(Missing) Concept Discovery in Heterogeneous Information Networks

Tobias Kötter and Michael R. Berthold

Nycomed-Chair for Bioinformatics and Information Mining, University of Konstanz,
78484 Konstanz, Germany
`Tobias.Koetter@uni-Konstanz.de`

Abstract. This article proposes a new approach to extract existing (or detect missing) concepts from a loosely integrated collection of information units by means of concept graph detection. Thereby a concept graph defines a concept by a quasi bipartite sub-graph of a bigger network with the members of the concept as the first vertex partition and their shared aspects as the second vertex partition. Once the concepts have been extracted they can be used to create higher level representations of the data. Concept graphs further allow the discovery of missing concepts, which could lead to new insights by connecting seemingly unrelated information units.

1 Introduction

The amount of data to which researchers have access is increasing at a breathtaking pace. The available data stems from heterogeneous sources from diverse domains with varying semantics and of various quality. It is a big challenge to integrate and reason from such an amount of data. However by integrating data from diverse domains, relations can be discovered spanning multiple domains, leading to new insights and thus a better understanding of complex systems. In this article we use a network-based approach to integrate data from diverse domains of varying quality. The network consists of vertices that represent information units such as objects, ideas or emotions, whereas edges represent the relations between these information units.

Once the data has been merged into a unifying model it needs to be analyzed. In this article we describe an approach based on concept graphs to extract semantical information from loosely integrated information fragments. This approach was presented at the International Conference on Computational Creativity [6]. Concept graphs allow for the detection of existing concepts, which can be used to create an abstraction of the underlying data. They define a concept by a quasi bipartite sub-graph consisting of two vertex partitions. The first partition contains the members of the concept and the second partition the aspects they have in common. By providing a higher level view on the data the user might obtain a better insight into the integrated data and discover new relations across diverse domains that have been hidden in the noise of the integrated data.

Concept graphs also allow for the detection of domain bridging concepts [8] that connect information units from various domains. These domain bridging concepts support creative thinking by connecting seemingly unrelated information units from diverse domains.

Another advantage of concept graphs is that they enable information units to be detected that share common properties but to which no concept has yet been assigned. This might lead to the discovery of concepts that are missing in the data or to the detection of new concepts.

The rest of the chapter is organized as follows: in the next section we will briefly review Bisociative Information Networks [7], which we use for the integration of heterogeneous data sources from diverse domains. We move on to introduce concept graphs and describe their detection, and subsequently discuss the discovery of concept graphs in a real world data set and show some example graphs. Finally we draw conclusions from our discussion and provide an outlook on future work.

2 Bisociative Information Networks

Bisociative Information Networks (BisoNets) [3,7] provide a framework for the integration of semantically meaningful information but also loosely coupled information fragments from heterogeneous data sources. The term *bisociation* [5] was coined by Arthur Koestler in 1964 to indicate the "...joining of unrelated, often conflicting information in a new way...".

BisoNets are based on a k-partite graph structure, whereby the most trivial partitioning consists of two partitions ($k = 2$), with the first vertex set representing units of information and the second set representing the relations among information units. By representing relations as vertices BisoNets support the modeling of relationships among any number of members.

However the role of a vertex is not fixed in the data. Depending on the point of view a vertex can represent an information unit or a relation describing the connection between units of information. Members of a relation are connected by an edge with the vertex describing the relation they share. One example is the representation of documents and authors where documents as well as authors are represented as vertices. Depending on the point of view, a document might play the role of the relation describing authorship or might be a member in the relation of documents written by the same author.

The unified modeling of information units and relations as vertices has many advantages e.g. they both support the assigning of attributes such as different labels. However these attributes do not carry any semantic information. Edges can be further marked as directed to explicit model relationships that are only valid in one direction. Vertices can also be assigned to partitions to distinguish between different domains such as biology, chemistry, etc.

In contrast to ontologies, semantic networks or topic maps, relations are assigned a weight that describes the reliability of the connection. This means that BisoNets support the integration not only of facts but also of pieces of evidence.

Thus units of information and their relations can be extracted from various information sources such as existing databases, ontologies or semantical networks. But also semistructured and noisy data such as literature or biological experiments can be integrated in order to provide a much richer and broader description of the information units. By applying different mining algorithms to the same information source, diverse relations and units of information can be extracted with each mining algorithm representing an alternative view that might highlight a different aspect of the same data.

BisoNets focus solely on the information units and their relations and do not store all the more detailed data underneath the pieces of information. However vertices do reference the detailed data they stem from. This allows BisoNets to integrate huge amounts of data and still be able to show the data from which a vertex originates.

