 JBD, vol.1, no.1, pp.25-38, 2019

Brief Talk About Big Data Graph Analysis and Visualization

Guang Sun[1, 2], Fenghua Li[1, *] and Wangdong Jiang[1]

Abstract: Graphical methods are used for construction. Data analysis and visualization are an important area of applications of big data. At the same time, visual analysis is also an important method for big data analysis. Data visualization refers to data that is presented in a visual form, such as a chart or map, to help people understand the meaning of the data. Data visualization helps people extract meaning from data quickly and easily. Visualization can be used to fully demonstrate the patterns, trends, and dependencies of your data, which can be found in other displays. Big data visualization analysis combines the advantages of computers, which can be static or interactive, interactive analysis methods and interactive technologies, which can directly help people and effectively understand the information behind big data. It is indispensable in the era of big data visualization, and it can be very intuitive if used properly. Graphical analysis also found that valuable information becomes a powerful tool in complex data relationships, and it represents a significant business opportunity. With the rise of big data, important technologies suitable for dealing with complex relationships have emerged. Graphics come in a variety of shapes and sizes for a variety of business problems. Graphic analysis is first in the visualization. The step is to get the right data and answer the goal. In short, to choose the right method, you must understand each relative strengths and weaknesses and understand the data. Key steps to get data: target; collect; clean; connect.

Keywords: Big data, visualization, information visualization, graph analysis.

1 Research background and significance

This article is about the application of graph analysis and visualization in the era of big data. The first part is the background and significance of this study. The second part is intended to introduce the value of the actuality of the art graphics for commercial analysis. The third part summarizes the effective principles and techniques of graphical analysis and visualization.

As the amount, diversity, and speed of available data increases, so does the demand for technology and technology. According to IDC, global production in 2009 was 0.8 ZB and in 2012 it was 2.8 ZB. According to historical data, it is estimated that the amount of data will reach 44 ZB in 2020, and the amount of data in 2030 will reach 2500 ZB. These predictions show that humans have entered an era of rapid data growth. The massive unstructured data generated in the context of big data, the importance of big data is not

[1] Hunan University of Finance and Economics, Changsha, 410001, China.

[2] The University of Alabama, Tuscaloosa 35401, USA.

[*] Corresponding Author: FengHua Li. Email: 1095145133@qq.com.

only reflected in how to store massive data. In recent years, the rapid development of technology of big data has also been favored by more and more scholars. The storage and calculation of massive data has also been resolved. At the same time, it is more important to analyze and visualize the vast amounts of data to obtain more valuable information. The sole purpose of information visualization is to obtain more information from massive chaotic data and achieve it in a shorter time [Du (2004); Yang and Liu (2014)]. More importantly, the chart provides a way to get highly unique and valuable insights from the data. This can discover complex relationships and provide effective decision.

Because the brain needs time to reflect the meaning of a feeling, visual images allow the brain to immediately understand its meaning. Yuan [Yuan (2014)] Also used in life sciences: on Taobao or Tmall, all products have larger pictures than words. On the highway, street signs are displayed in the form of symbols rather than words. The scientific principle behind it is that people's vision is first captured by a pleasant image and then carefully read the details. Even road traffic sign recognition is an important task in intelligent transportation systems [Zhou, Liang and Li (2018)]. In meteorological and hydrological research, the accuracy of existing snow-covered products in mountainous areas is significantly lower than that in plain areas due to the confusion of clouds caused by high altitude and complex terrain during severe snow.In response to this problem, Xi Kan proposed an improved method about mapping snows in the mountains [Chen and Lin (2016)].

In the era of big data, many data and wealthy companies around the world are looking for new ways to understand a large number of complex, irregular, sometimes unverifiable and interrelated data [Dai, Chen and Hang (2013); Zen (2014)]. Graphical analysis and visualization, as a tool to help achieve this goal, is gaining momentum. Charts are particularly good at describing complex relationships that are difficult to describe in black and white. They are also a natural choice for display networks, and display networks are an increasingly important part of many commercial data sets.

Computer visualization of business decisions is a relatively new phenomenon. With the maturity of the hardware graphics function, the sense of visual value has gradually matured. Timely, accurate, and rapid understanding of events and trends are all critical to making lightning decisions for decision makers. In busy analysis, the graphical representation of information helps decision makers understand the strategic value of decision making [Wu and Wang (2010)].

As information problems and complexity increase in the business, basic lines, bars, and pie charts rarely express all useful information and use them to make decisions. A richer combination of forms and forms is needed. At the same time, because of the latest developments in open source graphical tools and libraries, charts are available to everyone, but knowledge of the principles and techniques of graphical analysis and visualization is still relatively limited.

In an environment of big data, companies must continue to absorb knowledge of big data and private knowledge through transfer multiple knowledge to maintain a competitive advantage [Wu, Zapevalova, Chen et al. (2018)]. has a visualization function that can fully display the patterns, trends, and correlations of data, which may be difficult to achieve in other representations.

