Maximal Ordinal Two-Factorizations

Dominik Dürrschnabel[1,2]([✉]) and Gerd Stumme[1,2]

[1] Knowledge & Data Engineering Group, University of Kassel, Kassel, Germany
{duerrschnabel,stumme}@cs.uni-kassel.de
[2] Interdisciplinary Research Center for Information System Design,
University of Kassel, Kassel, Germany

Abstract. Given a formal context, an ordinal factor is a subset of its incidence relation that forms a chain in the concept lattice, i.e., a part of the dataset that corresponds to a linear order. To visualize the data in a formal context, Ganter and Glodeanu proposed a biplot based on two ordinal factors. For the biplot to be useful, it is important that these factors comprise as much data points as possible, i.e., that they cover a large part of the incidence relation. In this work, we investigate such ordinal two-factorizations. First, we investigate for formal contexts that omit ordinal two-factorizations the disjointness of the two factors. Then, we show that deciding on the existence of two-factorizations of a given size is an NP-complete problem which makes computing maximal factorizations computationally expensive. Finally, we provide the algorithm Ord2Factor that allows us to compute large ordinal two-factorizations.

Keywords: Formal concept analysis · Ordinal factor analysis · Ordinal two-factorization · Disjoint factorizations

1 Introduction

A common way to analyze datasets with binary attributes is to treat them as numerical, i.e., the value 1 is assigned to each attribute-object incidence and 0 to each missing one. Then, dimension reduction methods such as principal component analysis can be applied. Thereby, multiple attributes are merged into a few number of axes while being weighted differently. The emerging axes thus condense the presence or absence of correlated features. The objects are embedded into these axes as follows. Each object is assigned to a real-valued number in each axis to represent the best position of the object in the respective axis. Thereby, the resulting placement of an object that only has some of these attributes yields an ambiguous representation. Therefore, the main issues of this method arise. Assigning a real value to an object is not consistent with the level of measurement of the underlying binary data. It promotes the perception that an element has, compared to others, a stronger bond to some attributes, which is not possible in a formal context. A method that encourages such comparisons and results in such an inaccurate representation of the original information is, in our opinion, not valid. An example of such a principal component analysis projection is given in Fig. 1.
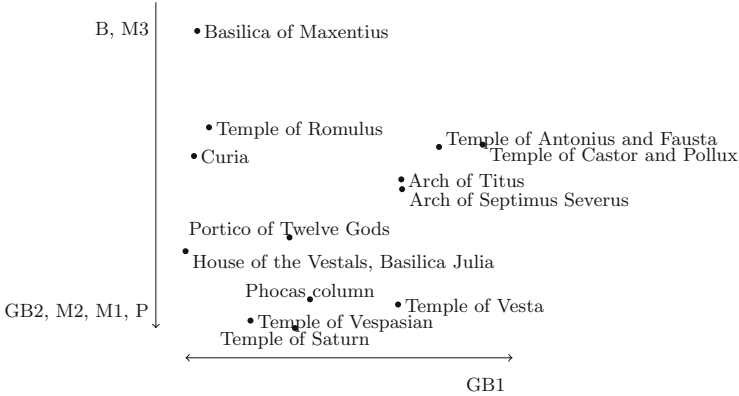
Fig. 1. A 2-dimensional projection of the objects from the dataset depicted in Fig. 3 using principal component analysis. This figure encourages the perception, that "Temple of Castor and Pollux" has a stronger bond to "GB1" than "Temple of Antonius and Fausta", which is false.

To address this problem, Ganter and Glodeanu [14] developed the method of *ordinal factor analysis.* It allows for a similar visualization technique while avoiding the problems that stem from the real-valued measurement on the binary attributes. Once again, multiple attributes are merged into a single factor. The computed projection in ordinal factor analysis thereby consists of linear orders of attributes, the so-called *ordinal factors.* Then, the method assigns each object, based on its attributes, a position in every factor. Compared to the principal component analysis approach, the positions assigned in the process are natural numbers instead of real-valued ones. Therefore, if interpreted correctly, the resulting projection does not express inaccurate and incorrect information. A desirable property of ordinal factorizations is completeness which allows the deduction of all original information. The positions of the objects are determined as follows. Each object in the two-dimensional coordinate system is placed at the last position of each axis such that it has all attributes until this position. Such a plot can be seen in Fig. 2. Reading the biplot is done as follows: Consider the "Portico of Twelve Gods" object. In the vertical factor, it is at position "GB1" which is preceeded by the attributes "M1", and "P" in this factor. Thus, the object has all three of these attributes. In the horizontal factor, it is at position "M1" which has no preceeding attributes. These incidences together precisely represents the incidences of the object. Deducing the same information from Fig. 1 is hardly possible.

However, Ganter and Glodeanu do not provide a method for the computation of ordinal two-factorizations. This is the point, where we step in with this work. We provide an algorithm to compute ordinal two-factorizations if they exist. We couple this algorithm with a method to compute a subset of the incidence relation of large size such that it admits an ordinal two-factorization. This enables the computation of ordinal two-factorizations of arbitrary datasets. This process

combined results in the algorithm ORD2FACTOR. Furthermore, we investigate ordinal two-factorizations with respect to their disjointness.
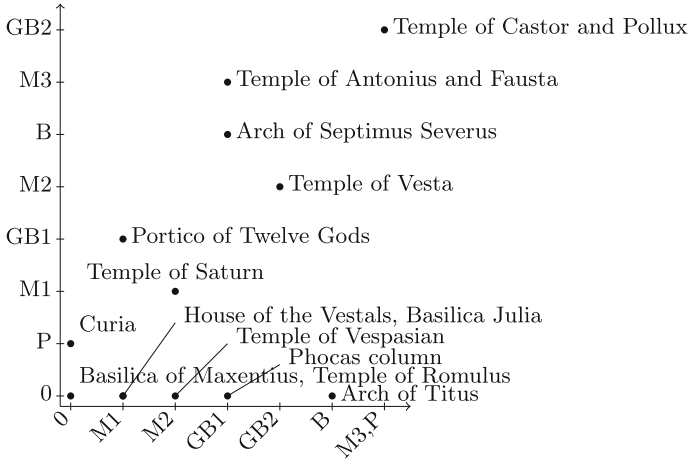

Fig. 2. A biplot of a maximal ordinal two-factorization of the data from Fig. 3. It represents all incidences from the formal context except (Temple of Romulus, GB1) and (Basilica of Maxentius, B).

