Alternative Diagrams for Concept Lattices

Uta Priss

Fakultät Informatik, Ostfalia University, Wolfenbüttel, Germany
www.upriss.org.uk

Abstract. This paper presents several types of diagrams for visualising concept lattices: tabular Euler diagrams with additional features, such as icons, nesting, factorisation and object frequencies, and also Repeated Attribute Lattice diagrams and diagrams for AFOC-posets and Relation Algebraic Concept Analysis. The aim is to provide a variety of means for visualising conceptual structures with carefully designed but easy to interpret diagrammatic representations.

1 Introduction

This paper continues previous research on using Euler diagrams for representing concept lattices (Priss and Dürrschnabel 2024) in the sense of Formal Concept Analysis (FCA) as defined by Ganter & Wille (1999). But it also presents other diagrams for concept lattices and related structures in order to showcase a variety of FCA related visualisations. It is not claimed that everything presented here is fundamentally novel considering that the literature on diagrams spans more than 100 years (eg. Neurath (1936) as an older example). But in our opinion, FCA still does not achieve its full potential for its diagrammatic representations. While Hasse diagrams[1] are favoured by mathematicians, they are not self-evident because, for example, their edges can be misread as vectors[2]. Because FCA is often applied to complex data structures, which are by nature challenging to represent, there is a danger for FCA software visualisations to contain an overload of features, sliders, buttons, windows and other options with a steep learning curve. In this paper a selection of diagram types is introduced that are intended to be less complex but self-explanatory and, furthermore, a few diagram types (AFOC-posets, APT, RAL and RaCA diagrams) that are novel at least for FCA.

Introductions to FCA and Euler diagrams are not included in this paper but can be found in Ganter & Wille (1999) and Priss & Dürrschnabel (2024), respectively. As a compromise between FCA and Euler diagram terminology, elements of sets are called *objects* and the sets themselves are called *attributes*. Elements of lattices are called *concepts* and correspond to *zones* of an Euler diagram. FCA contexts are called *matrices*. All sets are finite. A Boolean lattice with 2^n elements or, in other words, n attributes is abbreviated as Bn in this paper.

[1] The diagrams should not be named after Hasse because he did not invent them but the name is established in the literature.

[2] Based on personal observations while teaching mathematics to first year university students.

2 Euler Diagrams

Euler diagrams are a means for visualising sets and their partial orders. Each set, or attribute, is represented by a closed curve, such as a circle or rectangle. Elements, or objects, can be written into the *zones*, which are the smallest areas that are surrounded by some curve segments. In particular, *tabular diagrams* are easy to read because they are essentially tables where row or column headings are allowed to extend to more than one neighbouring row or column. A tabular diagram should have as many non-shaded zones as there are concepts in a lattice but not all lattices are drawable as Euler diagrams of a certain type. For example, a B2 is drawable as a linear diagram, a B3 with rectangles or circles and a B4 with rectangles or ellipses but not with circles. Mechanisms exist for increasing the drawability such as shading areas which are not meant to exist, omitting the bottom concept, *splitting attributes* into more than one curve, using curves of other shapes or other visual cues, for example, colours and icons.

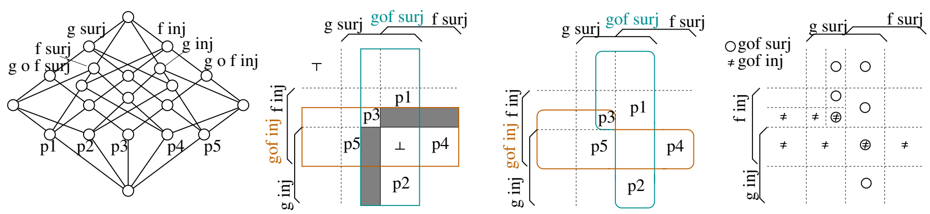

Fig. 1. A Hasse diagram of a lattice and three versions of a corresponding tabular diagram

