Dense rectangles in object-attribute data

Radim Bělohlávek and Vilém Vychodil

***Abstract*— We study dense rectangles in data tables with binary attributes, i.e. subtables which are "almost full of 1's". Dense rectangles represent interesting patterns which an be thought of as particular granules in data tables. Rectangles which are "full of 1's" appear as natural patterns in several areas and have been widely studied in computer science and data analysis. Our paper presents a study in which we loosen the criterion of a density of a rectangle. Instead of rectangles full of 1's, we are interested in rectangles which may contain a few 0's. This way, one can capture different kinds of patterns in data. These patterns elude methods which extract only rectangles "full of 1's". We propose several ways to define density of a rectangle. We concentrate on column-like (and dually, row-like) conditions which say that a rectangle is dense if each of its columns contains at most a given (small) number of 0's. For this case, we develop theoretical insight resembling that one behind rectangles "full of 1's", present illustrative examples and experiments, and outline further issues and future research.**

***Index Terms*— Algorithms, Artificial intelligence, Clustering methods, Information retrieval**

I. Introduction

A. Problem setting

DATA tables with rows labeled by objects $x \in X$, columns labeled by attributes $y \in Y$, and table entries containing $\times$ or blank (alternatively, 1 or 0) depending on whether the corresponding object does or does not have the corresponding attribute, appear in many situations. Fig. 1 presents such a data table. There has been a lot of effort to extract various patterns from such data. Examples of these patterns are rectangles (subtables, rectangular areas) of the table. A rectangle can be identified with a pair $\langle A, B\rangle$ where $A \subseteq X$ is a set of its rows (objects) and $B \subseteq Y$ is a set of its columns (attributes). In particular, maximal rectangles which are full of $\times$'s (1's) belong to the most useful patters. Maximal full rectangles have been studied long ago (see e.g. [5]). As an example, $\langle\{x_4\}, \{y_2, \dots, y_9\}\rangle$, $\langle\{x_2, \dots, x_7\}, \{y_4, y_6, y_7\}\rangle$, and $\langle\{x_2, x_4, x_5, x_6\}, \{y_6, \dots, y_9\}\rangle$ are maximal full rectangles in table from Fig. 1. There is an alternative way to look at maximal full rectangles. Namely, $\langle A, B\rangle$ is a maximal full rectangle iff A consists of all objects common to all attributes from B and, conversely, B consists of all attributes common to all objects from A.

	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}
x_1										
x_2			×	×	×	×	×	×	×	
x_3		×	×	×	×	×	×		×	
x_4		×	×	×	×	×	×	×	×	
x_5		×		×	×	×	×	×	×	
x_6		×	×	×		×	×	×	×	
x_7		×	×	×	×	×	×	×		
x_8										

Fig. 1. Data table with binary attributes with one rectangle.

Due to the alternative point of view, maximal full rectangles $\langle A, B\rangle$ can be interpreted as interesting clusters/granules of objects and attributes drawn together by having common attributes and having common objects. This interpretation is adopted in formal concept analysis (FCA), a method of data analysis based on extraction of maximal full rectangles, attribute dependencies, and related information from data tables, see [2], [3]. Data analysis based on maximal full rectangles proved to be useful in several areas, e.g. data mining (extraction of collections of maximal rectangles for direct data analysis or as method of preprocessing, e.g. in mining of non-redundant association rules), software engineering (class hierarchies, software structure), information retrieval (structured queries), and several further particular areas like textual data, psychology, civil engineering, education, see [2] for further information and references.

Full rectangles $\langle A, B\rangle$ do not allow for blanks (0's) inside. Equivalently, this means that objects from A are shared by *all* attributes from B (and *vice versa*). However, there might be several reasons for allowing blanks inside rectangles. First, data may be noisy and then accidentally switching $\times$ to blank (1 to 0) may lead to loss of a possibly interesting rectangle. More importantly, there might be quite interesting rectangular patterns in data which contain a reasonably small number of blanks, i.e. rectangles which are "dense". For illustration, consider the table from Fig. 1. We have seen that the table contains several maximal full rectangles. Now, if "dense rectangle" means a rectangle in which each column contains at most one blank, then the table contains just one non-trivial maximal dense rectangle, namely rectangle $\langle A, B\rangle$ with $A = \{x_2, \dots, x_7\}$ and $B = \{y_2, \dots, y_9\}$. Notice that this rectangle is "the one" (the only one) we can see in the table. If one does not allow for blanks inside rectangles, this rectangle eludes us and remains hidden. Note that considering dense rectangles instead of full rectangles corresponds to replacing "…common to *all* objects/attributes" by "…common to *most* of the objects/attributes".

This paper brings results on "dense rectangles" in data tables with binary attributes. First, we provide theoretical insight. Namely, in case of maximal full rectangles, theoretical insight makes possible an efficient implementation of analysis of rectangular patterns [3]. We concentrate on criteria of denseness which limit the number of blanks in columns, or dually, in rows of rectangles. We present results on closure structures which are behind dense rectangles. The well-known results from formal concept analysis follow from our results for the particular case when no blanks are allowed (i.e. "dense" means "full"). Second, we present illustrative examples. Third, we sketch topics which did not fit into the scope of the present paper and topics for future research.

Due to the limited scope of the paper, we omit proofs.

B. Basics from data tables with binary attributes and formal concept analysis

In the basic setting, formal concept analysis (FCA) deals with input data in the form of a table with rows corresponding to objects and columns corresponding to attributes which describes a relationship between the objects and attributes. *A data table* (*with binary attributes*) can be identified with a triplet $\langle X, Y, I\rangle$ where X is a non-empty finite set (of *objects*), Y is a non-empty finite set

Manuscript received December 28, 2005. This work was supported by grant No. 1ET101370417 of GA AV ČR, by grant No. 201/05/0079 of the Czech Science Foundation, and by institutional support, research plan MSM 6198959214.

