Use of association rules for cause-effects relationships analysis of collision accidents in the Yangtze River

B.Wu, J.H.Zhang & X.P.Yan
Intelligent Transportation System Research Center (ITSC),
National Engineering Research Center for Water Transport Safety,
Wuhan University of Technology, Wuhan, China
T.L.Yip
Department of Logistics and Maritime Studies, Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

ABSTRACT: In order to discover cause–effect relationships in collision accidents, an association rule-based method is applied to analyze the historical accident data in the Jiangsu section of the Yangtze River from 2012 to 2016. First, the Apriori algorithm is introduced for interesting rules mining, and three types of measures of significance and interest are considered, which are support, confidence and lift. Second, the data are discretized based on previous studies and work experience, and the R software is introduced to facilitate the modeling process. Third, the contributing factors are discovered from the cause-effect relationship analysis. Finally, the generated rules are visualized using the Gephi software to further analysis the unknown relationships and patterns. The observed patterns of collision accidents can be avoided by cutting off some factors in the sequential chain of collision accidents, which is beneficial for prevention of collision accidents. Consequently, this paper provides a data-driven method for accident analysis and prevention.

1 INTRODUCTION

Ship collision is a typical accident and accounts for a lot among all maritime accidents (Cai et al. 2017, Sedova et al. 2018). For example, from 1953 to 2002, 461 serious maritime accidents occurred in the Istanbul Straits, the majority of them, i.e. 45.34%, were ship collisions (Akten 1999). In the Gulf of Finland, collision accidents rank second from 1997 to 2006 (Kujala et al. 2009). From 2008 to 2013, the collision accidents accounted for 65% Tianjin Port (Zhang et al. 2018). Similarly, in the Jiangsu section of the Yangtze River, the ship collision accidents account for more than 60% (Jiang 2010).

Owing to the relatively high occurrence of ship collision accidents, many studies focused on ship collision risk analysis and mitigation. Wang & Yang (2018) proposed a novel method to calculate the severity of water transport accidents based on Bayesian network. Sii et al. (2001) developed a fuzzy logic-based model by considering various variables in the concept design stage. Sedova et al. (2018) presented a fuzzy neural system for collision avoidance in busy waterways. Moreover, Hu et al. (2007) proposed a safety assessment method for risk management of waterborne transport after defining four unique criteria.

From previous studies, valuable insights have been gained. However, it can be seen that the previous models require the practitioners have good knowledge and understanding of the accident development, which may cause some uncertainty. However, association rule is a data mining method, which means it manages to discover the patterns of collision accidents only from the accident data without any prior knowledge. Therefore, the motivation of this paper is to use association rules to discover the cause-effect relationships from a variety of causation factors.

The remainder of this paper is organized as follows. The definitions of association rules are introduced in Section 2. The cause-effect relationship analysis of ships collision accidents is introduced in Section 3, which includes data sources, association rules of ship collision accidents, analysis of high support and lift, and visualization of association rules. Conclusions are drawn in Section 4.

2 ASSOCIATION RULES

2.1 *Introduction of association rules*

Association rule learning is a rule-based machine learning method for discovering hidden relationships between variables in a database from the perspective of data mining. When introducing it to ship collision accident analysis, after discovering the patterns of ship collision, it is meaningful to take countermeasures to cut off the necessary causation factors in an association rule. For example, an association rule for ship collisions in the Yangtze River is {accident area = anchorage} => {encounter scenarios = collision with stationary ship}. It shows that the collision with stationary ships will have a large probability to occur when ships are anchored at an anchorage. Therefore, the officer on watch should always take sharp lookout to prevent the occurrence of collision accidents when anchoring in the anchorage.

