


# Simplifying Contextual Structures

Ivo Düntsch<sup>1</sup> and Günther Gediga<sup>2</sup>

<sup>1</sup> Brock University, St. Catharines, ON L2S 3A1, Canada

[duentsch@brocku.ca](mailto:duentsch@brocku.ca)

<sup>2</sup> Department of Psychology, Institut IV, Universität Münster, Fliednerstr. 21,  
Münster, Germany

[gediga@uni-muenster.de](mailto:gediga@uni-muenster.de)

**Abstract.** We present a method to simplify a formal context while retaining much of its information content. Although simple, our ICRA approach offers an effective way to reduce the complexity of a concept lattice and/or a knowledge space by changing only little information in comparison to a competing model which uses fuzzy K-Means clustering.

## 1 Introduction

A very simple data structure is a triple  $\mathfrak{C} = \langle U, V, R \rangle$  where  $R$  is a binary relation between elements of  $U$  and elements of  $V$  which is sometimes called a *formal context* [6, 19]. From this, various data models can be obtained, one of the more popular ones being the *concept lattice* obtained from  $\mathfrak{C}$  introduced by Wille [19]. With each concept a line diagram can be associated which depicts the concept lattice in a consolidated way. For lack of space we shall not describe this further; for details we invite the reader to consult, for example, [20] or [6].

As a context  $\mathfrak{C}$  grows large, the construction of the concept lattice is costly and it is difficult to interpret the structure and its associated line diagram. Therefore, various techniques have been proposed to simplify a formal context  $\mathfrak{C}$  or its associated concept lattice such as stability indices [1, 11, 14, 15] which only consider only part of the concept lattice, simplification using fuzzy K-Means clustering (FKM) [13] or object similarity [2], or selection of relevant concepts in the presence of noisy data [11]. All these techniques can be subsumed under one of the following strategies:

1. Omit attributes (or objects), or
2. Merge attributes (or objects) which are similar according to some criterion, or
3. Remove concepts with low index values.

In each case, the adjacency matrix of  $R$  is changed. However, reducing the matrix does not guarantee that the associated concept lattice will be reduced as well, see Example 3 of [12]. In this paper we propose a simple algorithm to simplify a concept which does not increase the size of its associated concept lattice.

---

The ordering of authors is alphabetical and equal authorship is implied.

The author gratefully acknowledges support by the Natural Sciences and Engineering Research Council of Canada.

## 2 Notation and Definitions

Throughout we suppose that  $U = \{p_1, \dots, p_n\}$  is a finite set of objects (such as problems) and  $V = \{s_1, \dots, s_k\}$  is a finite set of attributes (such as skills).  $R \subseteq U \times V$  is a binary relation between elements of  $U$  and elements of  $V$ . For each  $p \in U$  we set  $R(u) \stackrel{\text{df}}{=} \{s \in V : pRs\}$ , and  $\mathcal{R} \stackrel{\text{df}}{=} \{R(u) : u \in U\}$ . The identity relation on  $U$  is denoted by  $1'_U$ . The relational converse of  $R$  is denoted by  $R^\smile$ , and  $-R$  is the complement of  $R$  in  $U \times V$ . The set  $\mathcal{R}$  is partially ordered by  $\subseteq$ . The *adjacency matrix of  $R$*  has rows labeled by the elements of  $U$ , and columns labeled with the elements of  $V$ . An entry  $\langle u, v \rangle$  is 1 if and only if  $u_i R s_j$ , otherwise, the entry in this cell is left empty. A formal context  $\langle U, V, R \rangle$  gives rise to several set operators frequently used in modal logics: Let  $X, X' \subseteq U$  and define

$$\langle R \rangle(X) = \{b \in V : (\exists a \in X) aRb\} = \{b \in B : R^\smile(b) \cap X \neq \emptyset\}, \quad (\text{Possibility})$$

$$[R](X) = \{b \in V : (\forall a \in U)[aRb \Rightarrow a \in X]\} = \{b \in B : R^\smile(b) \subseteq X\}, \quad (\text{Necessity})$$

$$[[R]](X) = \{b \in V : (\forall a \in U)[a \in X \Rightarrow aRb]\} = \{b \in B : X \subseteq R^\smile(b)\}. \quad (\text{Sufficiency})$$