Authors' address: Dept. Computer Science, Palacký University, Tomkova 40, CZ-779 00, Olomouc, Czech Republic, email: radim.belohlavek@upol.cz, vilem.vychodil@upol.cz

(of *attributes*), and $I \subseteq X \times Y$ is an (object-attribute) relation with $\langle x, y\rangle \in I$ indicating that object x has attribute y. For each $A \subseteq X$ denote by $A^{\Uparrow}$ a subset of Y defined by

$$A^{\Uparrow} = \{y \in Y \mid \text{for each } x \in A\colon \langle x, y\rangle \in I\}. \qquad (1)$$

Similarly, for $B \subseteq Y$ denote by $B^{\Downarrow}$ a subset of X defined by

$$B^{\Downarrow} = \{x \in X \mid \text{for each } y \in B\colon \langle x, y\rangle \in I\}. \qquad (2)$$

That is, $A^{\Uparrow}$ is the set of all attributes from Y shared by all objects from A (and similarly for $B^{\Downarrow}$). A *formal concept* in $\langle X, Y, I\rangle$ is a pair $\langle A, B\rangle$ of $A \subseteq X$ and $B \subseteq Y$ satisfying $A^{\Uparrow} = B$ and $B^{\Downarrow} = A$. That is, a formal concept consists of a set A (so-called *extent*) of objects covered by the concept and a set B (so-called *intent*) of attributes covered by the concept such that A is the set of all objects sharing all attributes from B and, conversely, B is the collection of all attributes from Y shared by all objects from A. As mentioned in Section I-A, formal concepts are maximal rectangles of $\langle X, Y, I\rangle$ which are full of $\times$'s (or 1's): For $A \subseteq X$ and $B \subseteq Y$, $\langle A, B\rangle$ is a formal concept in $\langle X, Y, I\rangle$ iff $A \times B \subseteq I$ (i.e., $\langle A, B\rangle$ is a rectangle full of $\times$'s) and there is no $A' \supset A$ or $B' \supset B$ such that $A' \times B \subseteq I$ or $A \times B' \subseteq I$ (i.e., $\langle A, B\rangle$ is maximal).

The set $\mathcal{B}(X, Y, I) = \{\langle A, B\rangle \mid A^{\Uparrow} = B, B^{\Downarrow} = A\}$ of all formal concepts in $\langle X, Y, I\rangle$ can be equipped with a partial order $\leq$ (modeling the subconcept-superconcept hierarchy, e.g. *dog* $\leq$ *mammal*) defined by

$$\langle A_1, B_1\rangle \leq \langle A_2, B_2\rangle \text{ iff } A_1 \subseteq A_2 \text{ (iff } B_2 \subseteq B_1).$$

Under $\leq$, $\mathcal{B}(X, Y, I)$ happens to be a complete lattice, called a concept lattice, the basic structure of which is described by the so-called main theorem of concept lattices [3], the first part of which is presented in the next theorem.

Theorem 1: The set $\mathcal{B}(X, Y, I)$ is under $\leq$ a complete lattice where the infimum and supremum of $\langle A_j, B_j\rangle$'s, $j \in J$, are given by

$$\textstyle\bigwedge_{j\in J}\langle A_j, B_j\rangle = \langle \bigcap_{j\in J} A_j, (\bigcup_{j\in J} B_j)^{\Downarrow\Uparrow}\rangle,$$
$$\textstyle\bigvee_{j\in J}\langle A_j, B_j\rangle = \langle (\bigcup_{j\in J} A_j)^{\Uparrow\Downarrow}, \bigcap_{j\in J} B_j\rangle.$$

II. Maximal dense rectangles

A. Definition

Definition 2: A *rectangle* over sets X and Y is a pair $\langle A, B\rangle$ with $A \subseteq X$ and $B \subseteq Y$. A rectangle $\langle A, B\rangle$ is a subrectangle of a rectangle $\langle C, D\rangle$ ($\langle C, D\rangle$ is a superrectangle of $\langle A, B\rangle$) if $A \subseteq C$ and $B \subseteq D$.

Occasionally, we also speak of a rectangle $\langle A, B\rangle$ in a data table $\langle X, Y, I\rangle$. A rectangle $\langle A, B\rangle$ in $\langle X, Y, I\rangle$ corresponds to a subtable (submatrix) of table $\langle X, Y, I\rangle$ delineated by rows given by objects from A and columns given by attributes from B. For brevity, the corresponding subtables will also be called rectangles. $\langle A, B\rangle$ is a *proper* subrectangle of a rectangle $\langle C, D\rangle$ if $\langle A, B\rangle$ is a subrectangle of $\langle C, D\rangle$, and $A \subset C$ or $B \subset D$.

We will consider properties of rectangles over given sets X and Y. For such a property $\mathcal{D}$ and a data table $\langle X, Y, I\rangle$, we denote by $\mathcal{D}_I(A, B)$ the fact that rectangle $\langle A, B\rangle$ has property $\mathcal{D}$ in data table $\langle X, Y, I\rangle$ (formally, for a given data table $\langle X, Y, I\rangle$, a property $\mathcal{D}$ can be understood as a mapping assigning to any $I \subseteq X \times Y$ and any rectangle $\langle A, B\rangle$ a truth value 1 or 0 indicating whether $\langle A, B\rangle$ has property $\mathcal{D}$ or not). If I is clear from the context, we write just $\mathcal{D}(A, B)$ instead of $\mathcal{D}_I(A, B)$.

We are interested in properties $\mathcal{D}$ such that $\mathcal{D}(A, B)$ means that $\langle A, B\rangle$ is a dense rectangle in $\langle X, Y, I\rangle$. By $\langle A, B\rangle$ being dense we mean that "almost all entries of a subtable of $\langle X, Y, I\rangle$ given by $\langle A, B\rangle$ contain 1 (or $\times$)". The following are some examples of properties $\mathcal{D}$.

	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}
x_1											
x_2			×	×							
x_3		×		×							
x_4		×	×								
x_5											
x_6							×	×	×	×	×
x_7							×		×	×	
x_8							×	×		×	×
x_9								×	×		×

Fig. 2. Maximal dense rectangles.

Example 1: (1) Denote by *full* a property of rectangles in $\langle X, Y, I\rangle$ such that *full*(A, B) means that for each $x \in A$ and $y \in B$ we have $\langle x, y\rangle \in I$, i.e. rectangle $\langle A, B\rangle$ is full of 1's.

(2) Let l be a non-negative integer. Denote by $col(l)$ a property of rectangles in $\langle X, Y, I\rangle$ such that $col(l)(A, B)$, written also $col(l, A, B)$, means for each $y \in B$ we have

$$|\{x \in A \mid \langle x, y\rangle \notin I\}| \leq l,$$

i.e. rectangle $\langle A, B\rangle$ contains at most l zeros in each of its columns. More generally, if $\mathbf{l} = \{l_y \mid y \in Y\}$ is an Y-indexed collection of non-negative integers l_y we might consider a property $col(\mathbf{l})$ such that $col(\mathbf{l}, A, B)$ means that for each $y \in B$ we have $|\{x \in A \mid \langle x, y\rangle \notin I\}| \leq l_y$, i.e. for each column y of B we allow for at most l_y zeros. Clearly, if $l_y = l$ for all $y \in Y$, $col(\mathbf{l})$ is equivalent to $col(l)$ introduced above.

(3) Symmetrically, let k be a non-negative integer. Denote by $row(k)$ a property of rectangles in $\langle X, Y, I\rangle$ such that $row(k)(A, B)$, written also $row(k, A, B)$, means that for each $x \in A$ we have

$$|\{y \in B \mid \langle x, y\rangle \notin I\}| \leq k,$$

i.e. rectangle $\langle A, B\rangle$ contains at most k zeros in each of its rows. As in (2), we can consider $row(\mathbf{k})$ for a collection $\mathbf{k} = \{k_x \mid x \in X\}$ of non-negative integers k_x.

