



HAL
open science

Feature Independence from the Point of View of Formal Concept Analysis

Alexandre Bazin

► **To cite this version:**

Alexandre Bazin. Feature Independence from the Point of View of Formal Concept Analysis. 2024.
hal-04561745

HAL Id: hal-04561745

<https://hal.science/hal-04561745>

Preprint submitted on 27 Apr 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Feature Independence from the Point of View of Formal Concept Analysis

Paper #2419

Abstract. Measuring the dependence of two features/variables in a dataset is a problem that finds applications in most sciences. It is generally either based on the probability theoretic definition of independence or done by evaluating how much a feature is a particular function of the other. In this paper, we introduce a definition of independence in formal concept analysis, a lattice theoretic framework, and we investigate whether it can be leveraged to measure the independence of numerical features. We exploit the connections between binary relations and algebraic and logical structures at the heart of formal concept analysis to propose three measures and we evaluate their potential using synthetic feature selection problems.

1 Introduction

Deciding whether features, or variables, are independent is one of the most important problem in science as it appears in most sciences that work with data. Many approaches have been proposed over the years to measure independence or association/correlation. They all either compare the observed data with what would be observed if the probability-theoretic definition of independence ($P(A, B) = P(A)P(B)$) applied or they try to evaluate whether one feature is a particular function (or class of function) of the other. However, many different approaches exist based on joint cumulative distribution functions and ranks [1, 3], pairwise distances [6, 9], kernels [8], copulas [4] and information theory [11, 15].

In this paper, we introduce a new definition of independence in formal concept analysis (FCA), a mathematical framework that aims at structuring the concepts and regularities in binary relations, and use it to build independence measures. Our proposed independence measures are distances between the co-occurrence relation of the values of features in the dataset and the FCA definition of independence. Our goal is not to outperform existing independence measures but to show that our measures sometimes recognise patterns others do not. In turn, we aim to show that this new definition of independence based on conceptual and logical structures instead of probabilities and functions is promising. We evaluate our proposition on numerical features from synthetic nonlinear regression datasets.

In Section 2, we provide the required FCA definitions. In Section 3, we present our definition of independence in the FCA framework under factual, conceptual and logical points of view. In Section 4, we briefly discuss how to apply our proposed definition to measuring the independence of numerical features. In Section 5, we present experimental results showing the our measures are usable. Finally, in the conclusion, we discuss the potential of our proposal and future work.

2 Formal Concept Analysis: Definitions

In this section, we present the necessary definitions of formal concept analysis [7]. Formal concept analysis (FCA) is a mathematical framework based on lattice theory that aims at structuring the information found in data. The data is assumed to take the form of a binary relation called a formal context.

Definition 1 (Formal context). A (formal) context is a triple $(\mathcal{O}, \mathcal{A}, \mathcal{R})$ in which \mathcal{O} is a set of objects, \mathcal{A} is a set of attributes and $\mathcal{R} \subseteq \mathcal{O} \times \mathcal{A}$ is an incidence relation between objects and attributes. When $(o, a) \in \mathcal{R}$ we say that the object o is described by the attribute a .

Formal contexts can be represented by crosstables, as illustrated in Fig. 1.

	a_1	a_2	a_3	a_4
o_1	×		×	×
o_2	×	×		
o_3			×	
o_4	×	×		

Figure 1. A formal context with four objects and four attributes.

A context gives rise to two *derivation operators*, both denoted by \cdot' . The first maps an object set to the set of attributes they have in common, i.e. $O' = \{a \in \mathcal{A} \mid \forall o \in O, (o, a) \in \mathcal{R}\}$, and the second maps an attribute set to the set of objects they all describe, i.e. $A' = \{o \in \mathcal{O} \mid \forall a \in A, (o, a) \in \mathcal{R}\}$. The compositions of both derivation operators form closure operators, denoted by \cdot'' . These derivation and closure operators are then used in the definition of the patterns that give the framework its name: concepts.

Definition 2 (Formal concept). A formal concept is a pair (E, I) where E is an object set and I is an attribute set such that $E = I'$ and $I = E'$. Hence, $E = E''$ and $I = I''$. We call E the extent and I the intent of the concept.

Formal concepts correspond to maximal rectangles of crosses, up to permutation of rows and columns, in the crosstable representation of the context. For instance, in the Fig. 1 example, the pair $(\{o_2, o_4\}, \{a_1, a_3\})$ is a concept while $(\{o_3\}, \{a_3\})$ is not because o_1 could be added to the first component.

