VISUAL REPRESENTATIONS IN FINANCIAL PROCESS MINING

Ilona Veitaitė[1], Audrius Lopata[2], Saulius Gudas[3],
[1] Institute of Social Sciences and Applied Informatics, Vilnius university
[2] Faculty of Informatics, Kaunas University of Technology
[3] Institute of Data Science and Digital Technologies, Vilnius university
ilona.veitaite@knf.vu.lt

Abstract.
Process mining, introduced by Aalst in 2004, extracts process-related information from historical event logs and has become an essential tool for data-driven process improvements. Business process mining, an emerging and rapidly growing field of research, aims to analyse business processes by applying data mining and machine learning techniques to event data. Despite its novelty, process mining has quickly gained broad support due to its ability to provide fast, reliable, and ongoing insights for discovering, monitoring, and optimising business processes. Process mining enhances traditional Business Intelligence (BI) tools by offering detailed, micro-level analysis of process behaviour that complements the macro-level insights BI provides on overall business operations. Additionally, it plays a crucial role in digital transformation efforts, delivering deep insights that drive operational excellence. By bridging the gap between process science and data science, process mining has become an indispensable tool for fast-growing and ambitious manufacturing organisations. While modern business systems like CRM, Finance Management Systems and ERP capture extensive event data, visualising this data presents significant challenges. Translating complex, abstract process data into intuitive visual formats is critical but difficult, as it requires balancing clarity, accuracy, and comprehensiveness. Effective visualisations must highlight patterns, detect anomalies, and simplify interpretation for auditors and decision-makers, but the complexity of processes and volume of data can lead to misleading visuals. Aligning visualisations with diverse stakeholder needs and ensuring they enhance communication across teams is another challenge. Each task is designed based on the specific problem at hand, with a focus on the visual representation and comprehension of data. Process visualisation involves animating process executions in various formats. It can also be depicted through process cube operations, visualising dimensions of financial data, displaying process models as process maps, or using statistical diagrams and other visual techniques. This research explores these challenges by presenting examples of process mining visualisations derived from historical event logs and discusses how visualisations can better support business decision-making.

With a multitude of process mining tools available, each with its own distinct environment, it becomes overly complex for a financial specialist to directly employ them in formulating data analysis tasks. In previous researches there were presented a user-friendly approach to PM technology implementation for financial data analysis using a multi-dimensional space of financial data (figure 1). According to the user's specific need for financial data analysis, the financial expert selects in the financial data space which FDS dimensions are relevant (will be visible to the PM tool environment) and which dimension members are important for specifying the PM project.

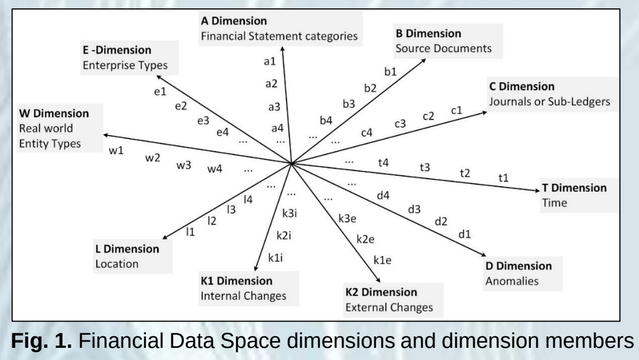

Fig. 1. Financial Data Space dimensions and dimension members

In this case, financial data refers to records of all activities and entities within the company, located in the Finance Data Space. The Finance Data Space encompasses financial accounting objects, which are any named file field (data record field) except for time attributes. It comprises a set of dimensions with meaningful names, each corresponding to a type of Financial Object (FO). The number of dimensions of FO can vary depending on the experts who classify FO. Each dimension represents one axis of the Finance Data Space and consists of dimension members that define the hierarchical structure of the FO. Dimension members are assigned attributes (identifiers) corresponding to the data record fields (quantities, values, or codes). Different dimensions can create combinations of members if they share at least one common attribute (identifier).

