Using Process Mining approaches for development and analyze of logistic network

Julia Rudnitckaia

Brno, University of Technology, Faculty of Information Technology, irudnickaia@fit.vutbr.cz

1

Abstract.

In this article it will be showed use-case of Process Mining in terms of existing logistic scheme of seaport. By help this approach, bottlenecks and other weak places of process will be detected. Moreover, resources and their interaction will be described. Therefore, not only main activities, but also attributes of cases take place in this article. The necessity of introduction business process management in the port system will be grounded. Also appropriate pattern for creation unified information logistic seaport system based on Workflow nets will be proposed. Then, the methods of analyze of existing transport seaport model will be considered. Finally, we will be able to see how using PM technology can change and improvement our processes in the seaport system.
Generally most of information here is based on Massive Open Online Course: "Process Mining: Data science in Action" and official process mining's website.

Keywords: Petri Nets, Process Mining, Workflow, Process Models, Information logistics system, Seaport infrastructure, Operational management, Prediction of processes

Introduction

Nowadays, there are various tools and methods to improve productivity and efficiency of manufactures, firms, various service services. Under the press of constantly competition they must search for new ways of optimization of work process. Reducing the working staff or buying more powerful hardware give temporary results and become vestiges of the past. Today, it's more necessary so to organize and possibly automatize workflows that can lead to maximize of main key performance characteristics. Rudimentary and repetitive tasks are increasingly being automated. For these purposes, such broad area as Data science was created. And, recently, one part of its namely Process Mining become more and more popular due of simplicity and flexibility using. Interesting in terms of Data science, especially in improvement quality and clearness of data analyze, it raises doubts and mistrust among entrepreneurs and stakeholders who are used to usual tools of data mining and business intelligence. Despite on simplicity, Process Mining is very powerful method and non-experienced user likely will have rough mistakes in a results interpretation. It also can be used in online modes that help to monitor processes in real time and to detect problems immediately.

In fact, the main obstacle to the development of maritime transportations in the Russian Federation is not so much the absence of the necessary infrastructure and capacities, as inefficient management of existing port complex, therefore the main task of the scientific community and representatives of maritime business can be considered the development of new approaches of management and control of seaports.

Applying Workflow-net to design seaport logistic system

The first business process management systems emerged over ten years ago, however, and today the situation in this field remains difficult and it's developing very dynamically. Many issues related to these systems are actively engaged in a variety of organizations: the leading software developers, international consortiums, standardization committees, as well as academics (especially mathematics – specialists in graph theory and process algebra).

Experts predict a significant increase in the share of business process management systems on the market of enterprise-wide information systems in the coming years.

At actual stage of development of science the process approach to company management is recognized as the most promising. According to him the entire activity of the company is represented as a set of business processes. Business process – a time-ordered set of tasks performed by both human and information technology systems of the enterprise, which is aimed at achieving the previously known targets in the given time limit. In the case of application of this approach to the seaport, purpose is handling and transshipment cargo traffic through the port logistics infrastructure.

As a result of the task there is a need in flexible information systems that implement the process approach to management - they are called business process management systems, or BPM-systems (Business Process Management). An important feature of such software is supporting of the rapid development and changes in business processes without updating specialized code, using a graphical development environment that is available to management personnel and operational control flows specialists.

Business Process Management – is actively developing field, and today an acute issue is standardization of definitions, theories and management approaches in the workflow systems. Various authors widely use such concepts as the BPM-systems, WorkFlow-systems, DocFlow-systems, EAI (Enterprise Application Integration) and so on.

As the process control systems at the seaport seems best to use a subclass of Petri nets – WorkFlow-systems.

Petri net PN (also P/T nets) is description of process in terms of places (P), transitions (T) and arcs. The semantics is always formally defined. For Petri nets the following conditions must be performed:

- there is only one input place «i» such, that not exist input arcs in «i»;
- there is only one output place «o» such, that not exist output arcs from «o»;
- every place given net is found between input place «i» and output place «o».

