Wolfgang Kollewe
Forschungsgruppe Begriffsanalyse
Fachbereich Mathematik
Technische Hochschule Darmstadt

Data Representation by Nested Line Diagrams Illustrated by a Survey of Pensioners

Kollewe, Wolfgang: Data Representation by Nested Line Diagrams Illustrated by a Survey of Pensioners
Int. Classif. 18(1991) No.1. p.xx – xx, 12refs.

With Formal Concept Analysis surveys are analyzable in the way that a meaningful picture of the answers of the interviewed persons is available. Line diagrams of large concept lattices might become less readable up to the point that it is impossible to pursue the line segments with the eyes. Nested line diagrams give the opportunity to overcome these difficulties. The main idea of nested line diagrams is to partition the line diagram into boxes so that line segments between two boxes are all parallel and may be replaced by one line segment. The possibility to draw line diagrams with more than two factors does allow it to describe concept lattices with many hundred or thousand concepts in a clear structure. In practice it has often been proven useful to take standardized scales for the single levels. (Author)

1 Introduction

A frequently used method of data analysis is to describe structures of data with graphic representations. In Formal Concept Analysis data are represented by line diagrams. These line diagrams of concept lattices should not only reflect the mathematical structure, but also give a meaningful presentation for the data. If data sets are more extensive – as often in the social sciences and economics – the concept lattices will become also larger. Line diagrams of large concept lattices might become less readable up to the point that it is difficult or impossible to identify line segments and their endpoints. In the following article we will show the method of the nested line diagrams to overcome these problems. Nested line diagrams are first described in [8]. Certainly it would exceed the extent of this article to describe and to explain all methods of Formal Concept Analysis; for further information see [3, 6, 8, 9, 11, 12, 13].

First (section 2) we represent the theoretical foundations of nested line diagrams. Then we demonstrate (section 3) with the help of an example the single steps which have to be done to come from a many-valued context to a nested line diagram. Finally (section 4) we illustrate the possibilities of this method of evaluation by using two examples, which were taken from a questioning of a large number of pensioners.

2 Nested Line Diagram

The main idea of nested line diagrams is to partition the line diagram into boxes so that line segments between two boxes are all parallel and may be replaced by one line segment between the two boxes. The original line segments can be recognized by their endpoints which are assumed to coincide after tanslating one box onto another. A double line segment between two boxes means that each element of the upper box is greater than all elements of the lower box. Figure 2 shows a normal line diagram. Figure 3 and 4 represent nested line diagrams of the same concept lattice.

From the viewpoint of data analysis, the question arises: How can one determine meaningful partitions of the concept lattices which yield nested line diagrams? An answer can be derived from the following proposition [8].

Proposition 1 *Let* $\mathbf{K} := (G, M, I)$ *be a context and let* $M = M_1 \cup \ldots \cup M_n$. *Then*

$$(A, B) \mapsto (((B \cap M_1)', B \cap M_1), \ldots, ((B \cap M_n)', B \cap M_n))$$

for all $(A, B) \in \underline{\mathfrak{B}}(\mathbf{K})$ *describes an isomorphism* φ *from* $\underline{\mathfrak{B}}(\mathbf{K})$ *onto a subdirect product of the*

$\bigvee$-*semilattices* $\underline{\mathfrak{B}} := (G, M_i, I \cap G \times M_i)$ *(i = 1,2, ..., n).*

A nested line diagram with two factors is designed in the following way: In the first step we partition the set of the attributes: $M = M_1 \cup M_2$. Thereby it is important for the interpretation of the data that the chosen sets of attributes M_i are meaningful with respect to the evaluation. In the second step the line diagrams of the subcontexts $\mathbb{K}_i := (G, M_i, I \cap G \times M_i)(i = 1,2)$ are drawn and labelled with the names of the objects and attributes, as usual. In the third step we draw as an auxiliary structure a nested line diagram of the direct product of the concept lattices $\underline{\mathfrak{B}}(\mathbb{K}_i)$. This is done in the following way: we take first the line diagram of the concept lattice $\underline{\mathfrak{B}}(\mathbb{K}_1)$ and blow up the small circles to congruent boxes in each we draw the same diagram of the second factor $\underline{\mathfrak{B}}(\mathbb{K}_2)$. Then the circles for every concept γg $(g \in G)$ are located as follows: In the box which represents the concept $((\{g\}' \cap M_1)', \{g\}' \cap M_1)$ we localize the circle of the concept $((\{g\}' \cap M_2)', \{g\}' \cap M_2)$. In accordance to the *'Basic Theorem on Concept Lattices'* γg is supremum-dense in $\underline{\mathfrak{B}}(\mathbb{K})$. One can get all circles of the nested line diagram if one constitutes all suprema of the described concepts γg $(g \in G)$. The concepts of the subdirect product are distinguished and the other circles and the superfluous line segments eventually are erased.

