Semantic Analysis and Complex Networks as Conjugated Techniques Supporting Decision Making

Pedro Ivo Lancellotta, Victor Ströele, Regina Braga, José Maria N. David and Fernanda Campos
Post-Graduate Program in Computer Science, Federal University of Juiz de Fora, Juiz de Fora, Brazil

Keywords: Business Intelligence, Complex Networks, Data Visualization, Decision Making, InfoVis, Ontology.

Abstract: The expansion scenario in information technology has led to the need for the analysis of huge amounts of data to manage information for quick decision making. This paper presents an architecture for data visualization that uses semantic and structural interpretation as conjugated techniques for data analysis in different domains, through an interface which supports visualization strategies. A case study was carried out with a specialist in real-world agricultural context using data from dairy cattle to answer our research question. The results demonstrate the feasibility of the proposal.

1 INTRODUCTION

Technology is increasingly present in the daily routine of companies and industries as devices and tools for data manipulation. The great amount of data makes manual analysis a complex activity. The main objective of Business Intelligence (BI) is to support decision making, reduce costs, optimize work efficiency to improve process management (Popovič et al., 2013). Among the techniques to support BI decision-making are Data Mining, Panels, Reports and Data Visualization (Ranjan, 2008).

The use of data visualization techniques, or information visualization, helps the cognitive process in data interpretation and knowledge discovery (Card et al., 2009). These aims to simplify the way these data are presented to managers as well as decision makers in order to be able to observe trends and implicit information in huge volumes of data. Efforts are made to create new mechanisms that optimize information visualization techniques, such as the identification of more appropriate visualization techniques for specific datasets, and the reduction in the amount of information displayed for a greater abstraction of information, assisting user perception (Jonker et al., 2013).

Among the efforts for the interpretation and semantic data analysis, we can highlight the use of ontologies, which have been applied in several domains (Miah et al., 2007) (Zillner et al., 2008) (da Silva & Cavalcanti, 2014) (Liu et al., 2016). Ontologies add semantic value to the data, defining explicit and implicit relationships among them. It can define the knowledge related to a given domain of data, allowing it to be processed computationally, making possible the extraction of new information and assignment of meaning to the data.

When defining an ontology for a domain, the relationships between the domain interest individuals and their characteristics, or in a simplified way, the actors that interact in the domain of application are identified. Thus, it is possible to delineate this ontology in the form of a network, representing individuals and their relationships (Heim et al., 2010). Still according to (Heim et al., 2010) these networks allow a more effective visualization of intrinsic relationships of data, observing the connections between the individuals that compose the network. By characterizing the ontology in the form of a network, it is possible to identify how data are related and to apply complex network metrics, such as centrality analysis, node degree, among other metrics to identify individuals that stand out in the data set.

Several works use ontologies and complex networks individually to provide interpretations of the data, but few explore both together. This paper explores the combination of ontology techniques and complex networks using them together, to complement the information on each one in an automated or semi-automated form.

Ontologies are able to offer benefits to complex networks, such as (i) Abstraction of network nodes with the use of inferences (ii) Aggregation of

Lancellotta, P., Ströele, V., Braga, R., David, J. and Campos, F.
Semantic Analysis and Complex Networks as Conjugated Techniques Supporting Decision Making.
DOI: 10.5220/0006662001950202
In *Proceedings of the 20th International Conference on Enterprise Information Systems (ICEIS 2018)*, pages 195-202
ISBN: 978-989-758-298-1

semantic values for the data network (iii) Classification of individuals, identifying semantic groups that can be analyzed separately. On the other hand, the complex network can also benefit the ontology, such as: (i) Identification of new groups of individuals, based on the analysis of ontology entities, enabling the creation of other groups with similar characteristics identified in the network (ii) Identification of new relationships between data, automating the process of creating new logical rules (iii) Structural analysis of the ontology and creation of new properties of relationships between data, from the use of complex network metrics.

The main objective of this paper is to present an architecture that defines the necessary components for the visualization of information and support for decision making. The architecture uses ontologies and complex networks in a conjugated form, supporting visualization for decision-making.

To evaluate the technical feasibility of the architecture, a visualization platform was developed to support the components of the proposed architecture. In addition, in order to evaluate whether the proposal assists decision making in a real environment, a case study was carried out for an agricultural research company involving data analysis of dairy cattle production.

