293800539084	06Q1	Nassrasur	dm-BrandK1 [redacted]
41	45	2076	0.119946091644205	06Q1	Kerzen	BrandH2
41	36	207	0.117250673854447	06Q1	Entferner	dm-BrandK1
41	6	1265	0.110512129380054	06Q1	Haushaltspap	BrandY1
41	20	219	0.105121293800539	06Q1	Nager	dm-BrandB1
41	119	1290	0.0983827493261455	06Q1	Augen Make u	[redacted]
41	5	66	0.0943396226415094	06Q1	Haarpflege	[redacted]
41	59	199	0.0929919137466307	06Q1	KLKD-Push	BrandG2
41	42	32	0.0929919137466307	06Q1	Mund/Zahn	[redacted]
41	119	29	0.0835579514824798	06Q1	Augen Make u	[redacted]
41	5	33	0.0822102425876011	06Q1	Haarpflege	BrandZ1
41	42	187	0.0795148247978437	06Q1	Mund/Zahn	[redacted]
41	131	1988	0.0795148247978437	06Q1	Grobstrick	[redacted]
41	110	1988	0.0795148247978437	06Q1	Feinstrick 1	[redacted]
41	9	219	0.0768194070080863	06Q1	Hund	dm-BrandB1
41	14	2106	0.0768194070080863	06Q1	Vogel	[redacted]
41	3	267	0.0768194070080863	06Q1	Damenhygiene	[redacted]
41	3	2191	0.0754716981132075	06Q1	Damenhygiene	BrandA2
41	60	1365	0.0741239892183288	06Q1	Gesicht	[redacted]
41	113	-1	0.0714285714285714	06Q1	Ostern mitte	?
41	45	2075	0.0714285714285714	06Q1	Kerzen	BrandR1
41	36	213	0.0714285714285714	06Q1	Entferner	dm-BrandO1
41	30	59	0.068733153638814	06Q1	Babypflege	[redacted]
41	60	174	0.068733153638814	06Q1	Gesicht	[redacted]
41	40	381	0.068733153638814	06Q1	Bonbon	[redacted]
41	75	381	0.0660377358490566	06Q1	Reform	[redacted]
41	18	1884	0.0646900269541779	06Q1	Haush.artik.	[redacted]
41	53	133	0.0646900269541779	06Q1	Posten/Saiso	[redacted]
41	72	1213	0.0646900269541779	06Q1	Tee	[redacted]
41	39	201	0.0633423180592992	06Q1	Hand	dm-green-BrandT1
41	15	270	0.0633423180592992	06Q1	WPR	[redacted]
41	19	1275	0.0619946091644205	06Q1	Hygienepapie	[redacted]
41	54	218	0.0606469002695418	06Q1	Bad	dm-BrandK1 [redacted]

41	42	58	0.0606469002695418	06Q1	Mund/Zahn	[redacted]
41	39	257	0.0592991913746631	06Q1	Hand	[redacted]
41	120	1290	0.0566037735849057	06Q1	Lippenpflege	[redacted]
41	15	439	0.0566037735849057	06Q1	WPR	[redacted]
41	34	-1	0.0566037735849057	06Q1	Fotozubehr	?
41	57	5	0.055256064690027	06Q1	Deo	[redacted]
41	15	249	0.055256064690027	06Q1	WPR	[redacted]
41	60	176	0.0539083557951483	06Q1	Gesicht	[redacted] [redacted]
41	40	1789	0.0539083557951483	06Q1	Bonbon	[redacted]
41	42	1140	0.0525606469002695	06Q1	Mund/Zahn	[redacted]

F.3.2 The *Premium/Traditional* Cluster

06Q2, Cluster 12

Cluster	S-A ID	Brand ID	*Centroid* Value	Quarter	Sub-Assortment	Brand
12	44	-1	0.340909090909091	06Q2	Presse	?
12	103	-1	0.181818181818182	06Q2	Tchibo	?
12	48	210	0.167355371900826	06Q2	Sonne	dm-BrandS1
12	61	4	0.159090909090909	06Q2	Bodylotion	[redacted]
12	30	59	0.144628099173554	06Q2	Babypflege	[redacted]
12	6	1265	0.136363636363636	06Q2	Haushaltspap	BrandY1
12	11	65	0.130165289256198	06Q2	Haarstyling	[redacted]
12	42	187	0.12603305785124	06Q2	Mund/Zahn	[redacted]
12	5	33	0.12396694214876	06Q2	Haarpflege	BrandZ1
12	51	207	0.12396694214876	06Q2	Fu	dm-BrandK1
12	42	32	0.12396694214876	06Q2	Mund/Zahn	[redacted]
12	3	2191	0.121900826446281	06Q2	Damenhygiene	BrandA2
12	40	1772	0.107438016528926	06Q2	Bonbon	BrandC2
12	60	176	0.105371900826446	06Q2	Gesicht	[redacted] [redacted]
12	36	213	0.0991735537190083	06Q2	Entferner	dm-BrandO1
12	127	200	0.0950413223140496	06Q2	Ebelin Body	[redacted]
12	18	1884	0.0867768595041322	06Q2	Haush.artik.	[redacted]
12	110	1988	0.0867768595041322	06Q2	Feinstrick 1	[redacted]
12	53	-1	0.0847107438016529	06Q2	Posten/Saiso	?
12	39	257	0.0826446280991736	06Q2	Hand	[redacted]
12	56	1243	0.0826446280991736	06Q2	Filter&Folien	BrandX1
12	5	66	0.0805785123966942	06Q2	Haarpflege	[redacted]
12	12	1881	0.0785123966942149	06Q2	Insekt/Pflan	[redacted]
12	12	212	0.0785123966942149	06Q2	Insekt/Pflan	[redacted]
12	5	257	0.0785123966942149	06Q2	Haarpflege	[redacted]
12	39	61	0.0743801652892562	06Q2	Hand	[redacted]

12	119	1290	0.0743801652892562	06Q2	Augen Make u	[redacted]
12	66	218	0.0723140495867769	06Q2	Nassrasur	dm-BrandK1 [redacted]
12	45	2075	0.0681818181818182	06Q2	Kerzen	BrandR1
12	39	1999	0.0681818181818182	06Q2	Hand	[redacted]
12	40	381	0.0681818181818182	06Q2	Bonbon	[redacted]
12	3	267	0.0661157024793388	06Q2	Damenhygiene	[redacted]
12	51	69	0.0661157024793388	06Q2	Fu	[redacted]
12	58	32	0.0661157024793388	06Q2	Erlebnis dm	[redacted]
12	103	490	0.0661157024793388	06Q2	Tchibo	[redacted]
12	60	284	0.0661157024793388	06Q2	Gesicht	[redacted]
12	36	207	0.0640495867768595	06Q2	Entferner	dm-BrandK1
12	4	1241	0.0640495867768595	06Q2	Kosmetik-Pap	BrandB2
12	56	1245	0.0619834710743802	06Q2	Filter&Folien	[redacted]
12	119	29	0.0619834710743802	06Q2	Augen Make u	[redacted]
12	66	71	0.0619834710743802	06Q2	Nassrasur	[redacted]
12	57	173	0.0619834710743802	06Q2	Deo	[redacted]
12	19	1275	0.0599173553719008	06Q2	Hygienepapie	[redacted]
12	120	1309	0.0599173553719008	06Q2	Lippenpflege	[redacted]
12	40	314	0.0599173553719008	06Q2	Bonbon	[redacted]
12	58	1265	0.0578512396694215	06Q2	Erlebnis dm	BrandY1
12	42	58	0.0578512396694215	06Q2	Mund/Zahn	[redacted]
12	75	306	0.0578512396694215	06Q2	Reform	[redacted]
12	78	897	0.0557851239669422	06Q2	Pharma	[redacted]
12	5	4	0.0557851239669422	06Q2	Haarpflege	[redacted]
12	5	91	0.0557851239669422	06Q2	Haarpflege	[redacted]
12	40	1789	0.0557851239669422	06Q2	Bonbon	[redacted]
12	12	223	0.0557851239669422	06Q2	Insekt/Pflan	[redacted]
12	54	218	0.0557851239669422	06Q2	Bad	dm-BrandK1 [redacted]
12	42	1140	0.0537190082644628	06Q2	Mund/Zahn	[redacted]
12	54	257	0.0537190082644628	06Q2	Bad	[redacted]
12	57	5	0.0537190082644628	06Q2	Deo	[redacted]
12	15	270	0.0537190082644628	06Q2	WPR	[redacted]
12	119	156	0.0537190082644628	06Q2	Augen Make u	[redacted]

***06Q3*, Cluster 31**

Cluster	S-A ID	Brand ID	*Centroid* Value	Quarter	Sub-Assortment	Brand
31	44	-1	0.353765323992995	06Q3	Presse	?
31	103	-1	0.320490367775832	06Q3	Tchibo	?
31	42	187	0.252189141856392	06Q3	Mund/Zahn	[redacted]
31	3	2191	0.145359019264448	06Q3	Damenhygiene	BrandA2
31	5	33	0.141856392294221	06Q3	Haarpflege	BrandZ1

31	6	1265	0.138353765323993	06Q3	Haushaltspap	BrandY1
31	48	210	0.134851138353765	06Q3	Sonne	dm-BrandS1
31	66	218	0.134851138353765	06Q3	Nassrasur	dm-BrandK1 [redacted]
31	51	207	0.120840630472855	06Q3	Fu	dm-BrandK1
31	40	1772	0.115586690017513	06Q3	Bonbon	BrandC2
31	30	59	0.113835376532399	06Q3	Babypflege	[redacted]
31	18	1884	0.110332749562172	06Q3	Haush.artik.	[redacted]
31	11	65	0.106830122591944	06Q3	Haarstyling	[redacted]
31	56	1245	0.106830122591944	06Q3	Filter&Folien	[redacted]
31	61	4	0.0928196147110333	06Q3	Bodylotion	[redacted]
31	3	267	0.0928196147110333	06Q3	Damenhygiene	[redacted]
31	127	200	0.0928196147110333	06Q3	Ebelin Body	[redacted]
31	15	262	0.0928196147110333	06Q3	WPR	[redacted]
31	57	5	0.0910683012259194	06Q3	Deo	[redacted]
31	60	176	0.0893169877408056	06Q3	Gesicht	[redacted] [redacted]
31	36	213	0.0893169877408056	06Q3	Entferner	dm-BrandO1
31	5	257	0.0875656742556918	06Q3	Haarpflege	[redacted]
31	119	29	0.0858143607705779	06Q3	Augen Make u	[redacted]
31	39	201	0.0823117338003503	06Q3	Hand	dm-green-BrandT1
31	103	490	0.0823117338003503	06Q3	Tchibo	[redacted]
31	40	314	0.0805604203152364	06Q3	Bonbon	[redacted]
31	36	207	0.0788091068301226	06Q3	Entferner	dm-BrandK1
31	39	257	0.0788091068301226	06Q3	Hand	[redacted]
31	39	61	0.0788091068301226	06Q3	Hand	[redacted]
31	42	32	0.0770577933450088	06Q3	Mund/Zahn	[redacted]
31	56	1243	0.0770577933450088	06Q3	Filter&Folien	BrandX1
31	53	-1	0.0770577933450088	06Q3	Posten/Saiso	?
31	19	1275	0.0753064798598949	06Q3	Hygienepapie	[redacted]
31	15	270	0.0665499124343257	06Q3	WPR	[redacted]
31	75	381	0.0647985989492119	06Q3	Reform	[redacted]
31	5	952	0.0647985989492119	06Q3	Haarpflege	[redacted]
31	61	257	0.0647985989492119	06Q3	Bodylotion	[redacted]
31	54	257	0.0630472854640981	06Q3	Bad	[redacted]
31	15	1273	0.0595446584938704	06Q3	WPR	[redacted]
31	66	71	0.0577933450087566	06Q3	Nassrasur	[redacted]
31	110	1988	0.0577933450087566	06Q3	Feinstrick 1	[redacted]
31	60	1365	0.0577933450087566	06Q3	Gesicht	[redacted]
31	42	58	0.0560420315236427	06Q3	Mund/Zahn	[redacted]
31	56	1244	0.0542907180385289	06Q3	Filter&Folien	[redacted]
31	58	1265	0.0525394045534151	06Q3	Erlebnis dm	BrandY1
31	40	1775	0.0507880910683012	06Q3	Bonbon	[redacted]
31	5	91	0.0507880910683012	06Q3	Haarpflege	[redacted]
31	40	381	0.0507880910683012	06Q3	Bonbon	[redacted]

31	54	20	0.0507880910683012	06Q3	Bad	[redacted]
31	15	212	0.0507880910683012	06Q3	WPR	[redacted]

(***06Q1***, Cluster 35)