(4) Let k and l be non-negative integers. Denote by $row(k)\&col(l)$ a property of rectangles in $\langle X, Y, I\rangle$ which results by a conjunction of properties $row(k)$ and $col(l)$ from (2) and (3). Then $[row(k)\&col(l)](A, B)$ means that each row of rectangle $\langle A, B\rangle$ contains at most k zeros and each column of rectangle $\langle A, B\rangle$ contains at most l zeros. More generally, we can consider $row(\mathbf{k})\&col(\mathbf{l})$, see (2) and (3).

(5) Notice that for any non-negative integers k and l, the properties $row(0)\&col(l)$, $row(k)\&col(0)$, and *full* are equivalent. Furthermore, $row(|Y|)\&col(|X|)$ is a property which is true for any rectangle $\langle A, B\rangle$ in $\langle X, Y, I\rangle$.

Definition 3: Let $\langle X, Y, I\rangle$ be a data table, $\mathcal{D}$ be a property of rectangles in $\langle X, Y, I\rangle$. A rectangle $\langle A, B\rangle$ in $\langle X, Y, I\rangle$ is called *maximal* w.r.t. $\mathcal{D}$ if $\langle A, B\rangle$ has $\mathcal{D}$ and no proper superrectangle of $\langle A, B\rangle$ has $\mathcal{D}$.

The following example illustrates various properties $\mathcal{D}$.

Example 2: Consider the data table in Fig. 2.

(1) Let $\mathcal{D}$ be *full*. There are 15 rectangles maximal w.r.t. $\mathcal{D}$. Note that none of $\langle\{x_2, x_3, x_4\}, \{y_2, y_3, y_4\}\rangle$ and $\langle\{x_6, \ldots, x_9\}, \{y_7, \ldots, y_{11}\}\rangle$ has $\mathcal{D}$.

(2) Let $\mathcal{D}$ be $col(1)$. Both $\langle\{x_2, x_3, x_4\}, \{y_2, y_3, y_4\}\rangle$ and $\langle\{x_6, \ldots, x_9\}, \{y_7, \ldots, y_{11}\}\rangle$ are maximal rectangles which have $\mathcal{D}$. Note that among the rectangles maximal w.r.t. $\mathcal{D}$ there are also those

with an empty row, e.g. $\langle\{x_5, x_6\}, \{y_7, \ldots, y_{11}\}\rangle$. Such rectangles can be considered not interesting and can be disregarded with an appropriate definition of interestingness, cf. Section IV.

(3) Let $\mathcal{D}$ be $col(1)\&row(1)$. While $\langle\{x_2, x_3, x_4\}, \{y_2, y_3, y_4\}\rangle$ is a maximal rectangle which has $\mathcal{D}$, $\langle\{x_6, \ldots, x_9\}, \{y_7, \ldots, y_{11}\}\rangle$ is not (it does not have $\mathcal{D}$).

B. Maximal rectangles w.r.t. column-like properties and row-like properties

In the rest of this paper we will be concerned with a particular type of properties $\mathcal{D}$ which result by imposing restrictions on the number of blanks in columns (or, dually, rows) of rectangles. We call these properties column-like (row-like) properties. We start by their characteristic features.

Definition 4: A property $\mathcal{D}$ of rectangles over X and Y is called a *column-like* property if for any data table $\langle X, Y, I\rangle$ we have

1) for each $A \subseteq X$ and $B \subseteq Y$ we have $\mathcal{D}(A, \emptyset)$ and $\mathcal{D}(\emptyset, B)$,
2) for each $A_1, A_2 \subseteq X$ and $B \subseteq Y$: if *full*(A_1, B) and $\mathcal{D}(A_2, B)$, then $\mathcal{D}(A_1 \cup A_2, B)$,
3) for each $A_1 \subseteq A_2 \subseteq X$ and $B \subseteq Y$: if $\mathcal{D}(A_2, B)$ then $\mathcal{D}(A_1, B)$,
4) for each $A \subseteq X$ and $B_1, B_2 \subseteq Y$, $\mathcal{D}(A, B_1 \cup B_2)$ iff $\mathcal{D}(A, B_1)$ and $\mathcal{D}(A, B_2)$,
5) for any permutation π of X and any rectangle $\langle A, B\rangle$ such that $\langle x, y\rangle \in I$ iff $\langle \pi(x), y\rangle \in I$ *(*$x \in A$ and $y \in B$*):* $\mathcal{D}(A, B)$ iff $\mathcal{D}(\pi(A), B)$, where $\pi(A) = \{\pi(x) \,|\, x \in A\}$.

Lemma 5: $\mathcal{D}$ is a column-like property iff for each $\langle X, Y, I\rangle$ there is a collection $\mathbf{l} = \{l_y \,|\, y \in Y\}$ of non-negative integers l_y such that $\mathcal{D}_I(A, B)$ is equivalent to the fact that for each $y \in B$ we have $|\{x \in A \,|\, \langle x, y\rangle \notin I\}| \leq l_y$, see Example 1 (2).

Remark 1: Suppose we are given a data table $\langle X, Y, I\rangle$. Then, for the sake of brevity, we say that a column-like property $\mathcal{D}$ is (equivalent to) $col(\mathbf{l})$ if $\mathbf{l}$ is the collection of non-negative integers corresponding to $\mathcal{D}$ and $\langle X, Y, I\rangle$ due to Lemma 5.

Remark 2: Row-like properties are defined in a dual way to Definition 4. Therefore, row-like properties obey all features which are dual to properties of column-like properties. For instance, Lemma 5) has an obvious version for row-like properties. We omit details due to lack of space.

C. $\mathcal{D}$-concept lattices and related structures

In case of maximal full rectangles in a table $\langle X, Y, I\rangle$, given $A \subseteq X$, there is a unique maximal (hence, the largest) B such that $\langle A, B\rangle$ is a rectangle full of $\times$'s, namely $B = A^{\Uparrow}$. Likewise, for $B \subseteq Y$, $B^{\Downarrow}$ is the largest one such that $\langle B^{\Downarrow}, B\rangle$ is a rectangle full of $\times$'s. The mappings $^{\Uparrow}$ and $^{\Downarrow}$ form a so-called Galois connection between $\langle 2^X, \subseteq\rangle$ and $\langle 2^Y, \subseteq\rangle$, i.e. between the sets 2^X and 2^Y of all subsets of X and Y equipped with set inclusion. Maximal full rectangles are just fixpoints of $^{\Uparrow}$ and $^{\Downarrow}$, i.e. pairs $\langle A, B\rangle$ satisfying $A^{\Uparrow} = B$ and $B^{\Downarrow} = A$. The set of all these fixpoints is just a concept lattice $\mathcal{B}(X, Y, I)$.