Considering the structure of the seaport organization, the start position «i» can be taken the raid through which each ship with installment of cargo passes of the consignment, the end position «o» – leaving the territory of the port by any form of transport. This distribution is acceptable in the processing of import flows, however, for export flow start and end positions are swapped without changing the overall structure of the network.

When using WF-nets, the movement of cargo through the seaport infrastructure can be represented as a path (set of nodes), connecting elements of intraport cargo-transferring chain. During the passage of the track the installment of cargo (token in the terminology of the WF networks) passes through the different elements of cargo-transferring process (places) by performing cargo handling operations (running of the tokens along arches that connecting nodes).

Since the process of passage of the cargo through the seaport is closely related to registration of the necessary documentation, it makes sense in the simultaneous using WorkFlow and DocFlow systems (document flow). Any activity can be represented in the form of documents transmitted between their editors on a certain scheme in accordance with the prescribed rules.

For DF-systems as well as systems for WF-systems, there are diagrams look like graphs, which consist of nodes interconnected arches. However, in this case not tokens are running, but set of documents. As a rule, in the DF-systems data are contained inside the documents that are moved directly according to document flow net.

DF-systems replaced the paper-based document flow. Thats why they have follow limits: the document can perform a limited set of actions: approve / reject, endorse, remove, edit, and so on. Usually, document management systems are complemented by systems for storage of patterns of paper documents. The main advantage of document management systems is the possibility of rapid implementation in the enterprise, surely, if it has adjusted workflow at a high abstract level.

Taking into account that for today the seaport already has a set of software for organizing, managing and accounting activities of the port, the purpose of the implementation of DF and WF-network is to combine all existing systems and standardization of processes within unified information environment.

By implementing a unified approach to managing workflows, any operation is seen as the follow set of activities: the required case, the resources for performing this case, assigned executer, timestamps for beginning and end of the case, and other attributes.

The main advantage of building information and logistics system for the seaport is possibility of deep analysis of the existing statistics within a common standard of the enterprise, as well as the possibility of an efficient management of the entire system on the basis of prediction the possible problems associated with the transport of cargo through the port infrastructure.

Further, we will consider the basic methods of data mining and extend the WF-net with additional information from event logs based on the specifics of the seaport.

Applying methods Decision tree and Decision points

Consider an arbitrary piece of WF-network with a choice between two available operations (Fig. 1).

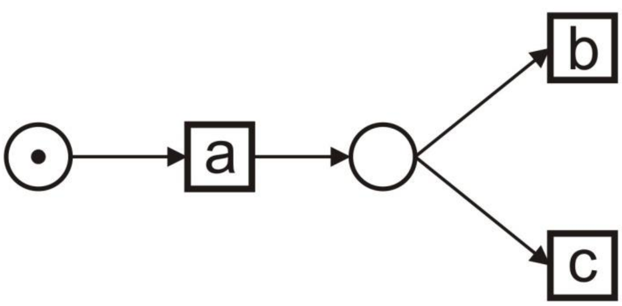

Figure 1: piece of WF-network

In order to successfully predict which of the available paths will be chosen for further performing operations, it's necessary to consider all cases in the event log that are connected with a given decision point in the algorithm. Consider the abstract event log, in which the data of cargo operations are listed (Tab. 1).

Table 1 Example of event log

CargoID	Activity	Executer	Start time	Cargo type	Tonnage (tn)
1	a	Executer №1	8 : 11	wood	5000
2	a	Executer №2	8 : 30	pipes	12300
2	d	Executer №3	9 : 05	pipes	12300
1	b	Executer №1	10 : 11	wood	5000
3	a	Executer №2	10 : 23	grain	9000
3	b	Executer №4	11 : 40	grain	9000
1	d	Executer №3	12 : 10	wood	5000
2	c	Executer №4	12 : 16	pipes	12300

During analysis it's important to identify patterns of selection a particular activities, that follow after activity «a», in other words, which set of attributes from event log led to the implementation of the activity «b» and which – to activity «c». For clearer understanding of this example, separate table, (Tab. 2) describing what activities and with what attributes were performed after completion activity a, is pointed below.