The main contributions of this work are (i) the proposal of a visualization architecture based on semantic and structural data analysis (ii) Use of ontologies and complex networks as conjugated techniques in order to improve the interpretation of strategic data (iii) Development of an ontology in the agricultural context, related to dairy control, and (iv) Development of a visualization platform to support decision making.

The paper is organized as follows: Section 2 details some related works. In section 3, we present the architecture. Section 4 presents a use case with real data from dairy farms. Finally, section 5 presents the conclusions and future work.

2 RELATED WORK

The main objective of this section is to identify works that are related to the proposed approach in this paper, as well as observing deficiencies found in the context of data visualization. We reviewed works that use information visualization, such as Business Intelligence techniques, to support decision making. We filtered the works that mainly use semantic and complex network analysis as an approach, contributing to the interpretation of information and strategic analyzes for datasets in any domain.

Analyzing huge volumes of data is a challenge in many research areas. Some research works, focus on reducing the amount of data that will be available in the visualization (Gu and Wang, 2011).

Ontologies have the ability to aggregate semantic information to analyzed data, in addition to incorporating several aids to the understanding of the specification of a dataset (Sumalatha et al., 2008). A recurrent use of the ontology is to consistently characterize an application domain. It is possible to represent actors that integrate it and its interrelations creating a unified and consolidated concept of the target domain (Jayaraman et al., 2013).

The use of complex networks allows a structural investigation of the data, so that it is possible to identify individuals who are more influential in the network as well as their intrinsic relationships (Lopes et al., 2010). Still according to (Lopes et al., 2010), in order to identify important individuals within the network, complex network metrics are used, such as node degree, centrality, among others.

Many recent studies have made efforts to identify algorithms and metrics for the identification of relevant individuals and implicit relationships among those individuals in complex networks, for example, social networks (Silva et al., 2017) (Muniz et al., 2017).

Some studies use ontologies and / or complex network metrics to better visualize data and identify important elements for analysis (Queiroz-Sousa et al., 2013) (Ströele et al., 2017), and the creation of visualization tools that behave in an automated or semi-automated way (Sobral et al., 2016). However, few studies use these techniques in conjunction with data interpretation (Heim et al., 2010). Others only use different views on data (Dörk et al., 2012) (Radics et al., 2015) (Soklakova et al., 2016).

The main goal of this article is the integrated use of ontology techniques and complex networks so that one complements the information of the other. This enables the development of a visualization platform capable of inferring new information and rules about a set of data analyzed in an automatic or semi-automatic way.

3 CONCEPTUAL ARCHITECTURE

Based on the literature review, we identified the need to create a visualization mechanism to assist business managers in making strategic decisions that lead to higher quality and productivity. A conceptual architecture that supports decision-making is proposed which can self-feed and produce new information. The general objective of the architecture is to enable data visualization and interpretation from a specific data source from the semantic and structural point of view.

The conceptual model of the architecture is presented in Figure 1. The architecture is divided into layers, presented in detail in the following subsections.

The operationalization of the architecture usage occurs according to the following steps. Initially, (i) data is collected from one or more repositories of the same domain (ii) these data are organized, translated into the ETL layer and then imported into the ontology, in which some domain rules are defined (iii) from the defined domain ontology, it is translated as a network, showing the main relationships between the main actors that constitute the application domain (iv) finally, a complex network is modeled with the individuals of the domain. In this network, the relations can be similar to the connections obtained from the ontology, or new connections between the individuals are defined considering their intrinsic characteristics. In this way, semantic analysis is done for data abstraction in order to summarize, reduce the data volume, highlight the most relevant data and identify new rules or groupings.

3.1 ETL Layer

The ETL layer is responsible for extracting, transforming and loading data for later processing. Data repositories can come from several different sources, but all of them refer to the same application domain, or at least have data related to the domain.

In this vein, the architecture will import data and load the ontology with the individuals, according to the domain rules. The main functionality of the ETL layer is the ability to interpret data from different repositories to a consolidated and well-defined data set through the ontology.

3.2 Processing Layer

The main purpose of the processing layer is to organize and model the data so that it can be sent to the visualization layer. Upon receiving data imported by the ETL layer, the domain ontology is consolidated with all individuals in the data set.

