Knowledge Discovery and Visualization in Healthcare Datasets using Formal Concept Analysis and Graph Databases

Diana Cristea and **Christian Săcărea** and **Diana-Florina Şotropa** [1]

Abstract. Among the major advances in Artificial Intelligence we can mention Knowledge Discovery, Processing and Representation. Since in our modern society the healthcare system plays an important role and has a major impact in our daily lives, it lies at hand to apply the aforementioned methods in order to discover relevant patterns in healthcare databases and then to represent them in a way which supports reasoning, decision making, and communication. We approach this task by using two complementary directions, which are then interlinked. On the one hand we make use of the graphical representation capabilities of Formal Concept Analysis (FCA) and its powerful algorithms for conceptual knowledge discovery and processing. On the other, we use graph databases as a complementary visualization method of the extracted knowledge patterns. We exemplify this approach on a particular medical dataset, highlighting a 3D representation of conceptual hierarchies by using virtual reality (VR).

1 Introduction

Formal Concept Analysis (FCA) is a prominent field of applied mathematics which formalizes the classical philosophical understanding of a concept as a unit of thought and provides powerful algorithms for knowledge discovery, processing and representation. FCA is well known for its expressive and intuitive graphical representation of knowledge. The basic data structure is a formal context, i.e., a universe of discourse, and knowledge extraction is restricted to concepts, particular patterns which constitute building blocks of the knowledge encapsulated in the dataset. Concepts are ordered and displayed in an order diagram, called concept lattice or conceptual hierarchy. Due to its elementary yet powerful formal theory, FCA can express other methods, and therefore has the potential to unify the methodology of data analysis. Summarizing, FCA is a human-centered method to structure and analyze data, as well as a method to visualize data and its inherent structures, implications and dependencies.

How well can healthcare systems be used in order to support physicians? As researchers, we cannot stop asking what we should do in order to improve them. When trying to assemble and analyze medical data, we all have the same purpose: to aid both patients and care providers, while improving the outcomes and offering personalised care. A common approach followed in order to extract knowledge from the large amount of collected data usually starts with data preprocessing and analysis, which is then usually continued with extraction of knowledge or patterns. Once extracted, this knowledge can be used in various ways, from improvement of medical systems, to understanding and prediction issues or for learning.

This paper is presenting some current research about using FCA and graph databases to discover knowledge in healthcare databases. The data is organized and represented in conceptual landscapes of knowledge, using a methodology developed by R. Wille [22]. These conceptual landscapes of knowledge can be used, for instance to understand the way how patterns are arising from medical data, to investigate analogies between symptoms and treatments, to support communication and they can be be integrated into a decision support system that assists doctors in the process of diagnosis. One major step forward is switching from 2D to 3D using VR and establishing virtual discussion rooms where multiplayers can navigate and explore conceptual knowledge. On the other hand, graph databases are offering a different perspective. They enable us to analyze different connections between data, using a graph based approach.

The contributions of our paper include detecting and extracting knowledge patterns from healthcare data as well as presenting some visualization techniques for these, both in 2D and 3D format.

The paper is structured as follows. Section 2 describes some related work, while Section 3 contains some preliminaries, introducing the method use to extract the knowledge from the dataset, namely Formal Concept Analysis, and graph databases. Section 4 presents some experiments while trying to show how new information about medical investigations can be discovered using knowledge graphs. Section 5 concludes the work presented and highlights some future research directions.

2 Related Work

Artificial Intelligence is a wide field comprising a large set of methods and algorithms that can be applied in multiple fields. Given the nature and the importance of medicine in our lives, a large number of researchers work on applications in the medical field. A lot of the work in this field is focused on prediction models for diseases using different data mining methods. For instance, Delen et al. present a comparison of three data mining methods (logistic regression, artificial neural networks and decision trees) for predicting breast cancer survivability [2]. However, an important part of the medical system is the diagnosis of the patients. In this sub-field there is a lot of work to be done in order to build systems that can aid practitioners in their decisions. For example, one of the previous authors identifies this in a subsequent paper, where different machine learning techniques are applied to build predictive models [23]. Their conclusions are

[1] Babeş-Bolyai University, Romania, email: dianat@cs.ubbcluj.ro, csacarea@math.ubbcluj.ro, diana.halita@ubbcluj.ro

that having more information about the patients' conditions can improve models' predictive power which then can help practitioners make better diagnostic and treatment decisions.

When dealing with electronic health record (EHR) it is well known that the volume of data can easily become too large for humans to process. Therefore, the need of implementing support systems that assist clinicians in examining the data has been previously identified and acknowledged by medical experts and researchers. For instance, Fujita et al. propose in their paper to improve the user experience by limiting the visual format. They define several screen designs based on some identified principles, such as: limiting a view to a single patient data, summarizing an overview of the data and give details on demand [3]. In this approach, the advantage of having a large collection of data is lost and becomes rather a disadvantage. Such approaches lose sight of important information correlating different cases, symptoms and diagnostics, which can give an important and useful insight into the data. For that reason we believe that a better approach is to find ways of taking advantage of all the patterns contained in the data and, instead of cutting down on the data visualized, find new methods of knowledge discovery and representation techniques that allow clinicians to have an overview of the data and at the same time to be able to infer knowledge from the data, such as useful correlations and patterns.

Through FCA, medical data will be scaled so that it can be modeled as objects possessing attributes, in order to allow the discovery and to visualize of implications between them. The formal concept is the unit of measure and the central point from which the pattern mining begins. The graphical representation is given by the concept lattice containing all formal concepts.

Formal Concept Analysis can be applied in multiple fields proving that it is a suitable information retrieval technique [13]. There are also some application of FCA in the medical field. Gupta et al. use context reduction techniques along with classification rules in order to find redundancies among various medical examination tests [5]. Jay et al. use FCA for mining and interpreting patient flows within a healthcare network [7]. Pan et al. propose to use FCA in order to provide a method for modeling and designing a multidisciplinary clinical process, in which medical specialists can coordinate the treatment of specific groups of patients [12].

