ECS 272 Information Visualization Project Report

Taiwan Stock Market Analysis and Visualization

Name: Shu-Wei Hsu

Student ID: 998569417

Introduction

Regarding to the significant amount of companies and value/volume of daily trading of stock market, it is always challenging to do the data analysis and further visualization to show the transaction network between companies and brokers efficiently. The daily exchanges of Taiwan stock market is made by centralized computer. The normal trading sessions from 9:00am to 1:30pm and post-market sessions from 02:00pm to 02:30pm on all days of the week except Saturdays, Sundays and holidays declared by the Exchange in advance. The fundamental exchange unit is one sheet, which comprises 1000 shares. There are totally 758 listed and 564 over-the-counter companies in Taiwan stock market. The average daily trading value is around 10 to 50 billion US dollars with up to average million sheets (1000 shares) volume of daily transactions.

Data Analysis

A snippet of a sample dataset of the daily trading records regarding to a single security with respective brokers in Taiwan stock market across ten years through 2001 to 2011 is as shown below (Figure 1). The daily trading records of each security with respect to each broker are labeled with the captions:

- Stkid: stock id
- Stknm: stock name
- Scid: broker id
- Scnm: broker name
- Buypic: the volume for purchasing the stock
- Sellpic: the volume for selling the stock
- Buymoney: the total amount of money for purchasing the stock
- Sellmoney: the total amount of money for selling the stock
- Buyrow: remarks for purchasing the stock
- Sellrow: remarks for selling the stock

Here the Buyrow and Sellrow indicate the volume of transactions of purchasing and selling via the broker, respectively. Furthermore, the instantaneous value of purchasing is defined by the total amount of money for purchasing the stock divided by the volume for purchasing the stock.

stkid	stknm	scid	scnm	buypic	sellpic	buymoney	sellmoney	buyrow	sellrow
2382	廣達	1020	合 庫	29250	25000	1714500	1459500	19	13
2382	廣達	1030	土 銀	21000	2000	1229500	117200	15	1
2382	廣達	1040	臺 銀	43000	4000	2526900	234300	29	4
2382	廣達	1090	工 銀	1000	5000	58100	292500	1	5
2382	廣達	1110	企 銀	23000	2000	1349100	118600	19	2
2382	廣達	1160	日 盛	122850	56000	7221900	3293900	92	40
2382	廣達	1230	彰 銀	10000	0	582000	0	1	0
2382	廣達	1260	宏 遠	36000	7000	2111400	410200	19	4
2382	廣達	1360	港麥格理	1071000	729000	62618300	43062000	422	348
2382	廣達	1380	港商里昂	217000	226000	12586000	13256900	32	114
2382	廣達	1440	美林	141000	159000	8349100	9316100	64	76
2382	廣達	1470	台灣摩根	59828	573000	3505024	33835400	41	296
2382	廣達	1480	美商高盛	30000	597000	1741500	34902800	3	250

Figure 1. Trading record of 2382 Quanta Computer (廣達 in Chinese)

Taking the a specifically daily report from Taiwan Stock Exchange Corp. (TWSE) in Figure 2 of Daily Trading Value/Volume of Security 2382 on March, 2011 for instance: the sum of buypic (the volume for purchasing the stock) is equal to the trade volume; the sum of each buymoney (the total amount of money for purchasing the stock) is equal to the trade value; and the sum of buyrow (transaction) is equal to the transaction. The change of value is simply calculated by difference of the price and previous price of the day before the specific day.

2011/03 2382								(NT$,share)
Date	Trade Volume	Trade Value	Opening Price	Highest Price	Lowest Price	Closing Price	Change	Transaction
2011/03/01	7,675,135	460,905,853	58.90	61.00	58.80	60.80	+2.70	3,442
2011/03/02	7,294,495	428,098,110	60.00	60.00	58.00	58.00	-2.80	3,166
2011/03/03	8,818,906	524,300,260	59.30	60.20	58.50	60.10	+2.10	3,558
2011/03/04	7,054,955	434,166,238	60.80	62.70	60.50	61.00	+0.90	3,194
2011/03/07	1,840,954	111,440,406	60.60	61.00	60.10	60.30	-0.70	1,004
2011/03/08	3,588,343	214,940,436	59.80	60.30	59.50	60.00	-0.30	1,777
2011/03/09	4,715,852	277,400,013	60.00	60.00	58.30	58.60	-1.40	2,046
2011/03/10	6,646,912	380,513,484	58.00	58.00	56.90	57.00	-1.60	2,483
2011/03/11	12,847,882	710,588,751	56.00	56.30	54.70	54.70	-2.30	4,681
2011/03/14	6,739,771	359,363,314	53.10	54.10	52.70	53.60	-1.10	2,892
2011/03/15	11,629,913	599,450,769	52.80	54.00	49.90	51.90	-1.70	4,092
2011/03/16	7,201,164	381,686,051	53.50	53.50	52.20	52.70	+0.80	2,905
2011/03/17	6,232,935	322,165,768	50.50	52.90	50.50	52.70	0.00	2,590
2011/03/18	6,076,568	321,562,406	52.50	53.70	52.10	53.20	+0.50	2,359
2011/03/21	3,205,747	171,822,227	53.50	54.20	52.90	53.60	+0.40	1,565
2011/03/22	3,525,317	187,062,255	53.80	53.90	52.80	52.90	-0.70	1,733
2011/03/23	4,376,826	231,021,378	53.60	53.60	52.20	53.00	+0.10	2,154
2011/03/24	7,687,606	406,361,418	53.20	53.80	52.50	52.60	-0.40	3,169
2011/03/25	8,970,887	485,829,166	53.10	55.20	53.00	54.30	+1.70	3,763
2011/03/28	7,142,772	379,919,144	54.30	54.30	52.30	53.70	-0.60	2,968
2011/03/29	5,581,551	300,819,056	52.80	54.70	52.70	54.30	+0.60	2,537
2011/03/30	11,707,099	659,965,769	54.30	57.60	54.30	56.40	+2.10	5,103
2011/03/31	10,771,397	595,376,778	56.40	56.40	54.90	55.50	-0.90	2,925

Figure 2. Daily Trading Value/Volume of Security 2382 on March, 2011

Furthermore, the correlation between stock brokers can be represented by the p-value, which is calculated by daily stock transaction data for a certain time using the statistical model Value at Risk (VAR). For the p-value matrix, it shows the significance of the correlation among the brokers. The p-value can be defined with some cut points:

- If the p-value < 0.01, it shows the correlation between the two brokers is very significant.
- If the 0.01<p-value < 0.05, it shows the correlation between the two brokers is significant.
- If the 0.05<p-value < 0.1, it shows the correlation between the two brokers is less significant.
- If the p-value > 0.1, it shows the correlation between the two brokers is not significant.

Related Works

In modern visualization of stock values and volumes, line charts and bar charts are still commonly used to represent the path of value of shares and volume of transactions. As in Figure 3, the stock chart and be scaled to days, weeks, months, and years, or ranged to any specific interval to display useful information for users through interactions. In Figure 4, MetaView group have designed a 3D graphic visualization to help people quickly see patterns and spot market trends and discover cause and effect in stock market. The 3D module displays the variation of different share value of companies with timelines in a more effective way. However, it lacks of data of transaction volumes and correlations between companies in the same industry, and trading between stock brokers.

Figure 3. A sample of stock chart of 2382 Quanta Computer Inc.