The inclusion relation on the extents (or intents) of concepts induces a partial order \leq : $(E_1, I_1) \leq (E_2, I_2)$ iff $E_1 \subseteq E_2$ (or $I_2 \subseteq I_1$). This corresponds to the subsumption relation between concepts. As per the basic theorem of FCA, the set of all concepts in a context ordered in such a way forms a complete lattice called a

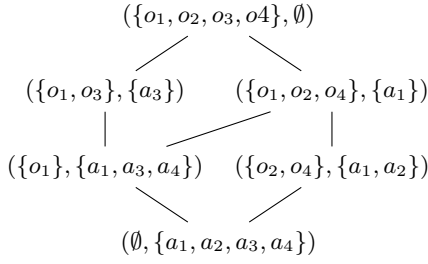


Figure 2. The concept lattice of the Fig. 1 context.

concept lattice. The concept lattice of the Fig. 1 example is depicted in Fig 2.

There is no loss of information between a context and its concept lattice. Indeed, one can reconstruct the context using the lattice.

Through the closure operator on attributes, one can also obtain regularities between the descriptions of objects called implications.

Definition 3 (Implication). An implication is a rule of the form $A \rightarrow B$, where A and B are attribute sets. It is said to hold in a context when all objects described by the objects in A are also described by the objects in B . In other words, $A \rightarrow B$ holds when $B \subseteq A''$.

The support of an implication $A \rightarrow B$ is A' . For instance, the following implications hold in the Fig 1 example: $\{a_2\} \rightarrow \{a_1, a_2\}$ (support o_2, o_4), $\{a_3, a_4\} \rightarrow \{a_1, a_3, a_4\}$ (support o_1). The following implication, however, does not hold: $\{a_2\} \rightarrow \{a_4\}$.

The set of all implications that hold in a context can be used to retrieve the intents of the associated concept lattice. This makes implications a close representation of the information in the context, with the identity of the objects as the only loss.

A formal context, its concept lattice and its corresponding set of implications are three points of views on the same phenomenon. The context is the factual point of view as it only contains observations. The concept lattice is the conceptual point of view as it is the ordered/algebraic structure of all ‘‘concepts’’ that exist in the observations. The set of implications is the logical point of view as it encapsulates the regularities in the observations in logical implications through which one can make inferences. Note that, here, we equate the closure to an inference.

3 Measuring Independence in Formal Concept Analysis

Formal concept analysis lends itself to an intuitive notion of independence between the sets of objects and attributes. In this section, we define this independence and propose three ways to measure it, corresponding to the three main structures in FCA.

3.1 Factual Independence

Let us consider the three formal contexts depicted in Fig. 3.

	a_1	a_2	a_3		a_1	a_2	a_3		a_1	a_2	a_3
o_1			×	o_1		×	×	o_1	×	×	×
o_2		×		o_2	×	×	×	o_2	×	×	×
o_3	×			o_3	×	×		o_3	×	×	×

Figure 3. Three formal contexts.

In the first one, the three objects each have a different attribute. As such, the presence of an attribute in a description completely depends

on which object is considered. Dually, the presence of an object in the derivation of an attribute also depends on the attribute. In the second one, all the objects share the attribute a_2 and all the attributes describe the object o_2 . However, the presence of a_1 and a_3 in a description still depends on which object is considered: one describes o_1 but not o_3 while the other describes o_3 but not o_1 ! The only case in which the presence of an object (resp. attribute) is independent of the attributes (resp. objects) is when all objects are described by all attributes. This corresponds to the third context, which represents our definition of independence in FCA.

Definition 4 (Independent context). In a formal context $(\mathcal{O}, \mathcal{A}, \mathcal{R})$ the sets of objects and attributes are independent when $\mathcal{R} = \mathcal{O} \times \mathcal{A}$.

We then quantify the independence of the object and attribute sets of a given context \mathcal{C} with a distance measure between \mathcal{C} and the independent context. As a measure of distance between contexts, we use a distance between the incidence relations as sets.

