Proceedings of the 3rd International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/6/20230678

# Analysis of network resilience global air transportation

Yaoxuan Liu

Telecommunications Engineering with Management, International School, Beijing University of Posts and Telecommunications, Beijing, 100000, PR China

2726298755@bupt.edu.cn

**Abstract**. Small-world network is a very common network structure, which is characterized by low average degree, small average path length and high centrality. At the same time, small-world networks have high resilience to random errors and low resilience to targeted attacks. In this study, the importance of nodes is represented by attributes such as degree and centrality, and attacks refer to the removal of important nodes. The network is attacked according to the degree, betweenness and closeness centrality to observe the power distribution. The data is mainly obtained from the open source OpenFlight. Gephi, Python, and Excel are used as tools. Gephi is used for network visualization and analysis. The third-party python libraries Pandas, Matplotlib, and NetworkX were used in this study to deal with the things that Gephi can't compute or represent well, and then plot the corresponding graphs with Matplotlib. The work of cleaning data is mainly done by excel.

**Keywords:** Small-World Network, Power-Law Distribution, Air Transport Network, Node Degree and Centrality, Network Attack, Geographical Distribution.

## 1. Introduction

There exists a class of small-world networks. Power-law distribution is satisfied by this class of small-world networks. This kind of network has the characteristics of low average degree, small average path length and high centrality. It is a very common kind of network structure and has high resilience to random errors and low resilience to targeted attacks. Therefore, in this study, the impact of directed attacks on key nodes is investigated. In this study, the importance of a node with properties such as its degree and centrality will be represented. An attack refers to removing important nodes. Important nodes will be removed in order of importance to simulate a targeted attack on the network.

The air transport network is a typical case of this type of network. A data set about the world's air transportation network was obtained from the Internet. The author will operate on it to simulate attacks and observe the impact of targeted attacks with different priorities on the network. The goal isto understand whether attacks on key nodes change the degree distribution of the network, thus breaking the power-law distribution, and whether properties such as small average path length resulting from the power-law distribution fluctuate significantly.

The air transport network is also an entity built by humans in the real world, so it will be related to something in the real world. In this study, the authorfocusses on the connection between the community distribution of global airports and their geographical distribution. The author would also like to understand the impact of attacks on this connection.

DOI: 10.54254/2755-2721/6/20230678

## 2. Methodology

### *2.1. Dataset*

OpenFlight is a type of open-source project that allows users map flights around the world, search and filter them in all sorts of interesting ways, calculate statistics automatically. Standard formats in virtual reality applications include OpenFlight [1]. Open Flight model is a data model for describing 3D virtual scenes introduced by Multigene-Paradigm Corporation (MPI). With the increase of the scale of 3D visual simulation system, 3D solid modeling takes more and more time, which occupies a lot of system development time and extends the development cycle of visual simulation system. Currently, there are many standard file formats in the field of visual simulation, and OpenFlight is one of the most widely used formats [2]. OpenFlights makes their data publicly available on GitHub. The public dataset contains data on airlines, airports, countries, locales, planes, and routes in CSV format with the extension ".dat". Therefore, for the convenience of subsequent data processing, we need to manually adjust the extension name of these files.

In study, airports and routes are used to provide the required information for the nodes and edges of the network respectively. The file routes.csv contains 67,663pieces of flight information, where the origin and destination of each flight are represented by the IATA code of the airport.In terms of data collection and acquisition users test hypotheses according to IATA guidance and follow a structured format to facilitate effective decision making [3]. The airports.csv file contains information for 7,698 airports, including 6,072 records with valid IATA codes. The IATA code is used as an index. Gephi and NetworkX are used to build the undirected graphs.

Some other information related to airports and flights is also provided in the dataset. The following information need to be emphasizedin study: the longitude and latitude of the airport in airports.csv and the number of stops of flights in routes.csv.Latitude and longitude information can help locate the nodes on the world map to visualize our results in a more intuitive way. For example, the distribution of the airport community around the world.