3 An Example

The following example is based on the data of a questioning of a large number of pensioners. In this survey a total of 340 pensioners — among them 260 women — were asked about their social situation. Most of the data in this questioning is of a qualitative nature, but some data is quantitative, like the age of the pensioners, the number of children and grandchildren etc.

The collected data can be interpreted as a formal context. The units of the questioning are the objects and the answers give rise to the attributes. This leads to a many-valued context, which, in general, is defined as a quadruple (G, M, W, I), where G, M and W are sets and I a tenary incidence relation (*i.e.* $I \subset G \times M \times W$) such that $(g, m, w) \in I$ and $(g, m, v) \in I$ always imply $v = w$ (see [3, 4]). If $(g, m, w) \in I$ we say: the object g has the attribute m with the value w.

The first example describes how many children and grandchildren the married female pensioners have. Therefore our many-valued context has only two attributes, namely children and grandchildren. The labels of the objects show a peculiarity. Normally every object has its own label. In this case the names of the objects give information about the frequency of a combination of the attributes existing in our data set. For example, the label 11 means in this context that 11 married female pensioners have neither a child nor a grandchild. Combinations with frequency 0 are not objects of the context.

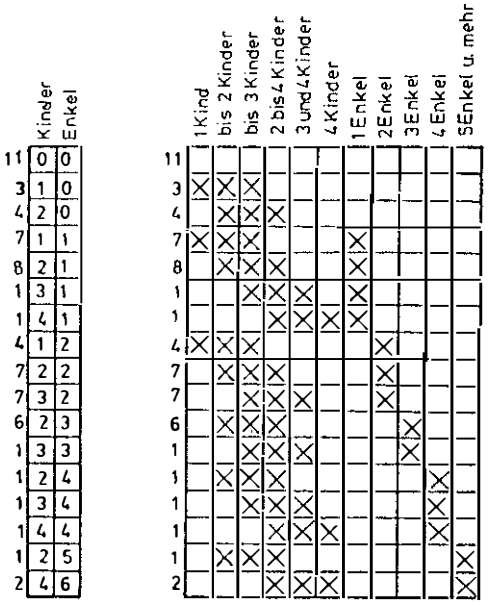

Figure 1: Many–valued context and derived context

To draw a line diagram we first have to transform the many-valued context into a one-valued context by the methods of conceptual scaling. The latter is called the derived context of the many-valued context. The methods of conceptual scaling are described in [2, 4, 5]. In the example the nominal scale $\mathbb{N}_5$ stands for the attribute 'Grandchildren' (Enkel) and the interordinal scale $\mathbb{I}_4$ stands for the attribute 'Children' (Kinder). After conceptual scaling the derived context has 11 one-valued attributes.

Now we have to decide about the partition of the one-valued attributes. In this case we settle on the partition which is given by the two selected scales, because another partition in our opinion is not meaningful. In a second step we have to decide which subcontext describes the outer line diagram. The determination depends on the aims of the evaluator. If we want a general view of the number of the children which the interviewed female pensioners have, we choose for the subcontext children as the outer line diagram. Is the general question, how many grandchildren the married female pensioners