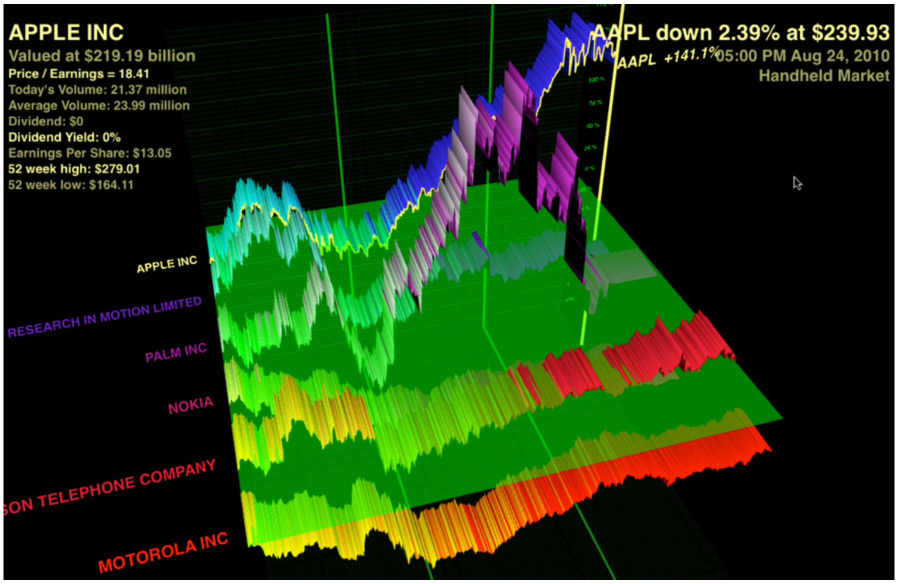

Figure 4. MetaView 3D stock market visualization

Grant et al. [3] have done an amazing work of visualizing US stock market with universe like approach as showed in Figure 5. Companies can be selected or searched to show the financial details such as value of shares or volume of transactions. Further classification by sectors/industries is also applied to reduce size and complexity of information visualization. However, no relationships and transactions of companies and brokers are illustrated, which make the visualization fabulous but randomly distributed and lack of usability.

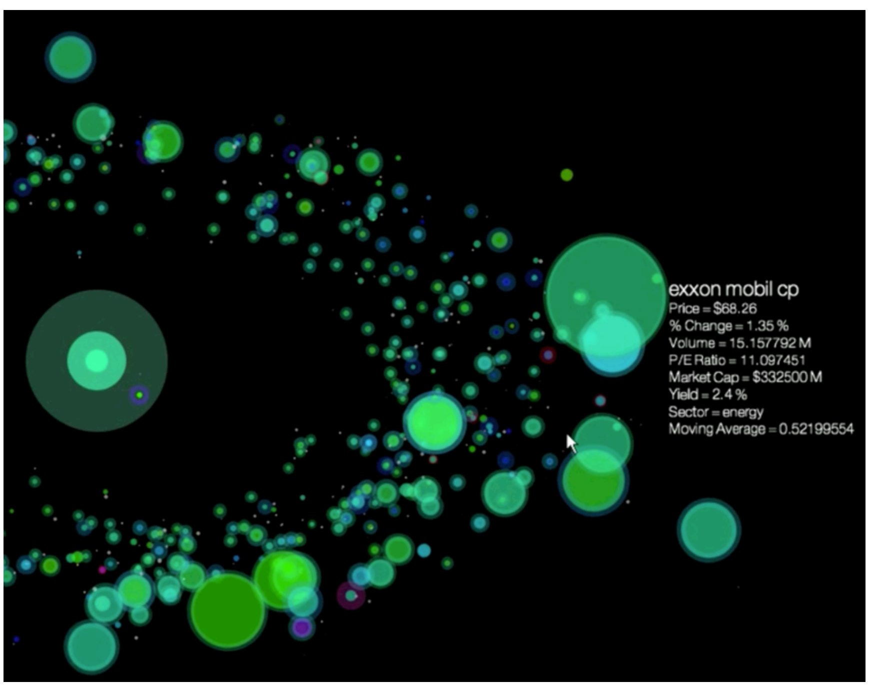

Figure 5. Stock visualization done by Grant et al.

D3.js [4] is a JavaScript library to display given data into graphic, dynamic forms using largely SVG, JavaScript, and CSS languages for data visualization. A hierarchical edge bundling with tension adjustment available is introduced as showed in Figure 6. The complexity of relationships between nodes represented by links is reduced significantly by grouping the edges from the same group together. The groups of links can be highlighted by simply moving the cursor over the nodes they originate from. But for visualizing stock market exchange network, more hierarchical visualization should be presented to scale to single broker with companies ranked by relationship significance.

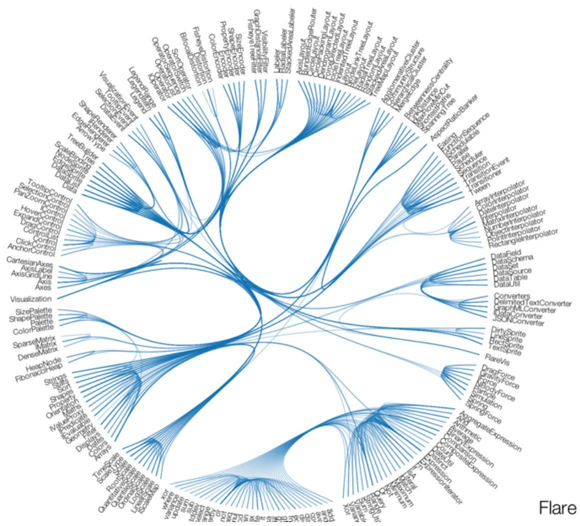

Figure 6. Edge bundling with high tension in D3js

Goals and Approaches

The goal of our visualization design of stock market data is to show the large scale stock data in a efficient way, and customize the visualized graph by user interaction to reduce the size and complexity of data in the graphs. We separate our visualization process into 3 phases in a cycle. First of all, we do the data analysis. Afterwards, the selected data are illustrated with the default scale of visualization. Thirdly, we collect the feedback or instruction from users to re-analyze the data and present next visualization for more useful or prospective data.

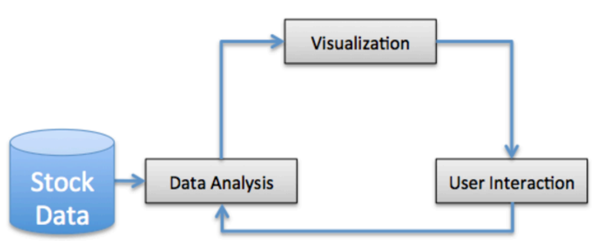

Figure 7. Visualization process

The larger quantity of Taiwan stock daily data set leads to the main difficulty of visualization. Large-scale data visualization usually contains too much information to present the data in effective and efficient way. Furthermore, when trying to reduce the size and complexity of data to visualize the information more specifically, it is also hard to do the statistical data analysis and make decision of classification. It is an intuitive way to utilize the technique of hierarchical visualization structure to organize the data into gradual displaying levels. A simple click or zoom in/out can switch the levels of information visualization in a fixed screen to optimize the user interface scalability through interactions. On the other hand, for applying large number node-link graphs to visualization of transaction of company shares through brokers, edge bundling introduces a state-of-art solution for clutter reduction. Furthermore, through analyzing the daily, monthly, and annual stock data and trading records of that individual company and related industry, our goal is to comprise the timeline of stock market data set to show the pattern of dynamical stock networks.