3.2.1 Domain Ontology

Semantic data interpretation is done through the ontology, providing the benefits of reliability, portability, reuse and maintenance.

The ontology is defined in order to consider the interactions between the entities which are present in the data set. *Property chains* are properties defined through a chain of object properties, defining inferences related to these data. The ontology has the power to make the architecture more robust and independent because it is able to interpret data, and infer new knowledge. From this semantic layer, the architecture covers different data domains. In addition, these inferences support analysis based on complex networks in order to reduce the number of nodes to be represented through visualization techniques, as well as to add semantic information to the nodes and edges.

3.2.2 Ontology Network

In this layer, a network was modeled that represents the relationships of the ontology and a heterogeneous complex network.

The connection between elements that compose the ontology naturally creates a network among the actors who are inherent to the observed data domain. The modeling and visualization of this network allows some approaches for data analysis, such as, (i) it illustrates to the decision maker the main links between the elements that compose the data domain, supporting the awareness and creation of new relations from direct observation, and (ii) the use of metrics (explained in section 3.2.4) to identify relevant elements in the ontology.

3.2.3 Individual Complex Network

Nodes in a complex network represent individuals and the connection between them represents different measures that are defined by the user at run time.

This layer uses inferred data from the ontology and reduces the amount of data to be presented to the user. As a result, it identifies most relevant individuals with complex network metrics.

Considering the semantic analyzes, the data chosen by the user are related to the individuals connected through the ontology. From the ontology, relationships between data as well as their *logical rules* are analyzed. *Logical rules* are responsible for synthesizing information, so that individuals can relate in a direct or indirect way.

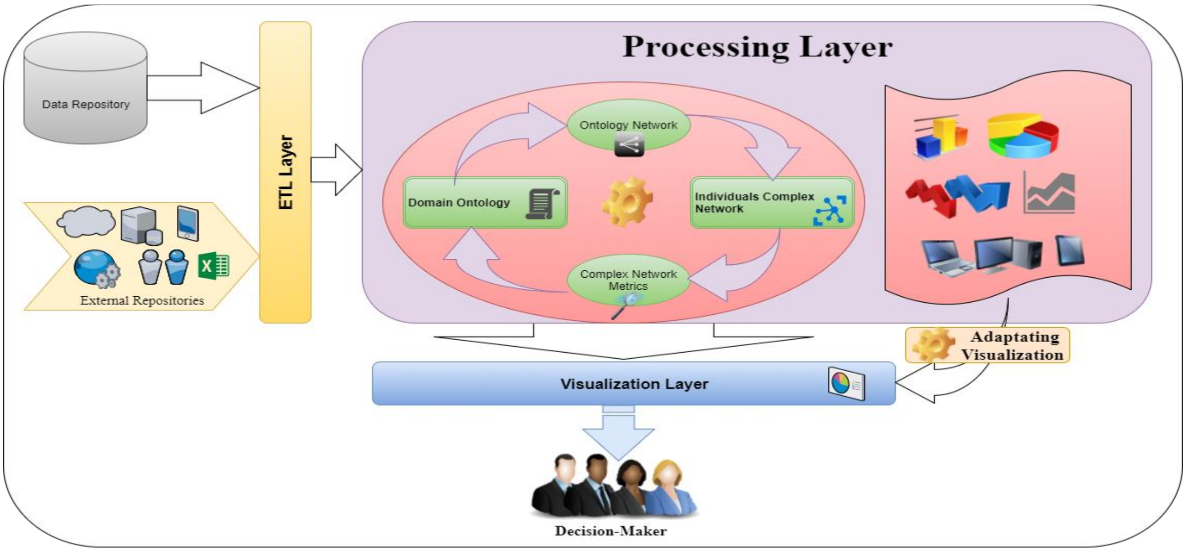

Figure 1: General overview of the architecture.

3.2.4 Complex Network Metrics

Complex network metrics can be used to perform the analysis of highlighted nodes in the network. The identification of more influential nodes is important for a more specific analysis of an element, in order to observe the causes of its prominence with respect to the other individuals. Two metrics are used in the case study in this work: *closeness centrality* and *betweenness centrality* (Wasserman & Faust, 1994).

