[1] Rui Huang
[2] Hailong Gai
[3] Rong Jing
[4,*] Fucheng Wan

Word Spectral Visualization Base on Fruchterman-Reingold Optimized ForceAtlas2

Journal of Electrical Systems

Abstract: - This article presents a visualization technique employing the BERT model similarity lexicon and delves into the method of picking and enhancing layout algorithms when generating similar term lexicon exhibitions. In a bid to achieve superior visualization outcomes, two distinct layout algorithms, namely ForceAtlas2 and Fruchterman-Reingold, are utilized to leverage their individual merits and augment visualization quality. The similar term lexicon's visual depiction is accomplished using the ForceAtlas2 and Fruchterman-Reingold layout algorithms, with the amalgamation of ForceAtlas2's initial layout and Fruchterman-Reingold's fine-tuning resulting in a presentation map of the similar term lexicon with a superior visualization impact. This tool equips us with an efficient method to profoundly interpret the semantic relationships of textual data. Through this method, we can perceive the connections and associations among words more explicitly, aiding in unveiling the concealed information and meanings in the text data. Further research can further refine the visualization and broaden the application domain to cater to the escalating demands. This method is anticipated to play a vital role in text-connected tasks, offering a potent tool for extracting knowledge and information from textual data

Keywords: Similar words, Visualization of lexicons, ForceAtlas2, Fruchterman-Reingold

I. INTRODUCTION

In the past few years, the extensive proliferation of information technology has given rise to a multitude of data across diverse industries. Mobile devices, social media platforms, e-commerce websites, and sensors are just a few examples of the many sources that continuously generate various types of data. This phenomenon has made big data a central topic in both research and application, with a primary challenge being the extraction of valuable insights from the deluge of data. However, the true value of big data does not lie solely in its volume; rather, it is found in its ability to be processed and mined efficiently for the insights it contains. As such, we highlight that the true value of big data is in its "valuableness" and not just its "size". Traditional statistical analysis methods are well-suited for small-scale data, but they become ineffective when dealing with the ever-expanding scale of data.

In recent years, data mining and machine learning methods have emerged that can uncover insights from large-scale data, but they often struggle to present these insights in an intuitive manner. This is where information visualization technology steps in. This technology presents data in a graphical format, offering a more intuitive means of understanding big data while also allowing users to interact with the data. A plethora of methods and techniques have been developed in information visualization, but many have not been widely adopted. This is primarily due to the fact that many visualization methods focus excessively on technological innovation, neglecting the cognitive needs of users. As such, designing visualization methods that adhere to the cognitive patterns of users and can intuitively convey the insights of big data remains a crucial challenge [1].

In intricate networks with a multitude of nodes and interconnections, conventional textual representation techniques are unable to satisfy users' demands for comprehending and extracting valuable insights from network structures. The field of computer science frequently employs graph structures for examining social networks, computer networks, transportation networks, and numerous other applications. These graph structures aid in graphically illustrating relationships among various projects. To enhance readers' understanding of the data presented in charts, we implement layout algorithms to manipulate and display the charts' structures [2]. Numerous network visualization layout algorithms are available, such as force-directed layout, circular layout, and map layout. Each method, however, has its strengths and weaknesses, with the primary objective of graph layout being

[1] Key Laboratory of Linguistic and Cultural Computing Ministry of Education, Northwest Minzu University, Lanzhou, Gansu 730030, China
[2] Key Laboratory of Linguistic and Cultural Computing Ministry of Education, Northwest Minzu University, Lanzhou, Gansu 730030, China
[3] Key Laboratory of Linguistic and Cultural Computing Ministry of Education, Northwest Minzu University, Lanzhou, Gansu 730030, China
[4] Key Laboratory of China's Ethnic Languages and Intelligent Processing of Gansu Province, Northwest Minzu University, Lanzhou, Gansu, China
*Corresponding author: Fucheng Wan

to present the crucial features of the network in an intuitive manner [3]. Graph visualization of network data rapidly uncovers vital information in a visual format, aiding individuals in comprehending network relationships more intuitively. Consequently, research into automatic visualization layout algorithms for complex networks has emerged as a popular field.