Our user interface design will be with at least two scales and timeline selection. The first layer is the correlation (association) between brokers by daily transaction of shares through them. The second layer is the graph of companies related to the specific broker selected by the user, which ranked by the volume of transactions. Furthermore, more details of one individual broker are illustrated when selected. The timeline below is considered as a scrollbar to change the data set scale. Afterwards, the process of data analysis and visualization will be done iteratively in our visualization process cycle.

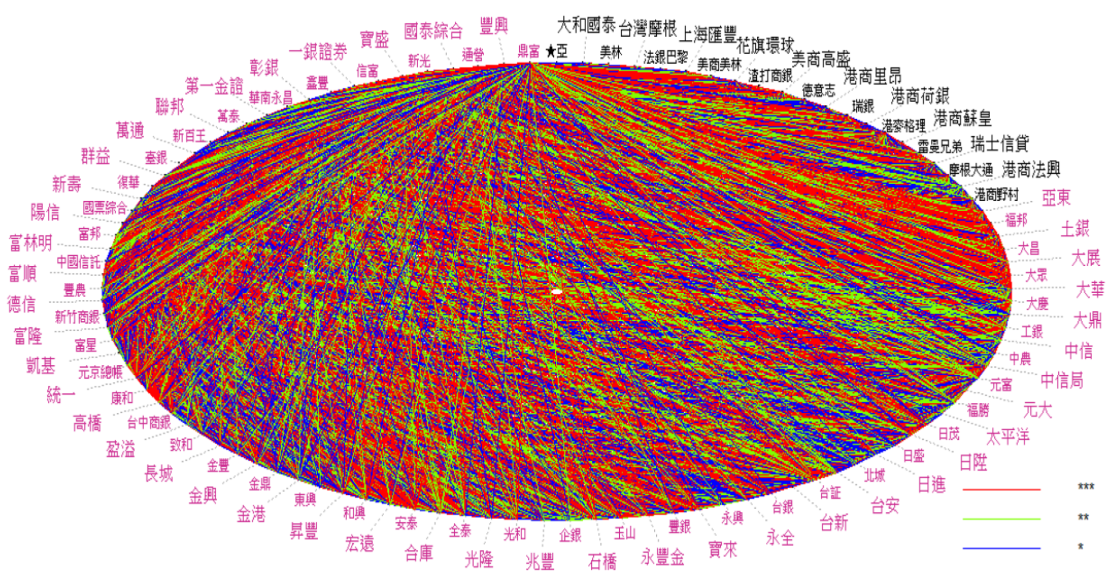

Figure 8. Showing brokers and associations with nodes and edges

Implementations

To draw the association among Taiwan stock brokers, the most straightforward method is to connect each node as one broker with edges illustrating the significance of associations by colors as showed in Figure 8. The higher association edges with P-value < 0.001 are indicated by red, the medium association edges with P-value 0.001 < P-value < 0.01 are indicated by green, and the low association edges with P-value 0.01 < P-value < 0.1 are indicated by blue. The black texts are the foreign brokers, while the others (gray texts) are all local brokers. Because of the large scale of association data, this visualization clutters in the middle with a great quantity of edges, which is almost useless for broker co-relationship analysis.

Firstly, we try to implement the conventional visualization without separating by local/foreign brokers showed in Figure 8. We adopt Qt cross-platform application GUI framework with OpenGL graphics library to apply to node-edge visualization. The nodes of brokers are showed with different colors because being not grouped in the first place. Note we use broker grouping later instead of marking foreign and local brokers with different font colors.

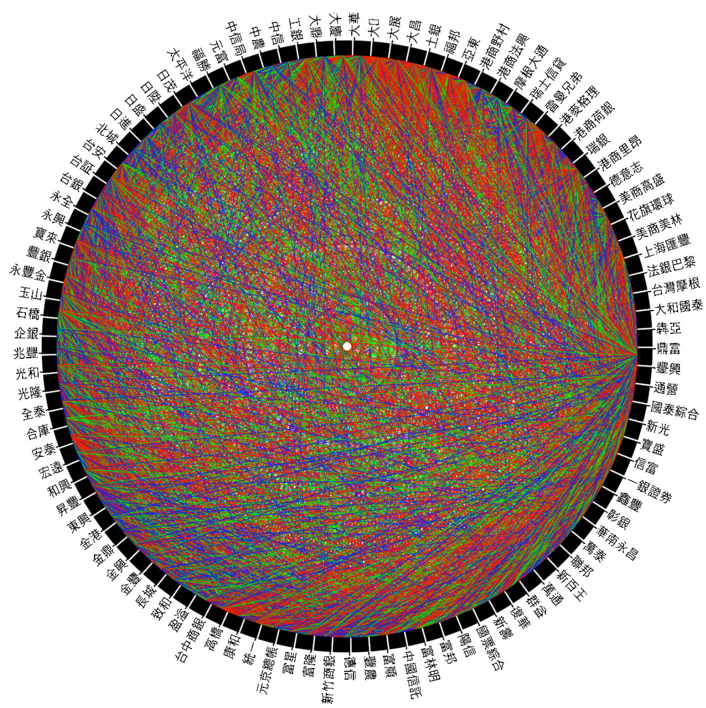

Figure 9. Conventional visualization using Qt and OpenGL

The technique we used for remodeling cluttering visualization is called hierarchical edge bundling with tension strength adjustment available. The bundling edges are defined by splines with bundling control point tree showed in Figure 10. The splines are constrained by the control points within different types of cutoff type. The Lowest Common Ancestor (LCA) cutoff type defines the splines of different communities from the root control point to the leaf points, while Bottom Up cutoff type defines in the opposite way. Note the colors of edges are defined by the colors in color map with corresponding gradient values of associations. The result of brokers and co-relationship visualization after edge bundling is more clear as showed in Figure 11. However, it is still impossible to tell the paths of edges. Note the red nodes are the control points that are unused/disabled.