Table 2 The state after completion activity a

CargoID	Executer	Start time	Cargo type	Tonnage (tn)	Next activity
1	Executer №1	10 : 11	wood	5000	b
2	Executer №4	11 : 40	grain	9000	b
3	Executer №4	12 : 16	pipes	12300	c

Therefore, it's seen from the table, that activity «a» was performed 3 times, after that activity «b» was executed 2 times and activity «c» - only once. Based on this table, we can apply famous method Decision tree (Fig. 2), that is filtering of activities' attributes.

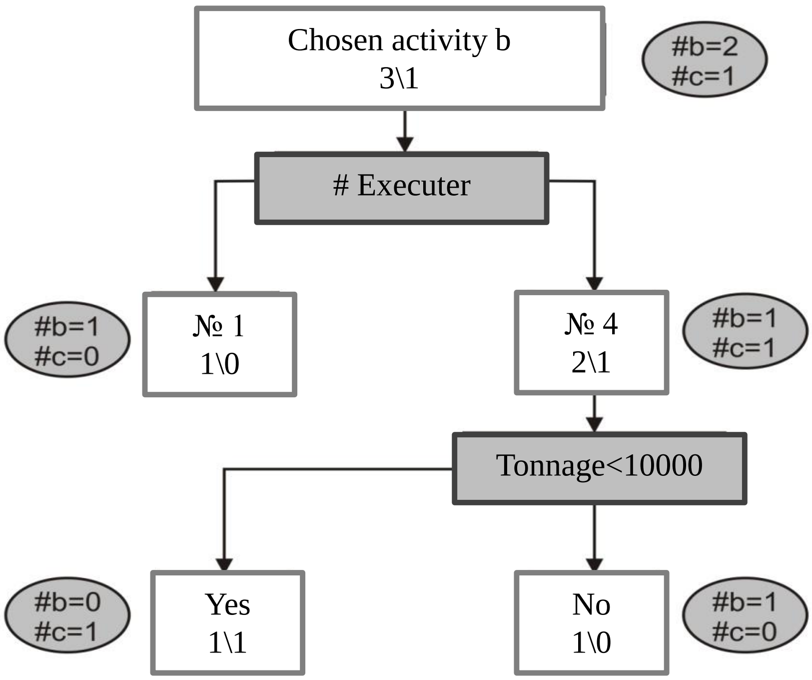

Figure 2 Decision tree for «b» and «c» activities's attributes.

Decision tree method is usually used to reduce the degree of uncertainty select the next activity based on atributes. The entropy is measured in bits of information and is defined by the formula (1):

$$E = -\sum_i^k p_i \log_2 p_i \qquad (1)$$

where, E – entropy – measure for the uncertainly,
p_i - probability of chosen variant

When using several filtration steps one should begin with the calculation of the total uncertainty of the final stage and then, on each following stage one needs calculate the arithmetic average of uncertainty.

When the final value of the uncertainty E = 1 we can talk about complete uncertainty, i. e. about the impossibility of processes prediction. If the value of the uncertainty E = 0, all events are clearly defined, thus prediction of processes will be sufficiently accurate. In practice, the value E is between 0 and 1, as in the analysis of large amounts of the event log data is often impossible to completely remove uncertainty because of the large number of attributes for each case. In our example, a measure of uncertainty before using decision tree method was E = 0.968, after the first step of filtration E = 0.5, after the second stage of filtration E = 0. This means that from this point availability of accurate prediction of the WF-net following activities becomes possible. It is believed that for the successful prediction, it's enough to have at least uncertainty coefficient E = 0.2.

Having examined thus piece of WF-network, it can be concluded that the activity «b» is performed by Executer №1 and for tonnage of less than 10,000 tons. Whereas activity «c» is performed with the tonnage of over 10,000 tons. Found data can be applied to the WF-net, as shown in Fig. 3.