The mappings  $\langle R \rangle$  and  $[[R]]$  are, respectively, the existential (disjunctive) and universal (conjunctive) extensions of the assignment  $x \mapsto R(x)$  to subsets of  $U$ , since it follows immediately from the definitions that for all  $x \in U, X \subseteq U$ ,

$$\langle R \rangle(\{x\}) = [[R]](\{x\}) = R(x), \quad (1)$$

$$\langle R \rangle(X) = \bigcup_{x \in X} R(x), \quad [[R]](X) = \bigcap_{x \in X} R(x). \quad (2)$$

The operators  $[[R]]$  and  $[R]$ , as well as  $\langle R \rangle$ , are related since

$$b \in [[R]](X) \iff X \subseteq R^\smile(b) \iff U - R^\smile(b) \subseteq U \setminus X \iff b \in [-R](U \setminus X). \quad (3)$$

For unexplained notation and concepts in lattice theory we refer the reader to [8].

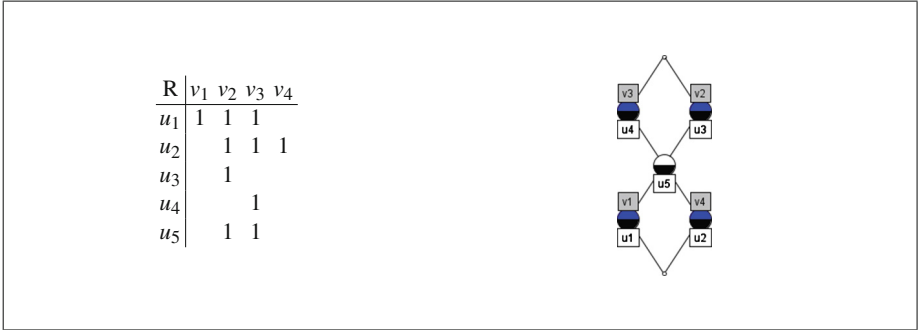
## 3 Data Models Based on Modal Operators

Suppose we have a formal context  $\mathfrak{C} = \langle U, V, R \rangle$  which we regard as “raw data”. The image sets  $R(x)$  are our basic constructs. As a first approach to a data model based on  $\langle U, V, R \rangle$ , which, in our view, is a structural representation of raw data, we define a quasiorder  $\preceq$  on  $U$  by setting  $x \preceq y$  if and only if  $R(x) \subseteq R(y)$ . We also define the *incomparability relation* by

$$x \# y \stackrel{\text{df}}{\iff} (x \not\preceq y) \text{ and } (y \not\preceq x). \quad (4)$$

From this starting point, several more involved data models can be developed. One of the better known models are those based on the sufficiency operators  $[[R]]$  (“intent”) and  $[[R^\smile]]$  (“extent”): For each  $X \subseteq U$ ,  $[[R]](X)$  is the set of all

attributes common to all elements of  $X$ , and for  $Y \subseteq V$ ,  $[[R^\vee]](Y)$  is the set of all objects which possess all attributes in  $Y$ . A pair  $\langle [[R^\vee]]([[R]](X)), [[R]](X) \rangle$  is called a *formal concept*. The set of all formal concepts can be made into a lattice which can be drawn as a consolidated line diagram [19] as in Fig. 1<sup>1</sup>. Each node of the diagram represents a formal concept, and for each object  $x$ ,  $R(x)$  is the set of all attributes above the node labelled  $x$  (we interpret “above” and “below” as reflexive relations). In the line diagram of  $R$ ,  $x \preceq y$  if and only if  $x$  and  $y$  label the same node or the node labelled by  $y$  is below the node labelled by  $x$ .



**Fig. 1.** A context and its line diagram

A data model which in some sense competes with concept lattices are the *knowledge spaces* introduced in [4]. These are set systems closed under union and can be related to the modal operator  $\langle R \rangle$  which is called the *span operator* in [3]. It was shown in [7] that the models arising from  $[[R]]$  and  $\langle R \rangle$  have the same expressive power and are useful in situations different from those where conjunctive assignments such as the (DINA) model [9, 10, 16] and the rule space model [18] are employed.

