Analysis of social communities with iceberg and stability-based concept lattices

Nicolas Jay, François Kohler, Amedeo Napoli

Analysis of Social Communities with Iceberg and Stability-Based Concept Lattices

Nicolas Jay[1,2], François Kohler[2], and Amedeo Napoli[1]

[1] Laboratoire Lorrain de Recherche en Informatique et ses Applications, Campus Scientifique - BP 239 - 54506 Vandoeuvre-lès-Nancy Cedex, France
jay@loria.fr
http://www.loria.fr/equipes/orpailleur
[2] Laboratoire SPI-EAO,
Faculté de Médecine - 9 Avenue de la Forêt de Haye -BP 184 54505 Vandoeuvre Cedex, France

Abstract. In this paper, we presents a research work based on formal concept analysis and interest measures associated with formal concepts. This work focuses on the ability of concept lattices to discover and represent special groups of individuals, called social communities. Concept lattices are very useful for the task of knowledge discovery in databases, but they are hard to analyze when their size become too large. We rely on concept stability and support measures to reduce the size of large concept lattices. We propose an example from real medical use cases and we discuss the meaning and the interest of concept stability for extracting and explaining social communities within a healthcare network.

1 Introduction

Knowledge Discovery in Databases (KDD) is an iterative and interactive process for identifying valid, novel, and potentially useful patterns in data [1]. It is usually divided into three main steps: data preparation, data mining, and interpretation of the extracted units. Data mining is often considered as the central step in the KDD process. However, interpretation of data-mining results is also an important step within the KDD process. Indeed, one of the the success keys in KDD practice relies on the ability of easily producing units understandable as knowledge units. One way of achieving such a goal is to provide an adapted organization and representation of the extracted units, especially when the KDD system has to be used by novice users.

In parallel, Formal Concept Analysis (FCA) is a theory of data analysis introduced in [2], that is tightly connected with KDD [3, 4], particularly regarding the search of frequent itemsets and the extraction of association rules [5]. Many algorithms relying on FCA central property of closure have been proposed to extract frequent closed itemsets: e.g. CLOSE [6], CLOSET [7], CHARM [8], TITANIC [9], and ZART [10]. The set of frequent closed itemsets may be used to determine the set of all frequent itemsets: closed itemsets are a loss less representation of frequent itemsets, while the set of closed itemsets can be orders of magnitude smaller than the set of all frequent itemsets.

R. Medina and S. Obiedkov (Eds.): ICFCA 2008, LNAI 4933, pp. 258–272, 2008.

FCA organizes information into a concept lattice representing inherent structures existing in data. A concept lattice can be visualized for analysis using graphical tools, e.g. TOSCANA [11], GALICIA [12]. FCA forms also the basis of a knowledge processing paradigm known as “Conceptual Landscapes” [13]. Furthermore, Stumme [9] has introduced the so-called iceberg lattices, which are concept lattices of frequent closed itemsets. Iceberg lattices serve as a support for visualization of association rules mined in large database. They can help analysts in selecting interesting patterns and organizing these patterns into understandable and reusable knowledge units. However, iceberg lattices may hide non frequent but still relevant concepts.

Following the idea of compact, reduced (loss less), and concise representation of extracted units (i.e. itemsets, association rules, or concepts), a number of numerical measures used for pruning itemsets, association rules, and in a certain sense concepts, have been proposed [14]. In this way, Kuznetsov has introduced stability as a new interest measure for concepts [15, 16]. Stability has been successfully used for pruning concept lattices, e.g. in the field of social networks [17–19]. Accordingly, in this article, we address the problem of exploring “social communities”. By “social communities”, we intend sets of agents or organizations whose members are linked by a common interest or objective [20]. One of our goals is to study the basis and to design a decision support system for assisting experts identifying social communities. The selection, organization, and discrimination of relevant units of knowledge, help to understand how agents interact in a social community and how they gather on specific topics. Moreover, we show in this paper that combining concept frequency together with concept stability provides a very efficient means for discovering and analyzing social communities.

The paper is organized as follows. Following the present first section, the second section introduces the definitions and the properties of FCA, of support and stability measures. The third section presents a qualitative discussion on stability and shows how stability enlighten concept with a high internal cohesion, i.e. stable and without exceptional individuals. Then, the fourth section gives details on an example of social community discovery within a healthcare network. A discussion on the example and on the knowledge units that can be extracted is proposed and precedes the conclusion of the paper.

2 Support and Stability: Interest Measures of Formal Concepts

2.1 Formal Concept Analysis

We describe here the FCA basics. FCA starts with a formal context $\mathbb{K} = (\mathtt{G}, \mathtt{M}, \mathtt{I})$ where $\mathtt{G}$ is a set of objects, $\mathtt{M}$ is a set of attributes, and the binary relation $\mathtt{I} = \mathtt{G} \times \mathtt{M}$ specifies which objects have which attributes. Two operators, both denoted by $'$, connect the power sets of objects $2^{\mathtt{G}}$ and attributes $2^{\mathtt{M}}$ as follows:

$$' : 2^{\mathtt{G}} \rightarrow 2^{\mathtt{M}}, \mathtt{X}' = \{\mathtt{m} \in \mathtt{M} | \forall \mathtt{g} \in \mathtt{X}, \mathtt{gIm}\}$$

The operator $'$ is dually defined on attributes. The pair of $'$ operators induces a Galois connection between $2^{\mathtt{G}}$ and $2^{\mathtt{M}}$. The composition operators $''$ are closure operators: they are idempotent, extensive and monotonous. For any $\mathtt{A} \subseteq \mathtt{G}$ and $\mathtt{B} \subseteq \mathtt{M}$, A'' and B'' are closed sets whenever $\mathtt{A} = \mathtt{A}''$ and $\mathtt{B} = \mathtt{B}''$.

