Journal of Theoretical and Applied Information Technology
31st January 2021. Vol.99. No 2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

PARALLEL IMPLEMENTATION OF FORCE ALGORITHMS FOR GRAPH VISUALIZATION

ABENOV ZHANSULTAN [1], AUBAKIROV SANZHAR [2], PAULO TRIGO [3]

[1]Designation, undergraduate University of International Business, Department of Business informatics, Kazakhstan
[2]Designation. Ph.D at the University of International Business, Department of Business informatics, Kazakhstan
[3]Designation. Instituto Superior de Engenharia de Lisboa, Department of Agent and Systems Modeling, Portugal
E-mail: [1]zhansultan996@gmail.com, [2]c0rp.aubakirov@gmail.com, [3] ptrigo@deetc.isel.ipl.pt

ABSTRACT

In this paper we select some graph generate-and-render algorithms and evaluate their performance. The experimental results ground our proposal of a parallelized version of an algorithm (the Force Atlas 2) and the corresponding performance analysis. Our work resorts to graph visualization tools such as D3.js, jgraph, NetworkX, Gephi, and sigma.js. We analyze how to use those tools, their advantages and disadvantages and compare their performance. Since each of those tools use their own algorithms, we also conducted a comparative review of graph visualization algorithms. We reviewed algorithms such as FA (Force Atlas), FA2 (Force Atlas 2), FR (Fruchterman & Reingold), LinLog and Dijkstra (Dijkstra's algorithm). The different algorithms were experimentally compared on the same data. The experiments enabled to build a performance ranked perspective. The fastest algorithm was chosen, and experiments were carried out to optimize the rate of generation of graph coordinates by introducing parallel computations during the generation of graph coordinates. During the experiments, the best visualization algorithm for Force Atlas 2 graphs was identified and selected. This algorithm was parallelized, and we achieved a relevant speedup ratio around 26.7% and parallel efficiency around 30%.

Keywords: *Graphs, Directed, Non-directed, D3.js, FA, FA2, Jgraph, NetworkX, Gephi, Sigma.js, FR, Dijkstra, LinLog*

1. INTRODUCTION

A graph is a non-linear data structure that consists of vertices (or nodes) connected by edges (or arcs), where edges can be directed or not directed. Graphs can be used as representations of communication systems, operational semantics and model validation, as well as for visualizing transactions. Graphs are an intuitive representation system states where transitions can be conveniently described by graph transformation rules. Information visualization to represent large, abstract, semi-structured data that is very difficult to interpret in a more understandable and human readable form. For our purpose we investigate force algorithms. These algorithms are well known given that they are: a) intuitive and appropriate for presenting discretionary data as a graph b) effectively deciphered and can be modified as they are portrayed in detail via pseudo-code, c) can be tweaked to fit our purposes, and d) the visual diagrams look stylishly wonderful. Traditional force algorithms are widely accepted for data analysis in the field of graph drawing. Suitable forces are applied to every component, for example, spring power, gravity for vertices and edges to keep them at a conceivable and meaningful separation, as presented in Figure 1.

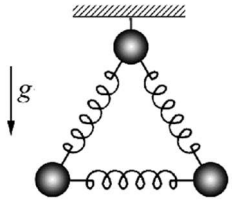


Figure 1: Representation of The Physical Graph System

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The idea was first presented by Tutte based on a barycentric image. The springs between the vertices work according to Hooke's Law [1]. Between the nodes there are repulsive forces as well as gravity forces. Gravity and repulsion forces can vary from the number of subgraphs in the graph. Graph coordinates constructed using force algorithms are visually pleasing and easy to interpret due to the harmonious balance between nodes and edges. The edges and nodes of the force-generated graph are at a comfortable distance from each other. However, force algorithms have a long execution time due to their iterative nature in O (n^2) calculations [1].

The huge flow of financial reporting and analytics complicates information processing. Fast collection from internal and external sources and fast processing of large streams of information using mobile devices and services can accelerate the entire business process from collection and storage to processing and reporting, as modern cloud services allow the use of powerful computing resources. Collecting information from different sources through cloud data services allows access to different sources regardless of the location of the data.

The financial technology industry is currently undergoing radical changes affecting the entire infrastructure of the financial sector. Almost any financial transaction can be carried out using a mobile device that provides personal financial management, biometric payments, and social benefits.

The number of transactions based on the exchange of products or on the use of alternative currencies within the framework of online platforms is actively growing; a completely new type of financial transactions between devices without human intervention is becoming widespread. The importance of cybersecurity issues, personal data protection, personal identification in the information space when making transactions is growing. The volume of venture capital investments in financial technologies has increased over the past 5 years, which confirms the growing interest in technological innovation in this area. Companies outside the traditional financial services industry are actively developing.

2. GRAPH RENDERING METHODS

This section presents the experimental methodology and associated algorithms, as well as experimental hypotheses.

This research includes the following aspects:

• comparison of available tools

• description of algorithms

• aesthetic criteria for graphic layouts.

Next, algorithms and available graph visualization tools will be described.

Force Algorithms. As a major aspect of the evaluation, four calculations were chosen: FR (Fruchtermann and Reinhold), FA Layout Algorithm (Bastian, Heymann and Jacomy, 2009), FA2 (ForceAtlas2), LinLog.