2   Related Work

In this work, we consider a method that represent the objects in a low number of dimensions which condense multiple merged attributes. A commonly applied approach with the same fundamental idea is principal component analysis [26], which is a method that minimizes the average squared distances from the data points to a line. An example for a principal component analysis projection is depicted in Fig. 3. For an extensive survey on dimensional reduction methods, we refer the reader to Espadoto et al. [11]. A comparison of principal component analysis with the methods from formal concept analysis was described by Spangenberg and Wolff [27].

Our work is located in the research area formal concept analysis [15] (FCA). One well-researched way to apply dimensionality reduction in FCA is Boolean factor analysis [2–4,21,22]. Thereby, the incidence relation is represented by families of incident attribute and object subsets. This idea is highly related to the notion of a formal concept in FCA. Based on this method, Ganter and Glodeanu propose in [14] to group multiple Boolean factors into many-valued factors. One research direction that they consider to be of special interest is grouping them into ordinal factors, which is the direction we follow in this work. The same authors also demonstrate [17] that the research method is useful in application.

	B	GB1	GB2	M1	M2	M3	P
Arch of Septimus Severus	×	×		×	×		×
Arch of Titus	×	×	×	×	×		
Basilica Julia				×			
Basilica of Maxentius	×						
Curia							×
House of the Vestals				×			
Phocas column		×		×	×		
Portico of Twelve Gods		×		×			×
Temple of Antonius and Fausta	×	×		×	×	×	×
Temple of Castor and Pollux	×	×	×	×	×	×	×
Temple of Romulus		×					
Temple of Saturn				×	×		×
Temple of Vespasian				×	×		
Temple of Vesta		×	×	×	×		×

Fig. 3. Running example: This dataset compares attributes of different social media platforms.

The theory was also lifted to the triadic case [16] and graded data [18]. In our previous work [10], we propose an algorithm to greedily compute ordinal factors in large formal contexts. In a broader sense, the discovery of substructures, such as induced contranominal scales [9] or ordinal motifs [20] is an often-considered problem in the context of structural investigations in formal concept analysis.

The methods developed in this paper can be considered dual to the DIM-DRAW-algorithm [7]. In both approaches, two linear orders are sought-after to represent the dataset which is done using a connection to a bipartite subgraph.

3   Foundations of Ordinal Factor Analysis

In this section we briefly recap the definitions and notions from graph theory, formal concept analysis and ordinal factor analysis that are necessary to understand this paper. A graph is a tuple (V, E) with $E \subseteq \binom{V}{2}$. The set V is called the set of *vertices* and E the set of *edges*. A path between two vertices v_1 and v_n is a sequence of vertices $v_1 \ldots v_n$ with $\{v_i, v_{i+1}\} \in E$. A (connected) component is a maximal subset of the vertices of a graph, such that between every pair of vertices in the component there is a path. An *ordered set* is a tuple $(B, \leq)$ where $\leq$ is a binary relation on $B \times B$ that is reflexive, antisymmetric and transitive. The cocomparability graph of the ordered set $(B, \leq)$ is the graph (B, E) where $\{a, b\} \in E$ if $a \not\leq b$ and $b \not\leq a$. A *formal context* is a triple (G, M, I), where G is called the set of *objects*, M is called the set of *attributes* and $I \subseteq G \times M$ is called the incidence relation. The two *derivation operators* between the power sets of attributes and objects are given by $A' = \{m \in M \mid \forall g \in A : (g, m) \in I\}$ and $B' = \{g \in G \mid \forall m \in B : (g, m) \in I\}$. A formal concept is a tuple (A, B) with $A \subseteq G$ and $B \subseteq M$ such that $A' = B$ and $B' = A$. The set of all formal concepts

with is denoted by $\mathfrak{B}$ and forms together with the order relation $\leq$ for which $(A_1, B_1) \leq (A_2, B_2)$ if and only if $A_1 \subseteq A_2$, the concept lattice $\underline{\mathfrak{B}} = (\mathfrak{B}, \leq)$.

From here on we introduce the notions that are important for the present work. A *Ferrers relation* is a binary relation $F \subseteq G \times M$ with the property that if $(g, m) \in F$ and $(h, n) \in F$, either $(g, n) \in F$ or $(h, m) \in F$. An *ordinal two-factorization* of a formal context (G, M, I) is a set of two Ferrers relations F_1 and F_2 such that $F_1 \cup F_2 = I$. The *incompatibility graph* of a formal context (G, M, I) is the graph (I, E) with $\{(g, m), (h, n)\} \in E$ if and only if $(g, n) \notin I$ and $(h, m) \notin I$. We also say that (g, n) and (h, m) are *incompatible*. It is known, that a formal context admits an ordinal two-factorization, if and only if the incompatibility graph is bipartite [6, Sec.2, Prop.2]. Also note, that it is possible to extend an ordinal factor, i.e., a Ferrers relation, to a chain of formal concepts [14, Prop.7].