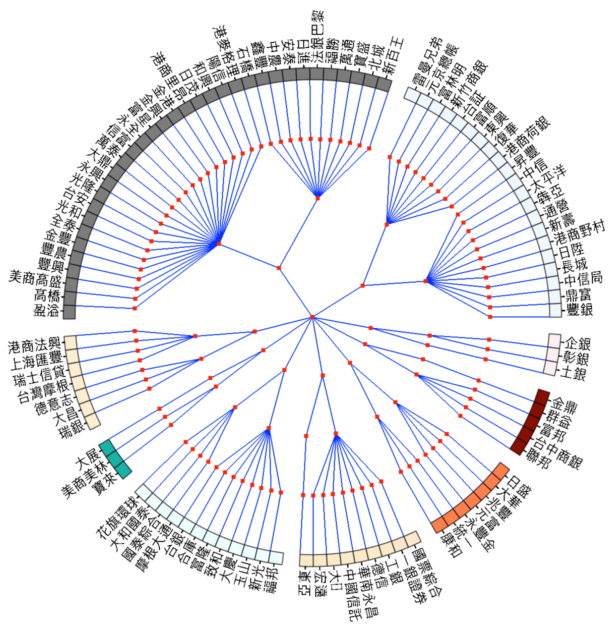

Figure 10. A sample of edge bundling control point tree

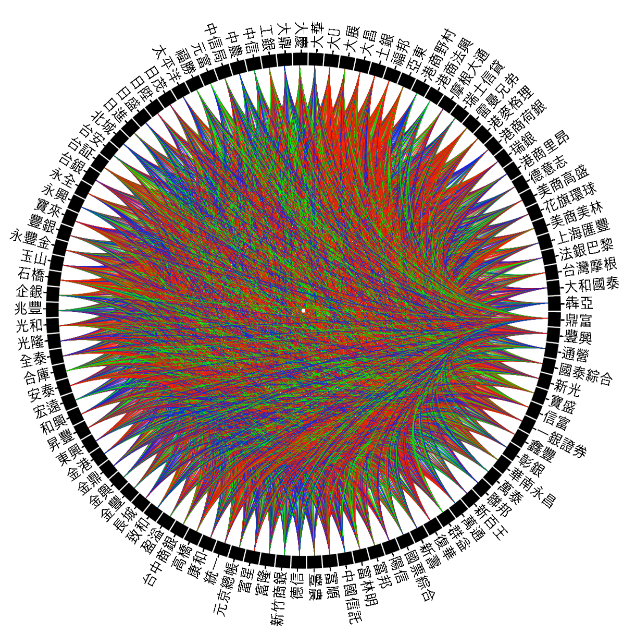

Figure 11. Visualization after edge bundling

Therefore, we separate the brokers into groups by foreign/local (Figure 12), capital (Figure 13), and star up date (Figure 14). According to the edge bundling techniques of different cutoff types, the edges can be clustered by groups. It is obvious that there are generally four independent groups which connect with each other to construct a networks when grouping the brokers to foreign or local.

Moreover, regarding the complexity of edges for multiple nodes because of great quantity of associations among brokers, it is believed that showing edges related to individual nodes can succinctly provides useful detail information. We draw the edges that are from or to the node selected when right clicking on one single node as showed in Figure 13. Note we only show the labels of nodes that are connected to the selected node. The different bundling strength in Figure 15 (0.95) and Figure 16 (0.75) make it a trade off between edge bundling clearance and complexity.

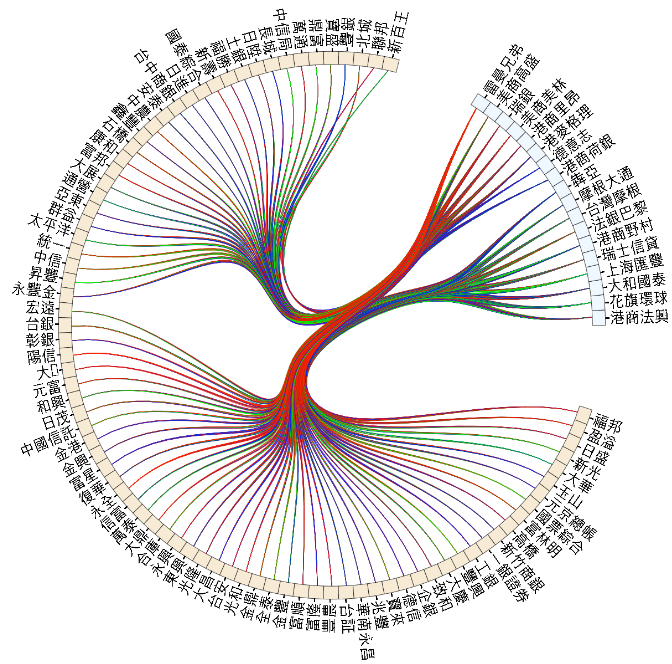

Figure 12. Visualization of edge bundling after grouping by foreign/local brokers

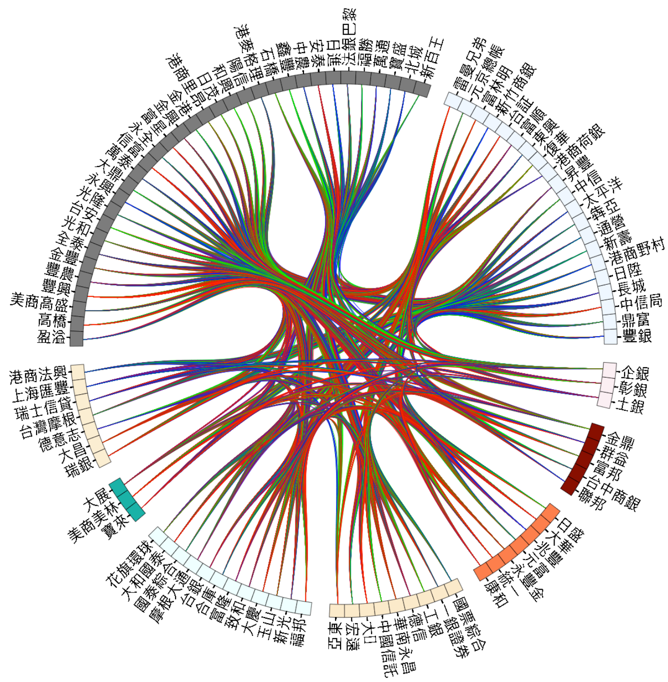

Figure 13. Visualization of edge bundling after grouping by capitals

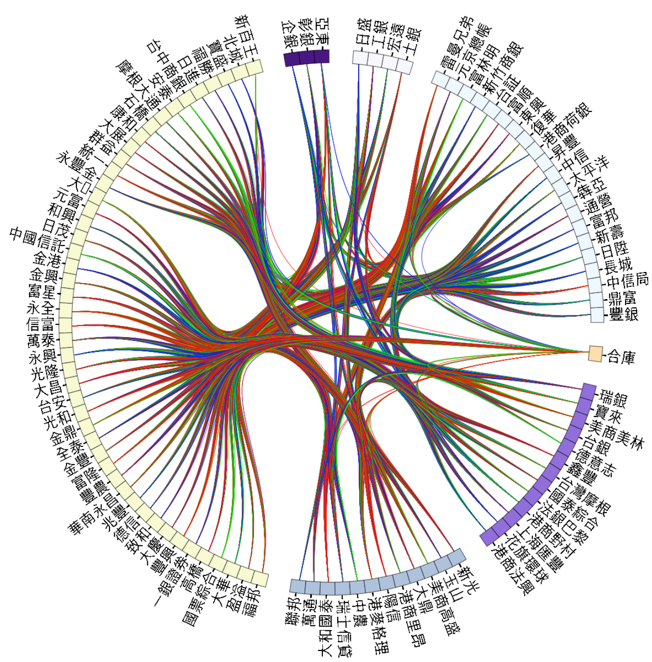

Figure 14. Visualization of edge bundling after grouping by start up date

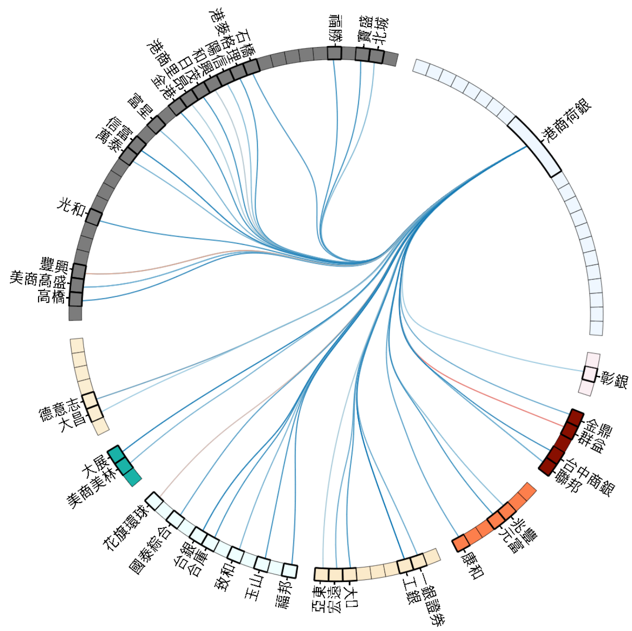

Figure 15. Visualization of edge bundling within one individual node with bundling strength 0.95