The FR is a conventional power calculation that is an adjustment of the Eades spring calculation model (Eades, 1984). This calculation makes stylishly satisfying two-dimensional pictures of charts by leading rearranged demonstrating of physical frameworks. In this method, the vertices are represented by steel rings and the ribs are treated as springs, the algorithm takes into account the force between every two vertices. The forces of attraction continue to change by moving the nodes until the energy is minimized by the completion state of the entire "mechanical system". Although the constant that determines how far one step of the node moves must be predetermined, the final layout may not reach equilibrium at all. FR is a basic, exquisite, natural, productive and having a similar edge length and sensible preparing time calculation [2]. This calculation has two standards portrayed underneath.

1. The vertices associated by an edge ought to be attracted close to one another.
2. The vertices ought not be drawn excessively near one another.

In this calculation, assume that fa and fr are the powers of attraction and repulsion, individually, d2 is the separation between two vertices and k is the span of the unfilled district around the vertex, at that point.

$$f_a(d) = \frac{d^2}{k} \quad (1)$$

$$f_r(\mathrm{d}) = \frac{-\mathrm{k}^2}{\mathrm{d}} \quad (2)$$

FA Layout Algorithm (Bastian, Heymann and Jacomy, 2009) is a spatial design strategy in the classification of intensity calculations planned for

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

giving the system a reasonable shape (spatial confinement), alongside mix between different techniques, for example, degree-subordinate Barnes-Hut demonstrating ghastly powers, nearby and worldwide versatile temperature. It is intended for genuine systems, for example, web systems, giving a general and natural approach to restrict systems. The FA format calculation gives more weight to the nature of the circuit than the speed at which it was determined [3]. This is particularly valid for enormous systems. The FA calculation offers constant settings, including speed, gravity, repugnance, auto-adjustment, dormancy, or size modification [4]. The FA power model is like FA2, as depicted underneath. FA2 (ForceAtlas2);

FA2 depends on FA, however offers more highlights and creative advancements that make it a quick format calculation.

The usage of versatile neighborhood and worldwide velocities gives great execution to a system of under 100,000 hubs. It was tentatively seen that FA2 was, best case scenario with exceptionally bunched systems.

The power of fascination fa between two associated hubs n1 and n2 straightly relies upon the separation d (n1, n2); the horrible power Fr is corresponding to the result of the point in addition to one (deg + 1) of the two hubs n1 and n2.

$$f_a(n_1, n_2) = d(n_1, n_2) \quad (3)$$

$$\text{fr}\,(\text{n1}, \text{n2}) = \text{kr}\frac{(\deg\,(\text{n1}) + 1)(\deg\,(\text{n2}) + 1)}{\text{d}\,(\text{n1}, \text{n2})} \quad (4)$$

$$\text{fr}\,(\text{n1}, \text{n2}) = \frac{\text{kr}\,(\deg(\text{n1}) + 1)(\deg(\text{n2}) + 1)}{\text{d}\,(\text{n1}, \text{n2})} \quad (5)$$

For a given graph G = (V, E), the combined force applied to the vertex v:

$$\text{F}\,(\text{v}) = \Sigma(\text{u}, \text{v}) \in \text{E}\ \text{fa}, \text{uv} + \Sigma(\text{u}, \text{v}) \in \text{V} * \text{V}\ \text{fr}, \text{uv} \quad (6)$$

LinLog. Noack proposed LinLog vitality models that can normally speak to the bunch structure of charts by gathering firmly associated hubs and isolating scanty hubs. Models incorporate the LinLog repugnance hub and the LinLog edge-aversion, whose base vitality plans mirror the group structure of the charts, following two all around characterized measures for bunching and edge shock in vitality models. LinLog maintains a strategic distance from or diminishes the predisposition towards gathering hubs with a serious extent when utilized rather than or notwithstanding repulsing the hub [5].

For a diagram G = (V, E) for a drawing p and two hubs u, v ∈ V, the length of the distinction vector p (u) −p (v) is known as the Euclidean separation u and v in p and is signified by || p (u) −p (v) || [6].

$$\text{ULinLog}\,(\text{p}) = \sum(\text{u}, \text{v}) \in \text{E}\,||\,\text{p}\,(\text{u}) - \text{p}\,(\text{v})\,|| + \sum(\text{u}, \text{v}) \in \text{V} * \text{V}\ln||\,\text{p}\,(\text{u}) - \text{p}\,(\text{v})|| \quad (7)$$

Noack characterizes the vitality model ((fascination, aversion) model) of the format as an example taken by the separation in the equations used to ascertain fascination and shock (the log is considered as 0 degree). (fascination, repugnance) - The FA model (1, - 1) possesses a moderate situation between LinLog Noack (0, - 1) and the Fruchterman and Reingold (FR) (2, - 1) calculation. ForceAtlas2 (FA2) depends on FA, however offers more highlights and imaginative enhancements that make it a quick format calculation. FA2's capacity to show groups is better than FR however more terrible than LinLog.

3. ABOUT ANTI-MONEY LAUNDERING

Money laundering (in Switzerland and Austria also: money laundering) describes the process of incorporating illicitly obtained money or illicitly obtained assets into the legitimate financial and economic cycle. Since money to be laundered comes from illegal activities such as corruption, bribery, robbery, extortion, drug trafficking, unauthorized arms trade, or tax evasion, its origin should be hidden.