In our previous work, we have applied FCA techniques on different medical datasets, such as Otorhinolaryngology data, cancer registry and drug adverse reaction. For all these cases the medical datasets are considered as many-valued contexts, and they are subject to conceptual scaling in order to build knowledge landscapes. For instance, we have used several methods of conceptual knowledge processing to build a logical information system for oncological databases [1]. In some other work the scaling effort of FCA was focused on attributes describing treatment options and their results, the type and location of cancerous cells and the adverse drug reactions [16]. The databases were analyzed from different perspectives, using dyadic formal contexts as well as triadic formal contexts [6] . The triadic setting offers conditions as a third dimensions which can lead to a better understanding for instance in the case of adverse drug reactions [15].

Furthermore, we have used analogical reasoning combined with FCA in order to offer valuable support for decision making in a medical setting. The purpose of this is to improve the interaction between clinicians and electronic health record systems [19]. In some of our previous papers we started analyzing how combining different mining techniques and visualization methods, such as analogical reasoning, FCA and graph databases, can bring a fresh perspective over the medical process and improve the task of knowledge discovery in EHR systems [17, 18]. The results obtained so far highlight the fact that FCA is suitable for improving electronic health record systems.

3 Preliminaries

3.1 Formal Concept Analysis

Formal Concept Analysis (FCA) was introduced by Bernhard Ganter and Rudolf Wille in the early 1980s [4]. The theory has its mathematical basis in general lattice theory created by Garrett Birkhoff in the 1930s. One advantage of the FCA analysis techniques is that the FCA tools do not require extensive knowledge on lattice theory in order to be used and interpreted, which makes FCA a suitable and accessible method for information retrieval.

There are three kinds of relations that exist among concepts: independence, intersection and inheritance. Based on these relations, knowledge about the data can be extracted, and often causal relations can be identified [12].

We will briefly recall some definitions introduced by Rudolf Wille in [22] regarding formal concept, formal context, many-valued contexts and conceptual scaling. A formal context is a triple (G, M, I) where G and M are sets and $I \subseteq G \times M$ is a binary relation, called incidence relation. A Galois connection on the powersets of G and M respectively is defined and is used as a concept forming operator. More precisely, for $A \subseteq G$, we define $A' := \{m \in M \mid \forall a \in A, (a, m) \in I\}$, and dually for $B \subseteq M$, we define $B' := \{g \in G \mid \forall b \in B, (g, b) \in I\}$. A formal concept is a pair (A, B) with $A \subseteq G$, $B \subseteq M$, and $A' = B$, $B' = A$. Concepts are ordered by the subconcept-superconcept relation and the resulting structure is a complete lattice, called concept lattice or conceptual hierarchy, and it can be graphically represented as an order diagram. Every node of this order diagram represents a concept, while the path connecting the nodes upwards or downwards are exactly the subconcept-superconcept relation. Using a reduced labeling, only some particular concepts are labeled with the elements from G and M, respectively, more exactly those which are supremum or infimum irreducible in the lattice.

A many-valued context (G, M, W, I) consists of sets G, M, and W and a ternary relation I between G, M and W (i.e., $I \subseteq G \times M \times W$) for which it holds that $(g, m, w) \in I$ and $(g, m, v) \in I$ always implies $w = v$. The triple $(g, m, w) \in I$ is read as "the attribute m has the value w for the object g". The many-valued attributes can be regarded as partial maps from G in W. Therefore, it seems reasonable to write $m(g) = w$ instead of $(g, m, w) \in I$. In order to derive the conceptual structure of a many-valued context, we need to scale every many-valued attribute. This process is called conceptual scaling and it is always driven by the semantics of the attribute values.

A scale for the attribute m of a many-valued context is a formal context $S_m := (G_m, M_m, I_m)$ with $m(G) \subseteq G_m$. The objects of a scale are called scale values, the attributes are called scale attribute. Every context can be used as a scale. Formally there is no difference between a scale and a context. However, we will use the term "scale" only for contexts which have a clear conceptual structure and which bear meaning. The set of scales can then be used to navigate within the conceptual structure of the many-valued context (and the subsequent scaled context). Some scales are predefined (like nominally, ordinally, etc.), while for more complex views, we need to define particular scales.

3.2 Graph Databases

Graphs are data structures containing nodes with pairwise relationships between them, represented as edges. When the edge corresponds to an ordered pair of nodes, then the graph is called a directed graph, otherwise it is an undirected graph. A strongly connected component in an undirected graph is a maximal region within which each node is reachable from any other node. When defining this notion for directed graphs the direction of the edges plays an important role. Hence, we can define a strongly connected component for a directed graph as a maximal subset of nodes such that there is a directed path from any node to any other node. Strongly connected components can be very useful in an early phase of the data analysis in order to see how the graph is structured and to identify clusters of data that have similar behavior. Graph algorithms provide one of the most powerful approaches to analyzing connected data since they are relationship-oriented [11, 21].

Graphdatabases use a graph datamodel as opposed to the relational data model used in most database management systems. The main advantage of graph databases is that representing the data as a graph structure gives a more intuitive representation of the data rather than the relational structured databases or other table structures. Another reason for considering graph databases rather than the well established and widely spread relational or NoSQL data models is performance. There are use cases when a graph database is much more efficient and flexible for the implementation, mostly because a graph database can use graph-specific algorithms which in a different setting have a higher complexity [14]. The flexibility is given by the fact that the graph model is easily extensible. In contrast, when dealing with changes in a relational database one must make structural changes that can affect the existing data.