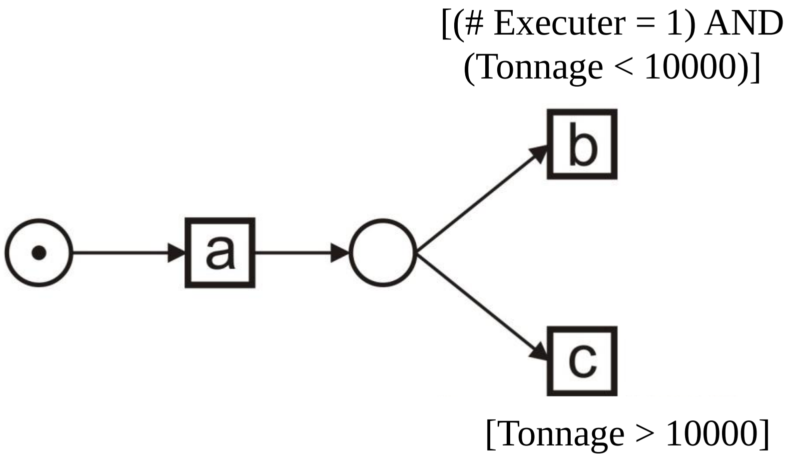

Figure 3 Piece of WF-net with decision point

A similar analysis can be performed for each decision point of information-logistic network, however, there may be cases where it is difficult to give clear preference to any branch of the net, as they are all, in some degree, probable. In this case, the model is added to a set of attribute, with which both paths of WF-net are possible. For the description of attribute sets used logic functions AND, OR, XOR.

It is also convenient for analysis to indicate probabilities of activity performing on information-logistic, that is calculated based on the event log, as shown in Fig. 4.

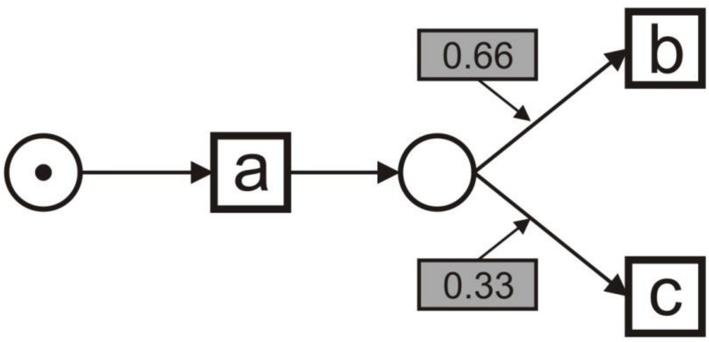

Figure 4 Piece of WF-net with indication of probabilities

The main problem of the decision tree method is redundant detail description of attributes and the ability to "glut" of data, which will lead to a lot of scenarios that will eventually be useful in practice, for prediction and manage freight flows in the seaport. To eliminate the negative effects it should be chosen the most significant attributes and be analyzed only information based on chosen group of event log data.

Bottlenecks mining by help timestamps

The main reason time limits violation is the existence of so-called "bottleneck" – phenomenon where the performance or capacity of an entire system is limited by a single or small number of components or resources. In our case, it's a narrow place where some logistic transportation routes converge and port services cannot handle the incoming flow of goods. Among the main causes of "bottleneck" and, accordingly, insufficient throughput of the cargo chain is usually both limited resources used to perform specified operation, or the lack of efficient use of resources, that is management or resource's issue.

To solve such problems initially one should make a detailed analysis. The sequence of steps for bottleneck analysis is:

1. Input data analysis: timestamps of start/finish and stop/restart for case, taking place in the intraport transport-logistics chain;
2. Creation process trace (an ordering activities) from event log, that has (not always) start/finish and stop/restart for case. Example of such path is showed below.

$$Trace\ №1:\ < a_{start}^{12}, a_{stop}^{27}, a_{restart}^{30}, a_{finish}^{33}, c_{start,}^{33} c_{finish}^{37} >$$

$$Trace\ №2 :< b_{start}^{17}, b_{finish}^{22} >$$

,where a_{start}^{12}- start time activity «a» at t=12;
a_{stop}^{27} – stop time activity «a» at t=27;
$a_{restart}^{30}$- restart time activity «a» at t=30;
a_{finish}^{33} – finish time activity «a» at t=33

3. Analysis of everey trace by timestamps and drawing up the summary table (Tab.3)

Table 3 Time perspective

TraceID	Activity	Start	Suspend	Resume	Complete
1	a	12	27	30	33
	c	35	-	-	37
2	b	17	-	-	22

From this table it is clear that the operation «a» in time between t = 27 and t = 30 has been suspended, so there is the problem of resource for the activity or the problem in person, who processes this activity at a given stage. It also shows that after the operation "a" in Δt = 2 starts operation «c», that says about suspension implementation of the trace 1. Ideally, the time between the various operations of a route should go to zero.