4   Disjointness of Ordinal Two-Factorizations

It is of interest wether the computed two ordinal factors are disjoint, i.e., if there is no information that is represented by both factors. Otherwise, the ordinal factorization contains redundant information as the two Ferrers relations share pairs of the incidence relation. However, as the formal context in Fig. 4 shows, achieving two disjoint ordinal factors is not always possible. The example is due to Das et al. [5], where they characterize the formal contexts that are factorizable into two disjoint ordinal factors as interval digraphs. The incompatibility graph of the example consists of two connected components, one consists of the single vertex $(6, f)$ and the other the vertices $I \setminus \{(6, f)\}$. We can assign the bipartition classes of this second component to factor 1 and 2 without loss of generality as shown on the right side of the Figure 4. But then the incidence pair then $(6, f)$ has to be in both factors, in factor 1 because as $(6, e) \in I$ and $(2, f) \in I$ but $(2, e) \notin I$ and in factor 2 because of $(5, f) \in I$, $(6, b) \in I$, and $(5, b) \notin I$.

We know that a formal context admits a two-factorization if its incompatibility graph is bipartite, and we can deduce from its incompatibility graph incidence pairs that cannot appear in the same ordinal factor. Still, we note that even in cases where disjoint two-factorizations do exist, the bipartition classes of the

	a	b	c	d	e	f	g
1				×	×	×	×
2						×	×
3							×
4	×						
5	×				×		
6	×	×			×	×	×
7	×	×	×		×		

	a	b	c	d	e	f	g
1				1	1	1	1
2						1	1
3							1
4	2						
5	2				2		
6	2	2			1	×	1
7	2	2	2		2		

Fig. 4. Example of a context with a maximal bipartite subgraph that does not give rise to an ordinal two-factorization. This example is due to Das et al. [5].
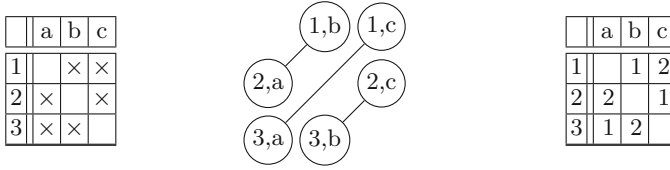
Fig. 5. *Left:* The formal context of a contranominal scale. *Middle:* its comparability graph. *Right:* A bipartition of the transitive comparability graph that does not give rise to an ordinal two-factorization.

incompatibility graph do not necessarily give rise to an ordinal two-factorization. To see this, refer to Fig. 5. The incompatibility graph (middle) of the formal context (left) consists of three components. An assignment of the incompatibility graph to bipartition classes can be seen on the right, however the elements $(2, a)$ and $(3, b)$ of factor 2 would imply, that the element $(3, a)$ also has to be in factor 2. However, this element is incompatible to element $(3, c)$, which is also in factor 2 by the assignment of the incompatibility graph. Thus, such an assignment does not always result in a valid ordinal two-factorization.

On the other hand, if the incompatibility graph is connected and bipartite, any assignment of the elements to bipartition classes of the incompatibility graph generates a valid ordinal two-factorization, as the following shows.

Proposition 1. *Let $\mathbb{K}$ be a formal context with a connected and bipartite incompatibility graph. Then there are two unique disjoint factors F_1 and F_2 that factorize $\mathbb{K}$. The sets F_1 and F_2 correspond to the bipartition classes of the incompatibility graph.*

Proof. As the cocomparability graph is bipartite, $\mathbb{K}$ admits an ordinal two-factorization. As the incompatibility graph is connected, it has unique bipartition classes. Finally, two elements in the same bipartition class cannot appear in the same ordinal factor. $\square$

On the other hand, if the incompatibility graph is not connected, it is of interest to further investigate the incidence pairs that can appear in both ordinal factors. We do so in the following.

Lemma 1. *Let $\mathbb{K} = (G, M, I)$ be a formal context with bipartite incompatibility graph (I, E). Let F_1, F_2 be an ordinal two-factorization of $\mathbb{K}$. For all elements $(g, m) \in F_1 \cap F_2$ it holds that $\{(g, m)\}$ is a connected component in (I, E).*

Proof. Assume not, i.e., there is an element $(g, m) \in F_1 \cap F_2$ that is not its own component. Then, there has to be some element $(h, n) \in I$ that is incompatible to (g, m), i.e., $(g, n) \notin I$ and $(h, m) \notin I$. But then (h, n) can be in neither F_1 nor F_2 which contradicts the definition of a Ferrers relation. $\square$

Thus, only the elements of the incompatibility graph that are not connected to another element can be in both ordinal factors. In the following, we further

characterize these isolated elements, as we show that for these elements it is always possible that they are in both factors, which then fully characterizes the intersection of the two ordinal factors.

Lemma 2. *Let* $\mathbb{K} = (G, M, I)$ *be a formal context with a two-factorization* F_1, F_2. *Let* C *be the set of all elements of* I *that are incompatible to no other element. Then* $F_1 \cup C$, $F_2 \cup C$ *is also an ordinal two-factorization.*

Proof. Let $G_i = F_i \setminus C$ and $\tilde{F}_i = F_i \cup C$ for $i \in \{1, 2\}$ Assume the statement is not true, i.e., either $\tilde{F}_1$ or $\tilde{F}_2$ is no ordinal factor. Without loss of generality, let $\tilde{F}_1 = F_1 \cup C$ be no ordinal factor. Then, there are two elements $(g, m), (h, n) \in \tilde{F}_1$ such that $(g, n) \notin \tilde{F}_1$ and $(h, m) \notin \tilde{F}_1$. As F_1 is an ordinal factor, at least one of the two elements has to be in C, let without loss of generality $(g, m) \in C$. We now do a case distinction whether one or both of them are in C.