3   Concept Graphs

Once all the data has been integrated, it has to be analyzed in order to find valuable information. We propose a new method to automatically extract semantic information from the loosely integrated collection of information units by means of concept graph detection.

A *concept graph* represents a *concept* that stands for a mental symbol. A concept consists of its *members*, which do not only refer to materialized objects but also to ideas, activities or events, and their common *aspects*, which represent the properties the concept members share. In philosophy and psychology, the concept members are also known as the extension of a concept, which consists of the things to which the concept applies - whereby the aspects are known as the intension of a concept, consisting of the idea or the properties of the concept. An example could be a concept representing birds with specific birds such as eagles or sparrows as members, which in turn are related to their common aspects such as feather, wing, and beak.

Concept graphs base on the assumption that similar information units share more properties than dissimilar information units. Therefore the more similar two information units are, the more properties they share. This assumption bases on the family resemblance proposed by Wittgenstein [12], which states that objects that already share some properties are likely to share further common properties. The theory of basic objects in natural categories from Rosch et al. [9] is also bases on the family resemblance. Rosch et al. define a basic category as the category that carries the most information; the basic categories consist of properties that are mostly connected to the members of the category. Thus family resemblance as well as the basic categories speak in favor of the assumption that a concept does not only possess one property, but many. These properties describe the members of a concept and distinguish the members of a concept from non-members.
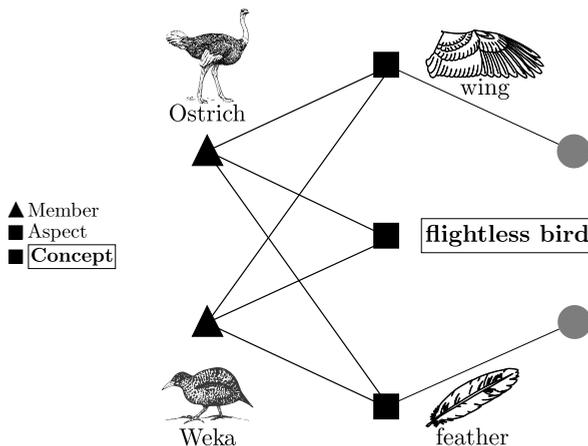
Fig. 1. Example of aconcept graph describing the concept *flightless bird* with its members *Ostrich* and *Weka* and their common aspects *wing* and *feather*

In addition to the concept members and their shared aspects, a concept graph might also contain the symbolic representation of the concept itself. This symbolic representation can be used to generate an abstract view on the data since it represents all members of the corresponding concept graph.

An example of a concept graph that represents the concept of *flightless birds* is depicted in Figure 1. It consists of the two concept members *Ostrich* and *Weka* and their shared aspects *wing* and *feather*. The graph also contains the symbolic representation of the *flightless bird* concept, which can be used as an abstract representation of this particular concept graph.

3.1   Preliminaries

As mentioned above the members of a concept graph are similar in that they share some aspects. In BisoNets the aspects of an information unit are represented by its direct neighbors. The more neighbors two information units share the more similar they are. This leads to the representation of a concept graph as a dense sub-graph in a BisoNet, consisting of two disjoint and fully connected vertex sets. Here the first vertex set represents the concept members and the second vertex set the aspects that are shared by all members of the concept graph. Thus a perfect concept graph would form a complete bipartite graph as depicted in Figure 1 with the concept members as the first partition and the aspects with the concept as the second partition. A concept graph might also contain relations among the vertices within a partition and thus does not necessarily form a perfect bipartite (sub) graph.

Once a dense sub-graph has been detected it needs to be analyzed in order to distinguish between the member set and the aspect set. We have developed heuristics to detect the different set types for directed and undirected networks. Both heuristics are based on the assumption that information units are described by their neighbors in the network. In addition, the heuristics for the directed network are based on the assumption that information units point to their aspects. Hence a relation in a directed network consists of an information unit as the source and an aspect as the target vertex.

The heuristics to identify the different vertex types are based on the following definitions:

Let $B(V, E)$ be the un/directed BisoNet that contains all information with V representing the vertices and $E \subseteq V \times V$ representing the edges. The edge $(u, v) \in E$ represents a directed edge with $u \in V$ as source and $v \in V$ as target vertex whereas $\{u, v\} \in E$ represents an undirected edge connecting the two vertices $u, v \in V$.

$C(V_A, V_M, E_C) \subseteq B$ defines the concept graph C in the BisoNet B. $V_A \subseteq V$ represents the aspect set and $V_M \subseteq V$ the member set of the concept graph C in which $V_A \cap V_M = \emptyset$. $V_C = V_A \cup V_M$ is the set of all vertices within the concept graph. $E_C \subseteq E$ is the subset of all edges that connect vertices within the concept graph $E_C = \{\{u, v\} \in E : u, v \in V_C\}$.