Cluster	S-A ID	Brand ID	*Centroid* Value	Quarter	Sub-Assortment	Brand
35	103	-1	0.347826086956522	06Q1	Tchibo	?
35	66	218	0.304347826086957	06Q1	Nassrasur	dm-BrandK1 [redacted]
35	18	1884	0.254658385093168	06Q1	Haush.artik.	[redacted]
35	54	218	0.173913043478261	06Q1	Bad	dm-BrandK1 [redacted]
35	40	381	0.124223602484472	06Q1	Bonbon	[redacted]
35	75	381	0.118012422360248	06Q1	Reform	[redacted]
35	42	58	0.111801242236025	06Q1	Mund/Zahn	[redacted]
35	36	207	0.111801242236025	06Q1	Entferner	dm-BrandK1
35	53	-1	0.111801242236025	06Q1	Posten/Saiso	?
35	78	897	0.111801242236025	06Q1	Pharma	[redacted]
35	42	32	0.105590062111801	06Q1	Mund/Zahn	[redacted]
35	3	267	0.105590062111801	06Q1	Damenhygiene	[redacted]
35	48	210	0.105590062111801	06Q1	Sonne	dm-BrandS1
35	5	952	0.105590062111801	06Q1	Haarpflege	[redacted]
35	5	113	0.105590062111801	06Q1	Haarpflege	[redacted]
35	42	187	0.093167701863354	06Q1	Mund/Zahn	[redacted]
35	20	219	0.093167701863354	06Q1	Nager	dm-BrandB1
35	39	201	0.093167701863354	06Q1	Hand	dm-green-BrandT1
35	60	176	0.093167701863354	06Q1	Gesicht	[redacted]
35	11	65	0.093167701863354	06Q1	Haarstyling	[redacted]
35	51	207	0.093167701863354	06Q1	Fu	dm-BrandK1
35	61	274	0.0807453416149068	06Q1	Bodylotion	[redacted]
35	11	756	0.0807453416149068	06Q1	Haarstyling	[redacted]
35	6	1265	0.0807453416149068	06Q1	Haushaltspap	BrandY1
35	16	433	0.0807453416149068	06Q1	Duftkerzen-M	[redacted]
35	53	133	0.0807453416149068	06Q1	Posten/Saiso	[redacted]
35	58	1265	0.0807453416149068	06Q1	Erlebnis dm	BrandY1
35	54	201	0.0807453416149068	06Q1	Bad	dm-green-BrandT1
35	14	2106	0.0745341614906832	06Q1	Vogel	[redacted]
35	15	238	0.0745341614906832	06Q1	WPR	BrandF2
35	45	2076	0.0745341614906832	06Q1	Kerzen	BrandH2
35	57	218	0.0683229813664596	06Q1	Deo	dm-BrandK1 [redacted]
35	40	314	0.0683229813664596	06Q1	Bonbon	[redacted]
35	37	1847	0.0683229813664596	06Q1	Babyglasnahr	BrandL1
35	39	61	0.062111801242236	06Q1	Hand	[redacted]
35	15	2098	0.062111801242236	06Q1	WPR	[redacted]

35	72	2000	0.062111801242236	06Q1	Tee	[redacted]
35	12	223	0.062111801242236	06Q1	Insekt/Pflan	[redacted]
35	54	782	0.062111801242236	06Q1	Bad	[redacted]
35	30	2067	0.062111801242236	06Q1	Babypflege	[redacted]
35	112	-1	0.062111801242236	06Q1	Ostern klein	?
35	39	257	0.0559006211180124	06Q1	Hand	[redacted]
35	40	1772	0.0559006211180124	06Q1	Bonbon	BrandC2
35	42	1140	0.0559006211180124	06Q1	Mund/Zahn	[redacted]
35	112	212	0.0559006211180124	06Q1	Ostern klein	[redacted]
35	26	206	0.0559006211180124	06Q1	Flaschen & S	dm-BrandQ1
35	34	205	0.0559006211180124	06Q1	Fotozubehr	[redacted]
35	15	221	0.0559006211180124	06Q1	WPR	[redacted]
35	66	71	0.0559006211180124	06Q1	Nassrasur	[redacted]
35	119	29	0.0559006211180124	06Q1	Augen Make u	[redacted]
35	3	2191	0.0559006211180124	06Q1	Damenhygiene	BrandA2
35	60	174	0.0559006211180124	06Q1	Gesicht	[redacted]
35	103	490	0.0559006211180124	06Q1	Tchibo	[redacted]

F.3.3 The *Baby/Young Families* Cluster

06Q1, Cluster 11

Cluster	S-A ID	Brand ID	*Centroid* Value	Quarter	Sub-Assortment	Brand
11	27	899	0.463675213675214	06Q1	Windeln	BrandE2
11	79	199	0.423076923076923	06Q1	Baby-Push	BrandG2
11	103	-1	0.412393162393162	06Q1	Tchibo	?
11	37	1846	0.273504273504274	06Q1	Babyglasnahr	BrandA1
11	33	1846	0.202991452991453	06Q1	Babynahrung	BrandA1
11	26	206	0.19017094017094	06Q1	Flaschen & S	dm-BrandQ1
11	76	199	0.188034188034188	06Q1	Baby-Pull	BrandG2
11	33	1847	0.168803418803419	06Q1	Babynahrung	BrandL1
11	59	199	0.153846153846154	06Q1	KLKD-Push	BrandG2
11	112	-1	0.138888888888889	06Q1	Ostern klein	?
11	37	1847	0.136752136752137	06Q1	Babyglasnahr	BrandL1
11	37	731	0.117521367521368	06Q1	Babyglasnahr	green-BrandU1
11	30	59	0.115384615384615	06Q1	Babypflege	[redacted]
11	42	187	0.113247863247863	06Q1	Mund/Zahn	[redacted]
11	76	217	0.106837606837607	06Q1	Baby-Pull	[redacted]
11	66	218	0.102564102564103	06Q1	Nassrasur	dm-BrandK1 [redacted]
11	53	-1	0.10042735042735	06Q1	Posten/Saiso	?
11	18	1884	0.0961538461538462	06Q1	Haush.artik.	[redacted]
11	3	267	0.0854700854700855	06Q1	Damenhygiene	[redacted]

11	6	1265	0.0833333333333333	06Q1	Haushaltspap	BrandY1
11	53	133	0.0811965811965812	06Q1	Posten/Saiso	[redacted]
11	15	262	0.0811965811965812	06Q1	WPR	[redacted]
11	15	238	0.0769230769230769	06Q1	WPR	BrandF2
11	40	381	0.0747863247863248	06Q1	Bonbon	[redacted]
11	16	433	0.0726495726495727	06Q1	Duftkerzen-M	[redacted]
11	5	33	0.0726495726495727	06Q1	Haarpflege	BrandZ1
11	56	1244	0.0705128205128205	06Q1	Filter&Folien	[redacted]
11	45	2076	0.0683760683760684	06Q1	Kerzen	BrandH2
11	34	205	0.0662393162393162	06Q1	Fotozubehr	[redacted]
11	42	58	0.0662393162393162	06Q1	Mund/Zahn	[redacted]
11	30	2067	0.061965811965812	06Q1	Babypflege	[redacted]
11	131	1988	0.061965811965812	06Q1	Grobstrick	[redacted]
11	11	65	0.061965811965812	06Q1	Haarstyling	[redacted]
11	5	952	0.0598290598290598	06Q1	Haarpflege	[redacted]
11	30	274	0.0598290598290598	06Q1	Babypflege	[redacted]
11	3	2191	0.0598290598290598	06Q1	Damenhygiene	BrandA2
11	72	2000	0.0576923076923077	06Q1	Tee	[redacted]
11	56	1245	0.0576923076923077	06Q1	Filter&Folien	[redacted]
11	40	1772	0.0555555555555556	06Q1	Bonbon	BrandC2
11	33	1855	0.0534188034188034	06Q1	Babynahrung	[redacted]
11	5	91	0.0512820512820513	06Q1	Haarpflege	[redacted]

***06Q2*, Cluster 13**

Cluster	S-A ID	Brand ID	*Centroid* Value	Quarter	Sub-Assortment	Brand
13	79	199	0.461764705882353	06Q2	Baby-Push	BrandG2
13	103	-1	0.386764705882353	06Q2	Tchibo	?
13	27	899	0.379411764705882	06Q2	Windeln	BrandE2
13	44	-1	0.302941176470588	06Q2	Presse	?
13	59	199	0.222058823529412	06Q2	KLKD-Push	BrandG2
13	37	1846	0.217647058823529	06Q2	Babyglasnahr	BrandA1
13	33	1846	0.175	06Q2	Babynahrung	BrandA1
13	37	1847	0.173529411764706	06Q2	Babyglasnahr	BrandL1
13	48	210	0.152941176470588	06Q2	Sonne	dm-BrandS1
13	26	206	0.148529411764706	06Q2	Flaschen & S	dm-BrandQ1
13	76	199	0.141176470588235	06Q2	Baby-Pull	BrandG2
13	53	-1	0.139705882352941	06Q2	Posten/Saiso	?
13	28	210	0.127941176470588	06Q2	Sonnenbrille	dm-BrandS1
13	33	1847	0.127941176470588	06Q2	Babynahrung	BrandL1
13	42	187	0.125	06Q2	Mund/Zahn	[redacted]
13	37	731	0.125	06Q2	Babyglasnahr	green-BrandU1

13	66	218	0.123529411764706	06Q2	Nassrasur	dm-BrandK1 [redacted]
13	30	59	0.122058823529412	06Q2	Babypflege	[redacted]
13	48	206	0.105882352941176	06Q2	Sonne	dm-BrandQ1
13	3	267	0.104411764705882	06Q2	Damenhygiene	[redacted]
13	30	2067	0.0985294117647059	06Q2	Babypflege	[redacted]
13	76	217	0.0970588235294118	06Q2	Baby-Pull	[redacted]
13	40	381	0.0955882352941176	06Q2	Bonbon	[redacted]
13	15	238	0.0882352941176471	06Q2	WPR	BrandF2
13	127	200	0.0852941176470588	06Q2	Ebelin Body	[redacted]
13	40	1772	0.0838235294117647	06Q2	Bonbon	BrandC2
13	51	207	0.0794117647058823	06Q2	Fu	dm-BrandK1
13	18	1884	0.0764705882352941	06Q2	Haush.artik.	[redacted]
13	39	201	0.075	06Q2	Hand	dm-green-BrandT1
13	42	32	0.0720588235294118	06Q2	Mund/Zahn	[redacted]
13	6	1265	0.0720588235294118	06Q2	Haushaltspap	BrandY1
13	34	-1	0.0691176470588235	06Q2	Fotozubehr	?
13	110	1988	0.0691176470588235	06Q2	Feinstrick 1	[redacted]
13	70	2359	0.0647058823529412	06Q2	Fotoarbeiten	[redacted] [redacted]
13	42	58	0.0617647058823529	06Q2	Mund/Zahn	[redacted]
13	54	218	0.0617647058823529	06Q2	Bad	dm-BrandK1 [redacted]
13	38	-1	0.0588235294117647	06Q2	Regenschirme	?
13	20	219	0.0544117647058824	06Q2	Nager	dm-BrandB1
13	40	1789	0.0544117647058824	06Q2	Bonbon	[redacted]
13	5	91	0.0544117647058824	06Q2	Haarpflege	[redacted]
13	103	490	0.0529411764705882	06Q2	Tchibo	[redacted]
13	12	1881	0.0529411764705882	06Q2	Insekt/Pflan	[redacted]
13	54	782	0.0514705882352941	06Q2	Bad	[redacted]
13	30	274	0.0514705882352941	06Q2	Babypflege	[redacted]
13	26	1843	0.0514705882352941	06Q2	Flaschen & S	[redacted]

***06Q3*, Cluster 24**

Cluster	S-A ID	Brand ID	*Centroid* Value	Quarter	Sub-Assortment	Brand
24	79	199	0.420647149460709	06Q3	Baby-Push	BrandG2
24	27	899	0.400616332819723	06Q3	Windeln	BrandE2
24	103	-1	0.382126348228043	06Q3	Tchibo	?
24	44	-1	0.329738058551618	06Q3	Presse	?
24	37	1846	0.248073959938367	06Q3	Babyglasnahr	BrandA1
24	33	1846	0.171032357473035	06Q3	Babynahrung	BrandA1
24	59	199	0.163328197226502	06Q3	KLKD-Push	BrandG2
24	26	206	0.152542372881356	06Q3	Flaschen & S	dm-BrandQ1
24	51	207	0.146379044684129	06Q3	Fu	dm-BrandK1