Money laundering is an international cross-border phenomenon and a criminal offense both under the criminal law of Germany and other countries. The fight against money laundering is seen as an important element in the fight against organized crime, including in relation to the financing of terrorism. In modern models of

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

economic growth, money laundering is seen as one of the long-term and sustainable growth inhibitors.

Purposes and methods of money laundering

The starting point is the possession of illegally obtained money, for example, from arms trafficking, drug trafficking, smuggling, corruption, bribery, human trafficking, robbery, extortion or tax evasion.

Purpose of money laundering

Money laundering activities aim to disguise the illegal origin of money. Money should be removed from law enforcement or tax authorities by transferring the proceeds of crime through commercial transactions that are as invisible as possible within the legitimate business cycle. This is done, for example, through the purchase of real estate, company stocks, artwork or securities, or through regular consumer transactions. In practice, parties hide their transactions, often through the involvement of mailbox companies, shadow banks, companies in tax havens, or hidden trusts. It also allows for the accumulation of economic resources from mafia-style crimes. Conversely, when uncovering money laundering, investigators uncover the criminal offenses and activities on which these cover-ups are based, and thus can continue to operate.

Money Laundering Procedure

The United Nations Office on Drugs and Crime identifies three stages in the money-laundering process:

- Accommodation
- Masking (layering)
- Integration (integration)
- Accommodation

The first step in money laundering is to incorporate the amount of money derived from criminal offenses into the financial or business cycle. This is usually done in small quantities so as not to attract attention (called "smurfing").

This involves visits to casinos, horse races, expensive hotels or exchange offices, payments to bank accounts, construction and acquisition of (especially flexibly traded) assets (for example, securities, luxury goods, works of art, cars, boats, yachts, etc. Jewelry, vouchers). For example, works of art bought with dirty money are pledged to generate new net capital, while works of art are kept in a customs warehouse. Often times, invoices are issued and paid for services that have not been completed. Online sports betting is used for money laundering in many countries. Hotels and restaurants that have a lot of money are also used as cover for money laundering. Sports clubs can also be used for money laundering.

Masking (layering)

In the second stage, the origin of these assets is hidden. To do this, money travels back and forth in a large number of transactions, so that the criminal origin can no longer be traced or proven. This serves to hide the tracks; with each additional wash, the masking becomes more successful.

Means of concealment include, for example, bogus transactions and foreign payments using offshore banks, custody companies, bogus companies and shells, often in countries with little protection against money laundering or bribery of officials. This also includes counterfeits or retroactive contracts. When setting up mailboxes or shell companies and raising capital, including further cover-ups, criminals need the help of the bank and its contacts. Nowadays, setting up a company of custodians may be legal, but it only becomes a crime when, for example, drug money is laundered. Classic examples of global money laundering in connection with trusteeship are many of the Panama Papers business relationships, Odebrechts Scandal and Russian Self Laundry. There are countless opportunities for money laundering in the art market. Works of art, for example, are sold at auctions at totally inflated prices to hide the path of money and, on the other hand, to bribe the seller. This type of bribery in China is called "yahui" - "elegant corruption".

In order to conceal assets, the global advisory and hidden asset conservation industry works not only with legal tax optimization, but also profitable circumvention of regulations, as well as criminal activities such as money laundering, corruption, arms and drug trafficking, and terrorist financing or hiding it. Members of this financial consulting industry create their own legal system through tax havens and loopholes, and sometimes engage in massive lobbying efforts to open new

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

loopholes and eliminate criminal offenses and formal requirements. For decades, little has been done to combat money laundering, corruption, or this lobbying, although it has done tremendous damage to states and taxpayers and has made only a few elites rich.

Internationally, such lobbyists seek to modify government controls in their own way to circumvent them or use financial service providers or banks themselves as controls, documentation or registration. Banking authorities sometimes get the right to sign documents for companies so that investors remain anonymous and not listed in transparency registers. Following the publication of the Panama Papers, Luxembourg Leaks, Swiss Leaks and Bahamian Leaks, an international reform process began as it became publicly known how vulnerable financial service providers are to money laundering and tax evasion and any involvement of bank officials in the establishment of companies with letterboxes and representation of companies and offshore companies, the representation or anonymity of customers or the issuance of official licenses and permits is internationally illegal. Many tax and legal systems also have very elaborate rules for evading taxes or transferring money abroad.

Transactions in cryptocurrencies (eg bitcoins) are also used to launder money. Since this infrastructure is organized in a decentralized manner, government authorities cannot easily control it. Cryptocurrency cash flows cannot be blocked and difficult to assign to real people. Conscientious consumers are often used as assistants to launder money from cybercrimes. According to the German Federal Criminal Police Office, in this country alone, thousands of consumers are used by criminals to launder money every year, and the amount of money laundering is growing steadily. Money travels through many countries or regional governments with the help of rapidly formed mailbox companies or internet startups to cover their tracks.

Terrorism financing, drug and arms trafficking, white collar crime and illegal tax evasion are crimes that are difficult to control due to the high profits, camouflage and lobbying of the offshore industry. Obfuscation and lobbying systems specifically relate to claims of “due diligence” or the duty to exercise caution in identifying and documenting legal transactions in the financial industry and abandoning effective practices (formal requirements, court filings, etc.) as “countering business ".

Integration (integration)

Once the origin of the money can no longer be determined, the laundered money is used as a result of legitimate business activities. For example, stocks of a company, real estate or life insurance are purchased.