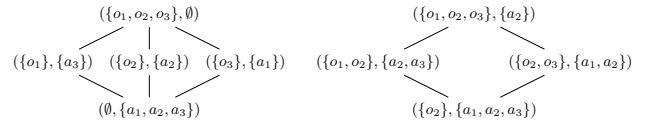
Definition 5 (Factual independence). In a formal context $(\mathcal{O}, \mathcal{A}, \mathcal{R})$, the factual independence of the object set \mathcal{O} and the attribute set \mathcal{A} is defined as

$$I_f(\mathcal{O}, \mathcal{A}) = 1 - \frac{|\mathcal{R}|}{|\mathcal{O} \times \mathcal{A}|}.$$

The factual independence takes its values in $[0, 1)$. It is 0 when objects and attributes are independent ($\mathcal{R} = \mathcal{O} \times \mathcal{A}$) and it is maximal when the incidence relation is a bijection.

3.2 Conceptual Independence

The concept lattice of the independent context consists of only one concept $(\mathcal{O}, \mathcal{A})$. By contrast, the first two formal contexts in Fig. 3 produce the following two concept lattices:



Definition 6 (Independent concept lattice). In the concept lattice \mathcal{L} of a formal context $(\mathcal{O}, \mathcal{A}, \mathcal{R})$ the sets of objects and attributes are independent when $\mathcal{L} = (\{(\mathcal{O}, \mathcal{A})\}, \emptyset)$.

Independence, from a conceptual point of view, is achieved when the world described by the context contains a single all-encompassing concept. The existence of other concepts is the mark of dependence, with more dependence resulting in more specific concepts. To quantify this dependence, we use a measure of distance between concept lattices. We consider concept lattices as representations of the derivation operators \cdot' that map each object set O (resp. attribute set A) to an attribute set O' (resp. object set A'). The independent lattice then maps all O to the set of all the attributes. To compute the independence measure, for each object o , we compare the number of attribute sets that have o in their derivations in the actual lattice ($|\{o'\}|$) and the independent lattice ($2^{|\mathcal{A}|}$). We then do the same for the attributes and keep the minimum distance.

Definition 7 (Conceptual independence). In a formal context $(\mathcal{O}, \mathcal{A}, \mathcal{R})$, the conceptual independence of the object set \mathcal{O} and the attribute set \mathcal{A} is defined as $I_c(\mathcal{O}, \mathcal{A}) =$

$$\max\left(\frac{\sqrt{\sum_{o \in \mathcal{O}} (2^{|\mathcal{A}|} - 2^{|\{o'\}|})^2}}{\sqrt{|\mathcal{O}|}}, \frac{\sqrt{\sum_{a \in \mathcal{A}} (2^{|\mathcal{O}|} - 2^{|\{a'\}'|})^2}}{\sqrt{|\mathcal{A}|}}\right).$$

147 The conceptual independence takes its values in $[0, 1)$. It is 0 when
 148 all the objects are described by all the attributes, i.e. there is only one
 149 concept $(\mathcal{O}, \mathcal{A})$, and it is maximal when $|\{o\}'| = |\{a\}'| = 1$ for all
 150 object o and attribute a , i.e. when the incidence relation is a bijection.

151 The conceptual independence can be visualised as a signal s on the
 152 powerset lattice of objects, such that $s(\mathcal{O}) = \frac{|\{o\}'|}{|\mathcal{A}|}$. For instance,
 153 for our three example contexts in Fig 3, we would obtain the drawings
 154 in Fig 4 by representing each object set by a dot and coloring it
 155 according to the strength of the signal (white is 0 and black is 1). The
 156 blacker the lattice, the more independent the dimensions.

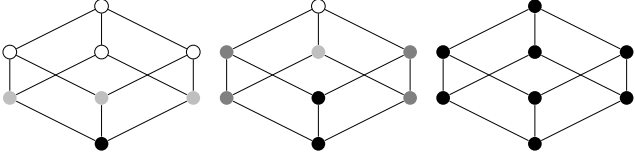


Figure 4. Conceptual independence visualised as a signal on a Boolean lattice.

157 3.3 Logical Independence

158 In the independent formal context, all the implications hold and have
 159 non-empty supports. By contrast, in the first context, the implications
 160 $\{a_1\} \rightarrow \{a_2, a_3\}$, $\{a_2\} \rightarrow \{a_1, a_3\}$ and $\{a_3\} \rightarrow \{a_1, a_2\}$
 161 do not hold and the implication $\{a_1, a_2\} \rightarrow \{a_3\}$ has the empty set as
 162 support.

