

Visualization Analysis for Business Performance of Chinese Listed Companies Based on Gephi

Guang Sun¹, Hongzhang Lv^{1,*}, Dianyu Wang², Xiaoping Fan^{1,3}, Yi Zuo¹,
Yanfei Xiao⁴, Xu Liu¹, Wenqian Xiang¹ and Ziyi Guo¹

Abstract: When conducting company performance evaluations, the traditional method cannot reflect the distribution characteristics of the company's operating conditions in the entire securities market. Gephi is an efficient tool for data analysis and visualization in the era of big data. It can convert the evaluation results of all listed companies into nodes and edges, and directly display them in the form of graphs, thus making up for the defects of traditional methods. This paper will take all the listed companies in the Shanghai and Shenzhen Stock Exchange as the analysis object. First uses tushare and web crawlers to collect the financial statement data of these companies. And then, uses the Economic Value Added model to calculate the EVA of each listed company and build graph data. Next, import the graph data into gephi to generate the distribution graph of all listed companies' performance, and summarize the distribution characteristics of business performance. Finally, select a listed company that you want to analyze in detail, using the traditional DuPont analysis method to conduct micro level visualization analysis of the business performance to find the main factors affecting the company's operating performance. Incorporating gephi into traditional performance analysis methods will make the results of traditional analytical methods more effective and complete.

Keywords: EVA, gephi, visualization, graph analysis, business performance evaluation.

1 Introduction

From the beginning of human society to the web 2.0, more people began to post their own information on the internet. By May 2011, the famous McKinsey Global Institute proposed that human beings enter the era of big data [Manyika, Chui, Brown et al. (2011); Shrote and Deorankar (2016)], until now the mobile internet has developed to 5G [Liu, Shi, Wang et al. (2017)]. Development of computer technology has been promoting the data on the

¹ Institute of Big Data, Hunan University of Finance and Economics, Changsha, 410205, China.

² Housheng School of International Education, Hunan University of Finance and Economics, Changsha, 410205, China.

³ School of Information Science and Engineering, Central South University, Changsha, 410075, China.

⁴ School of Finance and Economics, Hunan University of Finance and Economics, Changsha, 410205, China.

* Corresponding Author: Hongzhang Lv. Email: godspeedlv@outlook.com.

Received: 16 September 2019; Accepted: 18 November 2019.

internet. In 2018, millions of data records are generated every minute on the Internet [Ahmad (2018)]. Certainly, the sheer volume of data is a feature of big data, and another feature is the complex connection between different datasets, which scholars have summed up as 5 'V's, namely 'Volume', 'Variety', 'Velocity', 'Veracity', and 'Value' [Huang and Handfield (2015)]. The concept of big data has gone deep into all walks of life [Guo, Liu, Zhang et al. (2019)]. Meanwhile, many kinds of big data process technologies are also generated or used in various industries, such as financial industry [Kshetri (2016)]. Therefore, all walks of life inevitably encounter two major obstacles in data analysis, one is how to deal with the large amount of data, and the other is how to show the distribution of data sets with complex relationships. In addition, as the number of listed companies increases, financial industry data will accumulate at a faster rate [Cockcroft and Russell (2018)]. Because traditional analytical tools and methods cannot cope with the two basic problems of big data, people need to solve the problems of huge data volume and complex data connection based on traditional models and tools. For the former, there is Hadoop big data analysis tool [White (2015)], and for the latter there is graph analysis and visualization tool gephi [Alhajj and Rokne (2014)]. This paper focuses on the latter, and it focuses on solving some of the problems in the financial field regarding the visualization analysis of listed companies' operating performance. We can use gephi to display the relationships contained in the data. Gephi is opening source software which can be used in diagram analysis and visualization. And gephi is also a big data visualization tool, which can handle a very large number of data nodes at a time [Yang, Chang, Shi et al. (2017)]. This tool can solve the problem of how to show the connection between datasets to a certain extent.

It is not enough to have gephi in this work. Some traditional performance evaluation models and tools are also necessary. These models and tools include the Balanced Score Card (BSC) model [Khomba (2015)], DuPont analysis [Soliman (2008)] and the Economic Value Added (EVA) model [Unlu, Saygin and Ersoy (2015)], among others. The research on business benefit evaluation can be divided into three stages: from the beginning of 19th century to the beginning of 20th century, the evaluation index of business benefit is the cost. In 1911, American accountant Harry designed the earliest standard cost system [Chatfield and Vangermeersch (1996)]. Therefore, this stage is the evaluation stage of cost index. From the beginning of 20th century to the 1980s, the cost index cannot meet the needs of performance evaluation, people also include three aspects of enterprise profitability and solvency and operational capability in the evaluation. During this period, the famous evaluation methods include the Wole score method [Feng (2012)] and DuPont analysis, and the valuation method based on economic profit, etc., and the US's famous Economic Value Added model is also established at this time. This stage is also called the financial indicator evaluation stage. From the 1990s to the present, with the development of management science, the evaluation index system has entered into the multi-index comprehensive evaluation stage. This stage of enterprise development evaluation is based on financial evaluation, taking into the requirements of enterprise development, management and other aspects of the comprehensive evaluation. During this period, the most representative is the United States Balanced Scorecard model (BSC model). In addition, the RFM model for customer management is also used in the business evaluation of listed companies [Stern, Stewart and Chew (1996); Sarvari,

Ustundag and Takci (2016)].

The reason why these tools or models are traditional is that they have already created or been used in business performances evaluation. They can play a role in the precise analysis of a small number of companies in a part of the industry at the micro level. However, it is not possible to reflect the operating performance level of these few companies from the macro level of the entire securities market. Taking the analysis of the operating performance of Chinese listed companies in ‘Shanghai Stock Exchange and Shenzhen Stock exchange’ as an example, if you want to use traditional analytical tools and models to analyze all the listed companies in these two securities markets, you need sample and analyze each company first, which likes a census. It is almost impossible to do this quickly and accurately using traditional analytical methods and tools for analysis. Therefore, it is necessary to use gephi to make up for the shortcomings of traditional corporate performance evaluation methods. By adding gephi to the analysis process and combining with the traditional analysis tools or models, the whole analysis works could be finished at least three steps. These steps are building graph data, macro-level visualization analysis and micro-level visualization analysis. In the first step of building the graph data, we could collect all the financial statements data of those listed companies to calculate EVA value of each listed company by the Economic Value Added (EVA) model, and then transform the data into the graph data. In the second step of the macro-level visualization analysis, the graph data will be imported into the gephi to generate the EVA values distribution graph of all listed companies. In the micro-level visualization of the third step, select a company that wants to analyze and use DuPont Analytics method to analyze the performance of the company’s four indicators of profitability and finding out the main factors that affect the company’s operating performance.