Anti-money laundering methods

Know your customer principle

The most important, most effective, and cheapest anti-money laundering tool is the application of good formal requirements to document legal transactions and prevent concealment (e.g., letterbox companies, tax havens, backdating contracts, blank signatures, hidden identities). In this way, legal transactions can be understood and controlled by the authorities through judicial or notarial protocols, significant registers, or the obligation to certify signatures so that the actual economic purpose of the legal transaction can be determined. These documentary formal requirements for legal transactions such as real estate, corporate capital or company shares are intended to prevent simple and opaque capital transfers. In many countries, certain legal transactions, such as changes in company structure, capital movements in companies, real estate or company registrations, must be registered in court or notarized to avoid fictitious transactions, counterfeiting or deferred contracts. Existing loopholes are immediately exploited for money laundering. In order to prevent money laundering, tax evasion and corruption in the long term, it is imperative that important legal transactions involve an independent state-controlled intermediary that monitors the interests of the state (identification, documenting legal transactions, tax fairness, transparency, etc.)) And is economically independent. Nobel Prize Winner in Economics Joseph Stiglitz and Basel-based criminal defense attorney Mark Peet note that the fight against corruption is not ongoing due to extensive lobbying work and that effective documentation requirements have been prevented as a result. However, in many countries, formal requirements have been relaxed or abolished in recent years to be able to conduct transactions in a “business-friendly” manner, according to business associations. Unlike society

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

and consumers, large companies or some politicians often have no real interest in effective anti-corruption, anti-money laundering and anti-corruption practices, or put pressure on legal security authorities. The state of Delaware in the United States represents a distinct model of money laundering paradise with its combination of minimal taxes and fees, the ability to register companies with secret owners and, most importantly, launch the Internet or mobile phones in free form without effectively identifying the parties involved.

As such, virtual currencies, online banking, online gambling, the creation or administration of online companies, online marketplaces and online auctions are increasingly important for money laundering and money laundering crimes such as tax evasion and fraud. For example, money laundering is also carried out through the anonymous buying and selling of art, real estate and company shares, including common voluntary loss-making transactions, possibly also through simulated artificial markets. This also applies to financial service providers selling risky securities and other financial products.

To combat money laundering, intensive work is underway to establish national registries in which companies must disclose their true beneficial owners. Implementation is very difficult because in cases where, for example, mailbox companies, tax havens, forms or online companies, or bogus managing directors are involved, the reporting obligation can be easily circumvented and is therefore de facto “voluntary ". Despite this registry, law enforcement often fails due to hidden custody. Many countries already have effective registries (for example, for the commercial register, land register, trade law, foundations, foundations, associations), in which the relevant natural or legal person or intended person can be clearly traced due to the mandatory identification of the intervening judicial authorities. the process itself is verified by the court, which prevents abuse. Money laundering can only be prevented if these authorities are legally and economically independent from the person guilty of money laundering and this independence is properly protected.

Internationally, these judicial, official or notary bodies or compliance departments in companies are often poorly paid, bound by regulations, not (economically) independent, or deliberately poorly organized to be unable to monitor compliance. Whether compliance departments embedded in a corporate structure can have a fundamentally positive effect on preventing illegal but profitable actions within a corporate purpose is controversial, because companies generally maximize their interests. The use of email, information from any financial agent, information from websites or even government registries based on unverified information, electronic signatures or online records is gross negligence. In particular, the internationally organized money laundering industry operates on the basis of the division of labor.

In addition, anonymous economic transactions and payments must be prevented. For this, the principle "know your customer" (KYC) is used. Banks, insurance companies, lawyers, auditors, online casinos, casinos, jewelers, etc. Are obliged to identify their customers before starting a business relationship (for the procedure see: verification of legitimacy) and to inquire about the beneficial owners. This method is believed to be easily circumvented, especially by mailbox companies and undercover guardianships. Internationally, there is no single mandatory standard for know your customer (KYC).

Basically, the problem is that companies like banks, insurance companies, lawyers, jewelers, real estate agents, art dealers, merchants, or even casinos naturally maximize their economic goals, and their self-interest is not in the fight money laundering or transparency or legal certainty. Accordingly, German banks also face significant money laundering charges despite extensive compliance activities. It is therefore no surprise that the global and secretive asset consulting and advisory industry has sprung up with the goal of hiding assets. These consulting firms serve not only to optimize taxes, but primarily to circumvent relaxed formalities and a large number of criminal activities such as money laundering and corruption.

The due diligence of banks in creating customer identities was described in 2001 by the Basel Committee on Banking Supervision. In this way, clients are given the entire leading name to be able to observe the opening of different accounts by the same person. In addition to establishing identity, the bank must also find out the reason for starting a business relationship and verify its accuracy. However, in international practice, offshore consultants or representatives of the financial industry engaged in offshore consulting deliberately

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ignore identification procedures and related rules. For many financial service providers, money laundering, despite all the compliance mechanisms or associated statements and rules, is rather covert. In addition, the risk of fraudulent manipulation increases sharply as banks and the financial sector outsource more and more IT areas to external service providers and outsource these areas themselves. This also applies to managing external data. It is also associated with businesses with fake, falsified, or falsified bank accounts. Online accounts, called "banking transactions," are created using fake IDs, fake documents or fabricated data, which are then used to launder money. At the same time, of course, legal certainty and balance, especially in view of new markets and technologies, require a high-quality and reasonable legal law on the complex line between suitable framework conditions and over-regulation.