Let

$$N(v) = \{u \in V : \{v, u\} \in E\}$$

be the neighbors of the vertex $v \in V$ in the BisoNet B. Whereby

$$N^+(v) = \{u \in V : (v, u) \in E\}$$

denotes its target neighbors and

$$N^-(v) = \{u \in V : (u, v) \in E\}$$

its source neighbors.

The neighbors of a vertex $v \in V$ in a given vertex set $U \subseteq V$ are denoted by

$$N(v, U) = N(v) \cap U = \{u \in U : \{u, v\} \in E\}.$$

In the directed case

$$N^+(v, U) = N^+(v) \cap U = \{u \in U : (v, u) \in E\}$$

denotes the target neighbors and

$$N^-(v, U) = N^-(v) \cap U = \{u \in U : (u, v) \in E\}$$

the source neighbors of a given vertex $v \in V$ within a given vertex set $U \subseteq V$.

Member Set. The concept members form the first of the two disjoint vertex sets of the concept graph. The heuristic that denotes the probability of a vertex $v \in V_C$ to be part of the member set V_M is denoted by the function $m : V_C \to [0, 1]$.

Given the set $V_C \subseteq V$ of all vertices within a concept graph C in an undirected network, $m(v)$ is defined as the ratio of neighbors inside and outside the concept graph for a given vertex $v \in V_C$

$$m(v) = \frac{|N(v, V_C)|}{|N(v)|}.$$

In a directed network the heuristic bases on the assumption that concept members point to their aspects. This assumption leads to the computation of the ratio of target neighbors inside and outside the concept graph for a given vertex $v \in V_C$

$$m(v) = \frac{|N^+(v, V_C)|}{|N^+(v)|}.$$

The set of information units V_M for a given BisoNet B is defined as

$$V_M = \max_{V' \in V_C} \frac{1}{|V'|} \sum_{v \in V'} m(v).$$

Aspect Set. The aspect set is the second vertex set of the concept graph that describes the members of the concept graph. Each aspect on its own might be related to other vertices as well but the set of aspects is only shared by the members of the concept graph. The vertices of the aspect set might differ considerably in the number of relations to vertices outside of the concept graph depending on their level of detail. More abstract aspects such as animals are likely to share more neighbors outside of the concept graph than more detailed aspects such as bird.

The heuristic that denotes the probability of a vertex $v \in V_C$ to belong to the aspect set V_A is denoted by the function $a : V_C \to [0, 1]$.

Given the set $V_C \subseteq V$ of all vertices within a concept graph C in an undirected network, $a(v)$ is defined as the inverse ratio of neighbors inside and outside the concept graph for a given vertex $v \in V_C$

$$a(v) = 1 - \frac{|N(v, V_C)|}{|N(v)|} = 1 - m(v).$$

In a directed network the heuristic is defined as the ratio of the source neighbors inside and outside the concept graph for a given vertex $v \in V_C$

$$a(v) = \frac{|N^-(v, V_C)|}{|N^-(v)|}.$$

The set of aspects V_A for a given BisoNet B is defined as

$$V_A = \max_{V' \in V_C} \frac{1}{|V'|} \sum_{v \in V'} a(v).$$

Concepts. The concept $c \in V_A$ is a vertex of the aspect set. A concept differs from the other vertices of the aspect set in that it should only be related to the vertices of the member set within the concept graph. Hence a perfect concept has no relations to vertices outside of the concept graph and can thus be used to represent the concept graph.

The heuristic c : $V_A \to [0,1]$ denotes the probability of a vertex being the concept that can represent concept graph C. The heuristic is based on the cue validity [2] which describes the relevance of an aspect for a given concept. More specific aspects have a higher cue validity than more general ones.

Given the set $V_C \subseteq V$ of all vertices within a concept graph C in an undirected network, the heuristic is defined as the ratio of the neighbors inside and outside the concept graph

$$c(v) = \frac{|N(v, V_C)|}{|N(v)|} = m(v), v \in V_A.$$

In a directed network the heuristic considers the ratio of the source neighbors inside and outside the concept graph