2.2 *Definition of association rules*

In association rules, the collision data record set is called the database, defined as *D*; the collection of all items is called the itemsets, which is defined as *I*; each accident record in the database is defined as *T*, where *T*∈*D*. The set of items that appear simultaneously in accident record is called an itemset, and it is defined as *K-th* itemset. The items on the left and right of the symbol "= >" are referred to the antecedent and the succedent of the association rules. The frequency that uses to measure the occurrence frequency of an item is defined as *support*.

$$support(A \Rightarrow B) = P(A \cap B) \qquad (1)$$

When *A* belongs to the data record *D*, the probability of *B* also belongs to *D* is defined as *confidence*, i.e. the conditional probability.

$$confidence(A \Rightarrow B) = P(B|A) = P(A \cap B)/P(A) \qquad (2)$$

Another method used for measuring the relationship is called *lift*, which is defined as the interest (Ochin et al. 2016, Grabot 2018).

$$lift(A \Rightarrow B) = confidence(A \Rightarrow B)/P(B) = (P(A \cap B))/(P(A)P(B)) \qquad (3)$$

From this definition, if lift = 1, *A* and *B* are independent and have no influence; if lift > 1, *A* and *B* are interdependent, mutually reinforcing and positively correlated; if lift < 1, *A* and *B* are mutually constrained, mutually reinforcing and negatively correlated (Barati et al. 2017). Moreover, if lift = 0, *A* and *B* will not occur simultaneously.

2.3 *Association rule mining*

There are three widely used algorithms for association rules, which are Apriori algorithm, Partition-based algorithm and FP-tree algorithm, respectively. Apriori algorithm, proposed by Rakesh & Ramakrishnan in 1994 (Srikant & Agrawal 1994), is a basic algorithm in association rules and widely used as a classical algorithm (Borah & Nath 2018, Weng & Li 2017, Xu et al. 2018). This paper uses this algorithm because it is intuitive, concise and easy to implement. The flow chart of Apriori algorithm is shown in Figure 1.

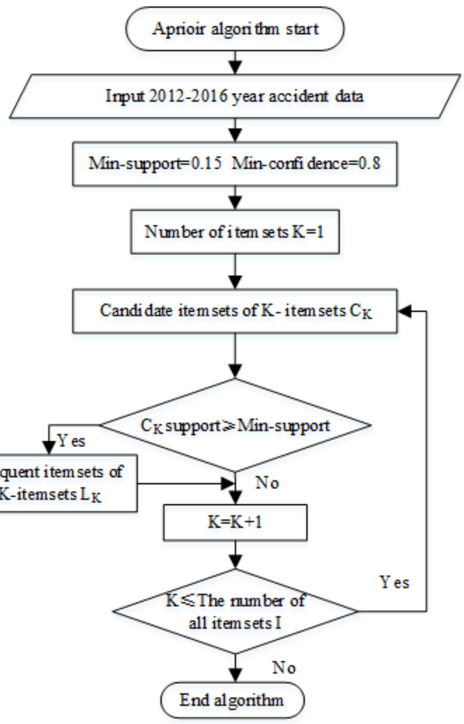

Figure 1. Flow chart of Apriori algorithm

3 CAUSE-EFFECTS RELATIONSHIPS OF INLAND SHIPS COLLISION ACCIDENTS

3.1 *Data Sources of ship collision accidents*

The ship collision accidents data were collected by Jiangsu Maritime Bureau. The database fully records the collision information of ships accidents in inland waterway from 2012 to 2016. The original collision data are discretized, and then the following set of accident features are obtained. The dataset contains the following information: the Yangtze River water period (i.e. flooding season, normal season, and dry season), time (i.e. daytime and nighttime), location of accidents (i.e. Changshu, Changzhou, Jiangyin, Nanjing, Nantong, Taicang, Taizhou, Yangzhou, Zhangjiagang and Zhenjiang), accident area (i.e. channel, anchorage and others), severity of consequences (i.e. negligible, minor and major), shipwreck (i.e. yes or no), number of fatalities (i.e. 0 fatalities, 1-2 fatalities and 3-9 fatalities),Causation factors (i.e. ship conditions, environmental factor, human factors and cargo factors), encounter scenarios (i.e. head-on situation, crossing situation, collision with stationary ships and overtaking).

From 2012 to 2016, 963 maritime accidents occurred in this waterway area. The proportions of different states for each variable are given Table 1. Specifically, the accident rate of flooding season is higher than normal season and dry season; the accident rate of nighttime is higher than daytime; the location of accident rate in Zhenjiang is obviously higher than other areas. In addition, the accident rate of channel is even higher than other water areas; the accident rate of accidents caused by human factors is around 80%, which coincides with previous statistics of accidents (Fan et al. 2017).