A formal concept of the context $\mathbb{K} = (\mathtt{G}, \mathtt{M}, \mathtt{I})$ is a pair $(\mathtt{A}, \mathtt{B}) \subseteq \mathtt{G} \times \mathtt{M}$ where $\mathtt{A}' = \mathtt{B}$ and $\mathtt{B}' = \mathtt{A}$. A is called the *extent* and B is called the *intent*. A concept $(\mathtt{A}_1, \mathtt{B}_1)$ is a *subconcept* of a concept $(\mathtt{A}_2, \mathtt{B}_2)$ if $\mathtt{A}_1 \subseteq \mathtt{A}_2$ (which is equivalent to $\mathtt{B}_2 \subseteq \mathtt{B}_1$) and we write $(\mathtt{A}_1, \mathtt{B}_1) \leq (\mathtt{A}_2, \mathtt{B}_2)$. The set $\mathfrak{B}$ of all concepts of a formal context $\mathbb{K}$ together with the partial order relation $\leq$ forms a lattice and is called concept lattice of $\mathbb{K}$.

2.2 Iceberg Concept Lattices

This paragraph is based on [9] and introduces basics of iceberg lattices.

Definition 1. *Let* $\mathtt{B} \subseteq \mathtt{M}$. *The support count of the attribute set* $\mathtt{B}$ *in* $\mathbb{K}$ *is*

$$\sigma(\mathtt{B}) = \frac{|\mathtt{B}'|}{|\mathtt{G}|} \tag{1}$$

Let $\mathtt{minsupp}$ be a threshold $\in [0, 1]$, then $\mathtt{B}$ is said to be a frequent itemset if $\sigma(\mathtt{B}) \geqslant \mathtt{minsupp}$.

A concept is called frequent concept if its intent is frequent.

Definition 2. *The set of all frequent concepts of a context* $\mathbb{K}$ *is called iceberg lattice of the context* $\mathbb{K}$.

The support function is monotonously decreasing: given two attribute sets $\mathtt{B}_1$ and $\mathtt{B}_2$, $\mathtt{B}_1 \subseteq \mathtt{B}_2 \Rightarrow \sigma(\mathtt{B}_1) \geq \sigma(\mathtt{B}_2)$. Thus an iceberg lattice is an order filter of the whole concept lattice and in general only a join-semi-lattice. Meanwhile, adding a bottom element makes it a lattice again.

Iceberg Lattices can be used to discover and visualize association rules. Within a formal context $\mathtt{K} = (\mathtt{G}, \mathtt{M}, \mathtt{I})$, the task of mining association rules is to determine all pairs $\mathtt{X} \rightarrow \mathtt{Y}$ of $\mathtt{M}$ such that $\sigma(\mathtt{X} \rightarrow \mathtt{Y}) = \sigma(\mathtt{X} \cup \mathtt{Y}) \geqslant \mathtt{minsupp}$, and the *confidence* $\mathtt{conf}(\mathtt{X} \rightarrow \mathtt{Y}) = \frac{\sigma(\mathtt{X} \cup \mathtt{Y})}{\sigma(\mathtt{X})}$ is above a given threshold $\mathtt{minconf} \in [0, 1]$.

Mining associations rules with FCA has two major advantages [21]. First, frequent closed itemsets are sufficient to deduce all frequent itemsets. Thus, algorithms can benefit from this property to reduce the search space. Second, iceberg lattices offer a reduced and lossless representation of association rules. They allow to directly read Luxenbourger basis for approximate association rules [22] from a line diagram.

2.3 Stability

Stability has been introduced (probably for the first time) in [15] and then revisited [16, 19]. Here, we rely on the definition given in [19].

Definition 3. *Let* $(\mathtt{A},\mathtt{B})$ *a formal concept of* $\mathfrak{B}(\mathbb{K})$*. Stability of* $(\mathtt{A},\mathtt{B})$ *is*

$$\gamma(\mathtt{A},\mathtt{B}) = \frac{|\{\mathtt{C} \subseteq \mathtt{A} | \mathtt{C}' = \mathtt{A}' = \mathtt{B}\}|}{2^{|\mathtt{A}|}} \tag{2}$$

The stability index of a concept indicates how much the concept intent depends on particular objects of the extent. Given a concept $(\mathtt{A},\mathtt{B})$, the stability index measures the number of elements of $\mathtt{G}$ that are in the same *equivalence class* of $\mathtt{A}$, where an equivalence class is defined as follows.