For general column-like properties $\mathcal{D}$, the situation is different. While for $A \subseteq X$ there is still a largest $B \subseteq Y$ such that $\mathcal{D}(A, B)$, for $B \subseteq Y$, there might be several maximal sets $A \subseteq X$ such that $\mathcal{D}(A, B)$. In the following, we propose an attempt in which interesting rectangles satisfying $\mathcal{D}$ are identified as fixed points of mappings resembling very much Galois connections. Also, we show a generalization of Theorem 1, describing the set of the fixpoints, in our setting.

Definition 6: Let $\leq$ be a binary relation defined on 2^{2^X} by

$$\mathcal{A}_1 \leq \mathcal{A}_2 \quad \text{iff} \quad \begin{array}{l}\text{for each } A_1 \in \mathcal{A}_1 \text{ there is } A_2 \in \mathcal{A}_2 \\ \text{such that } A_1 \subseteq A_2,\end{array}$$

for $\mathcal{A}_1, \mathcal{A}_2 \in 2^{2^X}$.

It is easy to see that $\langle 2^{2^X}, \leq\rangle$ is a quasiordered set. That is $\leq$ is a quasiorder (i.e., a reflexive and transitive relation) on 2^{2^X}. For a quasiordered set $\langle U, \leq\rangle$, introduce a binary relation $\equiv_{\leq}$ on U by $u \equiv_{\leq} v$ iff $u \leq v$ and $v \leq u$. It is well known that $\equiv_{\leq}$ is an equivalence in U. Furthermore, an element $u \in U$ is called an infimum of $V \subseteq U$ if the following conditions are satisfied: for each $v \in V$ we have $u \leq v$; if there is some $w \in U$ such that for each $v \in V$ we have $w \leq v$ then $w \leq u$. The notion of a supremum in a quasiordered set can be introduced dually. Note that u is an infimum of V iff each v such that $u \equiv_{\leq} v$ is an infimum of V; the same for suprema. Note also that if $\leq$ is a partial order on U then if $V \subseteq U$ has an infimum, the infimum is unique; the same for suprema. Therefore, for brevity we denote any of the possibly several existing infima of $V \subseteq U$ by $\bigwedge_U V$; and use analogously $\bigvee_U V$ for suprema. Call a quasiordered set $\langle U, \leq\rangle$ a complete lattice if for each $V \subseteq U$ there exist both an infimum and a supremum of V.

Lemma 7: $\langle 2^{2^X}, \leq\rangle$ is a quasiordered set which is a complete lattice with infima $\bigwedge_U$ and suprema $\bigvee_U$ given by

$$\bigwedge_{U, j \in J} \mathcal{A}_j = \Big\{\bigcap_{j \in J} A_j \,\Big|\, \text{ for each } j \in J : A_j \in \mathcal{A}_j\Big\},$$
$$\bigvee_{U, j \in J} \mathcal{A}_j = \bigcup_{j \in J} \mathcal{A}_j.$$

Remark 3: In the above expression for infima, $\{\bigcap_{j \in J} A_j \,|\, \text{for each } j \in J : A_j \in \mathcal{A}_j\}$, is the system of intersections of all J-indexed collections $\{A_j \,|\, j \in J\}$ such that for any $j \in J$ we have $A_j \in \mathcal{A}_j$.

Note that if $X \neq \emptyset$ then $\leq$ defined on 2^{2^X} is not antisymmetric, i.e. it is not a partial order (consider $\{\emptyset, \{x\}\}$ and $\{\{x\}\}$).

Consider now a column-like property $\mathcal{D}$ over sets X and Y. Let $\langle X, Y, I\rangle$ be a data table. For $A \subseteq X$ there is a largest $B \subseteq Y$ such that $\mathcal{D}(A, B)$. This follows from 1. and 4. of Definition 4. Let thus the largest B with $\mathcal{D}(A, B)$ be denoted by $A^{\uparrow}$.

For $B \subseteq Y$, however, there might be several sets $A \subseteq X$ which are maximal such that $\mathcal{D}(A, B)$ (i.e., we have $\mathcal{D}(A, B)$ and there is no $A' \supset A$ such that $\mathcal{D}(A', B)$). Put therefore for $B \subseteq Y$,

$$B^{\downarrow} = \{A \subseteq X \,|\, A \text{ is maximal such that } \mathcal{D}(A, B)\}. \qquad (3)$$

We have defined $^{\uparrow}$ as a mapping assigning subsets of Y to subsets of X. We now extend $^{\uparrow}$ to assign subsets of Y to collections of subsets of X. This is because for $B \subseteq Y$, there might be several maximal $A \in B^{\downarrow}$, see above.

For a system $\mathcal{A} \in 2^{2^X}$ of subsets of X, put

$$\mathcal{A}^{\uparrow} = \bigcap_{A \in \mathcal{A}} A^{\uparrow}. \qquad (4)$$

Remark 4: Note that we use $^{\uparrow}$ for a mapping of 2^{2^X} to 2^Y as well as for a mapping of 2^X to 2^Y. It will always be clear from the context which mapping we mean by $^{\uparrow}$. Note also that $^{\uparrow} : 2^{2^X} \to 2^Y$ is an extension of $^{\uparrow} : 2^X \to 2^Y$ in that for $A \subseteq X$ we have $A^{\uparrow} = \{A\}^{\uparrow}$.

We have defined a pair of mappings $^{\uparrow} : 2^{2^X} \to 2^Y$, see (4), and $^{\downarrow} : 2^Y \to 2^{2^X}$, see (3). The following theorem shows their basic properties.

Theorem 8: We have

$$\mathcal{A}_1 \le \mathcal{A}_2 \text{ implies } \mathcal{A}_2^{\uparrow} \subseteq \mathcal{A}_1^{\uparrow}, \quad (5)$$
$$B_1 \subseteq B_2 \text{ implies } B_2^{\downarrow} \le B_1^{\downarrow}, \quad (6)$$
$$\mathcal{A} \le \mathcal{A}^{\uparrow\downarrow}, \quad (7)$$
$$B \subseteq B^{\downarrow\uparrow}. \quad (8)$$

Call any pair of mappings $^{\uparrow} : U \to V$ and $^{\downarrow} : V \to U$ between a quasiordered set $\langle U, \le\rangle$ and a partially ordered set $\langle V, \subseteq\rangle$ satisfying (5)–(8) a *Galois connection* between $\langle U, \le\rangle$ and $\langle V, \subseteq\rangle$. Therefore, $^{\uparrow}$ and $^{\downarrow}$ introduced above form a Galois connection between $\langle 2^{2^X}, \le\rangle$ and $\langle 2^Y, \subseteq\rangle$. Note that the difference from the ordinary notion of a Galois connection [3], [6] is that we allow $\langle U, \le\rangle$ be a quasiordered set while for the usual notion of a Galois connection, $\langle U, \le\rangle$ is required to be a partially ordered set (i.e. a quasiordered set such that $u_1 \le u_2$ and $u_2 \le u_1$ imply $u_1 = u_2$). Basic properties of Galois connections between a quasiordered set and a partially ordered set are different from the ordinary case and are shown in the next lemma.

