VISUALISATION ASSIGNMENT 1

Nguyen Thi Phuong Thao

19332749

Western Sydney University

Visualisation

April 4,2018

Dr. Quang Vinh Nguyen

Introduction

In recent years, as the emerging of information and technology is remarkable, data becomes to play a crucial role in different corners of human life (Hajirahimova, M. S., & Aliyeva, A. S. , 2017). With an enormous amount of data, there is a need of developing tools that indicate data in a more interesting and friendly-interaction way, which is known as visualization. Data visualization is defined as a graphical display of abstract information that leverages the powers of human perception to discover and understand important hidden stories in data (Martinez, R., Ordunez, P., Soliz, P. N., & Ballesteros, M. F. , 2016). As a technical terms, visualization process can be splitted into two semantic contexts: viewing and seeing (Chen, M., & Floridi, L. ,2013). “Viewing” prefer to a preattentive process which is really important for cognition. In the literature on preattentive process in Visualization, the Gestalt Law has been discussed to lay a foundation for visual designs (Courtright, A. ,2002). It provides understanding about pattern perception of human and would be used to apply for author’s visualization work in this paper. Unlike “viewing”, “seeing” require more thought processes and cognitive experiences to interpreting the stories of received information visualization. When the visualization was first displayed, readers are able to “view” the data at a glance and then they will do more cognition process as “seeing” to understand data deeply. According to visualization process understanding, there are many tools and technique have been introduced for visualizing data such as Hierarchical structures, graphs and networks.

This paper will have a brief description as well as its advantages/disadvantages of visualization techniques, namely Spring Embedding algorithm, Circular drawing technique and Degree of Interest tree (DOI tree). Using both Cytoscape software and DOI tree java application for Western Sydney University’s Courses dataset, the author will make an analysis of Western Sydney University's courses base on the visualizations. In addition, there will be a discussion on other aspects related to the work and literature reviews in this paper.

Visualizing and Analyzing

Cytoscape

Cytoscape (Cytoscape 2018) was firstly introduced as a software for visualizing biological research. Its ability is to visualize mocular interaction networks and integrate annotations, gene expression profile and other state data with these networks. However, by time this software was developed and widely used for visualizing and analyzing complex network.

There are roughly four version of Cytoscape which are designed suitable for different target users such as casual users, app developers and web application developers. As the author is just a casual user, this paper will harness Cytoscape v3 for analyzing and visualizing the data.

Although Cytoscape is assessed as not really prefer to beginners, its friendly interaction workspace will help users who eager to explore gain experience quickly (Elisa E. Beshero-Bondar, 2018). Cytoscape requires users to define sources nodes, target nodes and interaction from the input data to generate the graph. Thus, Cytoscape is more ideal for those who want to graph a network or hierarchical treemap. There are many options for users to customize their visualization. They can change color, size, shape of edges, nodes and even labels. Therefore, it allows users to have an aesthetic visualization. Moreover, Cytoscape is equipped with many useful layout algorithm for making data better visualized and analyzed. Besides, the table panel integrated below the work screen showing data detail foster users in reading, keep tracking and analyzing any notes/edges selected.

However, Cytoscape has a disadvantage of processing time. It might slow down users' computers when processing any demand as well as take time to generate the result. Especially when working with big data, Cytoscape will require a lot memory and power. Besides, the original graph generated by Cytoscape looks quite messy. Thus, users need to do some manual styling step to redesign a better aesthetic visualization which is considered as time consuming.

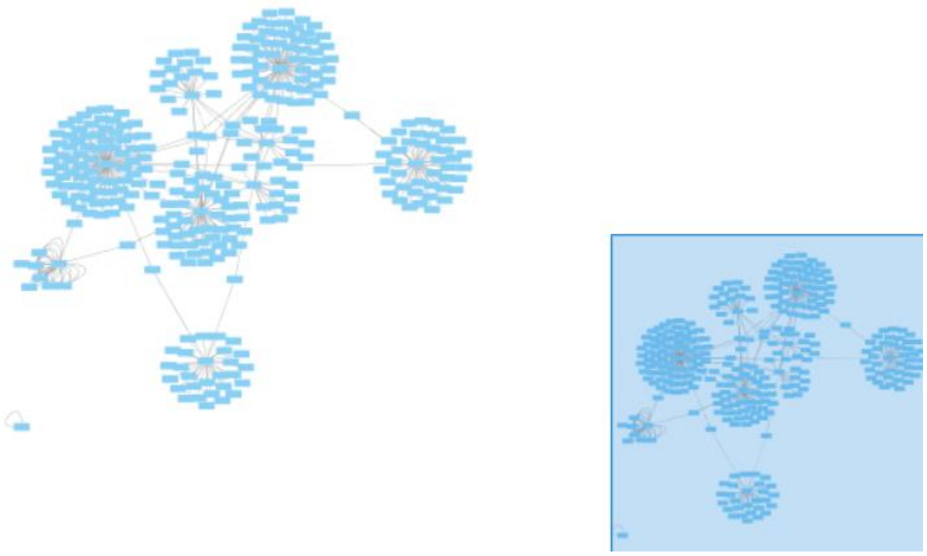

Figure 1: Initial Cytoscape visualization

The initial Cytoscape visualisation from Western Sydney University's courses dataset showing all connection between each school of Western Sydney University and links of courses offered by each school. At a first glance, we can see that there are five schools offer overweight number of courses than others and there is a single node that are disconnected to a big network. In order to have a closer look, the author use zoom-in tool to see the label clearly.