Definition 4. *Let* $\mathtt{X} \subseteq \mathtt{G}$*, we denote by* $\langle \mathtt{X} \rangle$ *the* equivalence class *of* $\mathtt{X}$ *where:*

$$\langle \mathtt{X} \rangle = \{\mathtt{Y} \subseteq \mathtt{G} | \mathtt{Y}' = \mathtt{X}'\} \tag{3}$$

Note that when $\mathtt{X}$ is closed, any $\mathtt{Y}$ in $\langle \mathtt{X} \rangle$ is a subset of $\mathtt{X}$. Thus, considering a formal concept $(\mathtt{A},\mathtt{B})$, definition 3 can be rewritten as:

$$\gamma(A,B) = \frac{|\langle A \rangle|}{2^{|A|}} \tag{4}$$

Then, the larger the equivalence class of an extent is (wrt to extent size), the more stable the concept is. The idea behind stability is that a stable concept is likely to have a real world interpretation even if the description of some its objects (i.e. elements in the extent) is "noisy". Figure 1 shows an example of stability in a concept lattice. Each concept is labelled by its extent, intent and stability. For example, for the concept $(\{\mathtt{1},\mathtt{5},\mathtt{6}\},\{\mathtt{a}\})$, we have:

$$\begin{aligned}
\emptyset' = \{\mathtt{a},\mathtt{b},\mathtt{c},\mathtt{d}\} &\neq \{\mathtt{a}\} \\
\{\mathtt{1}\}' &= \{\mathtt{a}\} \\
\{\mathtt{5}\}' &= \{\mathtt{a}\} \\
\{\mathtt{6}\}' = \{\mathtt{a},\mathtt{b},\mathtt{c}\} &\neq \{\mathtt{a}\} \\
\{\mathtt{1},\mathtt{5}\}' &= \{\mathtt{a}\} \\
\{\mathtt{1},\mathtt{6}\}' &= \{\mathtt{a}\} \\
\{\mathtt{5},\mathtt{6}\}' &= \{\mathtt{a}\} \\
\{\mathtt{1},\mathtt{5},\mathtt{6}\}' &= \{\mathtt{a}\}
\end{aligned}$$

Thus $\gamma(\{\mathtt{1},\mathtt{5},\mathtt{6}\},\{\mathtt{a}\}) = \frac{\mathtt{6}}{\mathtt{8}} = \mathtt{0.75}$. It can be noticed that stability is (by definition) always between 0 and 1. It can be still noticed that $\gamma(\bot) = \mathtt{1}$. This is always true, since for any subset $\mathtt{X}$ from the extent of $\bot$, $\mathtt{X}'$ is included in the intent of $\bot$.

Computing stability has been shown to be a #P-complete problem [16]. Meanwhile, once the concept lattice has been computed, a bottom-up traversal algorithm can efficiently compute stability [18]. Actually, a concept stability depends on the stability of its subconcepts. This can be shown as follows:

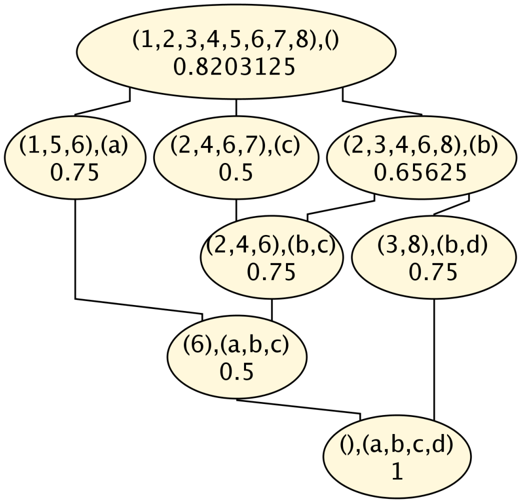


Fig. 1. Stability example

Proposition 1. *Let* $(\mathtt{A},\mathtt{B})$ *a formal concept of* $\mathfrak{B}(\mathbb{K})$*.*

$$\gamma(\mathtt{A},\mathtt{B}) = 1 - \sum_{\mathtt{X}\subset\mathtt{A},\mathtt{X}=\mathtt{X}''} \gamma(\mathtt{X},\mathtt{X}')2^{|\mathtt{X}|-|\mathtt{A}|} \tag{5}$$

Proof. For a formal concept $(\mathtt{A},\mathtt{B})$, from (4), we have:

$$\gamma_{(\mathtt{A},\mathtt{B})} = \frac{|\langle\mathtt{A}\rangle|}{2^{|\mathtt{A}|}}$$

Let $\mathtt{I}_\mathtt{A}$ be the set of subintents of $\mathtt{A}$: $\mathtt{I}_\mathtt{A} = \{\mathtt{X} \subseteq \mathtt{A} | \mathtt{X} = \mathtt{X}''\}$. The set of equivalent classes $\{\langle\mathtt{X}\rangle | \mathtt{X} \in \mathtt{I}_\mathtt{A}\}$ forms a partition of $2^\mathtt{A}$. Thus $|2^\mathtt{A}| = \sum_{\mathtt{X}\in\mathtt{I}_\mathtt{A}|\langle\mathtt{X}\rangle|}$, which gives:

$$|\langle\mathtt{A}\rangle| = |2^\mathtt{A}| - \sum_{\mathtt{X}\in\mathtt{I}_\mathtt{A},\mathtt{X}\neq\mathtt{A}} |\langle\mathtt{X}\rangle|$$

Dividing by $|2^\mathtt{A}|$ we obtain:

$$\frac{|\langle\mathtt{A}\rangle|}{|2^\mathtt{A}|} = 1 - \sum_{\mathtt{X}\in\mathtt{I}_\mathtt{A},\mathtt{X}\neq\mathtt{A}} \frac{|\langle\mathtt{X}\rangle|}{|2^\mathtt{A}|}$$

$$\gamma_{(\mathtt{A},\mathtt{B})} = 1 - \sum_{\mathtt{X}\subset\mathtt{A},\mathtt{X}=\mathtt{X}''} \gamma(\mathtt{X},\mathtt{X}')2^{|\mathtt{X}|-|\mathtt{A}|}$$