From the table based on the event log, that brings together a large amount of data, you can calculate the average execution time of each operation, downtime port resources, employment's workload, including the port specialists' workload.

4. Adding information about runtime of transport-logistic traces on general process model, that can be presented in other notations such Petri-nets, Workflow-nets, BPMN, YAWL and so forth. Thereby we are extending model (Fig. 5). Besides timestamps, for solving different tasks, one can also add organizational perspective (role, social network), case perspective (decision mining, see Fig. 3) and other perspectives (risks, costs).

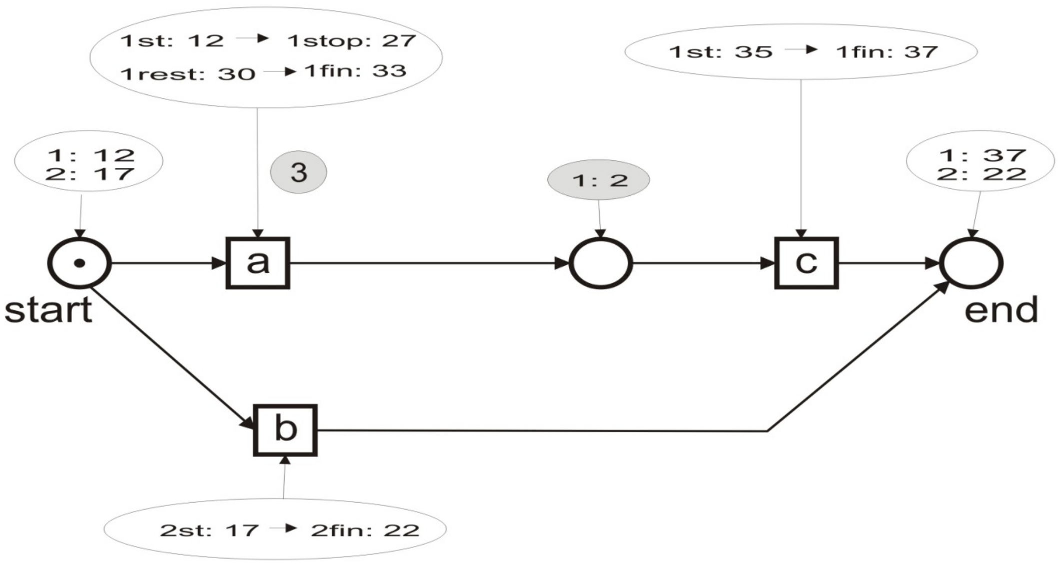

Figure 5 Extended process model

Total transport logistics scheme, built by help the event log allows you to visualize all the processes inside the port infrastructure, to determine at what point a delay of cargo flow is, and to understand the reasons for the delay. Thus, considered logistical scheme is a kind of «road map» of intraport logistics complex.

After identifying bottlenecks in the logistic scheme seaport, one should analyze the resources, that can help you resolve the issue. For these purposes, the most convenient method of analysis is the compilation chart of the activities and diagrams of resources used (see Fig. 6, 7).