The *closeness centrality* measure represents the average distance from the minimum paths of one node to all other nodes in the network. Thus, the more centered the node is, the closer it will be to other nodes in the network. It is also possible to observe the nodes with greater *betweenness centrality*, a metric that totalizes how many minimum paths they pass through a specific individual, allowing the identification of entities that manage a greater range of other entities and properties. Using these measures, it is possible to identify new groups of individuals for the creation of new semantic groups in the ontology, that is, elaborating the semi-automated creation of new *logical rules*.

3.2.5 Adapating Visualization

The processing layer will identify visualization techniques according to the data types obtained from the ontology and complex networks. It aims to represent the data as clearly as possible to the decision maker. Thus, the user can adapt the views to obtain better understanding about the data. After choosing the best visualization, the processing layer send this information to visualization layer, which shows to the user.

3.3 Visualization Layer

After the interpretation of the data by the ontology and the structural analysis of the complex networks, there is an interface layer in which the user makes the choice of the data that is pertinent to his/her analysis. He/she also defines the measurements of the data he/she intends to observe, such as numerical values that represent productivity, some attribute that suggests accounting for something, or any other attribute that is measurable.

4 EVALUATION

This section presents an initial evaluation of the proposed architecture, which was made from the first version of the information system for decision-making support. The objective of this evaluation is to answer the research question considering the use of the proposed solution in a real-world context.

The evaluation method adopted in this paper is based on a case study carried out in the context of an agricultural research company with real data extracted from dairy cattle production. A case study is an empirical investigation that relies on multiple sources of evidence to investigate an instance (or a small number of cases) of a contemporary phenomenon within its actual context, especially when the boundary between phenomenon and context cannot be clearly identified (Runeson et al., 2012).

4.1 Case Study Definition and Planning

A survey reported by MilkPoint (*https://www.milkpoint.com.br*) highlights the managerial aspect as the major limiting factor in milk production. There is a lack of approaches that assist the managers in the visualization of data on this field (Miah et al., 2007).

The evaluation scope was defined based on the GQM method (Basili et al., 1994) with the "**purpose of** characterizing **the effect of** semantic and structural analyzes, as aggregated techniques, on data visualization approaches **in** decision making **from the point of view of** business managers and milk institutions **in the context of** milk quality".

The research question to be evaluated in this study is: *How can the semantic and structural data analysis as aggregated techniques, combined with visualization techniques support decision-making*? To collect evidence and respond to the proposed research question, direct observations and interviews were used as data sources.

4.2 Carrying out the Case Study

This section describes the semantic and complex network analysis obtained from the analyzes performed in the developed system. The interpretable elements, observed from the use of ontologies and networks, as well as their objectives and main characteristics are described. The ontology developed for the context of dairy cattle is presented, detailing some of its logical rules and how its use can aid decision-making.

4.2.1 Designing the Domain Ontology

As far as we are aware, no work in the literature was found that used ontologies that represent the process of milk control and quality from dairy cattle. Therefore, an ontology was created using *Protégé* (*protege.stanford.edu*). The creation of this ontology was done with the help of specialists in the area of agricultural research to better detail the process and their needs. The ontology created has a total of 30 classes and 41 logical rules. The classes of *Institution, RegionalCenter, Coordinator, Consultant and Producer* represent managerial aspects and classes, such as: *DairyControl, Production, AnimalBreed, Tank* and *Herd* approach the productive aspect.

Through *property chains* (OWL 2.0), it is possible to obtain the *Farms* that are served by *Institutions* or other *Entities*, in addition to the productions or other information related to their respective *Farms*. In this context, it is possible to associate semantic analysis with managerial processes.

4.2.2 Data Analysis

Initially, a comparative analysis was performed through tables in relation to managerial data. The domain specialists observed the hierarchical relationship between some entities (*Regional Center*, *Coordinators* and *Consultants*). Without the use of ontologies, only the inter-relationships between individuals was identified. Due to the data set volume, it is costly to perform joins between these data in order to identify the attributes and their associations, allowing specialists to analyze only direct relationships. The use of ontologies is able to fill this gap, so that from the inferences made, it is possible to aggregate values to this information. From the platform expert's action to obtain managerial information, a table with the data obtained with the logical rules of the ontology is presented. In this case, the production values of all properties served by management entities were associated. As a result, it was possible to make previously complex observations, such as identifying the production of properties served by a specific *Regional Center*.