Data visualization can be static or interactive. Static data visualizations, such as charts and maps, have been used for centuries. According to the Pew research center, when a designer has a lot of information and can't cut it, they usually put them in a data table to make it more compact. Replace with a histogram. Because in the map, the viewer can see every media channel on the spectrum. In the distribution map, the distance between the media channels is particularly significant. If these points are only listed in the table, the viewer will not be able to see the comparison between each channel. Fig. 1 shows the distribution of channels about audience of political news.

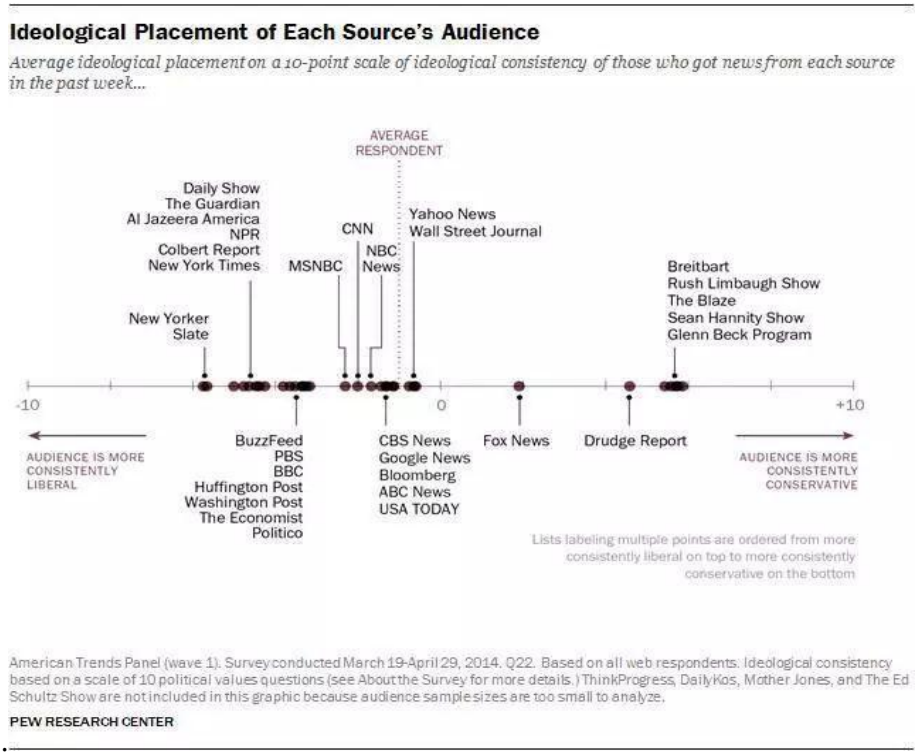

Figure 1: Distribution map of channels about audience of political news

In recent years, interactive data visualization has become a hot topic. Interactive data visualization is more advanced: people can use computers and mobile devices to delve into the details of these charts and graphs and then change data and data processing interactively. As shown in Fig. 2.

Figure 2: World language map

This interaction, designed by high density, is an impressive result that shows the non-verbal way of many languages in the world. There are 2,678.

In this picture, users can browse families using the same common language, and which languages are most commonly used, and which languages are used around the world. This is an extraordinary visual narrative: a profound subject that is easy to understand.

2 The value of graph analysis and visualization for commercial analysis

The first typical graph problem in the history of graph analysis was solved by Leonhard Euler, who proposed that the seven bridges in Prussian Königsberg (now Kaliningrad) span only once, as shown in Fig. 3. Euler reduced the problem to a map, and since then, more and more problems have been analyzed in the commercial and scientific fields.