The example in Fig.1 is based on a lattice of properties of functions from Ganter (2013, p. 76) and contains some objects (p1, p2, ... p5) inserted into zones. The top concept can be notated by $\top$ and the bottom concept by $\perp$. If the bottom concept is omitted its neighbours can instead be marked with $\perp$. In a tabular diagram, containment indicates a subset relationship, such as ‘g $\circ$ f surj’ being contained in‘g surj’, which according to FCA means that ‘g $\circ$ f surj’ implies ‘g surj’. Presumably, such containment is more intuitive to read from Euler diagrams. Most likely only experienced users will immediately observe that the structure in Fig.1 is a B4 with 2 further attributes (‘g $\circ$ f surj’ and ‘g $\circ$ f inj’), which however, is even more difficult to see in the Hasse diagram. The tabular diagrams on the right in Fig.1 showcase other visual cues such as other shapes and icons in order to avoid shading. Figure 2 displays a well-known example from Ganter & Wille (1999) converted into an Euler diagram. In general, tabular Euler diagrams are feasible for many kinds of lattices and in particular for those that are a nominal scale, crown-like or tree-like where parts of the tree can be B2s, B3s or even B4s and the bottom concept may be omitted.

3 Geometrical Diagrams, Nesting and Factorisation

This section discusses several different types of FCA diagrams. Ganter & Wille (1999) describe a special type of diagram called *geometrical diagram*, which is essentially a

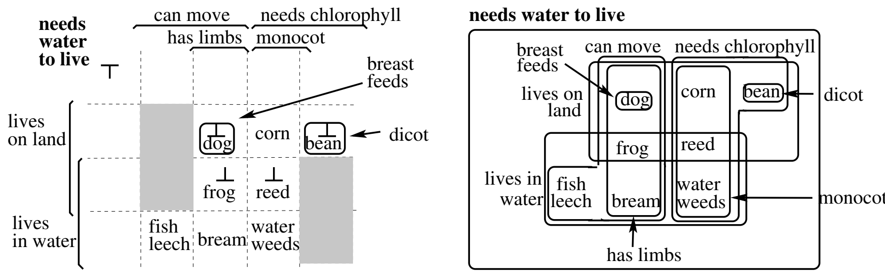

Fig. 2. 'Live in water' as a tabular diagram and a standard Euler diagram

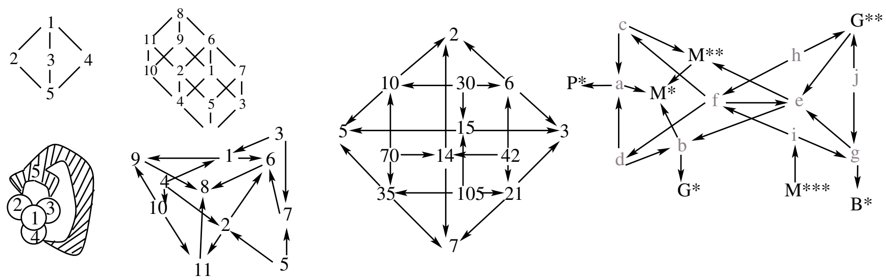

Fig. 3. Geometrical diagrams: nominal scale, two connected B3s, B4 and Forum Romanum

view from above or below onto a lattice. The bottom left of Fig. 3 shows a geometrical diagram of a nominal scale in Ganter & Wille (GW) notation. In GW notation the top and bottom concepts should be omitted but because any lattice can be a substructure of another lattice, top and bottom concepts must be drawable as well and are included in Fig.3. In GW notation concepts with a single upper neighbour are drawn adjacent but slightly behind the neighbour as shown for the concepts 2, 3 and 4. Concepts with more than one upper neighbour are drawn as n-simplexes (lines, triangles and so on). In the nominal scale concept 5 must be connected to concepts 2, 3 and 4, which is only possible with a distorted triangle (shown as a shape that is striped).