163 **Definition 8** (Independent implication set). *In the set of implications*
 164 *with non-empty supports \mathcal{I} of a formal context $(\mathcal{O}, \mathcal{A}, \mathcal{R})$ the sets of*
 165 *objects and attributes are independent when $\mathcal{I} = \{A \rightarrow \mathcal{A} \mid A \subseteq$*
 166 *$\mathcal{A}\}$.*

167 Independence, from a logical point of view, is achieved when all
 168 implications hold. As such, dependence is marked by those implica-
 169 tions that do not hold. We consider implications as representations of
 170 the closure operator \cdot'' that maps each attribute set A to the attribute
 171 set A'' . The independent implication set then maps all attribute sets
 172 to \mathcal{A} . To compute the independence measure, for each attribute a , we
 173 compare the number of attribute sets that imply a in the actual impli-
 174 cation set \mathcal{I} ($|\mathcal{I}_a|$) and in the independent implication set ($|2^{\mathcal{A}}|$).

175 **Definition 9** (Logical independence). *In a formal context*
 176 *$(\mathcal{O}, \mathcal{A}, \mathcal{R})$, let \mathcal{I}_a denote the set of attributes sets X such that*
 177 *$X \rightarrow \{a\}$ and define analogously \mathcal{I}_o in the context in which objects*
 178 *and attributes are swapped. The logical independence of the object*
 179 *set \mathcal{O} and the attribute set \mathcal{A} is then defined as $I_l(\mathcal{O}, \mathcal{A}) =$*

$$\max\left(\frac{\sqrt{\sum_{a \in \mathcal{A}} (2^{|\mathcal{A}|} - |\mathcal{I}_a|)^2}}{\sqrt{\mathcal{A}}}, \frac{\sqrt{\sum_{o \in \mathcal{O}} (2^{|\mathcal{O}|} - |\mathcal{I}_o|)^2}}{\sqrt{\mathcal{O}}}\right).$$

175 The logical independence takes its values in $[0, 1)$. It is 0 when all
 176 the attributes imply all the attributes and it is maximal when $|\{o\}'| =$
 177 $|\{a\}'| = 1$ for all object o and attribute a , i.e. when the incidence
 178 relation is a bijection.

179 4 Application to the Independence of Numerical Features

180 We now want to apply the FCA definition of independence to the
 181 measurement of the independence of numerical dataset features.
 182

183 Let X and Y be two numerical features. We construct a formal
 184 context such that the values X takes in the dataset are the objects,
 185 the values Y takes are the attributes and a pair of values is in the
 186 incidence relation if and only if they co-occur. Fig. 5 depicts a small
 187 example with two features X and Y and the corresponding context.
 188 If we were to compute one of the previously defined independence
 189 measures in this context, X and Y would be deemed completely de-
 190 pendent.

X	Y		17.9	8.0	17.4	7.9
1.32	17.9	1.32	×			
1.31	8.0	1.31		×		
2.08	17.4	2.08			×	
2.09	7.9	2.09				×

Figure 5. Toy dataset with two numerical features X and Y , and the formal context of their co-occurrence relation.

191 Another option would be to first discretise the features, for instance
 192 as depicted in Fig 6. In this case, X and Y would be deemed
 193 completely independent. Thus, the partitioning of the values of the
 194 features is crucial as it completely changes the result.

X	Y		≤ 8.0	≥ 17.4
≤ 1.32	≥ 17.4	≤ 1.32	×	×
≤ 1.32	≤ 8.0	≥ 2.08	×	×
≥ 2.08	≥ 17.4			
≥ 2.08	≤ 8.0			

Figure 6. Toy dataset with the two discretised features X and Y , and the formal context of their co-occurrence relation.

195 5 Experimental Evaluation

196 In order to evaluate whether the FCA definition of independence can
 197 be leveraged to measure the independence of numerical features, we
 198 compared our proposed measures to existing independence and cor-
 199 relation measures on the problem of recognising relevant features in
 200 synthetic regression datasets.

201 5.1 Measures

202 As discussed in Section 4, our proposed measure behave differently
 203 depending on the partitioning of the values of the features. To evalu-
 204 ate this difference, we use three versions of our three measures in
 205 which the values are partitioned into 15, 20 and 25 clusters using
 206 a k-means algorithm. We respectively call these measures I_f^{15} , I_f^{20} ,
 207 I_f^{25} , I_c^{15} , I_c^{20} , I_c^{25} , I_l^{15} , I_l^{20} and I_l^{25} .

208 On the side of existing measures, we selected the classic *pear-*
 209 *son's* and *spearman's* coefficients, the *mutual information* [10], the
 210 *f_regression* measure implemented in the scikit-learn Python library
 211 and the recently introduced ξ coefficient [2].