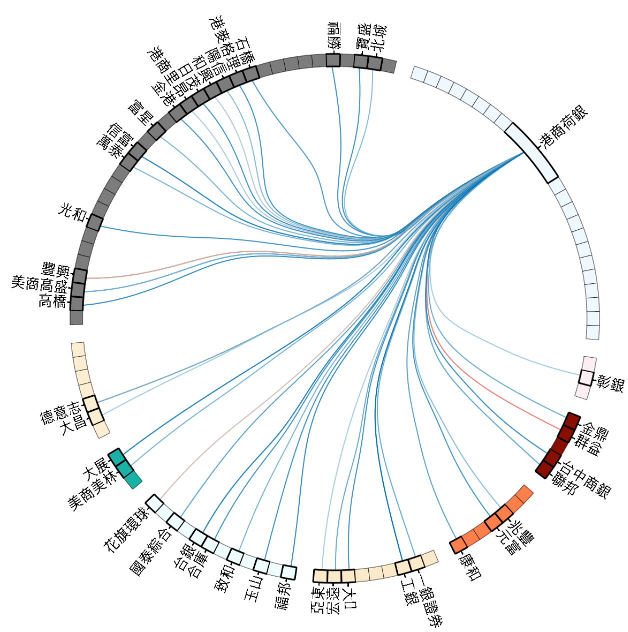

Figure 16. Visualization of edge bundling within one individual node with bundling strength 0.75

Furthermore, in this work, we accumulated and filtered the P-value by five fashions: True P-value, 3-level (0.1, 0.01, 0.001) P-value, 5-level (0.1, 0.05, 0.01, 0.005, 0.001) P-value, absolute values of 2 based logarithm P-values (|log2(P-value)|), and absolute values of 10 based logarithm P-values (|log2(P-value)|).

For higher succinctness of edges instead of only changing the bundling strength, we also exercise edge filtering. The Naive operating interface is as showed in Figure 17. We separate the edges by dividing the value of association to 100 threshold values (100 sectors) by the minimum to maximum. There are three modes for edge grouping: Group, Stack Up, and Stack Down. The Group mode shows the edges with the association value ranging from threshold ± 0.5 sector range. The Stack Up modes shows/accumulates the edges gradually when tuning up the threshold value by the scrollbar showed in Figure 15, while the Stack Down does in the opposite way. Figure 18, Figure 19, and Figure 20 illustrates edge accumulating in Stack Up mode when tuning up the threshold value to 5, 11, and 18, respectively. Note the values are defined by absolute values of 2 based logarithm P-values (|log2(P-value)|).

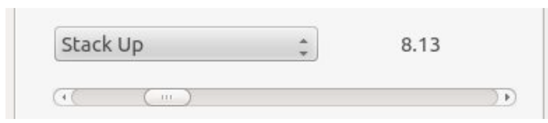

Figure 17. Naive edge grouping and stacking operating interface

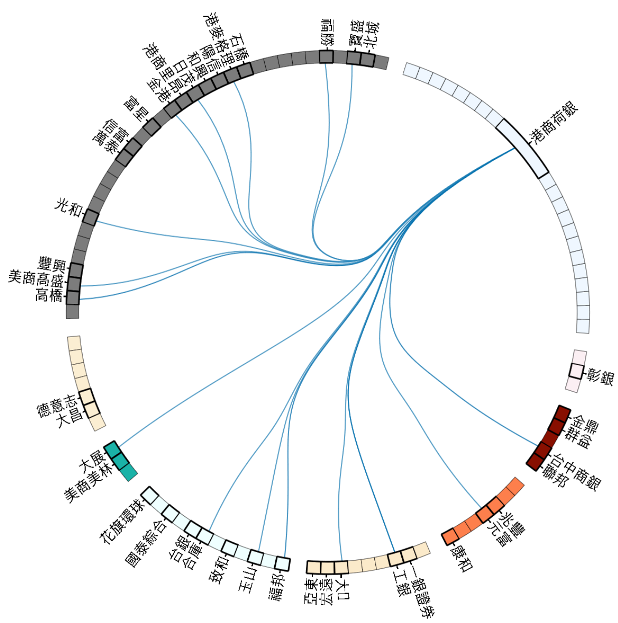

Figure 18. Visualization of edge bundling within one individual node in Stack Up mode when setting the threshold 5

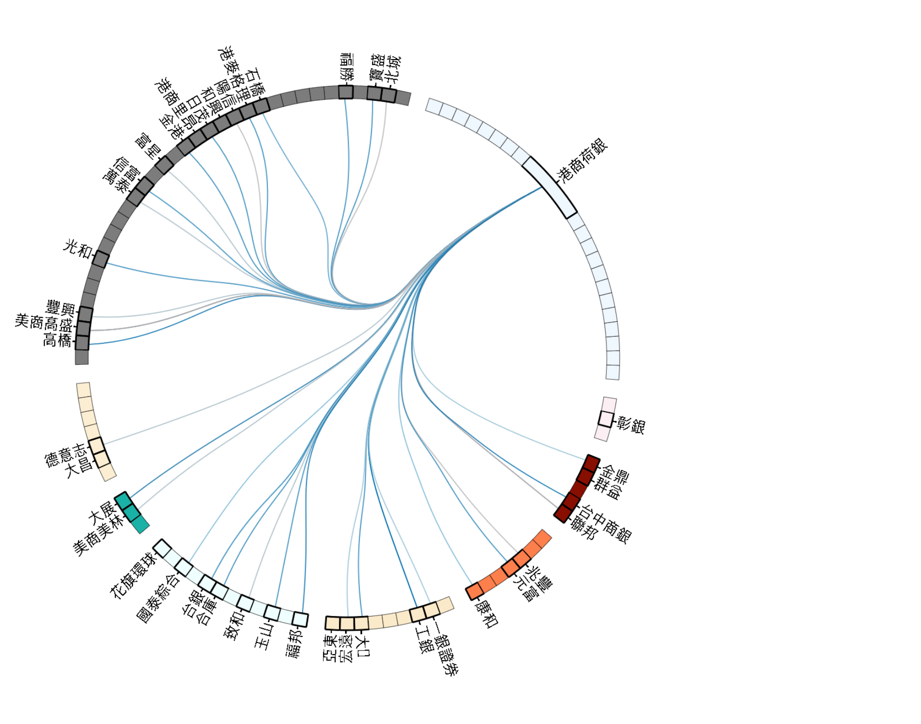

Figure 19. Visualization of edge bundling within one individual node in Stack Up mode when setting the threshold 11