Clearly already a nominal scale is almost impossible to draw if the top and bottom concepts are included because of the directly adjacent concepts. Therefore we are introducing an *arrow notation* as graphs with arrows between concepts and their direct upper neighbours as in the other examples in Fig.3. The set of arrows pointing away from a single concept corresponds to a simplex in GW notation. The GW notation can be translated into the arrow notation with the exact same positioning of the concepts but the arrow notation can also represent lattices which are challenging in GW notation. Figure 3 shows further examples including the Forum Romanum of Ganter & Wille (1999) which is further discussed in Fig.6 below. A B3 can be drawn in a triangular layout with the top and bottom concepts in the middle. A B4 without top and bottom concept can be drawn as a square with triangles for its B3 sublattices.

The layout of a geometrical diagram is not the same as for an Euler diagram because, in the former, higher and lower concepts can both be placed between other concepts whereas in the latter higher concepts are placed outside of and lower concepts between other concepts. As a result, geometrical diagrams are better than Euler diagrams at showing which concepts are neighbours whereas such information can be lost for some zones in an Euler diagram. But both types of diagrams suffer from the challenges caused by forcing complex structures into a 2-dimensional representation. The top and bottom concepts of the Forum Romanum in Fig.3 are missing because otherwise P*, M*, G* and B* would need to be connected to it leading to more line crossings if the top concept is placed in the middle or curved lines if the top concept is placed on the outside. In summary, geometrical diagrams display interesting symmetries and patterns but also have drawing limitations.

If a lattice is not drawable as an Euler diagram, nesting and factorisation provide means for compacting or splitting a diagram. Figure 4 displays different versions of a B7 where an inner B3 with split attributes is *nested* into an outer B4. In the tabular diagrams, circles are used for the inner B3 for better visual differentiation. Each attribute of the inner lattice is split 16 times. The tabular diagrams provide significantly more space for adding objects than the Hasse diagram. Ideally, the selection of which attributes to use in the inner and outer lattice and their left-right, top-down sequence should be based on some semantic difference or to highlight patterns, for example, if colouring is used for object frequencies.

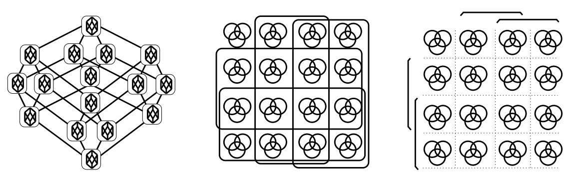

Fig. 4. A Boolean lattice with 7 attributes as a nested diagram and two types of Euler diagrams

A *factorisation* of a concept lattice is a selection of factor concepts that carry sufficient information to construct the lattice because the union of the factor concept matrices equals the matrix of the lattice (and so does a product of 2 matrices which have the factors as columns/rows (Kim 1982)). Determining a minimal factorisation is an NP-complete problem (Belohlavek & Vychodil 2010). Figure 5 shows an example[3] of a minimal factorisation of B6 as a union of 4 factor concept matrices.

An *AOC-poset* is a subset of a lattice consisting of attribute and object concepts and their connections (Osswald & Petersen 2002). An AOC-poset contains sufficient information to reconstruct a lattice but usually has many edge crossing, which are difficult to visually discern. We are suggesting to create an *AFOC-poset* instead by including factors in addition to attribute and object concepts. The factors chosen can be those of a minimal factorisation, such as the blue concepts in the middle layer in Fig.5. Or all factors can be selected that cover a given threshold number of objects and attributes under the condition that the union of the factor concept matrices equals the lattice matrix.