Graph databases basically consist of a labeled property graph model. In a graph database entities are represented as nodes of the graph and labels are used to express that a certain node belongs to a particular category. Nodes contain properties in form of key-value pairs. The structure of the graph is given by relationships among nodes. Relationships have a direction and a role property, i.e. a name, which together give the meaning of that relationship and show how two nodes are associated. For the implementation we used Neo4j [8] which enables us to build a knowledge graph for the analyzed dataset. Neo4j is a highly scalable and easy to manage graph database that offers an efficient query language implementation called Cypher.

4 Knowledge discovery in medical data

Medical diagnosis is regarded as an important yet difficult task that needs to be executed accurately and efficiently. Regarding accuracy, in practice, one can still find a high number of wrong diagnostics. Regarding efficiency, sometimes even if reaching the correct diagnostic, a lot of tests are performed on the patient, some of which may be irrelevant for the condition of the patient. This can be a time-consuming and costly process which can be optimized with the help of technologies that assist the doctors in their decisions. For this reason, we use FCA as a mining technique that has the potential to generate conceptual structures that can improve the quality of clinical decisions.

We are considering a collection of data from the Otorhinolaryngology department from a teaching hospital in Romania. This department is specialized in the diagnosis and treatment of ear, nose and throat disorders. The data collected for multiple patients contains information about symptoms presented by the patients and the diagnostics given by the doctors following a set of test and investigations. According to the importance of the symptoms and diagnostics, they are each divided into two categories: principal and secondary symptoms, respectively diagnostics.

Our analysis focuses on finding and visualizing patterns among different pairs of these elements, for instance analyzing correlations among principal and secondary symptoms, or among principal symptoms and principal diagnostics.

Using these data, we show how new knowledge about medical investigations can be discovered, by following 3 steps: finding concepts, finding relations between concepts and building knowledge concept lattices. Datasets are interpreted as many-valued contexts. We use `FCA Tools Bundle`[2] system ([9, 10]) to build conceptual scales and to visualize knowledge clusters.

Figure 1 reveals the correlations that exists in the dataset considering the dyadic case of diagnostics and symptoms. Due to the huge amount of data and for the purpose of the article, we have filtered our data by selecting only patients who had `Deviated Septum` and `Chronic Sinusitis` among the secondary diagnostics list. We chose to do that in order to exemplify our theory on a relatively small dataset. Afterward, we have selected diagnostics as objects and symptoms as attributes in order to build the formal context.

Deviated Septum	occurs when the thin wall (nasal septum) between your nasal passages is displaced to one side
Chronic Sinusitis	occurs when the spaces inside your nose and head (sinuses) are swollen and inflamed for three months or longer, despite treatment.
Chronic Otitis Media	describes some long-term problems with the middle ear, such as a hole (perforation) in the eardrum that does not heal or a middle ear infection (otitis media) that doesn't improve or keeps returning.
Otosclerosis	is a condition where one or more foci of irregularly laid spongy bone replace part of normally dense enchondral layer of bony otic capsule in the bony labyrinth.
Chronic Pharyngitis	is the chronic inflammation of the pharynx.

Table 1. Diagnostics from the extent of the highlighted node and their medical definitions

Autophony	the unusually loud hearing of a person's own voice
Ear Fullness	a sensation of pressure within the middle ears. This sensation is similar to the fullness we experience going up and down in airplanes or sensation of being deep under water. This pressure is horribly uncomfortable and severely distracting from everyday responsibilities and enjoyments.
Hearing Loss	deafness, or hard of hearing
Headache	the symptom of pain anywhere in the region of the head or neck

Table 2. Symptoms from the intent of the highlighted node and their medical definitions

On the generated context that can be seen in Figure 1, we have highlighted a concept having the extent {`Chronic Otitis Media`, `Otosclerosis`, `Chronic Pharyngitis`,

[2] `https://fca-tools-bundle.com/`

Chronic Sinusitis and Deviated Septum} and the intent {Autophony, Ear Fullness, Hearing Loss, Headache}. When looking at a node, the extent of the corresponding concept contains all the objects from the lattice reachable when going (only) downward. Similarly, the intent of the corresponding concept contains all the attributes from the lattice reachable when going (only) upward. Tables 1 and 2 show a detailed description of the extent and intent of the highlighted formal concept.

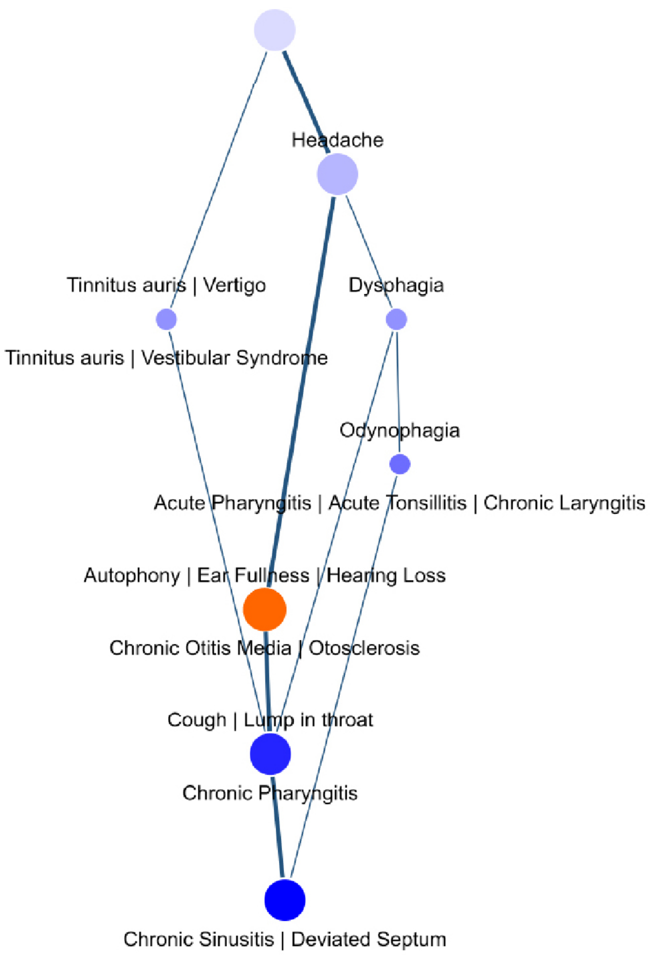

Figure 1. Deviated Septum and Chronic Sinusitis as secondary diagnostics in relation to corresponding symptoms: revealed diagnostics after investigations - LATTICE