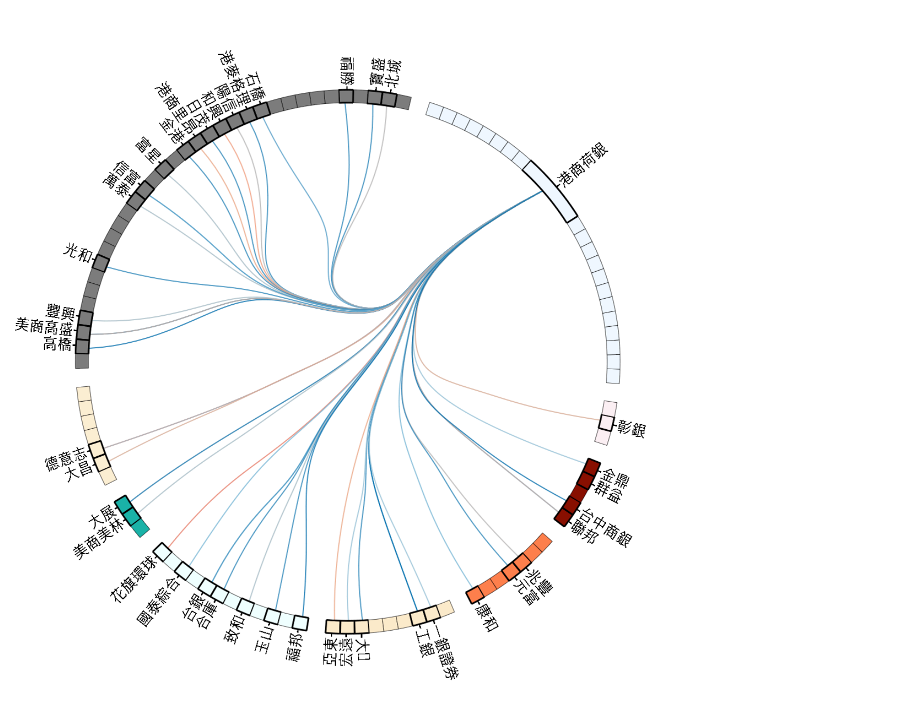

Figure 20. Visualization of edge bundling within one individual node in Stack Up mode when setting the threshold 18

Color Map

We extracted the color map from the website "Color Brewer" and build new color map for better user experience as showed in Figure 21. We used diverging color map (on the right in Figure 21) instead of sequential color map (on the left in Figure 21) for sequential color map's color being too light when the value is small since we have the background in white. We utilize the colormap "BuRd" showed on the right in Figure 21 with the white color in the center removed regarding to the possibility of misrecognizing the edges with that color as low values. All the color is remodeled to gradient form for the continuous characteristics of broker associations. Figure 22 and Figure 23 show the difference between two visualizations with different color maps.

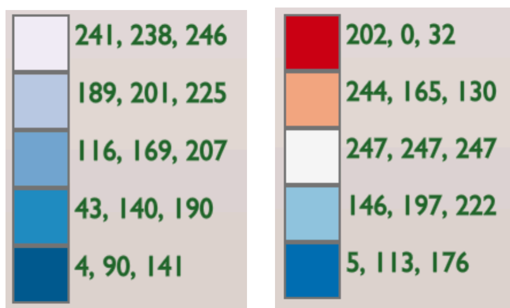

Figure 21. Sequential (left) and diverging color map (right) from Color Brewer

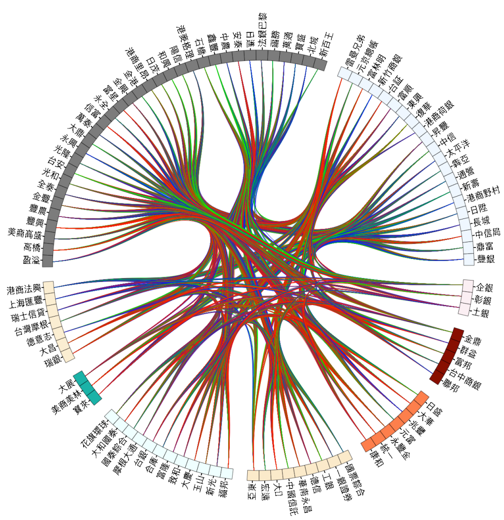

Figure 22. Visualization from original blue, green, and red colors

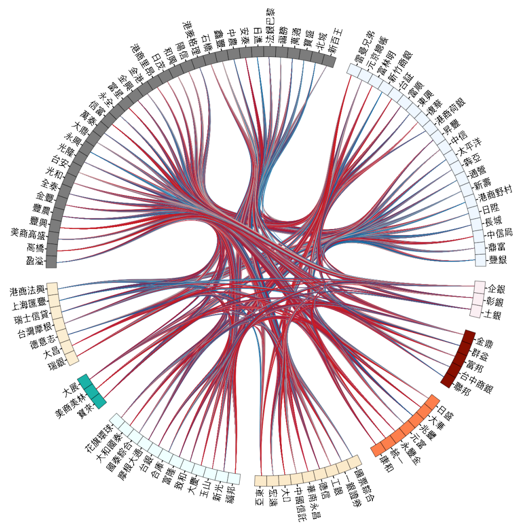

Figure 23. Visualization from customized BuRd color map

Visual Cut

In this work, we further implemented the cutoff type "Visual Cut" with two inner dashed circles defining enabling/disabling the control points in the control point tree showed in Figure 24, Figure 25, and Figure 26 (enabled: blue nodes, disabled: red nodes). When two dashed circles for Visual Cut are far apart, all the control points are enabled (with minor nodes unused with red). When two dashed circles for Visual Cut are close in the same level of control point tree, all the control points are disabled.