II. DEVELOPMENT STATUS OF LEXICON VISUALIZATION

Lexicon visualization is a highly attention-grabbing area within data visualization. Layout assumes a vital role in complex networks, encompassing a wide array of network visualization algorithms. Visual Complexity documents over 600 applications of network visualization. Based on Shneiderman's classification criteria [4], network graph visualization can implement various layout techniques, such as force-directed layout, map layout, ring layout, relative space layout, and clustering layout, among others [5]. These visualizations facilitate users' accurate and comprehensive understanding of text data, opening up more opportunities for textual analysis and comprehension. An effective network layout can not only present a substantial amount of data within a confined space but also enhance users' ease of comprehending network structures. Empirical results indicate that the layout of a graph significantly influences people's perception and comprehension of data. Consequently, it is essential to assess the graph's layout algorithm, and a layout that adheres to human cognitive principles and mental imagery is easier to comprehend, aiding users in gaining a deeper appreciation of the graph's structure, thereby enticing more users to utilize it [6]..

III. VISUALIZATION OF GRAPH LAYOUTS

A Force-Directed Layout:

Force-directed layout is a prevalent approach for visualizing general meshed structures [7]. This layout method is rooted in a physical model wherein the connections among nodes are depicted by springs, and attractive and repulsive forces exist between these nodes. The length of the connections is determined by the node-to-node relationship; nodes with attractive forces tend to move closer, guaranteeing that the distance between nodes and their adjacent nodes remains within a reasonable range. On the other hand, nodes with repulsive forces push each other away, stopping any two nodes from overlapping due to their close proximity, and as a result, forestalling a densely packed network that could impede visualization.

B Circular Layout:

The circular layout uniformly disperses nodes around a circle, accentuating the balance among nodes. This layout is particularly suitable for small-scale networks. By evenly arranging nodes around the circle, it becomes possible to arrange them based on a specific sequence or node attributes, with the smallest nodes located inside the circle, ensuring efficient space usage within the circle [8].

C Hierarchical Layout:

Hierarchical layout algorithms generally consist of three stages: network coarsening, initial layout, and network refinement [9]. This layout method segregates the graphical network into various levels, where each level signifies a distinct structural tier. It is employed for presenting tree-like or hierarchical structures. Nodes are organized based on their position within the hierarchical structure, with parent nodes situated above and child nodes below.

D Tree Layout:

In 1991, Johnson introduced the tree graph [10]. This layout represents the network in a hierarchical tree structure, making it well-suited for displaying hierarchical structures like organizational charts or website navigation maps. The nodes are organized based on their position within the tree structure, with each node having a single parent node and potentially zero or more child nodes.

E Kamada-Kawai Layout:

The Kamada-Kawai layout positions graphical networks by optimizing the path length between nodes to minimize edge crossings and improve visual clarity. This is achieved by minimizing the objective function, commonly employing iterative techniques.

IV. Force-Directed Layout

The force-directed layout offers several benefits for visualizing graphical networks. It can take into account the overall structure of the entire network by simulating forces between nodes to adjust their positions, showcasing the network's overall characteristics and topological structure, and facilitating user comprehension of the relationships and organization within the network. The force-directed layout promotes a uniform distribution of nodes, preventing crowding and overlapping, and enhancing visual clarity, making nodes and edges easier to distinguish and identify. The force-directed layout is versatile and can be applied to various types of networks, including social networks, biological networks, geographical networks, among others, making it extensively useful across different fields [11]. Moreover, it naturally represents relationships between nodes, with attractive forces between connected nodes and repulsive forces between non-connected nodes simulating the strength of actual relationships, making the visualization more intuitive. A unique aspect of the force-directed layout is its node placement process. This process solely depends on the connections between nodes, and the final attributes of the nodes are never taken into account. Its essence lies in converting structural proximity into visual proximity, making it convenient for analysis, particularly when examining social networks [12].

[1] *ForceAtlas2*

ForceAtlas2 is a force-directed layout that calculates the positions of nodes by modeling various forces between them. These forces encompass attraction (attracting connected nodes), repulsion (repelling other nodes), and damping forces between nodes (decelerating the movement of nodes), among others. Nodes exhibit a magnetic-like repulsion, while edges exert a spring-like attraction, as illustrated in Figure 1. These forces establish a dynamic system that converges to an equilibrium state [13].