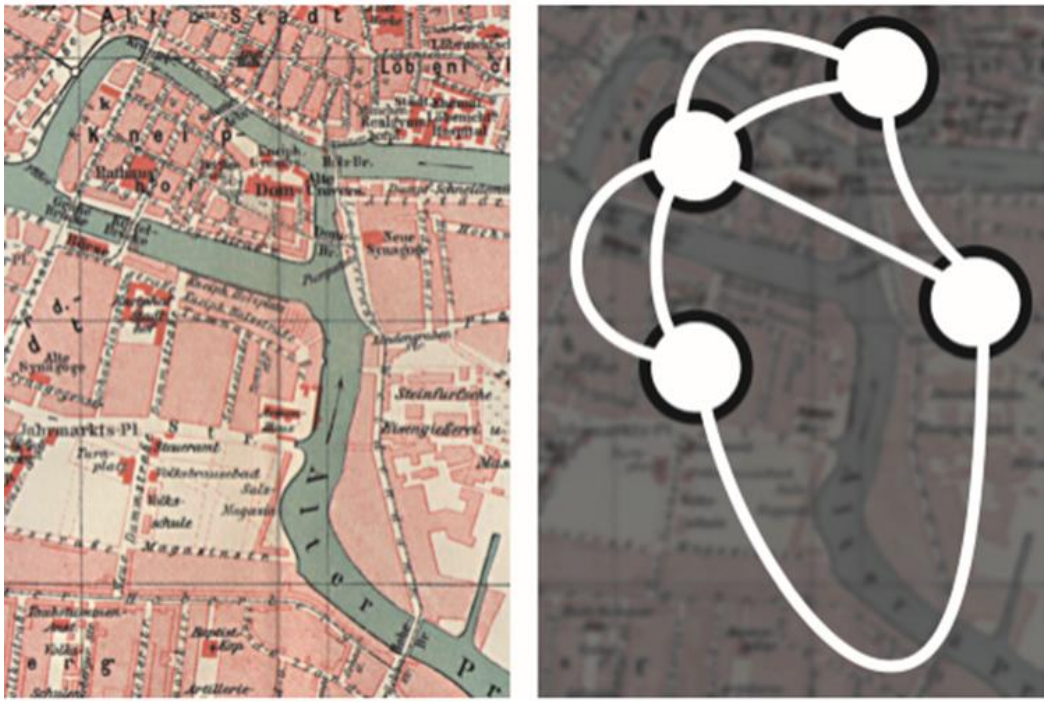

Figure 3: Euler's schematic diagram of seven bridge problems

Figure 3: Leonhard Euler explores whether each of the seven Bridges in the Konigsberg problem can only be traversed once. On the left is a map with seven Bridges, and on the right is a map transformed by problems.

2.1 Find associated

The most important and basic point of the diagram is to represent a relational model that builds a defined world or system. The chart shows a visual factor model that can sometimes help decision makers make informed business decisions.

Charts can be generated by a computer. Although the form varies from method to method, in essence, any form of graphic structure, body and object are represented by nodes, and the relationship is represented by a connecting line. The representation of the relationship in the figure can be simplified to a straight line, sometimes with a specific weight to indicate strength or volume. But in reality, potential relationships tend to exhibit more subtle or broader characteristics than simple lines. If the domain to be expressed is very small, then the visual representation of the connecting lines and their nodes can more fully explain the nature of the relationships between the entities.

If research shows that a relationship is critical to data science, it is interrelated in any business. Relevance tells people when and how the world is relevant and can help them make decisions. By controlling the factors that can be controlled within the scope, and by affecting the outcome of a given likelihood, understanding the conditions that are most favorable for a given outcome provides a practical basis for action-taking strategies.

Depending on the industry, this strategy may become targeted advertising for decision makers, who adjust premiums based on risk assessments or other actions.

However, when information is frequently exchanged in both directions, it is necessary to use another layout method-a string diagram. A string graph is mainly used to represent the connection between multiple elements, usually divided into two parts, an external node and an internal string. An arc from one element to another represents the correlation between two elements and is used to show the interaction between the entities. As shown in Fig. 4.

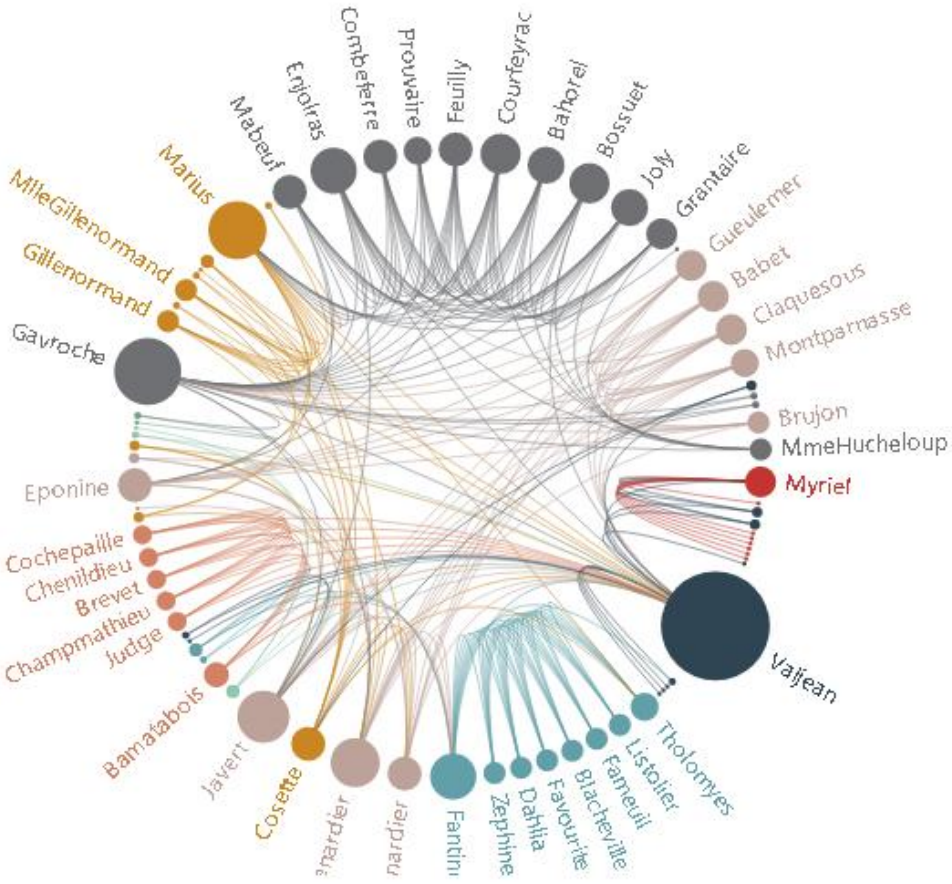

Figure 4: A simple string diagram of a tragic world figure (http:// www.PoPodv.com)

The visualization analysis of network has many applications. Social networks can be used to analyze relationships. Fig. 5 shows the figure of the tragic world (The case from the www.PoPodv.com).