212 5.2 Synthetic Datasets and Recognition Problem

213 We generated three kinds of synthetic datasets for Friedman's nonlin-
 214 ear regression problems [5]. In those datasets, the input features X_i
 215 are independent and uniformly distributed on $[0, 1]$ and the output y
 216 is a function of some of the input features plus some Gaussian noise
 217 $N(\mu = 0, \sigma = 1)$, the other features being irrelevant. In the three
 218 kinds of datasets, respectively called *Friedman1*, *Friedman2* and
 219 *Friedman3*, the outputs are such that

- *Friedman1*:

$$y = 10\sin(\pi X_1 X_2) + 20(X_3 - 0.5)^2 + 10X_4 + 5X_5 + N$$

- *Friedman2*:

$$y = (X_1^2 + (X_2 X_3 - 1/(X_2 X_4))^2)^{0.5} + N$$

- *Friedman3*:

$$y = \arctan((X_2 * X_3 - 1/(X_2 * X_4))/X_1) + N$$

A good independence measure should be able to recognise the input features that were involved in the creation of the output. We constructed each dataset with $4k$ input features, where k is the number of relevant features in the dataset (5 in *Friedman1* and 4 in the others). Our experiments also compare the measures on their resistance to the number of samples and to noise. To do this, we generated three versions of each dataset containing 200, 2000 and 10000 samples and, for each one, multiplied the noise by 0, 2, 10 and 100.

For each dataset and each measure, we ranked the input features according to the dependence of the output to them and kept the top- k . The measures are evaluated using two metrics:

- the precision, defined by the number of relevant features in the top- k over the total number of relevant features:

$$\frac{|\text{relevant features identified}|}{k}$$

- a ranking distance, produced by the measure and the ideal ranking in which the relevant features are scored above the irrelevant ones:

$$\frac{|\{(X_i, X_j) \mid j \leq k, \text{score}(X_i) \geq \text{score}(X_j)\}|}{3k^2}$$

5.3 Experimental Results

5.3.1 Precision

Table 1 presents the precision values obtained by each measure on each version of each dataset. Higher precisions (in bold) are better. We observe that our measures perform well on the *Friedman1* and *Friedman3* datasets regardless of noise and number of samples. However, on the *Friedman2* dataset, the performances are much lower. Classical measures are more regularly efficient but these results show that our measures can be used and sometimes discover dependence relations that elude existing approaches. For instance, the logical independence using 20 clusters perfectly recognised the four relevant features in the most noisy *Friedman3* with 10000 samples where the second best measures only found one.

5.3.2 Ranking Distance

Table 2 presents the ranking distances obtained by each measure on each version of each dataset. Lower distances (in bold) are better. The results are similar to the results on the precision, with the *Friedman2* dataset being harder for measures. However, we observe that our measures produce worse rankings on the *Friedman3*. Our interpretation is that, when our measure fail to recognise dependence, they fail harder than classical measures. The distances 1 found in the *Friedman3* datasets mean that all features have been ranked the same, i.e. they have all been deemed completely independent of the output.

6 Discussion and Conclusion

We proposed new independence measures based on the definition of independence in formal concept analysis that respect Rényi's B , C and D axioms [14]. Our aim was not to outperform other measures in any way but to propose an alternative to existing approaches. The experimental results show that there are cases in which our measures are just as good, and sometimes slightly better, at recognising dependent numerical features in a regression dataset. This suggests that there is a place for these measures as they are able to detect relations that may elude others approaches. As such, we believe that they should be used in complement to more traditional independence measures and correlation coefficients in the analysis of data. In addition, it is potentially interpretable, as exemplified in Fig 4. As with much of artificial intelligence and data science, the interpretability of dependence measures is a pressing topic [13].

More importantly, our measures offer a new point of view on independence based not on probability distributions and functions but on the co-occurrence relation of values, and the concepts and logical rules that it contains. These results show that this relation alone can be leveraged to detect independence and, thus, the question of the contribution of this co-occurrence relation compared to that of the frequencies of values arises. Future work shall be done on taking the frequency of each co-occurring pair of values into account in the FCA structures in order to both improve the measures and study the impact of both visions. Other future work will include a study of the interpretation of the values taken by the measures so as to propose a threshold to decide independence and a evaluation of the trust one can have in this decision, analogous to p-values.

Lastly, we intend to generalise our measures using to polyadic concept analysis [16], a generalisation of FCA in which there are more than two dimensions in the context. This will allow us to detect n -ary and conditional independencies, opening the way to a study of the use of these measures in the construction of symbolic "Bayesian networks" and causal inference [12].