24	76	199	0.14175654853621	06Q3	Baby-Pull	BrandG2
24	53	-1	0.138674884437596	06Q3	Posten/Saiso	?
24	37	1847	0.130970724191063	06Q3	Babyglasnahr	BrandL1
24	33	1847	0.129429892141757	06Q3	Babynahrung	BrandL1
24	42	187	0.12788906009245	06Q3	Mund/Zahn	[redacted]
24	66	218	0.10939907550077	06Q3	Nassrasur	dm-BrandK1 [redacted]
24	30	59	0.10939907550077	06Q3	Babypflege	[redacted]
24	48	210	0.103235747303544	06Q3	Sonne	dm-BrandS1
24	40	381	0.0970724191063174	06Q3	Bonbon	[redacted]
24	54	218	0.0970724191063174	06Q3	Bad	dm-BrandK1 [redacted]
24	37	731	0.0955315870570108	06Q3	Babyglasnahr	green-BrandU1
24	18	1884	0.0924499229583975	06Q3	Haush.artik.	[redacted]
24	30	2067	0.0847457627118644	06Q3	Babypflege	[redacted]
24	76	217	0.0832049306625578	06Q3	Baby-Pull	[redacted]
24	42	32	0.0832049306625578	06Q3	Mund/Zahn	[redacted]
24	15	238	0.0816640986132511	06Q3	WPR	BrandF2
24	42	58	0.0801232665639445	06Q3	Mund/Zahn	[redacted]
24	127	200	0.0770416024653313	06Q3	Ebelin Body	[redacted]
24	34	-1	0.073959938366718	06Q3	Fotozubehr	?
24	34	205	0.0724191063174114	06Q3	Fotozubehr	[redacted]
24	5	33	0.0724191063174114	06Q3	Haarpflege	BrandZ1
24	39	201	0.0693374422187981	06Q3	Hand	dm-green-BrandT1
24	40	1772	0.0677966101694915	06Q3	Bonbon	BrandC2
24	70	2359	0.0677966101694915	06Q3	Fotoarbeiten	[redacted]
24	30	899	0.0647149460708783	06Q3	Babypflege	BrandE2
24	15	221	0.0647149460708783	06Q3	WPR	[redacted]
24	6	1265	0.0647149460708783	06Q3	Haushaltspap	BrandY1
24	11	65	0.0600924499229584	06Q3	Haarstyling	[redacted]
24	3	2191	0.0600924499229584	06Q3	Damenhygiene	BrandA2
24	3	267	0.0585516178736518	06Q3	Damenhygiene	[redacted]
24	56	1245	0.0570107858243451	06Q3	Filter&Folien	[redacted]
24	74	861	0.0570107858243451	06Q3	Filme	[redacted]
24	20	219	0.0554699537750385	06Q3	Nager	dm-BrandB1
24	75	306	0.0523882896764253	06Q3	Reform	[redacted]
24	39	61	0.0508474576271186	06Q3	Hand	[redacted]

06Q4, Cluster 2

Cluster	S-A ID	Brand ID	*Centroid* Value	Quarter	Sub-Assortment	Brand
2	103	-1	0.53862660944206	06Q4	Tchibo	?
2	79	199	0.412017167381974	06Q4	Baby-Push	BrandG2
2	44	-1	0.405579399141631	06Q4	Presse	?

2	106	-1	0.343347639484979	06Q4	Modul klein	?
2	27	899	0.298283261802575	06Q4	Windeln	BrandE2
2	116	-1	0.283261802575107	06Q4	Modul li	?
2	59	199	0.223175965665236	06Q4	KLKD-Push	BrandG2
2	37	1846	0.208154506437768	06Q4	Babyglasnahr	BrandA1
2	76	199	0.199570815450644	06Q4	Baby-Pull	BrandG2
2	26	206	0.195278969957082	06Q4	Flaschen & S	dm-BrandQ1
2	42	187	0.163090128755365	06Q4	Mund/Zahn	[redacted]
2	33	1846	0.163090128755365	06Q4	Babynahrung	BrandA1
2	107	-1	0.160944206008584	06Q4	Modul mittel	?
2	37	1847	0.148068669527897	06Q4	Babyglasnahr	BrandL1
2	70	2359	0.137339055793991	06Q4	Fotoarbeiten	[redacted]
2	40	381	0.130901287553648	06Q4	Bonbon	[redacted]
2	6	1265	0.130901287553648	06Q4	Haushaltspap	BrandY1
2	16	212	0.124463519313305	06Q4	Duftkerzen-M	[redacted]
2	5	33	0.122317596566524	06Q4	Haarpflege	BrandZ1
2	56	1245	0.122317596566524	06Q4	Filter&Folien	[redacted]
2	66	218	0.11587982832618	06Q4	Nassrasur	dm-BrandK1 [redacted]
2	30	2067	0.111587982832618	06Q4	Babypflege	[redacted]
2	33	1847	0.109442060085837	06Q4	Babynahrung	BrandL1
2	54	218	0.103004291845494	06Q4	Bad	dm-BrandK1 [redacted]
2	53	-1	0.100858369098712	06Q4	Posten/Saiso	?
2	40	1772	0.0987124463519313	06Q4	Bonbon	BrandC2
2	127	200	0.0987124463519313	06Q4	Ebelin Body	[redacted]
2	45	2075	0.0944206008583691	06Q4	Kerzen	BrandR1
2	34	205	0.092274678111588	06Q4	Fotozubehr	[redacted]
2	36	207	0.092274678111588	06Q4	Entferner	dm-BrandK1
2	30	59	0.0901287553648069	06Q4	Babypflege	[redacted]
2	76	217	0.0858369098712446	06Q4	Baby-Pull	[redacted]
2	3	267	0.0858369098712446	06Q4	Damenhygiene	[redacted]
2	37	731	0.0858369098712446	06Q4	Babyglasnahr	green-BrandU1
2	18	1884	0.0815450643776824	06Q4	Haush.artik.	[redacted]
2	110	1988	0.0793991416309013	06Q4	Feinstrick 1	[redacted]
2	40	314	0.0729613733905579	06Q4	Bonbon	[redacted]
2	131	1988	0.0686695278969957	06Q4	Grobstrick	[redacted]
2	42	32	0.0686695278969957	06Q4	Mund/Zahn	[redacted]
2	106	731	0.0665236051502146	06Q4	Modul klein	green-BrandU1
2	15	238	0.0665236051502146	06Q4	WPR	BrandF2
2	61	4	0.0622317596566524	06Q4	Bodylotion	[redacted]
2	42	58	0.0622317596566524	06Q4	Mund/Zahn	[redacted]
2	119	1290	0.0622317596566524	06Q4	Augen Make u	[redacted]
2	5	257	0.0622317596566524	06Q4	Haarpflege	[redacted]
2	15	221	0.0600858369098713	06Q4	WPR	[redacted]

2	20	219	0.0600858369098713	06Q4	Nager	dm-BrandB1
2	34	-1	0.0579399141630901	06Q4	Fotozubehr	?
2	45	2076	0.0579399141630901	06Q4	Kerzen	BrandH2
2	72	2000	0.0579399141630901	06Q4	Tee	[redacted]
2	58	1265	0.055793991416309	06Q4	Erlebnis dm	BrandY1
2	3	2191	0.055793991416309	06Q4	Damenhygiene	BrandA2
2	40	1789	0.0536480686695279	06Q4	Bonbon	[redacted]
2	15	270	0.0536480686695279	06Q4	WPR	[redacted]
2	33	1856	0.0536480686695279	06Q4	Babynahrung	[redacted]
2	39	201	0.0536480686695279	06Q4	Hand	dm-green-BrandT1
2	30	899	0.0515021459227468	06Q4	Babypflege	BrandE2
2	9	219	0.0515021459227468	06Q4	Hund	dm-BrandB1
2	30	274	0.0515021459227468	06Q4	Babypflege	[redacted]
2	30	81	0.0515021459227468	06Q4	Babypflege	[redacted]
2	116	212	0.0515021459227468	06Q4	Modul li	[redacted]
2	56	1244	0.0515021459227468	06Q4	Filter&Folien	[redacted]

F.3.4 The *GreenConcerns* Cluster

06Q2, Cluster 7

Cluster	S-A ID	Brand ID	*Centroid* Value	Quarter	Sub-Assortment	Brand
7	39	201	0.457627118644068	06Q2	Hand	dm-green-BrandT1
7	54	201	0.372881355932203	06Q2	Bad	dm-green-BrandT1
7	44	-1	0.288135593220339	06Q2	Presse	?
7	15	238	0.288135593220339	06Q2	WPR	BrandF2
7	5	201	0.254237288135593	06Q2	Haarpflege	dm-green-BrandT1
7	103	-1	0.254237288135593	06Q2	Tchibo	?
7	20	219	0.203389830508475	06Q2	Nager	dm-BrandB1
7	66	218	0.186440677966102	06Q2	Nassrasur	dm-BrandK1 [redacted]
7	40	1772	0.186440677966102	06Q2	Bonbon	BrandC2
7	61	201	0.169491525423729	06Q2	Bodylotion	dm-green-BrandT1
7	12	223	0.152542372881356	06Q2	Insekt/Pflan	[redacted]
7	56	1245	0.135593220338983	06Q2	Filter&Folien	[redacted]
7	18	1884	0.11864406779661	06Q2	Haush.artik.	[redacted]
7	61	274	0.11864406779661	06Q2	Bodylotion	[redacted]
7	60	176	0.11864406779661	06Q2	Gesicht	[redacted] [redacted]
7	6	1265	0.11864406779661	06Q2	Haushaltspap	BrandY1
7	48	210	0.11864406779661	06Q2	Sonne	dm-BrandS1
7	34	-1	0.101694915254237	06Q2	Fotozubehr	?
7	131	1988	0.101694915254237	06Q2	Grobstrick	[redacted]
7	75	306	0.101694915254237	06Q2	Reform	[redacted]

7	18	870	0.0847457627118644	06Q2	Haush.artik.	
7	51	69	0.0847457627118644	06Q2	Fu	
7	42	58	0.0847457627118644	06Q2	Mund/Zahn	
7	40	1775	0.0847457627118644	06Q2	Bonbon	
7	42	1132	0.0847457627118644	06Q2	Mund/Zahn	
7	58	731	0.0847457627118644	06Q2	Erlebnis dm	green-BrandU1
7	110	1988	0.0847457627118644	06Q2	Feinstrick 1	
7	45	2075	0.0677966101694915	06Q2	Kerzen	BrandR1
7	40	381	0.0677966101694915	06Q2	Bonbon	
7	36	213	0.0677966101694915	06Q2	Entferner	dm-BrandO1
7	51	207	0.0677966101694915	06Q2	Fu	dm-BrandK1
7	42	187	0.0677966101694915	06Q2	Mund/Zahn	
7	58	32	0.0677966101694915	06Q2	Erlebnis dm	
7	15	212	0.0677966101694915	06Q2	WPR	
7	2	2213	0.0677966101694915	06Q2	Colorationen	
7	51	2055	0.0677966101694915	06Q2	Fu	
7	42	32	0.0677966101694915	06Q2	Mund/Zahn	
7	45	2076	0.0677966101694915	06Q2	Kerzen	BrandH2
7	9	219	0.0677966101694915	06Q2	Hund	dm-BrandB1
7	4	1241	0.0508474576271186	06Q2	Kosmetik-Pap	BrandB2
7	59	199	0.0508474576271186	06Q2	KLKD-Push	BrandG2
7	78	361	0.0508474576271186	06Q2	Pharma	
7	78	897	0.0508474576271186	06Q2	Pharma	
7	3	267	0.0508474576271186	06Q2	Damenhygiene	
7	18	1885	0.0508474576271186	06Q2	Haush.artik.	
7	5	952	0.0508474576271186	06Q2	Haarpflege	
7	6	1263	0.0508474576271186	06Q2	Haushaltspap	
7	36	207	0.0508474576271186	06Q2	Entferner	dm-BrandK1
7	11	65	0.0508474576271186	06Q2	Haarstyling	
7	57	21	0.0508474576271186	06Q2	Deo	
7	40	1781	0.0508474576271186	06Q2	Bonbon	
7	42	1140	0.0508474576271186	06Q2	Mund/Zahn	
7	57	218	0.0508474576271186	06Q2	Deo	dm-BrandK1
7	5	967	0.0508474576271186	06Q2	Haarpflege	
7	19	1275	0.0508474576271186	06Q2	Hygienepapie	
7	42	88	0.0508474576271186	06Q2	Mund/Zahn	
7	10	-1	0.0508474576271186	06Q2	Smereien-M	?
7	2	1377	0.0508474576271186	06Q2	Colorationen	
7	57	14	0.0508474576271186	06Q2	Deo	
7	127	200	0.0508474576271186	06Q2	Ebelin Body	
7	3	2191	0.0508474576271186	06Q2	Damenhygiene	BrandA2
7	5	257	0.0508474576271186	06Q2	Haarpflege	
7	56	1243	0.0508474576271186	06Q2	Filter&Folien	BrandX1