[3] Thanks to Dmitry I. Ignatov for sending me the factorisation of this example. His algorithm (Ignatov & Yakovleva 2021) provides a factorisation, however, not always minimal.

$$\begin{matrix}111110\\111101\\110111\\111011\\101111\\011111\end{matrix} = \begin{matrix}110100\\110100\\110100\\000000\\000000\\000000\end{matrix} \cup \begin{matrix}101010\\000000\\000000\\101010\\101010\\000000\end{matrix} \cup \begin{matrix}000000\\011001\\000000\\011001\\000000\\011001\end{matrix} \cup \begin{matrix}000000\\000000\\000111\\000000\\000111\\000111\end{matrix}$$

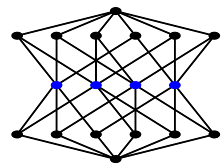

Fig. 5. Factorisation of a B6 and its AFOC-poset

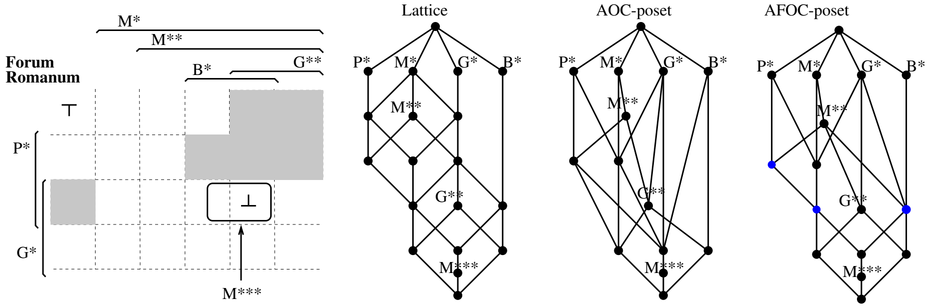

Fig. 6. Forum Romanum as a tabular diagram, lattice, AOC- and AFOC-poset

In the AFOC-poset in Fig.6 each of the 3 factors covers at least 3 objects and 3 attributes. Figures 5 and 6 demonstrate that an AFOC-poset constitutes a compromise between a lattice and an AOC-poset because it contains fewer elements than a lattice and fewer edge crossing than an AOC-poset thus reducing the visual complexity while still maintaining sufficient information about the object-attribute relationships. While Hasse diagrams of lattices might contain more potentially visually appealing parallel edges, it can be questioned in how far such patterns are meaningful for an application. AFOC-posets may be less ‘pretty’ but will be more concise.

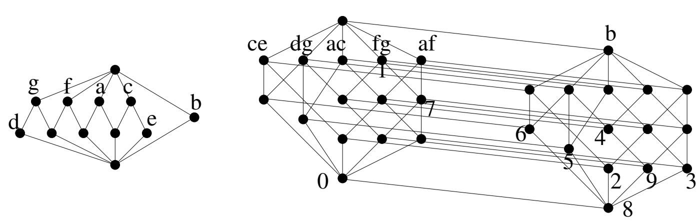

Fig. 7. Factorisation of the Digits example

For discussing factorisations Ganter & Wille (1999) utilise the Digits example, which is divided in Fig.7 into the concepts above and below the factors, because its lattice is quite complex. The bottom half is probably best presented by separating the concepts with and without attribute ‘b’ as shown in Fig.7 as a Hasse diagram and in Fig.8 as an Euler diagram. While we would argue that the Hasse and Euler diagrams in

Figs. 7 and 8 are both well readable, the AFOC-poset (of top and bottom combined) is probably more informative for most application purposes and more concise.

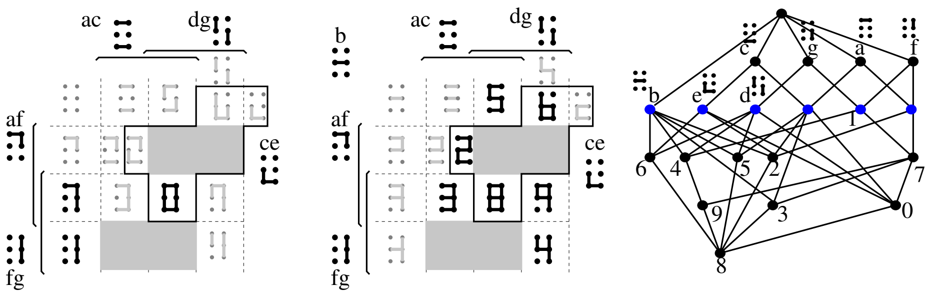

Fig. 8. Digits: bottom half as a 2-part Euler diagram and complete as an AFOC-poset