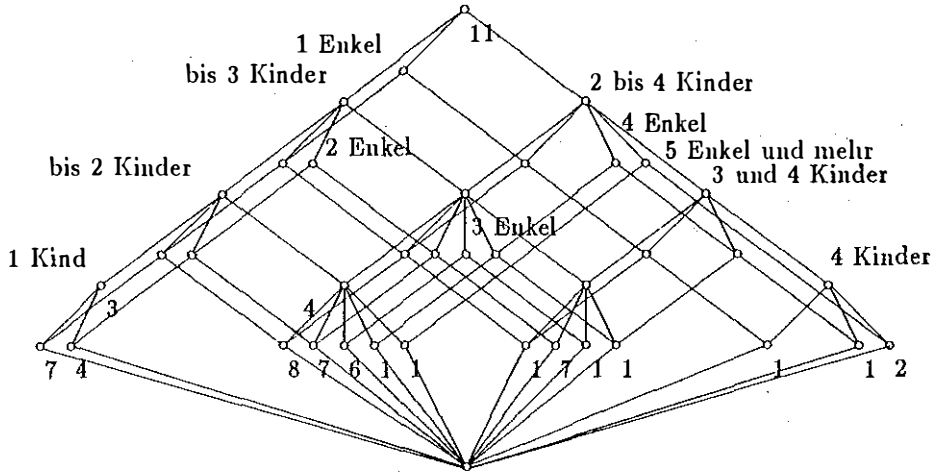

Figure 2: Offspring of the female pensioners 60 years and older

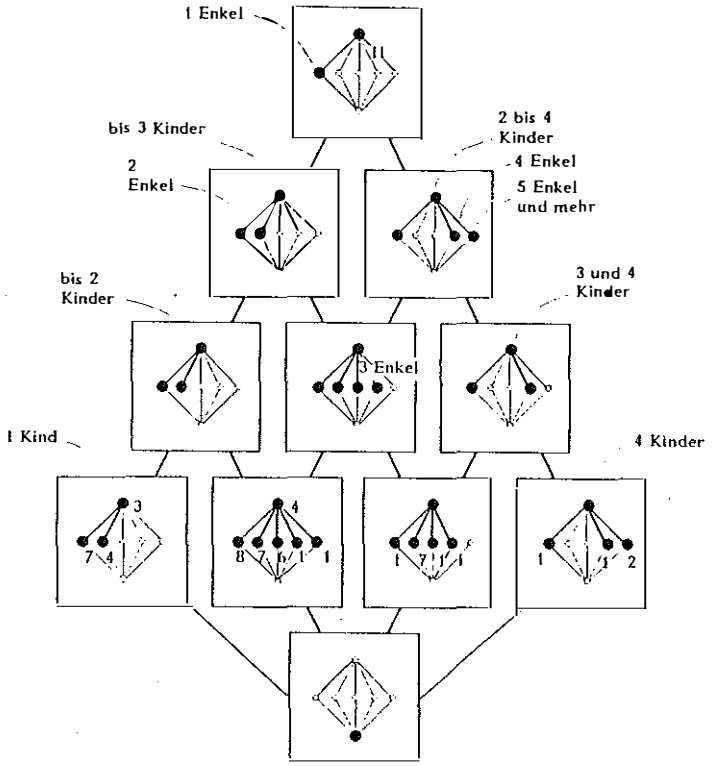

Figure 3: Children and grandchildren of the married female pensioners 60 years and older

have, than we select the subcontext grandchildren as the outer line diagram. For both decisions the nested line diagrams are drawn: figure 3 shows the first and figure 4 the second alternative.

In contrast to the normal line diagram the nested line diagrams are drawn with two kinds of circles. The bigger ones are the concepts of the main context. The small circles show the concepts of the direct product which are not concepts of the presented context. For doing we have two reasons. The more formal one is that a nested line diagram with these auxiliary line segments and circles has a better readability. The second one is meaningful: it is important to have this information explicitly. This shows for the evaluation which concepts are relevant for the interviewed persons and which are not. For example we can see explicitely in figure 3 that there exists no female pensioner who has one child and three grandchildren.

4 The Use of Nested Line Diagrams

The second example will give a deeper insight into the possibilities of evaluation with nested line diagrams. In this example we want to analyze the activities of the female pensioners in their private life. It is our aim to evaluate this question in connection with the marital status.

In the questionaire for the marital status there are three given categories for the interviewed persons to answer: 'Married' (verheiratet), 'Single' (alleinstehend) and 'Widowed' (verwitwet). For the activities in private life there are five given opportunities for answering: 'Hobby' (Hobby), 'Assistance in the garden' (Mithilfe im Garten), 'Assistance in the house or family' (Mithilfe im Haus bzw. in der Familie), 'Assistance in the kitchen' (Mithilfe bei der Küchenarbeit) and 'Assistance in caring for the children' (Mithilfe bei der Kinderbetreuung). The interviewed persons only need to mark these possibilities, which play a part in their private life. On that account we have not a many-valued context with 2 attributes, but a one-valued context with 8 attributes.

The construction of a nested line diagram has to follow two main principles, which sometimes compete with one another:

1. To have a clear and well readable structure.
2. To be a meaningful structure with respect to the aims of the evaluation.

To perform these principles as far as possible the attributes of the example are partitioned into the following two subsets: the marital status and the activities in private life.