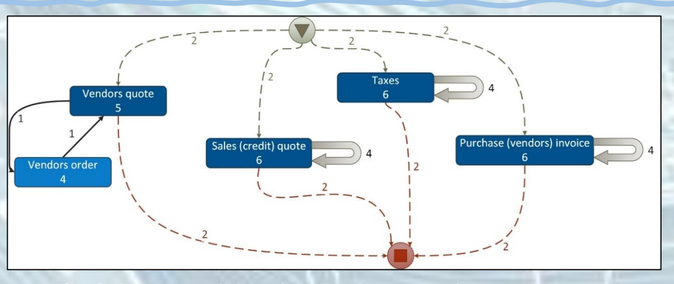

Fig. 2. Process mining visualization example – process map

The process mining execution results are the discovered process model, which is represented graphically as process map and the process model parameters (static data). The process mining model parameters are calculated by the PM tool. An example of discovered process model-map, as one of the examples of process mining visualization is presented in figure 2.

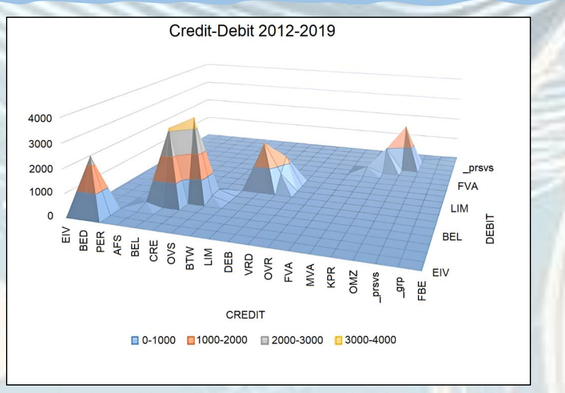

Fig. 3. Debit-Credit transactions by time periods

In the experiments with the expenditure event log there were developed two projects: developing expenditure meta-model Nr.1 of the mandatory sequence of records of Source Documents in Data Base (in JournalNames and Account Names) and developing Expenditure meta-model, Nr.2 of the mandatory sequence of records in [Journal-Names + Account Names] in Data Base (associated with the Source Documents). Despite the visualization result process map (figure 2), another visual result was presented (figure 3). The pattern discovered by the process mining tool is related to a real dataset, making it possible to visualize Debit-Credit transactions over time periods. This finding is a big advantage for financial analyst.

By its nature, records of the General Ledger correspond to the Directed Graph from Graph Theory, where accounts of the General Ledger could be considered as nodes and operations as edges. Therefore, some typical algorithms from Graph Theory could be used for Financial Accounting data analysis.

However, a graph containing all flows between accounts could be very complicated and hardly readable. Therefore, certain algorithms could be used to simplify the graph by providing only essential information, making it more beneficial. One such algorithm is the minimum spanning tree.

A minimum spanning tree is a subset of the edges of a connected, edge-weighted graph that connects all the vertices together without any cycles and with the minimal possible total edge weight.

In the case of financial data from the General Ledger, the sum of operations between accounts (nodes) should be used as the edge weight, and for analysis, it could be more interesting to evaluate the maximum spanning tree – flows with the largest amounts between accounts.