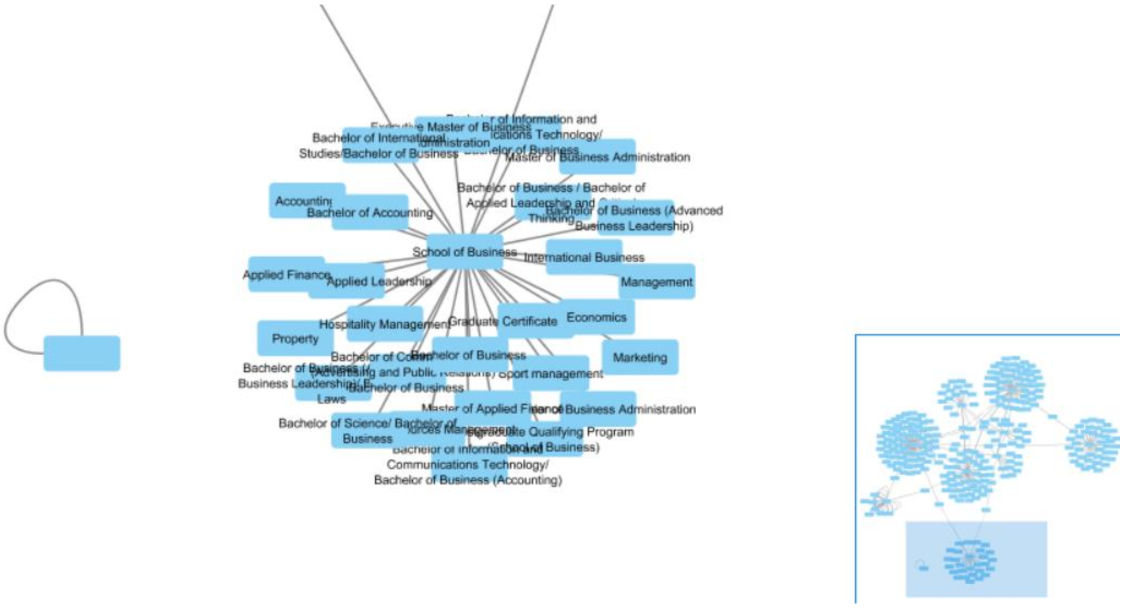

Figure 2: Zoom-in Initial Cytoscape Visualisation

With a closer look, the author find the isolate node is a missing out value which is irrelative to the dataset. Therefore, the author delete this node to clean the data. After having the data cleaned, the author use Spring Embedding Algorithm Method and circular drawing Method to cluster and graph data via Spring Embedded Layout and Circular Layout in Cytoscape.

Spring Embedding Algorithm method

Battista et al., 1994 defined spring embedding algorithm as a heuristic stimulating physical bodies, where nodes are seen as “rings” and edges are seen as “springs” providing forces between “rings”. It was successfully designed especially for general undirected graph, basing around the idea to minimize the total spring forces of the system. In this algorithm, “springs” are being calculated by partial differential equations of energy function that leads to attractive or repulsive forces which move “rings” accordingly (Kamada T, Kawai S.,1989).

Spring Embedding Algorithm have been widely used and developed in visualization. In 2007, Lin and Yu governed Spring Embedding Algorithms to visualize Phylogenetic Tree.

Their work reveals the useful of Spring Embedding Algorithm method in re-structuring the initial graph to be more aesthetic with nodes distributing evenly, lengths of edge being the same, etc. Jennifer Golbeck & Paul Mutton applys this method for graphing a Semantic Web in 2005 and analyzing data. Their work proved the ability of easier understanding of Web using this structure than that of using text or other visual displays.

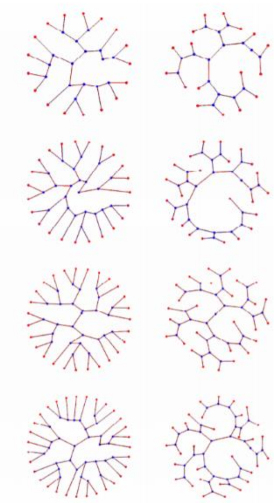

Figure 3 Lin & Yu demonstrates Spring embedding method for Phylogenetic tree model

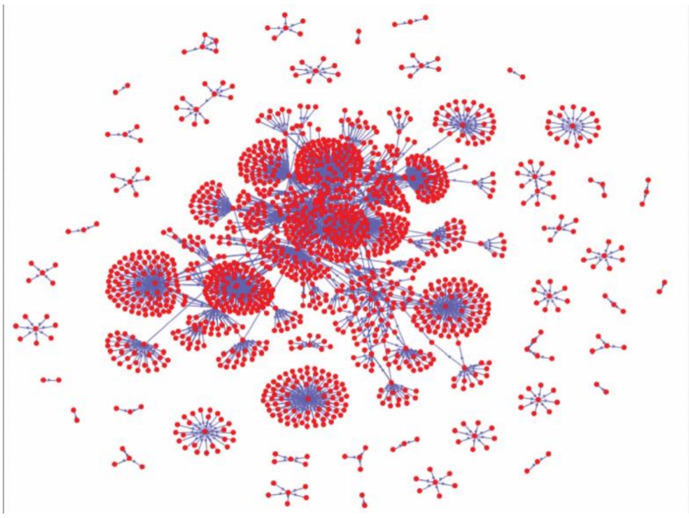

Figure 4 Jennifer Golbeck & Paul Mutton Semantic Web-based social network visualization using spring embedding method

This visualization method appears to work reasonably well with small and medium sized networks (example 50-1000 nodes), but not really work with larger networks as it can result in a "hairball network layout" where the connectivity structure are dense or too much clustered (Tuikkala, J., Vähämaa, H., Salmela, P., Nevalainen, O. S., & Aittokallio, T., 2012).