In some cases, banks also work with offshore gambling providers if money laundering is suspected. An estimated 85% of money laundering takes place through banks, although there are still many carriers of cash, especially in relation to terrorist financing. According to the US Drug Enforcement Agency DEA, the popularity of bitcoin as a money laundering agent is set to rise in the international drug trade. From an international point of view, there are anonymous payments on the Internet, the use of digital currencies such as bitcoin, hiding by fake companies or false identities created for this purpose, the use of companies registered with financial service providers, money transfers via bank transfers or banking transactions hawala or couriers, but also anonymous online auctions are a tried and tested way to bypass verification. To prevent this, important legal transactions require strict verification of the identity of the parties involved, the establishment of appropriate formal requirements, and the use of an independent state-controlled intermediary.

4. ANTI-MONEY LAUNDERING METHODS

Large financial companies carry out up to 20 million transactions per day, while the detection of suspicious transactions involves manual work, which can lead to the risk of errors and false positives. That is why it is extremely important to develop and implement high-tech AML systems that meet all modern trends and requirements of the digital economy. Modern AML tools include big data, machine learning, AI. This toolkit helps improve risk-based approaches and makes it possible to keep pace with the development of money laundering technologies. One of the solutions to these problems is the creation of an intelligent automated system for collecting, analyzing and distributing large unstructured data of financial flows in the cloud environment. Using data mining techniques, several strategies can be identified to identify nodes with an increased risk of fraud. The most popular method is the use of measures of centralization - degrees (in-degree, out-degree, all-degree), closeness and betweenness centrality - which are ways to identify the most significant social actors [8]. Degrees of Centralization In real life, we often consider people with many connections to be important. Likewise, in the social graph, actors who have a lot of contacts are considered important. Therefore, centralization is a measure of the importance of a node in a network. The degree of centrality C d for a node v i in an undirected graph is

$$C_d\,(v_i) = \Sigma^n\, d_i\ \ (1) \qquad (8)$$

where di is the degree (number of adjacent edges) of the node vi.

In-degree - This measure counts the number of inbound links, that is, how many times the customer has acted as a salesperson. In the weighted version, the values of the incoming arcs are summed. The (weighted) input power for the common node gi is represented by the equation:

$$(D_i(g_i)) = \Sigma^{\mathrm{n}}\, x_{ji} \qquad (9)$$

Out-degree - this measure counts the number of outgoing connections, that is, the number of cases when the client acts as a debtor. The weighted outgoing power of the common node g i is represented by the equation:

$$(Do(g_i)) = \Sigma\, x_{ji} \qquad (10)$$

Overall degree - this measure calculates the number of links connecting a node with other nodes, regardless of their direction. In the weighted version, the risk factor is calculated as the sum of the weights of all arcs associated with one node. This measure replaces inbound and outbound degrees in undirected networks. The (weighted) total power for a common node g i in a directed graph is represented by equation [9]:

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$\left(DA(g_i)\right) = \Sigma^{n}\left(x_{ji} + x_{ji}\right) \quad (11)$$

There are different ways of identifying communities [10]. The following is a listing of the main methods. Betweenness is an algorithm based on the coefficient of "centralization of mediation", defined as the number of shortest paths between all pairs of vertices passing through a certain edge [10]. If there are N shortest paths between the vertices, then each edge is added N 1 to the value of the coefficient. The higher the value, the more likely it is that the edge connects vertices from different communities. For example, when different groups are connected by one edge, all the shortest paths from one community to another go through that edge. This algorithm is one of the first algorithms for identifying communities in networks. This method was developed by Newman and Girvan and works according to the following scheme:

Calculation of the coefficients of “mediation centrality” on all edges of the graph.

2. Alternate removal of ribs with the highest coefficient.

3. Communities are the remaining connectivity components.

4. The procedure for removing links is completed when the modularity of the resulting partition reaches its maximum.

Data processing includes setting settings, choosing attributes, grouping nodes into Communities. This analysis makes it possible to trace the ways of withdrawing money abroad, diluting money in numerous accounts, reintroducing it into the country, crediting it to other accounts. Each of these stages is accompanied by a group of closely related bank accounts that will shape communities.

5. GRAPH VISUALIZATION TOOLS

To implement the graph visualization algorithms, the following visualization tools were chosen:

- D3.js
- NetworkX for python
- Jgraph for python
- Gephi (FA, FA2, LinLog, FR)
- Sigma.js.

D3.js is a JavaScript library for overseeing information reports. D3 helps make information dynamic utilizing HTML, SVG, and CSS. D3's accentuation on web guidelines gives all the capacities of present-day programs without binds itself to an exclusive structure, consolidating amazing perception parts and an information driven way to deal with controlling DOM.

D3 enables to tie discretionary information to the record object model (DOM), and afterward apply the information driven change to the archive. For instance, you can utilize D3 to produce a HTML table from a variety of numbers. Or then again utilize similar information to make an intuitive SVG histogram with smooth advances and communications.

D3 is certifiably not a solid structure that tries to give all capacities. Rather, D3 takes care of the center of the issue: effective information control of records. This keeps away from the restrictive introduction and gives excellent adaptability by uncovering all the conceivable outcomes of web norms, for example, HTML, SVG and CSS. With negligible overhead, D3 works exceptionally quick, supporting enormous datasets and dynamic conduct for cooperation and movement. The utilitarian style of D3 permits you to reuse code in a different assortment of authority and network created modules.