7	53	-1	0.0508474576271186	06Q2	Posten/Saiso	?
7	15	221	0.0508474576271186	06Q2	WPR	[redacted]
7	5	91	0.0508474576271186	06Q2	Haarpflege	[redacted]
7	11	952	0.0508474576271186	06Q2	Haarstyling	[redacted]
7	15	2098	0.0508474576271186	06Q2	WPR	[redacted]

***06Q3*, Cluster 0**

Cluster	S-A ID	Brand ID	*Centroid* Value	Quarter	Sub-Assortment	Brand
0	48	210	0.307692307692308	06Q3	Sonne	dm-BrandS1
0	54	201	0.294871794871795	06Q3	Bad	dm-green-BrandT1
0	5	201	0.282051282051282	06Q3	Haarpflege	dm-green-BrandT1
0	39	201	0.269230769230769	06Q3	Hand	dm-green-BrandT1
0	53	-1	0.243589743589744	06Q3	Posten/Saiso	?
0	15	238	0.217948717948718	06Q3	WPR	BrandF2
0	40	1789	0.192307692307692	06Q3	Bonbon	[redacted]
0	3	267	0.141025641025641	06Q3	Damenhygiene	[redacted]
0	103	-1	0.141025641025641	06Q3	Tchibo	?
0	51	207	0.128205128205128	06Q3	Fu	dm-BrandK1
0	34	205	0.128205128205128	06Q3	Fotozubehr	[redacted]
0	127	200	0.128205128205128	06Q3	Ebelin Body	[redacted]
0	66	218	0.128205128205128	06Q3	Nassrasur	dm-BrandK1 [redacted]
0	60	284	0.128205128205128	06Q3	Gesicht	[redacted]
0	61	201	0.102564102564103	06Q3	Bodylotion	dm-green-BrandT1
0	78	897	0.102564102564103	06Q3	Pharma	[redacted]
0	20	219	0.102564102564103	06Q3	Nager	dm-BrandB1
0	56	1245	0.102564102564103	06Q3	Filter&Folien	[redacted]
0	36	207	0.102564102564103	06Q3	Entferner	dm-BrandK1
0	27	899	0.102564102564103	06Q3	Windeln	BrandE2
0	103	490	0.0897435897435897	06Q3	Tchibo	[redacted]
0	60	176	0.0897435897435897	06Q3	Gesicht	[redacted]
0	15	212	0.0897435897435897	06Q3	WPR	[redacted]
0	3	2191	0.0897435897435897	06Q3	Damenhygiene	BrandA2
0	5	952	0.0897435897435897	06Q3	Haarpflege	[redacted]
0	18	1884	0.0897435897435897	06Q3	Haush.artik.	[redacted]
0	42	187	0.0897435897435897	06Q3	Mund/Zahn	[redacted]
0	54	218	0.0897435897435897	06Q3	Bad	dm-BrandK1 [redacted]
0	42	1140	0.0769230769230769	06Q3	Mund/Zahn	[redacted]
0	61	274	0.0769230769230769	06Q3	Bodylotion	[redacted]
0	12	1881	0.0769230769230769	06Q3	Insekt/Pflan	[redacted]
0	5	4	0.0769230769230769	06Q3	Haarpflege	[redacted]
0	5	33	0.0769230769230769	06Q3	Haarpflege	BrandZ1

0	110	1988	0.0769230769230769	06Q3	Feinstrick 1	
0	42	1189	0.0769230769230769	06Q3	Mund/Zahn	
0	30	2067	0.0769230769230769	06Q3	Babypflege	
0	6	1263	0.0769230769230769	06Q3	Haushaltspap	
0	12	212	0.0641025641025641	06Q3	Insekt/Pflan	
0	42	891	0.0641025641025641	06Q3	Mund/Zahn	
0	30	59	0.0641025641025641	06Q3	Babypflege	
0	11	65	0.0641025641025641	06Q3	Haarstyling	
0	103	497	0.0641025641025641	06Q3	Tchibo	
0	119	1290	0.0641025641025641	06Q3	Augen Make u	
0	40	1797	0.0512820512820513	06Q3	Bonbon	
0	54	381	0.0512820512820513	06Q3	Bad	
0	40	1793	0.0512820512820513	06Q3	Bonbon	
0	75	1937	0.0512820512820513	06Q3	Reform	
0	66	71	0.0512820512820513	06Q3	Nassrasur	
0	119	158	0.0512820512820513	06Q3	Augen Make u	
0	44	-1	0.0512820512820513	06Q3	Presse	?
0	5	91	0.0512820512820513	06Q3	Haarpflege	
0	30	274	0.0512820512820513	06Q3	Babypflege	
0	39	10	0.0512820512820513	06Q3	Hand	
0	119	79	0.0512820512820513	06Q3	Augen Make u	
0	60	274	0.0512820512820513	06Q3	Gesicht	
0	36	213	0.0512820512820513	06Q3	Entferner	dm-BrandO1
0	131	1988	0.0512820512820513	06Q3	Grobstrick	
0	75	306	0.0512820512820513	06Q3	Reform	
0	56	1243	0.0512820512820513	06Q3	Filter&Folien	BrandX1
0	51	2055	0.0512820512820513	06Q3	Fu	
0	15	272	0.0512820512820513	06Q3	WPR	
0	15	262	0.0512820512820513	06Q3	WPR	
0	42	32	0.0512820512820513	06Q3	Mund/Zahn	

Appendix G

Details to Stable Clusters, Store 22

G.1 *Centroid* values of *KhK* graph *06Q1*

See Section 4.5.1.

Cluster	S-A ID	Brand ID	*Centroid* Value	Quarter	Sub-Assortment	Brand
2	19	214	0.838862559241706	06Q1	Hygienepapie	dm-BrandN1
1	81	731	0.817375886524823	06Q1	green-BrandU1	green-BrandU1
5	19	214	0.815972222222222	06Q1	Hygienepapie	dm-BrandN1
2	6	214	0.797788309636651	06Q1	Haushaltspap	dm-BrandN1
5	6	214	0.788194444444444	06Q1	Haushaltspap	dm-BrandN1
5	81	731	0.784722222222222	06Q1	green-BrandU1	green-BrandU1
1	19	214	0.767730496453901	06Q1	Hygienepapie	dm-BrandN1
2	4	213	0.759873617693523	06Q1	Kosmetik-Pap	dm-BrandO1
1	6	214	0.737588652482269	06Q1	Haushaltspap	dm-BrandN1
5	56	212	0.673611111111111	06Q1	Filter&Folien	[redacted]
2	56	212	0.671406003159558	06Q1	Filter&Folien	[redacted]
1	56	212	0.652482269503546	06Q1	Filter&Folien	[redacted]
1	4	213	0.647163120567376	06Q1	Kosmetik-Pap	dm-BrandO1
5	4	213	0.625	06Q1	Kosmetik-Pap	dm-BrandO1
2	39	207	0.624012638230648	06Q1	Hand	dm-BrandK1
5	18	212	0.618055555555556	06Q1	Haush.artik.	[redacted]
2	81	731	0.598736176935229	06Q1	green-BrandU1	green-BrandU1
1	18	212	0.593971631205674	06Q1	Haush.artik.	[redacted]
2	18	212	0.55608214849921	06Q1	Haush.artik.	[redacted]
1	30	206	0.514184397163121	06Q1	Babypflege	dm-BrandQ1
5	45	212	0.510416666666667	06Q1	Kerzen	[redacted]
5	84	-1	0.506944444444444	06Q1	Sonstiges/Te	?
2	15	202	0.483412322274881	06Q1	WPR	dm-BrandD
1	70	-1	0.480496453900709	06Q1	Fotoarbeiten	?

2	3	208	0.456556082148499	06Q1	Damenhygiene	[redacted]
2	54	207	0.445497630331754	06Q1	Bad	dm-BrandK1
1	27	206	0.445035460992908	06Q1	Windeln	dm-BrandQ1
1	84	-1	0.434397163120567	06Q1	Sonstiges/Te	?
1	15	202	0.423758865248227	06Q1	WPR	dm-BrandD
2	60	207	0.420221169036335	06Q1	Gesicht	dm-BrandK1
2	42	211	0.39652448657188	06Q1	Mund/Zahn	dm-BrandP1

G.2 *Bestmatch*- and *Best CD*-graphs of Store 22, with Eliminations

See Section 4.5.2.

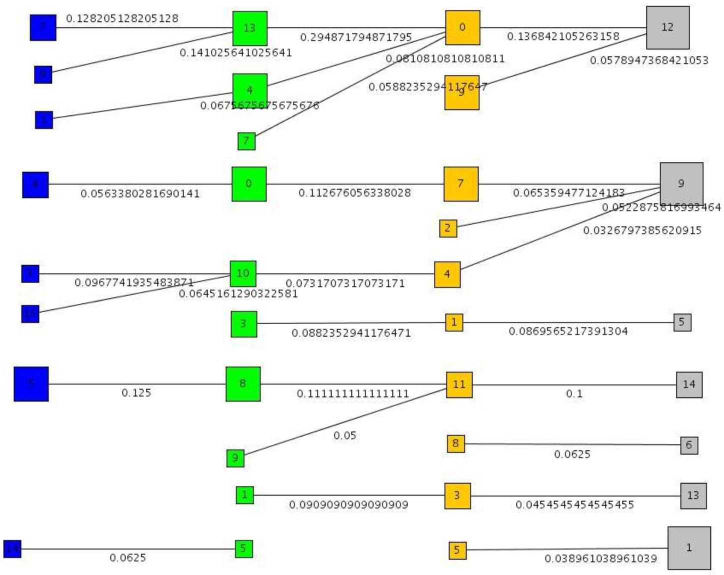


Figure G.1: *BestMatch*-graph with eliminations and customer intersection, store 22.

G.3 *Centroids* of Stable Clusters

All *centroid* values greater than 5%.

G.3.1 The *Traditional/Premium* Cluster

06Q1, Cluster 12

Cluster	S-A ID	Brand ID	*Centroid* Value	Quarter	Sub-Assortment	Brand
12	56	1243	0.169161676646707	06Q1	Filter&Folien	BrandX1
12	61	4	0.152694610778443	06Q1	Bodylotion	[redacted]
12	42	187	0.151197604790419	06Q1	Mund/Zahn	[redacted]
12	4	1241	0.148203592814371	06Q1	Kosmetik-Pap	BrandB2
12	6	1265	0.146706586826347	06Q1	Haushaltspap	BrandY1
12	36	207	0.145209580838323	06Q1	Entferner	dm-BrandK1
12	127	200	0.142215568862275	06Q1	Ebelin Body	[redacted]
12	5	33	0.139221556886228	06Q1	Haarpflege	BrandZ1
12	3	267	0.131736526946108	06Q1	Damenhygiene	[redacted]
12	11	65	0.125748502994012	06Q1	Haarstyling	[redacted]
12	42	32	0.124251497005988	06Q1	Mund/Zahn	[redacted]
12	60	176	0.116766467065868	06Q1	Gesicht	[redacted] [redacted]
12	30	59	0.110778443113772	06Q1	Babypflege	[redacted]
12	72	1213	0.109281437125748	06Q1	Tee	[redacted]
12	39	257	0.104790419161677	06Q1	Hand	[redacted]
12	75	306	0.103293413173653	06Q1	Reform	[redacted]
12	40	1772	0.103293413173653	06Q1	Bonbon	BrandC2
12	18	1884	0.103293413173653	06Q1	Haush.artik.	[redacted]
12	58	1265	0.0973053892215569	06Q1	Erlebnis dm	BrandY1
12	15	262	0.094311377245509	06Q1	WPR	[redacted]
12	45	2076	0.092814371257485	06Q1	Kerzen	BrandH2
12	40	1793	0.092814371257485	06Q1	Bonbon	[redacted]
12	51	207	0.0913173652694611	06Q1	Fu	dm-BrandK1
12	110	1988	0.0883233532934132	06Q1	Feinstrick 1	[redacted]
12	3	2191	0.0853293413173653	06Q1	Damenhygiene	BrandA2
12	19	1264	0.0853293413173653	06Q1	Hygienepapie	[redacted]
12	66	218	0.0823353293413174	06Q1	Nassrasur	dm-BrandK1 [redacted]
12	6	1263	0.0823353293413174	06Q1	Haushaltspap	[redacted]
12	112	212	0.0823353293413174	06Q1	Ostern klein	[redacted]
12	48	210	0.0808383233532934	06Q1	Sonne	dm-BrandS1
12	60	174	0.0808383233532934	06Q1	Gesicht	[redacted]
12	58	32	0.0808383233532934	06Q1	Erlebnis dm	[redacted]
12	40	381	0.0778443113772455	06Q1	Bonbon	[redacted]
12	119	1290	0.0763473053892216	06Q1	Augen Make u	[redacted]
12	5	952	0.0748502994011976	06Q1	Haarpflege	[redacted]
12	42	1140	0.0733532934131736	06Q1	Mund/Zahn	[redacted]
12	119	79	0.0703592814371257	06Q1	Augen Make u	[redacted]
12	19	1275	0.0688622754491018	06Q1	Hygienepapie	[redacted]
12	15	439	0.0688622754491018	06Q1	WPR	[redacted]
12	16	433	0.0673652694610778	06Q1	Duftkerzen-M	[redacted]
12	5	66	0.0673652694610778	06Q1	Haarpflege	[redacted]
12	54	782	0.0673652694610778	06Q1	Bad	[redacted]