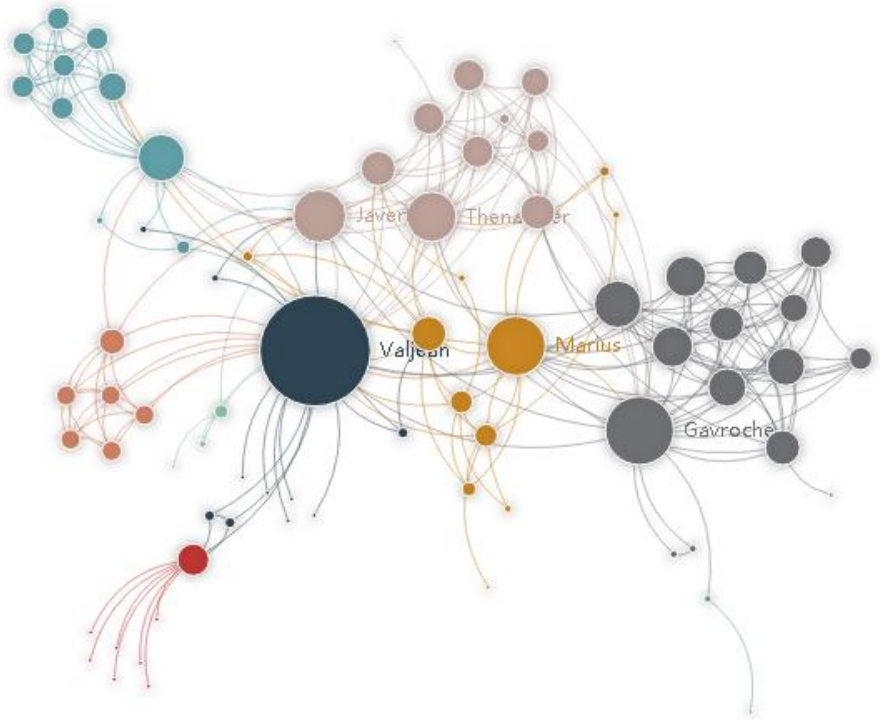

Figure 5: The tragic world figure chart (http://www.PoPodv.com)

Mapping social networks goes back hundreds of years. Fig. 6 shows the French royal family tree from Louis xiv to Louis xvi. This organizational structure visualization shows direct rulers, spouses, descendants, and remerged branches. Each node represents a person, the king a crown, the man a filling circle, and the woman a transparent diamond. According to the time line from top to bottom, the horizontal line style is divided between the children of a married couple (regular line) or the mistress (diamond line).

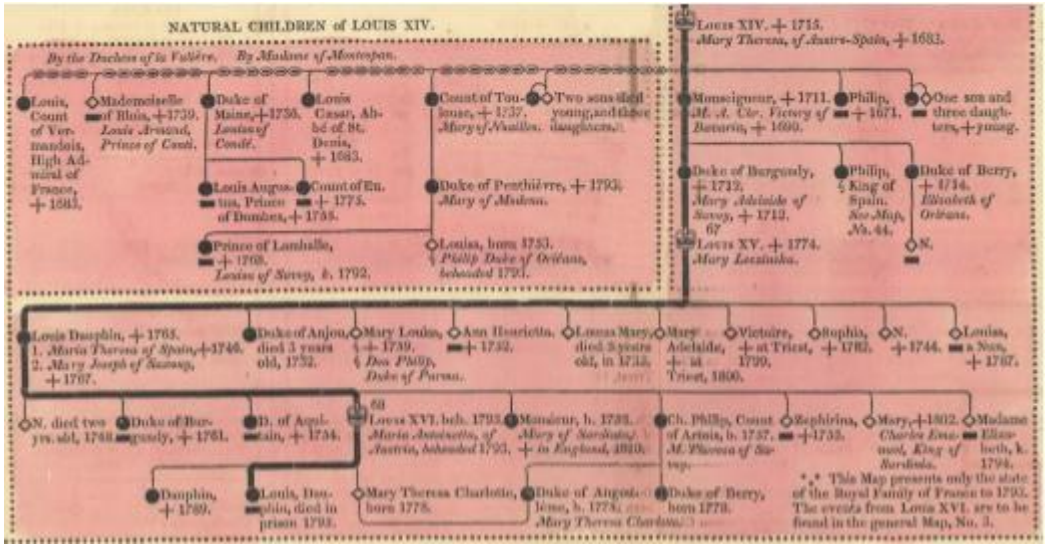

Figure 6: family tree of the French royal family (from Louis xiv to Louis xvi)