The stop-over information of flights indicates that the edges in our constructed network do not reflect the adjacency relationship between airports completely accurately, and some records may only reflect the connectivity relationship between airports rather than the adjacency relationship. Upon inspection, the existence of a total of 11 flights with stops does not significantly affect the overall accuracy of the study, but we think it is still worth our brief explanation.

### *2.2. Tools*

*2.2.1. Gephi.* In network visualization and analysis, the open source software Gephi is often used. Through visualizing and analysing networks, Gephican reveal trends and unveil stories with the data [4]. Data analysts use Gephi to visually reveal patterns and trends，highlight outliers.Gephi combines built-in functionalities and flexible architecture to explore, analyze, spatialize, filter, cluster, manipulate or export all types of networks. Gephi is often used as a network visualization software in many fields such as social network analysis [5]. Gephi has a dedicated 3D rendering engine that can be used to assist the visualization module in real-time rendering, so that the graphics card can be used by the graphics processing department, while the CPU usage is low[6].Gephi has strong visualization and dynamic analysis capabilities. Dynamic big data observability analysis works well with Gephi [7].

*2.2.2. Python.* Interpreting and analysing big data has always been a challenging task.Python is an interpreted programming language[8]. Python is easy to use, and you can use IPython or Jupyter Notebook for interactive programming, which makes working with data intuitive and clear. At the same time, Python can easily access a large number of powerful third-party libraries using the package manager pip, which makes some complex operations acceptable in our experiments. The third-party libraries Pandas, Matplotlib, and NetworkX are used. Pandas is used for data processing.However, many features of Pandas were not used in the study because the data obtained basically met the

DOI: 10.54254/2755-2721/6/20230678

requirements.Matplotlib working library is a powerful and open-source plotting tool. Matplotlib is developedbasedonpython. Matplotlib has many advantages, such as high plotting accuracy and simple code [9]. Different from other software or tools, Matplotlib working library is a simple programming software that integrates various powerful functions and advantages. To this end, based on the analysis of various visualization software, the Matplotlib module under Python is very important for the visualization of data analysis. Networkx is developed based on Python language, which is a powerful tool for analyzing complex networks. It provides widely used graph and complex network algorithms. Users can use NetworkX to understand how information spreads and generate intuition about communities formed through interactions [10]. NetworkX is based on graph theory and statistical design and covers many commonly used graph models and network models. NetworkX has a number of graph and network related algorithmic implementations built in, which can significantly reduce the difficulty of analyzing data from a network perspective. Some of the results we obtained can be visualized more clearly and intuitively, so we used Matplotlib to visualize some of the results.

*2.2.3. Excel.* As described in 2.1, there are illegal records in the data set, and data cleaning steps are needed to ensure the normal use of the data. Since the data set is relatively small and there is no requirement for reusability of operations, our data cleaning work is mainly done by using Excel.

### *2.3. Experimental design*

The distribution of airports and flights around the world is visualized using Gephi data. The average path length, average clustering coefficient and network diameter were calculated by using Gephi's built-in algorithm. In addition, the community detection algorithm should be used to calculate the modularity of the network, and the network should be divided into communities and colored separately to observe the clustering situation among global airports.

The filtering function of Gephi is used to simulate the attack on a specific node. After the attack, the network attributes are recalculated and compared with the attributes before the attack to understand the impact of the attack on the network. We will also examine the visual changes caused by the attack.

For things that cannot be computed or represented well with Gephi, Pandas and NetworkX can be used to handle them and draw them with Matplotlib. Simply check whether the degree distribution of the network satisfies a power-law distribution by using Python. The key nodes are screened and compared from three aspects: degree, mediation centrality and proximity centrality. In addition, it is necessary to analyze the impact of attacks on the network structure decomposition and study the impact of targeted attacks on the network degree distribution.

## 3. Result

### *3.1. Properties of original network*

The average degree of the network is 6.177, the average clustering coefficient is 0.628, the diameter is 13, and the average path length is 4.103.

Scatter plots are plotted in linear and logarithmic coordinates according to the degree distribution of the network. Through observation, it can be found that the degree distribution of the network generally satisfies the power-law distribution, as shown in Figure 1.

