

Investigation of Graph Mining for Business Processes

Azeem Lodhi, Gamal Kassem, Veit Koeppen, Gunter Saake
Department of Technical and Business Information Systems
Faculty of Computer Science
Otto-von-Guericke-University Magdeburg
Universitätsplatz 2, 39106 Magdeburg, Germany
e-mail: {firstname.lastname}@ovgu.de

Abstract—Business process management and business intelligence are fields which gain a lot of attention in recent years. These techniques try to improve not only efficiency of processes but also save considerable cost. Graph based representation of concepts (objects, data) are also used in business domain to support aforementioned techniques. Graph mining methods are successful in many fields for discovery of new relations, knowledge, and visualization. In this paper, we briefly discuss the fields in which graph mining is successfully applied. We also discuss challenges of applying graph mining in business processes and what are the benefits.

Keywords—Graph mining, business processes, graph mining applications, business intelligence, business process analysis

I. INTRODUCTION

Business processes are important elements for the success of companies and to perform operations in a organization. Due to this reason, enterprises always try to perform business processes in an effective and efficient manner. Different technologies are used to implement and manage business processes successfully. Workflow management systems are used to define, manage, execute, and monitor the business processes where IT can support in an efficient way. Workflow, itself, is defined as automation of a business process in a whole or a part during which documents, information, and other elements are passed through resources according to procedural rules [1]. To understand the execution of business processes and rules, a considerable amount of research is carried out in this field. Modeling languages and analytical methods are proposed to provide better capabilities for understanding and improvement in business processes.

The representation of concepts (objects, data) based on graphs is successful in many fields, for example, chemical structural analysis, bioinformatics, predictions like successful-movie predictions [2], and successful termination of processes [3]. In such fields, the relational information between entities and their attributes is very important as it helps for discovery of new knowledge. In business process management, this new knowledge can give answers to various questions related to business process improvement as stated in the following.

Which activities are frequently executed? Which path is

frequently used to accomplish defined tasks? What are the common characteristics and relationships between activities, business objects, and their flow or relation (business process executions)? What common features can we discover between successful and unsuccessful scenarios? Prediction on a certain business flow whether it will lead to a desired state or failure. What commonalities can we find between the executions of business applications by users? How should an organization be structured to get maximum benefits from the employees (dealing with social network analysis)? What would be the next information request from the user of a system during business process execution?

Some of these questions are investigated by researchers and we discuss them in Section V while others are still needed to be investigated. This paper demands further investigation in graphical representation of business process elements (objects, data). Mining the relational knowledge in business processes requires data to be presented in such a way that it preserves not only the relational information but also supports mining tasks effectively and efficiently [4].

The structure of the paper is as follows: we briefly discuss the basics of business processes and graph mining in Section II followed by Section III which discusses possible benefits of applying graph mining in the business process management domain. Challenges of applying graph mining on business processes are discussed in Section IV. Section V, provides a brief overview of related work of graph mining in business processes and in some other fields followed by Section VI where we discuss some other issues of graph mining. Section VII summarizes the paper and provides the outlook of our work.

II. GRAPH MINING AND BUSINESS PROCESSES

A graph consists of nodes and edges which connect nodes. Graphical representation is used for the visual representation of concepts. A graph provides a convenient way to represent relation between entities and respective data. In graph mining, we exploit relational information to extract patterns and discover new knowledge. New knowledge can be a pattern such as sub graph, an undiscovered relation to other elements, or on a more abstract level expression of trends in data [4].

Graph based representation and graph mining have been used in various domains where visualization helps for better understanding and management of tasks. For example, in communication networks, it is used for monitoring and management of complex networks, defining connection policies, capturing recurrent patterns in traffic, routing instabilities detections, and extrapolation of networks (designing next generation protocols, etc.). It is also used in the social network analysis domain where various applications are developed for analyzing the relations between different individuals like homeland security (crime and narcotics trafficking connections), computation biology, and web-search [4].