Taking  $\{R(x) : x \in U\}$  as a starting point, the set of spans and the set of intent go into different directions: It follows from (1) and (2) that  $\mathcal{K}_R \stackrel{\text{df}}{=} \{\langle R \rangle(X) : X \subseteq U\}$  is the  $\cup$ -semilattice generated by  $\{R(x) : x \in U\}$ , and  $\mathcal{I}_R \stackrel{\text{df}}{=} \{[[R]](X) : X \subseteq U\}$  is the  $\cap$ -semilattice generated by  $\{R(x) : x \in U\}$ . For  $X \subseteq U$ ,  $[[R]]$  is the set of all attributes lying above all objects in  $X$ , and  $\langle R \rangle(\{x\})$  is the set of all attributes not upwards reachable from object  $x$  in the line diagram of  $-R$ .

## 4 Reducing the Complexity

The simplest way to change the adjacency matrix is to change one bit at a time, according to a given criterion. The question arises which criterion we shall

<sup>1</sup> The diagrams were drawn by the ConExp package [21].

use. If  $\preceq$  is a linear quasi order – i.e. if any two objects of  $U$  are comparable – then  $\mathcal{K}_R$  and  $\mathcal{S}_R$  coincide and are equal to  $\langle \mathcal{K}_R, \subseteq \rangle$  (possibly with added  $\emptyset$  or  $V$ ); nothing is gained by going from the simple model  $\langle |C, \preceq \rangle$  to one of the more involved ones. At the other extreme, if no two different elements of  $U$  are comparable with respect to  $\#$ , then the representations obtained from  $\mathfrak{C}$  very strongly depend on the modal operator used and may widely differ. Consider the simple relation depicted in Fig. 2. There,  $\mathcal{S}_R$  consists of the singletons  $\{v_i\}$  and the empty set, while  $\mathcal{K}_R$  is the set of all nonempty subsets of  $V$ . If we consider the complement of  $-R$ , then situation is reversed, see Fig. 3.

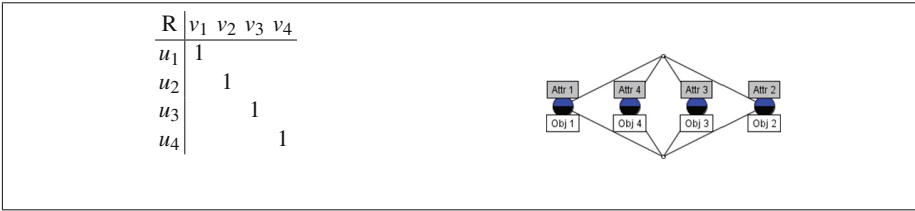


Fig. 2.  $\# = U^2 \setminus 1'_U$ , 1st example

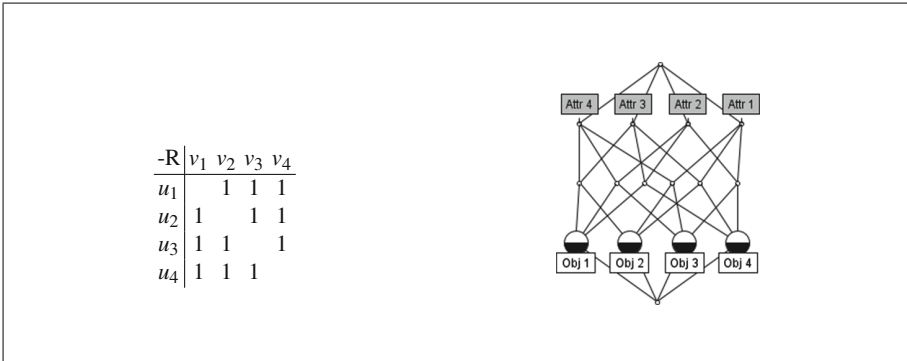


Fig. 3.  $\# = U^2 \setminus 1'_U$ , 2nd example

Therefore, if the incomparability relation is large, choosing one operator over the other may not provide a meaningful interpretation, and it may not be the wisest choice at the outset to prefer one over the other. Keeping in mind the problem/skill situation, we suggest the *relative incomparability* of objects as a measure of context complexity which we aim to reduce: If  $\mathfrak{C} = \langle U, V, R \rangle$  is a formal context and  $u \in U$ , then we let