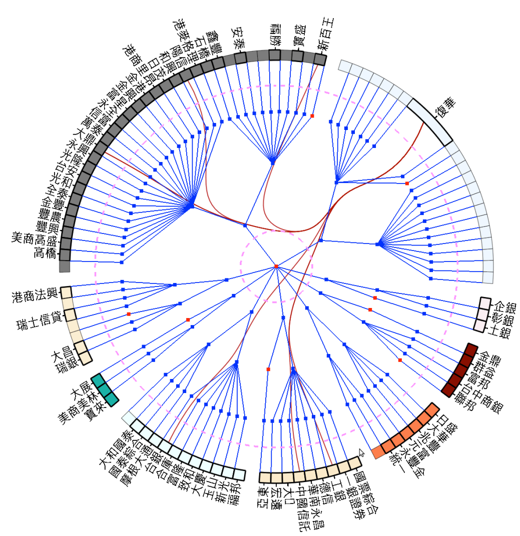

Figure 24. Two dashed circles for Visual Cut are far apart

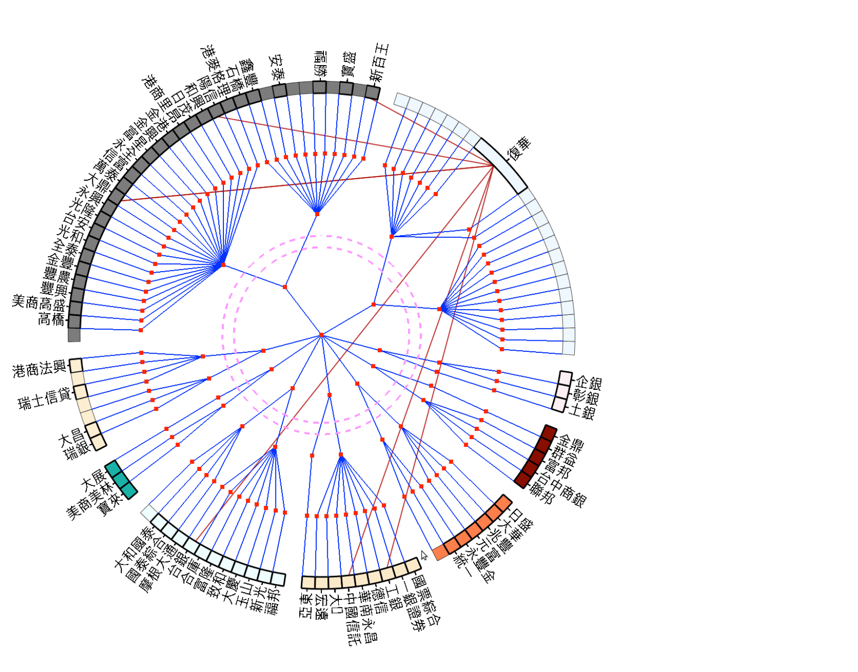

Figure 25. Two dashed circles for Visual Cut are close

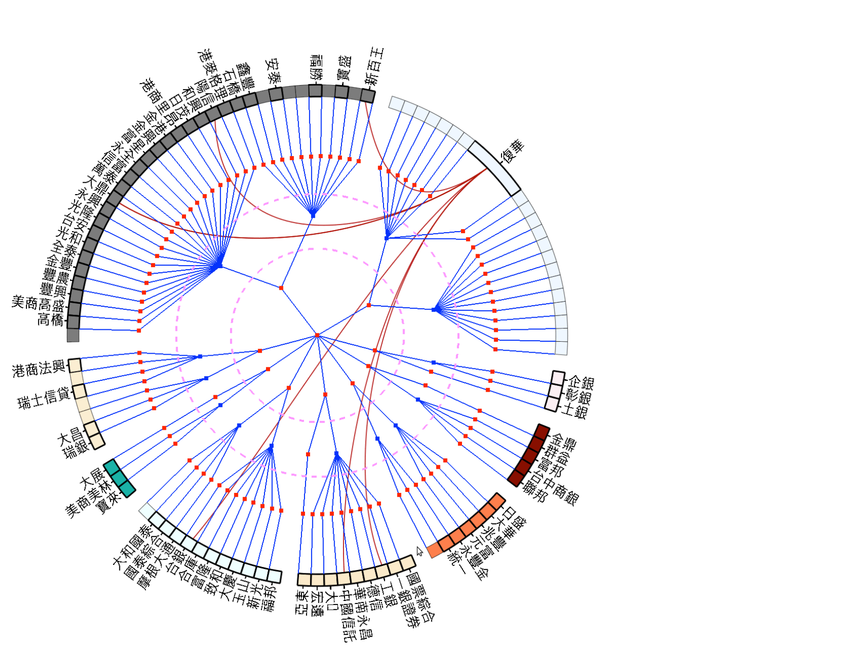

Figure 26. The control points between two dashed circles for Visual Cut are enabled

Distribution Histogram

One down side of edge bundling is lacking of statistic analysis of edge values. For better understanding of the distribution of P-values, we couple the edge bundling graph with interactive P-value distribution histogram. The data is separated into 100 categories by values as the number of grouping in edge filter technique for interaction between two graphs. Figure 28, Figure 29, and Figure 30 show the interaction of Edge bundling and P-value distribution histogram with the changes of threshold value of edge filtering (in stack up mode). Note that the distribution histogram is zoomable in vertical direction for better visualization of bar scaling as showed in Figure 28.