Case 1. Let first $(g, m) \in C$ and $(h, n) \notin C$. One of the elements (g, n) or (h, m) has to be in I, otherwise (g, m) and (h, n) are incompatible, without loss of generality, let $(h, m) \in I$. As $(h, m) \notin C$, it has to hold that $(h, m) \in G_2$ and there has to be some $(x, y) \in G_1$ with $(h, y) \notin I$ and $(x, m) \notin I$. As $(x, y) \in G_1$, $(h, n) \in G_1$, and $(h, y) \notin I$ and F_1 is an ordinal factor, $(x, n) \in F_1$. As $(g, m) \in C$ it is incompatible with no element and thus not incompatible with (x, n) in particular, but $(x, m) \notin I$, it holds that $(g, n) \in I$. The element (g, n) has to be in G_2, as otherwise (g, m) and (h, n) are not incompatible. Thus, there also has to be some element in $(a, b) \in G_1$ with $(a, n) \notin I$ and $(g, b) \notin I$. As F_1 is an ordinal factor and $(x, n) \in F_1, (a, b) \in F_1$ and $(a, n) \notin F_1$, the element $(x, b) \in F_1$. But then (x, b) is incompatible to (g, m) which is a contradiction to (g, m) being in C.

Case 2. Let $(g, m) \in C$ and $(h, n) \in C$. Either $(g, n) \in I$ or $(h, m) \in I$, let without loss of generality $(g, n) \in I$. Then it has to be hold more specifically that in $(g, n) \in G_2$. Thus, there is some element $(x, y) \in G_1$ with $(h, y) \notin I$ and $(x, m) \notin I$. Because (x, y) has to be compatible with (h, n), it has to hold that $(x, n) \in I$. On the other hand, (x, n) has to be compatible with (g, m), thus $(g, n) \in I$. It then has to hold that $(g, n) \in G_2$ and thus some element $(a, b) \in G_1$ has to exist with $(g, b) \notin I$ and $(a, n) \notin I$. As $(x, y) \in F_1$ and $(a, b) \in F_1$ and F_1 is an ordinal factor, either $(a, y) \in F_1$ or $(x, b) \in F_1$. If $(a, y) \in F_1$, it would be incompatible to (g, m), if $(x, b) \in F_1$ it would be incompatible to (h, n). Both would be a contradiction to the respective element being in C.

This theorem finishes a complete characterization of the non-disjoint part of ordinal factors. Therefore, a partition of the incidence as follows always exists, if the context is ordinal two-factorizable.

Theorem 1. *Let* $\mathbb{K} = (G, M, I)$ *be a two-factorizable formal context with* (I, E) *its incompatibility graph. Then there is a partition of* I *into* F_1, F_2, C *with* $C = \{D \mid D$ *connected component of* $(I, E), |D| = 1\}$ *and* $F_1 \cup C$ *and* $F_2 \cup C$ *are Ferrers relations.*

Proof. This theorem directly follows from the previous lemma. Let $\tilde{F}_1, \tilde{F}_2$ be a two-factorization and $C = \{D \mid D$ connected component of $(I, E), |D| = 1\}$. Then the partition is given by $\tilde{F}_1 \setminus C$, $\tilde{F}_2 \setminus C$, and C.

5   An Algorithm for Ordinal Two-Factorizations

Now, we propose an algorithm to compute ordinal two-factorizations if they exist. As we saw in the last section, the bipartition classes of the incompatibility graph do not directly give rise to an ordinal two-factorizations. An important observation [15, Thm.46] is that a formal context can be described by the intersection of two Ferrers relations, if and only if its corresponding concept lattice can be described as the intersection of two linear orders, i.e., if it has order dimension two. As the complement of a Ferrers relation is once again a Ferrers relation, a formal context is two-factorizable if and only if the concept lattice of its complement context has order dimension two. We leverage this relationship with the following theorem, to explicitly compute the ordinal two-factorization.

Theorem 2. *Let* $\mathbb{K} = (G, M, I)$ *be a formal context. Let* $(B, \leq) = \underline{\mathfrak{B}}(\mathbb{K}^c)$. *If* $\mathbb{K}$ *is two-factorizable, then* $\leq$ *is two-dimensional and there is a conjugate order* $\leq_c$. *The sets*

$$F_1 = \{(g, m) \in G \times M \mid \nexists (A, B), (C, D) \in \mathfrak{B}(\mathbb{K}^c), g \in A, m \in D,$$
$$((A, B), (C, D)) \in (\leq \cup \leq_c)\}$$

and

$$F_2 = \{(g, m) \in G \times M \mid \nexists (A, B), (C, D) \in \mathfrak{B}(\mathbb{K}^c), g \in A, m \in D,$$
$$((A, B), (C, D)) \in (\leq \cup \geq_c)\}$$

give rise to an ordinal factorization of $\mathbb{K}$.