Lemma 9: Mappings $^{\uparrow} : 2^{2^X} \to 2^Y$ and $^{\downarrow} : 2^Y \to 2^{2^X}$ satisfying *(5)–(8)* have the following properties.

- (i) $\mathcal{A}^{\uparrow} = \mathcal{A}^{\uparrow\downarrow\uparrow}$ for each $\mathcal{A} \in 2^{2^X}$.
- (ii) $B^{\downarrow} \equiv_{\le} B^{\downarrow\uparrow\downarrow}$ and $B^{\downarrow\uparrow\downarrow} = B^{\downarrow\uparrow\downarrow\uparrow\downarrow}$ for each $B \in 2^Y$.
- (iii) A mapping $C_X : 2^{2^X} \to 2^{2^X}$ defined by $C_X(\mathcal{A}) = \mathcal{A}^{\uparrow\downarrow}$ satisfies: $\mathcal{A} \le C_X(\mathcal{A})$, $\mathcal{A}_1 \le \mathcal{A}_2$ implies $C_X(\mathcal{A}_1) \le C_X(\mathcal{A}_2)$; $C_X(\mathcal{A}) \equiv_{\le} C_X(C_X(\mathcal{A}))$ and $C_X(C_X(\mathcal{A})) = C_X(C_X(C_X(\mathcal{A})))$.
- (iv) A mapping $C_Y : 2^Y \to 2^Y$ defined by $C_Y(B) = B^{\downarrow\uparrow}$ is a closure operator in $\langle 2^Y, \subseteq\rangle$, i.e. C_Y satisfies: $B \subseteq C_Y(B)$, $B_1 \subseteq B_2$ implies $C_Y(B_1) \subseteq C_Y(B_2)$; $C_Y(B) = C_Y(C_Y(B))$.

Definition 10: A *$\mathcal{D}$-concept lattice* of $\langle X, Y, I\rangle$ is a set

$$\mathcal{B}_{\mathcal{D}}(X, Y, I) = \{\langle \mathcal{A}, B\rangle \in 2^{2^X} \times 2^Y \mid \mathcal{A}^{\uparrow} = B,\ B^{\downarrow} = \mathcal{A}\}$$

equipped with a binary relation $\le$ defined by

$$\langle \mathcal{A}_1, B_1\rangle \le \langle \mathcal{A}_2, B_2\rangle \quad \text{iff} \quad \mathcal{A}_1 \le \mathcal{A}_2 \text{ (iff } B_1 \supseteq B_2\text{)}.$$

Following further the terminology of formal concept analysis, pairs $\langle \mathcal{A}, B\rangle \in \mathcal{B}_{\mathcal{D}}(X, Y, I)$ are called *formal $\mathcal{D}$-concepts*.

We are now going to show a theorem describing a structure of $\mathcal{D}$-concept lattices. We need the following lemma.

Theorem 11: For a column-like property $\mathcal{D}$ and a data table $\langle X, Y, I\rangle$, $\mathcal{B}_{\mathcal{D}}(X, Y, I)$ equipped with $\le$ is a complete lattice with infima $\bigwedge$ and suprema $\bigvee$ given by

$$\textstyle\bigwedge_{j\in J}\langle \mathcal{A}_j, B_j\rangle = \langle \{\bigcap_{j\in J} A_j \mid A_j \in \mathcal{A}_j\}^{\uparrow\downarrow\uparrow\downarrow}, (\bigcup_{j\in J} B_j)^{\downarrow\uparrow}\rangle, \quad (9)$$

$$\textstyle\bigvee_{j\in J}\langle \mathcal{A}_j, B_j\rangle = \langle (\bigcup_{j\in J} \mathcal{A}_j)^{\uparrow\downarrow\uparrow\downarrow}, \bigcap_{j\in J} B_j\rangle. \quad (10)$$

Remark 5: Note that the well-known Theorem 1 follows almost directly from Theorem 11. Indeed, consider col(0) as a property $\mathcal{D}$. It is easy to see that each $\langle \mathcal{A}, B\rangle \in \mathcal{B}_{\mathcal{D}}(X, Y, I)$ is of the form $\langle \mathcal{A}, B\rangle = \langle \{A\}, B\rangle$ for some $A \subseteq X$ and that $\langle \{A\}, B\rangle \in \mathcal{B}_{\mathcal{D}}(X, Y, I)$ iff $\langle A, B\rangle \in \mathcal{B}(X, Y, I)$ where $\mathcal{B}(X, Y, I)$ is the ordinary concept lattice of $\langle X, Y, I\rangle$. Therefore, $\mathcal{B}_{\mathcal{D}}(X, Y, I)$ is isomorphic to $\mathcal{B}(X, Y, I)$. Moreover, for $\langle \{A_j\}, B_j\rangle \in \mathcal{B}_{\mathcal{D}}(X, Y, I)$ we have $\{\bigcap_{j\in J} A_j\}^{\uparrow\downarrow\uparrow\downarrow} = \{\bigcap_{j\in J} A_j\}$ and $(\bigcup_{j\in J}\{A_j\})^{\uparrow\downarrow\uparrow\downarrow} = (\{\bigcup_{j\in J} A_j\})^{\uparrow\downarrow}$. This shows that Theorem 1 follows from Theorem 11.

D. Computing $\mathcal{D}$-concept lattices

Observe that $\mathcal{B}_{\mathcal{D}}(X, Y, I) = \{\langle B^{\downarrow}, B\rangle \mid B \in \text{fix}(C_Y)\}$ where $\text{fix}(C_Y) = \{B \subseteq Y \mid B = C_Y(B)\}$ is a set of all fixpoints of C_Y (note that $C_Y(B) = B^{\downarrow\uparrow}$). Therefore, in order to obtain $\mathcal{B}_{\mathcal{D}}(X, Y, I)$ it suffices to compute $\text{fix}(C_Y)$. Lemma 9 (iv) says that C_Y is a closure operator in $\langle 2^Y, \subseteq\rangle$. A set of all fixpoints of such a closure operator can be efficiently computed using Ganter's NextClosure algorithm [3] provided we can efficiently compute $C_Y(B)$ (for $B \subseteq Y$).

To find an efficient algorithm for computation of C_Y for a general column-like property $\mathcal{D}$ seems to be an interesting problem. Note that for particular choices of $\mathcal{D}$, we can use the following idea.