References

- [1] W. Bergsma and A. Dassios. A consistent test of independence based on a sign covariance related to kendall's tau. *Bernoulli*, pages 1006–1028, 2014.
- [2] S. Chatterjee. A new coefficient of correlation. *Journal of the American Statistical Association*, 116(536):2009–2022, 2021.
- [3] N. Deb and B. Sen. Multivariate rank-based distribution-free nonparametric testing using measure transportation. *Journal of the American Statistical Association*, 118(541):192–207, 2023.
- [4] H. Dette, K. F. Siburg, and P. A. Stoimenov. A copula-based nonparametric measure of regression dependence. *Scandinavian Journal of Statistics*, 40(1):21–41, 2013.
- [5] J. H. Friedman. Multivariate adaptive regression splines. *The annals of statistics*, 19(1):1–67, 1991.
- [6] J. H. Friedman and L. C. Rafsky. Graph-theoretic measures of multivariate association and prediction. *The Annals of Statistics*, pages 377–391, 1983.
- [7] B. Ganter and R. Wille. *Formal Concept Analysis - Mathematical Foundations*. Springer, 1999. ISBN 978-3-540-62771-5. doi: 10.1007/978-3-642-59830-2. URL <https://doi.org/10.1007/978-3-642-59830-2>.
- [8] A. Gretton, K. Fukumizu, C. Teo, L. Song, B. Schölkopf, and A. Smola. A kernel statistical test of independence. *Advances in neural information processing systems*, 20, 2007.
- [9] R. Heller, Y. Heller, and M. Gorfine. A consistent multivariate test of association based on ranks of distances. *Biometrika*, 100(2):503–510, 2013.
- [10] A. Kraskov, H. Stögbauer, and P. Grassberger. Estimating mutual information. *Physical review E*, 69(6):066138, 2004.
- [11] E. H. Linfoot. An informational measure of correlation. *Information and control*, 1(1):85–89, 1957.

Dataset	Noise	Samples	I_f^{15}	I_f^{20}	I_f^{25}	I_c^{15}	I_c^{20}	I_c^{25}	I_l^{15}	I_l^{20}	I_l^{25}	pearson	spearman	mutual info	ξ	f_{reg}
Friedman1	0	200	0.6	0.8	0.4	0.4	0.4	0.2	0.4	0.4	0.6	0.8	0.8	0.8	0.6	0.8
		2000	1	1	1	0.8	1	0.8	0.8	0.6	0.6	0.8	0.8	0.8	0.6	0.8
		10000	1	1	1	1	1	1	1	1	1	0.6	0.8	0.8	1	0.8