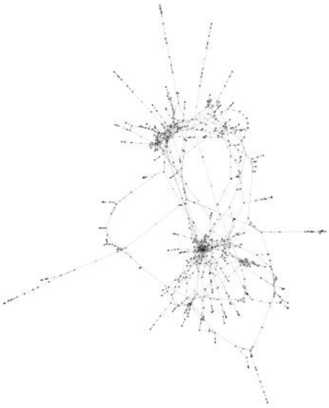

Figure 1: Force-directed Layout Using the FA Model

GravityF_a Connecting nodes $n_1 and\ n_2$, It is linearly related to the distance $d(n_1, n_2)$。

$$F_a(n_1, n_2) = kd(n_1, n_2) \tag{1}$$

A prevalent characteristic of networks utilizing ForceAtlas2 is the existence of numerous "leaves" (nodes with just one neighbor). This is attributed to the fact that many real-world data adheres to a power-law distribution. Modifying the repulsion force aids in reducing this specific visual clutter, enabling poorly connected nodes and nodes with exceedingly close connections to repel each other to a lesser extent, thus bringing them closer to equilibrium [14]. The repulsion force F_r between disconnected nodes intensifies as they move closer to one another, and its purpose is to forestall disconnected nodes from getting too close to each other, thus preventing node overlap. The force F_r is proportional to the product of the degrees plus one $(deg + 1)$ of the two nodes, with k_r representing the strength of the repulsion force. An elevated k_r value amplifies the repulsion, making it more challenging for nodes to come close to each other [15].

$$F_r(n_1, n_2) = k_r \frac{(deg(n_1)+1)(deg(n_2)+1)}{d(n_1,n_2)} \tag{2}$$

The damping force F_d is used to slow down the movement of nodes to ensure that the layout converges to an equilibrium state. F_d is proportional to the node's velocity v, where β is a constant representing the strength of the damping force.

$$F_d(v) = -\beta * v \tag{3}$$

Nodes are influenced by the simultaneous effects of attractive, repulsive, and damping forces, with the net force F_i denoting the overall force acting on node i. ForceAtlas2 computes the total force on each node and recursively alters the node positions to drive the overall force approach equilibrium, thus enhancing the node position layout. The fundamental concept of ForceAtlas2 entails continuously refining the adjustment of node positions by calculating the total force on each node, enabling the layout among nodes to ultimately reach a stable state, thereby accomplishing the network visualization layout. These equations constitute the backbone of the ForceAtlas2 algorithm, and by suitably adjusting the parameters, it can be employed in various types of networks to obtain optimal layout outcomes.

$$F_i = F_a + F_r + F_d \tag{4}$$

B Fruchterman-Reingold

The Fruchterman-Reingold algorithm, a member of the force-directed graph layout algorithm family, is among the most renowned graph layout methods. This approach adopts elements from the spring model to emulate the node layout process: it employs springs to represent relationships between nodes, wherein nodes that are too close are repelled by spring forces, while distant nodes are attracted closer together, as illustrated in Figure 2. Through iterative refinement, the overall layout converges to dynamic equilibrium, gradually stabilizing and evolving into the FR algorithm, also known as the force-directed layout algorithm. This enhanced algorithm incorporates a physical model between two nodes and introduces electrostatic forces. By minimizing the total energy of the system, it aims to optimize node layout. It adheres to two fundamental principles: nodes connected by edges should be in close proximity, and nodes should not be excessively dense. This algorithm computes node displacements by considering the interactions of attractive and repulsive forces between nodes, following principles akin to the motion of atoms or planets. Eventually, the system reaches a dynamic equilibrium state.

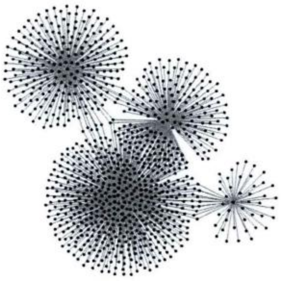

Figure 2: Force-Directed Layout Using the FR Model

The height of the display area is denoted as H, the width as W, the position of nodes as pos, and the position offset as dis.