The highlighted concept shows that all five diagnostics {Chronic Otitis Media, Otosclerosis, Chronic Pharyngitis, Chronic Sinusitis and Deviated Septum} have a set of common symptoms that need to be taken into consideration: {Autophony, Ear Fullness, Hearing Loss, Headache}. At the same time we can notice in the lattice that Chronic Otitis Media and Otosclerosis have the same symptoms, which makes the diagnosis difficult. However, three of the diagnostics, namely Chronic Pharyngitis, Chronic Sinusitis and Deviated Septum differentiate themselves by some additional symptoms, such as Cough, Lump in throat, and a few others that can be read from the lattice for each concept.

Therefore, Figure 1 highlights all the diagnostics that a physician should consider when treating a patient, together with a full list of symptoms (either principal or secondary) which may appear.

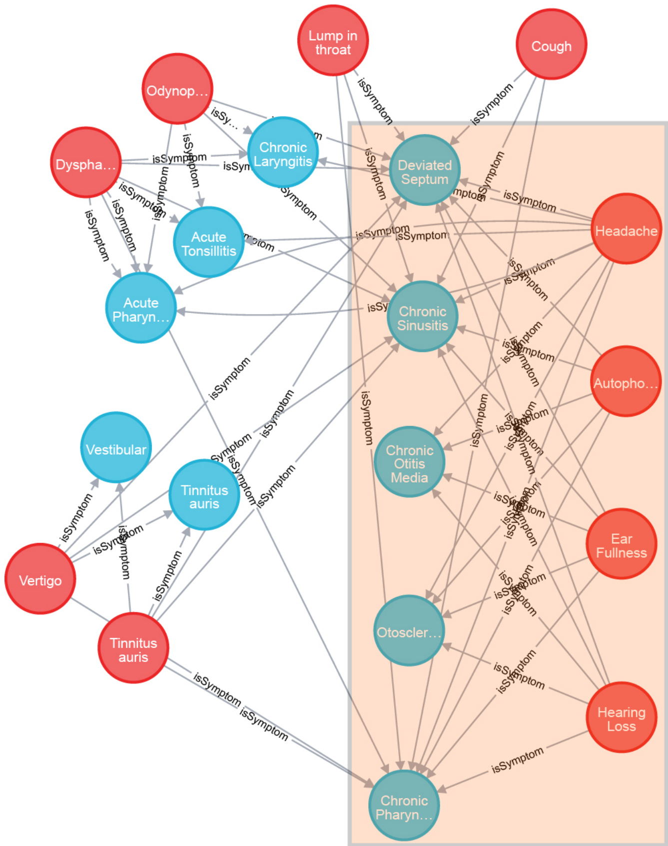

Figure 2. Deviated Septum and Chronic Sinusitis as secondary diagnostics in relation to corresponding symptoms: revealed diagnostics after investigations - NEO4j

In order to gain more information about symptoms and diagnostics, we switch our perspective to a different one, by choosing graph databases. We have processed and stored our medical datasets in the Neo4j graph database. Our purpose was to enrich our knowledge about the medical data, while analyzing different connections between data in the form of correlated nodes. The nodes of the graph correspond to the objects and attributes from the formal concept, while the binary relation is modeled as the directed edges from the graph database. Let us observe that in a formal concept the binary relation is not directed, meaning that saying an object has an attribute or that an attribute belongs to the object is exactly the same thing. However, it does not make any sense to clutter the graph by adding two type of relationships, one from the object to the attribute, and one the other way around. Therefore, we chose to add a single relationship, which in this case can be considered as unordered edges with respect to the graph properties.

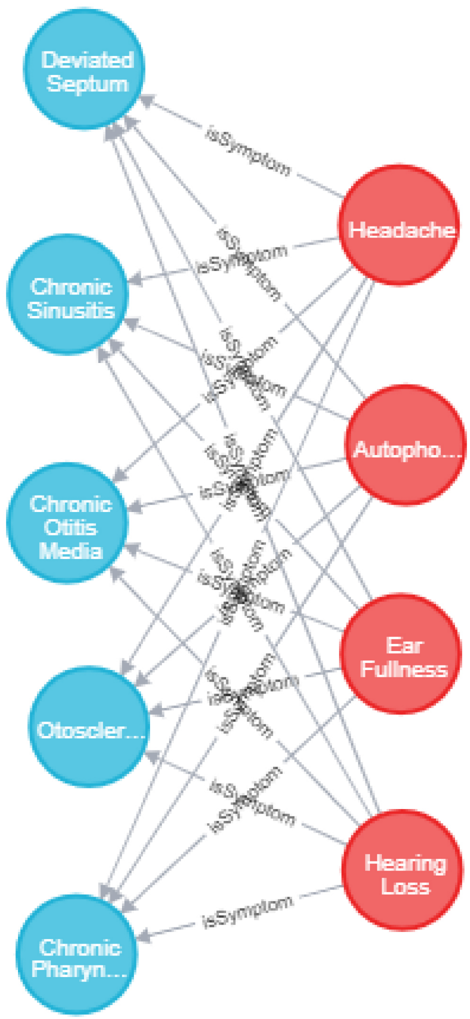

Figure 3. Deviated Septum and Chronic Sinusitis as secondary diagnostics in relation to corresponding symptoms: revealed diagnostics after investigations - NEO4j - zoom only on the concept