A business process is a set of inter-related activities which are executed in a specific order to accomplish a given task. It can be as simple as fulfilling an order form or complex such as performing the whole order process from selecting suppliers up to paying the bill. Business process management is concerned to define, design, execute, and evaluate the business activities efficiently and utilizing its resources effectively such that the organization's goal is achieved. A support for understanding this complexity is done by modeling business processes. They are graphically modeled in various languages for communication, analysis, and understanding during the business process management life cycle.

Besides different methods to represent business process data, we discuss two methods of representation for a formal analysis of workflows. One method is used in [3], where activities are represented as nodes and other notations like bold and dotted edges are used to distinguish between *AND* and *XOR* structures. The second method is using bipartite graphs called Petri nets [5] in which activities are denoted as places and actions as transition. In this paper we suggest to use the latter notation for graph mining because of mathematical formalism and other reasons discussed in [6]. One extension of Petri nets is proposed in [7] for analysis of business processes with business objects.

A Petri nets is bipartite directed graph, which consists of places (denoted by circles), transitions (denoted by rectangles), and arcs for connection between them. Places represent buffers, storage places where conditions are evaluated and decisions are made, and transitions represent tasks/events/actions. Places can contain tokens which describe the state of the system and values of related attributes at a certain stage [8].

III. BENEFITS OF GRAPH MINING IN BUSINESS PROCESSES

Graphs are useful to preserve the structural relationship between elements, as in the case of chemical compound structure analysis and in bioinformatics, whereas in other conventional approaches, there is a loss of relational information between elements. One of the most interesting benefits that can be achieved by applying graph based represen-

tation in business processes is the discovery of new relations between activities, their attributes, and an overall impact on activities in business processes. Representing a business process and its attributes in a graph saves the relationship between data objects and activities, which is extremely important for discovering relational knowledge. Discovery of relational knowledge using graph mining methods has been discovered by researchers in media industry also [9]. An often used example of graph mining in media industry uses IMDB¹ (Internet Movie Database) website as a resource website. This website contains information about movies and television programs. It provides movies and programs detail information to users through online queries freely. Graph mining is used on IMDB movie database in which movies attributes (actors, director, producer, etc.) are represented as graph nodes, and then graph mining methods are applied to discover new relations and knowledge. This knowledge is used to make a predictions like how much business will a certain movie do during a time-span? Will it be nominated for an award or not? The results of applying graph mining on IMDB and to some other fields are discussed in [4].

Graph mining methods can also be used to predict the information requirements of a user during the execution of activities in enterprise applications. Graph mining is explored for user navigation behavior over the web in [10]. Information generated from business intelligence tools can be combined with graphical models to provide better analytical facilities for decision making. For example, it can be used to predict whether a certain execution will lead to achieve the enterprise's objective or not. Analysis can help users to devise company's best practices that how tasks/activities should be routed based on attribute values.

Graph mining methods can be applied to find out the effect of new business collaboration by analyzing relations and collaboration among existing activities, new activities, and their participants. Similarly, graph based methods can be used to improve the organizational structure, e.g., which employees should work together, see [11]. Application of graph mining in application usage mining [12] can find the deficiencies between user and system interaction and help to improve business-IT alignment.

IV. CHALLENGES

Representation of business processes in graphs is a challenging task as several elements are related with business processes like inputs, outputs, resources, rules, and their attributes. Graph mining requires all involved objects to be represented in graph elements. We discuss some of the issues related with representation of business processes for graph mining as follow.

¹www.imdb.com

A. Granularity

In enterprises, business processes are very large and complex, involving hundreds of resources and activities. Analysts require a different level of granularity to understand the business processes. Business process models are viewed by different stakeholders and so at different levels of granularity (abstract or in detail). Executives are interested in an overall picture of business processes, thus they require a suitable abstraction of BP models, whereas operational managers look for specific details of processes and activities of their responsibility.

Both modeling approaches described in Section II (graph and Petri nets based) do not provide a suitable flexibility in granularity. A new representation of business processes should be proposed, which has enough flexibility and semantics to fulfill the granularity demands of an analyst and at the same time it is formal enough for analyses. This new representation has to be flexible enough that it can represent business process data in graphs so that graph mining methods can be applied within the business domain.