4 Object Frequency and Repeated Attributes

	a	a, b	b		
	0	20	50	40	110
c	20	40	10	20	90
	20	60	60	60	200

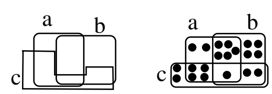

Fig. 9. Contingency table, area-proportional diagram and APT diagram

In some applications only object frequencies are of relevance. Euler diagrams are called *area-proportional* if zone sizes correspond to object frequencies as shown in the middle of Fig.9 although not exactly drawn to scale. The contingency table on the left of Fig.9 shows the object frequencies but with ordinal row and column headers. Unfortunately, area-proportionality provides another constraint on drawability. Therefore one of the shapes in the middle diagram is not a rectangle. Furthermore, area-proportionality requires visual size comparisons, which can be misleading if the shapes are different (Neurath 1936). Therefore, we are introducing *axis-proportional tabular diagrams with icons (APT diagrams)* as on the right in Fig.9. so that the width of the columns and the height of the rows are proportional to the frequencies along the two axes. Each icon (here: circle) represents a certain number of objects (here: 10) so that object frequencies can be compared. If the distribution of the objects is statistically independent then cell percentages are a product of the row and column percentages and an APT diagram is area-proportional. Otherwise it is easy to see which combinations are lower or higher than expected by evaluating how empty or crowded the zones are. APT diagrams are not an entirely new invention because they are related to Mosaic plots as explained below and somewhat similar displays have also been utilised in existing FCA software.

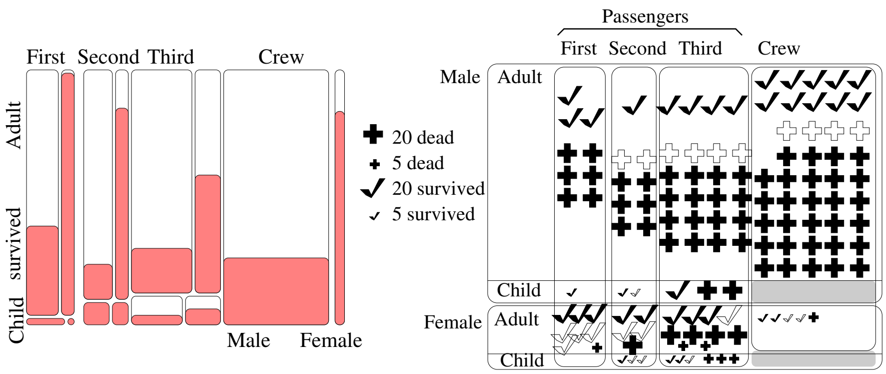

Fig. 10. Survival on the Titanic as a mosaic plot (left) and as an APT diagram (right)

Mosaic plots as invented by Hartigan and Kleiner (1981) and extended by Friendly (1994) for visualising contingency tables are essentially area-proportional Euler diagrams with split attributes. The left side of Fig.10 displays a mosaic plot of the well-known Titanic data for the many-valued attributes class, gender, age and survival. The diagram shows for example the better chance of survival for women and children. But it is basically impossible to compare percentages across attributes because of the different shapes. The APT diagram on the right side of Fig.10 with four split attributes (adult, child, dead and survived) displays the same data but with more information. Icons for survival and death have two sizes (for groups of 20 and groups of 5) and a white colour for higher counts than expected. Only the APT diagram shows clearly that the survival rate of first class male adults and third class male children corresponds exactly to the 30% survival rate of all passengers. Males had higher survival chances if they were crew members than second or third class passengers. Zones for the adult female passengers are crowded because of small numbers of female crew and zones for crew children are shaded because there were none.