12	75	381	0.0658682634730539	06Q1	Reform	[redacted]
12	60	1365	0.0643712574850299	06Q1	Gesicht	[redacted]
12	18	1349	0.0643712574850299	06Q1	Haush.artik.	[redacted]
12	42	58	0.061377245508982	06Q1	Mund/Zahn	[redacted]
12	15	895	0.0583832335329341	06Q1	WPR	[redacted]
12	57	173	0.0583832335329341	06Q1	Deo	[redacted]
12	15	270	0.0583832335329341	06Q1	WPR	[redacted]
12	66	71	0.0553892215568862	06Q1	Nassrasur	[redacted]
12	36	213	0.0553892215568862	06Q1	Entferner	dm-BrandO1
12	45	2075	0.0553892215568862	06Q1	Kerzen	BrandR1
12	5	257	0.0538922155688623	06Q1	Haarpflege	[redacted]
12	60	162	0.0538922155688623	06Q1	Gesicht	[redacted]
12	78	897	0.0538922155688623	06Q1	Pharma	[redacted]
12	61	257	0.0538922155688623	06Q1	Bodylotion	[redacted]
12	54	86	0.0538922155688623	06Q1	Bad	[redacted]
12	20	219	0.0523952095808383	06Q1	Nager	dm-BrandB1
12	54	257	0.0508982035928144	06Q1	Bad	[redacted]
12	40	314	0.0508982035928144	06Q1	Bonbon	[redacted]
12	54	218	0.0508982035928144	06Q1	Bad	dm-BrandK1 [redacted]
12	119	29	0.0508982035928144	06Q1	Augen Make u	[redacted]
12	5	25	0.0508982035928144	06Q1	Haarpflege	[redacted]
12	15	249	0.0508982035928144	06Q1	WPR	[redacted]

***06Q2*, Cluster 8**

Cluster	S-A ID	Brand ID	*Centroid* Value	Quarter	Sub-Assortment	Brand
8	61	4	0.246458923512748	06Q2	Bodylotion	[redacted]
8	48	210	0.209631728045326	06Q2	Sonne	dm-BrandS1
8	6	1265	0.192634560906516	06Q2	Haushaltspap	BrandY1
8	56	1243	0.172804532577904	06Q2	Filter&Folien	BrandX1
8	5	33	0.168555240793201	06Q2	Haarpflege	BrandZ1
8	60	176	0.160056657223796	06Q2	Gesicht	[redacted] [redacted]
8	42	187	0.154390934844193	06Q2	Mund/Zahn	[redacted]
8	18	1884	0.148725212464589	06Q2	Haush.artik.	[redacted]
8	3	2191	0.148725212464589	06Q2	Damenhygiene	BrandA2
8	75	306	0.147308781869688	06Q2	Reform	[redacted]
8	40	1793	0.128895184135977	06Q2	Bonbon	[redacted]
8	4	1241	0.121813031161473	06Q2	Kosmetik-Pap	BrandB2
8	58	1265	0.117563739376771	06Q2	Erlebnis dm	BrandY1
8	74	861	0.110481586402266	06Q2	Filme	[redacted]
8	15	262	0.0963172804532578	06Q2	WPR	[redacted]
8	53	-1	0.0934844192634561	06Q2	Posten/Saiso	?

8	12	223	0.0892351274787535	06Q2	Insekt/Pflan	[redacted]
8	42	32	0.0892351274787535	06Q2	Mund/Zahn	[redacted]
8	127	200	0.0878186968838527	06Q2	Ebelin Body	[redacted]
8	11	65	0.0878186968838527	06Q2	Haarstyling	[redacted]
8	42	1140	0.0878186968838527	06Q2	Mund/Zahn	[redacted]
8	51	207	0.0864022662889518	06Q2	Fu	dm-BrandK1
8	78	897	0.084985835694051	06Q2	Pharma	[redacted]
8	15	895	0.0835694050991501	06Q2	WPR	[redacted]
8	36	213	0.0821529745042493	06Q2	Entferner	dm-BrandO1
8	56	1245	0.0764872521246459	06Q2	Filter&Folien	[redacted]
8	60	1365	0.0764872521246459	06Q2	Gesicht	[redacted]
8	39	201	0.075070821529745	06Q2	Hand	dm-green-BrandT1
8	5	91	0.0736543909348442	06Q2	Haarpflege	[redacted]
8	39	257	0.0722379603399433	06Q2	Hand	[redacted]
8	3	267	0.0708215297450425	06Q2	Damenhygiene	[redacted]
8	19	1264	0.0679886685552408	06Q2	Hygienepapie	[redacted]
8	54	811	0.0679886685552408	06Q2	Bad	[redacted]
8	61	257	0.0665722379603399	06Q2	Bodylotion	[redacted]
8	36	207	0.0651558073654391	06Q2	Entferner	dm-BrandK1
8	39	61	0.0651558073654391	06Q2	Hand	[redacted]
8	15	249	0.0651558073654391	06Q2	WPR	[redacted]
8	60	174	0.0637393767705382	06Q2	Gesicht	[redacted]
8	42	1189	0.0623229461756374	06Q2	Mund/Zahn	[redacted]
8	51	69	0.0623229461756374	06Q2	Fu	[redacted]
8	30	59	0.0623229461756374	06Q2	Babypflege	[redacted]
8	40	381	0.0623229461756374	06Q2	Bonbon	[redacted]
8	12	212	0.0609065155807365	06Q2	Insekt/Pflan	[redacted]
8	19	1275	0.0609065155807365	06Q2	Hygienepapie	[redacted]
8	39	172	0.0594900849858357	06Q2	Hand	[redacted] [redacted]
8	119	29	0.0580736543909348	06Q2	Augen Make u	[redacted]
8	58	32	0.056657223796034	06Q2	Erlebnis dm	[redacted]
8	45	2076	0.056657223796034	06Q2	Kerzen	BrandH2
8	72	1213	0.056657223796034	06Q2	Tee	[redacted]
8	40	314	0.0552407932011331	06Q2	Bonbon	[redacted]
8	5	113	0.0538243626062323	06Q2	Haarpflege	[redacted]
8	11	908	0.0538243626062323	06Q2	Haarstyling	[redacted]
8	110	1988	0.0538243626062323	06Q2	Feinstrick 1	[redacted]
8	48	4	0.0538243626062323	06Q2	Sonne	[redacted]
8	40	1772	0.0524079320113314	06Q2	Bonbon	BrandC2
8	5	25	0.0509915014164306	06Q2	Haarpflege	[redacted]
8	48	74	0.0509915014164306	06Q2	Sonne	[redacted]

(*06Q3*, Cluster 2)

Cluster	S-A ID	Brand ID	*Centroid* Value	Quarter	Sub-Assortment	Brand
2	61	4	0.269814502529511	06Q3	Bodylotion	[redacted]
2	42	187	0.197301854974705	06Q3	Mund/Zahn	[redacted]
2	6	1265	0.18043844856661	06Q3	Haushaltspap	BrandY1
2	39	201	0.160202360876897	06Q3	Hand	dm-green-BrandT1
2	127	200	0.158516020236088	06Q3	Ebelin Body	[redacted]
2	56	1243	0.158516020236088	06Q3	Filter&Folien	BrandX1
2	18	1884	0.146711635750422	06Q3	Haush.artik.	[redacted]
2	58	1265	0.143338954468803	06Q3	Erlebnis dm	BrandY1
2	75	306	0.134907251264755	06Q3	Reform	[redacted]
2	74	861	0.133220910623946	06Q3	Filme	[redacted]
2	48	210	0.133220910623946	06Q3	Sonne	dm-BrandS1
2	60	176	0.131534569983137	06Q3	Gesicht	[redacted] [redacted]
2	5	33	0.131534569983137	06Q3	Haarpflege	BrandZ1
2	3	2191	0.114671163575042	06Q3	Damenhygiene	BrandA2
2	42	32	0.104553119730186	06Q3	Mund/Zahn	[redacted]
2	11	65	0.102866779089376	06Q3	Haarstyling	[redacted]
2	40	1793	0.0994940978077572	06Q3	Bonbon	[redacted]
2	40	381	0.0961214165261383	06Q3	Bonbon	[redacted]
2	51	207	0.0927487352445194	06Q3	Fu	dm-BrandK1
2	53	-1	0.0876897133220911	06Q3	Posten/Saiso	?
2	42	1189	0.0876897133220911	06Q3	Mund/Zahn	[redacted]
2	3	267	0.0860033726812816	06Q3	Damenhygiene	[redacted]
2	4	1241	0.0860033726812816	06Q3	Kosmetik-Pap	BrandB2
2	20	219	0.0826306913996627	06Q3	Nager	dm-BrandB1
2	15	895	0.0826306913996627	06Q3	WPR	[redacted]
2	61	201	0.0792580101180438	06Q3	Bodylotion	dm-green-BrandT1
2	54	82	0.0792580101180438	06Q3	Bad	[redacted]
2	45	2076	0.0741989881956155	06Q3	Kerzen	BrandH2
2	110	1988	0.0708263069139966	06Q3	Feinstrick 1	[redacted]
2	66	218	0.0691399662731872	06Q3	Nassrasur	dm-BrandK1 [redacted]
2	78	897	0.0691399662731872	06Q3	Pharma	[redacted]
2	36	213	0.0691399662731872	06Q3	Entferner	dm-BrandO1
2	30	59	0.0674536256323778	06Q3	Babypflege	[redacted]
2	39	172	0.0657672849915683	06Q3	Hand	[redacted] [redacted]
2	56	1245	0.0657672849915683	06Q3	Filter&Folien	[redacted]
2	54	201	0.0640809443507589	06Q3	Bad	dm-green-BrandT1
2	12	212	0.0640809443507589	06Q3	Insekt/Pflan	[redacted]
2	60	1365	0.0640809443507589	06Q3	Gesicht	[redacted]
2	5	25	0.0623946037099494	06Q3	Haarpflege	[redacted]
2	5	4	0.06070826306914	06Q3	Haarpflege	[redacted]

2	15	262	0.06070826306914	06Q3	WPR	[redacted]
2	42	891	0.0590219224283305	06Q3	Mund/Zahn	[redacted]
2	19	1264	0.0590219224283305	06Q3	Hygienepapie	[redacted]
2	40	1772	0.0573355817875211	06Q3	Bonbon	BrandC2
2	12	223	0.0573355817875211	06Q3	Insekt/Pflan	[redacted]
2	15	221	0.0556492411467116	06Q3	WPR	[redacted]
2	57	173	0.0556492411467116	06Q3	Deo	[redacted]
2	16	433	0.0556492411467116	06Q3	Duftkerzen-M	[redacted]
2	78	358	0.0556492411467116	06Q3	Pharma	[redacted]
2	5	201	0.0556492411467116	06Q3	Haarpflege	dm-green-BrandT1
2	119	79	0.0539629005059022	06Q3	Augen Make u	[redacted]
2	60	174	0.0539629005059022	06Q3	Gesicht	[redacted]
2	48	4	0.0539629005059022	06Q3	Sonne	[redacted]
2	15	238	0.0539629005059022	06Q3	WPR	BrandF2
2	72	1213	0.0539629005059022	06Q3	Tee	[redacted]
2	11	756	0.0522765598650928	06Q3	Haarstyling	[redacted]
2	5	66	0.0522765598650928	06Q3	Haarpflege	[redacted]
2	19	1275	0.0505902192242833	06Q3	Hygienepapie	[redacted]
2	51	69	0.0505902192242833	06Q3	Fu	[redacted]
2	45	2075	0.0505902192242833	06Q3	Kerzen	BrandR1
2	57	5	0.0505902192242833	06Q3	Deo	[redacted]
2	39	885	0.0505902192242833	06Q3	Hand	[redacted]