Let $\mathcal{D}$ be $col(\mathbf{l})$, where $\mathbf{l} = \{l_y \mid y \in Y,\ l_y = 0 \text{ or } l_y = 1\}$, see Example 1. That is, we allow for at most one zero in columns y with $l_y = 1$. In order to compute $C_Y(B) = B^{\downarrow\uparrow}$, we need to compute $\mathcal{A} = B^{\downarrow}$ and $\mathcal{A}^{\uparrow}$ (namely, $C_Y(B) = \mathcal{A}^{\uparrow}$).

Given $\mathcal{A} \in 2^{2^X}$, computation of $\mathcal{A}^{\uparrow}$ is obvious. Namely, we have $\mathcal{A}^{\uparrow} = \bigcap_{A\in\mathcal{A}} A^{\uparrow}$ and $A^{\uparrow} = \{y \in Y \mid \mathcal{D}(A, \{y\})\}$ due to 4. of Definition 4.

In order to compute $B^{\downarrow}$, put $B_0 = \{y \in B \mid l_y = 0\}$, $B_1 = \{y \in B \mid l_y = 1\}$, and consider the following undirected graph G. Vertices: The set of vertices of G is the set $X - (B^{\Downarrow} \cup Z)$ where

$$Z = \{x \in X \mid \text{there is } y \in B_0 : \langle x, y\rangle \notin I\}.$$

That is, vertices are particular objects from X. Edges: There is an edge between vertices x_1 and x_2 of G iff there is no $y \in B_1$ such that $\langle x_1, y\rangle \notin I$ and $\langle x_2, y\rangle \notin I$, i.e. neither of x_1 and x_2 has attribute y.

Recall that a clique in G is any set M of vertices of G such that for each $x_1, x_2 \in M$ there is an edge between x_1 and x_2. A clique M is maximal if no other vertex can be added to M so that M be still a clique. For technical reasons, if the set of vertices of G is empty, we consider $\emptyset$ (empty set) as a clique of G (this is then the only maximal clique of G). It is then easy to see the following assertion.

Lemma 12: For $B \subseteq Y$ we have

$$B^{\downarrow} = \{B^{\Downarrow} \cup M \mid M \text{ is a maximal clique in } G\}.$$

Recall that efficient algorithms for listing all maximal cliques exist (see e.g. [1], [4]). For more general properties $\mathcal{D}$, the same idea leads to analogous clique-characterizations. For instance, for D being $col(l)$, we get maximal cliques in uniform hypergraph with edges of size $l - 1$. These topics need to be explored both theoretically and experimentally.

III. Illustrative examples

In this section we present illustrative examples of $\mathcal{D}$-concept lattices and show experimental results with randomly generated data tables.

For brevity, we adopt the following convention for denoting column-like properties. Given a data table $\langle X, Y, I\rangle$, we assume that $Y = \{y_1, \dots, y_n\}$ is ordered by $y_1 < y_2 < \cdots < y_n$. Then, each column-like property $\mathcal{D}$ for $\langle X, Y, I\rangle$ is uniquely given by a sequence $l_{y_1}, l_{y_2}, \dots, l_{y_n}$ of nonnegative integers, meaning that $\mathcal{D}$ is equivalent to $col(\mathbf{l})$, where $\mathbf{l} = \{l_{y_1}, \dots, l_{y_n}\}$, see Example 1. If there is no danger of confusion, we write $l_{y_1} l_{y_2} \cdots l_{y_n}$ instead of $l_{y_1}, l_{y_2}, \dots, l_{y_n}$ and we denote $\mathcal{D}$ by $col(l_{y_1} l_{y_2} \cdots l_{y_n})$. For instance, if $Y = \{y_1, \dots, y_4\}$, then $col(0101)$ represents column-like property $\mathcal{D}$ which allows one blank in columns y_2 and y_4 and disallow any blanks elsewhere. This notation will be used in the following examples.

		a	b	c	d	e	f	g	h	i
leech	1	×	×					×		
bream	2	×	×					×	×	
frog	3	×	×	×				×	×	
dog	4	×		×				×	×	×
spike-weed	5	×	×		×		×			
reed	6	×	×	×	×		×			
bean	7	×		×	×	×				
maize	8	×		×	×		×			

Fig. 3. Data table [3]; the attributes are: a: needs water to live, b: lives in water, c: lives on land, d: needs chlorophyll to produce food, e: two seed leaves, f: one seed leaf, g: can move around, h: has limbs, i: suckles its offspring.

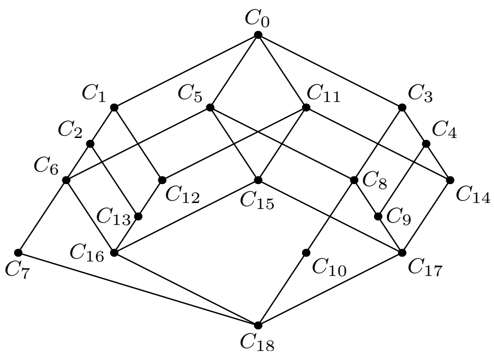

Fig. 4. Hierarchy of concepts

Example 3: Consider a data table $\langle X, Y, I\rangle$ presented in Fig. 3. This is a standard toy-example taken from [3]. The set X of objects contains objects $1, 2, \dots$ denoting organisms "leech", "bream", …; the set Y contains attributes $a, b, \dots$ denoting certain properties of organisms, see the comment under Fig. 3. The concept lattice $\mathcal{B}(X, Y, I)$ induced by $\langle X, Y, I\rangle$ has 19 formal concepts (maximal retangles), which are denoted by $C_0, \dots, C_{18}$:
$C_0 = \langle X, \{a\}\rangle$, $C_1 = \langle\{1, 2, 3, 4\}, \{a, g\}\rangle$,
$C_2 = \langle\{2, 3, 4\}, \{a, g, h\}\rangle$, $C_3 = \langle\{5, 6, 7, 8\}, \{a, d\}\rangle$,
$C_4 = \langle\{5, 6, 8\}, \{a, d, f\}\rangle$, $C_5 = \langle\{3, 4, 6, 7, 8\}, \{a, c\}\rangle$,
$C_6 = \langle\{3, 4\}, \{a, c, g, h\}\rangle$, $C_7 = \langle\{4\}, \{a, c, g, h, i\}\rangle$,
$C_8 = \langle\{6, 7, 8\}, \{a, c, d\}\rangle$, $C_9 = \langle\{6, 8\}, \{a, c, d, f\}\rangle$,
$C_{10} = \langle\{7\}, \{a, c, d, e\}\rangle$, $C_{11} = \langle\{1, 2, 3, 5, 6\}, \{a, b\}\rangle$,
$C_{12} = \langle\{1, 2, 3\}, \{a, b, g\}\rangle$, $C_{13} = \langle\{2, 3\}, \{a, b, g, h\}\rangle$,
$C_{14} = \langle\{5, 6\}, \{a, b, d, f\}\rangle$, $C_{15} = \langle\{3, 6\}, \{a, b, c\}\rangle$,
$C_{16} = \langle\{3\}, \{a, b, c, g, h\}\rangle$, $C_{17} = \langle\{6\}, \{a, b, c, d, f\}\rangle$,
$C_{18} = \langle\{\}, Y\rangle$.
Fig. 4 depicts the concept lattice $\mathcal{B}(X, Y, I)$ [3], i.e. the partially ordered hierarchy of formal concepts $C_0, \dots, C_{18}$. As mentioned above (cf. Remark 5) if we take $\mathcal{D}$ to be $col(0)$ (no blanks allowed), the $\mathcal{D}$-concept lattice $\mathcal{B}_{col(0)}(X, Y, I)$ is "the same" as the ordinary concept lattice $\mathcal{B}(X, Y, I)$. In more detail, we have