For the definition of levels of abstraction and structural analysis a complex network was modeled considering the relationships between the individuals. The management network aims to visualize the mean, total or other aggregations on the values related to the measurement of the dairy controls of animals from one or more herds of a farm. In this case study, the sum of the milk production values of the animals during the last dairy control was used. The values of these measurements are assigned as the weight of the network edges.

With the help of this network, the specialist was able to analyze which farms produce more milk for a *Regional Center* chosen by him/her according to his/her needs of analysis and decision making

In the context of milk quality and the ontology used, there are two main networks that make up the system: management and milk production. The analyzes made by the specialist which focused on the study of the managerial network (Figure 2) were composed of the following entities: *Institution, Regional Center*, *Coordinator*, *Consultant* and *Farm*.

In the management network presented in Figure 2, the following elements are shown: regional centers (in purple), coordinators (in pink), consultants (in yellow) and, finally, farms (in gray). It is possible to observe that, with the increase in the number of institutions, the visualization of the network is compromised according to the volume of data presented to the user. The objective is to highlight the entities that serve farms with a higher milk yield.

For this, the complex network metrics defined above are used. It was observed that nodes with greater *closeness centrality* are related to more *Farms* or with more *Entities*. Thus, comparing the entities of the same type (*Coordinator*, *Consultant*, *Regional Center*, and *Institution*) those that are most influential in the network were identified and presented to the specialist. This is due to the fact that they deal with a higher volume of milk production. The *betweenness centrality* metric was used to identify the nodes with the greatest influence on entity and / or property management. Thus, when measuring the minimum paths between regional centers and farms, it was possible to identify which entities are present in these paths, showing that this entity manages a greater range of other entities and properties.

In Figure 2, the ontology action can be observed in the abstraction of the managerial network. With this abstraction, the manager is able to briefly visualize the measures identified with the total analysis of the complex network. In this case, the edge between two nodes (a consultant and a coordinator) was selected, and the total production of farms consulted by this one is 304829.3 liters of milk.

Thus, it is possible to observe, through the flow that runs in the edges which coordinators or entities influence more, considering the productivity of a certain regional center or any other entity that one wishes to take as the central element of the network. The reduction in the number of nodes represented on the screen (Figure 2) facilitates the visualization by the business managers without compromising the information contained in the complex network.

Finally, this case study demonstrated the use of the first version of the system for visualization and decision making that covers semantic and structural aspects of the data. All the actions carried out in the system, as well as the analyzes and metrics adopted in this case study, were made by a specialist in the agricultural field.

In addition to the direct observations, an interview was conducted. Among the answers given by the specialist, after the use of the developed system, three can be highlighted.

(I) Is the information clearly presented by the system, helping in the decision-making process?

"*These charts enabled me to do an immediate analysis of the production of specific farms comparing to a national scenario, presenting information that supported me in the decision-making process with the milk producers*".

(II) How do you evaluate the use of different visual resources (graphs, tables and network) for decision-making?

"*The graphs are important, but the need to analyze them together stands out*".

(III) Is it possible to analyze data from different institutions and their contribution to national cattle production?