B. Formalism and Structural Relations

Formal representation of business processes has provided various benefits for simulation purposes, non-ambiguous behaviour and conflict resolution, and for other analytical techniques [6]. Besides formal representation of business processes in modeling languages, structural relationships should also be incorporated. In this way, different characteristics of activities and processes can be attached to them, and new relations and knowledge can be extracted. This relational knowledge may take many forms from periodic patterns of transactions to complicate structural patterns of interrelated transactions. Extracting such knowledge requires the data to be represented in a form that it not only captures the relational information but supports mining of this data and comprehensibility of the resulting knowledge. Therefore, before applying graph mining for knowledge discovery and analyzing business processes, modeling languages should be extended with structural information of elements.

C. Sequence Transformation and Semantics

In enterprises where business processes are supported by information systems that record lots of transactional data about the execution of activities, applying graph mining is time consuming, requires high computational effort, and faces complexity problems. Similarly, graphical business models are complex as well. This demands other representation techniques to discover new relations and knowledge. Just applying graph mining methods on business processes and their data sets, i.e., activities/attributes, is not feasible. Therefore, it should be investigated how these methods should be applied on such large data sets and how much would be the information loss if applied on specific data sets.

Another challenge which also demands further investigation is the transformation of graphs (containing relational information of business processes) into other representations like sequences. However, another question arises how should the transformation be carried out in such a way that there is no semantic loss (relational information)? How this can be evaluated whether there is any information loss or not? Such questions are still needed to be investigated, so mining techniques can be applied efficiently to discover new knowledge of process structures.

V. RELATED WORK

Graph based representation of data gains a lots of attention from researchers and is applied to many fields like chemical industry, bioinformatics, and business process management. Due to the scope of this paper which is business process analysis and graph mining techniques, this section provides a very brief overview of related work in this field.

Graph based representation is used in social network analysis in [11] where the authors discover the collaboration patterns and organizational structural information from event logs (business process executions). Different questions can be answered from their method like which persons/groups are working together to accomplish a certain task in business processes? How can the efficiency of employees be improved? Furthermore, social network analysis can be extended to other areas like crime investigations and fraud detection. In [13], authors propose an algorithm to transform the transactional data into graph based representation for graph mining. Further investigation of this transformation and application of graph mining methods is still missing.

In [3], the authors represent activities of business processes as the graph for pattern mining and propose two algorithms to determine frequent patterns of activities. On the basis of their proposed method, authors claim that a user can predict whether a certain execution of the business process will successfully terminate or not. This could be used to determine which tasks need to be executed to provide a benefit to the enterprise. Combining this work with social network analysis and performance analysis, different questions can be answered like how should we allocate resources to business processes for maximum efficiency. Graph mining has also been used to determine user navigational behavioral patterns on the web [10], where authors used probabilistic hypertext grammar and N-Grams to predict the likelihood of a web page to be seen based on previous page visits.

Graph mining is also applied to extract knowledge from consumer behavior in market [14] by using active mining technique in which dynamic process evolvement is dealt with spiral model and levels of user interests. The result of the study shows that graph based representation discovers new and useful relations between item sets with respect to different constraints, which was not discovered earlier. Similarly, IMDB (internet movie database) is also investigated by many

researchers [4] for several purposes by representing movies in the graph to discover new relations between movies, for movie recommendations to users, and prediction of the movie about its success or failure based on their attributes and relation with other movies.

VI. DISCUSSION

Besides the issues we discussed in the earlier sections, we can also apply graph mining in business processes over selected business process executions. For example, it is useful if we first detect frequent transactions occurring in the system and then represent the frequent transaction pattern in form of a graph to discover more knowledge and the relations with other elements which will save some computational over-head. It should be noted, that frequent patterns do not represent business process and potentially useful knowledge is not associated with frequent pattern tasks. Therefore, domain expert knowledge is required to be include to select activities which are not frequent but important for analysis in relation and knowledge discovery.