The attributes of the marital status create the outer line diagram, which is isomorphic to the boolean lattice $\mathfrak{B}_3$. The attributes of the activities in private life form a lattice, which is embedable in the boolean lattice $\mathfrak{B}_5$.

Before interpreting the data, it has to be pointed out that the collected data is not corrected in any way, which is a principle of Formal Concept Analysis. In other words the drawn line diagrams show all combinations of answers which are given by the interviewed pensioners. So the nested line diagram has to fulfill two important functions:

1. To show incorrect answers or misunderstandings by the interviewed persons.
2. To uncover the structure of the collected data.

The clear structure of the line diagram (figure 5) immediately shows that a large number of the interviewed persons have misunderstood the intention of the question marital status. The intention of this question was to get information about the actual marital status. But a part of the questioned pensioners marked more than one alternative; they give an answer about every marital status they have had in their life. So we can interpret the answer of a female pensioner who has marked widowed and single in the following way: this woman was formerly married, her husband died and now she lives alone. In our nested line diagram we find a group of pensioners who have answered in the described way. About this group of persons we have not only the knowledge of the actual marital status but also the information about every marital status they have had in their life. An early correction of the data via a check of plausibility would have deprived us of this knowledge.

The nested line diagram shows that a big part of female pensioners neither have a hobby nor they do practice any other activity in their private life. This fact is relatively independent of the actual marital status. We can recognize this fact in the clear structure of the outer line diagram. Another detail, which is also observed very quickly is that very few persons have marked all possibilities. Furthermore the nested line diagram shows that if female pensioners practice any kind of activity that this activity is not the only one they carry out. But the attribute 'Hobby' chosen by the female pensioners often stands alone.

We recognize in this nested line diagram with two factors that the line diagram inside the boxes has a complex structure, which is still readable. Nevertheless this example also shows the limitation of the

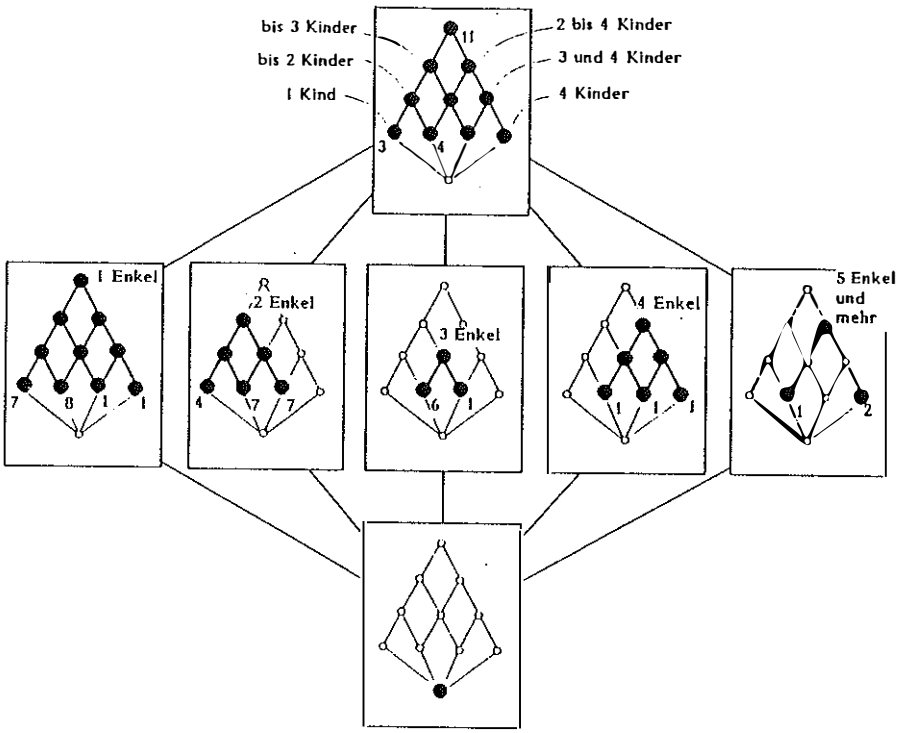

Figure 4: Grandchildren and children of the married female pensioners 60 years and older

applicability of the two fold nested line diagrams. To overcome this difficulty and to be able to represent contexts with several thousand of concepts we have to partition the attributes in more than two subsets. Our next example will give an impression how a manifold nested line diagram can be constructed.