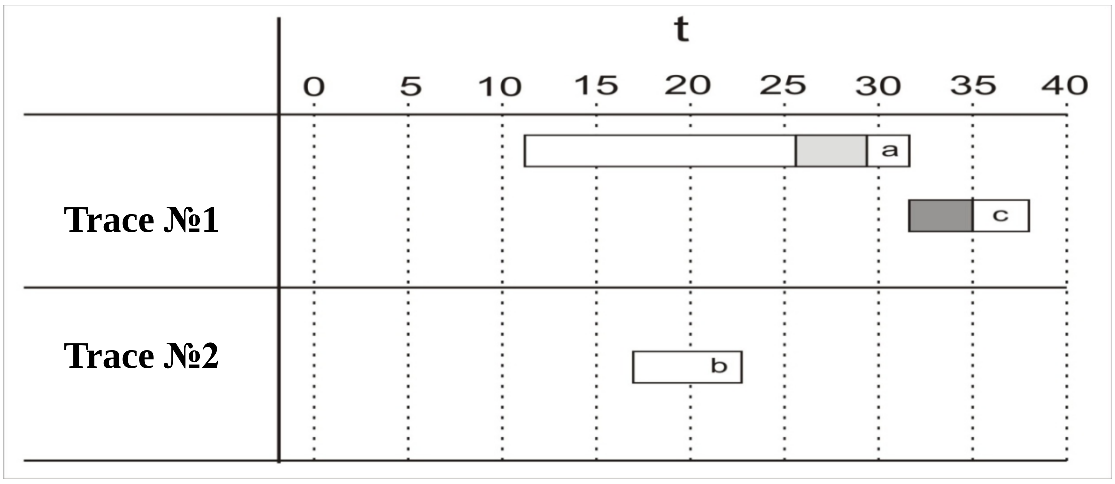

Figure 6 Distribution of acrivities within trace

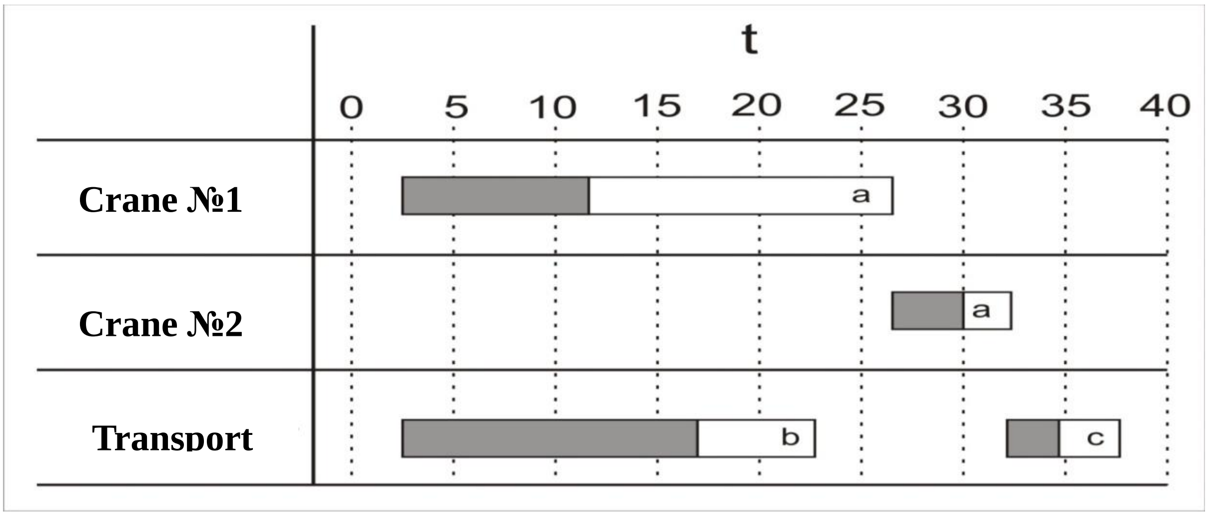

Figure 7 Distribution of resource

Analyzing Fig. 6 and Fig. 7, it can be concluded on the use of allocated resources. Resource "Transport" was able to start the activity «b» much earlier than set to work. Suspending the activity «a» associated with the transfer of work from the resource "Crane №1» to the resource "Crane №2». Thus, building a chart, you can make a full analysis of the employment of resources.

Analysis of employment executors and use of resources based on social network

Another effective and ostensive method of presenting information about the use of sea port resources is a method of creating a social network based on handover of work at the level of roles. This network shows the behavior of the executers during the freight and transport operations within the port.