$$c(v) = \frac{|N^-(v, V_C)|}{|N^-(v)|} = a(v), v \in V_A.$$

The concept c that can represent the concept graph is the vertex $v \in V_A$ with the highest value for c(v)
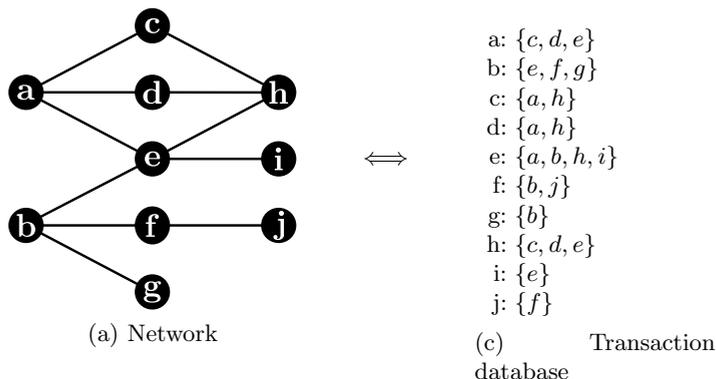
$$c = \max_{v \in V_A} c(v).$$

Depending on a user-given threshold we are able to detect a concept graph without a concept. The concept graph lacks a concept if the concept value c(v) of all vertices of its aspect set is below the given threshold. This might be an indication of an unknown relation among information units that has not been discovered yet and to which no concept has been assigned.

3.2 Detection

In this chapter we use a frequent item set mining algorithm [1] to detect concept graphs in BisoNets. By using frequent item set algorithms we are able to detect concept graphs of different sizes and specificity.

Frequent item set mining has been developed for the analysis of market baskets in order to find sets of products that are frequently bought together. It operates on a transaction database that consists of a transaction identifier and the products that have been bought together in the transaction. Represented as a graph, the overlapping transactions form a complete bipartite graph, which is the basis of our concept graphs.

In order to apply frequent item set mining algorithms to find concept graphs in BisoNets we use the adjacency list of the network as transaction database. Therefore, for each vertex in the BisoNet, we create an entry in the transaction database with the vertex as the identifier and its direct neighbors as the products

a: $\{c, d, e\}$
b: $\{e, f, g\}$
c: $\{a, h\}$
d: $\{a, h\}$
e: $\{a, b, h, i\}$
f: $\{b, j\}$
g: $\{b\}$
h: $\{c, d, e\}$
i: $\{e\}$
j: $\{f\}$

(a) Network

(c)   Transaction
database

Fig. 2. Network and the corresponding adjacency list which serves as the transaction database for the frequent item set mining algorithm

(Figure 2). Once the database has been created we can apply frequent item set mining algorithms to detect vertices that share some neighbors.

Frequent item set mining algorithms allow the selection of a minimum support that defines the minimum number of transactions containing a given item set in order to make it frequent. They also allow a minimum size to be set for the item set itself in order to discard all item sets that contain fewer items than the given threshold. By setting these two thresholds we are able to define the minimum size of the concept graph.

Since we want to find concept graphs of different specificity we need an additional threshold that takes the general overlap of the transactions into account. To achieve this we used an adaption of the Eclat [13] algorithm called Jaccard Item Set Mining (JIM) [10]. JIM uses the Jaccard index [4] as an additional threshold for pruning the frequent item sets. For two arbitrary sets A and B the Jaccard index is defined as

$$\mathrm{j}(A, B) = \frac{|A \cap B|}{|A \cup B|}.$$

Obviously, $\mathrm{j}(A, B)$ is 1 if the sets coincide (i.e. $A = B$) and 0 if they are disjoint (i.e. $A \cap B = \emptyset$).

By setting the threshold for the JIM algorithm between 0 and 1 we are able to detect concept graphs of different specificity. By setting the threshold to 1 only those vertices that share all of their neighbors are retained by the algorithm. This results in the detection of more specific concept graphs, which contain information units or aspects that exclusively belong to the detected concept graph. Relaxing the threshold by setting a smaller value results in the detection of more general concept graphs where the information units share some but not all of their aspects. Varying thresholds might lead to the detection of overlapping concept graphs. This can be used to create a hierarchy among the concepts.

4   Application

The 2008/09 Wikipedia Selection for schools[1] (Schools Wikipedia) is a free, hand-checked, non-commercial selection of the English Wikipedia[2] funded by SOS Children's Villages. It has been created with the intention of building a child safe encyclopedia. It has about 5500 articles and is about the size of a twenty volume encyclopedia (34,000 images and 20 million words). The encyclopedia contains 154 subjects, which are grouped into 16 main subjects such as countries, religion and science. The network has been created from the Schools Wikipedia version created in October 2008. Each article is represented by a vertex and the subjects are represented by domains. Every article is assigned to one or more domains depending on the assigned subjects. Hyperlinks are represented by links connecting the article that contains the hyperlink and the referenced article.