Proof. We have to show that F_1 and F_2 are Ferrers relations and $F_1 \cup F_2 = I$. We first show that F_1 is a Ferrers relation. By definition $\prec := \leq \cup \leq_c$ is a chain ordering all formal concepts of $\mathfrak{B}(K^c)$. Let (g, m) and (h, n) be two pairs in F_1. Assume that $(g, n) \notin F_1$ and $(h, m) \notin F_1$. Then there have to be two concept (A_1, B_1) and (A_2, B_2) with $g \in A_1$, $n \in B_2$ such that $(A_1, B_1) \prec (A_2, B_2)$. Similarly, there have to be two concepts (A_3, B_3) and (A_4, B_4) with $h \in A_3$, $m \in B_4$ such that $(A_3, B_3) \prec (A_4, B_4)$. For the concepts (A_2, B_2) and (A_3, B_3) it holds that $(A_2, B_2) \prec (A_3, B_3)$, as $(A_2, B_2) \neq (A_3, B_3)$ and $(A_3, B_3) \nprec (A_2, B_2)$ because $(g, m) \notin F_1$. By the same argument, $(A_4, B_4) \prec (A_1, B_1)$. Thus, $(A_1, B_1) \prec (A_2, B_2) \prec (A_3, B_3) \prec (A_4, B_4) \prec (A_1, B_1)$, which would imply that these three concepts are equal and is thus a contradiction. This proves that F_1 is a Ferrers relation. The argument to shows that F_2 is a Ferrers relation is dual. We now show that $F_1 \cup F_2 = I$. Let $(g, m) \in F_1 \cup F_2$, without loss of generality let it be an element of F_1. Then there are no two concepts $(A, B), (C, D) \in \mathfrak{B}(\mathbb{K}^c)$, with $g \in A$, $m \in D$ and $((A, B), (C, D)) \in (\leq \cup \leq_c)$. Consider the concept (g'', g') with the derivation from the complement context. By definition $((g'', g'), (g'', g')) \in (\leq \cup \leq_c)$ and thus $m \notin g'$ when using the derivation from the complement context, i.e., $(g, m) \notin I^c$. But then, $(g, m) \in I$. Now, let $(g, m) \in I$ and assume that $(g, m) \notin F_1$. Let (A, B) and (C, D) be arbitrary concepts of $\mathfrak{B}(\mathbb{K}^c)$ with $g \in A$ and $m \in D$. As $(g, m) \in I$, it is not in the incidence of

Algorithm 1. Compute Ordinal Two-Factorization

Input: Ordinal Two-Factorizable Formal Context $\mathbb{K} = (G, M, I)$
Output: Ordinal Two-Factorization F_1, F_2

```
def two_factor(G, M, I):
    (𝕭, ≤) = 𝕭(𝕂ᶜ)
    (𝕭, E) = co_comparability_graph(𝕭, ≤)
    ≤_c = transitive_orientation(𝕭, E)
    ≺₁ = ≤ ∪ ≤_c
    ≺₂ = ≤ ∪ ≥_c
    for i in {1, 2}:
        L_i = {}
        Ã = {}
        for (A, B) in 𝕭 ordered by ≺_i:
            Ã = Ã ∪ A
            L_i = L_i ∪ (Ã × B)
        F_i = (G × M) \ L_i
    return F_1, F_2
```

$\mathbb{K}^{\mathsf{c}}$ and thus (A, B) and (C, D) are not comparable with $\leq$. As $(g, m) \notin F_1$ it holds $(A, B) \not\leq_c (C, D)$ and as they are incomparable with $\leq$ it has to hold that $(A, B) \geq_c (C, D)$. Thus, $(g, m) \in F_2$. This shows that $F_1 \cup F_2 = I$ and thus concludes the proof.

This proof gives rise to the routine in Algorithm 1 where this information is used to compute an ordinal two-factorization of the formal context. It computes the two sets F_1 and F_2 from the previous theorem. To do so, it has to be paired with an algorithm to compute the concept lattice. An algorithm that is suitable is due to Lindig [24], as it computes the covering relation together with the concept lattice. Furthermore, an algorithm for transitive orientations [19] is required to compute the conjugate order. As we will discuss later, for both these algorithms the runtime is not critical, as we are interested in ordinal two-factorizations of small formal contexts. Modern computers are easily able to deal with such contexts. Still, if suitable supporting algorithms are chosen, this procedure results in a polynomial-time algorithm that computes ordinal two-factorizations.

6 Maximal Ordinal Two-Factorizations

In this section, we propose an algorithm to compute an ordinal two-factorizations of a formal context that covers a large part of the incidence relation.

Definition 1 (Maximal Ordinal Two-Factorizations). *For a formal context $\mathbb{K} = (G, M, I)$ a maximal ordinal two-factorization is a pair of Ferrers relations $F_1, F_2 \subseteq I$ such that there are no Ferrers relations $\tilde{F}_1, \tilde{F}_2 \subseteq I$ with $|\tilde{F}_1 \cup \tilde{F}_2| \geq |F_1 \cup F_2|$.*

While there are various thinkable ways, how to define maximal ordinal two-factorizations, we have chosen to do so by minimizing the size of not-covered incidence, as each element in the incidence can be viewed as a data point that would thus be lost. This definition also aligns with Ganter's suggestion in his textbook to maximize the size of the union [13]. Note, that a maximal two-factorization always exists, as a single element in the incidences is a Ferrers relation by itself.

6.1   Maximal Ordinal Two-Factorizations Are Hard

First, we investigate the computational complexity of the MAXIMAL ORDINAL TWO-FACTORIZATION PROBLEM. To do so, we perform a reduction from the TWO-DIMENSION EXTENSION PROBLEM. The problem requires finding the minimum number of pairs that need to be added to an order relation to make it two-dimensional. Formally, given an ordered set $(X, \leq)$ and a $k \in \mathbb{N}$ requests the decision whether there is a set $\tilde{\leq} \supset \leq$ such that is an order and $(X, \tilde{\leq})$ has order dimension two and $|\tilde{\leq}| - |\leq| = k$. Felsner and Reuter [12] showed that deciding this problem is NP-complete.

The decision problem, that is linked to the minimum number of pairs that we need to add to a relation to make it two-dimensional is the ORDINAL TWO-FACTORIZATION PROBLEM. For it, a formal context (G, M, I) and a $k \in \mathbb{N}$ are given. It is requested to decide the existance of a formal context $(G, M, \tilde{I})$ with $\tilde{I} \subseteq I$ and $|I| - |\tilde{I}| = k$ that has an ordinal two-factorization.

The relation between these two problems gives rise to the computational complexity of computing a two-factorization.