$$\mathcal{B}_{col(0)}(X, Y, I) = \{\langle\{A\}, B\rangle \mid \langle A, B\rangle \in \mathcal{B}(X, Y, I)\},$$

i.e. formal $col(0)$-concepts of $\mathcal{B}_{col(0)}(X, Y, I)$ are in the form $\langle\{A\}, B\rangle$ and correspond in a one-to-one way to formal concepts $\langle A, B\rangle$ of $\mathcal{B}(X, Y, I)$. Therefore, the structure of $\mathcal{B}_{col(0)}(X, Y, I)$ is given by Fig 4. On the other hand, by various choices of $\mathcal{D}$ we can get simplified or extended sets (hierarchies) of $\mathcal{D}$-concepts. For instance, if $\mathcal{D}$ is $col(011000000)$, i.e. if we allow one blank in columns corresponding to attributes "lives in water" and "lives on land", we get the following set of $\mathcal{D}$-concepts:
$C_0 = \langle\{X\}, \{a\}\rangle$,
$C_5 = \langle\{\{3, 4, 5, 6, 7, 8\}, \{2, 3, 4, 6, 7, 8\},$

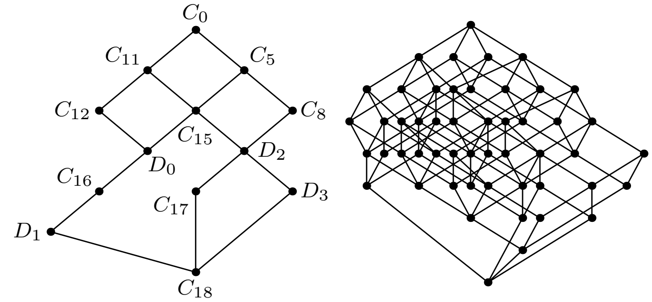

Fig. 5. Hierarchies of $\mathcal{D}$-concepts

	y_1	y_2	y_3	y_4	y_5
x_1	×		×	×	
x_2	×		×		×
x_3		×		×	×
x_4	×	×	×		×
x_5			×	×	×

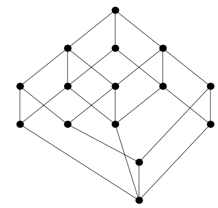

Fig. 6. Data table and the corresponding $\mathcal{B}(X, Y, I)$.

$\{1, 3, 4, 6, 7, 8\}\}, \{a, c\}\rangle$,
$C_8 = \langle\{\{5, 6, 7, 8\}\}, \{a, c, d\}\rangle$,
$C_{11} = \langle\{\{1, 2, 3, 5, 6, 8\}, \{1, 2, 3, 5, 6, 7\},$
$\{1, 2, 3, 4, 5, 6\}\}, \{a, b\}\rangle$,
$C_{12} = \langle\{\{1, 2, 3, 4\}\}, \{a, b, g\}\rangle$,
$C_{15} = \langle\{\{3, 5, 6, 8\}, \{3, 5, 6, 7\}, \{3, 4, 5, 6\}, \{2, 3, 6, 8\},$
$\{2, 3, 6, 7\}, \{2, 3, 4, 6\}, \{1, 3, 6, 8\}, \{1, 3, 6, 7\},$
$\{1, 3, 4, 6\}\}, \{a, b, c\}\rangle$,
$D_0 = \langle\{\{2, 3, 4\}, \{1, 3, 4\}\}, \{a, b, c, g\}\rangle$,
$C_{16} = \langle\{\{2, 3, 4\}\}, \{a, b, c, g, h\}\rangle$,
$D_1 = \langle\{\{4\}\}, \{a, b, c, g, h, i\}\rangle$,
$D_2 = \langle\{\{5, 6, 8\}, \{5, 6, 7\}\}, \{a, b, c, d\}\rangle$,
$C_{17} = \langle\{\{5, 6, 8\}\}, \{a, b, c, d, f\}\rangle$,
$D_3 = \langle\{\{7\}\}, \{a, b, c, d, e\}\rangle$,
$C_{18} = \langle\{\{\}\}, Y\rangle$.
The hierarchy of $\mathcal{D}$-concepts is depicted in Fig. 5 (left). Observe that $\mathcal{D}$-concepts denoted by C_i have the same intents as the corresponding $col(0)$-concepts. Extents of the corresponding $\mathcal{D}$-concepts and $col(0)$-concepts do not coincide in general because we use two different column-like properties. $\mathcal{B}_{\mathcal{D}}(X, Y, I)$ is smaller than $\mathcal{B}(X, Y, I)$. Thus, $\mathcal{B}_{\mathcal{D}}(X, Y, I)$ can be seen as a simplified view on $\mathcal{B}(X, Y, I)$ in which we allow $\mathcal{D}$-concepts which are not represented by rectangles full of 1's. $\mathcal{B}_{\mathcal{D}}(X, Y, I)$ contains four $\mathcal{D}$-concepts which do not have their analogies in $\mathcal{B}(X, Y, I)$: D_0 ($\mathcal{D}$-concept of organisms living in water and on land which can move around), D_1 ($\mathcal{D}$-concept of a dog), D_2 ($\mathcal{D}$-concept of organisms living in water and on land which need chlorophyll to produce food), D_3 ($\mathcal{D}$-concept of a bean). Extents of $\mathcal{D}$-concepts D_1 (a dog) and D_3 (a bean) are contained in $\mathcal{B}(X, Y, I)$ (see C_7 and C_{10}), however, intents of concepts C_7 and C_{10} differ from intents of D_1 and D_3.

Let us mention that other choices of $\mathcal{D}$ may extend the structure. As an example, consider column-like property $col(1)$ (one blank in each column). In this particular case, we have 51 $\mathcal{D}$-concepts, see Fig. 5 (right).