The question which had led to the construction of this nested line diagram was: Who nurses the pensioners needing care and how much additional physical strain does the nurse has to stand in her/his private life. The first factor of this fivefold nested line diagram is the age of the pensioners. This many-valued attribute is scaled with the nominal scale $\mathbb{N}_7$ with intervals of five years. The second factor, the sex of the questioned persons, is scaled with the dichotomic scale $\mathbb{D}$. The third factor shows by whom the pensioners are nursed, if they are not able to look after themselves. In the questionaire there are three alternatives: 'Social ward' (Sozialstation), 'Relatives' (Angehörige) and 'Daughter/Daughter-in-law' (Tochter/Schwiegertochter). These possibilities are one-valued and it was possible to mark more than one. The fourth factor is of a more complicated structure. There we have splitted the attribute 'Daughter/Daughter-in-law into three attributes: 'Daughter', 'Daughter-in-law' and 'Daughter or daughter-in-law'. The third possibility was for those persons who have not specified the attribute more precisely. At the same level the attributes 'Nurse has to support a family' (Pflegende(r) hat Familie zu versorgen) and 'Nurse has another occupation' (Pflegende(r) geht einem Beruf nach) are represented. The last factor of our nested line diagram gives information about the time which the person, who nurses a pensioner, works in his normal occupation. This is an important fact for the evaluation if the pensioner is nursed in his family.

To split the attributes in such a way is surely meaningful. Another evaluator may partition the attributes in another way and gets another nested line diagram. We think the selected partition is useful for the given evaluation aims. One can easily see that the most female pensioners at the age of 70 – 74 years do not need any nursing. Next we see that the female pensioners of this age, who are not able to look after themselves, are nursed mostly by the daughter or daughter-in-law. Very often this daughter or daughter-in-law is a housewife or mother or both and in the same time has an occupation. In some cases she has to work more than twenty hours a week in her job.

The direct product of this five concept lattices has 12 960 concepts, but as we have seen the connected nested line diagram is still easy to read. If the concept lattice has even more concepts than our example, there is the possibility to include further factors to describe it. To take scales as factors for the single level is a natural choice. We use the designation 'scale' for such contexts which have a clear