2 Tools and methods used for visualization analysis

The main tools used in this paper to visualize the operating performance of listed companies are Python and gephi, the main methods and models are Economic Value Added model and DuPont analysis. Other auxiliary tool is financial software, called Choice. This section describes these tools and how to use them. There are something need to noted that visualization analysis is not a separate work, it is the final work in data analysis. Before visualization accomplished, four tasks need to be finished, those are data collecting, data cleaning, data storing and data analysis. These five steps are basic steps of data analyzing work. As mentioned above, the visualization analysis of the operating performance for listed companies in this paper needs graph data construction, macro-level visualization and micro-level visualization of three steps. In fact, the first four steps of data analysis work are the preparation of the graph data in this paper, and the fifth step is the macro-level visualization and micro-level visualization of this paper. Relations of these steps could be illustrated by followed Fig. 1. Describing the relations between these steps makes it clearer that how these tools, methods, and usage processes are described.

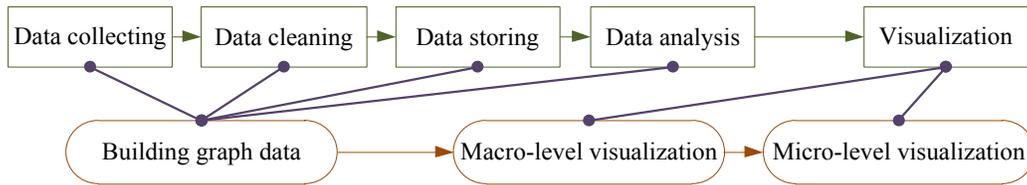


Figure 1: Relations between basic data analyzing works and steps of analyzing for business performance in this paper

2.1 Building graph data

Building graph data is the first thing we need to do. It is a time-consuming and laborious thing, but a most basic and crucial step. It includes the first four steps of data analysis work. The tools required for this step are python and the Economic Value Added model. Python is a powerful computer programming language which can be used to write programs to perform most of the data analysis and visualization operations. In this paper, python will be used in collecting basic financial data of listed companies, calculating financial indicators and converting graph data. The Economic Value Added model is used to calculate the EVA metrics for each company. The core idea of this index is that the profit obtained from the use of capital only by deducting the risk borne by the investor is the value of creation in the true sense. These risks could be regard as the cost of capital. It is a value evaluation indicator developed by ‘Stern Stewart Management Consulting Co., Ltd.’. And promoted in the mid-to-late 1990s [Bahri, St-Pierre and Sakka (2011)]. In general, the EVA indicator needs to be adjusted more than 160 financial indicators [Bauer, Braun, Elezi et al. (2014)], but the basic economic value added can be calculated using unadjusted indicators. Although this can lead to errors, it has little effect on the macro-level visualization of the second step. When the EVA values are calculated, they have to be standardized and the error can be further reduced. The size of the visual operation value is reflected in the color gradient, so the small part of the error on the overall visualization effect is not very large. One thing that needs to be explained is the basic financial data. These basic data include basic financial information about all listed the companies. Those include the Shanghai Composite Index for the past decades, the balance sheet, the cash flow statement and the income statement for nearly a decade. This step can be achieved through the tushare package in python and the web crawler programs.

Once you have access to these basic data, you can calculate each of the financial indicators utilized in this paper. Those include β coefficient, the Weighted Average Capital Cost (WACC), the basic Economic Value Added (EVA), and so on. The calculation of these indicators requires some calculation formal. Before introducing these formulas, it is necessary to define and explain some notations. It should be explained that the cost of debt capital K_d and risk-free yield R_f are taken from Choice financial software, $K_d = 6.825\%$, $R_f = 3.19\%$. The details are shown in the following Tab. 1:

Table 1: Notations and the meaning of notations

Notations	Meaning of Notations
β	Beta coefficient of stock
r	Stock daily rate of return
p, p_index	Closing price of stock, closing index
R_m	Stock market annual rate of return
rm	Stock market daily yield
R_f	Risk-free rate of return
K_e, K_d	Cost of equity capital, debt capital cost
WACC	Weighted average cost of capital
NOPAT	Net operating profit after tax
totalAssets	Reported total assets
EVA	Basic Economic Value Added
std_EVA	Z-score standardized EVA
D, E	Total liabilities, Total owner’s equity
E(X)	Mathematical expectations of X
Var(X)	The variance of X

The first thing that needs to be calculated is the daily yield r of each company’s stock.

$$r = \frac{p_1 - p_0}{p_0} \tag{1}$$

Among Eq. (1), p_1 and p_0 respectively refer to the closing price and the closing price of the previous day. They can be obtained by the tushare package. Then, we have to calculate the market’s daily rate of return rm.

$$rm = \frac{p_index_1 - p_index_0}{p_index_0} \tag{2}$$

Among Eq. (2), p_index_1 and p_index_0 , respectively refer to the closing index and the previous day’s closing index, which comes from the Shanghai Composite Index for the past ten years and can be obtained by tushare.

Based on the results calculated by Eqs. (1) and (2), the value of β can be calculated by using linear regression method [Lin and Chen (1990)].

$$\beta = \frac{n \sum_{i=1}^n r_i rm_i - \sum_{i=1}^n r_i \times \sum_{i=1}^n rm_i}{n \sum_{i=1}^n rm_i^2 - (\sum_{i=1}^n rm_i)^2} \tag{3}$$

The next step that needs to be calculated is the stock market annual rate of return R_m . One thing to note is that p_index_b and p_index_a respectively refer to the price index at the end of the previous year and the price closing index at the end of the year. The index is derived from the Shanghai index for nearly a decade and can be obtained by tushare.

$$R_m = \frac{p_index_a - p_index_b}{p_index_b} \tag{4}$$

According to the Eqs. (3) and (4), the Capital Asset Pricing Model (CAPM) can be used

to derive the equity capital cost of a company K_e [French (2017)].

$$K_e = Rf + \beta \times (E(Rm) - Rf) \quad (5)$$

After we get K_e of each listed company, the Weighted Average Capital Cost (WACC) of each company can be calculated [Fernandes (2014)].

$$WACC = \frac{D \times K_d}{D+E} + \frac{E \times K_e}{D+E} \quad (6)$$

K_d is from Choice financial software. According to Eq. (6), each listed company's basic EVA value can be concluded.

$$EVA = NOPAT - totalAssets \times WACC \quad (7)$$

NOPAT and totalAssets are derived from each company's profit statement and balance sheet, respectively, and these financial statements can be obtained by tushare. Finally, according to Eq. (7), the basic economic added value of Z_standardization can be calculated std_EVA [Mendenhall and Sincich (2006)].

$$std_EVA = \frac{EVA - E(EVA)}{\sqrt{Var(EVA)}} \quad (8)$$

The reason for standardizing the basic economic value added is to narrow the gap of each original EVA value and to facilitate the subsequent visualization analysis.