3 A Qualitative Analysis of Stability

3.1 Stability and Cohesion

As stated in [19], a concept is stable if its intent does not depend much Âon each particular object of the extent. Stability is aimed at measuring how much a concept extent depends on some of its individual members. This may be useful in analyzing a dataset with a concept lattice, having a special attention to social communities. Here, a social community can be thought as a group of agents –human, software, or resource agents– sharing the same interests, or ideas, or needs [17]. For example, patients visiting the same hospitals with similar medical problems can be identified as a special social community. In the associated formal context, objects correspond to patients and hospital stays correspond to attributes (see hereafter). Always following the line of [19], an actual community has to be "internally cohesive" enough: a stable concept continues to be a concept even if a few members stop being members. This means also that a stable concept is resistant to noise and will not collapse when some members will be removed from its extent.

In this way, a stable concept is a meaningful concept, in the sense that it covers a group of objects, that considered together, have a high internal cohesion. The most stable concepts determine the most interesting groups of objects, that constitute their extents.

3.2 Stable Concepts Are of High Interest

FCA provides a powerful framework for identifying social communities [17, 23, 24]. Relations between agents and common interests can be modeled within a formal context. The associated concept lattice will allow to discover and identify which agents do share common interests and what are these interests. However, as the size of a formal context increases, the number of formal concepts in the lattice may grow dramatically. In this case, interest measures such as stability and support can reduce the complexity of the analysis of the concept lattice. Filtering concepts by support relies on the assumption that useful knowledge is represented by frequent patterns. But, this is not always true as pointed out in studies on rare itemsets, as in e.g. in [25], where it is shown that association rules with a low support but a high confidence may be of high interest.

Stability gives an alternative point of view on formal concepts. It indicates the probability of preserving a concept intent while removing some objects of its extent. Considering social communities, stability helps to identify groups of commons interest that dot not entirely depend on some specific agents. As stability is somewhat independent from support, it can be used to discriminate low-support concepts and detect small communities of strongly related agents. Moreover, stability also detects frequent concepts only depending on a small number of objects. For example, considering a lattice composed of the two following concepts $C_1 = (\{g_1, ..., g_{n-1}, g_n\}, \{m_1\})$ and $C_2 = (\{g_1, ..., g_{n-1}\}, \{m_1, m_2\})$ with n high, then it can be noticed that C_1 depends solely on the object g_n. Although C_1 and C_2 have both a support close to 1, stability of C_2 is 1 while stability of C_1 is $\frac{1}{2}$. In terms of social communities, the group of individuals $\{g_1, ..., g_n\}$ has not a sufficient "internal cohesion" or has not a "real existence".

Hence, stability, together with support, are a convenient means for identifying two types of concepts:

– *rare stable concepts*: concepts with a low support and a high stability,
– *frequent unstable concepts*: concepts with a high support and low stability.

In section 4, this point of view is discussed and illustrated within a real-world application aimed at detecting communities of patients, i.e. groups of patients being treated in the same groups of hospitals.

3.3 Stable Concepts Are Monothetic Rather Than Polythetic

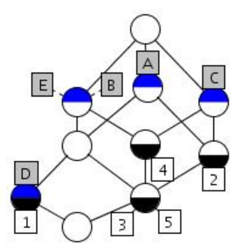
As introduced above, stability can also be linked with exceptions, and, furthermore, with the so-called *monothetic* and *polythetic* characters of a class of individuals [26–28]. When building a concept lattice and analyzing groups of individuals through the extents of concepts, one problem is to recognize and explain *exceptions*. A subsequent question is to understand whether exceptions are linked to monothetic or polythetic classes.

A class of individuals `C` is said to be *monothetic* if and only if there exists a set of attributes `Att` that determines the membership of an individual to the class `C` (`Att` is a set of necessary and sufficient membership conditions). By contrast, given a set of attributes $\texttt{Att} = \{\texttt{a}_1, ..., \texttt{a}_\texttt{n}\}$, a class of individuals `C` is said to be *polythetic* if and only if:

– Every object that is an instance of the class `C` has an "important" –not necessarily fixed– number of attributes of `Att`.
– Every attribute of `Att` belongs to an "important" –not necessarily fixed– number of instances of `C`.
– There is not necessarily an attribute of `Att` belonging to every instance of `C`.

Relying on the fact that a stable concept has a high "internal cohesion", is resistant to noise, and does not collapse when some members stop being members of its extent, the more a concept is stable, the more it does not represent exceptional individuals, and, accordingly, the more the concept is able to represent cohesive groups of individuals, such as social communities. This means that stable concepts are rather monothetic and that unstable concepts are rather "exceptional" or polythetic, i.e. they include some exceptional character, shared by only a few individuals. For illustrating this view, let us consider the following example.

	A	B	C	D	E
1	x	x		x	x
2	x		x		
3	x	x	x		x
4		x	x		x
5	x	x	x		x



Here, attribute `D` can be considered as a "necessary and sufficient condition" for the membership of an individual to the class including individual **1**. Indeed, the concept lattice includes the concept $(\mathtt{A}, \mathtt{B}) = (\mathtt{1}, \mathtt{abde})$. However, this concept has a very low stability ($\frac{4}{16}$) and also a low support for its intent ($\frac{1}{5}$ when abde is considered as an itemset). This is in agreement with the view that a stable concept appears to be monothetic while an unstable concept tends to be polythetic.