Lemma 3. *There is a polynomial-time reduction from the* TWO-DIMENSION EXTENSION PROBLEM *to the* ORDINAL TWO-FACTORIZATION PROBLEM.

Proof. Let $(X, \leq)$ and $k \in \mathbb{N}$ be an instance of the TWO-DIMENSION EXTENSION PROBLEM. We claim, that the problem has a solution if and only if ORDINAL TWO-FACTORIZATION PROBLEM $(X, X, \not\leq)$ with k has a solution.

Let $(X, \leq)$ be an ordered set with a two-dimension-extension C of size k. Let L_1, L_2 be a realizer, i.e., two linear extensions of $\leq \cup C$ with $L_1 \cap L_2 = \leq \cup C$. Then the relations $F_1 := (X \times X) \setminus L_1$ and $F_2 := (X \times X) \setminus L_2$ are Ferrers relations. Furthermore, $F_1 \subseteq \not\leq$ and $F_2 \subseteq \not\leq$ by definition. Assume that $F_1 \cup F_2 \cup C \neq \not\leq$. Then there has to be a pair $a, b \in X$ with $a \not\leq b$ and $(a, b) \notin C$. Then $(b, a) \in L_1$, or $(b, a) \in L_2$, or both, without loss of generality let $(b, a) \in L_1$. But this implies that $(a, b) \in L_1$ which is a contradiction.

Now, let for the formal context $(X, X, \not\leq)$ be C a set of cardinality k such that there are two Ferrers relations F_1, F_2 with $|F_1 \cup F_2 \cup C| = |\not\leq|$. Now, let $L_1 = (X \times X) \setminus F_1$ and $L_2 = (X \times X) \setminus F_2$. Then it holds that $L_1 \cap L_2 = \leq \cup C$. L_1 and L_2 are supersets of $\leq$ and Ferrers relations, thus they are transitive, reflexive and for all elements $a, b \in X$ it holds that $a, b \in L_i$ or $b, a \in L_i$. By definition for both $i \in \{1, 2\}$ there is a $\tilde{L}_i \subseteq L_i$ that has all these properties and is also antisymmetric. The existence of $\tilde{L}_i$ follows from placing a linear order on

each of the equivalence classes of L_i. Both $\tilde{L}_1$ and $\tilde{L}_2$ are linear extensions of $\leq$ and $|\tilde{L}_1 \cap \tilde{L}_2| \leq |L_1 \cap L_2| = |\leq| + k$.

Therefore, the claim is proven. and we reduced the Two-Dimension Extension Problem to the Ordinal Two-Factorization Problem.

Lemma 4. *Validation of a solution of the* Ordinal Two-Factorization *can be done in polynomial time.*

Proof. Given a formal context (G, M, I) and a set $C \subseteq I$ of size k, to check whether $(G, M, I \setminus C)$ admits an ordinal two-factorization is equivalent to the check whether the incompatibility graph of $(G, M, I \setminus C)$ is bipartite.

Thus, the Ordinal Two-Factorization Problem is in the same complexity class as the Two-Dimension Extension Problem.

Theorem 3. *The* Ordinal Two-Factorization Problem *is* NP-*complete.*

Proof. Follows from Lemmas 3 and 4.

6.2 Maximal Bipartite Subgraphs Are Not Sufficient

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r
1		×	×	×		×		×		×	×	×	×	×	×	×		
2	×		×	×	×	×	×	×	×	×	×	×	×	×	×		×	
3	×	×		×	×	×	×	×	×	×	×	×	×	×	×			×
4	×	×	×		×	×			×	×		×		×		×	×	×
5	×	×	×	×		×		×		×	×	×	×	×	×	×	×	×
6	×	×	×	×	×		×			×		×		×		×	×	×
7	×	×	×	×	×	×		×	×	×	×	×	×	×	×	×	×	×
8	×	×	×	×	×	×	×		×	×		×		×		×	×	×
9	×	×	×	×	×	×	×	×		×	×	×	×	×	×	×	×	×
10	×	×	×	×	×	×	×	×	×		×	×		×		×	×	×
11	×	×	×	×	×	×	×	×	×	×		×		×		×	×	×
12	×	×	×	×	×	×	×	×	×	×	×	×		×		×	×	×
13	×	×	×	×	×	×	×	×	×	×	×	×		×	×	×	×	×
14	×	×	×	×	×	×	×	×	×	×	×	×	×		×	×	×	×
15	×	×	×	×	×	×	×	×	×	×	×	×	×	×		×	×	×
16	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×		×	×
17	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×		×
18	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	

Fig. 6. Example of a context with a maximal bipartite subgraph that does not give rise to an ordinal two-factorization.

The structure of the incompatibility graph provides an interesting foundation to compute ordinal two-factorizations. It seems to be a tempting idea to compute

---

Algorithm 2. ORD2FACTOR to Compute Large Ordinal Two-Factorization

---

Input: Formal Context (G, M, I)
Output: Ordinal Factors F_1 and F_2

```
def Ord2Factor(G, M, I):
    (I, E) = incompatibility_graph(G, M, I)
    while (I, E) not bipartite:
        I = maximal_bipartite_inducing_vertex_set(I, E)
        (I, E) = incompatibility_graph(G, M, I)
    return two_factor(G, M, I)
```

---

the maximal induced bipartite subgraph of a formal context. The vertex set that induces such a maximal bipartite subgraph could than be used for the ordinal two-factorization. However, it turns out a bipartite subgraph does not always give rise to an ordinal two-factorization as Fig. 6 demonstrates. Its incompatibility graph has an odd cycle, and it is thus not bipartite. This fact makes the formal context not two-factorizable. An inclusion-minimal set that can be removed to make it bipartite is given by

$$C = \{(6, j), (4, n), (7, p), (18, p), (6, p), (6, n), (12, k), (10, g), (6, g), (5, p), (2, i),$$
$$(4, p), (12, m), (3, i), (12, h), (1, p), (2, q)\}.$$

However, the formal context $(G, M, I \setminus C)$ is again not two-factorizable as its incompatibility graph contains once again an odd cycle. This is possible as new incompatibilities can arise with the removal of incidences.