	2	200	0.8	0.4	0.4	0.6	0.6	0	0.2	0.2	0.6	0.8	0.8	0.8	0.4	0.2
		2000	1	1	0.8	1	1	0.6	0.4	0.4	0.4	0.8	0.8	0.8	0.6	0.8
		10000	1	1	1	1	1	1	1	1	0.6	0.8	0.8	1	0.8	
	10	200	0.4	0.4	0.2	0.2	0	0.4	0.4	0.2	0.6	0.6	0.6	0.4	0.6	
		2000	0.6	0.2	0.6	0.8	0.2	0.4	0.4	0.4	0.4	0.8	0.8	0.6	0.4	
		10000	0.2	0.8	0.8	0.2	0.8	0.4	0.2	0.4	0.8	0.8	0.8	1	0.6	
	100	200	0.4	0.4	0	0.2	0.4	0.2	0.2	0.2	0.6	0.4	0.2	0.2	0.4	
		2000	0	0.4	0.6	0.2	0.2	0.4	0.2	0.2	0.2	0.4	0.4	0.2	0.4	
		10000	0.2	0.2	0.2	0.2	0	0	0.2	0.8	0.4	0.6	0.6	0.4	0.2	
Friedman2	0	200	0.75	0.5	0.5	0.5	0	0	0.25	0.25	0.5	0.5	0.5	0.5	0.5	
		2000	0.75	0.75	0.75	0.75	0.75	0.5	0.5	0.25	0.5	1	1	0.75	0.5	
		10000	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	1	0.75	1	1	0.75	
	2	200	0.5	0.5	0.5	0.5	0	0.5	0.5	0	0.5	0.5	0.5	0.75	0.5	
		2000	0.75	0.75	0.5	0.75	0.75	0.75	0.5	0.25	0.5	1	1	0.75	0.5	
		10000	0.5	0.75	0.75	0.5	0.75	0.75	0.5	1	0.75	0.75	1	0.75	0.75	
	10	200	0.75	0.5	0.5	0.5	0.25	0.25	0.5	0	0.75	0.5	0.5	0.75	0.5	
		2000	0.75	0.75	0.75	0.5	0.5	0.5	0.75	0.5	0.5	1	1	0.75	0.5	
		10000	0.5	0.5	0.75	0.5	0.5	0.75	0.5	1	0.5	0.75	1	0.75	0.75	
	100	200	0.75	1	0.5	0.5	0.25	0	0.25	0	0.75	0.5	0.5	0.75	0.5	
		2000	0.5	0.75	0.5	0.5	0.5	0.5	0.75	0.25	0.25	1	1	0.75	0.5	
		10000	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.25	0.75	0.75	0.5	0.5	
Friedman3	0	200	0.25	0.5	0.25	0.25	0.25	0.5	0.25	0.25	0.5	0.5	0.5	0.5	0.25	
		2000	0.25	0.5	0.75	0.25	0.5	0.25	0.5	0.5	0.75	0.75	0.75	0.25		
		10000	0.5	0.5	0.75	0	0.25	0.25	0	0.75	0.5	0.75	0.75	0.5		
	2	200	0.25	0.5	0.75	0.25	0.5	0.25	0.25	0.25	0.75	0	0	0.25	0	
		2000	0.5	0.25	0.25	0.25	0.5	0.25	0.5	0.25	0.25	0.5	0.5	0.75	0.5	
		10000	0	0.25	0.25	0	0.25	0.25	0	0.5	0.5	1	1	0.75	0.25	
	10	200	0.5	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0	0.25	
		2000	0.25	0.25	0.25	0.25	0	0.25	0	0.5	0.5	0.5	0.5	0.5	0.25	
		10000	0	0.75	0.25	0	0.75	0.25	0	0.75	0.25	0.75	0.75	0	0.5	
	100	200	0	0.25	0.25	0.25	0.25	0.25	0	0.25	0.25	0.25	0	0	0.25	
		2000	0.5	0.25	0.5	0.25	0.25	0.5	0.25	0	0.75	0	0	0.25	0	
		10000	0	0.25	0.25	0	0	0.25	0	1	0	0	0.25	0	0	