Define 'a' as the display area.

$$a = W * H \tag{5}$$

Equilibrium distance，Where $|v|$ represents the number of nodes in the graph:

$$k = \sqrt{\frac{a}{|v|}} \tag{6}$$

Geometric distance between two points:

$$dis(u,v) = \sqrt{(u.pos_x - v.pos_x)^2 + \left(u.pos_y - v.pos_y\right)^2} \tag{7}$$

Repulsive force f_r between two points:

$$f_r(u,v) = k^2/dis(u,v) \tag{8}$$

In the Fruchterman-Reingold layout, the spring force simulates the attractive effect between connected nodes, causing adjacent nodes to be drawn closer to each other. The calculation of spring force typically follows Hooke's law:

$$f_a(n_1,n_2) = kd(n_1,n_2) \tag{9}$$

V. Application of Lexical Spectrum Visualization

In the course of this research, an in-depth exploration of layout algorithm selection and optimization processes for the creation of lexical spectrum visualizations was carried out. In order to achieve the most favorable visualization outcomes, two distinct layout algorithms, ForceAtlas2 and Fruchterman-Reingold, were employed to capitalize on their individual strengths and augment the visualization's quality.

Initially, ForceAtlas2 was chosen as the foundation due to its notable performance benefits when handling large-scale and intricate networks. This algorithm effectively captures the overall structure of networks and boasts high scalability, making it suitable for a wide range of network types. By utilizing ForceAtlas2, a suitable initial layout can be obtained, ensuring that the lexical spectrum visualization of cognate words exhibits coherence and clarity.

Following this, the Fruchterman-Reingold layout was applied to further enhance the visualization. The Fruchterman-Reingold layout accentuates repulsive and spring forces between nodes, creating a well-balanced layout by emulating these forces. This serves to emphasize the relationships between similar words. By implementing Fruchterman-Reingold on top of the ForceAtlas2 layout, the visualization's aesthetics and readability were further optimized, enabling the audience to comprehend the relationships and structure between similar words more easily.

By integrating these two layout algorithms, their individual strengths are fully harnessed while preserving the global structural characteristics of ForceAtlas2 and enhancing the local layout effectiveness of Fruchterman-Reingold. This hybrid approach offers a more comprehensive and visually appealing representation of the lexical spectrum, ultimately enhancing the overall visualization quality and facilitating a better understanding of the relationships between words for the intended audience.

A Selection of Similar Words Spectrum

Incorporating the BERT model for identifying analogous terms capitalizes on the model's comprehensive linguistic representations, enabling a more accurate depiction of semantic associations among words and yielding a superior collection of similar terms.

The degree of resemblance among word embeddings is determined by examining the cosine similarity of word embedding vectors yielded by the BERT model. The mathematical expression for calculating the cosine similarity between two word vectors, A and B, is as follows [17]:

$$Cosine\ Similarity(A,B) = \frac{A*B}{||A||*||B||} \tag{10}$$

In this context, $A * B$ represents the dot product of vectors, and $||A||$ and $||B||$ respectively denote the magnitudes (or norms) of vectors A and B.

Compute the resemblance between the target word and the remaining words within the corpus, and retain words with a similarity score surpassing a predetermined cutoff. For instance, as illustrated in Table I:

Table 1: Partial Lexicon of Similar Words

core word	**Similar words 1**	**Similar words 2**	**Similar words 3**	**Similar words 4**	**Similar words 5**
content	study for practical applications	mode	specific	stage	recruit
today	good morning	Full	birthday	really	days
phone	find	Meizu	Taobao	client	Flash payment
client	download	Landscape screen	mobile phone	Sina.com	address
Cloud atmosphere	fine with occasional clouds	Mining and excavation	Inconveniences	as a topic	Tycoon

B Generating Initial Layout with ForceAtlas2

In comparison with the Fruchterman-Reingold layout, ForceAtlas2 exhibits more potent attractive forces in the overall layout, accompanied by reduced repulsive interactions between nodes. This facilitates its ability to efficiently apprehend the global architecture of the network [16], thereby enabling macro-level control. Utilizing the lexicon of analogous terms, ForceAtlas2 was employed to generate the initial layout for a total of 300 words, as illustrated in Figure 3:

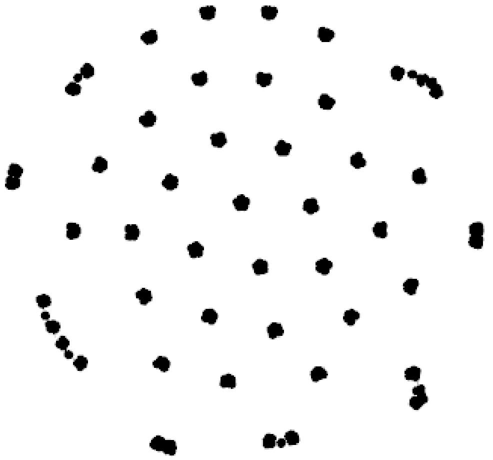

Figure 3: Initial layout of ForceAtlas2

In Figure 3, the arrangement of nodes follows the principles of the ForceAtlas2 algorithm, wherein nodes with numerous connections are positioned towards the periphery of the graph, while those with fewer connections tend to aggregate near the central core. This strategic positioning is a direct result of the algorithm's balancing act, where it fosters attractive forces among the highly connected nodes and repulsive forces among the less connected nodes[18]. Consequently, this layout naturally creates clusters and communities, mirroring the semantic relationships that exist within the lexical spectrum. The visual representation that emerges from this arrangement offers enhanced clarity. This outcome aligns with the algorithm's primary objective, which is to optimize lexical analysis visualization[19].

C Fine-Tuning with Fruchterman-Reingold Layout

Building upon the ForceAtlas2 layout, the fine-tuning process employs the Fruchterman-Reingold layout. This layout is characterized by local repulsive forces that are more robust than attractive forces. As a result, connected nodes tend to distance themselves from one another due to the overwhelming repulsive forces outweighing the attractive ones. This phenomenon elucidates distinct lexical spectrum relationships, as illustrated in Figure 4.:

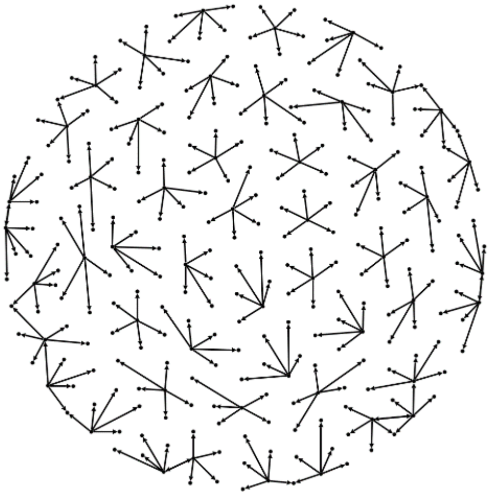

Figure 4: Fine-Tuning with Fruchterman-Reingold

Upon implementing the Fruchterman-Reingold (FR) algorithm for fine-tuning, the primary layout of the lexical spectrum continues to be reliant on the ForceAtlas2 (FA) layout. Nonetheless, the employment of lexical spectrum space is significantly improved, leading to a more dense filling of the graph. Consequently, this approach culminates in a more visually intuitive and clear-cut depiction of the lexical spectrum, thereby enhancing overall comprehension and interpretation of the presented data[20].

VI. CONCLUSION

Upon completing the initial layout using the ForceAtlas2 method and subsequent fine-tuning with the Fruchterman-Reingold algorithm, further processing of the lexical spectrum representation produces Figure 5. Within the lexicon of similar words, those that are closer to the core word 1 exhibit a stronger relationship and possess greater significance, in contrast to word 5, which is further away from the core word and carries less influence. In the lexical spectrum visualization, the weight of similar words is depicted through the thickness and color saturation of the directed edges that emanate from the core word and point towards these similar words. This unique representation allows for a more discernible and comprehensive understanding of the lexical relationships, thereby enhancing the overall interpretability of the presented data.