6.3   Computing Maximal Ordinal Two-Factorizations

In the last sections, we did structural investigations on ordinal two-factorizations and provided an algorithm to compute them. We now use this algorithm to compute a large ordinal two-factorization. To this end, we propose the algorithm ORD2FACTOR in Algorithm 2. Thereby, the induced bipartite subgraph of the incompatibility graph is computed. As new incompatibilities can arise by the removal of crosses, as noted previous section, we might have to repeat this procedure.

The induced bipartite subgraph can be computed using the methods proposed by Dürrschnabel et al. [8]. We are not aware of a formal context where the SAT-sovler approach described in this paper requires a second repetition of the algorithm which motivates the following open question.

OPEN QUESTION 1. Is there a formal context (G, M, I), such that the maximal set $\tilde{I}$ that induces a bipartite subgraph on the incompatibility graph does not give rise to a two-factorizable formal context?

This open problem is of special interest, because it would allow our approach to compute globally maximal two-factorization, as the following shows.

Theorem 4. *Let $\mathbb{K} = (G, M, I)$ be a formal context and $\tilde{I}$ the subset of I that induces a maximal bipartite subgraph on the incompatibility graph. If the formal context $\mathbb{K} = (G, M, \tilde{I})$ admits an ordinal two-factorization, its factors are the maximal ordinal factors of $\mathbb{K}$.*

Proof. Assume there are two ordinal factors F_1 and F_2 of $\mathbb{K}$ and $|F_1 \cup F_2| > |\tilde{I}|$. But then the context $(G, M, F_1 \cup F_2)$ is a two-factorizable and thus the graph induced by $F_1 \cup F_2$ on the incompatibility graph is bipartite, a contradiction.

7   Runtime Discussion

If a formal context has order dimension two, it cannot contain a contranominal of dimension three as an induced subcontext. From a result by Albano [1], it follows that a context without a contranominal scale of dimension three and thus especially for all two-dimensional formal contexts, the number of concepts is bounded from above by $\frac{3}{2}|G|^2$ or dually $\frac{3}{2}|M|^2$. There are algorithms that compute the set of all concept of a formal context with polynomial delay [23] and the computation of a conjugate order can be performed in quadratic time [25]. Thus, the algorithm to compute ordinal two-factorizations has polynomial runtime if it paired with the these algorithms.

For the computation of large ordinal factorizations of formal contexts that do not omit ordinal two-factorizations by their structure, the runtime-obstacle is the computation of the large induced bipartite subgraph. If the exact problem is solved, the algorithm thus has exponential runtime. Our previous work [8] also discusses three heuristics for the computation of bipartite subgraphs which can be plugged to achieve an algorithm that has polynomial runtime. Usually, we are interested in two-factorizations of formal context with limited size, as a human can otherwise not grasp the connections encoded in the dataset. Thus, the runtime of these algorithms are usually not the critical limitation and thus a method that is computationally expensive can be employed.

8   Conclusion

In this paper, we expanded on the work done on ordinal two-factorizations in the realm of ordinal factor analysis. First, we performed some structural investigations about the disjointness of the two ordinal factors. Thereby, we were able to characterize the incidence pairs that can appear in both ordinal factors as the isolated elements of the incompatibility graph. Then, we proposed an algorithm for the computation of maximal ordinal two-factorizations. To this end, we developed a polynomial time algorithm to compute a two-factorization of a formal context that has a bipartite incompatibility graph. We showed, that the problem to compute maximal ordinal two-factorizations is NP-complete and proposed our approach ORD2FACTOR to compute large ordinal two-factorizations. As we demonstrated that the problem entails an NP-complete problem, the resulting algorithm is exponential.

Datasets often consist not only of binary but also already ordinal data. The ordinal factor analysis in its current form can only deal with this data by interpreting it as binary. While scaling in formal concept analysis is a tool to deal with this data, factors will not necessarily respect the order encapsulated in the data. In our opinion, the next step should be to extend this method to deal with this kind of non-binary data directly.