$$\text{incomp}(u) \stackrel{\text{df}}{=} \{v \in U : u \# v\}, \quad \text{incomp}(\mathfrak{C}) \stackrel{\text{df}}{=} \frac{|\{(u, v) : u \# v\}|}{n^2 - n},$$

where  $n = |U|$ . Now,  $\text{incomp}(\mathfrak{C}) = 0$  if and only if  $\preceq$  is a linear quasiorder, and  $\text{incomp}(\mathfrak{C}) = 1$  if no two different elements are  $\preceq$ -comparable. The measure of success is the reduction of  $\text{incomp}(\mathfrak{C})$  relative to the number of bit changes.

Our *InComparability Reduction Analysis* algorithm (ICRA)<sup>2</sup> is based on a simple steepest descent method: We consider objects  $u$  for which  $|\text{incomp}(u)|$  is maximal and then invert a bit – i.e. an entry in the adjacency matrix of the relation under consideration – for which the drop of the number of overall incomparable pairs is maximal. This will increase the comparability of objects with respect to  $\preceq$  or, equivalently, of sets  $R(x)$  without increasing the number of intents, respectively, knowledge states. Indeed, in most cases we have looked at, the complexity of the concept lattice was significantly reduced. If one bit is inverted, so that the resulting relation is  $R'$  and  $x \preceq_{R'} y$ , then there will be a path from  $y$  to  $x$  in the line diagram of  $R'$  as well, so that the new representation is closer to the data as represented by  $R$ .

The basic concept is that we assume some of the data to be faulty, but we do not know which entries. More concretely, we assume that some (or all) incomparabilities are caused by faulty data. In this sense, our proposed procedure is a trade – off measure.

The stop criterion is a predetermined relative value of incomparable pairs, i.e. a value for  $\text{incomp}(\mathfrak{C})$ , where  $\mathfrak{C}$  is the current context, or no more complexity reduction is possible. As a rule of thumb we suggest to require that 50% of

```

noEntry:=FALSE.
pout:=p                                ▷ Initialize stop criterion to  $0 \leq p \leq 1$ .
Unmark all object-attribute-pairs.      ▷ No pair changed yet.
repeat
  Find the set OBJ of objects belonging to unmarked pairs for which  $\text{incomp}(u)$  is maximal.
  if  $\text{incomp}(\mathfrak{C}) \leq p$  then                                ▷ Goalreached
    NoEntry:=TRUE
  else
    Using OBJ find the object-attribute-pairs, which maximally reduce their comparability,
    when inverting one bit of the matrix under consideration.
    if no reduction is achieved for any of these then
      NoEntry:=TRUE
    else
      Invert the entry of one of the maximal object-attribute-pairs and use the new rela-
tion.
      Mark the chosen object-attribute-pair.
      Replace  $\mathfrak{C}$  with there vised context.
    end if
  end if
until NoEntry=TRUE.

```

**Fig. 4.** Pseudocode of the algorithm

<sup>2</sup> The algorithm is implemented in R [17] and the source code is available at <http://roughsets.net/FCred.R>.

pairs with different components should be comparable (*Median InComparability Reduction Analysis*). An overview of the pseudocode the ICRA algorithm is shown in Fig. 4.

### 5 Experiments

Even though our procedure is simple, it compares well with other simplification measures. As a case in point we shall consider the reduction using fuzzy K-Means clustering (FKM) proposed in [13]. This method is based on partitioning a set of vectors into  $k$  fuzzy clusters, specifying to what degree a vector belongs to the cluster centre. Owing to lack of space we cannot explain their method in detail and refer the reader to [13]. The context  $\mathcal{C}$  of their first example relates documents with keywords and it is shown in Fig. 5 along with its context lattice. The relative incomparability of  $\mathcal{C}$  is 94 %.