Graph analysing

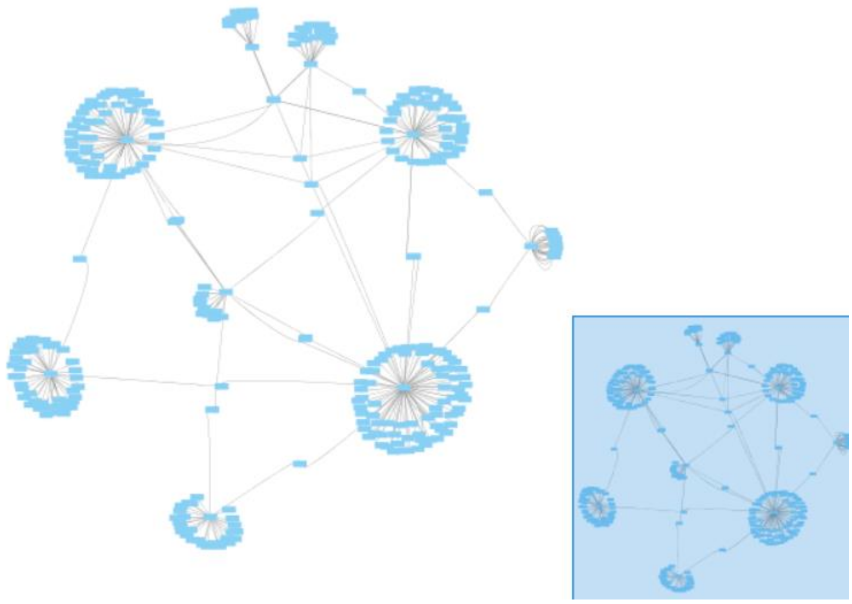

Figure 5 WSU's courses visualization applying Spring Embedding Layout

After applying spring embedding algorithm layout to the Western Sydney University's courses dataset, the visualization become easier to be observed:

- Each schools and courses which belonging to that school are clustered into a group. There are 9 different groups represent to 9 different schools of Western Sydney University. Each school offer different number of courses according to the visual group size.
- There are some courses offered by different schools (Courses linked with more than 1 school).
- All schools have connection between them.

- Some schools have high relation to each other (for example: School of Law have connection with School of Social Science and Psychology, School of Humanity and Communication art, School of Business, School of Science and Health).

In order to distinguish different programs offered by one school demonstrated in the visualization, the author apply spring embedding method again for each school to group and classify courses of undergraduate program, post graduate program and higher degree program. After that, the author makes use of Proximity and Similarity in Gestalt Law to style the visualization, meaning that similar elements get group together with a close distance and the same shape as well as color.