In Figure 2 nodes colored in blue are Diagnostics, while nodes colored in red are Symptoms. In this particular case, the relation between Symptoms and Diagnostics is represented with an arrow labeled with `isSymptom`. By following the arrows, i.e. the relationships between different nodes, we can find out different correlations hidden in the medical dataset. Considering the orange highlighted concept presented in Figure 1, we have looked at the graphical representation to identify the same pattern in the generated data graph. Figure 2 presents the same filtered medical dataset where the nodes corresponding to the highlighted concept from Figure 1 are the ones highlighted in the rectangle. This shows how formal contexts and data graphs can be correlated. In this case, we can read the extent and the intent of the corresponding formal concept directly from the graph. We observe that there is no other red node, i.e. symptom, which is in relation to all five diagnostic nodes. Similarly, one can see that no other blue node, i.e. diagnostic, is in relation to all four symptoms identified. Basically we can imagine that a formal concept corresponds to a specific type of strongly connected component in the graph, where there is a relation between all pairs of nodes, with the property that the nodes are of different types, i.e. one is a diagnostic and one is a symptom (obviously it wouldn't make sense to have the relationship "is symptom of" between two diagnostics or between two symptoms).

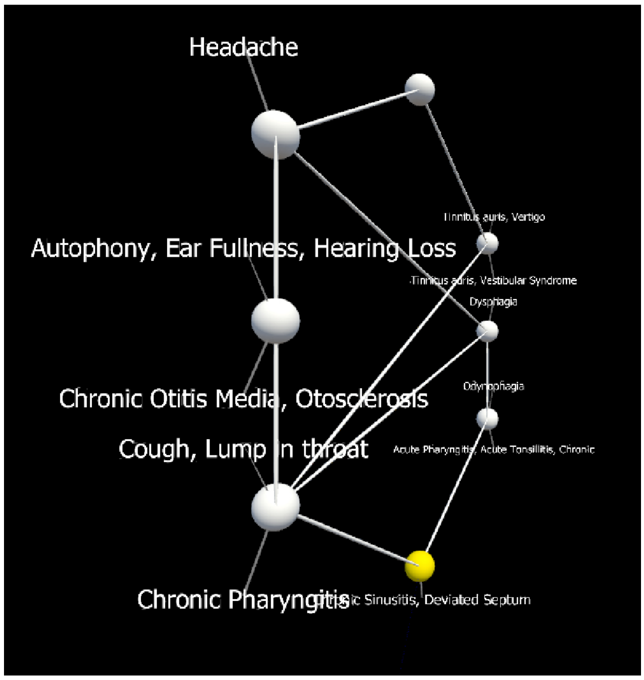

Figure 4. Deviated Septum and Chronic Sinusitis as secondary diagnostics in relation to corresponding symptoms: revealed diagnostics after investigations -3D visualisation of the lattice

Although some data graphs seem a bit hard to read with all the relationships represented, in practice one can choose to exclude or include certain vertices in order to focus on the aspects of interest. Therefore, if we want to analyze the relationship between Symptoms and Diagnostics which are related to `Deviated Septum` and `Chronic Sinusitis` we can choose to exclude nodes which are not connected to all symptoms and diagnostics of interest. In that way, in each of the presented data graphs, we chose to visualize certain relationships between diagnostics and symptoms of the patients. Figure 3 presents the graph containing only the elements corresponding to the formal concept, after choosing to exclude the nodes which are not of interest for this particular example.

By comparing the two representations, the concept lattice obtained with FCA and the data graph obtained with `Neo4j`, we can observe an important advantage of the graph database approach, namely the quantitative information of the clusters which can be easily observed in the data graph, while it is not straightforward in the concept lattice.

When analyzing medical data, interesting facts might stand out, such as rare connections between symptoms and diagnostics. Facts that stand out like this should be analyzed on patients' records over a large period of time and, if they persist, it can lead to the formulation of some hypotheses which can then be researched in more details by medical staff. For instance it would be important to know if there are diagnostics with very similar symptoms, especially if the diagnostics correspond to different medical departments. In that case doctors can be alerted that there is a high chance of a misplaced diagnostic and that, before making the treatment decision, they should consult a doctor with a different specialization in order to exclude diagnostics