4 Social Communities in a Regional Healthcare System

4.1 Motivations

Healthcare management and planning play a key role for improving the overall health level of the population. From a population point of view, even the best and state-of-the-art therapy is not effective if it cannot be delivered in the right conditions. Actually, many determinants influence the effective delivery of healthcare services: availability of trained personnel, availability of equipments, security constraints, costs, proximity... All of these should meet economics, demographics, and epidemiological needs in a given area. This issue is especially acute in the field of cancer care where many institutions and professionals must cooperate to deliver high level, long term, and costly care. Therefore, it is crucial for healthcare managers and decisions makers to be assisted by decision support systems that give strategic insights about the intrinsic behavior of the healthcare system.

On the one hand, healthcare systems can be considered as "data rich" as they produce massive amounts of data such as electronic medical records, clinical trial data, hospital records, administrative data, and so on. On the other hand, they can be regarded as "knowledge poor" as these data are rarely embedded into a strategic decision-support resource [29]. In France, the PMSI database is a national information system used to describe hospital activity with both an economical and medical point of view. In a previous work, we used this system together with iceberg lattices to discover how several institutions organize themselves into an implicit network to provide coordinated care at a regional level [30]. Our method has been used in real world by healthcare managers. It appeared that support based pruning had some limits, for example in analysing small institutions interactions.

4.2 The Difficulty of Choosing the Good Support Threshold

In this section we present an example of an iceberg lattice showing cooperations between hospitals in the field of cancer. We then discuss the choice of `minsupp` by studying concept support distribution. In our approach, we build a context in which objects are patients suffering from cancer and attributes are hospitals. A patient and a hospital are related if the hospital has delivered cancer care to this patient. In our experiment, the resulting context has 6036 patients and 170 hospitals. While the whole concept lattice holds 865 concepts, an iceberg lattice built with a `minsupp` of 0.0033 (20 patients) gives 93 frequent concepts. A small excerpt of this iceberg is shown in figure 2.

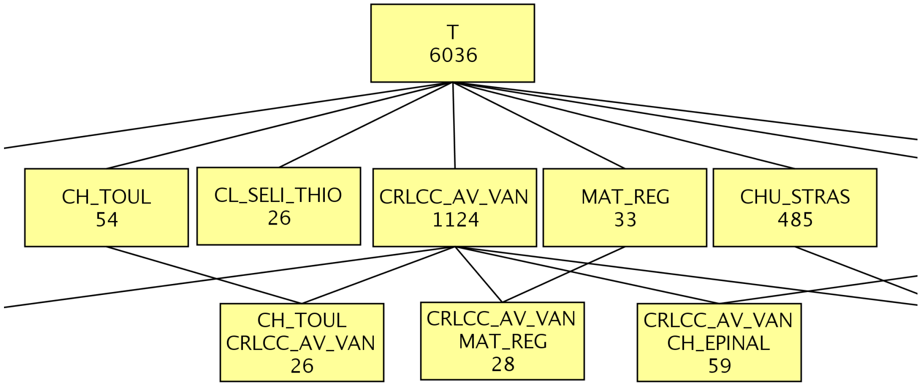


Fig. 2. Iceberg of cancer treatment cooperations

For the sake of clarity, $\bot$ was removed. Although its right and leftmost parts are not drawn, the iceberg lattice is much more wide than deep because the context is sparse and data are poorly correlated. This means that cooperations are most of the time tightly partitioned, and that patients are rarely hospitalized in more than two hospitals. The intent of co-atoms, i.e. immediate descendants of $\top$, is always a singleton, indicating that a hospital never shares all of its patients with another one, or if it is so, less than 20 patients are involved in the interaction. The intent of atoms, i.e. the immediate ascendant of $\bot$, is always a pair. The extent of atoms gives an idea of the strength of the cooperation between the two hospitals lying in the intent: the larger is the cardinal of the extent, the higher is the strength of the cooperation (i.e. the more patients are shared between the two hospitals). The examination of the iceberg brings different types of knowledge:

– some concepts that are both atoms and co-atoms (for example : `CL-SELI-THIO`). They represent institutions that share a few patients with others. This is that either they treat a few patients, or they work in a relative autonomy, or cooperation is split with many other hospitals.
– other concepts have at least a sub-concept (different from $\bot$). They represent a hospital receiving a significant number of patients, and having collaborations with at least another establishment. For example `MAT-REG` and `CLCC-AV-VAN` share 28 patients.
– The concept representing the `CLCC-AV-VAN` hospital has a high support and many sub-concepts. This hospital is a specialized anti-cancer center. It employs highly skilled and specialized personnel. Treatments given there rely on state-of-art technology. Furthermore, It actively participates in anti-cancer research programs and thus can be considered as a reference institution.

The choice of the minimum support strongly influences the interpretation of the iceberg. It must be sufficiently low to convey meaningful knowledge and sufficiently high to keep this knowledge readable for a human expert. Here, the

Table 1. Concept support count and intent size

Number of patients \ Intent size	0	1	2	3	> 4
< 10	0	40	285	297	76
$[10, 20[$	0	13	52	7	0
$[20, 30[$	0	10	20	0	0
$[30, 40[$	0	6	2	0	0
> 40	1	34	20	0	0

whole lattice holds 865 concepts. Table 1 shows the distribution of their support according to the size of their intent. By choosing a *minsupp* of corresponding to 20 patients, we see we can miss interesting knowledge:

- 13 hospitals treat at least 10 patients,
- 52 cooperations of two hospitals involve at least 10 patients,
- 7 concepts represent cooperations between three hospitals and involve at least 10 patients.