Drawing social network maps can be traced back hundreds of years. Fig. 6 shows the genealogy of the French royal family from Louis XIV to Louis XVI. This organizational structure visualizes the direct ruler, spouse, descendants, and recombined branches. Each node represents a person, the king represents the crown, the man represents the circle, and the woman represents the transparent diamond. According to the timeline from top to bottom, the horizontal line style is divided into the children of the couple (ordinary line) and the mistress (diamond line). Analysis of these social networks often reveals dense clusters. Identifying these people and their relationships can help companies identify different customer groups and better understand their internal interactions. For example, Chinese Mobile's model of customer segmentation. An analysis of social networks can provide a detailed analysis of the impacts of individuals or organizations and within and between these groups. An in-depth survey of individuals in the population indicates that most people in the group have the same characteristics. This allows researchers to study a small group of people in a large group to achieve the same behavior. These analysis can also be used to detect network traffic (such as virus propagation) and connection-based customer segmentation. Save time, labor and money.

2.2 Classification

Graphics are also the best choice for getting information from hierarchical data. In the computer world, hierarchical data is usually represented by trees. The tree is also perfect for understanding the organization. Genealogy is a typical example of an organization-level visualization technique in which an ancestor is on the root node and a child node branches off from the parent node. It is often referred to as an organization chart or an organization chart. It is used to indicate the relationship between work and life. A more novel number is the rising sky.

The sun burst is one of the most detailed images, showing the connection between the two. The daily storm map is a circular mosaic. Each ring represents the same level of

scale data. The closer the circle is to the origin, the higher the level. Fig. 7 is an example of a PoPo site showing the same organization containing global coffee flavor survey information and other detailed information. In this case, the root node is centered and the branches are radial. The sibling node represents the child sector of its parent node and represents the exact percentage of the entire node.

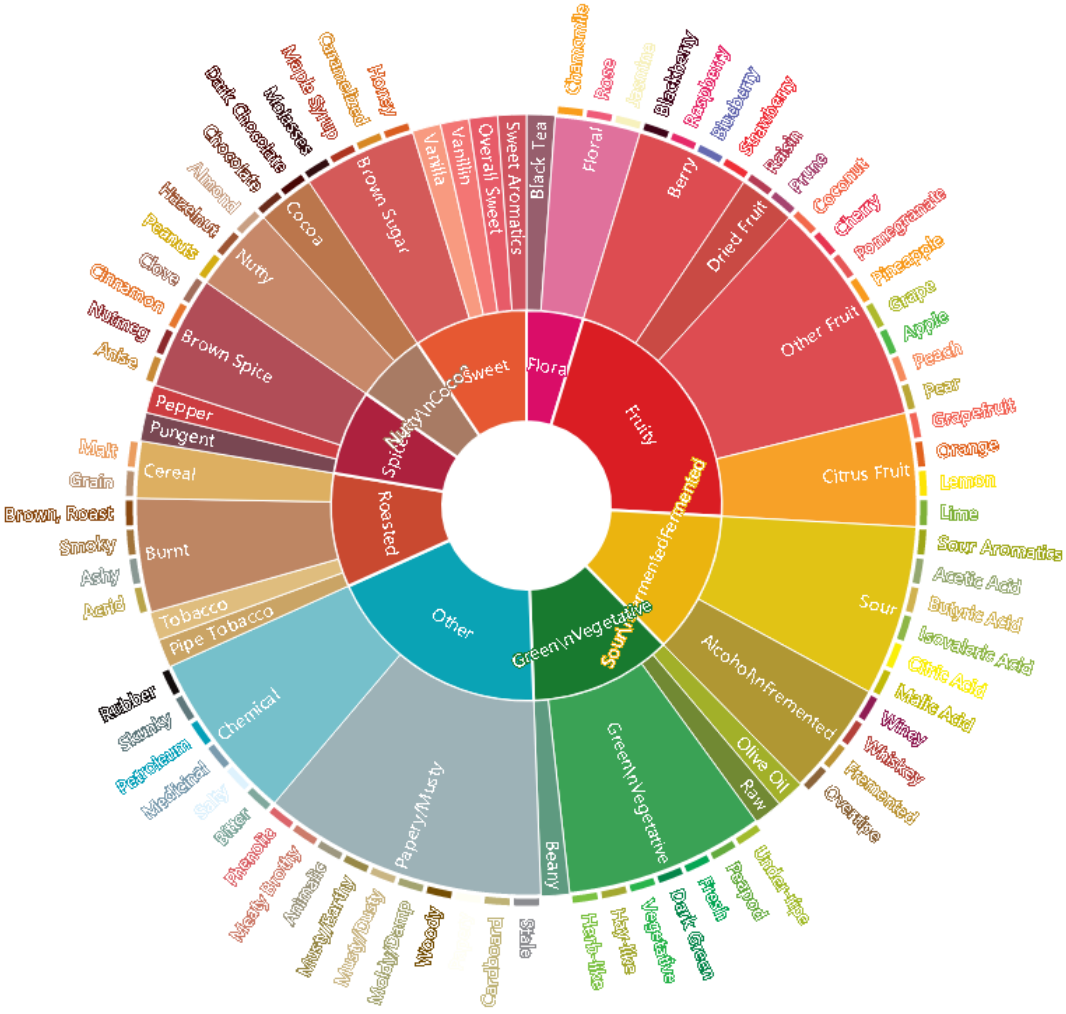

Figure 7: Rising sunburst chart of global coffee taste survey information (http://www.PoPodv.com)