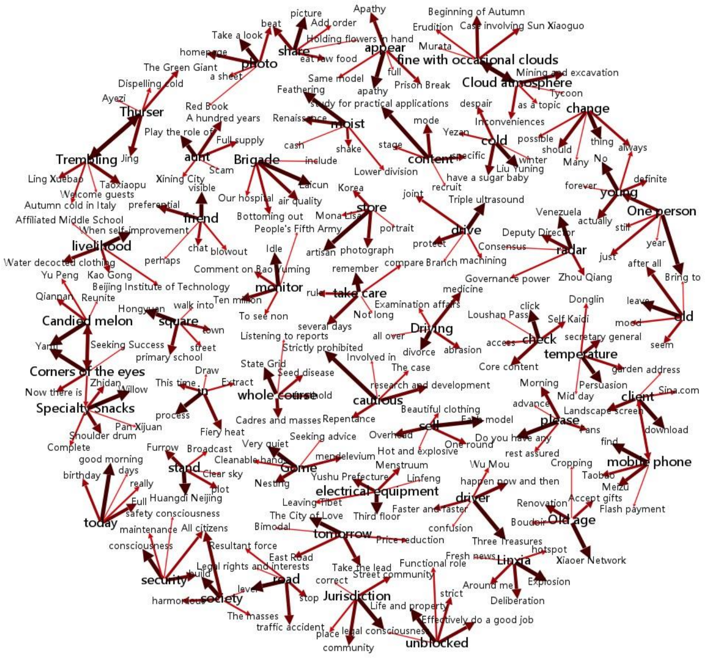

Figure 5: Lexical Spectrum Display

In Figure 6, a segment of the graph is presented as illustrated in Figure 5. The core words "Candied melon", "Corners of the eyes", and "Specialty Snacks" are featured, with "Yanti" being a similar word 1 for both "Candied melon" and "Corners of the eyes". Simultaneously, "Candied melon" and "Corners of the eyes" serve as mutual similar word 2 for each other, while "Specialty Snacks" is identified as a similar word 4 for "Corners of the eyes". The lexical spectrum display offers a clear and intuitive depiction of the relationships that exist within the vocabulary of the lexical spectrum.

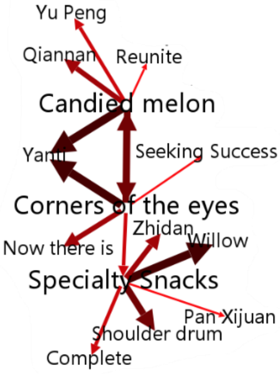

Figure 6: Part of the lyric sheet

In essence, the visual representation introduced in this study represents a momentous leap forward in our ability to decipher the intricate network of word connections within a lexical spectrum. The augmented clarity and depth it affords enable a more profound recognition of complex patterns and structures embedded in the data, thereby enriching our understanding of language and its subtleties.

However, it is crucial to recognize that there are certain limitations to this methodology. One significant drawback is the potential demand for considerable computational resources, which could present difficulties for users with restricted access to high-performance computing environments. Moreover, fine-tuning the Fruchterman-Reingold algorithm may necessitate a certain level of expertise, rendering it less accessible to individuals lacking specialized knowledge in data visualization. Notwithstanding these constraints, the fusion of the ForceAtlas2 and Fruchterman-Reingold algorithms remains a formidable tool for lexical analysis. It markedly enhances the quality and expressiveness of visualizations, offering a comprehensive and meticulous depiction of similar word lexicons. This approach sets a new benchmark for the field, providing valuable insights and guidance for future scholarly endeavors.

In conclusion, the captivating behavior of nodes within our visualization, wherein highly connected nodes tend to move towards the periphery and less connected nodes congregate closer to the core, significantly contributes to enhancing the clarity and practicality of the visual representation. This inherent organizational pattern eerily echoes real-world semantic relationships, typically characterized by central, prominent concepts surrounded by peripheral, albeit less significant, elements. By accurately emulating this natural structure, our visualization enables viewers to extract valuable insights more efficiently from the labyrinthine lexical spectrum. The inherent organization, however, is not just a visual curiosity but a potent instrument for comprehension. It enables viewers to swiftly pinpoint central concepts and appreciate the broader context in which they exist, much like how pivotal words shape the overall meaning of a text. Simultaneously, the clustering of less connected nodes at the core facilitates the identification of less prevalent or specialized terms, presenting a nuanced view of the lexical landscape. In essence, this behavior establishes a harmonious visual hierarchy, akin to the hierarchical structure of concepts in language, empowering researchers and practitioners to navigate and decipher complex textual data with greater accuracy and depth.