Table 1. Classification statistics of accident data

Variable	Description	Count	Percent
Water period (WP)	Dry season (DS)	295	30.63
	Flooding season (FS)	345	35.83
	Normal season (NS)	323	33.54
Time (T)	Daytime (DT)	390	40.50
	Nighttime (NT)	573	59.50
Location of accidents (LOA)	Zhenjiang	270	28.04
	Jiangyin	121	12.56
	Nanjing	118	12.25
	Nantong	118	12.25
	Taizhou	96	9.97
	Taicang	81	8.41
	Others	159	16.51
Accident area (AA)	Anchorage (ANCH)	158	16.41
	Channel (CHAN)	569	59.09
	Others	236	24.51
Severity of consequences (SOC)	Major	16	1.66
	Minor	52	5.40
	Negligible (NEG)	895	92.94
Shipwreck (SW)	No	847	87.95
	Yes	116	12.05
Number of fatalities (NOF)	No	916	95.12
	1-2 fatalities	38	3.95
	Over 2 fatalities	9	0.93
Causation factors (CF)	Cargo factors (FOC)	3	0.31
	Environmental factors(EF)	123	12.77
	Human factors (HF)	788	81.83
	Ship conditions (SC)	49	5.09
Encounter scenarios (ES)	Collision with stationary ship (CWSS)	399	41.43
	Crossing situation (CS)	350	36.34
	Head-on situation (HOS)	79	8.20
	Overtaking (OT)	129	13.40

3.2 *Rules of ship collision accidents*

In order to derive the association rules for ship collision, the Aprior algorithm is introduced by using "arules" package in R software. In this case study, the original data of 963 marine accidents were divided into several sections and changed into CSV formats. As shown in Figure 1, the thresholds of minimum support and minimum confidence are defined as 0.15 and 0.8, respectively, and 1243 association rules are generated after using this definition.

Note that some rules share the same semantic measure or statistical measure in the extracted rule set, which is considered as redundant rules (Sahoo et al. 2015). In other words, when an association rule is the parent rule of another rule, and the parent rule has the same or lower lift, the parent rule is a redundant rule. For example, a parent rule: {location of accidents = Zhenjiang, severity of consequences = negligible, number of fatalities = 0 fatalities} => {accident area = channel}, where the lift = 1.357, and the child rule: {location of accidents = Zhenjiang, severity of consequences = negligible} => {accident area = channel}, where the lift = 1.357. This parent rule is redundant because redundant rules cannot provide more detailed and effective information, the existence or absence of the item "number of fatalities = 0 fatalities" has no practical assistance on the child rule. By removing the redundant rules, there are only 231 effective rules used for the next step.

Afterwards, the "arulesViz" package in the R software is utilized to visualize the obtained association rules. As the factors are classified into 37 categories, the items with low support tend to have negligible influence on the rules. In order to derive the high frequency items of collision accidents, the first 20 items of frequency distribution are selected and the frequency distribution is shown in Figure 2. It can be seen that more than 95% of the accidents haven't caused casualties, over 90% of the accidents are minor accidents, 87.95% of them haven't caused shipwreck, human factors is prominent, and the occurrence of collision in Zhenjiang section is high.