Apart from objects, attributes can also occur with higher frequencies in some applications. Hasse diagrams of concept lattices are sometimes drawn to contain parallel edges, which might improve readability but carry no further meaning. In some applications attributes can apply more than once to an object in which case all parallel edges can represent a single attribute. We are calling such diagrams *Repeated Attribute Lattice diagrams (RAL diagram)* and each attribute a *vector*. The far left of Fig.11 displays a RAL diagram of numbers and divisors where one vector represents ‘multiply 2’ and the other ‘multiply 3’. The length of a vector does not carry any meaning but vector addition corresponds to multiplication, such as $2 \cdot 2 \cdot 3 = 12$. Unfortunately, only small parts of RAL diagrams can be drawn for more than 2 vectors because eventually crossing will occur that are not valid multiples. Espinosa-Aldama and Mendoza (2024) construct lattices for physical units without considering how often units are used. We are suggesting to use RAL diagrams instead. Figure 11 displays parts of a RAL diagram for physical units that simultaneously show valid products, such as $F \cdot \Omega = s$, $\Omega \cdot A = V$, $A \cdot s = C$, $V \cdot s = Wb$, $A \cdot V \cdot s = J$ and many more.

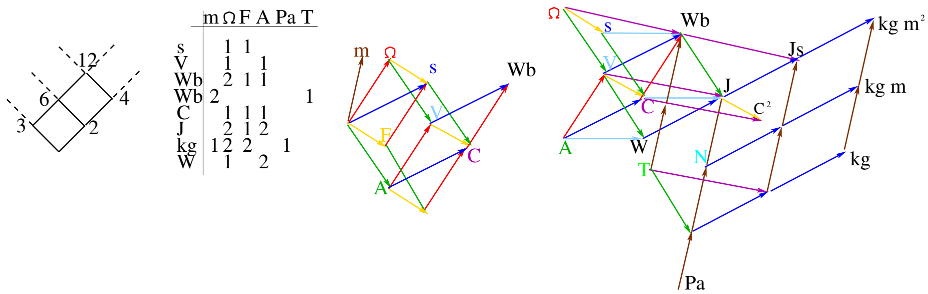

Fig. 11. RAL diagrams: multiples, matrix and parts of a RAL diagram for physical units

5 Relation Algebraic Concept Analysis

If a concept lattice is combined with a relation, many additional concepts can be defined in an ad-hoc manner. For example, attributes ‘male’ and ‘female’ together with a parent-child relationship yield concepts such as ‘male parent’ and ‘parent of at least one female child’. Relational concepts are either lexicalised (‘father’, ’daughter’) or can be described ad hoc, such as ‘has parent who has daughter’, which might be renamed to ‘has or is sister’. Together with quantifiers this leads to many possibilities. Extensions of FCA such as RCA (Dolques, Le Ber & Huchard 2013) and GFCA (Ferre & Cellier 2019) exist, which study the formation of relational concepts. They were preceded by what used to be called RCA (Priss 1999) but is renamed into Relation Algebraic Concept Analysis (RaCA) in order to avoid confusion with the unrelated later RCA. While RCA and GFCA compute all possibilities for some relation and develop algorithms, the focus of RaCA is on deriving diagrams in a flexible manner using standard Relation Algebra and matrix operations (Priss 2006). In this paper RaCA is demonstrated using an example of some members of the British royal family (Fig.12) originally from GFCA (Ferre & Cellier 2019).

Table P:

name	gender	birth
Charles	male	1948
Diana	female	1961
William	male	1982
Harry	male	1984
Kate	female	1982
George	male	2013
Charlotte	female	2015
Louis	male	2018

Table PC:

parent	child
Charles	William
Charles	Harry
Diana	William
Diana	Harry
William	George
William	Charlotte
William	Louis
Kate	George
Kate	Charlotte
Kate	Louis