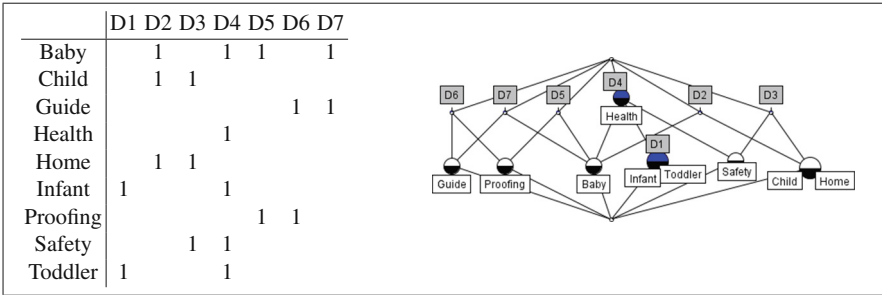


Fig. 5. Example from [13], p. 2699

After applying FKM based clustering with  $k = 2$ , the columns D1 – D2 are identified and the entry  $\langle T_i, D1' - -D4 \rangle$  of the resulting adjacency matrix is  $\max\{\langle T_i, D1 \rangle, \dots, \langle T_i, D4 \rangle\}$ . The simplified context  $\mathcal{C}_1$  and its concept lattice are shown in Fig. 6.

To achieve the FKM result  $\mathcal{C}_1$  from  $\mathcal{C}$  requires to change 15 bits for a relative incomparability of 49 %; this includes the effort to identify columns. In comparison, our algorithm needs only 4 bits for a 50 % incomparability, and 9 bits for 0 % incomparability. The resulting context along with its line diagram is shown in Fig. 7. It has the same number of concepts as the concept lattice obtained from FKM (9), and the same number of edges (14).

In classification tasks, there is often a trade – off between the (relative) number of correctly classified objects and, for example, the (relative) cost of obtaining the classification or the clarity of a pictorial representation. In some instances, this may be expressed as the amount of errors we are prepared to allow to achieve another aim. A case in point are curves based on receiver operating characteristics (ROC), where the sensitivity (benefit) of a binary classifier is