This example data set and the representation as a hyperlink graph has been chosen since it can be validated manually by reading the Schools Wikipedia articles and inspecting their hyperlinks.

4.1   Results

This section illustrates concept graphs discovered in the Schools Wikipedia data set using the JIM algorithm. The concept graphs consist of the discovered item sets that form the first vertex set and the corresponding root vertices of the transaction that build the second vertex set. Once we have discovered both vertex sets and determined their types we can display them as a graph.

The following graphs display the information units with triangular vertices. Both aspects and the concept are represented by a squared vertex whereas the concept has a box around its label.

Figure 3 depicts such a discovered concept graph that represents the *dinosaur* concept from the biological section of the Schools Wikipedia. The members of the concept graph consist of the orders (e.g. *Ornithischia*) and genera (e.g. *Triceratops*) of the dinosaurs. The members are described by their aspects *Animal*, phylum *Chordate* and the reference to the *Biological classification* as well as the concept *Dinosaur* itself.

Detection and Expansion of Existing Hierarchies. This section demonstrates the ability of concept graphs to detect and expand existing hierarchies in the integrated data. Figure 4 depicts the *Saurischia* order (see Fig. 4a) and one of its suborders *Sauropodomorpha* (see Fig. 4b), which where discovered in the integrated data with the help of the detected concept graphs. These concept graphs benefit from the structure of the Schools Wikipedia pages of the animal section, as they include an information box with the Kingdom, Phylum etc. of the animal.

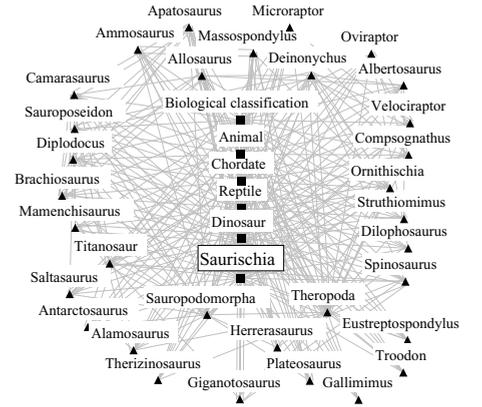[1] http://schools-wikipedia.org/
[2] http://en.wikipedia.org
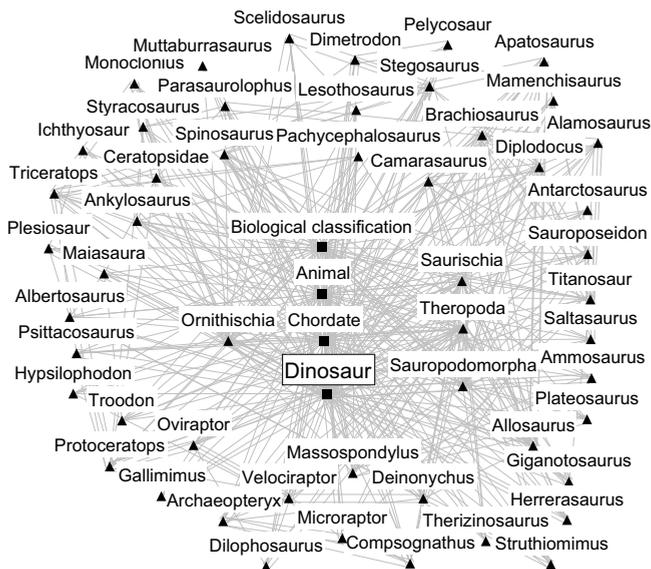
Fig. 3. *Dinosaur* concept graph with the orders and genera of the dinosaurs as members and their biological classification as their common aspects

Figure 5 depicts two different bird categories which were also extracted from the animal section of the Schools Wikipedia data set but which are not part of the standard information box of the corresponding Schools Wikipedia pages.

The concept graph in Figure 5a represents the group of *Waders*. *Waders* are long-legged wading birds such as *Herons*, *Flamingos* and *Plovers*. The concept graph also contains *Terns* even though they are only distantly related to *Waders*. However Schools Wikipedia states that studies in 2004 showed that some of the gene sequences of *Terns* showed a close relationship between *Terns* and the Thinocori, a species of aberrant *Waders*.

The concept graph in Figure 5b represents the *Bird of prey* group. *Birds of prey* or raptors hunt for food on the wing. The graph includes different sub families such as *Hawk*, *Kite* and *Falcon* as well as members of these sub families such as the *Harrier Hawk*. The *Common Cuckoo* is not a *Bird of prey* but is included in the concept graph since it looks like a small *Bird of prey* in flight as stated in its article in Schools Wikipedia. The concept graph contains the *Great Black-backed Gull* which is a gull that behaves more like a *Bird of prey* than a typical gull by frequently hunting any prey that is smaller than itself.