NetworkX is a python library written in a similar language. Permits you to envision the relationship of chart hubs by interfacing them with edges. Consequently tallies the directions of diagram components. The library is proposed to picture charts and system structures. The library highlights are:

- functions for working with non-oriented, oriented and weighted graphs
- embedded functions for creating graphs
- functions for detecting subgraphs
- obtaining characteristics of the coordinates of the graph
- visualization of networks in the form of 2D and 3D.

This library is the fastest among the libraries for python, since it uses a low-level data structure of the language. According to the description of the developers, using this tool you can visualize from 10 million to 100 million nodes and edges.

In the course of work, it was found that this library also requires large resources for visualizing large graphs.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Jgraph (formerly igraph). This python library is written in C. Using this library, you can visualize graphs in a python application. The tool allows you to control the colors, sizes of nodes, edges and other parameters for drawing a graph. The disadvantage of this tool is the excessive consumption of resources, and the inability to design an automated, cloud system for online graph visualization, without installing any software.

Gephi is an open source software that makes it easy to create beautiful layouts of graphs and networks. The graphs generated by Gephi can be studied, analyzed, filtered and modified.

Gephi offers many options for formatting graphs so that they most effectively show the relationship that needs to be identified. This tool is a ready-made software that needs to be installed as a regular program. The tool can be extended by installing new modules. The basic functionality of the program is quite enough for importing and visualizing data. Gephi allows you to control graph parameters such as degree, betweenness, network diameter, edge color, visualization algorithm, and other parameters in the graphical interface. Another advantage of the tool is the ability to directly connect to the database to obtain data. In addition to connecting to the database, the tool allows you to import graph description files from gexf, json, gdf, gml, xlsx, csv files. The last two formats can be imported in the form of a table with two columns, in the form of sender and receiver. The disadvantage for our work was the inability to create an online visualization tool. Gephi allows you to visualize and manage visualization processes very extensively, but it is impossible to build a web tool using it, since this tool was created as a ready-made program for visualizing graphs, and does not imply interaction with third-party software.

Sigma is a JavaScript library for drawing graphs and graphs. This facilitates the publication of networks on web pages and allows developers to integrate network research into multifunctional web applications.

The tool was designed as a mechanism that can be configured and used to develop highly interactive web applications that display graphic visualizations. Some of the features:

Using sigma, you can control the rendering of the graph and the data supplied for manual rendering, which is convenient when creating a self-written tool for visualizing graphs. Canvas or WebGL built-in renderers that help you render interactive 2D and 3D graphics in any compatible web browser without using plug-ins. And the built-in visualization tools also provide many ways to customize the rendering. Focus on interactivity: you can find out when the user clicks or reels the mouse on the node. You can find out when the user drags the chart or zooms in, and always knows the position of the chart relative to the screen, and much more. Powerful graph model: Sigma is just a rendering engine, but you can do more, for example, run your own graph algorithms. For this, the sigma graph model is customizable, and you can add your own indexes to the data. Expandable: Easily develop plugins or simple snippets to expand sigma capabilities. Some of them are already available in the main repository to read some popular graphic file formats or, for example, to run complex layout algorithms. Compatibility: Sigma works in all modern browsers that support Canvas and works faster in browsers with WebGL support.

Among the features that make the interesting Sigma.js library:

• ability to bind methods

• event management

• the ability to add plugins, use GEXF files, ForceAtlas2 algorithms

• API simple and affordable

• graphs are drawn in a readable form

• insert using frames.

As a result of using this library, it was revealed that the library is lightweight (less than 3 mb), it is easily customizable, you can add your own functions, and using it you can create an online tool for graph visualization.

Table 1: Table Type Styles

Tool name	Type		Language	Customizable
D3	Web		JavaScript	+
NetworkX	Web		Python	-
Jgraph	Web		Python	-
Gephi	Desktop program		Java	-

6. EXPERIMENTAL APPROACH

After receiving the results of comparing the algorithms and graph visualization tools, the results of which will be shown later in the next section, it was decided to generate the coordinates for the graph on the backend side, and draw the graph on the frontend side from the ready coordinates. This action

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

allows us to check whether dividing tasks into different subtasks will help to increase speed.

The Java language was chosen for the backend and the Spring MVC framework was used. When developing the backend of the application, the algorithm was parallelized using the API Java Streams and functional programming to improve performance. As an instrument on the front end side, D3.js.

7. RESULTS

During experiments with graph visualization, the following tools were used:

- D3.js (Dijkstra's algorithm)
- NetworkX for python (Betweenness Centrality, FA, FA2, LinLog algorithm)
- Gephi (FA, FA2, LinLog, FR).

During the study, the following results were obtained (Core i7, 8GB RAM, AMD Radeon HD 8750M 2GB) shown in Table 1.