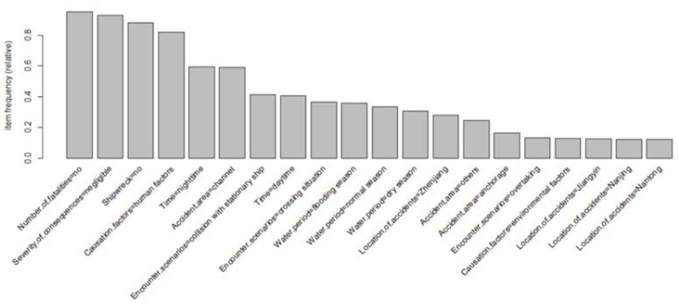

Figure 2. Frequency distribution of items for marine accidents

The degree of lift is greater than 1 indicates that there is a positive correlation between the antecedent and the succedent (Geurts et al. 2005). By setting the threshold of the lift degree as 1, 231 positive correlation rules are obtained. The above rules are visualized in Figure 3. The default aggregate function takes the average value of the group of association rule and is represented by the color and size of the graph (Weng et al. 2016). Similar association rules are divided into one group so as to extract the general characteristics of association rules. The vertical bar represents the antecedent of the positive association rules and is clustered into 58 groups. Due to the limitation of the size of the graph, part of the antecedent is omitted, and the number of omitted items is reflected in the graph. The transverse bar represents the succedent of the positive association rules which is clustered into 6 groups. The size of the balloon indicates the support level. The larger the balloon, the higher the support degree of the association rules. Moreover, the color of the balloon indicates the lift level. The deeper the balloon color, the higher the lift degree of association rule, and more closely relationship between the antecedent and the succedent. For example, it can be seen from Figure 3 that ships in anchoring areas are more likely to collide with stationary ships.

Figure 3. Grouping matrix diagram for 231 rules

Note that from Figure 3, the groups of high lift rules and high support rules do not coincide with each other, and there are still some hidden information needs to be further analyzed. Therefore, the scatter plot is introduced for these 231 rules and the result is shown in Figure 4. This figure can clearly show the difference and the relation among the three measures of support, lift and confidence. Each point in Figure 4 represents an association rule. The majority of support is below 0.4 while the majority of lift is between 1.0 and 1.3 with some rules more than 2.

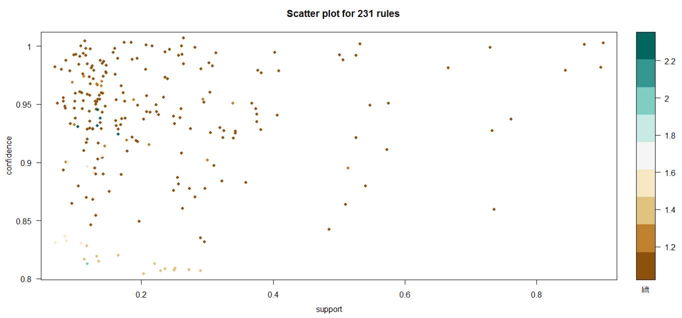

Figure 4. Scatter plot for 231 association rules of collision accidents

3.3 *Analysis of high support association rules*

The higher support, the greater probability of occurrence of the item. In order to further explore the relationship between high support rules and high lift rules, the thresholds for setting high support rules and high lifting rules are 0.3 and 1.1 respectively, and only 46 rules are derived, which is shown in Figure 5. The source in Figure 5 represents the antecedent, the arrow indicates the direction of the relationship, and the end of the arrow indicates the succedent. It is discovered from this figure that the high support rules are mainly distributed in {severity of consequences = negligible}, {shipwreck = yes}, {causation factors = human factors}, {time = nighttime}, {accident area = channel}. To a certain extent, this indicates the general principle of the collision accidents in the Yangtze River section in recent years.

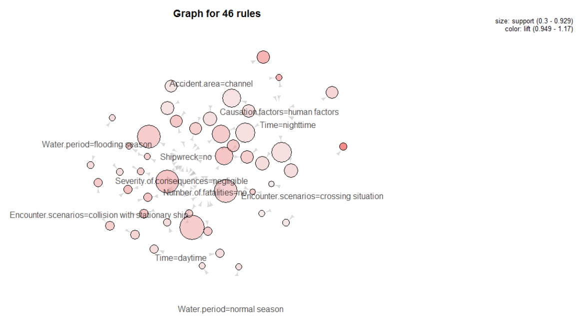

Figure 5. Graph for 46 rules of high support rules

To further understand the rules expressed in the above graphs, the high support rules are shown in Table 2. It can be seen that the confidence of the rule {water period = flooding season} => {number of fatalities = 0 fatalities} is 0.954, which indicates that the flooding season of the Yangtze River will not have a great impact on casualties. It also shows that the characteristic of the Yangtze River collision accident is that: minor accidents occupy a high proportion, the majority of them occur in the channel and nighttime. And human factor is the main causation factor.