***06Q4*, Cluster 9**

Cluster	S-A ID	Brand ID	*Centroid* Value	Quarter	Sub-Assortment	Brand
9	116	-1	0.199692780337942	06Q4	Modul li	?
9	106	-1	0.19047619047619	06Q4	Modul klein	?
9	45	2076	0.179723502304147	06Q4	Kerzen	BrandH2
9	56	1243	0.139784946236559	06Q4	Filter&Folien	BrandX1
9	60	176	0.125960061443932	06Q4	Gesicht	[redacted] [redacted]
9	61	4	0.125960061443932	06Q4	Bodylotion	[redacted]
9	16	212	0.122887864823349	06Q4	Duftkerzen-M	[redacted]
9	36	207	0.119815668202765	06Q4	Entferner	dm-BrandK1
9	40	1793	0.119815668202765	06Q4	Bonbon	[redacted]
9	58	1265	0.113671274961598	06Q4	Erlebnis dm	BrandY1
9	51	207	0.113671274961598	06Q4	Fu	dm-BrandK1
9	3	267	0.113671274961598	06Q4	Damenhygiene	[redacted]
9	11	65	0.112135176651306	06Q4	Haarstyling	[redacted]
9	40	1772	0.109062980030722	06Q4	Bonbon	BrandC2
9	18	1884	0.104454685099846	06Q4	Haush.artik.	[redacted]
9	127	200	0.0998463901689708	06Q4	Ebelin Body	[redacted]

9	42	32	0.0998463901689708	06Q4	Mund/Zahn	[redacted]
9	45	2075	0.098310291858679	06Q4	Kerzen	BrandR1
9	107	-1	0.098310291858679	06Q4	Modul mittel	?
9	66	218	0.0967741935483871	06Q4	Nassrasur	dm-BrandK1 [redacted]
9	60	1365	0.0937019969278034	06Q4	Gesicht	[redacted]
9	110	1988	0.0921658986175115	06Q4	Feinstrick 1	[redacted]
9	119	1290	0.0906298003072197	06Q4	Augen Make u	[redacted]
9	72	1213	0.0814132104454685	06Q4	Tee	[redacted]
9	5	33	0.0783410138248848	06Q4	Haarpflege	BrandZ1
9	57	5	0.0752688172043011	06Q4	Deo	[redacted]
9	48	210	0.0737327188940092	06Q4	Sonne	dm-BrandS1
9	6	1265	0.0737327188940092	06Q4	Haushaltspap	BrandY1
9	5	66	0.0737327188940092	06Q4	Haarpflege	[redacted]
9	36	213	0.0737327188940092	06Q4	Entferner	dm-BrandO1
9	42	58	0.0721966205837174	06Q4	Mund/Zahn	[redacted]
9	42	1140	0.0706605222734255	06Q4	Mund/Zahn	[redacted]
9	40	381	0.0691244239631336	06Q4	Bonbon	[redacted]
9	42	187	0.0691244239631336	06Q4	Mund/Zahn	[redacted]
9	56	1245	0.0675883256528418	06Q4	Filter&Folien	[redacted]
9	4	1241	0.0660522273425499	06Q4	Kosmetik-Pap	BrandB2
9	54	218	0.0645161290322581	06Q4	Bad	dm-BrandK1 [redacted]
9	60	174	0.0645161290322581	06Q4	Gesicht	[redacted]
9	20	219	0.0645161290322581	06Q4	Nager	dm-BrandB1
9	53	-1	0.0629800307219662	06Q4	Posten/Saiso	?
9	39	201	0.0629800307219662	06Q4	Hand	dm-green-BrandT1
9	54	782	0.0629800307219662	06Q4	Bad	[redacted]
9	3	2191	0.0614439324116744	06Q4	Damenhygiene	BrandA2
9	58	32	0.0599078341013825	06Q4	Erlebnis dm	[redacted]
9	5	113	0.0583717357910906	06Q4	Haarpflege	[redacted]
9	75	306	0.0583717357910906	06Q4	Reform	[redacted]
9	5	952	0.0583717357910906	06Q4	Haarpflege	[redacted]
9	6	1263	0.0583717357910906	06Q4	Haushaltspap	[redacted]
9	119	79	0.0568356374807988	06Q4	Augen Make u	[redacted]
9	40	1775	0.0568356374807988	06Q4	Bonbon	[redacted]
9	116	212	0.0552995391705069	06Q4	Modul li	[redacted]
9	15	895	0.0537634408602151	06Q4	WPR	[redacted]
9	119	49	0.0537634408602151	06Q4	Augen Make u	[redacted]
9	18	1349	0.0537634408602151	06Q4	Haush.artik.	[redacted]
9	19	1275	0.0522273425499232	06Q4	Hygienepapie	[redacted]
9	39	61	0.0506912442396313	06Q4	Hand	[redacted]
9	119	158	0.0506912442396313	06Q4	Augen Make u	[redacted]

G.3.2 The *Mainstream* Cluster

06Q2, Cluster 6

Cluster	S-A ID	Brand ID	*Centroid* Value	Quarter	Sub-Assortment	Brand
6	48	210	0.411829134720701	06Q2	Sonne	dm-BrandS1
6	51	207	0.245345016429354	06Q2	Fu	dm-BrandK1
6	66	218	0.200438116100767	06Q2	Nassrasur	dm-BrandK1 [redacted]
6	53	-1	0.178532311062432	06Q2	Posten/Saiso	?
6	110	1988	0.164293537787514	06Q2	Feinstrick 1	[redacted]
6	11	65	0.161007667031763	06Q2	Haarstyling	[redacted]
6	40	1772	0.138006571741512	06Q2	Bonbon	BrandC2
6	39	201	0.136911281489595	06Q2	Hand	dm-green-BrandT1
6	3	267	0.132530120481928	06Q2	Damenhygiene	[redacted]
6	54	218	0.130339539978094	06Q2	Bad	dm-BrandK1 [redacted]
6	36	213	0.130339539978094	06Q2	Entferner	dm-BrandO1
6	42	32	0.127053669222344	06Q2	Mund/Zahn	[redacted]
6	36	207	0.112814895947426	06Q2	Entferner	dm-BrandK1
6	119	1290	0.110624315443593	06Q2	Augen Make u	[redacted]
6	127	200	0.104052573932092	06Q2	Ebelin Body	[redacted]
6	40	381	0.0974808324205915	06Q2	Bonbon	[redacted]
6	20	219	0.0952902519167579	06Q2	Nager	dm-BrandB1
6	5	952	0.0920043811610077	06Q2	Haarpflege	[redacted]
6	112	-1	0.0832420591456736	06Q2	Ostern klein	?
6	57	5	0.0832420591456736	06Q2	Deo	[redacted]
6	42	58	0.0832420591456736	06Q2	Mund/Zahn	[redacted]
6	5	66	0.0799561883899233	06Q2	Haarpflege	[redacted]
6	121	1309	0.0799561883899233	06Q2	Nagelpflege	[redacted]
6	40	314	0.0777656078860898	06Q2	Bonbon	[redacted]
6	56	1245	0.0755750273822563	06Q2	Filter&Folien	[redacted]
6	42	187	0.0744797371303395	06Q2	Mund/Zahn	[redacted]
6	3	2191	0.0711938663745893	06Q2	Damenhygiene	BrandA2
6	54	201	0.0700985761226725	06Q2	Bad	dm-green-BrandT1
6	39	257	0.0700985761226725	06Q2	Hand	[redacted]
6	34	-1	0.0690032858707558	06Q2	Fotozubehr	?
6	61	82	0.0690032858707558	06Q2	Bodylotion	[redacted]
6	5	257	0.0690032858707558	06Q2	Haarpflege	[redacted]
6	16	433	0.067907995618839	06Q2	Duftkerzen-M	[redacted]
6	75	306	0.0668127053669222	06Q2	Reform	[redacted]
6	18	1884	0.0668127053669222	06Q2	Haush.artik.	[redacted]
6	119	29	0.0657174151150055	06Q2	Augen Make u	[redacted]
6	5	33	0.0646221248630887	06Q2	Haarpflege	BrandZ1
6	34	2189	0.0646221248630887	06Q2	Fotozubehr	[redacted]

6	61	201	0.0646221248630887	06Q2	Bodylotion	dm-green-BrandT1
6	120	1290	0.063526834611172	06Q2	Lippenpflege	[redacted]
6	15	439	0.0624315443592552	06Q2	WPR	[redacted]
6	10	-1	0.0624315443592552	06Q2	Smereien-M	?
6	38	-1	0.0613362541073384	06Q2	Regenschirme	?
6	45	2076	0.0602409638554217	06Q2	Kerzen	BrandH2
6	30	59	0.0591456736035049	06Q2	Babypflege	[redacted]
6	30	81	0.0591456736035049	06Q2	Babypflege	[redacted]
6	61	4	0.0591456736035049	06Q2	Bodylotion	[redacted]
6	57	218	0.0580503833515882	06Q2	Deo	dm-BrandK1 [redacted]
6	60	162	0.0569550930996714	06Q2	Gesicht	[redacted]
6	12	212	0.0558598028477547	06Q2	Insekt/Pflan	[redacted]
6	72	1213	0.0558598028477547	06Q2	Tee	[redacted]
6	75	381	0.0547645125958379	06Q2	Reform	[redacted]
6	12	1881	0.0547645125958379	06Q2	Insekt/Pflan	[redacted]
6	6	1263	0.0547645125958379	06Q2	Haushaltspap	[redacted]
6	78	897	0.0536692223439211	06Q2	Pharma	[redacted]
6	6	1265	0.0536692223439211	06Q2	Haushaltspap	BrandY1
6	58	32	0.0536692223439211	06Q2	Erlebnis dm	[redacted]
6	54	782	0.0536692223439211	06Q2	Bad	[redacted]
6	5	201	0.0525739320920044	06Q2	Haarpflege	dm-green-BrandT1
6	18	1349	0.0514786418400876	06Q2	Haush.artik.	[redacted]
6	15	238	0.0503833515881709	06Q2	WPR	BrandF2
6	119	79	0.0503833515881709	06Q2	Augen Make u	[redacted]

***06Q3*, Cluster 5**

Cluster	S-A ID	Brand ID	*Centroid* Value	Quarter	Sub-Assortment	Brand
5	48	210	0.396174863387978	06Q3	Sonne	dm-BrandS1
5	51	207	0.259562841530055	06Q3	Fu	dm-BrandK1
5	11	65	0.229508196721311	06Q3	Haarstyling	[redacted]
5	66	218	0.222677595628415	06Q3	Nassrasur	dm-BrandK1 [redacted]
5	36	213	0.218579234972678	06Q3	Entferner	dm-BrandO1
5	40	1772	0.183060109289617	06Q3	Bonbon	BrandC2
5	36	207	0.173497267759563	06Q3	Entferner	dm-BrandK1
5	54	218	0.154371584699454	06Q3	Bad	dm-BrandK1 [redacted]
5	53	-1	0.150273224043716	06Q3	Posten/Saiso	?
5	127	200	0.147540983606557	06Q3	Ebelin Body	[redacted]
5	42	187	0.129781420765027	06Q3	Mund/Zahn	[redacted]
5	121	1309	0.121584699453552	06Q3	Nagelpflege	[redacted]
5	110	1988	0.117486338797814	06Q3	Feinstrick 1	[redacted]
5	3	267	0.116120218579235	06Q3	Damenhygiene	[redacted]