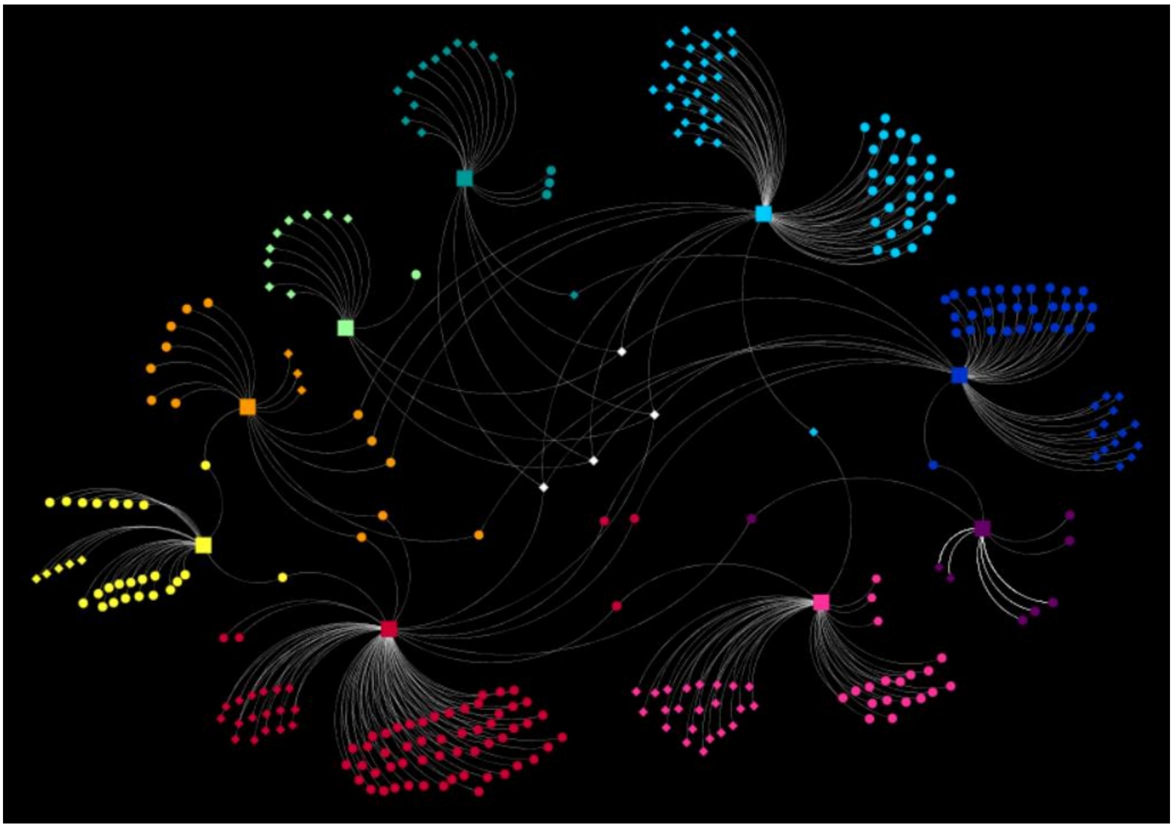

Figure 6 Western Sydney University courses Visualisation

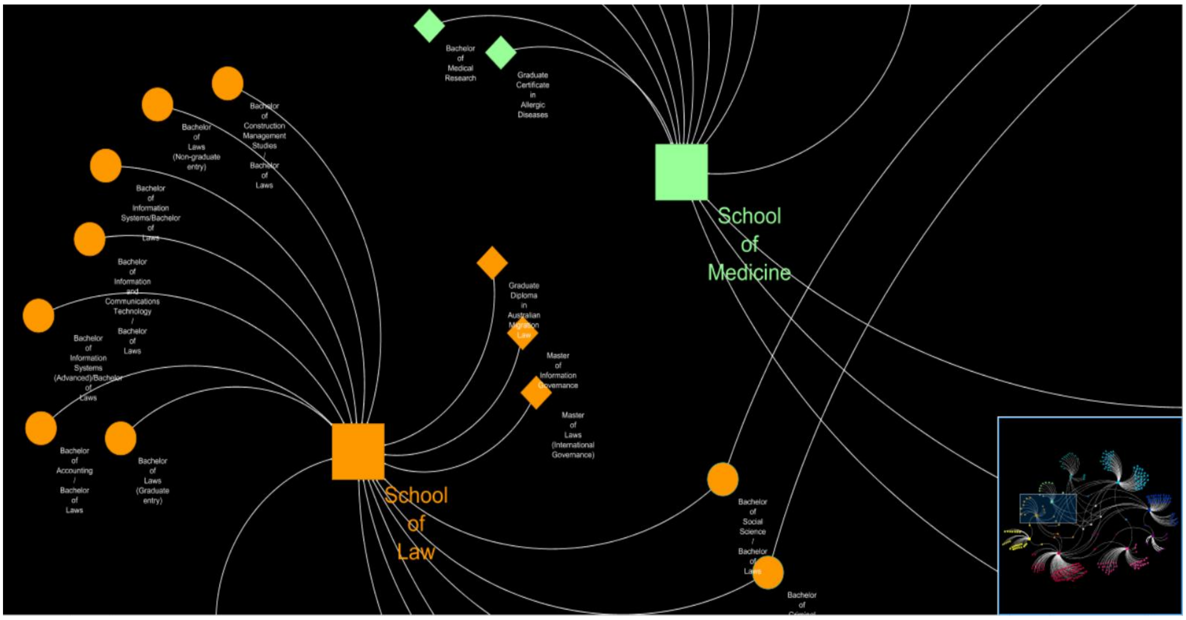

Figure 7 Zoom-in part of WSU's courses Visualisation

The figure 6 and 7 give readers a perceptual ideas that:

- Schools related to Medicare and Health industry such as School of Medicine, School of Nursing and Midwifery have more postgraduate courses than undergraduate courses.
- All schools provide undergraduate and postgraduate courses. Haft of them offer higher degree courses.
- The number of courses of School of Humanity and Communication Arts, School of Social Science, School of Science and Health, School of Computing, Engineering and Mathematics are remarkably overweight than that of remaining schools. Therefore, probability of a student who from one of these four school is supposed to be higher.

According to the graph, readers are able to have some assumption of the popularity of Western Sydney University according to the background academic.

Circular drawing technique

Circular drawing technique is used to generate a graph that is clustered into different groups where every nodes of each groups are organized to form a circle with each circle has different circumference and each edge is a straight line. It keeps the number of crossing edge low to demonstrate the graph efficiently (Janet M. Six & Ioannis G. Tollis,?). It is believed that this technique works well in reality although it is really complex to solve the number of edges problem in a large data visualization. This technique shows the connection between vertices and organized to utilize space which is ideal for big data that cannot be done well with linear layout (Krzywinski, M. et al, 2004-2016).

Taking to previous work, Kar, Madden, and Gilbert (11/1988) employed Heuristic layout algorithms, algorithms used for clustering and node placing as well as some extra calculation method to create circular visualization for networks management. Using circular drawing technique, they achieved their target of generating aesthetic visualization that qualify two conditions: (1) sites and notes are distributed as evenly as possible; (2) the number of crossing edges is minimized.

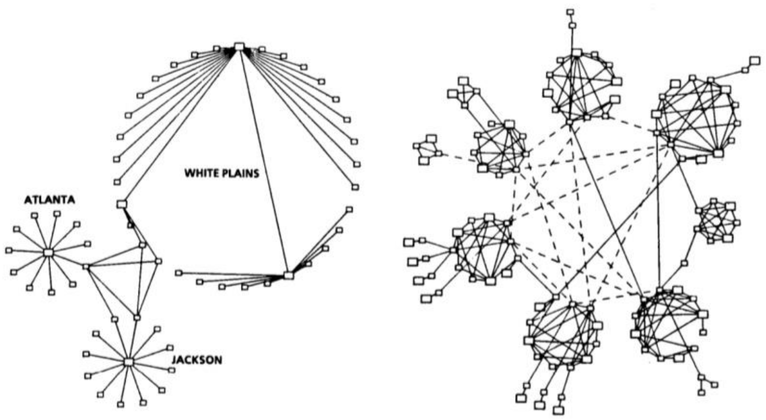

Figure 8 Kar, Madden & Gilbert complex network visualization applying circular drawing methods

While Kar, Madden, and Gilbert succeeded in minimizing the number of crossing edges, Tollis and Xia contributed to circular drawing technique with several linear time algorithms and tree structure including (1) outside drawing: placing two rings that share the same node side by side, (2) inside drawing: placing children ring inside its parent ring, and (3) mixed drawing: separating graph using inside and outside drawing.