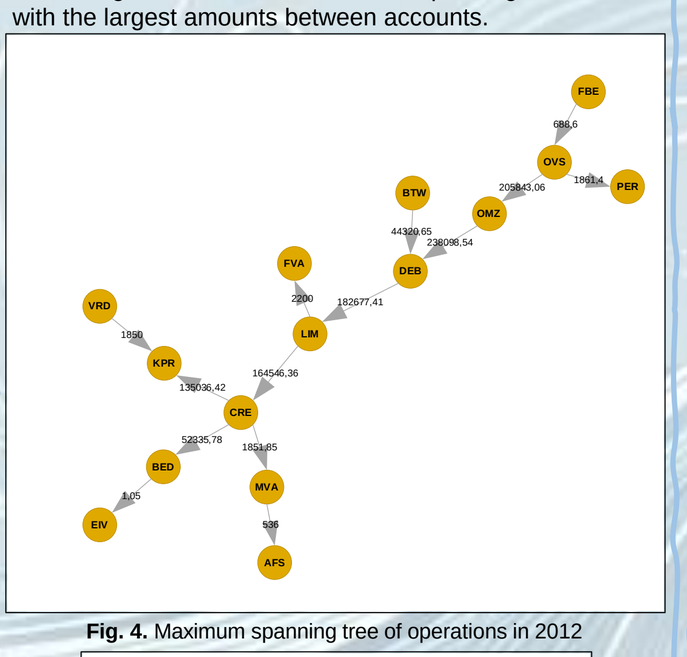

Fig. 4. Maximum spanning tree of operations in 2012

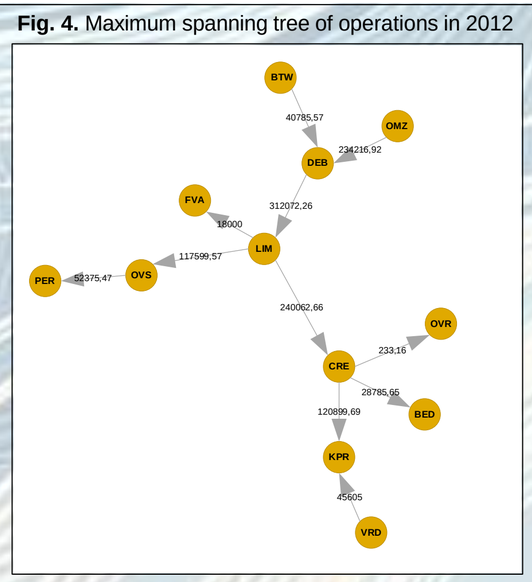

Fig. 5. Maximum spanning tree of operations in 2013

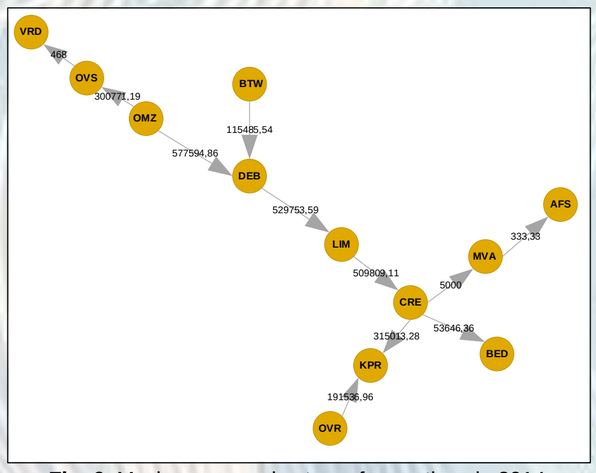

Fig. 6. Maximum spanning tree of operations in 2014

As figures 4, 5 and 6 represent examples of the maximum spanning trees for the one company's operation for 3 years, maximum spanning trees provide an easy-to-understand depiction of the main funds flows between accounts. By comparing graphs from different periods, changes in accounting or business activities could be identified.

Sankey diagrams could be used to visualize graphically accounting operations. Such diagrams allow financial auditors to have a rapid overview of funds transfer between General Ledger accounts and get better understanding of activities during selected period. These diagrams could provide insights of business cycles of the company. There were analyzed company's 7 periods/first months of the year. Figures 7, 8, 9 represent Sankey diagrams for the same company for 1, 4, 7 periods.

During the 1st period, few type of operations were dominating – from Accounts Payable to Cost of Sale and from cash and cash equivalents to Financial Fixed Assets.