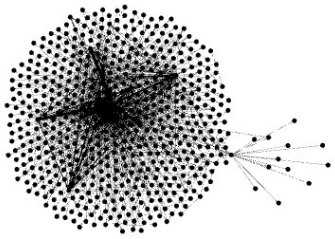


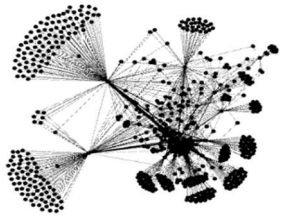
Figure 2: The Result of Visualization Using the LinLog Algorithms

Table 2: Table Type Styles

Algorithm	Tool	Number of nodes	Lead time
Force Atlas	Gephi	~ 20 000	~ 8 min
Force Atlas 2	Gephi	~ 20 000	~ 8 min
Fruchterman & Reingold	Gephi	~ 20 000	~ 9 min
LinLog	Gephi	~ 20 000	~ 9 min
Dijkstra	D3	~ 20 000	~ 9 min

Figures 2, 3, 4, 5 illustrate the results of visualization at 500 nodes of the graph.

Figure 3: Visualization Result Using Force Atlas 2 Algorithm



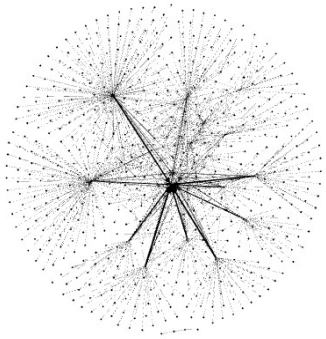


Figure 4: Visualization Result Using the Fruchterman & Reingold Algorithm

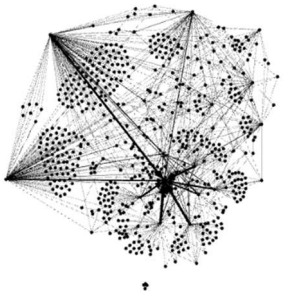


Figure 5: Visualization Result Using Force Atlas Algorithm

The Force Atlas 2 algorithm was programmed in the backend of the part for generating the coordinates of the graph. During the experiments, it was revealed that 29.75 seconds were spent on generating coordinates for a graph of 307 thousand nodes.

7. ALGORITHMIC PARALLELIZATION APPROACH

During the analysis of the resource-consuming methods of the algorithm, the Profiler tool revealed that the most resource-consuming method is the

Journal of Theoretical and Applied Information Technology
31st January 2021. Vol.99. No 2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

applyForceOnNodes function, which consumes up to 40% of all CPU resources. The pseudocode for this function is shown in Figure 6.

```
function applyForceOnNodes(List<Node> nodes, double theta)
        for Node n : nodes
                if nodes.size < 2
                        calculate node coordinate
                endIf
                else
                        calculate nodes coordiante for subregion
                endElse
endFunction
```

Figure 6: ApplyForceOnNodes Function Pseudocode

Based on this evidence, it was decided to parallelize this method's computation. Parallelization of the method was implemented by the java.util.concurrent library. Also, in all parts of the code where possible, Java Streams methods were applied. For calculations, the method was parallelized from 1 to 12 threads.

When implementing parallelization for the described pseudocode, streams and the function java.util.concurrent executorservice were applied. After implementing the algorithm, several methods were identified that consumed the most resources. The method that consumed the most resources was applyForceOnNodes. This method took 69% of the total code execution time. This is approximately 21 seconds. It was decided to parallelize these methods to optimize resources. Figures 7 and 8 show the comparative results between non-parallel and parallel implementations of the init () and buildSubRegions () methods.

The call to this method was implemented using the java.util.concurrent library. The code image shows that all the elements of the graph pass through the applyForceOnNodes method. This method calculates the distance between the nodes of the graph. Before parallelization, all calculations were performed in a single thread. After parallelization, the number of threads became configurable. The experiments were carried out on 1, 2, 4, 8 and 12 streams. On each number of threads on the same data, calculations were made 3 times to calculate the average execution time. The fastest calculation was performed on 12 threads.

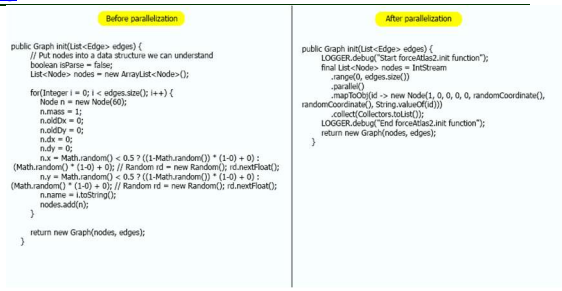


Figure 7: Comparative Result of Non-parallel and Parallel Implementation of the init () Method

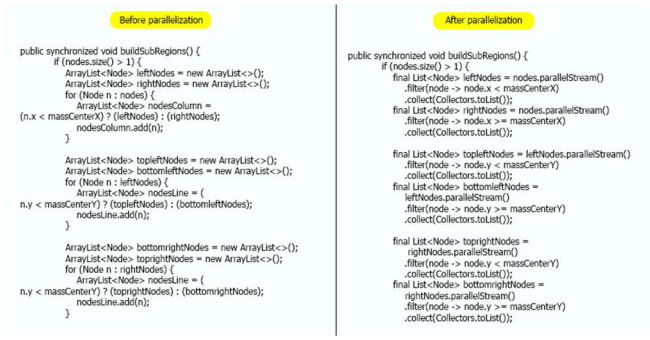


Figure 8: Comparative result of non-parallel and parallel implementation of the buildSubRegions() method

The computation was executed in a laptop with a processor of 12 cores. After using parallel computing, the query execution speed increased from 29.75 to 9.256 seconds and the following results were obtained: Speedup Ratio - 3.2 and Parallel Efficiency - 0.3. Figure 9 shows the parallelization results using the Speedup Ratio and Parallel Efficiency metrics.