The sunburst chart shows a more detailed hierarchy of elements that can be viewed in proportion to the total. According to the color depth and the size of the fan shaped area, the flavor information of each brand is revealed. The difference between the conventional organization chart and the sunburst chart is that the organization chart can make the qualitative aspect of the tree clearer, while the sunburst chart tends to provide clearer information to the quantitative aspect.

2.3 Spatial analysis

As a density map, thermodynamic data is usually represented by large chromatic aberration. In thermal maps, bright colors generally indicate a higher frequency of events or a higher density of objects, while darker colors indicate the opposite. When decision-makers need to present density problems in various big data information in a more intuitive and effective form, thermal map is undoubtedly a good way. It is worth mentioning that the final effect of thermal maps is often better than the direct display of discrete points, which can intuitively show the density or frequency of spatial data on a two-dimensional plane or map. See Fig. 8 (image from Tencent wedsite).

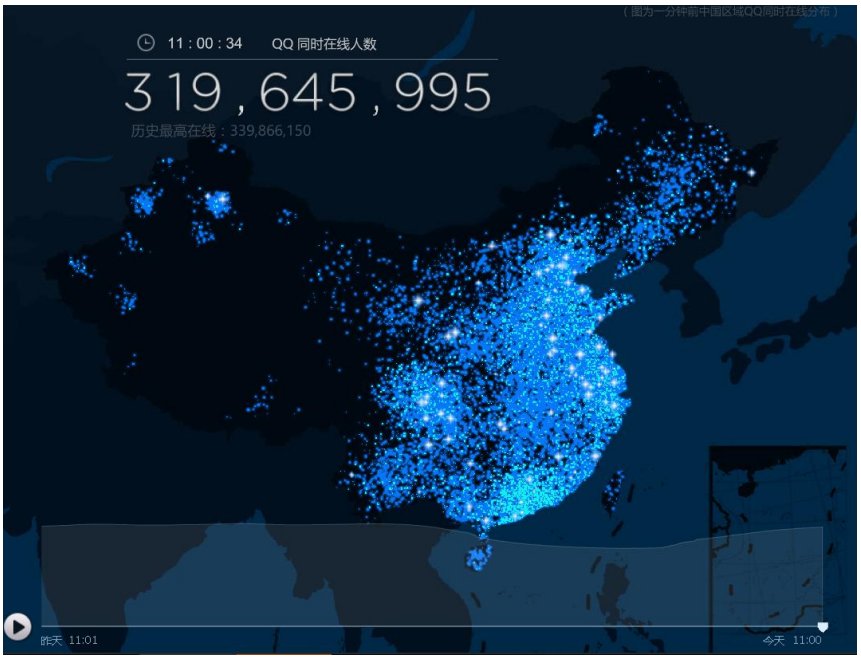

Figure 8: Online population heat map(https://im.qq.com/index.shtml)

Fig. 8 shows the number of people online in China's QQ in the early morning, showing the “night scene” in the busy area of China.

Decision makers can intuitively make geographically relevant strategic arrangements, such as the entry of goods into the market or the choice of investment market.

2.4 Spatial network display

The graph is also the best choice for displaying spatial network data. In a spatial network, the nodes already have a meaningful physical location. One of the challenges, however, is the limited ability of mobile nodes to make links easier to read. In Fig. 9, the subway map shows how the spatial network diagram can abstract the Angle and position. The advantage of abstract spatial layout of schematic network diagram is its readability, as shown in Fig. 9.