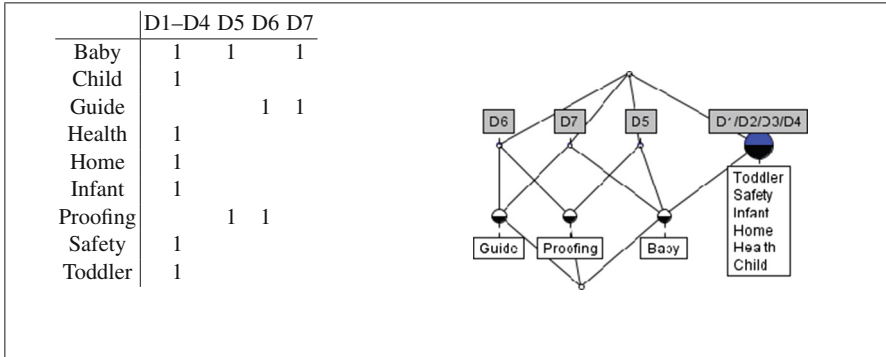


Fig. 6. Example from [13], p. 2699, reduced

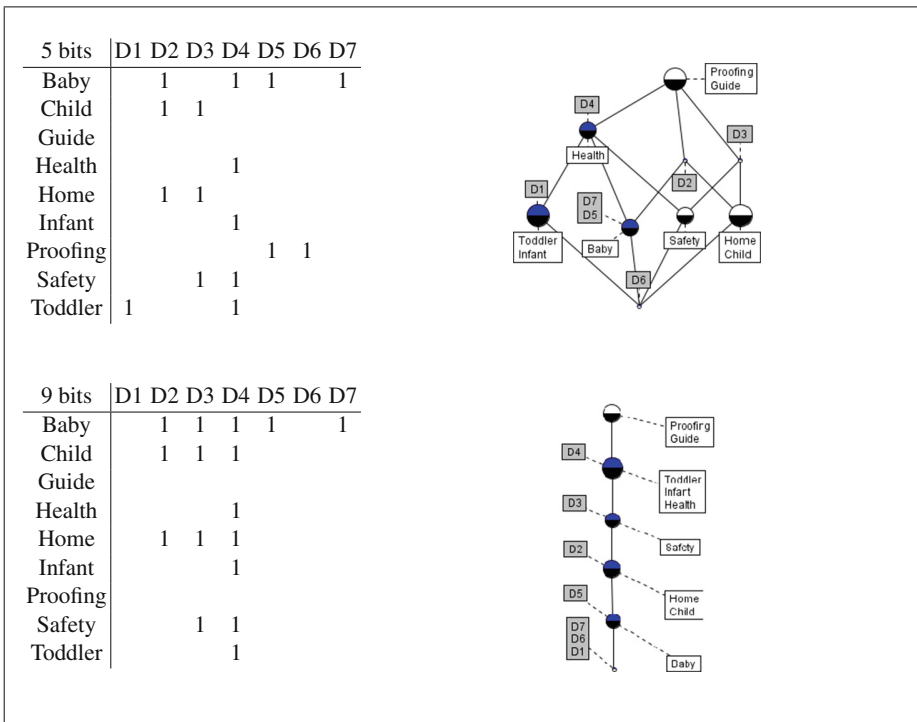
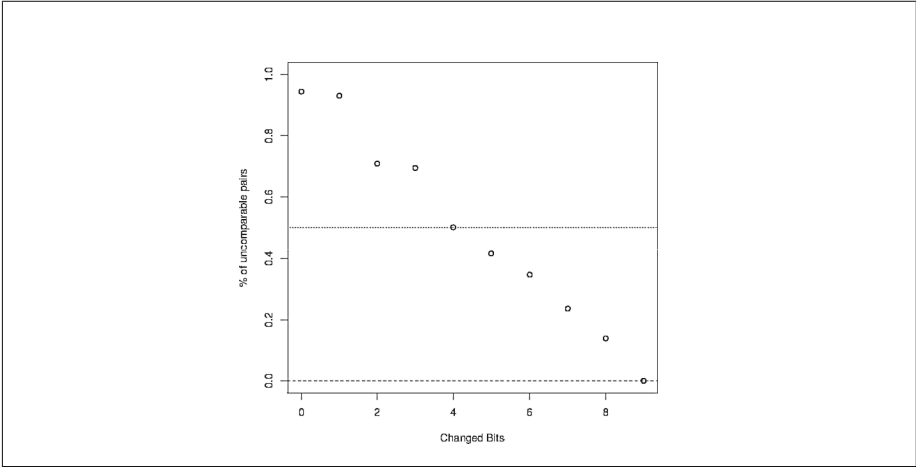


Fig. 7. Reduction of Example 1 from [13] using ICRA

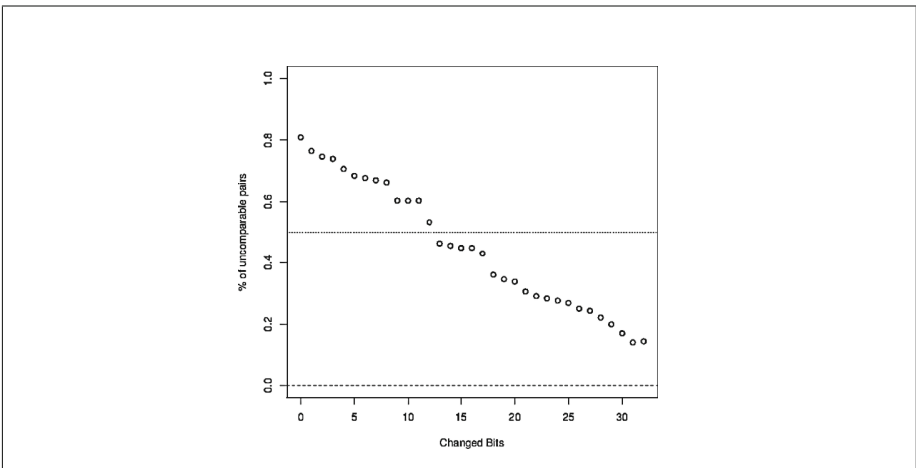
plotted as a function of its FP rate (cost), see [5] for an overview. We can plot the relative incomparability as a function of the number of bits changed to achieve it, see the graph in Fig. 8. If we interpret (in-)comparability as sensitivity and the number of changed bits as cost to retrieve the original data, this can be interpreted as a ROC curve.