After the financial indicator is calculated, we will also need to convert these calculations to graph data that gephi can process. gephi supports importing data from a spreadsheet in a csv file format. But the first line of the spreadsheet must contain some fixed column names. Only the data that satisfies these column names, can they be called graph data and be recognized gephi. There are two types of graph data. One is used in describing the information for the node and the other for describing the edges. The data that describes the node information is called node data. Another data that describes the edge information is called edge data. The node data table must contain 'Id' and 'Label', which are used to represent the number and name of the node, respectively. The edge data table must contain 'Source', 'Target', 'Type', 'Id', 'Label' and 'Weight'. They respectively represent each edge's source node, target node, type, number, name and weight [Cherven (2013)]. In these fixed columns, some columns can take NULL value, but these column names must exist. After converting these data into graph data, the first step of visual analysis of the operating performance of listed companies is completed.

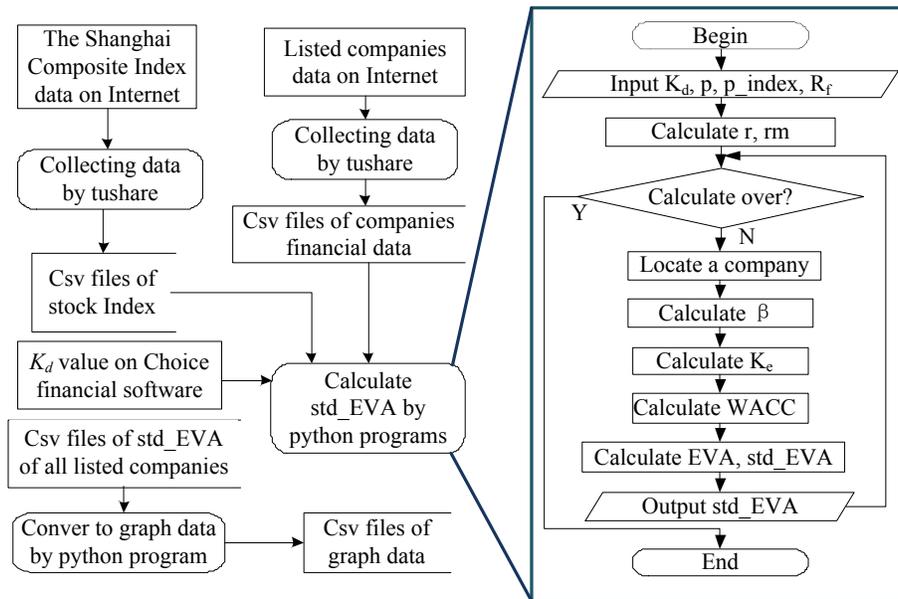


Figure 2: Process of building graph data

2.2 Visualization at the macro-level and visualization at the micro-level

After the first step of the diagram data construction is complete, the visual analysis work can be started. This section sets out the work of the second step and the work of the third step. The reason for this is that both steps are for visual analysis. In the second step of the macro-level visualization analysis, the main tool used is gephi. This is a visualization tool that can handle thousands of node data. When doing big data analysis, gephi is often used to visualize the results. Visualization using gephi generally follow the six steps. Those are importing diagram data, overall formatting adjustment, partial format adjustment, visual preview, generating final images and macroscopic visualization analysis. When importing diagram data, the users only need to be aware of whether the imported table matches the file format described earlier. If the column names in the table do not meet the requirements of the diagram data, the table cannot be imported into gephi. In the overall formatting step, the user needs to first determine an appropriate overall layout, and then set the color or size of the nodes and edges. gephi offers 12 layout algorithms, commonly used by ‘Force Atlas’, ‘Force Atlas 2’, ‘Fruchterman Reingold’, ‘OpenOrd’, ‘Yifan Hu’. These common layout algorithms are based on the theory of particle physics. When gephi runs these algorithms, these nodes will be simulated into atoms. And the position of each node is determined by the gravitational and repulsive force between the atoms [Zhuhadar and Yang (2012)]. Different layouts produce different visualizations, and we can try to use multiple layouts to sift through the visualization under each layout to select the best layout. There are also some auxiliary layouts that are primarily used to adjust visualization effects, such as ‘Noverlap’ which is to make no overlap between nodes. These layouts are generally used to determine the common layout, and then use the auxiliary layout to improve. Once the overall layout is determined, the user can adjust the color or size of the nodes and edges. In terms of color filling, gephi provides two color

filling methods 'Partition' and 'Ranking'. In terms of size adjustment, gephi only provides a 'Ranking' method [Liu and Du (2017)]. For these colors filling methods, if the user selects the 'Partition' method, gephi will divide all nodes into different intervals by the size of a property value. The nodes or labels with different intervals are applied to different colors. In addition, the color users of each interval can set themselves. If the user selects the 'Ranking' method, gephi will match the range of values to the gradient range of a color, thus filling different colors for different nodes or labels. Of course, users can also set the range of colors according to their own needs. During size adjustment, we can use 'Ranking' method to finish it. 'Ranking' includes color adjustment which could match the range of a node's property value to the range of dimensions, so that nodes with different property values render different sizes. The maximum and minimum values for the display dimensions can also be set by the user. After the overall layout, the color or size of the node or edge is set. The user can select some nodes for local adjustment. This is primarily for some nodes with special meanings. For example, some nodes only serve as connection. We can right click to select it to modify the parameters of some nodes, such as color, size, and so on.