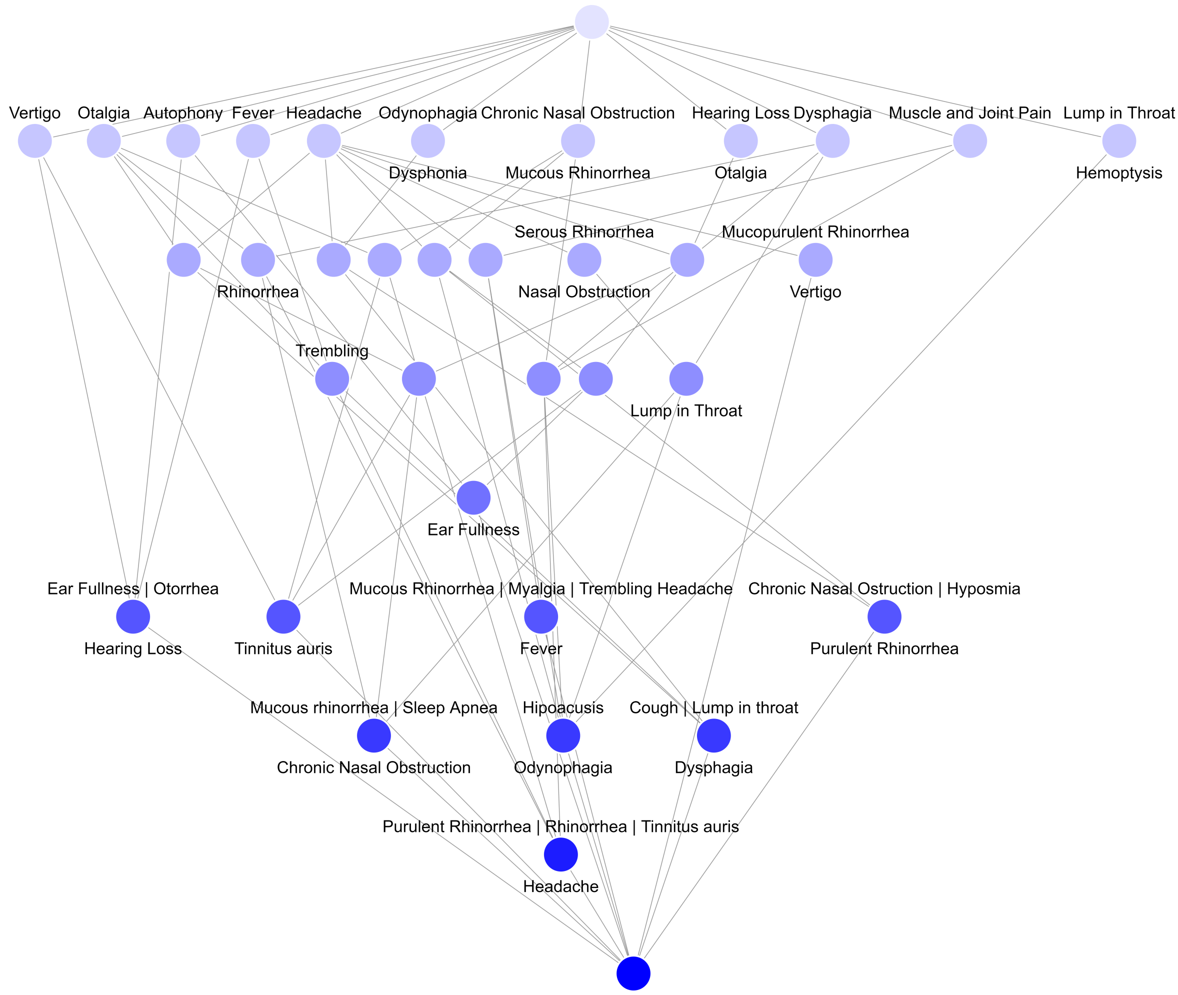

Figure 5. The relation between Principal and Secondary Symptoms - LATTICE

with similar symptoms.

We are especially interested in finding new correlations between symptoms in order to understand which is the cause that led to the diagnostics. For that reason we have represented in Figure 5 the concept lattice, considering principal symptoms as objects and secondary symptoms as attributes. By analyzing the obtained results, doctors can visualize correlations between their patients and coordinate treatment or analyze the differences between them and potential disease progressions.

Due to the fact that usually medical datasets consist of a huge amount of data, visualizing patterns or discovering knowledge is not always an easy task. With the development of new technologies and game engines, the modern graphic capabilities of these technologies increased dramatically. Therefore, we propose a novel approach of navigating through a concept lattice by combining the effectiveness of conceptual scaling with virtual reality. The tool that we have implemented for this purpose is `TOSCANA goes 3D` [20], which allows us to visualize concept lattices by using HTC VIVE HR headsets. As far as we know, this is the first time when knowledge discovery in medical data is enhanced with a virtual reality perspective.

The concept lattices are represented in 3D by using a circular cone like view of the nodes which are at the same depth in the lattice. While exploring the lattice there is the possibility to "move" around the lattice (teleport or fly) in order to view the lattice from different perspectives or be closer to some nodes (i.e. fly to some node). Moreover, one can choose to rotate the lattice or move the nodes around. This movement options implemented in `TOSCANA goes 3D` offer an important advantage for focusing on a desired concept or analyzing a formal concept through its extent or intent. These might give a valuable perspective over the relations between diagnostics and symptoms and how they are correlated. Moreover, the hidden information extracted in the dyadic case or in the graph based visualization, might be further analyzed in connection with other information found in the analyzed dataset.

Figure 4 shows a printscreen from the 3D visualization of the same concept lattice represented in Figure 1, namely `Deviated`