These examples partially benefit from the structure of the Schools Wikipedia pages of the animal section. They all contain an information box specifying the Kingdom, Phylum etc. of the animal. However this demonstrates that our method is able to discover ontologies, such as the biological classification of the dinosaurs (see Fig. 4), if they are available in the integrated data. Furthermore the examples demonstrate the capability of the method to detect further categories such as *Waders* or *Birds of prey* (see Fig. 5) even though they are not part

(a) Concept graph of the dinosaur order *Saurischia*.
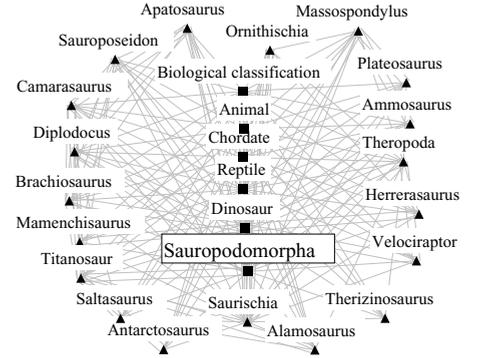
(b) Concept graph of the dinosaur suborder *Sauropodomorpha*.

Fig. 4. Concept graphs of the dinosaur order and one of its suborders

of the ontology structure in Schools Wikipedia. By including the *Great Black-backed Gull* and the *Common Cuckoo* the concept graph for the *Bird of prey* also demonstrates the ability to detect information units that are not typically related to the concept.

Missing Concept Detection. This section demonstrates the ability of concept graphs to detect groups of information units that share common aspects but to which no concept has been assigned. These concept graphs might be the result of incomplete or erroneous data. They might also be a hint of groups of information units that share certain aspects that have not been discovered yet.

The concept graph in Figure 6 is an example of such a concept graph, which lacks an appropriate concept. The graph describes battles between the *United States* and the *Imperial Japanese Navy* during *World War II*. The various battles represent the information units of the concept graph whereas the combatants and
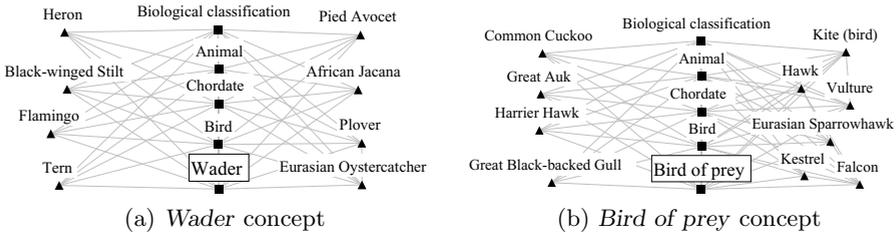
(a) *Wader* concept

(b) *Bird of prey* concept

Fig. 5. Concept graphs that expand the existing birds hierarchy
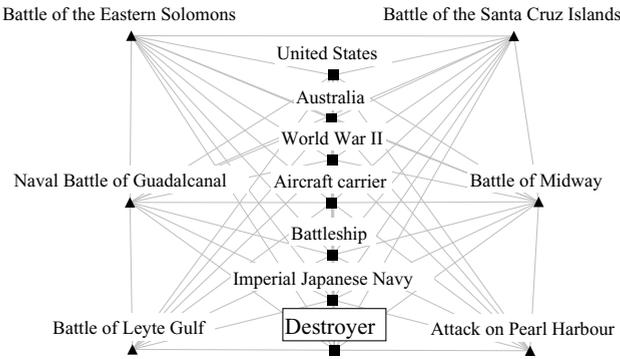

Fig. 6. Example concept graph with a low concept confidence

some of their warships form its aspects. The concept is missing since the Schools Wikipedia data set does not contain an article that groups the described battles.

The following concept graphs contain more aspects than information units. The information units represented by triangular vertices are therefore depicted in the center whereas the aspects and the concepts represented by squared vertices form the outer circle.

Overlapping Concept Graphs. This section describes concept graphs that overlap by sharing common vertices. These shared vertices might have different types since their type depends on the role they possess in a given concept graph. These vertices, which belong to several concept graphs and possess different types, are represented by circular vertices in the following figures.