After completing the local formatting, the effect of macro-level visualization is roughly out. But the image also needs to be adjusted in the preview box, because the result of the final output is the result of the preview box display. In the preview interface, users can also modify the gephi processed images, including partial adjustments to nodes and edges, such as the curvature and straightness of edges, the font and size of labels, and so on. It is important to note that you need to click the Refresh button to see the effect for each time's change of the graph. When the user is satisfied with the display effect of the image, the image can be output, and the output file format has three kinds, they are 'Portable Document format (PDF)', 'Portable Network Graphics format (PNG)' and 'Scalable Vector graphics format (SVG)'. If the size of the diagram is small, you can select PNG format. If the size of the graph is large, it is better to select the output in PDF or SVG format, because the images in these two formats are not easily distorted when zoomed in. With the visualization results, people can analyze the visualization results of gephi. Generally speaking, the analysis of the image is in accordance with the first overall after the local, from large to small, from many to less order. Following such a rule, macroscopic visualization analysis has a chapter to follow. For example, this paper analyzes the operating performance of listed companies in China. The result of visualization is the distribution of business performance. Firstly, from the overall observation, it is possible to know how the operating performance of all companies in the entire securities market is, which industry has more companies in the entire securities market, and which has fewer, which part of the business is good and which part is not good. And then, we can enlarge the picture and narrow the scope of the analysis to an industry. We could know of how well-run and poorly-run in this industry is distributed or which part is more and which part is less.

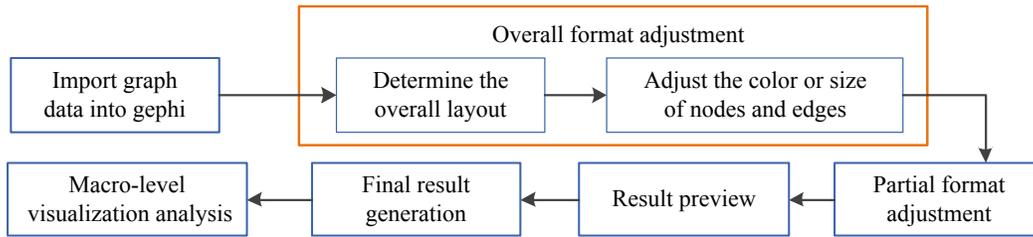


Figure 3: Using gephi to process graph data

It should be explained that the selection of the industry or company is not sole. You can choose the one whose standardize EVA value performs extreme. You can also select industries of interest to users for analysis. When analyzing to individual enterprises, the user’s analysis work has begun to move away from the macroscopic concept, into the micro-level analysis.

The next paragraph describes the micro-level visualization tools, methods, and operational steps. And the micro-level visualization is roughly the same as the statistical analysis of the general meaning. We can use bar charts, pie charts, line charts, and so on. The methods that can be taken are the wall scoring method, DuPont analysis and Factor analysis approach [Wang (1996)]. In the micro-level analysis of this paper, the tool used is the line chart, and the method used is DuPont analysis method and factor analysis method. Users can use specialized data analysis software such as Excel, SPSS to draw line diagrams, or they can do so by writing programs. In this paper, we still the python programming language. Among them, the line chart depicts the changes of indicators which are get from the DuPont analysis method. They are ‘Return on Equity (ROE)’, ‘Profit Margin’, ‘Asset Turnover’ and ‘the Equity Multiplier’. The DuPont Analytics is derived from DuPont Corporation, which was invented in 1912 by a salesperson, Donaldson Brown, and began to be used by the company in the 1920s. This method is often used to evaluate the financial situation and operating results of enterprises in an integrated system. One feature of this approach is it starts from the core index, ROE, and is broken down from top to bottom [Bodie, Kane and Marcus (2004)]. The more layers of decomposition, the more detailed the level of analysis of the company. The reason why the equity net interest rate is the core indicator is that the indicator reflects the net profit earned by each shareholder’s equity and can measure the overall profitability of the enterprise. The following describes the index decomposition process of DuPont analysis. The first is ROE [Hargrave (2019)] :

$$ROE = \frac{Net\ Income}{Total\ Equity} \tag{9}$$

Based on formula 9, we can multiply by Total Assets on the numerator and denominator to get 2 sub-indicators, ROA and ‘the Equity Multiplier’:

$$\begin{aligned}
 ROE &= \frac{Net\ Income \times Total\ Assets}{Total\ Assets \times Total\ Equity} \\
 &= \frac{Net\ Income}{Total\ Assets} \times \frac{Total\ Assets}{Total\ Equity}
 \end{aligned}$$

$$= ROA \times \text{Equity multiplier} \quad (10)$$

The ROA in Eq. (10) is an abbreviation for return on assets. The ROA indicator reflects the net profit created by each dollar's total assets [Crosson, Needles and Powers (2008)]. In addition, it is also the basic way to increase the equity interest rate (ROE). Multiplying the numerator and denominator of the ROA by the sales revenue and decomposing again can give the following formula [Beattie, Edwards and Goodacre (1998); Kumar and Lee (2001)]:

$$ROA = \frac{\text{Net Income} \times \text{Revenue}}{\text{Revenue} \times \text{Total Assets}} = \text{Profit margin} \times \text{Asset turnover} \quad (11)$$

After the decomposition of the above formula 11, the return on equity (ROE) can be expressed by the following expression:

$$ROE = \text{Profit margin} \times \text{Asset turnover} \times \text{Equity multiplier} \quad (12)$$

The decomposition of the indicator can continue, in the microscopic visualization work of this paper, we will decompose the number of hierarchies to 3 layers. Therefore, we can get the basic framework of DuPont Analysis:

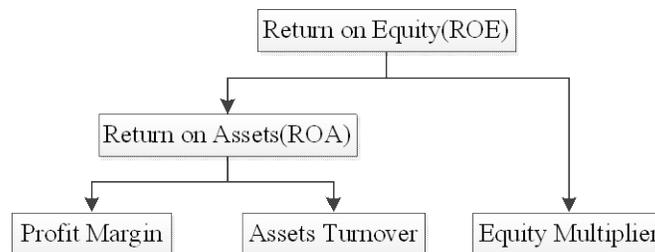


Figure 4: The basic framework of DuPont analysis in this paper

Through the decomposition of indicators, we can find three factors that affect the company's operating conditions at the micro level. These three indicators, together with the core indicator ROE, reflect the company's recent status in the profitability of owners' equity, profitability of sales, asset management capacity and financial leverage. Therefore, we can use the historical data of these four indicators to compare with the current data. And we can analyze which of the three indicators in the third layer has the greatest impact on the core indicator ROE. This requires the use of factor analysis method. Factor analysis method is to determine the influence degree of each factor on the index quantitatively. With the help of factor analysis, an index can be decomposed into several indexes, and the result of multiplying each index after decomposition is still equal to the value of the decomposed index. We can calculate the degree of influence of each factor by substituting various factors. To illustrate the factor analysis' process, we can assume that S is the object to be analyzed, and the amount of its factors is n . $F[i]$ is one of the factor in S , and $i = 1, 2, \dots, n$. Accordingly, we can get the equation as follow:

$$S = \prod_{i=1}^n F[i] \quad (13)$$

We could assume the values of each current indicator are $S_t, F_t[i]$. And the historical data of the previous period are $S_0, F_0[i]$, ($i = 1, 2, \dots, n$.) Respectively. The following whole process of factor analysis could be showed:

First, the S_0 is calculated using Eq. (13) and the historical data of the previous period $F_0[i]$:

$$S_0 = \prod_{i=1}^n F_0[i] \tag{14}$$

Then, starting from the first factor $F_0[1]$, replace all the elements of S_0 with the current value $F_1[1]$, $F_1[2]$, ... , $F_1[n]$ individually, and calculate the temp value H_i after each replacement, where $i=1, 2, \dots, n$. When all the elements are replaced, you can get a set of data:

$$\begin{cases} H_1 = F_1[1] \times F_0[2] \times F_0[3] \times \dots \times F_0[n] \\ H_2 = F_1[1] \times F_1[2] \times F_0[3] \times \dots \times F_0[n] \\ H_3 = F_1[1] \times F_1[2] \times F_1[3] \times \dots \times F_0[n] \\ \dots \\ H_n = F_1[1] \times F_1[2] \times F_1[3] \times \dots \times F_1[n] \end{cases} \tag{15}$$

We can define the degree of influence as Δ , which represents the change of each factor to S_0 . Then the degree of influence from the first factor $F[1]$ Δ_1 to the last factor $F[n]$ Δ_n can be expressed as:

$$\begin{cases} \Delta_1 = H_1 - S_0 \\ \Delta_2 = H_2 - H_1 \\ \Delta_3 = H_3 - H_2 \\ \dots \\ \Delta_n = H_n - H_{n-1} \end{cases} \tag{16}$$

We only need to find out the main factors that affect the S indicator in these factors, depending on the size of the $\Delta_1, \Delta_2, \dots, \Delta_n$. In the introduction of DuPont analysis, this paper has mentioned that ROE indicators can be broken down into three financial indicators multiplied. Therefore, we can use factor analysis to analyze ROE indicators, and find out the main factors affecting ROE indicators [Chen (1987)].

3 The process of visualization analysis of business performance

The previous chapter already describes the methods and tools used for visualization analysis of business performance. This section will introduce the entire process of visual analysis of the operating performance of listed companies in China. As mentioned above, the whole process is divided into three steps. The first is to build graph data, and in this step, we need to collect basic data from all listed companies and use the EVA model to build graph data. Then there is the macro-level visualization, the main task of this step is to use gephi to produce the EVA map of listed companies. Finally, micro-level visualization, the main task of this step is to select a company, using DuPont analysis and factor analysis, to analyze the changes in the company's operating capacity, and to identify the main factors affecting the company's operating performance. It is worth mentioning that the criteria for choosing a company are not unique. The selected company can be a higher EVA enterprise, or it can be a lower EVA enterprise. The entire process is shown in the following Fig. 5.

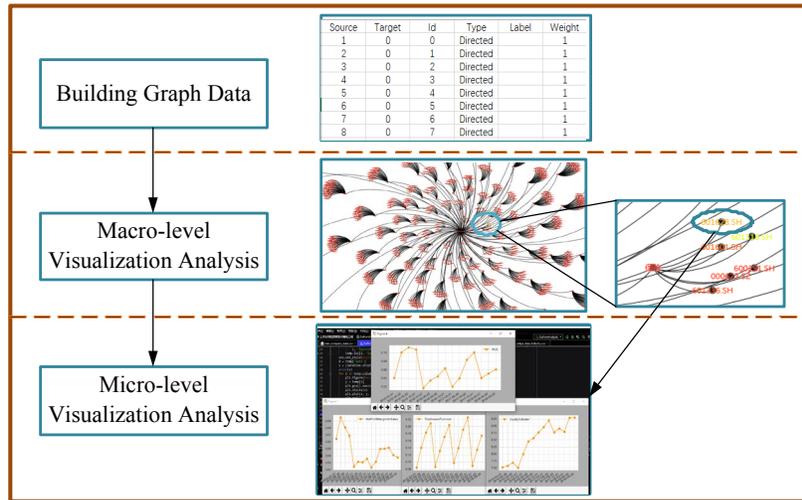


Figure 5: Three steps of business performance analysis based on gephi

Firstly, the whole process of graph data construction is introduced, and it should be reiterated that all listed companies in ‘Shanghai and Shenzhen Stock Exchange’ are representatives of listed companies in China. Therefore, we first need to use python package tushare to collect all Shanghai and Shenzhen listed companies including stock list, financial statements and Shanghai index, including the basic data. Note that the order of collection must be to collect the stock list first, then collect the financial statements, and finally collect the Shanghai index. The reason for this is that the acquisition of financial statements depends on the stock list. In addition, the values of K_d and R_f can be obtained in the WACC calculator in Choice financial software. After the basic data acquisition is complete, the EVA value of each company can be calculated using the EVA model and the second chapter process. Due to these values are large, some EVA values are even as high as 16 billion yuan. So, we need to use the Z_score standardization method to standardize each company’s EVA for subsequent gephi visualization. At this point, all the company node data is ready. However, just preparing company node data is not enough. On the one hand, under the action of force-guided layout algorithm in gephi, only nodes without edges will make the nodes shrink or repel each other to an infinite distance. On the other hand, only the company’s node data does not reflect its industry-owned relationship. Therefore, in view of the first aspect of the problem we need to construct the root node. For the second aspect of the problem, we need to construct the industry node. In addition, we need to prepare node data for the industry. Due to there is only one root node, we can define it as numbered 0. Industry nodes can be filtered out of the stock list data using the python programs. With industry node data and root nodes, we can define edge data. The edge data here consists of two types, one that points from the company node to the edge data of the industry node, and the other that points to the edge data from the industry node to the root node. Based on the definitions of these two types of edges, we can also construct them by python programs. The only thing to note is that you must construct diagram data that contains the listed names inherent in the diagram data. The column name of the node data must contain ‘Id’ and ‘Label’. The side data

column names must contain ‘Source’, ‘Target’, ‘Type’, ‘Id’, ‘Label’ and ‘Weight’.