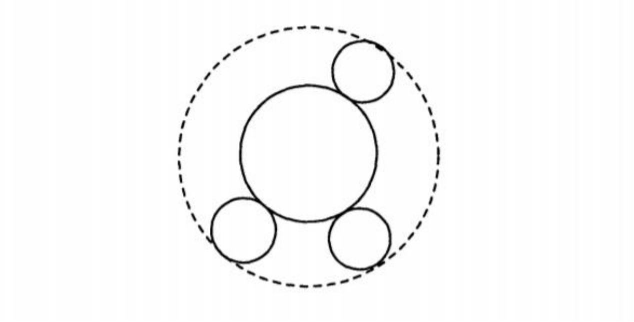

Figure 9 Tollis & Xia inside structure drawing

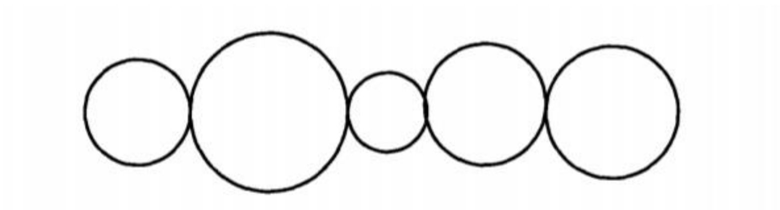

Figure 10 Tollis & Xia outside structure drawing

Graph analysing

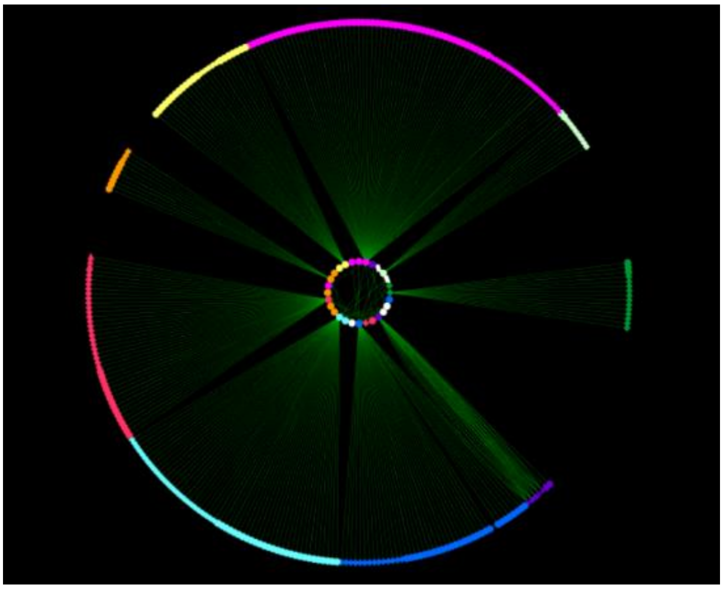

Figure 11 WSU's courses initial visualization using Circular Layout

The figure 11 is an initial visualization generated with circular layout algorithms. At a first glance, we can see that schools and course are partitioned into two clusters and there are connection between parent nodes (schools) and children nodes (courses). We can also easy to compare the number of courses among schools via weight of edge sets. To be specific, because the range of edges connected from School of Humanity and Communication Art to its courses is wider than that of School of Law, there must be more courses provided by School of Humanity and Communication Art than School of Law. However, courses provided by multi-schools are misinterpreted to be classed in the same group with schools, so the author do some manual step to distinguish them. Next, each group of courses provided by each schools is partitioned again into clusters according to its course program, using group attributes layout which is similar to Attribute circle layout. The result shown in figure 12.

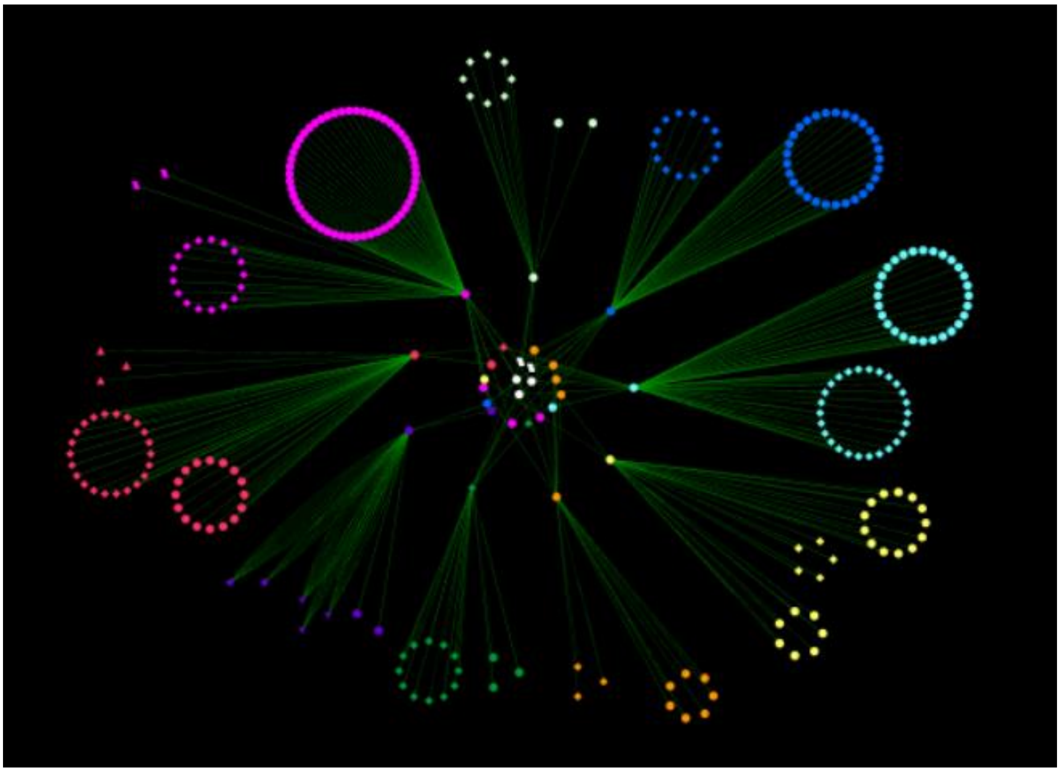

Figure 12 WSU's Courses Visualzation applying Circular layout for each cluster

The figure 12 illustrates information such as:

- There are nine different schools in Western Sydney University.
- Each school provides at least two course program (undergraduate and postgraduate).
- School of Humanity and Communication Art provides the biggest number of undergraduate courses.
- School of Education is the smallest school with the lowest number of courses and also have different style of dividing course program from the remaining.
- 4 out of 9 schools occupy the majority of Western Sydney University courses, namely School of Humanity and Communication Art, School of Science and Health, School of Social Science and Psychology, School of Computing, Engineering and Mathematics. According to that, we can make the assumption about which fields Western Sydney University focus on and have advantages.

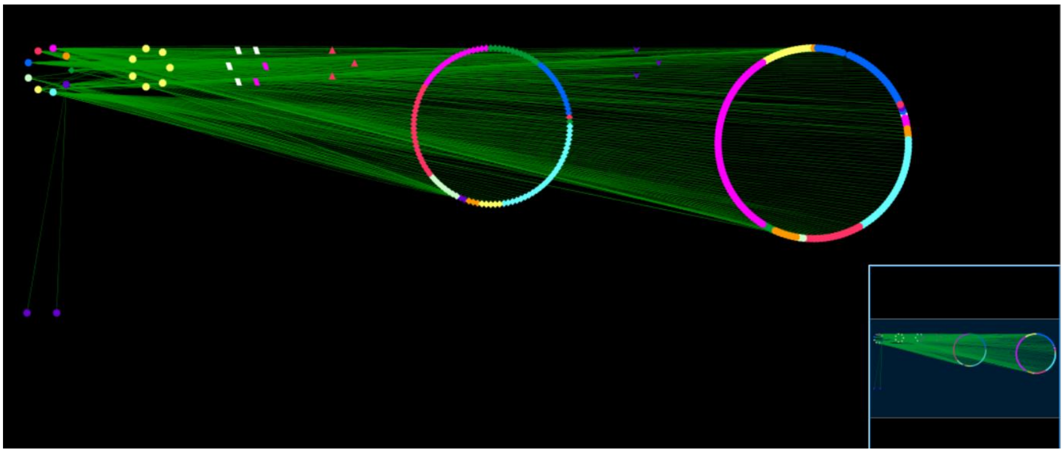

Figure 13 WSU's courses circular visualisation clustering according to courses program

When applying group attributes circle layout to whole nodes, we read further information from the data in different approach:

- According to the circle size, the biggest circle represents that Western Sydney University provides most is undergraduate courses, following is postgraduates courses.
- There are 7 types of course programs: undergraduate, postgraduate, higher degree research, double degree, Honours courses, Undergraduate teacher and teacher education.
- Only School of Business provides double degree. Therefore, if students want to have a double degree they should register to School of Business. In other words, any students who graduated from Western Sydney University with double degree, they must be a students from School of Business.
- Only School of Computing, Engineering and Mathematics provide honour courses program.

Degree of Interest tree (DOI tree)

Degree of Interest tree (DOI tree) is a technique to visualize hierarchical data using connection approach (Heer, Jeffrey and Stuart K.Card,2004). It is assessed to overcome the space displacing problem of general node-link diagram (Guanqun Wang, Tsuneo Nakanishi and Akira Fukuda, 2016). The concept of degree of Interest is firstly introduced by Furnas (1986), basing around the idea that every different observers has their different interests with items, so that the importance of items is different. He made use of the differences in interest to weight the importance of each item that indicates in its degree of interest and generate the graph that allows users only view their interest information.

After that, thanks to authors such as Card & Nation (2002) who encouraged to employ focus+context displays, Heer&Card (2004) who optimized and scalabled DOI tree by adding advanced search browser and filter, DOI tree time by time improved much compared to the original one. It now has interaction and navigation to allow users view multi-nodes, branches at the same time. It also provides smooth fade-in and fade-out animations among transition. Moreover, DOI tree has friendly interface so beginners can access and build visualization rapidly.

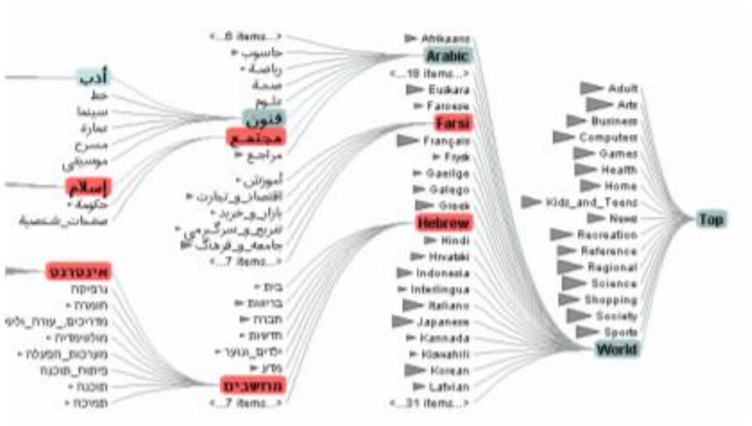

Figure 14 DOITree visualization of the Open Directory Project containing over 600,000 nodes by Heer&Card

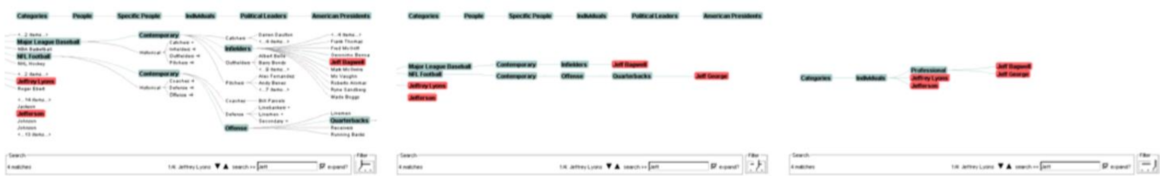

Figure 15 Search & Filtering DOItree Project by Heer&Card,creating more compact views by reducing contexts.