ACKNOWLEDGMENT

This work is supported by Lanzhou Chengguan District Science and Technology Plan Project(2021RCCX0016)and the Fundamental Research Funds for the Central Universities (NO. 31920230004).

REFERENCES

[1] Ren Lei, Du Yi, Ma Shuai, et al. A Survey of Big Data Visual Analytics. Journal of Software, 2014(9): 1909-1936.

[2] P . Gajdos , T . Jezowicz , V . Uher , P . Dohnalek , A parallel fruchtermanreingold algorithm optimized for fast visualization of large graphs and swarms of data , Swarm and Evolutionary Computation 26(2016)56-63.

[3] Chen L, Jian X. Research on methods and tools of social network visualization. Data Analysis and Knowledge Discovery, 2012, 28(5): 7-15.

[4] Readings in information visualization: using vision to think. Morgan Kaufmann, 1999.

[5] Shneiderman B , Aris A . Network Visualization by Semantic Substrates .IEEE Transactions on Visualization & Computer Graphics ,2006,12(5):733.

[6] Purchase H C, Carrington D, Allder J A. Empirical evaluation of aesthetics-based graph layout. Empirical Software Engineering, 2002, 7: 233-255.

[7] Kamada T, Kawai S. An algorithm for drawing general undirected graphs. Information processing letters, 1989, 31(1): 7-15.

[8] Fucheng Wan, Dongjiao Zhang, Lei Zhang, Ao Zhu. Question Similarity calculating method towards medical question answering system, basic clin pharmacol,2021,127(3):278-293.

[9] Walshaw C . A multilevel algorithm for force - directed graph drawing// Proc of the 19th International Symposium on Grap h Drawing . Heidelberg . Berlin : Springer ,2011:171-182.

[10] Johnson , B . and B . Shneiderman . Tree - maps : A space - filling approach to the visualization of hierarchical information structu res . in Visualization ,1991.Visualization& apos ;91, Proceedings ., IEEE Conference on .1991.IEEE

[11] Cheong S H, Si Y W. Force-directed algorithms for schematic drawings and placement: A survey. Information Visualization, 2020, 19(1): 65-91.

[12] Jacomy M, Venturini T, Heymann S, et al. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. PloS one, 2014, 9(6): e98679.

[13] Mathieu J, Tommaso V. ForceAtlas2, A Graph Layout Algorithm for Handy Network Visualization. Draft article, 2011.

[14] Newman M E J. Analysis of weighted networks. Physical review E, 2004, 70(5): 056131.

[15] T.M.J. Fruchterman, E M. Reingold, Graph drawing by force-directed placement, Software-Practice & Experience 21 (1991) 1129-1164.

[16] Jiang X, Song C, Xu Y, et al. Research on sentiment classification for netizens based on the BERT-BiLSTM-TextCNN model. PeerJ Computer Science, 2022, 8: e1005.

[17] Jacomy M, Heymann S, Venturini T, et al. ForceAtlas2, a graph layout algorithm for handy network visualization. Sciences Po, medialab, 2011.

[18] Fucheng Wan. Medical Information Extraction Technology Based on Association Rules. Indian Journal of Pharmaceutical Sciences,2018[3]

[19] Wan Fucheng,Yang Yimin, Zhu Dengyun, et al. Semantic Role Labeling Integrated with Multilevel Linguistic Cues and Bi-LSTM-CRF.Mathematical Problems in Engineering, 2022, 2022.

[20] Wan Fucheng , Yang Fangtao , Wu Titantian ,et al. Chinese shallow semantic parsing based on multilevel linguistic clues.Journal of Computational Methods in Sciences and Engineering, 2020(2):1-10.DOI:10.3233/JCM-194111.