Next, what we need to do is macro-level visualization analysis. The first step is to import the graph data into the gephi. When we arrive at ‘Overview’ window in gephi, click the ‘File’ button, and import data by clicking ‘spreadsheet’ button. And then select the previously prepared graph data file. When the data import works are complete, an irregular initial diagram is displayed in the ‘Overview’ window. Next, choose a suitable overall layout. After trying out the various layout algorithms, we choose the ‘Yifan Hu’ layout. We can select ‘Label Color’ in the ‘Nodes’ tab, and select ‘Ranking’ mode, finally select std_EVA. Once this is done, you can associate the size of the std_EVA with the gradient range of a color. Alternatively, you can choose ‘Label Size’ to associate the size of the ‘In Degree’ with the size of the node label in the same way. It is worth mentioning that in adjusting the color, we selected a different spline to dilute the std_EVA performance of the node. After the above steps, the approximate shape of the macro-level visualization results comes out. Switch to the ‘Preview’ window, and then do the final trimming of the diagram. The distribution of listed companies in various industries can be obtained, as shown in the following Fig. 6.

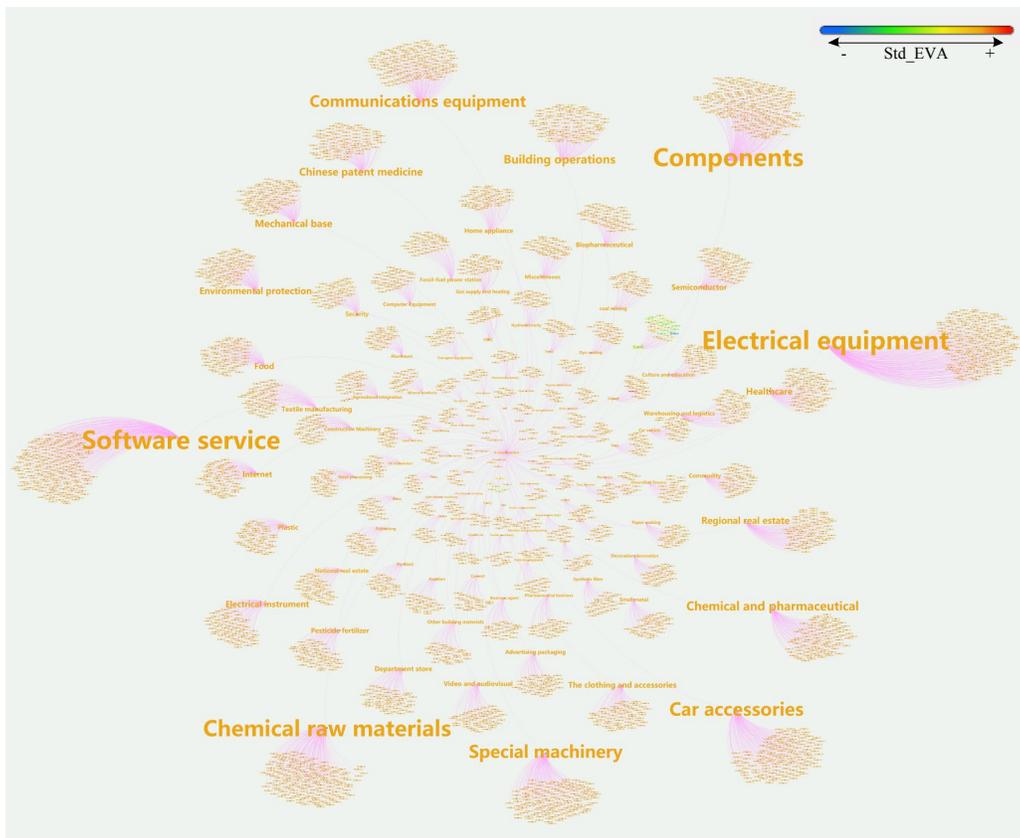


Figure 6: Distribution status of listed companies in various industries

Let's analyze the Fig. 6. Since we relate 'In Degree' to the size of the label when the overall format is adjusted, the larger the label, the more companies the industry belongs to. The root node 'SHSZStockMarket' is excluded here because the root node only acts as a connection to each industry node. Therefore, the first conclusion we can draw is that the four industries that include the largest number of listed companies are 'Software Service' industry, 'Component' industry, 'Chemical Raw Materials' industry and 'Electrical Equipment' industry. And those followed by 'Communication Equipment' industry, 'Car Accessories' industry and 'Special Machinery' industry. There is a color gradient bar at the top right of the picture. The more the color of the label tends to be blue, the smaller the std_EVA value is. As std_EVA value increases, the label's color will tend to be red. It should be noted that the std_EVA value of the industry node is the average of the std_EAV value of all companies in the industry. Therefore, the label of the industry node also comes with color. On the whole, most of the std_EVA of these listed companies are concentrated in the orange red area, and their business performance is relatively good. But we can enlarge the whole picture, through the color to find out which industries have good operating performance, and which industries have poor business. We intercepted some parts of Fig. 6 for display.

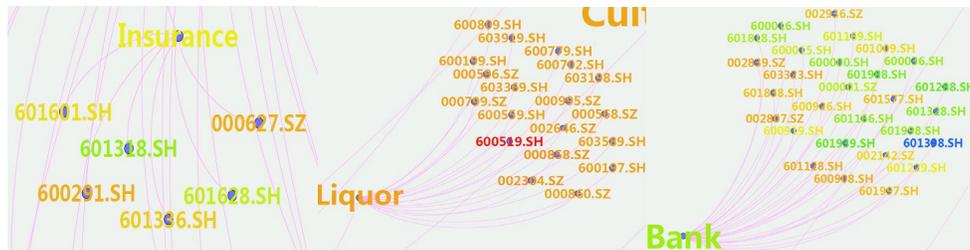


Figure 7: Operating performance of some industries and companies

Based on the information in Fig. 7, we can draw a second conclusion. In the last two quarters of the time, the operating performance of 'Liquor' industry is better. In this industry, std_EVA of 'KWEICHOW MOUTAI', stock code is '600519.SH', is the highest. That means this enterprise has the best operating performance. However, the 'Bank' industry and Insurance industry are underperforming. Among them, 'Industrial and Commercial Bank of China Limited (ICBC)', with '601398.SH' as the stock code, has the lowest std_EVA , which indicates operating result of ICBC is not optimistic. At this stage, the macro-level visualization of the business performance of listed companies has been completed.