5	119	1290	0.109289617486339	06Q3	Augen Make u	[redacted]
5	34	-1	0.109289617486339	06Q3	Fotozubehr	?
5	40	381	0.0997267759562841	06Q3	Bonbon	[redacted]
5	5	257	0.0997267759562841	06Q3	Haarpflege	[redacted]
5	39	257	0.092896174863388	06Q3	Hand	[redacted]
5	5	33	0.0860655737704918	06Q3	Haarpflege	BrandZ1
5	5	66	0.0860655737704918	06Q3	Haarpflege	[redacted]
5	119	29	0.0846994535519126	06Q3	Augen Make u	[redacted]
5	39	1999	0.0833333333333333	06Q3	Hand	[redacted]
5	42	32	0.0833333333333333	06Q3	Mund/Zahn	[redacted]
5	56	1245	0.0819672131147541	06Q3	Filter&Folien	[redacted]
5	54	782	0.0792349726775956	06Q3	Bad	[redacted]
5	15	439	0.0778688524590164	06Q3	WPR	[redacted]
5	20	219	0.0737704918032787	06Q3	Nager	dm-BrandB1
5	16	433	0.0737704918032787	06Q3	Duftkerzen-M	[redacted]
5	57	5	0.0724043715846995	06Q3	Deo	[redacted]
5	42	58	0.0710382513661202	06Q3	Mund/Zahn	[redacted]
5	57	218	0.069672131147541	06Q3	Deo	dm-BrandK1 [redacted]
5	15	249	0.069672131147541	06Q3	WPR	[redacted]
5	6	1265	0.0669398907103825	06Q3	Haushaltspap	BrandY1
5	60	162	0.0669398907103825	06Q3	Gesicht	[redacted]
5	18	1349	0.0655737704918033	06Q3	Haush.artik.	[redacted]
5	60	1365	0.064207650273224	06Q3	Gesicht	[redacted]
5	30	59	0.064207650273224	06Q3	Babypflege	[redacted]
5	119	79	0.0601092896174863	06Q3	Augen Make u	[redacted]
5	70	2359	0.0601092896174863	06Q3	Fotoarbeiten	[redacted] [redacted]
5	75	381	0.0601092896174863	06Q3	Reform	[redacted]
5	5	952	0.0601092896174863	06Q3	Haarpflege	[redacted]
5	74	861	0.0587431693989071	06Q3	Filme	[redacted]
5	15	262	0.0587431693989071	06Q3	WPR	[redacted]
5	18	1884	0.0560109289617486	06Q3	Haush.artik.	[redacted]
5	119	49	0.0560109289617486	06Q3	Augen Make u	[redacted]
5	40	314	0.0560109289617486	06Q3	Bonbon	[redacted]
5	11	756	0.0560109289617486	06Q3	Haarstyling	[redacted]
5	78	897	0.0560109289617486	06Q3	Pharma	[redacted]
5	30	81	0.0546448087431694	06Q3	Babypflege	[redacted]
5	3	2191	0.0546448087431694	06Q3	Damenhygiene	BrandA2
5	121	79	0.0532786885245902	06Q3	Nagelpflege	[redacted]
5	5	967	0.0532786885245902	06Q3	Haarpflege	[redacted]
5	39	61	0.0532786885245902	06Q3	Hand	[redacted]
5	5	1916	0.0532786885245902	06Q3	Haarpflege	[redacted]
5	119	158	0.0532786885245902	06Q3	Augen Make u	[redacted]
5	42	1140	0.0519125683060109	06Q3	Mund/Zahn	[redacted]

5	54	811	0.0519125683060109	06Q3	Bad	[redacted]
5	5	91	0.0519125683060109	06Q3	Haarpflege	[redacted]
5	120	1290	0.0519125683060109	06Q3	Lippenpflege	[redacted]
5	75	306	0.0519125683060109	06Q3	Reform	[redacted]
5	118	1290	0.0519125683060109	06Q3	Make up	[redacted]
5	72	1213	0.0505464480874317	06Q3	Tee	[redacted]
5	6	1263	0.0505464480874317	06Q3	Haushaltspap	[redacted]

(*06Q1*, Cluster 15)

Cluster	S-A ID	Brand ID	*Centroid* Value	Quarter	Sub-Assortment	Brand
15	112	-1	0.416398713826367	06Q1	Ostern klein	?
15	53	-1	0.233118971061093	06Q1	Posten/Saiso	?
15	45	2076	0.173633440514469	06Q1	Kerzen	BrandH2
15	16	433	0.168810289389068	06Q1	Duftkerzen-M	[redacted]
15	11	65	0.155948553054662	06Q1	Haarstyling	[redacted]
15	39	201	0.147909967845659	06Q1	Hand	dm-green-BrandT1
15	51	207	0.135048231511254	06Q1	Fu	dm-BrandK1
15	127	200	0.12540192926045	06Q1	Ebelin Body	[redacted]
15	56	1245	0.122186495176849	06Q1	Filter&Folien	[redacted]
15	5	33	0.120578778135048	06Q1	Haarpflege	BrandZ1
15	113	-1	0.120578778135048	06Q1	Ostern mitte	?
15	48	210	0.120578778135048	06Q1	Sonne	dm-BrandS1
15	66	218	0.117363344051447	06Q1	Nassrasur	dm-BrandK1 [redacted]
15	75	306	0.114147909967846	06Q1	Reform	[redacted]
15	61	4	0.106109324758842	06Q1	Bodylotion	[redacted]
15	42	187	0.104501607717042	06Q1	Mund/Zahn	[redacted]
15	40	1772	0.104501607717042	06Q1	Bonbon	BrandC2
15	110	1988	0.0964630225080386	06Q1	Feinstrick 1	[redacted]
15	18	1349	0.0932475884244373	06Q1	Haush.artik.	[redacted]
15	40	381	0.0916398713826367	06Q1	Bonbon	[redacted]
15	42	32	0.0916398713826367	06Q1	Mund/Zahn	[redacted]
15	6	1263	0.0916398713826367	06Q1	Haushaltspap	[redacted]
15	60	1365	0.0852090032154341	06Q1	Gesicht	[redacted]
15	3	267	0.0852090032154341	06Q1	Damenhygiene	[redacted]
15	20	219	0.0852090032154341	06Q1	Nager	dm-BrandB1
15	40	314	0.0852090032154341	06Q1	Bonbon	[redacted]
15	18	1884	0.0836012861736334	06Q1	Haush.artik.	[redacted]
15	3	2191	0.0819935691318328	06Q1	Damenhygiene	BrandA2
15	4	1241	0.0755627009646302	06Q1	Kosmetik-Pap	BrandB2
15	15	439	0.0739549839228296	06Q1	WPR	[redacted]
15	36	213	0.0739549839228296	06Q1	Entferner	dm-BrandO1

15	15	249	0.0739549839228296	06Q1	WPR	[redacted]
15	5	201	0.0707395498392283	06Q1	Haarpflege	dm-green-BrandT1
15	72	1213	0.0691318327974277	06Q1	Tee	[redacted]
15	54	201	0.0659163987138264	06Q1	Bad	dm-green-BrandT1
15	6	1265	0.0659163987138264	06Q1	Haushaltspap	BrandY1
15	75	381	0.0659163987138264	06Q1	Reform	[redacted]
15	57	5	0.0643086816720257	06Q1	Deo	[redacted]
15	56	1243	0.0643086816720257	06Q1	Filter&Folien	BrandX1
15	36	207	0.0643086816720257	06Q1	Entferner	dm-BrandK1
15	60	176	0.0643086816720257	06Q1	Gesicht	[redacted] [redacted]
15	119	1290	0.0610932475884244	06Q1	Augen Make u	[redacted]
15	15	270	0.0610932475884244	06Q1	WPR	[redacted]
15	5	952	0.0594855305466238	06Q1	Haarpflege	[redacted]
15	121	1309	0.0594855305466238	06Q1	Nagelpflege	[redacted]
15	60	162	0.0578778135048231	06Q1	Gesicht	[redacted]
15	120	1290	0.0578778135048231	06Q1	Lippenpflege	[redacted]
15	5	66	0.0578778135048231	06Q1	Haarpflege	[redacted]
15	30	59	0.0578778135048231	06Q1	Babypflege	[redacted]
15	42	58	0.0578778135048231	06Q1	Mund/Zahn	[redacted]
15	61	201	0.0578778135048231	06Q1	Bodylotion	dm-green-BrandT1
15	57	218	0.0562700964630225	06Q1	Deo	dm-BrandK1 [redacted]
15	34	-1	0.0562700964630225	06Q1	Fotozubehr	?
15	42	1140	0.0562700964630225	06Q1	Mund/Zahn	[redacted]
15	58	731	0.0562700964630225	06Q1	Erlebnis dm	green-BrandU1
15	58	1265	0.0562700964630225	06Q1	Erlebnis dm	BrandY1
15	34	2189	0.0562700964630225	06Q1	Fotozubehr	[redacted]
15	54	782	0.0546623794212219	06Q1	Bad	[redacted]
15	15	262	0.0530546623794212	06Q1	WPR	[redacted]
15	40	1793	0.0530546623794212	06Q1	Bonbon	[redacted]
15	61	274	0.0514469453376206	06Q1	Bodylotion	[redacted]
15	72	2000	0.0514469453376206	06Q1	Tee	[redacted]
15	18	1885	0.0514469453376206	06Q1	Haush.artik.	[redacted]

G.3.3 The *Baby/Young Families* Cluster

06Q1, Cluster 4

Cluster	S-A ID	Brand ID	*Centroid* Value	Quarter	Sub-Assortment	Brand
4	79	199	0.477406679764244	06Q1	Baby-Push	BrandG2
4	112	-1	0.341846758349705	06Q1	Ostern klein	?
4	27	899	0.337917485265226	06Q1	Windeln	BrandE2
4	59	199	0.212180746561886	06Q1	KLKD-Push	BrandG2

4	40	381	0.212180746561886	06Q1	Bonbon	[redacted]
4	76	199	0.202357563850688	06Q1	Baby-Pull	BrandG2
4	37	1846	0.200392927308448	06Q1	Babyglasnahr	BrandA1
4	33	1846	0.180746561886051	06Q1	Babynahrung	BrandA1
4	37	1847	0.161100196463654	06Q1	Babyglasnahr	BrandL1
4	26	206	0.157170923379175	06Q1	Flaschen & S	dm-BrandQ1
4	33	1847	0.151277013752456	06Q1	Babynahrung	BrandL1
4	53	-1	0.143418467583497	06Q1	Posten/Saiso	?
4	37	731	0.131630648330059	06Q1	Babyglasnahr	green-BrandU1
4	6	1265	0.131630648330059	06Q1	Haushaltspap	BrandY1
4	76	217	0.121807465618861	06Q1	Baby-Pull	[redacted]
4	42	187	0.117878192534381	06Q1	Mund/Zahn	[redacted]
4	30	2067	0.111984282907662	06Q1	Babypflege	[redacted]
4	3	267	0.104125736738703	06Q1	Damenhygiene	[redacted]
4	66	218	0.0962671905697446	06Q1	Nassrasur	dm-BrandK1 [redacted]
4	34	-1	0.0962671905697446	06Q1	Fotozubehr	?
4	58	1265	0.0943025540275049	06Q1	Erlebnis dm	BrandY1
4	113	-1	0.0943025540275049	06Q1	Ostern mitte	?
4	18	1884	0.0884086444007859	06Q1	Haush.artik.	[redacted]
4	5	33	0.0864440078585462	06Q1	Haarpflege	BrandZ1
4	42	58	0.0864440078585462	06Q1	Mund/Zahn	[redacted]
4	72	1213	0.0785854616895874	06Q1	Tee	[redacted]
4	42	32	0.0766208251473477	06Q1	Mund/Zahn	[redacted]
4	34	205	0.0766208251473477	06Q1	Fotozubehr	[redacted]
4	30	59	0.0746561886051081	06Q1	Babypflege	[redacted]
4	39	201	0.0746561886051081	06Q1	Hand	dm-green-BrandT1
4	60	174	0.0726915520628684	06Q1	Gesicht	[redacted]
4	11	65	0.0707269155206287	06Q1	Haarstyling	[redacted]
4	127	200	0.0707269155206287	06Q1	Ebelin Body	[redacted]
4	30	2111	0.0667976424361493	06Q1	Babypflege	[redacted]
4	15	238	0.0628683693516699	06Q1	WPR	BrandF2
4	15	262	0.0609037328094303	06Q1	WPR	[redacted]
4	74	861	0.0609037328094303	06Q1	Filme	[redacted]
4	18	1349	0.0609037328094303	06Q1	Haush.artik.	[redacted]
4	56	1244	0.0609037328094303	06Q1	Filter&Folien	[redacted]
4	51	207	0.0589390962671906	06Q1	Fu	dm-BrandK1
4	56	1245	0.0589390962671906	06Q1	Filter&Folien	[redacted]
4	75	306	0.0589390962671906	06Q1	Reform	[redacted]
4	110	1988	0.0569744597249509	06Q1	Feinstrick 1	[redacted]
4	60	1365	0.0569744597249509	06Q1	Gesicht	[redacted]
4	75	381	0.0569744597249509	06Q1	Reform	[redacted]
4	40	1793	0.0569744597249509	06Q1	Bonbon	[redacted]
4	5	952	0.0550098231827112	06Q1	Haarpflege	[redacted]

4	70	2359	0.0550098231827112	06Q1	Fotoarbeiten	[redacted] [redacted]
4	15	221	0.0550098231827112	06Q1	WPR	[redacted]
4	48	210	0.0530451866404715	06Q1	Sonne	dm-BrandS1
4	30	274	0.0530451866404715	06Q1	Babypflege	[redacted]
4	15	439	0.0530451866404715	06Q1	WPR	[redacted]
4	40	314	0.0530451866404715	06Q1	Bonbon	[redacted]
4	56	1243	0.0530451866404715	06Q1	Filter&Folien	BrandX1
4	38	-1	0.0510805500982318	06Q1	Regenschirme	?
4	3	2191	0.0510805500982318	06Q1	Damenhygiene	BrandA2