FRAUEN 60 JAHRE UND ÄLTER GEGLIEDERT NACH FAMILIENSTAND UND TÄTIGKEITEN IM PERSÖNLICHEN BEREICH

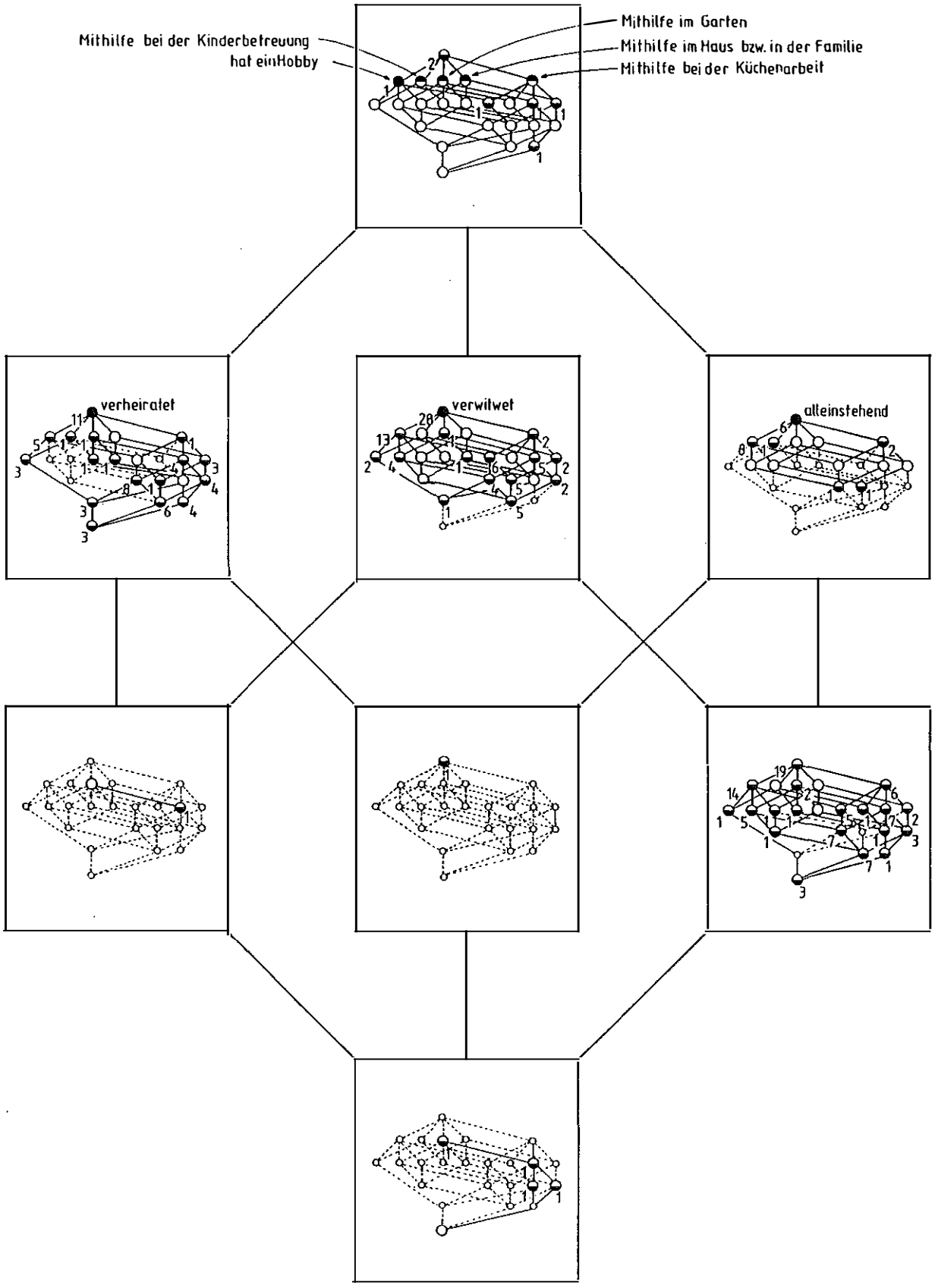

Figure 5: Activities in private life women 60 years and older

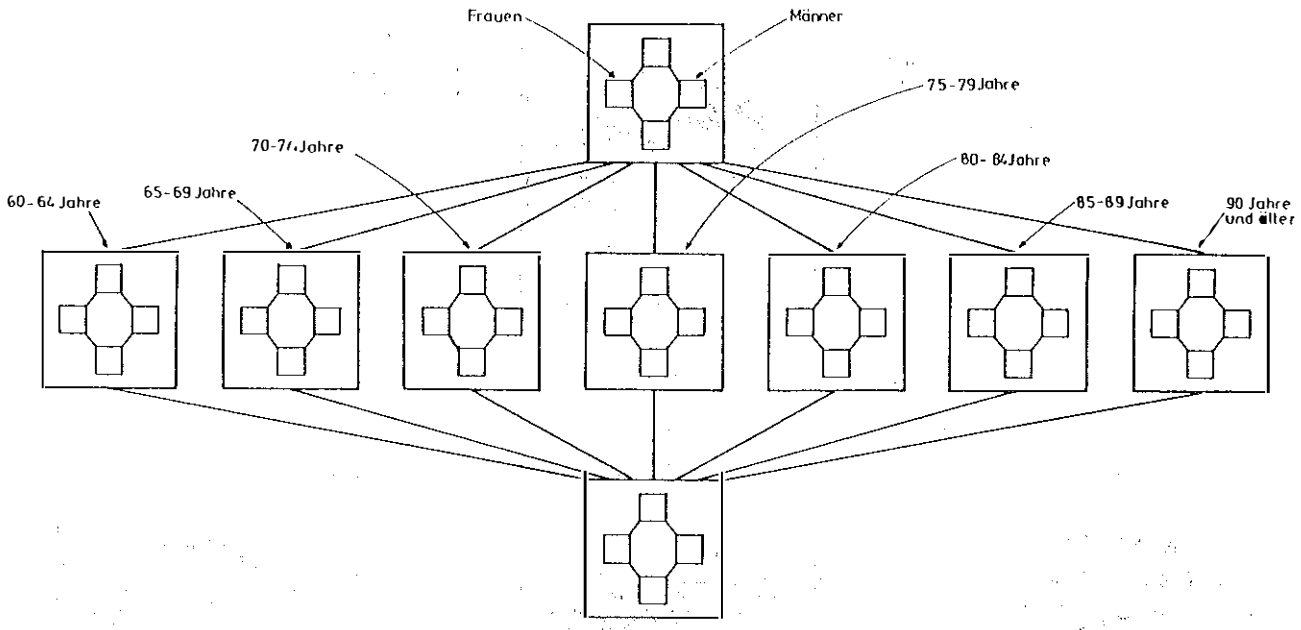

Figure 6: Pensioners who are needing care

structure and reflect some meaning. Thereby scales fulfill our above mentioned principles to be clear and to have a meaning. For many kinds of data standardized scales have been proven as useful. This also shows our example. We have taken standardized scales of ordinal type for the first, the second and the fifth factor.