In our case, support is a weak mean for discriminating "not-so-frequent" concepts. Some concepts not appearing in the iceberg may be of interest for different reasons:

- They illustrate a cooperation of one hospital sharing almost all of its patients with another one.
- They concern a hospital treating few patients but not sharing them with any other one.
- They concern a 3-hospitals interaction.

Lowering the support threshold can let these concepts appear in the iceberg, but at the expense of readability. Moreover support measure is not specific enough to discriminate concepts with the above characteristics.

4.3 Stability Analysis

In this section, we study stability of concepts within the whole concept lattice. Figure 3 shows the histogram of concept stability.

Most of the concepts hold only one object in their extent. As $\perp$ extent is empty, they have a stability of 0.5. The next important group consists of concepts having stability very close to 1. It corresponds generally to the most frequent concepts. Indeed, 93% of the concepts having support greater than 10 have stability greater than 0.99. Figure 4 is a scatter-plot of stability and support, also featuring intent size.

Values are presented on a log-scale for better visualization of both support and stability ranges. Moreover, we prefer to use the stability odds defined as follows:

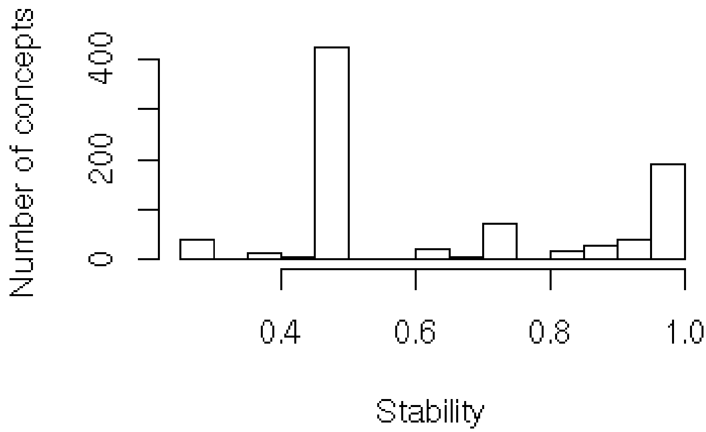


Fig. 3. Concept stability histogram

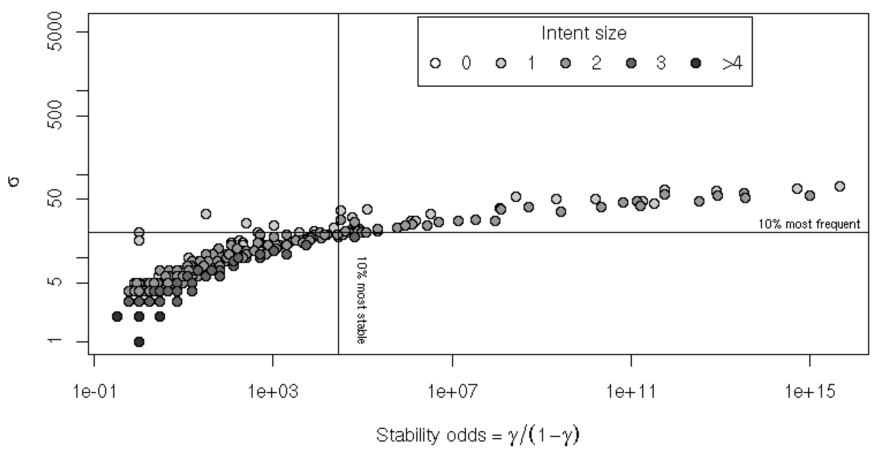


Fig. 4. Concept stability and support

Definition 5. *Given a concept* $(\mathtt{A},\mathtt{B}) \neq \bot$*, we call stability odds of* $(\mathtt{A},\mathtt{B})$*:*

$$\mathtt{o}_\gamma(\mathtt{A},\mathtt{B}) = \frac{\gamma(\mathtt{A},\mathtt{B})}{1-\gamma(\mathtt{A},\mathtt{B})} \tag{6}$$

Stability odds are the ratio between the number of subsets of an extent $\mathtt{A}$ *which belong to the equivalence class of* $\mathtt{A}$*, and of those which do not.*

Stability odds illustrate more clearly the distribution of stability for values close to 1 and show that stability has a better discriminant power than support. The whole set of concepts has 87 distinct support values and 152 distinct stability values. For concept having support between 10 and 19, we observe 57 distinct stability values. Figure 4 displays thresholds for the 10% most frequent concepts and 10% most stable concepts. In the next sections we discuss the differences between those two thresholds.

4.4 Frequent Unstable Concepts

The upper left quarter of figure 4 shows the 9 concepts having support greater than the 10% most frequent threshold and stability less than the 10% most stable threshold, i.e. frequent unstable concepts. They represent institutions playing a significant role in cancer care as secondary or tertiary care[1] centers. They share most of their patients with other hospitals. This gives rise to cooperations of 2 kinds.