Despite the benefits of graph based representation of data, we should be careful in applying graphs on different applications has to be carefully considered because for some problems its techniques are NP-hard problems. It is not a good idea to use graph data for the problem of finding whether company's best practice is followed or not?, instead string matching methods are more efficient. The maximum common subgraph can be used to find similarities in executions and patterns. Although this is NP-complete problem but there are optimal and sub-optimal algorithms as discussed in [15] where a comparison of algorithms is presented. Sequence mining techniques can also be considered for such problems.

VII. SUMMARY & OUTLOOK

We provided a brief overview of graph mining in various domains. In this paper, we also discuss possible benefits of applying graph mining in the business process domain. Challenges of applying graph mining in business processes and the need of a new representation for this purpose is also discussed. Our future work involves the specification of new representation which allows the application of graph mining methods for business process analysis. Using the new representation and graph mining we will able to answer the questions raised in this paper.

ACKNOWLEDGMENT

Azeem Lodhi is supported by a grant from the federal state of Saxony-Anhalt in Germany. Veit Koeppen is supported by the German Ministry of Education and Science (BMBF), within the VIERforES project no. 01IM08003C.

REFERENCES

- [1] D. Hollingsworth, "The workflow reference model," Workflow Management Coalition, January 1995. [Online]. Available: <http://www.wfmc.org/standards/docs/tc003v11.pdf>
- [2] D. Jensen and J. Neville, "Data mining in social networks," in *National Academy of Sciences Symposium on Dynamic Social Network Modeling and Analysis*, Washington DC, USA, 2002.
- [3] G. Greco, A. Guzzo, and G. Manco, "Mining and reasoning on workflows," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 4, pp. 519–534, 2005.
- [4] D. J. Cook and L. B. Holder, *Mining Graph Data*. Wiley & Sons, 2006.
- [5] C. A. Petri, "Kommunikation mit automaten," Ph.D. dissertation, University of Bonn, Bonn, Germany, 1962, (In German).
- [6] W. van der Aalst, "Three good reasons for using a Petri net based workflow management system," in *Proceedings of the International Working Conference on Information and Process Integration in Enterprises*, S. Navathe and T. Wakayama, Eds., Camebridge, Massachusetts, 1996, pp. 179–201.
- [7] A. Lodhi, G. Kassem, and C. Rautenstrauch, "Modeling and analysis of business processes using business objects," in *Proceedings of The 2nd IEEE International Conference on Computer, Control, and Communication*, Karachi, Pakistan, Feb. 2009.
- [8] W. van der Aalst and K. van Hee, *Workflow Management: Models, Methods, and Systems*. MIT Press, 2002.
- [9] "The Internet Movies Database," <http://www.imdb.com>.
- [10] J. Borges and M. Levene, "Data mining of user navigation patterns," in *WEBKDD '99: Revised Papers from the International Workshop on Web Usage Analysis and User Profiling*. London, UK: Springer-Verlag, 2000, pp. 92–111.
- [11] W. van der Aalst, H. A. Reijers, and M. Song, "Discovering social networks from event logs," *Computer Supported Cooperative Work*, vol. 14, no. 6, pp. 549–593, 2005.
- [12] G. Kassem, "Application usage mining: Grundlagen und verfahren," Ph.D. dissertation, Otto von Guericke Universität Magdeburg, 2007.
- [13] S. Pradhan, S. Chakravarthy, and A. Telang, "Modeling relational data as graphs for mining," in *Proceedings of 15th International Conference on Management of Data*. Mysore, India: Computer Society of India, Dec. 2009.
- [14] M. Kuroda, K. Yada, H. Motoda, and T. Washio, "Knowledge discovery from consumer behavior in an alcohol market by using graph mining technique," *IEICE technical report. Artificial intelligence and knowledge-based processing*, vol. 104, no. 486, pp. 21–26, 2004.
- [15] D. Conte, P. Foggia, and M. Vento, "Challenging complexity of maximum common subgraph detection algorithms: A performance analysis of three algorithms on a wide database of graphs," *Journal of Graph Algorithms and Applications*, vol. 11, no. 1, pp. 99–143, 2007.