However, there is still a room for DOI tree to improve. To be specific, when the amount of data exposed and processed in DOI browser is large, it would come to result in distracting users from finding solution to their query (Raluca Budiu, Peter Pirolli, Michael Fletwood, 2016). Sometimes, cramming a lot information on the screen does not help improve performance. Although Van Ham, F., & Perer, A. (2009) suggested that DOI tree can be used for managing complex and large data by adapting additional term in DOI function to extract and expand sub graph, there is still a problem that it could lead to a bias interface (Card &Nation, 2002), meaning that there are such non-interest nodes that could be potential interest nodes but not displayed. However, this technique should be considered when visualizing big data.

Graph analyzing

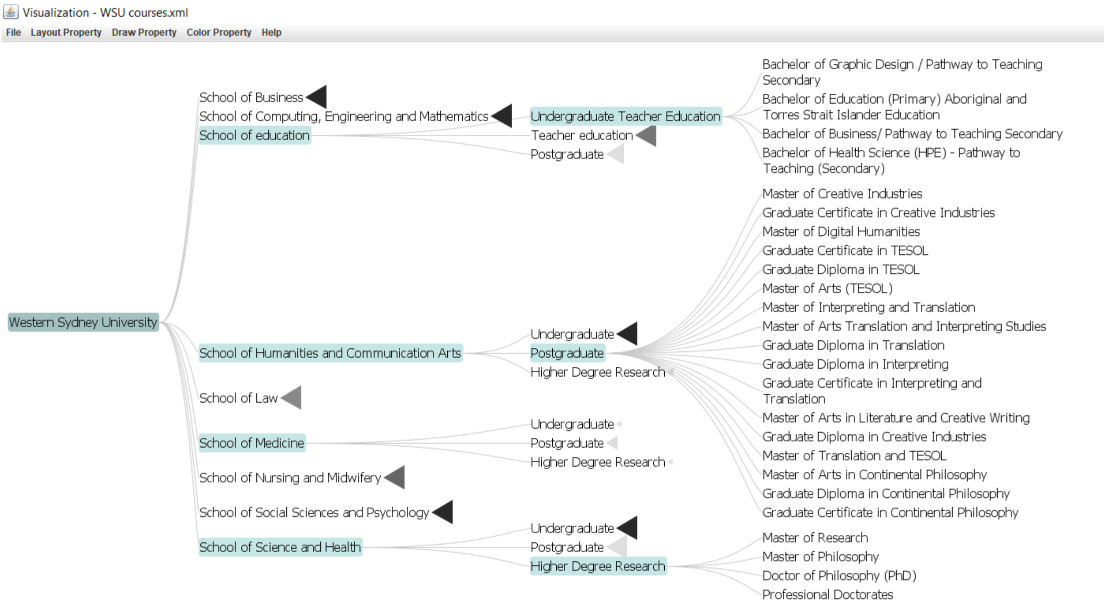

Figure 16: DOI tree visualization

The visualization generated by DOI tree make information easily read because of the label details of each nodes. Parent and children hierarchical structure is performed in a universal accepted way. Thus, it can be considered as good at listing. However, unlike previous visualization, DOI tree visualization neither show the connection between schools nor the edges weight to compare the number of courses offered by different schools within the university. Via the visualization, we can observe the content clearly despite of the contents’ length. Readers are able to choose viewing any fields they interest in. Moreover, they can also view more than one course programs they interest in at one time with a quick processing. This DOI tree visualization from WSU’s courses dataset is ideal for upcoming students who want to study in WSU but have not decided which schools/ courses they want to study yet. This visualization will allow them to have a quick draft of diplomas they are going to get after graduation in fields they are interested in and then figuring out which one they will register to achieve their career dream.

Discussion

Three visualization methods used above work well in visualizing parent and children dataset. They are all using connection approach to demonstrate the connection between nodes. While spring embedding algorithm calculates the forces between nodes to repel and attract vertices in order to keep the total energy of the system, circular drawing technique aims to keep the minimum crossing edges. Thus, when comparing the readability between these two methods, graph using circular drawing technique seems to be more advantage with the minimum numbers of crossing edges, which qualify aesthetic rules (mentioned in Lin &Yu,2007, p 88). Different with Spring Embedding Method and Circular Drawing Method, DOI tree shows contents as a list so it not suitable for networks such as social network. Although DOI trees does not view whole nodes at the same time like other methods, it allows users to view only which users interest in and hide the non-interested one. This has both pros and cons, as users can be overlooked the promising interested nodes. Taking into visualize big data, DOI tree and Circular Drawing Algorithm is supposed to be more applicable as they utilize space well.

Conclusion

In conclusion, this paper using three visualization techniques namely Spring Embedding Algorithm, Circular Drawing technique and DOI tree to visualize Western Sydney University’s course dataset. The result is that three techniques are suitable to visualize connections between nodes of a parent-children dataset. They help to group the data into clusters for comparing and analyzing. Each method used generated different layouts for the dataset, basing on each layouts, there are slightly different information revealed. Therefore, when using the visualization technique to visualize data, beside the type of dataset, we should also consider purposes of authors and/or the aim of readers, meaning that we should define which messages we want to convey via the visualization and the target readers who will use our visualization when choosing appropriate visualization techniques/tools.