Figure 5 shows the order ideals of two frequent unstable concepts. Two hospitals have tight interaction each with another tertiary care center, i.e. `MAT-REG-NANCY` with `CRLCC-AV-VAN`, `CL-ARCENCIEL` with `CH-EPINAL`. These two institutions are secondary care centers usually delivering surgery and referring patients to a tertiary care center for radiotherapy. Their activity can be almost entirely explained by an exclusive cooperation. This induces a form of dependency with the tertiary care center. Patient not concerned by the cooperation can be considered as exceptions, i.e. patients not following the usual care pathway for some reason (for example because they suffer from a very specific pathology). After pruning according to stability, only the concepts in grey on figure 5 will remain.

One center, as shown on Figure 6, is a highly specialized anti-cancer institution. `IGR-PARIS` is an international well-known anti-cancer center located in Paris. Many Lorraine local hospitals refer patients to `IGR` for rare tumors. If we apply stability pruning here, the whole sub-lattice will disappear, which may be desirable for readability. Meanwhile, its suggests that searching for this type of unstable frequent concept can also be an interesting knowledge mining task.

4.5 Rare Stable Concepts

Three concepts are located in the right lower corner of figure 4.

Two have an intent of size 2 and illustrate thus cooperations between two hospitals sharing 18 patients. These cooperations differ from others of the same support in that they are not split themselves in cooperations involving a third institution. Thus, similarity of concerned patients is entirely explained by those cooperations. The last concept has a size 3 intent and illustrates the cooperation between three large specialized hospitals located in the same city of Nancy. This is the most frequent and most stable concept of that type.

Stability allows to distinguish rare concepts that cannot be separated from others by support. Rare stable concepts differ from other rare concepts in that their attributes suffice to explain the similarity of their objects.

[1] Secondary care is delivered by a broadly skilled specialist (e.g. a general surgeon, a general internist, or an obstetrician). Tertiary care is provided by a sub-specialist(e.g. an orthopedic surgeon, a neurologist, or neonatologist).

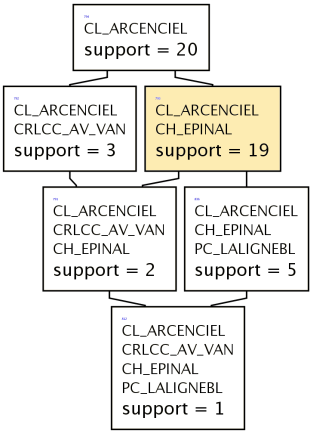


Fig. 5. Frequent unstable concepts: exclusive cooperations

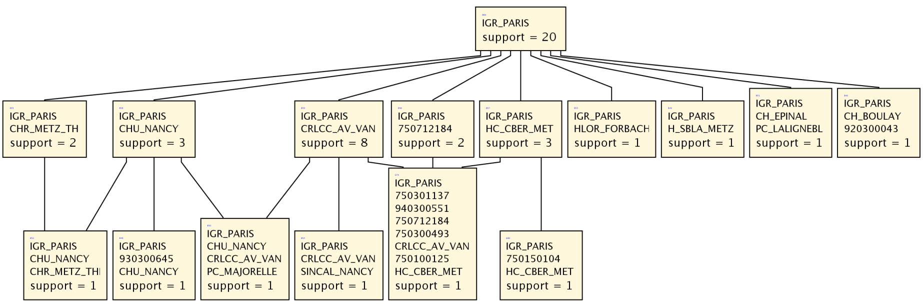


Fig. 6. Frequent unstable concepts : `IGR-PARIS` as a national referred care center

5 Synthesis

Our system is used in real world by healthcare managers. First, at the individual level, it helps healthcare professionals to assess their activity in the regional landscape. While physicians are able to cite the names of people they are used to cooperate with, they cannot measure the strength of these cooperations. And it is even harder for hospital managers to count patients shared with other institutions due to the gaps and lacks of adapted processes in information systems. Second, at the regional level, it provides for the administrative staff a decision support to reorganize care resources according to the implicit behavior of the healthcare system. Actually, French law establishes activity thresholds in the field of cancer: e.g. an hospital must treat at least 30 patients a year to be authorized for digestive cancer surgery. Our system has allowed to enlighten and accordingly to promote cooperations between institutions that could not reach the thresholds alone. It has also demonstrated how administrative decisions could impact the social healthcare network, given the existence of many dependencies between structures.

6 Conclusion

Our main objective is to build a system allowing for visualization of a social healthcare network in the field of cancer care. This system is used today by healthcare managers. Things must be kept simple while conveying enough information for assisting strategic decisions. The use of concept interest measures has a strong impact both on readability and semantics of discovered knowledge. Together with support, stability can successfully identify two kind of concepts: frequent unstable concepts and rare stable concepts. In our experiment, the formal context is sparse and we need to mine concepts with very low support. Stability brings additional knowledge that helps to discover interesting rare concepts that can not be discriminated by support. We believe that it could have significant implications in the field of rare itemsets mining [25]. Furthermore, stability enhances lattice visualization when pruning frequent unstable concepts. Besides, frequent unstable concepts may also be a subject of interest.