***06Q2*, Cluster 10**

Cluster	S-A ID	Brand ID	*Centroid* Value	Quarter	Sub-Assortment	Brand
10	79	199	0.561122244488978	06Q2	Baby-Push	BrandG2
10	27	899	0.372745490981964	06Q2	Windeln	BrandE2
10	37	1846	0.244488977955912	06Q2	Babyglasnahr	BrandA1
10	59	199	0.226452905811623	06Q2	KLKD-Push	BrandG2
10	48	210	0.202404809619238	06Q2	Sonne	dm-BrandS1
10	76	199	0.190380761523046	06Q2	Baby-Pull	BrandG2
10	26	206	0.180360721442886	06Q2	Flaschen & S	dm-BrandQ1
10	53	-1	0.180360721442886	06Q2	Posten/Saiso	?
10	42	187	0.178356713426854	06Q2	Mund/Zahn	[redacted]
10	40	381	0.170340681362725	06Q2	Bonbon	[redacted]
10	33	1846	0.162324649298597	06Q2	Babynahrung	BrandA1
10	37	1847	0.148296593186373	06Q2	Babyglasnahr	BrandL1
10	33	1847	0.13627254509018	06Q2	Babynahrung	BrandL1
10	37	731	0.130260521042084	06Q2	Babyglasnahr	green-BrandU1
10	76	217	0.124248496993988	06Q2	Baby-Pull	[redacted]
10	3	267	0.118236472945892	06Q2	Damenhygiene	[redacted]
10	6	1265	0.114228456913828	06Q2	Haushaltspap	BrandY1
10	30	2067	0.100200400801603	06Q2	Babypflege	[redacted]
10	5	33	0.100200400801603	06Q2	Haarpflege	BrandZ1
10	30	59	0.0961923847695391	06Q2	Babypflege	[redacted]
10	112	-1	0.094188376753507	06Q2	Ostern klein	?
10	42	58	0.094188376753507	06Q2	Mund/Zahn	[redacted]
10	3	2191	0.0901803607214429	06Q2	Damenhygiene	BrandA2
10	18	1884	0.0861723446893787	06Q2	Haush.artik.	[redacted]
10	11	65	0.0841683366733467	06Q2	Haarstyling	[redacted]
10	48	206	0.0841683366733467	06Q2	Sonne	dm-BrandQ1
10	42	32	0.0781563126252505	06Q2	Mund/Zahn	[redacted]
10	34	205	0.0781563126252505	06Q2	Fotozubehr	[redacted]
10	34	-1	0.0761523046092184	06Q2	Fotozubehr	?

10	51	207	0.0741482965931864	06Q2	Fu	dm-BrandK1
10	66	218	0.0721442885771543	06Q2	Nassrasur	dm-BrandK1 [redacted]
10	39	201	0.0721442885771543	06Q2	Hand	dm-green-BrandT1
10	110	1988	0.0721442885771543	06Q2	Feinstrick 1	[redacted]
10	56	1245	0.0701402805611222	06Q2	Filter&Folien	[redacted]
10	75	306	0.0701402805611222	06Q2	Reform	[redacted]
10	70	2359	0.0701402805611222	06Q2	Fotoarbeiten	[redacted] [redacted]
10	58	1265	0.0701402805611222	06Q2	Erlebnis dm	BrandY1
10	74	861	0.0681362725450902	06Q2	Filme	[redacted]
10	127	200	0.0661322645290581	06Q2	Ebelin Body	[redacted]
10	10	-1	0.064128256513026	06Q2	Smereien-M	?
10	75	381	0.064128256513026	06Q2	Reform	[redacted]
10	5	952	0.062124248496994	06Q2	Haarpflege	[redacted]
10	15	238	0.062124248496994	06Q2	WPR	BrandF2
10	30	2111	0.0601202404809619	06Q2	Babypflege	[redacted]
10	30	274	0.0601202404809619	06Q2	Babypflege	[redacted]
10	30	81	0.0601202404809619	06Q2	Babypflege	[redacted]
10	4	1241	0.0601202404809619	06Q2	Kosmetik-Pap	BrandB2
10	36	207	0.0581162324649299	06Q2	Entferner	dm-BrandK1
10	15	221	0.0581162324649299	06Q2	WPR	[redacted]
10	42	1140	0.0581162324649299	06Q2	Mund/Zahn	[redacted]
10	30	899	0.0561122244488978	06Q2	Babypflege	BrandE2
10	58	731	0.0541082164328657	06Q2	Erlebnis dm	green-BrandU1
10	61	4	0.0541082164328657	06Q2	Bodylotion	[redacted]
10	5	4	0.0541082164328657	06Q2	Haarpflege	[redacted]
10	33	1855	0.0541082164328657	06Q2	Babynahrung	[redacted]
10	5	91	0.0541082164328657	06Q2	Haarpflege	[redacted]
10	54	782	0.0541082164328657	06Q2	Bad	[redacted]
10	121	1309	0.0521042084168337	06Q2	Nagelpflege	[redacted]
10	40	1772	0.0521042084168337	06Q2	Bonbon	BrandC2
10	15	262	0.0501002004008016	06Q2	WPR	[redacted]
10	40	1789	0.0501002004008016	06Q2	Bonbon	[redacted]

06Q3, Cluster 23

Cluster	S-A ID	Brand ID	*Centroid* Value	Quarter	Sub-Assortment	Brand
23	79	199	0.469043151969981	06Q3	Baby-Push	BrandG2
23	27	899	0.322701688555347	06Q3	Windeln	BrandE2
23	59	199	0.221388367729831	06Q3	KLKD-Push	BrandG2
23	40	381	0.221388367729831	06Q3	Bonbon	[redacted]
23	42	187	0.213883677298311	06Q3	Mund/Zahn	[redacted]
23	53	-1	0.204502814258912	06Q3	Posten/Saiso	?

23	26	206	0.204502814258912	06Q3	Flaschen & S	dm-BrandQ1
23	37	1846	0.170731707317073	06Q3	Babyglasnahr	BrandA1
23	48	210	0.153846153846154	06Q3	Sonne	dm-BrandS1
23	76	199	0.150093808630394	06Q3	Baby-Pull	BrandG2
23	37	1847	0.148217636022514	06Q3	Babyglasnahr	BrandL1
23	33	1846	0.146341463414634	06Q3	Babynahrung	BrandA1
23	34	-1	0.142589118198874	06Q3	Fotozubehr	?
23	33	1847	0.140712945590994	06Q3	Babynahrung	BrandL1
23	3	267	0.118198874296435	06Q3	Damenhygiene	[redacted]
23	30	2067	0.116322701688555	06Q3	Babypflege	[redacted]
23	76	217	0.106941838649156	06Q3	Baby-Pull	[redacted]
23	30	59	0.101313320825516	06Q3	Babypflege	[redacted]
23	70	2359	0.099437148217636	06Q3	Fotoarbeiten	[redacted] [redacted]
23	6	1265	0.099437148217636	06Q3	Haushaltspap	BrandY1
23	66	218	0.0938086303939962	06Q3	Nassrasur	dm-BrandK1 [redacted]
23	51	207	0.0938086303939962	06Q3	Fu	dm-BrandK1
23	37	731	0.0919324577861163	06Q3	Babyglasnahr	green-BrandU1
23	56	1245	0.0919324577861163	06Q3	Filter&Folien	[redacted]
23	75	306	0.0863039399624765	06Q3	Reform	[redacted]
23	34	205	0.0844277673545966	06Q3	Fotozubehr	[redacted]
23	127	200	0.0825515947467167	06Q3	Ebelin Body	[redacted]
23	36	213	0.0787992495309568	06Q3	Entferner	dm-BrandO1
23	30	899	0.075046904315197	06Q3	Babypflege	BrandE2
23	3	2191	0.075046904315197	06Q3	Damenhygiene	BrandA2
23	42	58	0.0731707317073171	06Q3	Mund/Zahn	[redacted]
23	18	1884	0.0712945590994371	06Q3	Haush.artik.	[redacted]
23	39	201	0.0694183864915572	06Q3	Hand	dm-green-BrandT1
23	54	218	0.0675422138836773	06Q3	Bad	dm-BrandK1 [redacted]
23	5	952	0.0675422138836773	06Q3	Haarpflege	[redacted]
23	40	314	0.0675422138836773	06Q3	Bonbon	[redacted]
23	5	33	0.0675422138836773	06Q3	Haarpflege	BrandZ1
23	11	65	0.0656660412757974	06Q3	Haarstyling	[redacted]
23	15	221	0.0656660412757974	06Q3	WPR	[redacted]
23	104	199	0.0637898686679174	06Q3	KLKD-Pull	BrandG2
23	5	91	0.0637898686679174	06Q3	Haarpflege	[redacted]
23	74	861	0.0637898686679174	06Q3	Filme	[redacted]
23	40	1772	0.0619136960600375	06Q3	Bonbon	BrandC2
23	61	4	0.0619136960600375	06Q3	Bodylotion	[redacted]
23	39	61	0.0619136960600375	06Q3	Hand	[redacted]
23	15	270	0.0600375234521576	06Q3	WPR	[redacted]
23	5	257	0.0600375234521576	06Q3	Haarpflege	[redacted]
23	30	274	0.0581613508442777	06Q3	Babypflege	[redacted]
23	30	2111	0.0581613508442777	06Q3	Babypflege	[redacted]

23	38	-1	0.0581613508442777	06Q3	Regenschirme	?
23	18	1349	0.0525328330206379	06Q3	Haush.artik.	[redacted]
23	75	381	0.0525328330206379	06Q3	Reform	[redacted]
23	36	207	0.0525328330206379	06Q3	Entferner	dm-BrandK1
23	30	81	0.050656660412758	06Q3	Babypflege	[redacted]

Bibliography

[BDG+08] Ulrik Brandes, Daniel Delling, Marco Gaertler, Robert Görke, Martin Höfer, Zoran Nikoloski, and Dorothea Wagner. On Modularity Clustering. *IEEE Transactions on Knowledge and Data Engineering*, 20:172–188, 2008.

[BE05] Ulrik Brandes and Thomas Erlebach, editors. *Network Analysis: Methodological Foundations*, volume 3418 of *Lecture Notes in Computer Science*. Springer-Verlag, February 2005.

[BGW03] Ulrik Brandes, Marco Gaertler, and Dorothea Wagner. Experiments on Graph Clustering Algorithms. In *Proceedings of the 11th Annual European Symposium on Algorithms (ESA'03)*, volume 2832 of *Lecture Notes in Computer Science*, pages 568–579. Springer-Verlag, September 2003.

[BGW07] Ulrik Brandes, Marco Gaertler, and Dorothea Wagner. Engineering Graph Clustering: Models and Experimental Evaluation. *ACM Journal of Experimental Algorithmics*, 12(1.1):1–26, 2007.

[Die00] Reinhard Diestel. *Graph Theory*. Graduate Texts in Mathematics. Springer-Verlag, 2nd edition, 2000.

[Fer03] Reginald Ferber. *Information Retrieval Suchmodelle und Data-Mining-Verfahren für Textsammlungen und das Web*. dpunkt.verlag Heidelberg, 2003.

[GGW07] Marco Gaertler, Robert Görke, and Dorothea Wagner. Significance-Driven Graph Clustering. In *Proceedings of the 3rd International Conference on Algorithmic Aspects in Information and Management (AAIM'07)*, Lecture Notes in Computer Science, pages 11–26. Springer-Verlag, June 2007.

[GGWW06] Marco Gaertler, Robert Görke, Dorothea Wagner, and Silke Wagner. How to Cluster Evolving Graphs. In *Proceedings of the European Conference of Complex Systems (ECCS'06)*, September 2006.

[Gla08] Dieter Glaser. Zeitexpandiertes graphenclustern – modellierung und experimente. Master's thesis, Universität Karlsruhe, 2008.

[Gmb08] Payback GmbH. markenwelten - marken und ihre macher. Europa-Fachpresse-Verlag GmbH, October 2008.

[Gro] Data Mining Group. http://www.dmg.org/. A independent, vendor led group which develops data mining standards, such as PMML.

[HK06] Jiawei Han and Micheline Kamber. *Data Mining Concepts and Techniques*. Morgan Kaufmann Publishers Inc,US, 2006.

[HKKS04] J. Hopcroft, O. Khan, B. Kulis, and B. Selman. Tracking evolving communities in large linked networks. *Proceedings of the National Academy of Science USA*, 101:Proceedings of the National Academy of Science USA, 2004.

[JF87] William P. Jones and George W. Furnas. Pictures of relevance: A geometric analysis of similarity measures. *Journal of the American Society for Information Science*, 38(6):420–442, 1987.

[MG08] MicroStrategy Deutschland GmbH M. Grossjung. Data mining workshop in daun, December 08.

[New04] Mark E. J. Newman. Detecting community structure in networks. *The European Physical Journal B*, 38, 2004.

[Tan57] T.T. Tanimoto. Ibm internal report. Technical report, Nov. 1957.

[vR79] C. J. van Rijsbergen. *INFORMATION RETRIEVAL*. London: Butterworths, 1979.