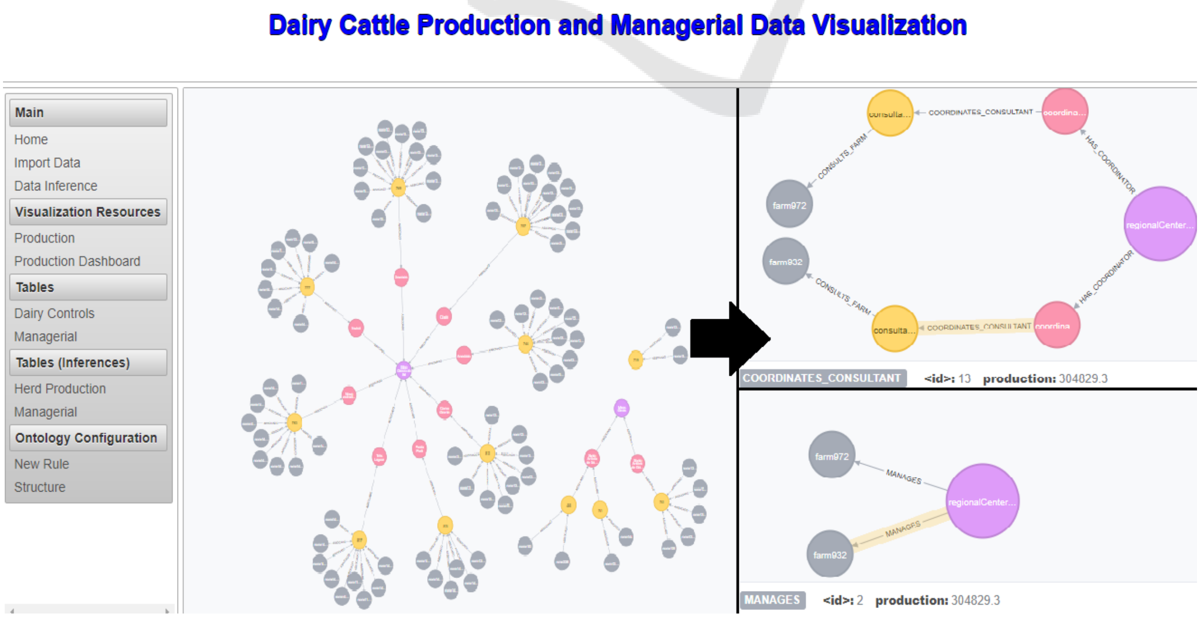

Figure 2: Complex Network generated from available data.

"Milk production presented in a network way allows us to identify and to compare the different entities in relation to its productivity. This is a good contribution of the system for data analysis".

Based on the first answer, there is evidence that the information presented in the visualization artifacts provides better awareness, so that dairy institutions can help milk producers in the administration of their herds and farms. According to the second answer, there is evidence that the components used helped decision making. However, it is still necessary to integrate these components. The third answer showed that the use of semantic analysis and complex networks as conjugated techniques can help in decision-making process.

In order to answer the research question, we observed (i) the specialists' interactions with the system to obtain information; (ii) the analyzes carried out by them for decision-making in the context of milk quality; (iii) and the answers given by them in the interview at the end of the case study. The specialist's answers go to what was observed during the case study, pointing to the proposal feasibility.

Considering these points, it is possible to affirm that there is evidence that the semantic and structural analysis of the data, together with visualization techniques are able to aid in the decision making, adding more information to the data. However, further case studies should be carried out for a wider evaluation of the proposal.

As **threats to the validity** of this case study we can cite the reduced number of participants. Through the evaluation it was possible to state that these results could not be generalized. However, it was possible to identify situations in which similar results could be obtained.

5 FINAL REMARKS

This article presented an architecture for an information system to support decision making, based on components that use complex networks and ontologies as conjugated techniques to support the decision-making process for business managers. The use of the ontology together with the complex networks provides a semantic and structural analysis of the data, so that the visualization strategy brings forward additional information from the data used. A classical problem in analyzing complex networks is the amount of information presented to users. In this sense, the ontology collaborates by providing abstractions in the network, so that it is possible to omit some nodes without the loss of important information for decision-making. Complex networks also collaborate in the creation of new logical rules for the previously defined ontology.

A first version of the system was developed, and a case study was carried out with the purpose of evaluating the architecture. The use of ontologies and complex networks of conjugated form adds value in decision making. The results demonstrate the feasibility of the proposal, but this is only an initial evaluation. As future work, we intend to expand the developed ontology, to carry out a case study considering a greater number of specialists and, in addition, adopt new metrics to support the analyzes related to the complex networks and identification of new logical rules.

ACKNOWLEDGEMENTS

We would like to thank the CAPES, CNPQ, FAPEMIG, UFJF and EMBRAPA.

REFERENCES

Basili, V.; Caldiera, G.; Rombach, H.: Goal Question Metric Approach, Encyclopedia of Software Engineering (1994), John Wiley & Sons, Inc., p. 528-532.

Card, S., Mackinlay, J., and Shneiderman, B. (2009). Information visualization. Human-computer interaction: design issues, solutions, and applications, 181.

Da Silva, M.A.A. and Cavalcanti, M.C. (2014). Combining ontology modules for scientific text annotation. Journal of Information and Data Management, 5(3):238.

Dörk, M.; Carpendale, S.; Williamson, C. (2012) Visualizing explicit and implicit relations of complex information spaces. Information Visualization, v. 11, n. 1, p. 5–21.