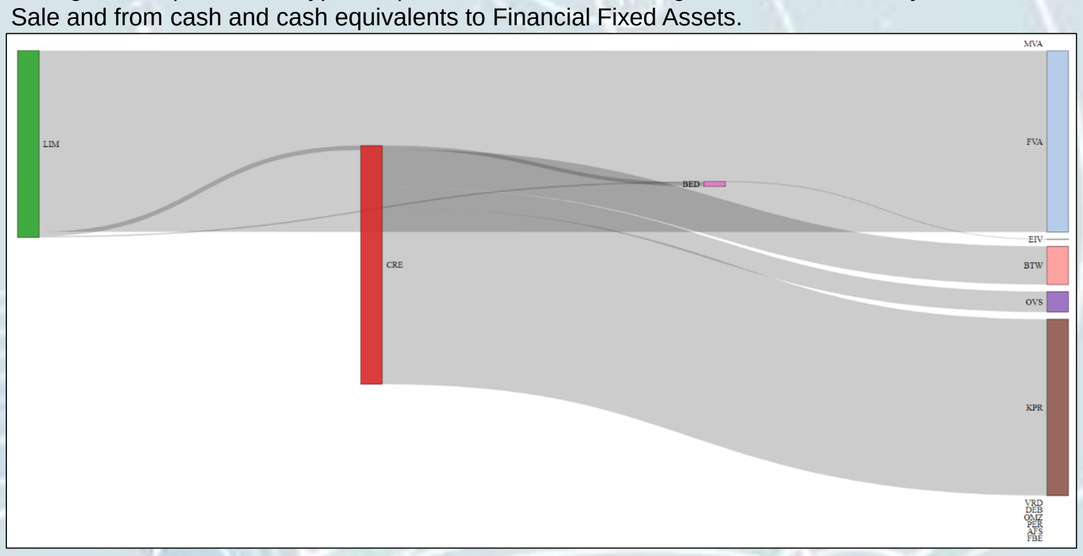

Fig. 7. Sankey diagram of operations in the period 1

In the 2nd period, the main funds' transfer between General Leger accounts demonstrates that the main operations were payments to the creditors (LIM -> CRE) and payment of the taxis (BTW). During the 3rd period, new invoices from the creditors were posted to the General Ledger. During the 4th period, the company sold goods/services and received payment from Customers (operations OMZ->DEB, DEB->LIM). Biggest part of the received Cash was used to pay to Creditors (LIM->CRE).

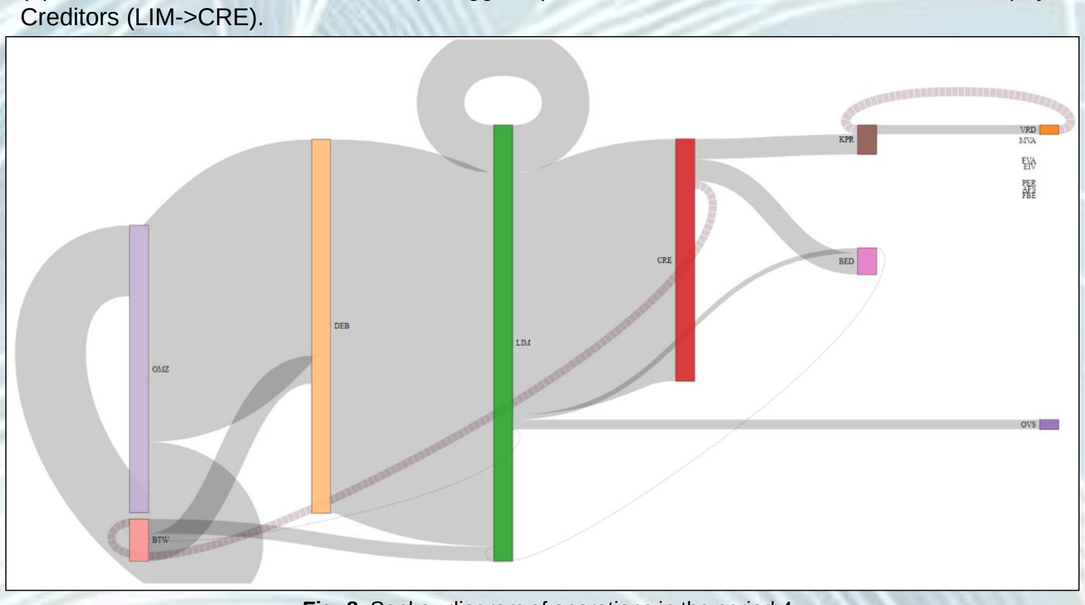

Fig. 8. Sankey diagram of operations in the period 4

During the 5th period, company sold goods/services to Customers (OMZ->DEB). During the 6th period, company issued new invoice to customer, as well as made invoice correction (OMZ->DEB). During the 7th period, company has received payment from Customers (DEB-> LIM) and large part of received funds used to pay to Creditors.