In this case, the main considered factors are not processes themselves, and executers, used resources and their execute time. Input data for this method are also in the seaport event log.

The seaport social networks represent the interaction between executers or resources, arcs define the relationship between them. The purpose of social network is to find the key executers or resources that have the largest number of applications and have the greatest number of conections with other elements of information-logistic network, as well as to integrate similar executers/resources in clusters. Thus, this technique can help to identify cliques – groups of entities that are strongly connected to each other while having few connections to entities outside the clique. Such type of agregation can show real organizational structure that can be modified and optimized in the future.

To build a social network one should at first determine network elements and their connections. Ussually, for solving this task, a matrix of social relations is used. The matrix shows how many times one executer transfers performing of operation to another within one case (e.g. shipments). Consider the abstract matrix of social relations (Tab. 4).

Table 4 Social network matrix

	Executer №1	Executer №2	Executer №3	Executer №4
Executer №1	0.135	0.225	0.09	0.06
Executer №2	0.225	0.375	0.15	0.1
Executer №3	0	0	0	0.77
Executer №4	0.15	0.6	0	0

The values entered in the table should be greater than or equal to zero, the upper limit value is not available, because work can be passed from one executer to another several times within one case. On the basis of the matrix we build a social network (Fig. 8).

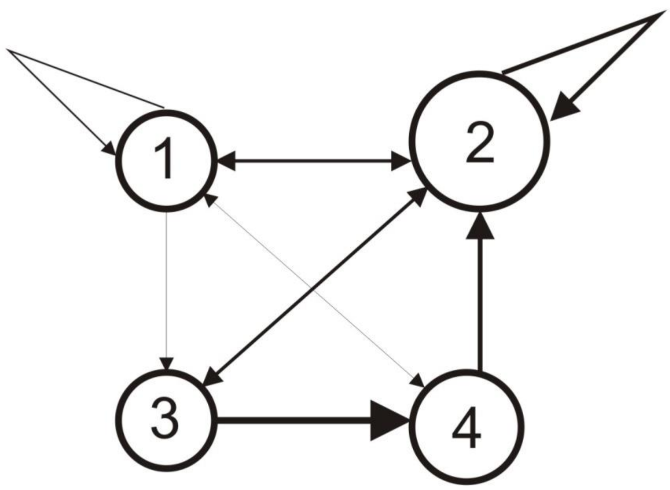

Figure 8 Social network

The weight/size of nodes/circle are based on the number of times a resource having the role performs an activity. the weight of the arcs are based of the average number of times a handover takes place from one role to another per case. the thickness of the arcs is based on the frequency of handovers from one person to another. Therefore, you can quickly and visually to obtain information about the relationship between the performers.

Connecting social network matrix and activities we can obtain resource-activity matrix showing the mean number of times a person performed an activity per case. Accordingly, it will be possible to see persons, who performes identical work (Tab. 5). Using such information, one can learn more about people, machines, organizational structuree (roles and departments), work distribution, and work patterns.

Table 5. Resource-activity matrix

Activity / Executer	a	b	c	d	e
Executer №1	0.3	0	0.34	0.1	0
Executer №2	0.4	0	0.9	0.2	0
Executer №3	0	0.8	0	0	1.1
Executer №4	0	0	0	0	0

From Tab. 5 it's clear, that Executer №1 and №2 perform identical set of operations. Record of operations performed by each executer can be formed as a vector that is the most convenient type of representation for subsequent computer processing. For example, for Executor №1 vector will be as follows:

Executers with an identical set of operations may be combined into a cliques. Elements within the clique are connected to each other much more than with elements outside the clique. So it is appropriate method for description the other seaport departments. Finally, a social network can be represented as follows (Fig. 9).