Finally, we can do the visualization analysis of business performance at the micro level. First choose a company to analyze. The standard for choosing a company is not unique. We can choose 'KWEICHOW MOUTAI Co., Ltd.' as the object of Analysis, hereinafter referred to as 'KWEICHOW MOUTAI'. Because the company's operating conditions are the best among all listed companies. We would like to use DuPont Analytics to conduct an in-depth analysis of the company's operating performance. Combined with the DuPont analysis theory in the second chapter, we can calculate the company's financial indicators of the company in each quarter and draw a line chart by python.

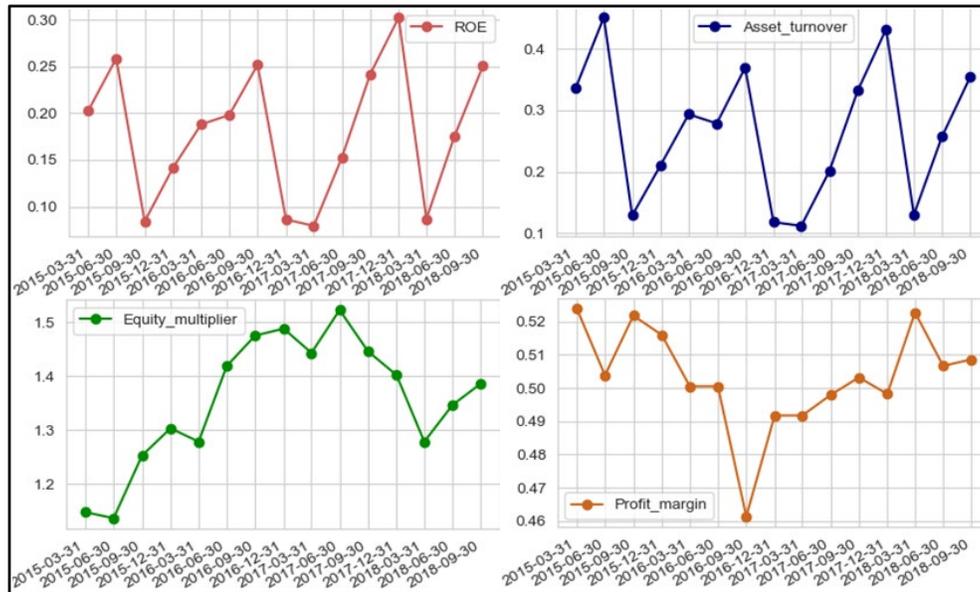


Figure 8: Changes in various indicators of KWEICHOW MOUTAI

As can be seen from Fig. 8, the ROE of ‘KWEICHOW MOUTAI’ has changed roughly between 0.025 and 0.3 over the past three years, and ROE has shown an upward trend in the last two quarters. ‘Assets turnover’ is almost the same as the change in ROE. ‘Profit Margin’, unlike other indicators, has fallen first and then increased in the last two quarters. In order to reflect the influence degree of each index numerically, we also use factor analysis method. The following calculation results are calculated by python program:

Table 2: Calculation results of index changes

Changes	Values
total_change	0.163469
Profit_margin_change	-0.002347
Asset_turnover_change	0.146331
Equity_multiplier_change	0.019485

According to the above calculation Tab. 2 and Fig. 8, the following conclusions can be drawn. First, in the last two quarters, operating effect of ‘KWEICHOW MOUTAI’ has gradually improved. The ROE index of the second ‘KWEICHOW MOUTAI’ is greatly influenced by ‘Assets turnover’, and the sales ability is the main factor affecting the company’s operating performance. Third, in the last two quarters, the occupations of debt for ‘KWEICHOW MOUTAI’ in total assets have gradually increased. Fourth, the changes of company’s profitability have little influence very much to ROE. Accordingly, at the investor’s perspective, the company should focus on improving sales capabilities to improve the ROE indicators, in order to further create new value for the company’s shareholders.

4 Conclusion

In the analysis of the performance of listed companies in China, this paper takes all the listed companies in the 'Shanghai and Shenzhen Stock Exchange' as the representative. Firstly, we use python program to collect all listed companies, and process the collected data into graph data. And then, we use the big data visualization tool, gephi, to generate the distribution graph of the management performance of the listed company, and complete a macro-level visualization of all listed companies. The largest four industries in the distribution of business performance in gephi production are the 'Software Service' industry, the 'Component' industry, the 'Chemical Raw Materials' industry and 'Electrical Equipment' industry. Under the help of EVA evaluation model, most of industries in the last two quarters have better operating performances. But the operating performance of 'Bank' and 'Insurance' industry is not good. Finally, take 'KWEICHOW MOUTAI' as an example, we not only analyze the recent performance of the company's owner's equity profitability, but also analyze the recent performance of the factors affecting ROE indicators, which include sales profitability, asset management capacity and financial leverage. The results of these changes are described in numerical form and presented in the form of line charts, respectively. And complete the micro-level analysis of the operating performance of listed companies. In our analysis of 'KWEICHOW MOUTAI', we have found that the most important factor affecting the company was 'Assets turnover'. In order to bring greater economic value to shareholders in the future, 'KWEICHOW MOUTAI' should pay attention to the improvement of sales ability in future operations.

This company performance evaluation process, which combines gephi tools with traditional corporate evaluation methods, can not only see the recent performance of the factors affecting the company's operating performance, but also see the performance of all listed companies from the macroscopic level. In order to more accurately reflect a company's operating performance of good or bad. From the trend of more and more listed companies, this kind of company performance visualization analysis process will be more conducive to the future company performance evaluation work.