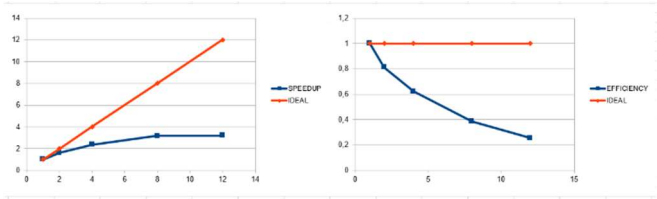


Figure 9: Parallelisation Results Using the Speedup Ratio and Parallel Efficiency Metrics

8. MONEY LAUNDERING DETECTION

For the money laundering study, data was provided by a mobile financial service. This data contained 20,451 nodes and 27,429 edges. In total, this data contained 40,908 transactions. As part of the study, the data obtained were visualized. In the process of rendering, the following result was obtained, which is shown in the figure. As part of the study, the data obtained were visualized. During the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

visualization, the following result was obtained, which is shown in Figure 10.

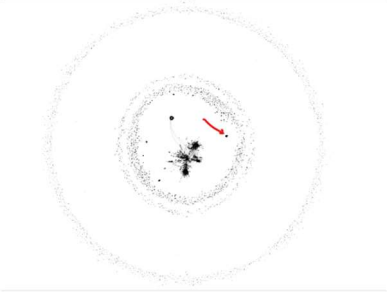


Figure 10: Visualization of Mobile Financial Service Transactions

During the analysis of this graph, it was revealed that there is a subgraph in transactions that does not contact the other subgraphs, but has connections only within the community. There is no legal entity in this column, which means that money laundering or illegal trade passes through this sub-column. Figure 11 shows a subgraph with suspicious transactions.

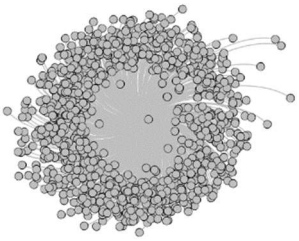


Figure 11: Subgraph With Suspicious Transactions

9. DISCUSSION

During the research, it was revealed that of all the graph visualization algorithms, the most readable is the Force Atlas 2 algorithm. The readability check was also revealed by interviewing undergraduates. The undergraduates were asked which of these images most reveals for them subgroups and connections between nodes. Therefore, this algorithm was used to generate the coordinates of the graph. Compared to other algorithms, this algorithm is pleasantly perceived due to a clear choice of subgroups and a more understandable determination of the coordinates of the node and edges compared to other algorithms.

The D3.js tool was chosen to draw the graph, because this tool is flexible and customizable.

When generating coordinates on the server side, a resource-intensive method was parallelized, after which the coordinate generation speed increased around 31.1%.

To proceed with our work, the next step is to use WebGL to efficiently draw a graph on a web interface.

10. ACKNOWLEDGMENT

I would also like to thank the University of International Business for the opportunity to conduct research within the framework of the project "Cloud applications in the management of information flows in financial markets" on the topic of project No. AP05136146.

During the analysis of the resource-consuming methods of the algorithm, the Profiler tool revealed that the most resource-consuming method is the applyForceOnNodes function, which consumes up to 40% of all CPU resources. The pseudocode for this function is shown in Figure 6.

REFERENCES:

[1]S. G. Kobourov, "Force-directed Drawing Algorithms," Handb. Graph Draw. Vis. (Discrete Math. Its Appl., pp. 383–408, 2013 (references)

[2] T. M. J. Fruchterman and E. M. Reingold, "Graph Drawing by Force-Directed Placement," Software Practice and Experience., vol. 21, no. 11, pp. 1129–1164, 1991.

[3] M. Bastian, S. Heymann, and M. Jacomy, (2009) "Gephi: An Open Source Software for Exploring and Manipulating Networks," Third International AAAI Conference on Weblogs and Social Media. doi: 10.1136/qshc.2004.010033.

[4] M. Jacomy, S. Heymann, T. Venturini, and M. Bastian, "Force Atlas 2, A Graph Layout Algorithm for Handy Network Visualization,", pp. 1–21, 2011.

[5] A. Noack, "Energy Models for Graph Clustering," J. Graph Algorithms Appl. JGAA, vol. 11, no. 112, pp. 453–480, 2007.

[6] A. Noack, "An Energy Model for Visual Graph Clustering," Proc. 11th Int. Symp. Graph Draw. (GD 2003), LNCS 2912, pp. 425–436, 2003.

[7] Official site of the international organization FATF.- http://www.fatf-gafi.org

[8] Iskakov U.M., Bokhaev D.T., Ruzieva E.A. Financial Markets and Intermediaries: A Textbook (Revised Edition). - Almaty: Economy, p. 297, 2012.

[9] Iskakova Z.D. Theoretical foundations of finance, credit and the role of the financial system in the development of the Strategy of Kazakhstan: scientific publication - Almaty, p. 254, 2014.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[10] Osbom R., Baughn C. Forms of Interorganizational Governance for Multinational Alliances // Academy of Management Journal.- № 33 (3).- pp. 503-519, 2010.

[11] Daribaeva M.Zh. Development of the financial market of the Republic of Kazakhstan in the context of globalization // Vestnik KNU-2013-№ 3.- pp. 132-137, 2013.