Table 2. High support rules for collision accidents

Antecedent	**Succedent**	**Supp**	**Conf**	**Lift**
{SOC=NEG}	{NOF=no}	0.929	1.000	1.051
{SW=no}	{NOF=no}	0.870	0.989	1.040
{SW=no, OF=no}	{SOC=NEG}	0.865	0.994	1.070
{SW=no}	{SOC=NEG}	0.865	0.983	1.058
{CF=HF}	{NOF=no}	0.773	0.944	0.993
{CF=HF}	{SOC=NEG}	0.755	0.923	0.993
{CF=HF}	{SW=no}	0.709	0.867	0.985
{SW=no, NOF=no, CF=HF}	{SOC=NEG}	0.698	0.996	1.071
{SW=no, CF=HF}	{SOC=NEG}	0.698	0.984	1.059
{T=NT}	{NOF=no}	0.567	0.953	1.002
{AA=CHAN}	{NOF=no}	0.560	0.947	0.996
{T=NT}	{SOC=NEG}	0.551	0.927	0.997
{AA=CHAN}	{SOC=NEG}	0.541	0.916	0.985
{AA=CHAN}	{CF=HF}	0.532	0.900	1.100
{T=NT}	{SW=no}	0.520	0.874	0.994
{T=NT, SW=no}	{NOF=no}	0.517	0.994	1.045
{T=NT, SW=no}	{SOC=NEG}	0.514	0.988	1.063
{AA=CHAN}	{SW=no}	0.506	0.856	0.973
{AA=CHAN, SW=no}	{NOF=no}	0.503	0.994	1.045
{T=NT}	{CF=HF}	0.501	0.841	1.028
{T=DT, NOF=no}	{SW=no}	0.353	0.919	1.045
{ES=CS}	{CF=HF}	0.348	0.957	1.170
{WP=FS}	{NOF=no}	0.342	0.954	1.003
{ES=CS}	{NOF=no}	0.335	0.923	0.970
{WP=FS, NOF=no}	{SOC=NEG}	0.334	0.979	1.053
{WP=FS}	{SOC=NEG}	0.334	0.933	1.004

{ES=CS}	{SOC=NEG}	0.324	0.891	0.959
{T=NT,	{CF=HF}	0.323	0.907	1.108
{WP=NS}	{NOF=no}	0.318	0.947	0.996
{WP=FS}	{SW=no}	0.317	0.884	1.005
{WP=FS, OF=no}	{SW=no}	0.314	0.918	1.044
{WP=FS, SOC=NEG}	{SW=no}	0.312	0.932	1.059
{WP=NS}	{SOC=NEG}	0.310	0.926	0.996
{ES=CS}	{SW=no}	0.303	0.834	0.949
{SW=no, ES=CS}	{NOF=no}	0.300	0.990	1.041

3.4 *Analysis of high lift association rules*

The lift is introduced to further discover the correlation of association rules. The higher the lift, the closer the relationship of rules is. Define the threshold value of lift as greater than 1.1, 42 rules are derived and the result is shown in Figure 6. As can be seen from Figure 6, the high lift rules are: {accident area = anchorage}, {shipwreck = yes}, {encounter scenarios = collision with stationary ships}. High support rules are: {causation factors = human factors}, {encounter scenarios = crossing situation}, {accident area = channel}. These factors are closely related to the characteristics of the collision accident in the Yangtze River, therefore, countermeasures should be taken, e.g. more attention should be paid in the anchorage as the majority of collision accident occurred in the anchorage from the rule {accident area = anchorage}.

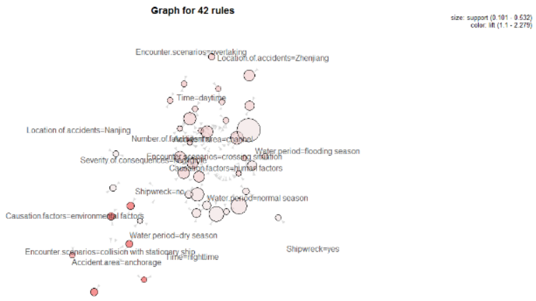

Figure 6. Graph for 42 rules of high lift rules

In order to explain Figure 6, the rules with high lift values are selected and listed in Table 3. It can be seen that the high lift rule includes {encounter scenarios = collision with stationary ship}, {accident area = channel}, {causation factors = human factors}. The causes of collision accidents in the Yangtze River are closely related to these factors. According to this result, it is found that there is a tendency to collide with a stationary ship in an anchorage area, but will not cause shipwreck and the severity of consequences is negligible because quick and effective emergency response actions can be taken.