I	m	f
Charles	1	0
Diana	0	1
William	1	0
Kate	0	1

J	m	f
William	1	0
Harry	1	0
George	1	0
Charlotte	0	1
Louis	1	0

R	W	H	G	C	L
C	1	1	0	0	0
D	1	1	0	0	0
W	0	0	1	1	1
K	0	0	1	1	1

$I \cup (R \circ J)$

	m	f	m	f
C	1	0	1	0
D	0	1	1	0
W	1	0	1	1
K	0	1	1	1

Scale:
< 1950, < 1980
< 2000, ≥ 2000

S				
1948	1	1	1	0
1961	0	1	1	0
1982	0	0	1	0
1984	0	0	1	0
2013	0	0	0	1
2015	0	0	0	1
2018	0	0	0	1

Fig. 12. A parent-child relationship amongst some members of the British royal family

The data of the example is contained in the tables P and PC in Fig.12, which are modelled as binary matrices I and J, R and a scale S for 4 different ranges of the time of birth. The left side of Fig.13 displays a modelling of the primary data as

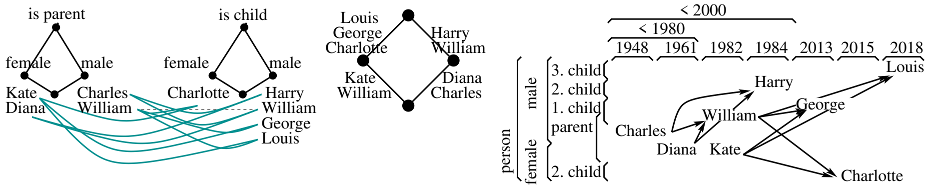

Fig. 13. A relation amongst objects in lattices and an Euler diagram

lattices where the lines that connect parents and children are difficult to visually trace. The dashed line is meant to indicate that 'William' refers to the same person. The right side of Fig.13 shows the relationships amongst individual objects with arrows, which are easier to observe. The attributes '1. child', '2. child', '3. child' are computed from background information for temporal data but can also be visually inferred from the diagram because the two generations, married couples and birth sequence form visual patterns. The attribute '2. child' is split.

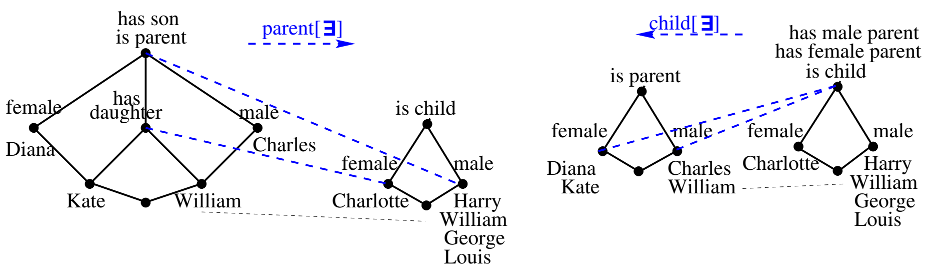

Fig. 14. RaCA diagrams: conceptual parent-child[∃] relation

In RaCA a relation amongst objects leads to relations amongst concepts based on quantifiers (Priss 1999). While tabular Euler diagrams may be more suitable for relations amongst objects, Hasse diagrams are more suited for relations amongst concepts. The data for the following diagrams is produced by binary matrix multiplication with complementation depending on the quantifiers as explained by Priss (2006). For example, in order to add the attributes of the children to the lattice of the parents the operation $I \cup (R \circ J)$ is performed on the matrices from Fig.12 with a disjoint union so that attributes get renamed ('female' for children becomes 'has female child' for adults) resulting in the lattices in Fig.14. The corresponding matrix operation for the children is $(I \circ R)^d \cup J$ where ^{d} indicates the dual matrix. Priss (1999) suggests that a relation amongst concepts can be visualised via a basis where ∀-quantifiers are inherited by subconcepts and ∃-quantifiers by superconcepts so that ∀-quantifiers only need to be shown for the largest concepts and ∃-quantifiers only for the smallest concepts. In Fig.14 *all* parents have *at least 1* son but only Kate and William have *at least 1* daughter. *All* children have *at least 1* male and female parent. Obviously, as soon as more attributes are added to the lattice of the parents these lead to more attributes of the children resulting in $((I \cup (R \circ J))^d \circ R)^d \cup J$ and so on. Repeated multiplication of a binary matrix will reach a fixed state once the product does not change anymore or has a period (Kim 1982). Together with a union this will lead to the lattices stabilising after a few iterations. The scale S of Fig.12 is only displayed in the Euler diagram in Fig.13 but applying a scale is also just a matter of matrix multiplication.