DOI: 10.54254/2755-2721/6/20230678

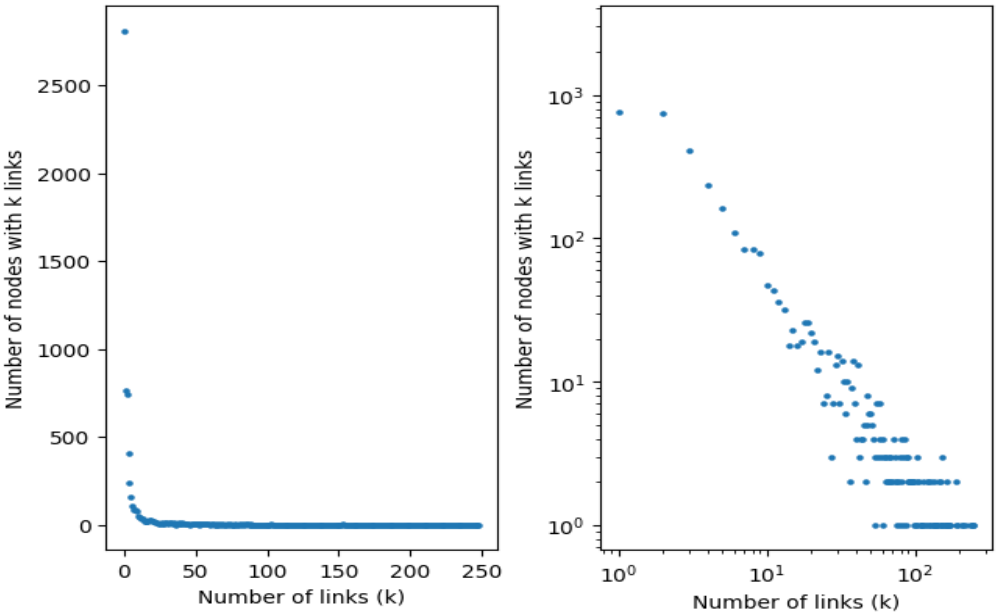


**Figure 1.** Scatter plots in linear and logarithmic coordinates.

*3.2. Communities of global airports*

A total of 7698 airports were recorded in the original data set, and 6072 airports were recorded after removing illegal records. After importing the flight data set, the modularity of the network was calculated to be 0.655. Based on this, communities were partitioned and colored respectively, and the effect was as shown as below in Figure 2:

**Figure 2.** Graph of airports divided into communities and colored.

DOI: 10.54254/2755-2721/6/20230678

*3.3. Targeted attacks according to different properties*

The impact of attacking key nodes on the number of components in the network and the impact of selecting key nodes from three different perspectives of node degree, betweenness centrality and closeness centrality on the attack effect are studied. The impact of selecting key nodes from three different perspectives of node degree, betweenness centrality and closeness centrality on the attack effect. In Figure 3, the abscissa represents the cumulative number of nodes according to the descending priority, and the ordinate represents the number of components present in the network at the corresponding time. Because there are isolated nodes in the dataset, the ordinate does not start at 1. It can be seen from Figure 3 that the number of components in the network first increases and then decreases. The number of components shows a maximum at around 1000 nodes removed. The effect of the attack based on betweenness centrality is more significant than the other two:

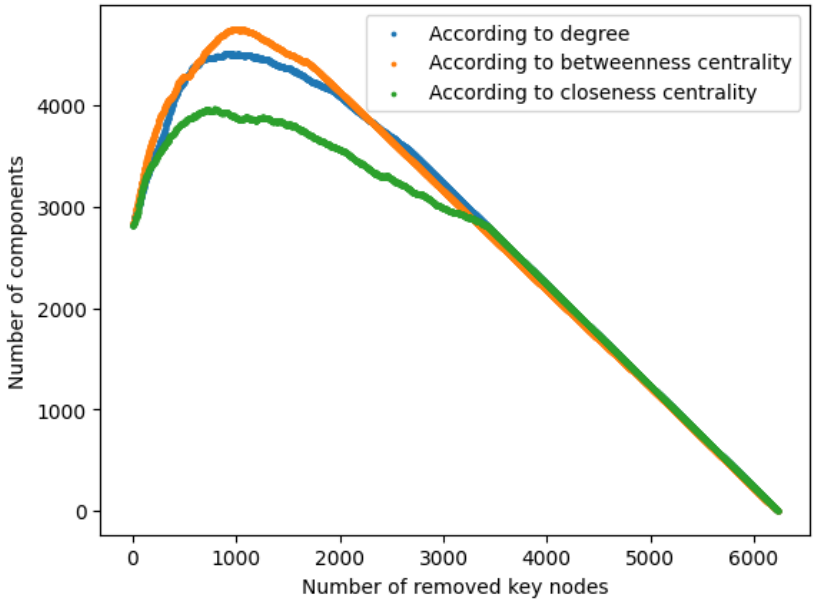


**Figure 3.** Trend graph of the number of components under different attacks.

*3.4. Attack in different circumstances*

*3.4.1. Attack according to degree.* As it can be seen from Figure 4, after removing the top 5% or so of the nodes in the network, a large number of edges in the network disappear, and many of the previously larger communities split into a large number of smaller communities that are rendered grey by the software (because there are not enough colours to label them one by one). The average degree of the network drops to 1.177, the average clustering coefficient becomes 0.458, the diameter surges to 35, and the average path length also increases to 12.368.