Figure 7 depicts the connection among the concepts of *Meteorology*, *Greenhouse effect*, *Attribution of recent climate change* and the *Ice sheet* of *Glaciers*. The concept graphs demonstrate the ability of the discussed method to assign different types to the same vertex depending on a given concept graph. The vertex *Earth's atmosphere* as an example is an information unit of the *Greenhouse effect* but an aspect of the concept *Meteorology*. The vertices *Global warming* and *Climate change* are information units of the concept of *Attribution of recent climate change* but aspects of the *Ice sheet* concept.
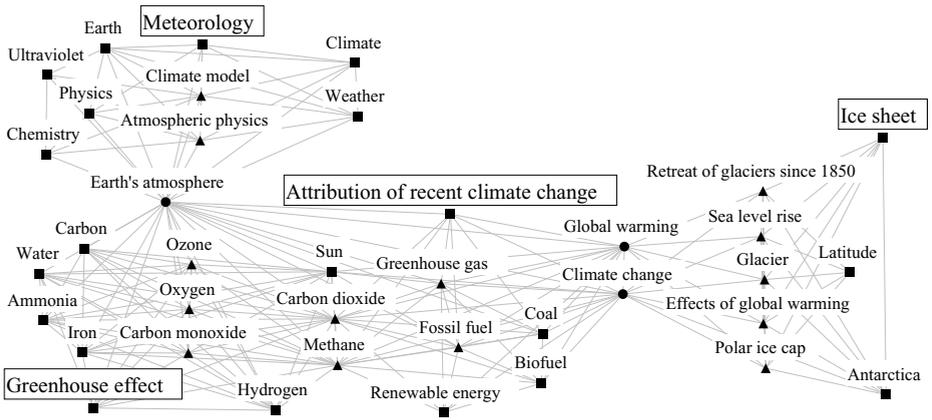
Fig. 7. Overlapping concept graphs reveal complex relations between different concepts

Domain Bridging Concept Graphs. The following examples demonstrate the ability of concept graphs to detect domain bridging concepts [8]. They depict concept graphs that contain either information units or aspects from diverse domains.

The first Figure 8a depicts the concept graph for *Blood pressure*. This concept contains two information units from the *Health and medicine* domain, which are described by aspects from various domains. The domains of the aspects range from *Health and medicine* (e.g. *Blood*), to *Chemical compounds* (e.g. *Glucose*), to *Sports* (e.g. *Sumo*) and *Recreation* (e.g. *Walking*) as well as *Plants* (e.g. *Garlic*).

The second Figure 8b groups information units from diverse domains such as *Cartoons* and *Military History and War*. The vertex *Donald Duck* represents a famous character from the *Cartoons* domain that was used as propaganda
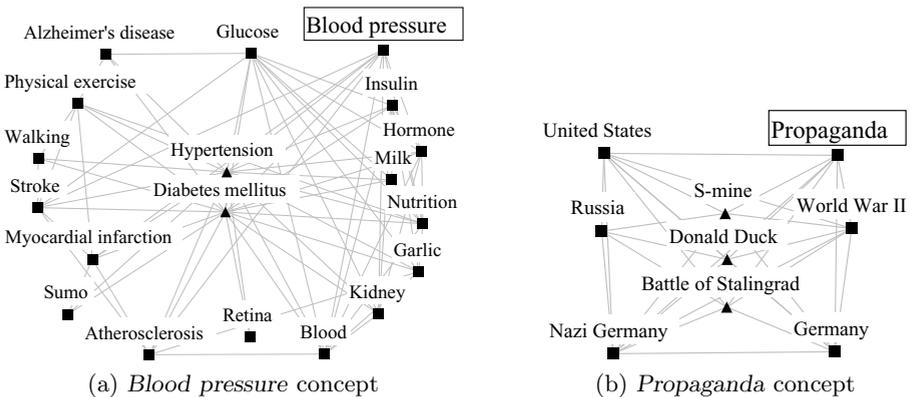

(a) *Blood pressure* concept

(b) *Propaganda* concept

Fig. 8. Two domain bridging concept graphs that connect aspects and information units from different domains
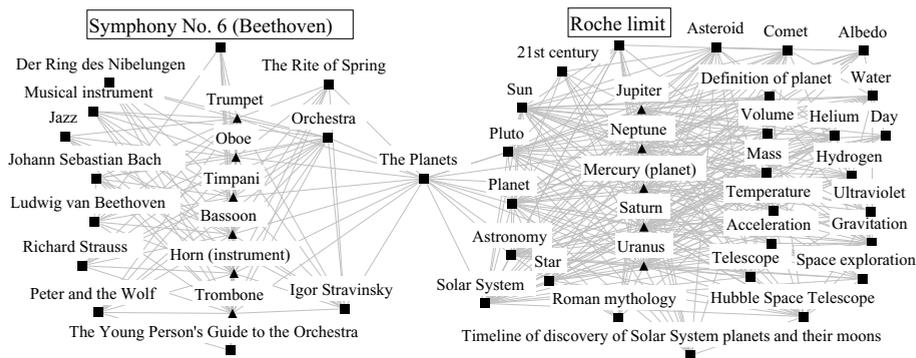
Fig. 9. Example of an overlapping concept graph that connects concepts from heterogeneous domains

whereas *S-mine* describes a weapon from the *Military History and War* domain that was used during *World War II*.