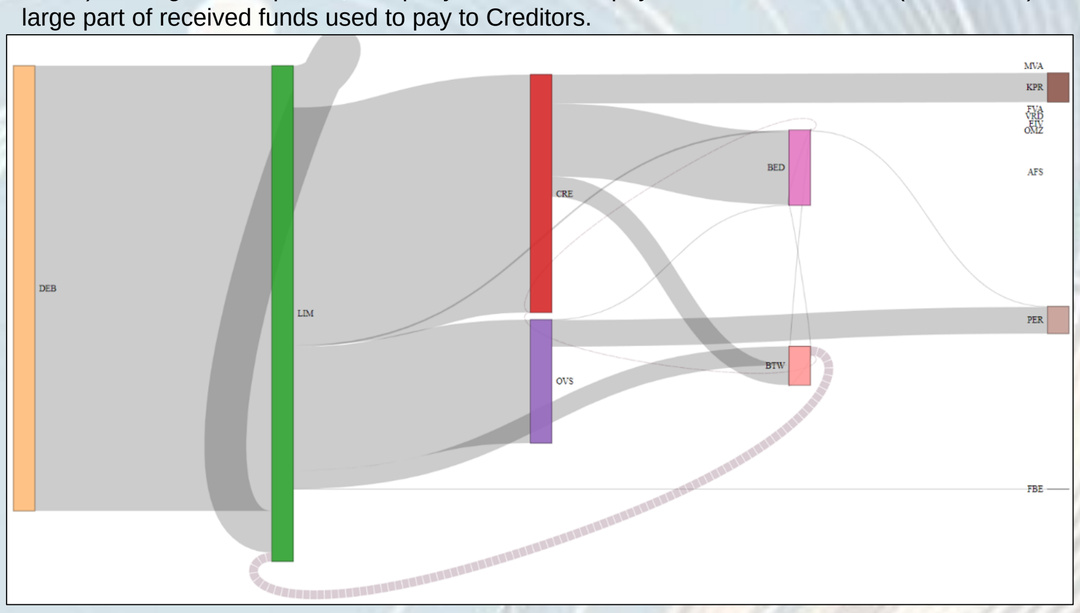

Fig. 9. Sankey diagram of operations in the period 7

As follows from diagrams of all 7 periods, following insights could be made:

1. Company purchases from creditors and pays next month;
2. Sales to Customers are done once in 2 months, after significant expenses to creditors;
3. Customers pay next month after invoices.

Sankey diagrams could be used for understanding accounting and business process from visual point of view and financial analysts could benefit from more clear picture due to visual presentation of operations.

Conclusions

- ✓ Every company is aware of the rapidly expanding volume of data they possess, and the challenge lies in preprocessing this data for subsequent analysis to gain competitive advantages. Process mining endeavors to uncover, monitor, and enhance actual processes by extracting insights from readily accessible event logs within contemporary information systems, gathering data from these logs sourced from a business's systems or a data warehouse, with the essential data prerequisites including the activity name, a unique case ID, and a timestamp for each case to map a process.
- ✓ In financial process mining, a viable way is: to develop normal models (patterns) for financial activity transactions (workflow), as they become criteria for evaluating real data, or to calculate indicators by recurring periods (years, months), because then it is possible to compare the values of indicators and determine changes (in absolute and relative values) by periods.
- ✓ Several visualization examples were presented: firstly, the process mining tool's discovery process map and discovery of patterns related to real datasets enable the visualization of Debit-Credit transactions over time periods, offering a significant advantage for financial analysts. Secondly, maximum spanning trees offer an easy-to-understand depiction of the primary funds flows between accounts, facilitating the identification of changes in accounting or business activities through graph comparisons. Thirdly, Sankey diagrams provide a clear visual representation of accounting and business processes, aiding in understanding operations. As visualization commonly focuses on the visual presentation of results and aids in improving the understanding of data, it is possible to assert that visualization may significantly enhance the work of financial analysts.