In conclusion, this paper presents a variety of diagrams for lattices with or without additional structures such as relations or repeated objects and attributes. It is argued that Euler diagrams are easier to read for untrained users and are suitable for nesting and some RaCA diagrams but without nesting only feasible for fairly small lattices which are mostly tree-like or crown-like. A certain type of Euler diagram called APT diagram is presented as an improvement over Mosaic plots. Euler diagrams provide more space for representing objects whereas Hasse diagrams are more suitable for some applications where individual objects are less relevant. Factorisation leads to a type of Hasse diagram called AFOC-poset, which is a compromise between a lattice and its minimal representation as an AOC-poset. RaCA diagrams are Hasse diagrams with additional relations amongst lattices and RAL diagrams are Hasse diagrams with repeated attributes. Last but not least, geometrical diagrams are neither Hasse, nor Euler diagrams but a type of graph and are interesting but most likely of limited practical use.

References

1. Belohlavek, R.; Vychodil, V. (2010). Discovery of optimal factors in binary data via a novel method of matrix decomposition. J. of Computer and System Sciences 76.1, p. 3-20.
2. Dolques, X.; Le Ber, F.; Huchard, M. (2013). AOC-posets: a scalable alternative to Concept Lattices for Relational Concept Analysis. CLA: Concept Lattices and their Applications.
3. Espinosa-Aldama, M., & Mendoza, S. (2024). Multidimensional representation of semantic relations between physical theories, fundamental constants and units of measurement with formal concept analysis. Symmetry 16.7.
4. Ferre, S.; Cellier, P. (2019). Graph-FCA: An extension of formal concept analysis to knowledge graphs. Discrete Applied Mathematics 273, p. 81–102
5. Friendly, M. (1994). Mosaic displays for multi-way contingency tables. J. of Amer. Stat. Assoc., 89, 425, p. 190-200.
6. Ganter, B.; Wille, R. (1999). Formal Concept Analysis. Mathematical Foundations. Springer.
7. Ganter, B. (2013). Diskrete Mathematik: Geordnete Mengen. Springer.
8. Hartigan, J. A.; Kleiner, B. (1981). Mosaics for Contingency Tables. Computer Science and Statistics: Proc. of the 13th Symposium on the Interface. Springer, p. 268-273.
9. Ignatov, D. I.; Yakovleva, A. (2021). On Suboptimality of GreConD for Boolean Matrix Factorisation of Contranominal Scales. FCA4AI@IJCAI 2021, p. 87-98.
10. Kim, K. H. (1982). Boolean Matrix Theory and Applications. Marcel Dekker, New York.
11. Neurath, O. (1936). International picture language. London: Kegan Paul.
12. Osswald, R.; Petersen, W. (2002). Induction of classifications from linguistic data. In: Proc. of the ECAI-Workshop on Advances in FCA for KDD.
13. Priss, U. (1999). Efficient Implementation of Semantic Relations in Lexical Databases. Computational Intelligence, Vol. 15, 1, p. 79-87.
14. Priss, U. (2006). An FCA interpretation of Relation Algebra. Formal Concept Analysis: 4th International Conference, ICFCA, Springer, LNCS 3874, p. 248-263.
15. Priss, U., Dürrschnabel, D. (2024). Rectangular Euler Diagrams and Order Theory In: Proceedings of Diagrams 2024, Springer, LNAI 14981, 2024, p. 165-181.