Acknowledgment: This research is funded by the National Social Science Fund Project, grant number 14BJL086. This research is funded by the Open Foundation for the University Innovation Platform in the Hunan Province, grant number 18K103; Hunan Provincial Natural Science Foundation of China, grant number 2017JJ2016; Accurate crawler design and implementation with a data cleaning function, National Students innovation and entrepreneurship of training program, grant number 201811532010; This research work is implemented at the 2011 Collaborative Innovation Center for Development and Utilization of Finance and Economics Big Data Property, Universities of Hunan Province. Open project, grant number 20181901CRP03, 20181901CRP04, 20181901CRP05; National Social Science Fund Project: Research on the Impact Mechanism of China's Capital Space Flow on Regional Economic Development (Project No. 14BJL086)

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- Ahmad, I.** (2018): This is what happens in an internet minute [infographic]. <https://www.socialmediatoday.com/news/this-is-what-happens-in-an-internet-minute-infographic/524426/>.
- Alhadj, R.; Rokne, J.** (2014): *Encyclopedia of Social Network Analysis and Mining*. Springer Publishing Company.
- Bahri, M.; St-Pierre, J.; Sakka, O.** (2011): Economic value added: a useful tool for sme performance management. *International Journal of Productivity and Performance Management*, vol. 60, no. 6, pp. 603-621.
- Bauer, W.; Braun, F.; Elezi, F.; Maurer, M.** (2014): Monetary evaluation of platform strategies from variant costs assessment to economic value added. *IEEE International Systems Conference Proceedings*, pp. 567-574.
- Beattie, V.; Edwards, K.; Goodacre, A.** (1998): The impact of constructive operating lease capitalisation on key accounting ratios. *Accounting and Business Research*, vol. 28, no. 4, pp. 233-254.
- Bodie, Z.; Kane, A.; Marcus, A. J.** (2004): *Essentials of Investments 5th Edition*, McGraw-Hill Irwin.
- Chatfield, M.; Vangermeersch, R.** (1996): *History of Accounting: An International Encyclopedia*. Garland Publishing, Inc. University of Mississippi Libraries. pp. 88-101. http://clio.lib.olemiss.edu/cdm/compoundobject/collection/aah/id/726/rec/3#img_view_container.
- Chen, Y.** (1987): Discussion on the basic principles of factor analysis, *Accounting Research*, no. 1 pp. 38-44.
- Cherven, K.** (2013): *Network Graph Analysis and Visualization with Gephi*. Packt Publishing Ltd.
- Liu, Q.; Shi, X.; Wang, X.; Li, J.** (2017): 5G development in china: from policy strategy to user-oriented architecture. *Mobile Information Systems*, pp. 1-11.
- Cockcroft, S.; Russell, M.** (2018): Big data opportunities for accounting and finance practice and research. *Australian Accounting Review*, vol. 28, no. 3, pp. 323-333.
- Crosson, S. V.; Needles, B. E.; Powers, M.** (2008): *Principles of Accounting*, Boston: Houghton Mifflin.
- Feng, H.** (2012): Discrimination and redefinition of relevant concepts of financial analysis. *Communication of Finance and Accounting*, no. 29, pp. 113-114.
- Fernandes, N.** (2014): *Finance for Executives-A Practical Guide for Managers*, NPV Publishing.
- French, J.** (2017): The one: a simulation of CAPM market returns. *Journal of Wealth Management*, vol. 20, no. 1, pp. 126-147.
- Guo, F.; Liu, P.; Zhang, C.; Chen, W.; Han, W. et al.** (2019): Research on the law of

garlic price based on big data. *Computers, Materials & Continua*, vol. 58, no. 3, pp. 795-808.

Hargrave, M. (2019): Return on equity-roe definition.

<https://www.investopedia.com/terms/r/returnonequity.asp>.

Huang, Y. Y.; Handfield, R. B. (2015): Measuring the benefits of ERP on supply management maturity model: a “big data” method. *International Journal of Operations & Production Management*, vol. 35, no. 1, pp. 2-25.

Khomba, J. K. (2015): Conceptualisation of the balanced scorecard (bsc) model. *International Journal of Commerce and Management*, vol. 25, no. 4, pp. 424-441.

Kshetri, N. (2016): Big data’s role in expanding access to financial services in China. *International Journal of Information Management*, vol. 36, no. 3, pp. 297-308.

Kumar, P.; Lee, B. S. (2001): Discrete dividend policy with permanent earnings. *Financial Management*, vol. 30, no. 3, pp. 55-76.

Lin, W. T.; Chen, Y. H. (1990): Investment horizon and beta coefficients. *Journal of Business Research*, vol. 21, no. 1, pp. 19-37.

Liu, Y.; Du, Y. (2017): *Network Data Visualization and Analysis Tool: Gephi Chinese Tutorial*, Publishing House of Electronics Industry.

Manyika, J.; Chui, M.; Brown, B.; Bughin, J.; Dobbs, R. et al. (2011): Big data: the next frontier for innovation, competition, and productivity.

https://www.mckinsey.com/~media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Big%20data%20The%20next%20frontier%20for%20innovation/MGI_big_data_full_report.ashx.

Mendenhall, W.; Sincich, T. (2006): *Statistics for Engineering and the Sciences 5th Edition*. Prentice-Hall, Inc.

Sarvari, P. A.; Ustundag, A.; Takci, H. (2016): Performance evaluation of different customer segmentation approaches based on RFM and demographics analysis. *Kybernetes*, vol. 45, no. 7, pp. 1129-1157.

Shrote, K. R.; Deorankar, A. V. (2016): Review based service recommendation for big data. *2nd International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics*, pp. 470-474.

Soliman, M. T. (2008): The use of DuPont analysis by market participants. *Accounting Review*, vol. 83, no. 3, pp. 823-853.

Unlu, U.; Saygin, O.; Ersoy, E. (2015): Wealth added index (WAI) in performance measurement: an empirical application on BIST 30 index. *Social Science Electronic Publishing*, vol. 2, no. 1, pp. 18-28.

Stern, J. M.; Stewart, G. B.; Chew, D. H. (1996): Eva: an integrated financial management system. *European Financial Management*, vol. 2, no. 2, pp. 223-245.

Wang, P. (1996): A novel method of factor analysis: differential and chain supersession approach. *Journal of Xi'an Jiaotong University*, vol. 30, no. 6, pp. 103-108.

White, T. (2015): *Hadoop: The Definitive Guide, 4th Edition*, O'Reilly Media, Inc.

Yang, J.; Chang, X.; Shi, S.; Yang, S. (2017): Comparison of complex network analysis software: citespace, SCI2 and Gephi. *IEEE 2nd International Conference on Big Data*

Analysis, pp. 169-172.

Zhuhadar, L.; Yang, R. (2012): Cyberlearners and learning resources. *2nd International Conference on Learning Analytics and Knowledge*, pp. 65-68.