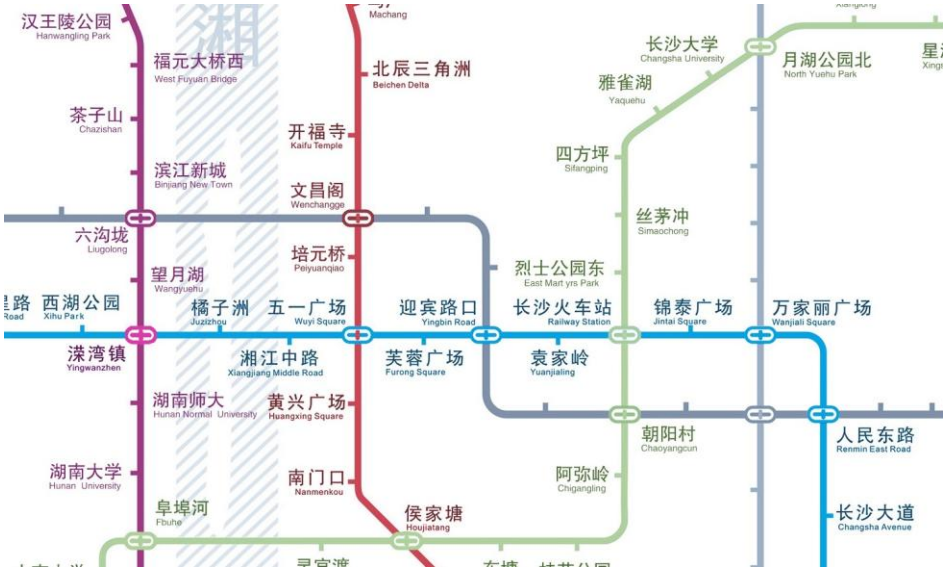

Figure 9: Changsha metro line of 2018

Graphics are an obvious fit for a visual spatial network. Selecting appropriate technologies based on goals will help overcome possible difficulties in space networks.

2.5 A Subsection Sample

Flow charts are often used to analyze the flow of goods through companies or around the world. As shown in Fig. 10, Joseph Minard presents an early example of a flow chart that looks at immigration around the world. It is easy to see the emigration from Britain to the colonies, from France to the United States, and from China to other places.

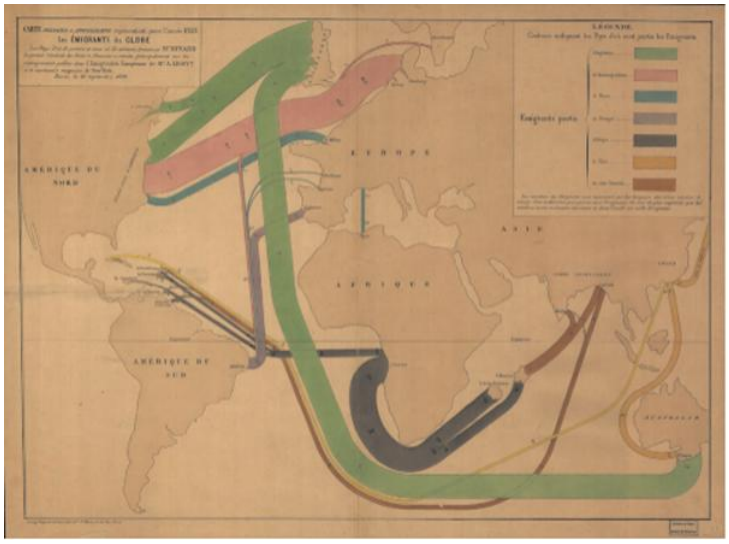

Figure 10: Flow chart of migration around the world in 1858

Flowcharts can be used to analyze the flow of material phenomena in real life, such as the flow of people, goods or money, whether through the world, processes or websites. In the era of the rise of big data, flowchart is the most commonly used graph for analyzing big data.Over time, streaming datasets can quickly become big data. In this case, it is even more necessary to use different strategies to deal with the dynamic flow data. So flowcharts can be used to detect and understand exceptions. Use flow charts to analyze and detect anomalies to improve efficiency. See Fig. 11(from www.PoPodv.com).

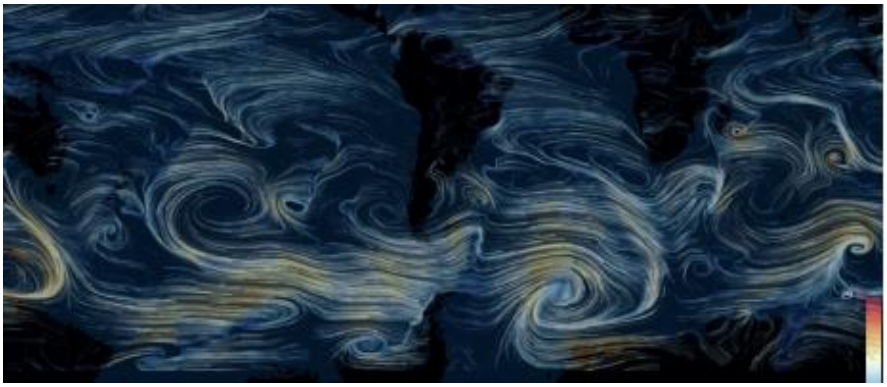

Figure 11: Some functions of x

Fig. 11 shows the flow of the global wind field, which can be used by the weather monitoring department to detect or look for anomalies, make predictions first, and then arrange for prevention

3 Conclusion

Graphical analysis is a powerful tool for finding valuable information about relationships in complex data and providing important business opportunities to observers [Chen (2015)]. Graphical visualization is essential and can be very intuitive if used correctly [Qi, and Gao (2014)]. Information visualization leverages natural perception, enabling analysts to see more information faster. As the amount of available data increases, the importance of visualization in various fields is widely recognized. With the rise of big data, technologies that are suitable for dealing with complex relationships are becoming more and more important. Technology is demand-driven, and graphics have evolved into tools for network monitoring, market basket analysis, influencing analysis, process and organizational representation, and now graphical tools are becoming a valuable resource for business analysts. A graph can be a solution to a problem or an organizational framework with a small amount of other types of visualization. Users can use a model that represents a problem or a high-level concept map to represent links and node relationships. [Chen (2015); Ping (2013); He and Huang (2015); Liu and Lei (2014); Xie (2014); Zhang, Zhang and Wang (2018)] You can use the sunburst diagram to effectively visualize the hierarchy. Nested and overlapping groups are shown in the layout shown in the figure and are represented by symbols or tags. Spatial networks use a method to locate nodes and routes, which helps to sort out the relationship between objects and discoveries. Business problems are multidimensional and cannot be fully illustrated by