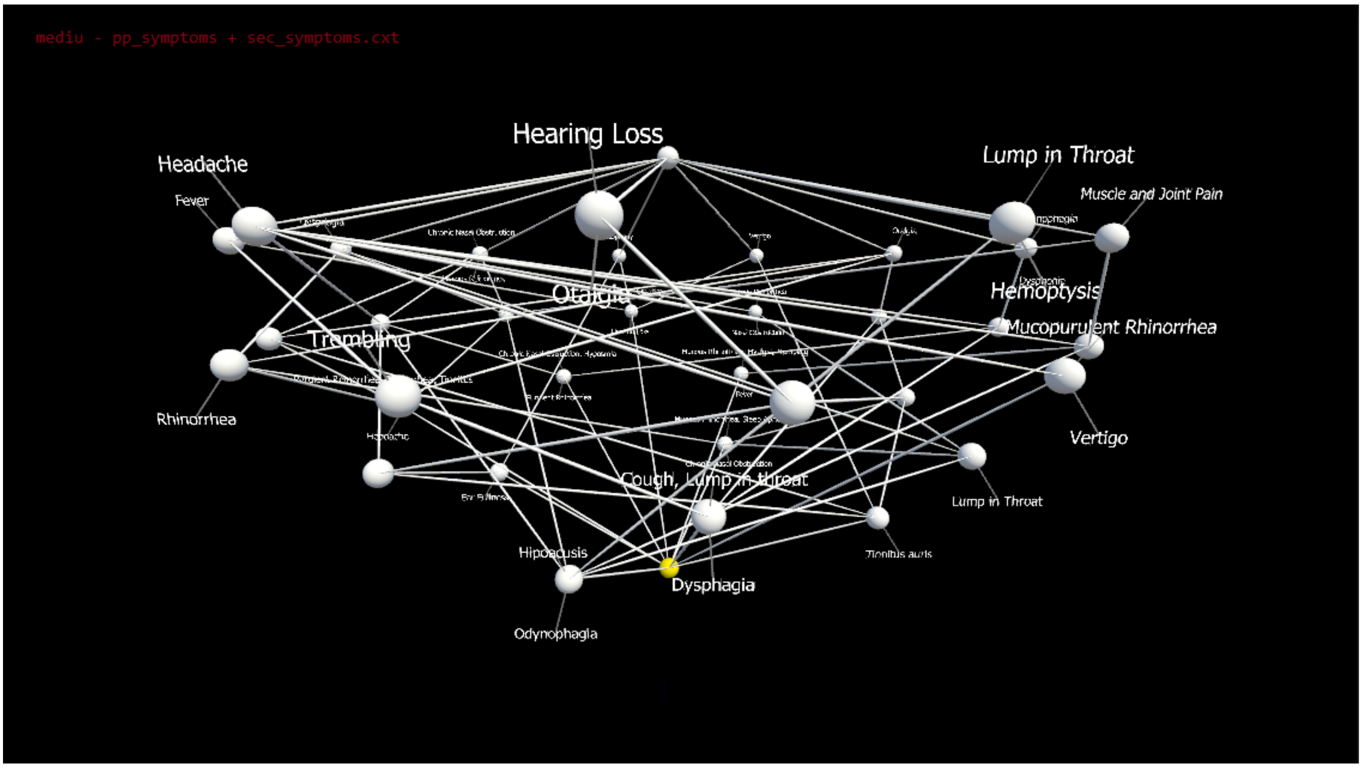

Figure 6. The relation between Principal and Secondary Symptoms -3D visualisation of the lattice

`Septum` and `Chronic Sinusitis` as secondary diagnostics in relation to corresponding symptoms. However, such a flat visualization of the 3D lattice is not conclusive and might seem hard to read, but the whole point of the 3D representation is to be "inside" the lattice, where you can see all the nodes and navigate among them. The difference between the 3D representation and the 2D representation of a lattice is that in 2D we represent the concepts and their links, so that they do not intersect in the two-dimensional space. In a three-dimensional space the representation looks completely different, since it tries to avoid intersections in the three-dimensional space. Therefore, in 3D it is possible that, looking from one perspective one can see a lot of line intersections in the diagram, while shifting the perspective might give you a clear view of the lattice structure. The corresponding 3D lattice of Figure 5 is represented in Figure 6. We can see that the "flat" image shown in Figure 6 is not conclusive and seem hardly readable, but, in order to give the reader a feeling of the 3D navigation, a recording can be found following this link[3].

5 Conclusions

Knowledge Discovery based on FCA and graph databases proves to be a valuable pattern extraction and visualization method which can be used in various ways outside of the scientific community. Switching from 2D to 3D, for instance, makes a more detailed navigation through these conceptual structures possible. This might not be very clear while looking at the 2D variant of a 3D structure, but zooming in, flying around, rotating and teleport are impressive new methods to navigate, explore or evaluate knowledge patterns. On the other hand, using Neo4j to explore graph databases proves that graph based knowledge representation can be used as a complementary exploration method.

Further work will focus on further developing of the 3D capabilities of our approach, including also temporal data and analyzing triadic datasets, i.e., datasets comprising objects, which have properties under some certain conditions. Then, the visualization methods developed will be validated and evaluated by experts of the field. Finally, with the help of experts we can compare our methods with other approaches.

[3] `http://www.cs.ubbcluj.ro/~fca/toscana-goes-3d/`

ACKNOWLEDGEMENTS

The present work has received financial support through the project: Entrepreneurship for innovation through doctoral and postdoctoral research, POCU/360/6/13/123886 co-financed by the European Social Fund, through the Operational Program for Human Capital 2014-2020.

REFERENCES

[1] Alexandru Brad, Luciana Neamţu, Silvia Răusanu, and Christian Săcărea, 'Conceptual knowledge processing grounded logical information system for oncological databases', in *Proceedings of the International Conference on Knowledge Engineering, Principles and Techniques, KEPT 2011, Cluj-Napoca, Romania, July 4-6, 2011*. Studia Informatica - Issue no. 2 / 2011, (2011).

[2] Dursun Delen, Glenn Walker, and Amit Kadam, 'Predicting breast cancer survivability: a comparison of three data mining methods', *Artif. Intell. Medicine*, **34**(2), 113–127, (2005).

[3] Kenichiro Fujita, Onishi Katsumi, Tadamasa Takemura, and Tomohiro Kuroda, 'The improvement of the electronic health record user experience by screen design principles', *J. Medical Systems*, **44**(1), 21, (2020).

[4] Bernhard Ganter and Rudolf Wille, *Formal Concept Analysis - Mathematical Foundations*, Springer, 1999.

[5] Anamika Gupta, Naveen Kumar, and Vasudha Bhatnagar, 'Analysis of medical data using data mining and formal concept analysis', *International Journal of Medical and Health Sciences*, **1**(11), 591 – 594, (2007).

[6] Diana Halita and Christian Sacarea, 'Is FCA suitable to improve electronic health record systems?', in *24th International Conference on Software, Telecommunications and Computer Networks, SoftCOM 2016, Split, Croatia, September 22-24, 2016*, pp. 1–5. IEEE, (2016).