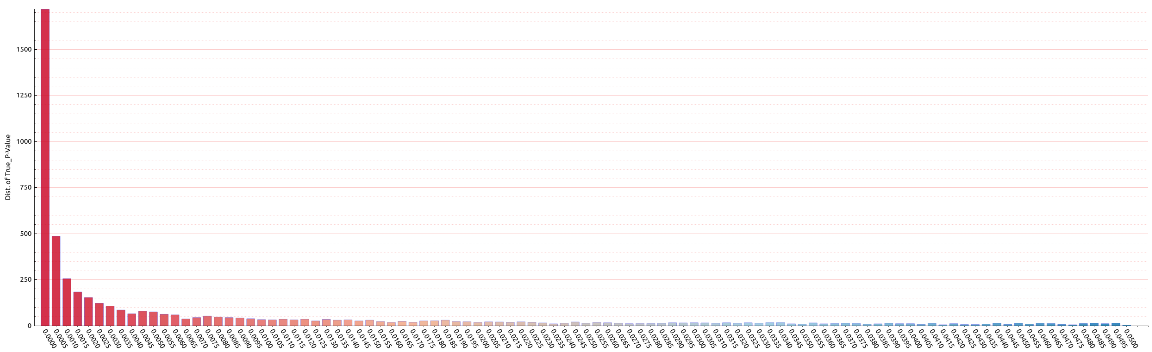

Figure 28. "True" P-value distribution histogram

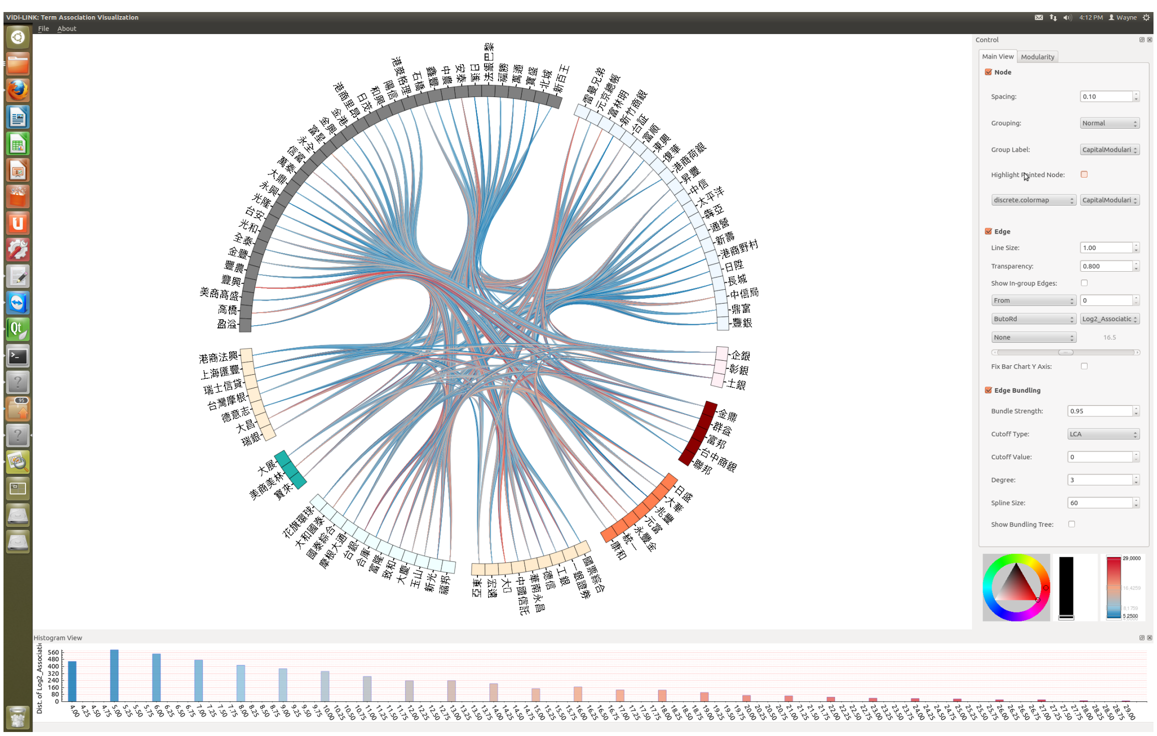

Figure 28. Edge bundling with P-value distribution histogram

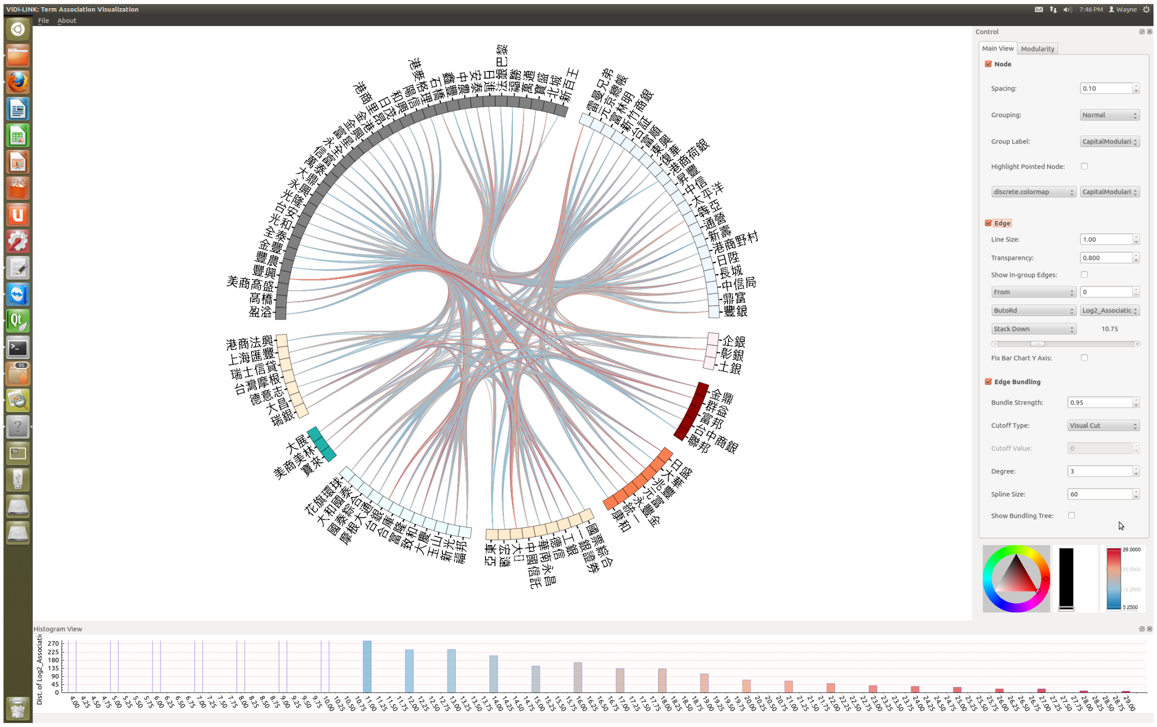

Figure 29. Edge bundling with P-value distribution histogram with stack up edge filtering - 1

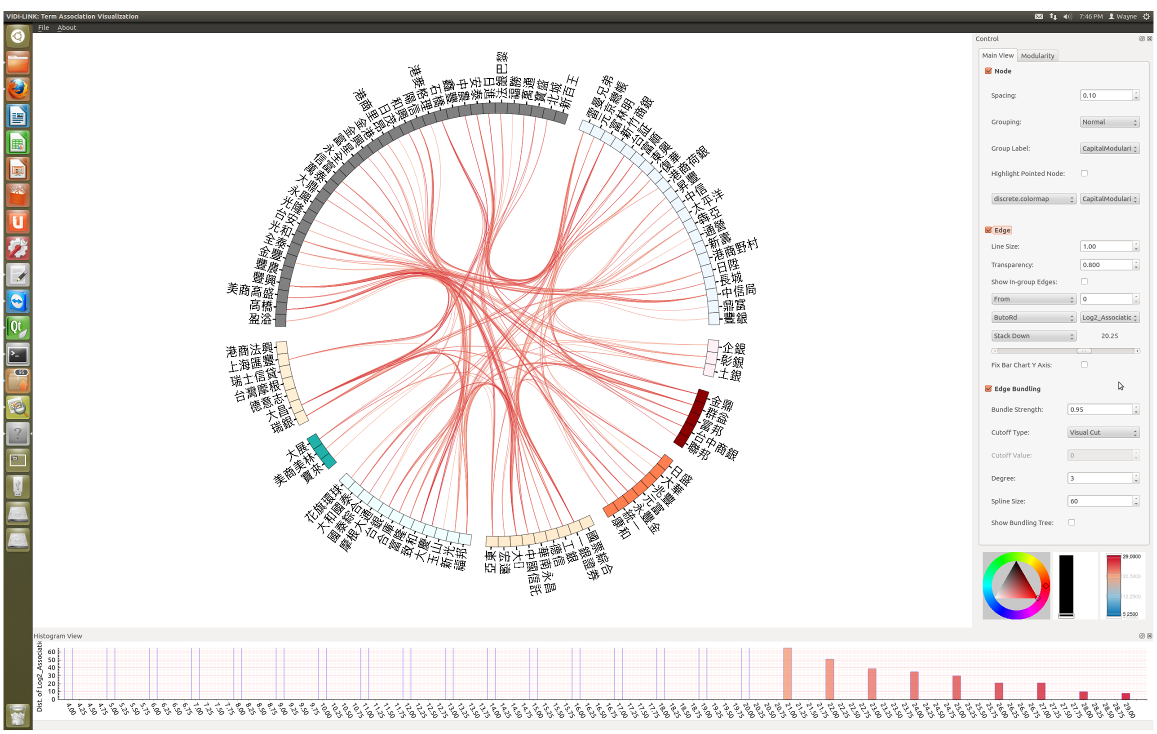

Figure 30. Edge bundling with P-value distribution histogram with stack up edge filtering - 2

Conclusion

In this work, we apply edge bundling to visualization of stock brokers associations defined by P-values, which represent the significance level of hypothesis of correlation among brokers. The P-values are recalculation into 4 types, 3-level, 5-level, 2 based logarithm and 10 based logarithm. The brokers are also clustered into groups by locations, capitals, and start up date for different bundling types. The changes of bundling strength, control points, transparency, degree improve the resolution and readability of paths of edges. Furthermore, edge filtering provides a user interactive method to details of paths and values in a state-of-art way. In the end, a distribution histogram is applied to help user a better understanding of the distribution of P-values. We hope that this work can help investors or analyzers come up with concrete concept of the trend of stock market in a effective and efficient way, and take it as a reference of volatility analysis for reducing investment losses in the stock market.

Future Works

In the future, our search topic for this work will focus on finding the paths with better user interaction techniques. Edges should be also filtered visually in a more succinct and effective way as a visualization for analysis. Active edge filtering by user is also recommended. Moreover, further detail information of nodes and edges may be provided in lists with user interactions, such as selecting or hovering on the edges.

References

[1] Danny Holten. "Hierarchical Edge Bundles: Visualization of Adjacency Relations of Hierarchical Data". In *TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS,* IEEE, 2006.

[2] Ulrik Brandes et al., "On Module Clustering". In *TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, IEEE, 2009

[3] Grant et al., "Stoc – Stock Ticker Orbital Comparison", In: http://uniformchaos.org/stoc/

[4] Jason Davies, "Edge Bundling with D3js", In: http://mbostock.github.com/d3/talk/20111116/bundle.html

[5] "Color Brewer", In: http://colorbrewer2.org/