References:

Battista, Eades, Tamassia, & Tollis. (1994). Algorithms for drawing graphs: An annotated bibliography. Computational Geometry: Theory and Applications, 4(5), 235-282. Retrieved from https://www-sciencedirect-com.ezproxy.uws.edu.au/science/article/pii/092577219400014X

Card, S., & Nation, D. (2002). Degree-of-interest trees: A component of an attention-reactive user interface. Proceedings of the Working Conference on Advanced Visual Interfaces, 231-245. Retrieved from http://delivery.acm.org.ezproxy.uws.edu.au/10.1145/1560000/1556300/p231-card.pdf?ip=137.154.19.149&id=1556300&acc=ACTIVE%20SERVICE&key=65D80644F295BC0D%2E638C67C1DB76B401%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35&__acm__=1524289946_e59dec45f8de45d28d964706b71747c4

Courtright, A. (2002, 03). Applying the basics: Gestalt rules of design. Inside Microsoft PowerPoint, 9, 13-14. Retrieved from https://search.proquest.com/docview/191079960?accountid=63189

Cytoscape, http://www.cytoscape.org/ (accessed date 20/04/2018)

Chen, M., & Floridi, L. (2013). An analysis of information visualisation. Synthese, 190(16), 3421-3438. doi:http://dx.doi.org/10.1007/s11229-012-0183-y

Elisa E. Beshero-Bondar. (01/02/2018). http://ebeshero.github.io/thalaba/cytosc.html (accessed date 21/04/2018)

G. Kar, B. Madden, R. Gilbert. (1988). Heuristic layout algorithms for network presentation services. Retrieved from https://ieeexplore-ieee-org.ezproxy.uws.edu.au/stamp/stamp.jsp?tp=&arnumber=10027

G.W. Furnas. (1986). Generalized fisheye views, Proceedings Human Factors in Computing Systems CHI 86, pp. 16-23. Retrieved from DOI: http://dx.doi.org/10.1145/22339.22342

Guanqun Wang, Tsuneo Nakanishi and Akira Fukuda. (2016). 2-D Layout for Tree Visualization: a survey. Retrieved from https://www.researchgate.net/publication/301672502_2D_Layout_for_Tree_Visualization_a_survey.

Hajirahimova, M. S., & Aliyeva, A. S. (2017). About big data measurement methodologies and indicators. International Journal of Modern Education and Computer Science, 9(10), 1. doi:http://dx.doi.org/10.5815/ijmecs.2017.10.01

Heer, J., & Card, S. (2004). DOITrees revisited: Scalable, space-constrained visualization of hierarchical data. Proceedings of the Working Conference on Advanced Visual Interfaces, 421-424. Retrieved from http://delivery.acm.org.ezproxy.uws.edu.au/10.1145/990000/989941/p421-heer.pdf?ip=137.154.19.149&id=989941&acc=ACTIVE%20SERVICE&key=65D80644F295BC0D%2E638C67C1DB76B401%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35&__acm__=1524290695_be3d734705543c357104ba27d42af1de

Heer, Jeffrey, and Stuart K. Card. (2004) . DOITrees revisited: scalable, space-constrained visualization of hierarchical data. Proceedings of the working conference on Advanced visual interfaces.

I.G. Tollis, C. Xia. (1997). Drawing telecommunication networks. Retrieved from https://link.springer.com/content/pdf/10.1007/3-540-58950-3_372.pdf

Janet M. Six & Ioannis G. Tollis. (?). Circular Drawing Algorithms. Retrieved from https://cs.brown.edu/~rt/gdhandbook/chapters/circular.pdf

Jennifer Golbeck & Paul Mutton. (2005). Spring-Embedded Graphs for Semantic Visualization. Retrieved from https://pdfs.semanticscholar.org/0c26/d3484ae650c321a864de3775dcb30a775279.pdf

Kamada T, Kawai S: An algorithm for drawing general undirected graphs. Inf Process Lett 1989, 31:7-15. Retrieved from https://pdfs.semanticscholar.org/b8d3/bca50ccc573c5cb99f7d201e8acce6618f04.pdf

Krzywinski, M. et al. (2004-2016). Circos: An information aesthetic for comparative genomics. Retrieved from http://circos.ca/intro/circular_approach/

Lin & Yu. (2007). Visualizing Phylogenetic Trees by Spring-Embedder Models. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.108.5680&rep=rep1&type=pdf

Martinez, R., Ordunez, P., Soliz, P. N., & Ballesteros, M. F. (2016). Data visualisation in surveillance for injury prevention and control: Conceptual bases and case studies. Injury Prevention, 22doi:http://dx.doi.org/10.1136/injuryprev-2015-041812

Tuikkala, J., Vähämaa, H., Salmela, P., Nevalainen, O. S., & Aittokallio, T. (2012). A multilevel layout algorithm for visualizing physical and genetic interaction networks, with emphasis on their modular organization. Biodata Mining, 5doi:http://dx.doi.org/10.1186/1756-0381-5-2

Tutte, W. T. (1963), "How to draw a graph", Proceedings of the London Mathematical Society, 13 (52): 743–768, doi:10.1112/plms/s3-13.1.743.

Van Ham, F., & Perer, A. (2009). "Search, Show Context, Expand on Demand": Supporting Large Graph Exploration with Degree-of-Interest. Ieee Transactions On Visualization And Computer Graphics, 15(6), 953-960. Retrieved from: https://ieeexplore-ieee-org.ezproxy.uws.edu.au/document/5290699/authors