DOI: 10.54254/2755-2721/6/20230678

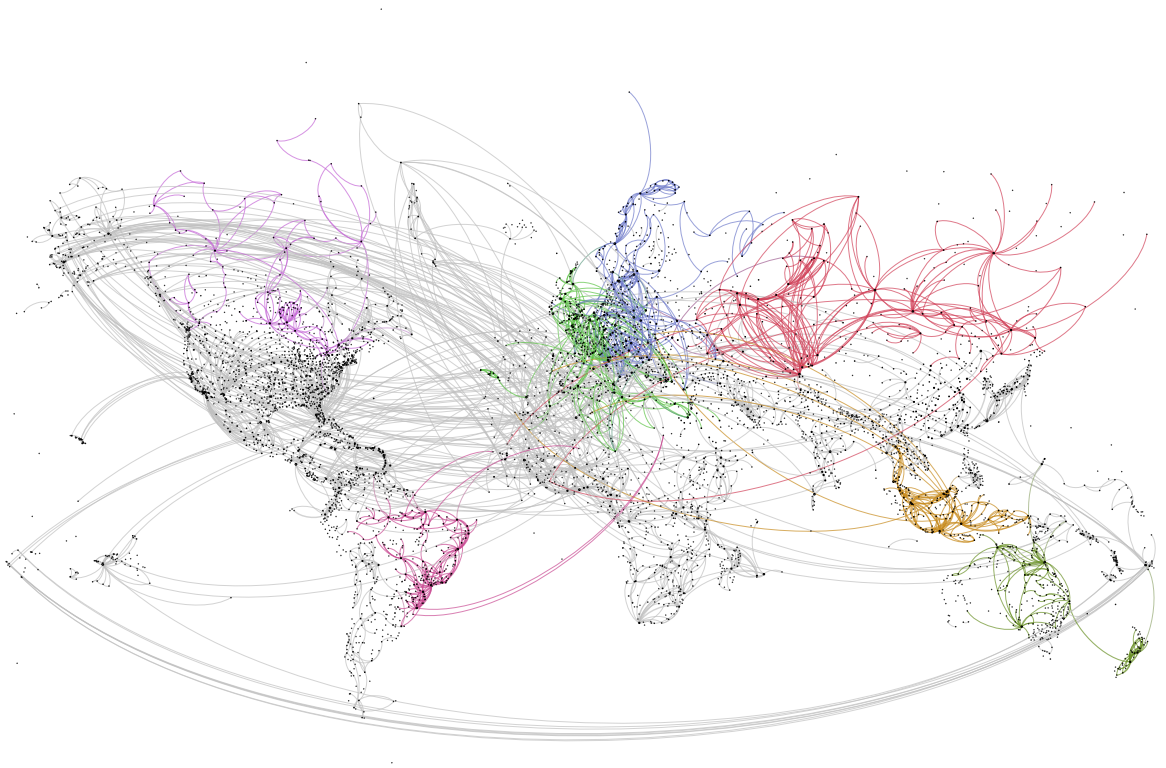

**Figure 4.** Graph of coloured airports under attack according to degree.

*3.4.2. Attack according to betweenness centrality.* As it can be seen from Figure5, after removing about 5% of nodes with the highest betweenness centrality in the network, some large communities split, however there are some communities that do not seem to be affected much. The average degree of the network drops to 2.104, the average clustering coefficient becomes 0.43, the diameter increases to 20, and the average path length increases to 6.299.

DOI: 10.54254/2755-2721/6/20230678

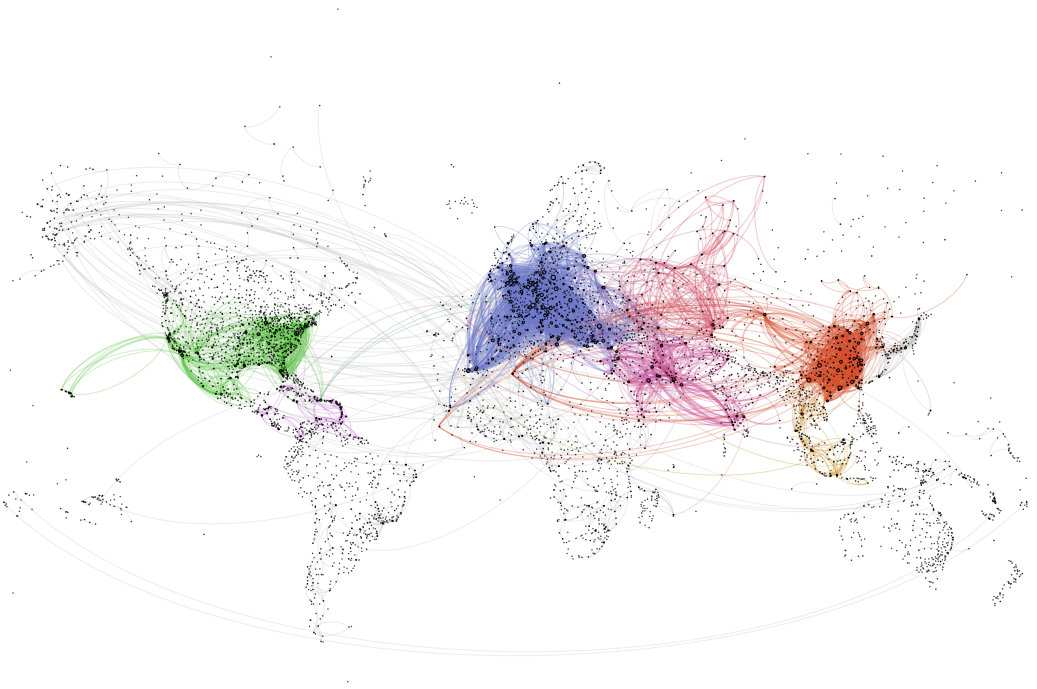

**Figure 5.** Graphofcolored airports underattack according to betweenness centrality.

*3.4.3. Attack according to closeness centrality.* As it can be seen from Figure6, after removing about 5% of highest closeness centrality nodes in the network, the average degree of the network becomes 6.308, the average clustering coefficient becomes 0.613, the diameter becomes 12, and the average path length becomes 3.928. The changes are very small, and even better than the original ones:

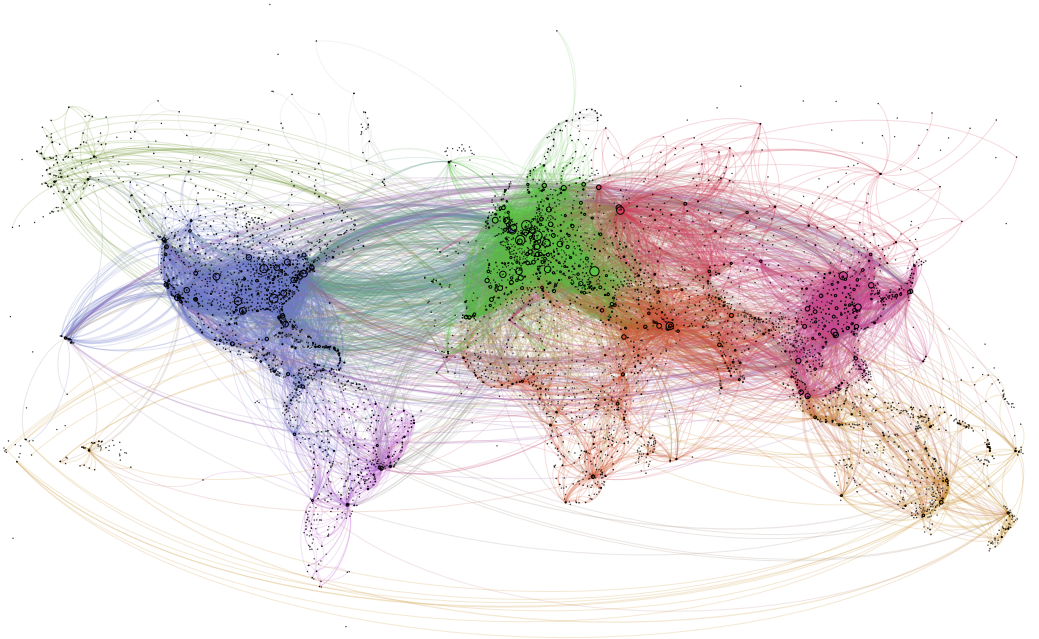

**Figure 6.** Graphofcolored airports underattack according to closeness centrality.

DOI: 10.54254/2755-2721/6/20230678

*3.5. Comparison between three ways of attacking*

As it can be seen from Figure7, after removing about 10% of the key nodes, the results show that no matter which criteria the targeted attack uses, it doesn't have a significant impact on the power-law distribution of the network.