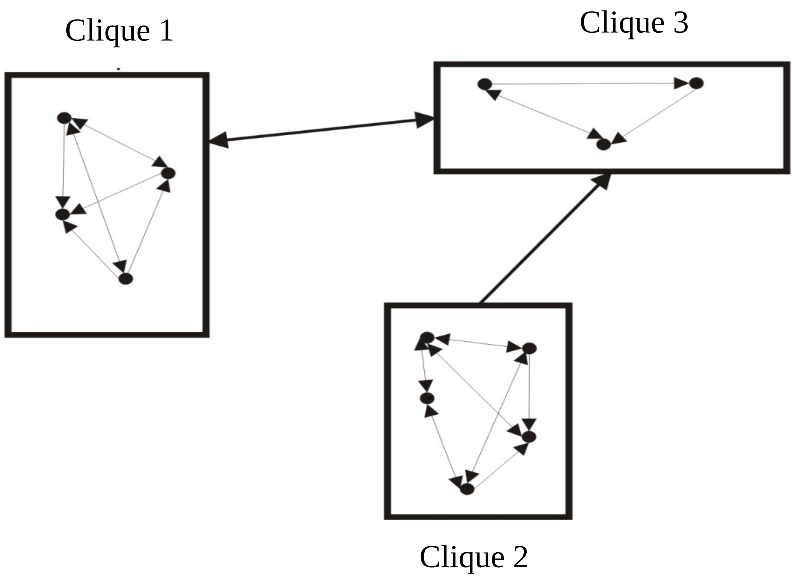

Figure 9 Social network based on similarity of work.

Grouping processes \ resources \ executers allows you to create more accurate specialized models than a common process model. The development of a number of specialized models improves the quality of the simulation of the entire system, due to the increased detail scrutiny of current processes and reduce the measure of the uncertainty in each model.

Conclusion

The implementation of unified information-logistic system based on Workflow-net into the seaport production process will allow to simulate the current work of intraport cargo operations, i.e. to create a process model "As is"; using the listed above analysis models tools and events log, it can be possible to evaluate the effectiveness of structure of the port and parts of the transport infrastructure as a whole, to find bottlenecks and to determine the causes of failures in cargo-transferring processes.

A particular advantage of the implementation of the unified information-logistic system is the ability to get instantly access to data of any seaport substructures. That makes possible to promptly respond to current challenges and to successfully predict the further operational activities of the port complex.

Furthermore, getting information will be used for rational utilization of the seaport resource base, as well as to determine the workload of resources.

References

1. 21st International Petri nets conference. Petri Nets. Introductory tutorial — www.daimi.au.dk/PetriNets/introductions/pn2000_introtut.pdf.
2. Van der Aalst, Process Mining: Discovery, Conformance and Enhancement // Springer Verlag, 2011
3. An introduction to Petri nets http:// viking.gmu.edu/ http/syst511 /vg511/AppC.html.
4. Фиошин М. Основы p-исчисления — http://progr.tsi.lv/research/picalc.pdf.
5. High-level Petri Nets. Международный стандарт ISO/IEC 15909, версия 4.7.1 —www.informatik.hu-berlin.de/top/PNX/pnstd-4.7.1.pdf.
6. Van der Aalst, Hofstede A. H. M. YAWL: Yet Another WorkFlow Language — www.citi.qut.edu.au/pubs/technical/yawlrevtech.pdf.
7. Van der Aalst W. M. P. Pi calculus versus Petri nets: Let us eat «humble pie" rather than further inflate the «Pi hype" — tmitwww.tm.tue.nl/staff/wvdaalst/pi-hype.pdf.
8. Michael zur Muehlen. Process Management Standards Overview — www.wfmc.org/standards/docs/Process_Management_Standards_files/frame.htm.
9. Manuscript WfMC. WorkFlow Reference Model — www.wfmc.org/standards/docs/ TC00-1003_10_1994.pdf.
10. Стрельникова И.А., Бачище А.В., Стрельников Д.Д. Физико-математическое моделирование КЭЭ портовых перевалочных комплексов // Вестник ГМУ им. адм. Ф.Ф. Ушакова 2(7) / ГМУ им. адм. Ф.Ф. Ушакова – Новороссийск, 2014 – С. 48-51
11. Широков А.П., Математическое моделирование транспортных процессов [учебное пособие], Хабаровск, 2012 г.
12. Van der Aalst, Kees van Hee. Workflow Management. Models, methods, and systems. The MIT Press, London.