Table 1. Precision values obtained by each measure for each dataset.

320 [12] J. Pearl. From bayesian networks to causal networks. In *Mathematical*
321 *models for handling partial knowledge in artificial intelligence*, pages
322 157–182. Springer, 1995.

323 [13] M. Reimherr and D. L. Nicolae. On quantifying dependence: A frame-
324 work for developing interpretable measures. *Statistical Science*, pages
325 116–130, 2013.

326 [14] A. Rényi. On measures of dependence. *Acta mathematica hungarica*,
327 10(3-4):441–451, 1959.

328 [15] D. N. Reshef, Y. A. Reshef, H. K. Finucane, S. R. Grossman,
329 G. McVean, P. J. Turnbaugh, E. S. Lander, M. Mitzenmacher, and P. C.
330 Sabeti. Detecting novel associations in large data sets. *science*, 334
331 (6062):1518–1524, 2011.

332 [16] G. Voutsadakis. Polyadic concept analysis. *Order*, 19:295–304, 2002.

Dataset	Noise	Samples	I_f^{15}	I_f^{20}	I_f^{25}	I_c^{15}	I_c^{20}	I_c^{25}	I_l^{15}	I_l^{20}	I_l^{25}	pearson	spearman	mutual info	ξ	f_{reg}
Friedman1	0	200	0.18	0.22	0.29	0.25	0.28	0.56	0.32	0.22	0.10	0.09	0.09	0.53	0.30	0.09
		2000	0	0	0	0.02	0	0.01	0.06	0.12	0.16	0.17	0.18	0.02	0.09	0.17
		10000	0	0	0	0	0	0	0	0	0.17	0.08	0.13	0	0	0.08
	2	200	0.12	0.42	0.58	0.30	0.29	0.42	0.24	0.33	0.16	0.12	0.13	0.01	0.17	0.12
		2000	0	0	0.06	0	0.01	0.09	0.14	0.54	0.26	0.17	0.18	0.13	0.09	0.17
		10000	0	0	0	0	0	0	0	0.13	0.24	0.13	0.12	0	0	0.13
	10	200	0.36	0.36	0.60	0.57	0.72	0.45	0.18	0.54	0.14	0.18	0.22	0.38	0.32	0.18
		2000	0.32	0.60	0.28	0.18	0.54	0.36	0.45	0.30	0.49	0.05	0.02	0.12	0.36	0.05
		10000	0.80	0.18	0.13	0.80	0.13	0.26	0.80	0.62	0.12	0.09	0.09	0	0.13	0.09
	100	200	0.52	0.50	0.76	0.33	0.37	0.52	0.57	0.62	0.29	0.49	0.52	0.82	0.48	0.49
		2000	0.52	0.42	0.25	0.52	0.46	0.30	0.62	0.38	0.62	0.26	0.22	0.41	0.33	0.26
		10000	0.82	0.85	0.52	0.82	0.8	0.61	0.82	0.84	0.41	0.21	0.22	0.48	0.73	0.21
Friedman2	0	200	0.22	0.08	0.29	0.35	0.72	0.64	0.37	0.54	0.27	0.37	0.29	0.12	0.31	0.37
		2000	0.25	0.20	0.08	0.25	0.06	0.18	0.41	0.60	0.35	0	0	0.06	0.14	0
		10000	0.25	0.18	0.18	0.25	0.18	0.18	0.25	0.37	0.29	0.04	0	0	0.10	0.41
	2	200	0.25	0.25	0.37	0.20	0.70	0.35	0.27	0.75	0.35	0.37	0.29	0.10	0.31	0.37
		2000	0.22	0.25	0.25	0.22	0.25	0.08	0.31	0.70	0.45	0	0	0.10	0.12	0
		10000	0.50	0.14	0.18	0.50	0.14	0.18	0.50	0.14	0.29	0.41	0	0.41	0.10	0.41
	10	200	0.10	0.20	0.29	0.18	0.33	0.56	0.12	0.58	0.20	0.37	0.35	0.25	0.33	0.37
		2000	0.16	0.29	0.27	0.20	0.29	0.33	0.25	0.35	0.37	0	0	0.04	0.12	0
		10000	0.50	0.33	0.20	0.50	0.33	0.20	0.50	0.33	0.31	0.41	0	0.25	0.14	0.41
	100	200	0.08	0.04	0.10	0.14	0.35	0.70	0.58	0.54	0.08	0.45	0.37	0.14	0.14	0.45
		2000	0.33	0.14	0.31	0.29	0.22	0.37	0.39	0.56	0	0	0	0.27	0.08	0
		10000	0.50	0.50	0.43	0.50	0.50	0.41	0.50	0.50	0.64	0.08	0.41	0.37	0.12	0.08
Friedman3	0	200	0.58	0.27	0.43	0.27	0.27	0.33	0.39	0.66	0.27	0.45	0.41	0.37	0.33	0.45
		2000	0.58	0.22	0.20	0.50	0.25	0.33	0.54	0.35	0.18	0.18	0.18	0.25	0.35	0.18
		10000	0.50	0.43	0.27	0.50	0.47	0.33	0.50	0.45	0.47	0.14	0.14	0.33	0.20	0.14
	2	200	0.54	0.45	0.20	0.39	0.56	0.52	0.56	0.56	0.06	0.54	0.56	0.81	0.41	0.54
		2000	0.64	0.56	0.37	0.75	0.50	0.35	0.50	0.43	0.56	0.27	0.35	0.31	0.54	0.27
		10000	1	0.50	0.58	1	0.45	0.54	1	0.68	0.37	0	0	0.25	0.60	0
	10	200	0.45	0.54	0.60	0.60	0.47	0.60	0.54	0.60	0.43	0.72	0.70	0.58	0.37	0.72
		2000	0.77	0.62	0.58	0.75	0.64	0.37	0.72	0.29	0.27	0.31	0.29	0.39	0.58	0.31
		10000	1	0.22	0.77	1	0.16	0.68	1	0.62	0.64	0.08	0.08	0.79	0.33	0.08
	100	200	0.68	0.66	0.47	0.43	0.47	0.64	0.62	0.68	0.39	0.79	0.77	0.68	0.50	0.79
		2000	0.43	0.39	0.41	0.39	0.33	0.31	0.58	0.54	0.27	0.70	0.64	0.39	0.47	0.70
		10000	1	0.54	0.62	1	0.60	0.54	1	0.37	0.75	0.54	0.62	0.75	0.62	0.54

Table 2. Ranking distances obtained by each measure for each dataset.