[7] Nicolas Jay, François Kohler, and Amedeo Napoli, 'Using formal concept analysis for mining and interpreting patient flows within a healthcare network', in *Concept Lattices and Their Applications, Fourth International Conference, CLA 2006, Tunis, Tunisia, October 30 - November 1, 2006, Selected Papers*, eds., Sadok Ben Yahia, Engelbert Mephu Nguifo, and Radim Belohlávek, volume 4923 of *Lecture Notes in Computer Science*, pp. 263–268. Springer, (2006).

[8] Chris Kemper, *Beginning Neo4j*, Apress,, 2015.

[9] Levente Lorand Kis, Christian Sacarea, and Diana-Florina Sotropa, 'Visualizing conceptual structures using FCA tools bundle', in *Graph-Based Representation and Reasoning - 23rd International Conference on Conceptual Structures, ICCS 2018, Edinburgh, UK, June 20-22, 2018, Proceedings*, eds., Peter Chapman, Dominik Endres, and Nathalie Pernelle, volume 10872 of *Lecture Notes in Computer Science*, pp. 193–196. Springer, (2018).

[10] Levente Lorand Kis, Christian Sacarea, and Diana Troanca, 'FCA tools bundle - A tool that enables dyadic and triadic conceptual navigation', in *Proceedings of the 5th International Workshop "What can FCA do for Artificial Intelligence"? co-located with the European Conference on Artificial Intelligence, FCA4AI collocated with ECAI 2016, The Hague, the Netherlands, August 30, 2016*, eds., Sergei O. Kuznetsov, Amedeo Napoli, and Sebastian Rudolph, volume 1703 of *CEUR Workshop Proceedings*, pp. 42–50, (2016).

[11] Mark Needham and Amy Hodler E., *Graph Algorithms: Practical Examples in Apache Spark and Neo4j*, O'Reilly Media, Inc., 2019.

[12] Telung Pan, Yunchun Tsai, and Kowting Fang, 'Using formal concept analysis to design and improve multidisciplinary clinical processes', *WSEAS Transactions on Information Science and Applications*, **5**(6), 880–890, (2008).

[13] Jonas Poelmans, Dmitry I. Ignatov, Sergei O. Kuznetsov, and Guido Dedene, 'Formal concept analysis in knowledge processing: A survey on applications', *Expert Syst. Appl.*, **40**(16), 6538–6560, (2013).

[14] Ian Robinson, Jim Webber, and Emil Eifrem, *Graph Databases: New Opportunities for Connected Data*, O'Reilly Media, Inc., 2nd edn., 2015.

[15] Christian Sacarea, 'Investigating oncological databases using conceptual landscapes', in *Graph-Based Representation and Reasoning - 21st International Conference on Conceptual Structures, ICCS 2014, Iaşi, Romania, July 27-30, 2014, Proceedings*, eds., Nathalie Hernandez, Robert Jäschke, and Madalina Croitoru, volume 8577 of *Lecture Notes in Computer Science*, pp. 299–304. Springer, (2014).

[16] Christian Sacarea, Silviu Boldeanu, Luciana Neamtiu, and Cezara Pastrav, 'Investigating adverse drug reactions using conceptual landscapes', in *IEEE International Conference on Intelligent Computer Communication and Processing, ICCP 2011, Cluj-Napoca, Romania, August 25-27, 2011*, pp. 41–48. IEEE, (2011).

[17] Christian Sacarea, Diana-Florina Sotropa, and Diana Troanca, 'Symptoms investigation by means of formal concept analysis for enhancing medical diagnoses', in *25th International Conference on Software, Telecommunications and Computer Networks, SoftCOM 2017, Split, Croatia, September 21-23, 2017*, eds., Dinko Begusic, Nikola Rozic, Josko Radic, and Matko Saric, pp. 1–5. IEEE, (2017).

[18] Christian Sacarea, Diana-Florina Sotropa, and Diana Troanca, 'Formal concept analysis grounded knowledge discovery in electronic health record systems', in *20th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, SYNASC 2018, Timisoara, Romania, September 20-23, 2018*, pp. 266–271. IEEE, (2018).

[19] Christian Sacarea, Diana-Florina Sotropa, and Diana Troanca, 'Using analogical complexes to improve human reasoning and decision making in electronic health record systems', in *Graph-Based Representation and Reasoning - 23rd International Conference on Conceptual Structures, ICCS 2018, Edinburgh, UK, June 20-22, 2018, Proceedings*, eds., Peter Chapman, Dominik Endres, and Nathalie Pernelle, volume 10872 of *Lecture Notes in Computer Science*, pp. 9–23. Springer, (2018).

[20] Christian Sacarea, Diana-Florina Sotropa, and Raul-Robert Zavaczki, 'Toscana goes 3d: Using VR to explore life tracks', in *Supplementary Proceedings of ICFCA 2019 Conference and Workshops, Frankfurt, Germany, June 25-28, 2019*, eds., Diana Cristea, Florence Le Ber, Rokia Missaoui, Léonard Kwuida, and Baris Sertkaya, volume 2378 of *CEUR Workshop Proceedings*, pp. 82–87. CEUR-WS.org, (2019).

[21] Douglas West, *Introduction to Graph Theory*, Pearson Education, Inc., 2nd edn., 2001.

[22] Rudolf Wille, *Classification in the Information Age: Proceedings of the 22nd Annual GfKl Conference, Dresden, March 4–6, 1998*, chapter Conceptual Landscapes of Knowledge: A Pragmatic Paradigm for Knowledge Processing, 344–356, Springer Berlin Heidelberg, Berlin, Heidelberg, 1999.

[23] Hamed Majidi Zolbanin, Dursun Delen, and Amir Hassan Zadeh, 'Predicting overall survivability in comorbidity of cancers: A data mining approach', *Decis. Support Syst.*, **74**, 150–161, (2015).