Gu, Y. and Wang, C. (2011). Transgraph: Hierarchical exploration of transition relationships in time-varying volumetric data. IEEE Transactions on Visualization and Computer Graphics, 17(12):2015–2024.

Heim, P., Lohmann, S., and Stegemann, T. (2010). Interactive relationship discovery via the semantic web. The Semantic Web: Research and Applications, 303-317.

Jayaraman, P. P., Gunasekera, K., Burstein, F., Haghighi, P. D., Soetikno, H. S., and Zaslavsky, A. (2013). An ontology-based framework for real-time collection and visualization of mobile field triage data in mass gatherings. In System Sciences (HICSS), 2013 46th Hawaii International Conference on (pp. 146-155).

Jonker, D., Langevin, S., Bozowsky, N., and Wright, W. (2013, January). Aperture: An open web 2.0 visualization framework. In System Sciences (HICSS), 2013 46th Hawaii International Conference on (pp. 1485-1494).

Kawasaki, M., Kawamura, S., Tsukahara, M., Morita, S., Komiya, M., and Natsuga, M. (2008). Near-infraredspectroscopicsensingsystemforon-linemilkqualityassessment in a milking robot. Computers and electronics in agriculture, 63(1):22–27.

Liu, Y., Wang, G., Jiang, Z., Qian, J., and Chen, Y. (2016, October). An integrated decision support system for emergency evacuation management. In Systems, Man, and Cybernetics (SMC), 2016.

Lopes, G. R., Moro, M. M., Wives, L. K., and de Oliveira, J. P. M. (2010). Cooperative Authorship Social Network. In AMW.

Miah, S. J., Gammack, J., and Kerr, D. (2007). Ontology development for context-sensitive decision support. In Semantics, Knowledge and Grid, Third International Conference on (pp. 475-478).

Muniz, C. P. M. T., Choren, R. and Goldschmidt, R. R. (2017). Using a Time-based Relationship Weighting Criterion to Improve Link Prediction in Social Networks.

Petersen, Kai. "Measuring and predicting software productivity: A systematic map and review." Information and Software Technology 53.4 (2011): 317-343.

Popovič, A., Hackney, R., Coelho, P. S., and Jaklič, J. (2012). Towards business intelligence systems success: Effects of maturity and culture on analytical decision making. Decision Support Systems, 54(1), 729-739.

Queiroz-Sousa, P. O., Salgado, A. C., and Pires, C. E. (2013). A method for building personalized ontology summaries. Journal of Information and Data Management, 4(3):236.

Ranjan, J. (2008). Business justification with business intelligence. Vine, 38(4), 461-475.

Radics, P., Polys, N., Neuman, S., and Lund, W. (2015). Osnap! introducing the open semantic network analysis platform. volume 9397.

Runeson, Per, et al. Case study research in software engineering: Guidelines and examples. John Wiley & Sons, 2012.

Silva, P. R., Dias, S. M., Brandao, W. C., Song, M. A., and Zárate, L. E. (2017). Formal Concept Analysis Applied to Professional Social Networks Analysis.

Sobral, T., Galvão, T., & Borges, J. (2016) VUMO: Towards an ontology of urban mobility events for supporting semi-automatic visualization tools. In Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on (pp. 1700-1705).

Soklakova, T., Ziarmand, A., and Osadchyieva, S. (2016, October). Big data visualization in smart cyber university. In East-West Design & Test Symposium (EWDTS), 2016 IEEE (pp. 1-9).

Ströele, V., Campos, F., David, J. M. N., Braga, R., Abdalla, A., Lancellotta, P. I., Zimbrao, G., and Souza, J. (2017). Data abstraction and centrality measures to scientific social network analysis. Computer Supported Cooperative Work in Design.

Sumalatha, M. R., Ravi, A., Aravind, M., & Prasanna, N. (2008). Collective Intelligence in Distributed Systems and Semantic Data Visualization. In Emerging Trends in Engineering and Technology, 2008.

Wasserman, S. and Faust, K. (1994). Socialnetworkanalysis: Methodsandapplications, volume 8. Cambridge university press.

Zillner, S., Hauer, T., Rogulin, D., Tsymbal, A., Huber, M., and Solomonides, T. (2008). Semantic visualization of patient information. In Computer-Based Medical Systems, 2008.