It is normal to draw the line diagrams in one picture. In the present example we have deviated from this principle. The line diagram was splitted into two line diagrams: one with two factors and one with three factors. The attributes 'Age' and 'Sex' of the pensioners construct the nested line diagram with two factors. The outer line diagram was built of the attribute 'Age'. In contrary to a normal two fold nested line diagram the line diagram inside the large boxes gives no information about the single objects. There we find only a reference to the (nested) line diagram, which is normally drawn inside this box. This method of proceeding is based on the idea of a general map where we find the references to the single town maps. The general map does not give any information about the streets of a city. This information is found in the city map. Our general map is the nested line diagram which is formed of the attributes age and sex of the pensioners. A small example to illustrate this fact: If we want to receive information about the situation of nursing of the female pensioners who are between 70 and 74 years old, we must look first to the large box for pensioners between 70 and 74 years and than inside the small box 'Women' (Frauen). There we find the reference to the nested line diagram which gives information to our question. The nested line diagram where we find our concepts is drawn in the normal way and can be read as usual. This drawing technique offers the possibility to construct nested line diagrams without any restriction of the size of the paper.

This kind of nested line diagram needs a special method of labelling. Normally the name of an attribute is written at the circle of the concept with the smallest intent containing the attribute. To increase the readability of the line diagram we deviate from this method. In the example the subdiagram was completely labelled with the names of the attributes. A possibility to differentiate between names which come from the above lying boxes and names which belong exactly to the concept where they are labelled is to take two types of characters for labelling. This differentiation is not done in our example.

Altogether it can be said that manifold nested line diagrams open up the possibility to visualize structures of large data sets. We have seen that the way to partition the attributes is of high importance for the clearness and readability of the nested line diagram. The last example has illustrated that the possibility to draw the nested line diagram onto several sheets of paper permits to visualize very large data sets with many thousand of concepts. The only