The last Figure 9 depicts not a single concept graph connecting vertices from diverse domains but two overlapping concept graphs that describe concepts from diverse domains. The first concept graph from the *Musical recordings and compositions* domain describes the concept of the *Symphony No. 6*, a famous symphony written by Ludwig van Beethoven. The second concept graph stems from the *Space(Astronomy)* domain describing the concept of the *Roche limit*, which defines the distance within a celestial body such as a planet held together by its own gravity alone. Both concepts are connected by their shared aspect, *The Planets*, which is an orchestral suite that consists of seven movements named after planets (e.g. *Jupiter*, *Neptune*, etc.).

5   Conclusion and Future work

In this chapter we have discussed a new approach to detect existing or missing concepts from a loosely integrated collection of information fragments that can lead to deeper insight into the underlying data. We have discussed concept graphs as a way to discover conceptual information in BisoNets. Concept graphs allow for the abstraction of the data by detecting existing concepts and producing a better overview of the integrated data. They further support the detection of missing concepts by discovering information units that share certain aspects but which have no concept and could be a hint of a previously unknown and potentially novel concept.

By using information networks as input data and retrieving existing as well as unknown concepts from the integrated data, the discussed approach supports creative thinking by improving the understanding of complex systems and the discovery of interesting and unexpected relationships.

This approach can also be expanded to detect domain bridging concepts [8], which might support creative thinking by connecting information units from diverse domains. Since BisoNets store the domain from which a vertex stems, we can use this information to find concept graphs that contain information units from diverse domains.

In addition to the discovery of concept graphs we plan to identify overlapping concept graphs, which can be used to create a hierarchy among the detected concepts using methods from formal concept analysis [11]. The hierarchy ranging from most specific to most general concepts can be created by detecting more specific concept graphs, which are included in more general concept graphs. The different levels of concept graphs can be detected by varying the threshold of the discussed Jaccard Item Set Mining algorithm.

References

1. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules in large databases. In: Proceedings of the 20th International Conference on Very Large Data Bases (1994)
2. Beach, L.R.: Cue probabilism and inference behavior. Psychological Monographs: General and Applied 78, 1–20 (1964)
3. Berthold, M.R., Dill, F., Kötter, T., Thiel, K.: Supporting Creativity: Towards Associative Discovery of New Insights. In: Washio, T., Suzuki, E., Ting, K.M., Inokuchi, A. (eds.) PAKDD 2008. LNCS (LNAI), vol. 5012, pp. 14–25. Springer, Heidelberg (2008)
4. Jaccard, P.: Étude comparative de la distribution florale dans une portion des alpes et des jura. Bulletin de la Société Vaudoise des Sciences Naturells 37, 547–579 (1901)
5. Koestler, A.: The Act of Creation. Macmillan (1964)
6. Kötter, T., Berthold, M.R.: (Missing) concept discovery in heterogeneous information networks. In: Proceedings of the 2nd International Conference on Computational Creativity, pp. 135–140 (2011)
7. Kötter, T., Berthold, M.R.: From Information Networks to Bisociative Information Networks. In: Berthold, M.R. (ed.) Bisociative Knowledge Discovery. LNCS (LNAI), vol. 7250, pp. 33–50. Springer, Heidelberg (2012)
8. Kötter, T., Thiel, K., Berthold, M.R.: Domain bridging associations support creativity. In: Proceedings of the International Conference on Computational Creativity, pp. 200–204 (2010)
9. Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M., Boyes-Braem, P.: Basic objects in natural categories. Cognitive Psychology 8, 382–439 (1976)
10. Segond, M., Borgelt, C.: Item Set Mining Based on Cover Similarity. In: Huang, J.Z., Cao, L., Srivastava, J. (eds.) PAKDD 2011, Part II. LNCS, vol. 6635, pp. 493–505. Springer, Heidelberg (2011)

11. Wille, R.: Restructuring lattice theory: An approach based on hierarchies of concepts. In: Ordered Sets, pp. 314–339 (1982)
12. Wittgenstein, L.: Philosophical Investigations. Blackwell, Oxford (1953)
13. Zaki, M.J., Parthasarathy, S., Ogihara, M., Li, W.: New algorithms for fast discovery of association rules. In: Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining (1997)