References

1. Albano, A., Chornomaz, B.: Why concept lattices are large: extremal theory for generators, concepts, and vc-dimension. Int. J. Gen Syst. **46**(5), 440–457 (2017). https://doi.org/10.1080/03081079.2017.1354798
2. Belohlávek, R., Vychodil, V.: Formal concepts as optimal factors in boolean factor analysis: Implications and experiments. In: Eklund, P.W., Diatta, J., Liquiere, M. (eds.) Proceedings of the Fifth International Conference on Concept Lattices and Their Applications, CLA 2007, Montpellier, France, October 24–26, 2007. CEUR Workshop Proceedings, vol. 331. CEUR-WS.org (2007)
3. Belohlávek, R., Vychodil, V.: Discovery of optimal factors in binary data via a novel method of matrix decomposition. J. Comput. Syst. Sci. **76**(1), 3–20 (2010). https://doi.org/10.1016/j.jcss.2009.05.002
4. Boeck, P.D., Rosenberg, S.: Hierarchical classes: model and data analysis. Psychometrika **53**(3), 361–381 (1988). https://doi.org/10.1007/BF02294218
5. Das, S., Sen, M.K., Roy, A.B., West, D.B.: Interval digraphs: an analogue of interval graphs. J. Graph Theor. **13**(2), 189–202 (1989). https://doi.org/10.1002/jgt.3190130206
6. Doignon, J.P., Ducamp, A., Falmagne, J.C.: On realizable biorders and the biorder dimension of a relation. J. Math. Psychol. **28**(1), 73–109 (1984). https://doi.org/10.1016/0022-2496(84)90020-8
7. Dürrschnabel, D., Hanika, T., Stumme, G.: Drawing order diagrams through two-dimension extension. CoRR abs/1906.06208 (2019), arxiv.org/abs/1906.06208
8. Dürrschnabel, D., Hanika, T., Stumme, G.: Discovering locally maximal bipartite subgraphs. CoRR abs/2211.10446 (2022). 10.48550/arXiv. 2211.10446, https://doi.org/10.48550/arXiv.2211.10446
9. Dürrschnabel, D., Koyda, M., Stumme, G.: Attribute selection using contranominal scales. In: Braun, T., Gehrke, M., Hanika, T., Hernandez, N. (eds.) Graph-Based Representation and Reasoning - 26th International Conference on Conceptual Structures, ICCS 2021, Virtual Event, September 20–22, 2021, Proceedings. Lecture Notes in Computer Science, vol. 12879, pp. 127–141. Springer (2021). https://doi.org/10.1007/978-3-030-86982-3_10
10. Dürrschnabel, D., Stumme, G.: Greedy discovery of ordinal factors. CoRR abs/2302.11554 (2023). https://doi.org/10.48550/arXiv.2302.11554
11. Espadoto, M., Martins, R.M., Kerren, A., Hirata, N.S.T., Telea, A.C.: Toward a quantitative survey of dimension reduction techniques. IEEE Trans. Vis. Comput. Graph. **27**(3), 2153–2173 (2021). https://doi.org/10.1109/TVCG.2019.2944182
12. Felsner, S., Reuter, K.: The linear extension diameter of a poset. SIAM J. Discret. Math. **12**(3), 360–373 (1999). https://doi.org/10.1137/S0895480197326139
13. Ganter, B.: Diskrete Mathematik: Geordnete Mengen. Springer-Lehrbuch, Springer Spektrum Berlin, Heidelberg, 1 edn. (2013). https://doi.org/10.1007/978-3-642-37500-2

14. Ganter, B., Glodeanu, C.V.: Ordinal Factor Analysis. In: Domenach, F., Ignatov, D.I., Poelmans, J. (eds.) ICFCA 2012. LNCS (LNAI), vol. 7278, pp. 128–139. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-29892-9_15
15. Ganter, B., Wille, R.: Formal Concept Analysis. Springer, Heidelberg (1999). https://doi.org/10.1007/978-3-642-59830-2
16. Glodeanu, C.V.: Tri-ordinal factor analysis. In: Cellier, P., Distel, F., Ganter, B. (eds.) Formal Concept Analysis, 11th International Conference, ICFCA 2013, Dresden, Germany, May 21–24, 2013. Proceedings. Lecture Notes in Computer Science, vol. 7880, pp. 125–140. Springer (2013). https://doi.org/10.1007/978-3-642-38317-5_8
17. Glodeanu, C.V., Ganter, B.: Applications of ordinal factor analysis. In: Cellier, P., Distel, F., Ganter, B. (eds.) Formal Concept Analysis, 11th International Conference, ICFCA 2013, Dresden, Germany, May 21–24, 2013. In: Proceedings. Lecture Notes in Computer Science, vol. 7880, pp. 109–124. Springer (2013). https://doi.org/10.1007/978-3-642-38317-5_7
18. Glodeanu, C.V., Konecny, J.: Ordinal factor analysis of graded data. In: Glodeanu, C.V., Kaytoue, M., Sacarea, C. (eds.) ICFCA 2014. LNCS (LNAI), vol. 8478, pp. 128–140. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07248-7_10
19. Golumbic, M.C.: The complexity of comparability graph recognition and coloring. Computing **18**(3), 199–208 (1977). https://doi.org/10.1007/BF02253207
20. Hirth, J., Horn, V., Stumme, G., Hanika, T.: Ordinal motifs in lattices. CoRR abs/2304.04827 (2023). arXiv:2304.04827
21. Keprt, A.: Algorithms for Binary Factor Analysis. Ph.D. thesis (2006)
22. Keprt, A., Snásel, V.: Binary factor analysis with help of formal concepts. In: Snásel, V., Belohlávek, R. (eds.) Proceedings of the CLA 2004 International Workshop on Concept Lattices and their Applications, Ostrava, Czech Republic, September 23–24, 2004. CEUR Workshop Proceedings, vol. 110. CEUR-WS.org (2004)
23. Kuznetsov, S.O., Obiedkov, S.A.: Comparing performance of algorithms for generating concept lattices. J. Exp. Theor. Artif. Intell. **14**(2–3), 189–216 (2002). https://doi.org/10.1080/09528130210164170
24. Lindig, C.: Fast concept analysis. Work. Conceptual Struct.-Contrib. ICCS **2000**, 152–161 (2000)
25. McConnell, R.M., Spinrad, J.P.: Linear-time transitive orientation. In: Saks, M.E. (ed.) Proceedings of the Eighth Annual ACM-SIAM Symposium on Discrete Algorithms, 5–7 January 1997, New Orleans, Louisiana, USA, pp. 19–25. ACM/SIAM (1997)
26. Pearson, K.: LIII. on lines and planes of closest fit to systems of points in space. London, Edinburgh, Dublin Philosophical Mag. J. Sci. **2**(11), 559–572 (1901). https://doi.org/10.1080/14786440109462720
27. Spangenberg, N., Wolff, K.E.: Comparison of biplot analysis and formal concept analysis in the case of a repertory grid. In: Bock, H.H., Ihm, P. (eds.) Classification, Data Analysis, and Knowledge Organization, pp. 104–112. Springer, Berlin Heidelberg, Berlin, Heidelberg (1991)