**Fig. 8.** Reducing relative incomparability with ICRA

The next example for [13] investigates a dataset consisting of various species of bacteria and 16 phenotypic characters, shown in Table 1.

For this context  $\mathcal{C}$ , the incomparability  $\text{incomp}(\mathcal{C})$  turns out to be 81%.  $\mathcal{C}$  is reduced with the FKM method for  $k = 5$  and  $k = 9$ , resulting in contexts  $\mathcal{C}_5$  and  $\mathcal{C}_9$  with  $\text{incomp}(\mathcal{C}_5) = 34.5\%$  and  $\text{incomp}(\mathcal{C}_9) = 64.7\%$ . 40 bits are required to reduce  $\mathcal{C}$  to  $\mathcal{C}_5$ , and the reduction to  $\mathcal{C}_9$  with 64.7% incomparability needs changing 11 bits. In contrast, our algorithm requires changing 19 bits to achieve an incomparability reduction to 34.6%, and 8 bits for a reduction to 66.1%. Changing 11 bits (as in the FKM reduction with  $k = 9$ ) results in a reduction to 60.2%. The ICRA reducibility graph is shown in Fig. 9.



**Fig. 9.** Reducing relative incomparability of the bacterial dataset with ICRA



**Table 1.** Bacterial dataset from [13]

	H2S	MAN	LYS	IND	ORN	CIT	URE	ONP	VPT	INO	LIP	PHE	MAL	ADO	ARA	RHA
ecoli1	0	1	1	1	0	0	0	1	0	0	0	0	0	0	1	1
ecoli2	0	1	0	1	1	0	0	1	0	0	0	0	0	0	1	0
ecoli3	1	1	0	1	1	0	0	1	0	0	0	0	0	0	1	1
stypHi1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0
stypHi2	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
stypHi3	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0
kpneu1	0	1	1	1	0	1	1	1	1	1	0	0	0	1	1	1
kpneu2	0	1	1	1	0	1	1	1	1	1	0	0	1	0	1	1
kpneu3	0	1	1	1	0	1	1	1	1	1	0	0	1	1	1	1
kpneu4	0	1	1	1	0	1	1	1	0	1	0	0	1	1	1	1
kpneu5	0	1	1	1	0	1	0	1	1	1	0	0	1	1	1	1
pvul1	1	0	0	1	0	1	1	0	0	0	0	1	0	0	0	0
pvul2	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0
pvul3	1	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0
pmor1	0	0	1	1	1	0	1	0	0	0	0	1	0	0	0	0
pmor2	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0
smar	0	1	1	0	1	1	0	1	1	0	1	0	0	0	0	0

## 6 Conclusion and Outlook

We have introduced a simple algorithm ICRA to simplify a formal context, the success criterion of which is a prescribed reduction of incomparable pairs. As a rule of thumb, we propose a relative frequency of incomparable pairs of objects of 50%. This seems a fair compromise between closeness to the data on the one hand, and the additional structure introduced by the chosen model on the other. We have compared the success of our algorithm with several examples of [13] and have found that fewer bits are needed than FKM to obtain similar incomparability ratios. Furthermore, the FKM algorithm requires much more effort and additional model assumptions so that its cost/benefit ratio is much smaller than for the median comparability algorithm. Furthermore, it is not clear which  $k$  should be used for the reduction.

In the available space, only an indication of the impact of the median comparability algorithm could be given. Further work will include investigation of the powers and limitations of the ICRA algorithm using both theoretical and practical analysis. In particular, we shall consider its effects on implication sets and association rules.