References

1. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P.: The kdd process for extracting useful knowledge from volumes of data. Communication of the ACM 29(11), 27–34 (1996)
2. Wille, R.: Restructuring lattice theory: an approach based on hierarchies of concepts. In: Rival, I. (ed.) Ordered Sets, Reidel (1982)
3. Wille, R.: Why can concept lattices support knowledge discovery in databases? J. Exp. Theor. Artif. Intell. 14(2-3), 81–92 (2002)
4. Valtchev, P., Missaoui, R., Godin, R.: Formal concept analysis for knowledge discovery and data mining: The new challenges. In: Eklund, P.W. (ed.) ICFCA 2004. LNCS (LNAI), vol. 2961, pp. 352–371. Springer, Heidelberg (2004)
5. Agrawal, R., Imielski, T., Swami, A.: Mining association rules between sets of items in large databases. In: Proceedings of the ACM SIGMOD Int'l Conference on Management of Data, pp. 207–216 (1993)
6. Pasquier, N., et al.: Efficient mining of association rules using closed itemset lattices. Journal of Information Systems 24, 25–46 (1999)
7. Pei, J., Han, J., Mao, R.: Closet: An efficient algorithm for mining frequent closed itemsets. In: Proceedings ACM SIGMOD Workshop DMKD 2000, Dallas, Texas, USA, May 2000, pp. 21–30 (2000)
8. Zaki, M.J., Hsiao, C.J.: Charm: An efficient algorithm for closed itemset mining. In: Grossman, R.L., et al. (eds.) SDM, SIAM, Philadelphia (2002)
9. Stumme, G.: Efficient Data Mining Based on Formal Concept Analysis. In: Hameurlain, A., Cicchetti, R., Traunmüller, R. (eds.) DEXA 2002. LNCS, vol. 2453, Springer, Heidelberg (2002)
10. Szathmary, L., Napoli, A., Kuznetsov, S.O.: ZART: A Multifunctional Itemset Mining Algorithm. In: Proc. of the 5th Intl. Conf. on Concept Lattices and Their Applications (CLA 2007), Montpellier, France (October 2007)

11. Vogt, F., Wille, R.: Toscana - a graphical tool for analyzing and exploring data. In: Tamassia, R., Tollis, I(Y.) G. (eds.) GD 1994. LNCS, vol. 894, pp. 226–233. Springer, Heidelberg (1995)
12. Valtchev, P., et al.: Galicia: an open platform for lattices. in using conceptual structures. In: Contributions to the 11th Intl. Conference on Conceptual Structures (ICCS 2003), pp. 241–254. Shaker Verlag (2003)
13. Wille, R.: Conceptual landscapes of knowledge: a pragmatic paradigm for knowledge processing. In: Gaul, W., Locarek-Junge (eds.) Classification in the Information Age, pp. 344–356. Springer, Heidelberg (1999)
14. Guillet, F., Hamilton, H.J.: Quality Measures in Data Mining (Studies in Computational Intelligence). Springer, Secaucus, NJ, USA (2007)
15. Kuznetsov, S.O.: Stability as an estimate of the degree of substantiation of hypotheses derived on the basis of operational similarity. Nauchn. Tekh. Inf., Ser.2 (Automat. Document. Math. Linguist.) 12, 21–29 (1990)
16. Kuznetsov, S.O.: On stability of a formal concept. Annals of Mathematics and Artificial Intelligence 49, 101–115 (2007)
17. Roth, C.: Co-evolution in epistemic networks – reconstructing social complex systems. Structure and Dynamics: eJournal of Anthropological and Related Sciences 1(3) (2006)
18. Roth, C., Obiedkov, S., Kourie, D.G.: Towards concise representation for taxonomies of epistemic communities. In: Yahia, S.B., Nguifo, E.M. (eds.) CLA 4th International Conference on Concept Lattices and their Applications, Tunis, Faculté des Sciences de Tunis, pp. 205–218 (2006)
19. Kuznetsov, S., Obiedkov, S., Roth, C.: Reducing the representation complexity of lattice-based taxonomies. In: Priss, U., Polovina, S., Hill, R. (eds.) ICCS 2007. LNCS (LNAI), vol. 4604, pp. 241–254. Springer, Heidelberg (2007)
20. Ahuja, M.K., Carley, K.M.: Network structure in virtual organizations. Journal of Computer-Mediated Communication 3(4) (1998)
21. Lakhal, L., Stumme, G.: Efficient mining of association rules based on formal concept analysis. In: Ganter, B., Stumme, G., Wille, R. (eds.) Formal Concept Analysis. LNCS (LNAI), vol. 3626, pp. 180–195. Springer, Heidelberg (2005)
22. Luxenbouger, M.: Implications partielles dans un contexte. Mathmatiques, Informatique et Sciences Humaines 29(113), 35–55 (1991)
23. Freeman, L.C., White, D.R.: Using galois lattices to represent network data. Sociological methodology 23, 127–146 (1993)
24. Duquenne, V.: Latticial structures in data analysis. Theoretical Computer Science 217, 407–436 (1999)
25. Szathmary, L., Napoli, A., Valtchev, P.: Towards rare itemset mining. In: Proceedings of the IEEE International Conference on Tools with Artificial Intelligence (ICTAI), Patras, Greece, IEEE Computer Society Press, Los Alamitos (2007)
26. Sokal, R.: Phenetic taxonomy: Theory and methods. Annual Review of Ecology and Systematics 17, 423–442 (1968)
27. Bailey, K.: Monothetic and polythetic typologies and their relation to conceptualization, measurement and scaling. American Sociological Review 38(1), 18–33 (1973)
28. Lerman, I.: Les bases de la classification automatique. Gauthier-Villars diteurs, Paris (1970)
29. Abidi, S.S.: Knowledge management in healthcare: towards 'knowledge-driven' decision-support services. Int. J. Med. Inform. 63(1-2), 5–18 (2001)
30. Jay, N., Kohler, F., Napoli, A.: Using formal concept analysis for mining and interpreting patient flows within a healthcare network. In: Proceedings of CLA 2006 (2006)