diagrams or charts. For example, some issues may involve spatial networks and streaming situations, and of course connections. When choosing a method, try to think about the most basic problems that need to be solved, and then the combination diagram will appear. Shapes and sizes can be used for a variety of business problems and can be grouped and summarized using computational clustering. In order to select the right map to find the problem, you must understand the relative advantages and disadvantages of each graph, understand the data, and most importantly, know what the target is.

Acknowledgement: This research work is supported by Hunan Provincial Education Science 13th Five Year Plan (Grant No. XJK016BXX001), Social Science Foundation of Hunan Province (Grant No. 17YBA049), Hunan Provincial Natural Science Foundation of China (Grant No. 2017JJ2016), National Students' platform for innovation and entrepreneurship training (Grant No. 201811532010). The work is also supported by Open foundation for University Innovation Platform from Hunan Province, China (Grand No. 16K013) and the 2011 Collaborative Innovation Center of Big Data for Financial and Economical Asset Development and Utility in Universities of Hunan Province. We also thank the anonymous reviewers for their valuable comments and insightful suggestions.

References

Chen, M. (2015): *Big Data Visualization Analysis*. Zhejiang University Press.

Chen, M. (2015): Visualization technology research based on complex data processing. *Computer and Digital Engineering*, vol. 43, no. 11, pp. 194-197.

Chen, Y.; Lin, X. L. (2016): An associated hierarchical data visualization method based on heat map and radiation loop. *Journal of computer-Aided Design and Graphics*, vol. 32, no. 7, pp. 12-15.

Dai, G. Z.; Chen, W.; Hang, W. X. (2013): Information visualization and visual analysis: challenges and opportunities. *Chinese science*, vol. 43, no. 1, pp. 178-184.

Du, X. L. (2004): Overview of big data research. *Computer Application Research*, vol. 31, no. 6, pp. 1612-1616.

Du, X. L. (2014): *The Concept of Data Visualization in the Context of Big Data Era (Ph.D. Thesis)*. Zhejiang University Press.

He, X. Q, Huang, L. (2015): The visual analysis of the evolution path of big data field, research hotspots and frontiers. *Modern Intelligence*, vol. 35, no. 4, pp. 46-51.

Liu, X. Q.; Lei, X. L. (2014): Review of big data systems and analysis techniques. *Journal of software*, vol. 25, no. 9, pp. 1889-1908.

Ping, Q. (2013): The major trend of data operation is mainstreaming. *Internet weekly*, vol. 13, no. 7, pp. 26-46.

Qi, C. H.; Gao, S. P. (2014): Hospital information data mining and data visualization.*China Science and Technology Information*, vol. 2014, no. 22, pp. 39-42.

Wang, R. S. (2016): Research on spatial and temporal multidimensional data visualization in big data environment. *Computer Education*, vol. 19, no. 5, pp. 94-97.

Wu, C. R.; Zapevalova, E.; Chen, Y. W.; Li, F. (2018): Time optimization of multiple

knowledge transfers in the big data environment. *Computers, Materials & Continua*, vol. 54, no. 3, pp. 269-285.

Wu, J. X.; Wang, J. H. (2010): Visual perception model based on situational awareness theory. *Modern library and information technology*, no. Z, pp. 9-14.

Wu, Y. B.; Guo, F. L. (2015): Big data visualization analysis. *Computer education*, vol. 12, no. 5, pp. 94-97.

Xiao, B.; Wang, Z.; Liu, Q.; Liu, X. D. (2018): An improved mini batch k-means algorithm based on mapreduce with big data. *Computers, Materials & Continua*, vol. 56, no. 3, pp. 365-379.

Xie, R. (2014): The beauty of big data visualization. *Internet weekly*, vol. 15, no. 6, pp. 32-34.

Yang, Y. B.; Liu, B. (2014): Review of research on information visualization. *Journal of Hebei university of science and technology*, pp. 18-23.

Yuan, S. Y. (2014): Vision, brain and image-the role and influence of neural mechanism in the process of painting. *Research on Art Education*, no. 7, pp. 56-57.

Zen, Y. (2014): Overview of big data research. *Computer application research*, vol. 11, no. 6, pp. 162-1616.

Zhang, H. J.; Zhang, X. L.; Wang, Y. P. (2018): Research on complex cognitive activities in visual analysis.*Journal of computer aided design and graphics*, vol. 30, no. 1, pp. 57-67.

Zhou, S.; Liang, W. L.; Li, J. G. (2018): Improved VGG model for road traffic sign recognition. *Computers, Materials & Continua*, vol. 57, no. 1, pp. 11-24.