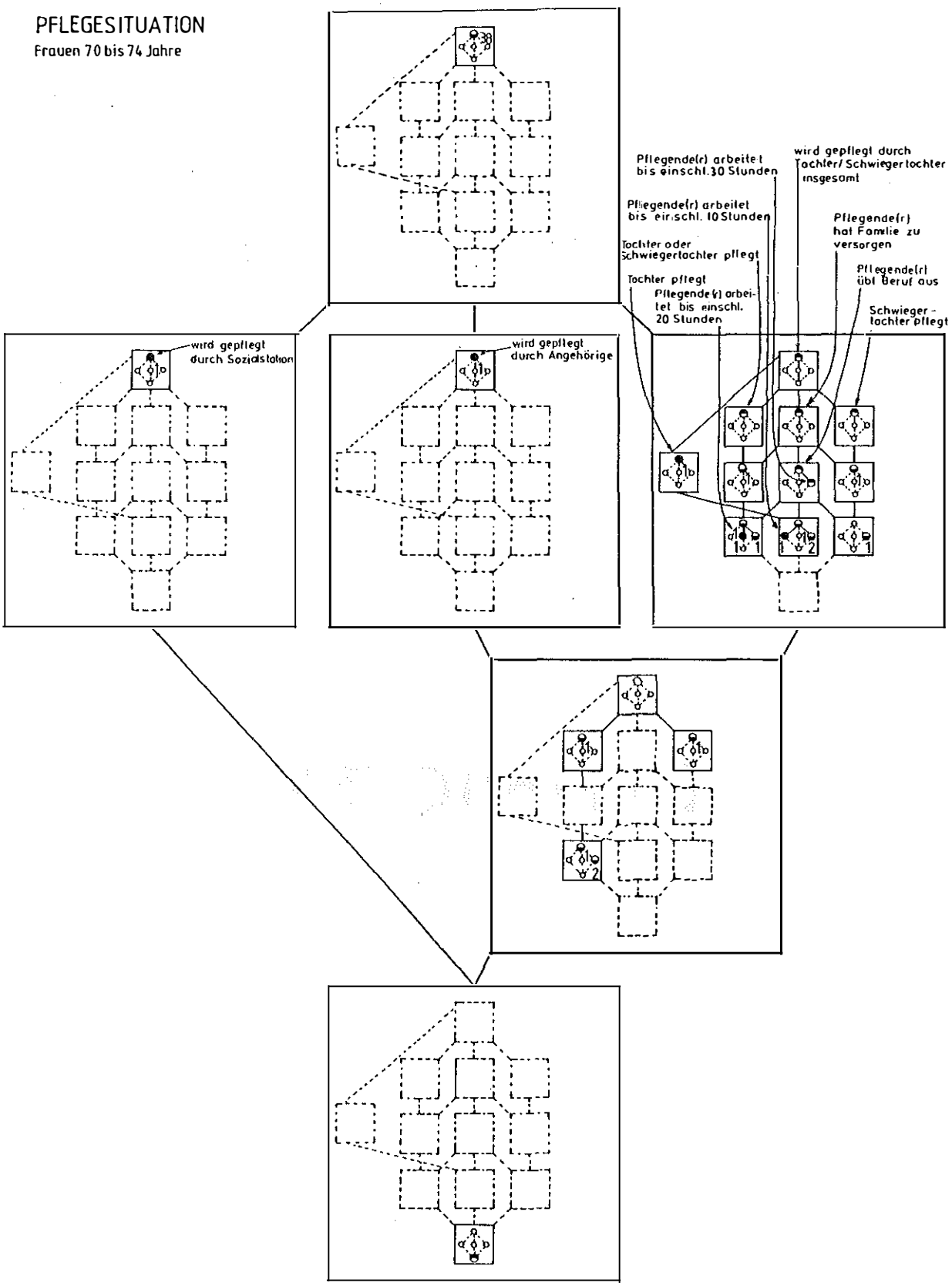

Figure 7: Female pensioners between 70 and 74 years who are needing care

restriction of this method is given by the wealth of imagination of the evaluator.

References

[1] Ganter, B.: Algorithmen zur formalen Begriffsanalyse. In: Ganter, B., Wille, R. and Wolff, K.E., editors, *Beiträge zur Begriffsanalyse*, pages 241–254, Mannheim, 1987, B.I. Wissenschaftsverlag

[2] Ganter, B., Stahl, J. and Wille, R.: Conceptual measurement and many-valued contexts. In: Gaul, W. and Schader, M., editors, *Classification as a tool of research*, pages 169–176, Amsterdam, 1986, North-Holland

[3] Ganter, B. and Wille, R.: Formale Begriffsanalyse, Manuscript 75p., B.I. Wissenschaftsverlag, (to appear)

[4] Ganter, B. and Wille, R.: Conceptual scaling. In: Roberts, F., editor, *Applications of combinatorics and graph theory in the biological science*, pages 139–167, New York, 1989, Springer Verlag

[5] Kollewe, W.: Evaluation of a survey with methods of formal concept analysis. In: Opitz, O., editor, *Conceptual and numerical analysis of data*, pages 123–134, Berlin–Heidelberg, 1989, Springer Verlag

[6] Luksch, P., Skorsky, M. and Wille, R.: On drawing concept lattices with a computer. In: Gaul, W. and Schader, M., editors, *Classification as a tool of research*, pages 269–274, Amsterdam, 1986, North-Holland

[7] Wille, R.: Restructuring lattice theory: an approach based on hierarchies of concepts. In: Rival, I., editor, *Ordered sets*, pages 445–470, Dordrecht–Boston, 1982, Reidel

[8] Wille, R.: Liniendiagramme hierarchischer Begriffssysteme. In: Bock, H.H, editor, *Anwendungen der Klassifikation: Datenanalyse und numerische Klassifikation*, pages 32–51, Frankfurt, 1984, Indeks Verlag

[9] Wille, R.: Bedeutungen von Begriffsverbänden. In: Ganter, B., Wille, R. and Wolff, K.E., editors, *Beiträge zur Begriffsanalyse*, pages 161–211, Mannheim, 1987, B.I. Wissenschaftsverlag

[10] Wille, R.: Dependencies of many valued attributes. In: Bock, H.H, editor, *Classification and related methods of data analysis*, pages 581–586, Amsterdam, 1988, North-Holland

[11] Wille, R.: Lattices in data analysis: how to draw them with a computer. In: Rival, I., editor, *Algorithms and order*, pages 33–58, Dordrecht–Boston, 1989, Reidel

[12] Wolff, K.E.: Vom begrifflichen Denken zur Begriffsanalyse, 13. Kolloquium Mathematik–Didaktik der Universität Bayreuth, 1988

[13] Wolff, K.E.: Einführung in die Formale Begriffsanalyse. Preprint, FH Darmstadt, 1988