Example 4: This example further demonstrates an effect of choices of $\mathcal{D}$ on the resulting structure of $\mathcal{D}$-concepts. We consider a data table given by Fig. 6 (left). The concept lattice $\mathcal{B}(X, Y, I)$ induced by the table is depicted in Fig. 6 (right). If we focus on column-

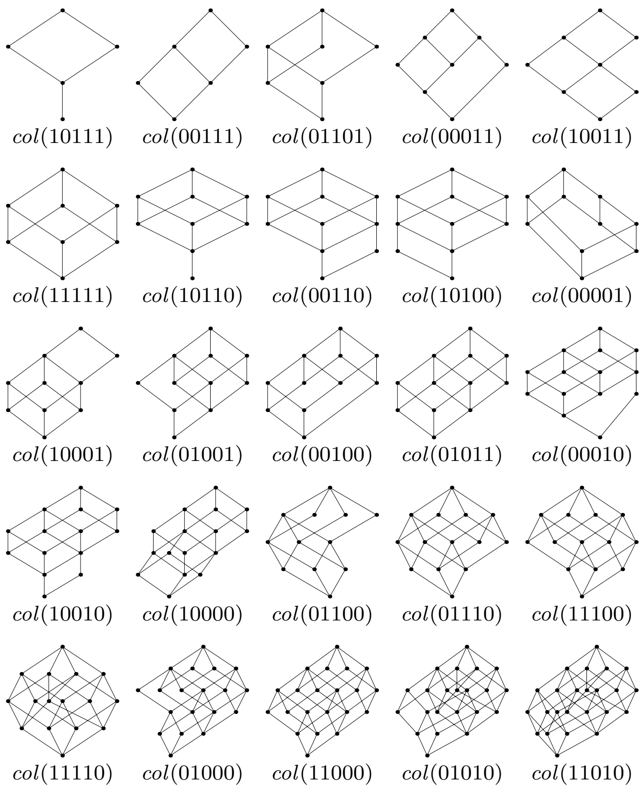

Fig. 7. Hierarchies of $\mathcal{D}$-concepts generated from a single data set using various column-like properties.

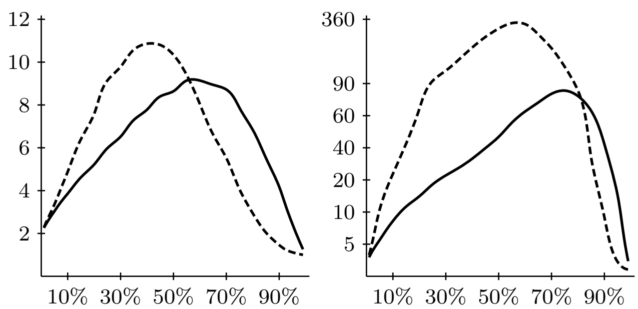

Fig. 8. Average number of $col(1)$-concepts according to density of input data tables.

like properties which allow at most one blank in a column, we have $2^5 = 32$ column-like properties $col(00000), \ldots, col(11111)$. In Fig. 7 we have displayed 25 different $\mathcal{D}$-concept lattices which result by choices of all $col(00001), \ldots, col(11111)$ (isomorphic $\mathcal{D}$-concept lattices were omitted, therefore we have only 25 lattices instead of 31). In addition to that, the $\mathcal{D}$-concept lattice corresponding to $col(00000)$ is isomorphic to the one from Fig. 6 (right), see above. Unlike the previous Example, now the collection of all $col(1)$-concepts is smaller than the collection of all $col(0)$-concepts (i.e., classical concepts).

Remark 6: We have seen from previous examples that in general we cannot say which of the number of all formal concepts of $\mathcal{B}_{col(0)}(X,Y,I)$ (maximal full rectangles) and the number of all formal concepts of $\mathcal{B}_{col(1)}(X,Y,I)$ (maximal dense rectangles) is smaller. Experiments have shown that in dense data tables (i.e., data tables which have only small proportion of blanks), $\mathcal{B}_{col(1)}(X,Y,I)$ is usually smaller than $\mathcal{B}_{col(0)}(X,Y,I)$. On the other hand, in data tables with average density the situation is the opposite. Consider the graphs in Fig. 8. Both graphs summarize average number of concepts (vertical axis) in dependence on the density of randomly generated data tables (density is displayed on the horizontal axis, $n\%$ means that $n\%$ of all table entries are 1's). The left-hand side table depicts the situation for data tables with 5 attributes, the right-hand side table depicts the situation for data tables with 10 attributes. Solid line in a graph represents average number of concepts ($col(0)$-concepts); dashed line represents average number of $col(1)$-concepts.

IV. Future research

- Investigation of other properties than column-like and row-like. A natural next step seems to be a conjunction $\mathcal{D}_1 \& \mathcal{D}_2$ where $\mathcal{D}_1$ is a column-like property and $\mathcal{D}_2$ is a row-like property.
- What are "the interesting" $\mathcal{D}$-concepts from $\mathcal{B}_{\mathcal{D}}(X,Y,I)$? The aim is to sharpen a condition of interestingness of $\mathcal{D}$-concepts, cf. Example 2. The intention is to leave out $\mathcal{D}$-concepts which are considered not interesting and to get this way a smaller number of extracted formal $\mathcal{D}$-concepts.
- Study of relationships between $\mathcal{B}_{\mathcal{D}}(X,Y,I)$ and $\mathcal{B}(X,Y,I)$ (ordinary concept lattice) and, in general, between $\mathcal{B}_{\mathcal{D}_1}(X,Y,I)$ and $\mathcal{B}_{\mathcal{D}_2}(X,Y,I)$ for various $\mathcal{D}_1$ and $\mathcal{D}_2$.
- Algorithms for computing $\mathcal{D}$-concepts and derived structures for other properties $\mathcal{D}$, see also Section II-D.

References

[1] Alexe G., Alexe S., Crama Y., Foldes S., Hammer P. L., Simeone B.: Consensus algorithms for the generation of all maximal bicliques. *Discrete Applied Mathematics*, Volume **145**, Issue 1 (2004), 11–21.

[2] Carpineto C., Romano G.: *Concept Data Analysis. Theory and Applications.* J. Wiley, 2004.

[3] Ganter B., Wille R.: *Formal Concept Analysis. Mathematical Foundations.* Springer, Berlin, 1999.

[4] Johnson D. S., Yannakakis M., Papadimitrou C. H.: On generating all maximal independent sets. *Inf. Processing Letters* **15**(1988), 129–133.

[5] Norris E. M.: An algorithm for computing the maximal rectangles of a binary relation. *Journal of ACM* 21:356–266, 1974.

[6] Ore O.: Galois connections. *Trans. Amer. Math. Soc.* 55:493–513, 1944.