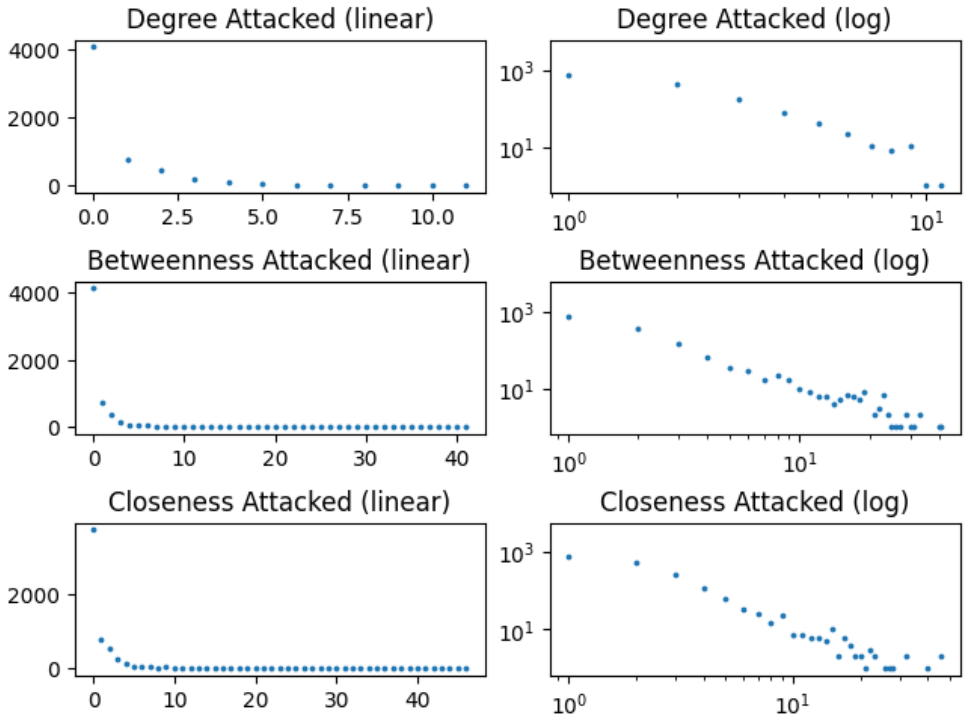


**Figure 7.** Results under three different attacks.

## 4. Conclusion

For the global air transport network, the networks with degree distribution satisfying power-law distribution are more resilient to random errors and less resilient to the targeted attacks which will destroy connectivity and small-world property easily and cause an increase in diameter and average path length. Meanwhile, the targeted attacks would not affect degree distribution of nodes significantly. The small-world network has low clustering coefficient, and it is harder to make it lower.

World air transportation involves transnational and transoceanic flights, so airports in the same continent or country tend to communicate frequently with each other, while the number of transnational and transoceanic routes is limited, resulting in the phenomenon that airports are clustered into communities according to geographical relationships. The targeted attacks on key nodes will cut off the connections between communities as well as the connections within communities, and the attacks against degrees tend to cut off the connections within communities. However, the targeted attacks against betweenness centrality are more likely to cut the ties between communities and create a large number of components and fragments in a short time.

## References

[1] Yuan Luo; Ai Zhu Ren.Integration and Conversion of 3D Models from 3DS to OpenFlight[J].Applied Mechanics and Materials, Volume 1439, Issue 88-89. 2011. PP 559-563

[2] Lei Song, Li-hao Yuan,Zhi-hui Dong College of Shipbuilding Engineering Harbin Engineering University Harbin,China. Research on Ship Virtual Reality Model Transformation Method Based on OSG [C]//.Proceedings of 2012 IEEE International Conference on Computer

DOI: 10.54254/2755-2721/6/20230678

Science and Automation Engineering（CSAE 2012）VOL01.Institute of Electrical and Electronics Engineers, 2012: 540-544.

[3] Gephi. (2017). About Gephi. [Online]. Available: https://gephi.org/about/

[4] Murphy Fiona A; Johnston Helinor J; Dekkers Susan; Bleeker Eric A J; Oomen Agnes G; Fernandes Teresa F; Rasmussen Kirsten; Jantunen Paula; Rauscher Hubert; Hunt Neil; di Cristos Luisana; Braakhuis Hedwig M; Haase Andrea; Hristozov Danail; Wohlleben Wendel; Sabella Stefania; Stone Vicki. How to formulate hypotheses and IATA to support grouping and read-across of nanoforms. [J] ALTEX: Alternatives to Animal Experimentation2022.

[5] Bruns A. How Long Is A Tweet? Mapping Dynamic Conversation Networks On Twitter Using Gawk And Gephi [J]. Info Commun Soc, 2012, 15(9):1323-51.

[6] Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks; proceedings of the Proceedings of the International AAAI Conference on Web and Social Media, F, 2009 [C].

[7] D. Jun. Comparative study of the social network analysis tools: Ucinet and Gephi. [Online]. Available: https://en.cnki.com.cn/Article_en/CJFDTotal-QBLL201408027.htm

[8] Lt Col Rahul Dutt Sharma：Python Tools for Big Data Analysis，[J]Journal of Research in Science and and Engineering, Volume 3, Issue 1. 2021.

[9] Cao Shengjia; Zeng Yunhan; Yang Shangru; Cao Songlin: Research on Python Data Visualization Technology, [J]Journal of Physics: Conference SeriesVolume 1757, Issue 1. 2021. PP 012122-

[10] Papadopoulou Olga; Makedas Themistoklis; Apostolidis Lazaros; Poldi Francesco; Papadopoulos Symeon; Kompatsiaris Ioannis: MeVer NetworkX: Network Analysis and Visualization for Tracing Disinformation, [J]Future InternetVolume 14, Issue 5. 2022. PP 147-147