**Acknowledgement.** We thank the referees for careful reading and constructive comments.

## References

1. Buzmakov, A., Kuznetsov, S.O., Napoli, A.: Scalable estimates of concept stability. In: Glodeanu, C.V., Kaytoue, M., Sacarea, C. (eds.) ICFCA 2014. LNCS, vol. 8478, pp. 157–172. Springer, Heidelberg (2014)
2. Dias, S.M., Vieira, N.J.: Reducing the size of concept lattices: the JBOS approach. In: Proceedings CLA, pp. 80–91 (2010)

3. Düntsch, I., Gediga, G.: Approximation operators in qualitative data analysis. In: de Swart, H., Orłowska, E., Schmidt, G., Roubens, M. (eds.) *Theory and Applications of Relational Structures as Knowledge Instruments*. LNCS, vol. 2929, pp. 214–230. Springer, Heidelberg (2003)
4. Falmagne, J.C., Koppen, M., Villano, M., Doignon, J.P., Johannesen, J.: Introduction to knowledge spaces: how to build, test and search them. *Psychol. Rev.* **97**(2), 201–224 (1990)
5. Fawcett, T.: An introduction to ROC analysis. *Pattern Recog. Lett.* **27**, 861–874 (2006)
6. Ganter, B., Wille, R.: *Formal Concept Analysis: Mathematical Foundations*. Springer, Berlin (1999)
7. Gediga, G., Düntsch, I.: Skill set analysis in knowledge structures. *Br. J. Math. Stat. Psychol.* **55**, 361–384 (2002). <http://www.cosc.brocku.ca/duentsch/archive/skills2.pdf>
8. Grätzer, G.: *General Lattice Theory*, 2nd edn. Birkhäuser, Basel (2000)
9. Haertel, E.H.: Using restricted latent class models to map the skill structure of achievement items. *J. Educ. Meas.* **26**, 301–324 (1989)
10. Junker, B.W., Sijtsma, K.: Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Appl. Psychol. Meas.* **25**, 258–272 (2001)
11. Klimushkin, M., Obiedkov, S., Roth, C.: Approaches to the selection of relevant concepts in the case of noisy data. In: Kwuida, L., Sertkaya, B. (eds.) *ICFCA 2010*. LNCS, vol. 5986, pp. 255–266. Springer, Heidelberg (2010)
12. Krupka, M.: On complexity reduction of concept lattices: three counterexamples. *Inf. Retr.* **15**(2), 151–156 (2012). <http://dx.doi.org/10.1007/s10791-011-9175-7>
13. Kumar, C.A., Srinivas, S.B.: Concept lattice reduction using fuzzy K-means clustering. *Exp. Syst. Appl.* **37**, 2696–2704 (2010)
14. Kuznetsov, S.O., Obiedkov, S., Roth, C.: Reducing the representation complexity of lattice-based taxonomies. In: Priss, U., Polovina, S., Hill, R. (eds.) *ICCS 2007*. LNCS (LNAI), vol. 4604, pp. 241–254. Springer, Heidelberg (2007)
15. Kuznetsov, S.: On stability of a formal concept. *Ann. Math. Artif. Intel.* **49**(1–4), 101–115 (2007). <http://dx.doi.org/10.1007/s10472-007-9053-6>
16. Macready, G.B., Dayton, C.M.: The use of probabilistic models in the assessment of mastery. *J. Edu. Stat.* **2**, 99–120 (1977)
17. R Core Team, R.: *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2014). <http://www.R-project.org/>
18. Tatsuoaka, K.K.: Rule space: an approach for dealing with misconceptions based on item response theory. *J. Edu. Meas.* **20**(4), 345–354 (1983)
19. Wille, R.: Restructuring lattice theory: an approach based on hierarchies of concepts. In: Rival, I. (ed.) *Ordered Sets*, pp. 445–470. NATO Advanced Studies Institute, Reidel, Dordrecht (1982)
20. Wolff, K.E.: A first course in formal concept analysis - how to understand line diagrams. In: Faulbaum, F. (ed.) *Softstat '93: Advances in Statistical Software 4*, pp. 429–438. Stuttgart, Fischer (1993)
21. Yevtushenko, S.: The concept explorer (2000). retrieved 24 December 2011. <http://conexp.sourceforge.net/index.html>