Table 3. High lift rules for collision accidents

Antecedent	**Succedent**	**Supp**	**Conf**	**Lift**
{AA=ANCH, SW=no}	{ES=CWSS}	0.141	0.944	2.279
{AA=ANCH, SOC=NEG}	{ES=CWSS}	0.147	0.934	2.255
{AA=ANCH, NOF=no}	{ES=CWSS}	0.147	0.934	2.255
{T=NT, AA= ANCH }	{ES=CWSS}	0.101	0.933	2.251
{AA=ANCH}	{ES=CWSS}	0.152	0.924	2.230
{CF=EF}	{ES=CWSS}	0.103	0.805	1.943
{SOC=NEG, ES=OT}	{AA=CHAN}	0.112	0.900	1.523
{NOF=no, ES=OT}	{AA=CHAN}	0.112	0.900	1.523
{CF=HF, ES=OT}	{AA=CHAN}	0.105	0.894	1.513
{ES=OT}	{AA=CHAN}	0.119	0.891	1.509
{WP=NS, CF= HF, ES=CS}	{AA=CHAN}	0.104	0.840	1.422
{T=DT, NOF=no, CF=HF, ES=CS}	{AA=CHAN}	0.104	0.833	1.410
{WP=NS, ES=CS}	{AA=CHAN}	0.106	0.829	1.403
{T=DT, SOC= NEG, CF=HF, ES=CS}	{AA=CHAN}	0.101	0.829	1.403
{T=DT, CF=HF, ES=CS}	{AA=CHAN}	0.112	0.824	1.395
{T=DT, NOF=no, ES=CS}	{AA=CHAN}	0.107	0.817	1.384
{LOA=Zhenjiang, CF=HF}	{AA=CHAN}	0.196	0.815	1.379
{T=DT, SOC= NEG, ES=CS}	{AA=CHAN}	0.104	0.813	1.376
{T=DT, ES=CS}	{AA=CHAN}	0.115	0.810	1.371
{SW=no, NOF=no, CF=HF, ES=CS}	{AA=CHAN}	0.233	0.809	1.369
{LOA=Zhen-jiang}	{AA=CHAN}	0.226	0.807	1.366
{SW=no, CF=HF, ES=CS}	{AA=CHAN}	0.235	0.807	1.366
{NOF=no, CF=HF, ES=CS}	{AA=CHAN}	0.258	0.805	1.363
{SW=no, NOF=no, ES=CS}	{AA=CHAN}	0.241	0.803	1.359
{NOF=no, ES=CS}	{AA=CHAN}	0.269	0.802	1.357
{SW=no, ES=CS}	{AA=CHAN}	0.243	0.801	1.356
{CF=HF, ES=CS}	{AA=CHAN}	0.278	0.800	1.354
{WP=NS, ES=CS}	{CF=HF}	0.124	0.967	1.182
{WP=FS, ES=CS}	{CF=HF}	0.110	0.964	1.178
{T=NT, SW=no, ES=CS}	{CF=HF}	0.178	0.961	1.174
{SW=no, ES= CS}	{CF=HF}	0.291	0.959	1.172
{ES=CS}	{CF=HF}	0.348	0.957	1.170
{WP=NS, T=NT, AA=CHAN}	{CF=HF}	0.115	0.933	1.140
{WP=DS, AA= CHAN, SW=no}	{CF=HF}	0.147	0.922	1.127
{WP=NS, AA=CHAN}	{CF=HF}	0.194	0.917	1.120

{WP=DS, AA=CHAN}	{CF=HF}	0.174	0.913	1.116

3.5 *Visualization of association rules for ship collisions*

Considering that the "arulesViz" package cannot express the relationship among all association rules, Gephi software is introduced into graphic modeling to make the rules fully visualized. The antecedent and the succedent of the rules are treated as separate nodes, therefore, the 231 rules generate 612 nodes and 1242 edges. In Figure 7, “Force Atlas” is used for layout and “Modularity Class” is used for module division. The size of nodes represents the degree of penetration, and the larger the number of connected nodes, the larger the nodes. Different colors represent different patterns, and there is a closer connection in the same patterns (Weng et al. 2016). As shown in Figure 7, the larger nodes reflect the characteristics of collision accidents more obviously, while some small nodes, which may indicate as ambiguous, is owing to the weak cause-effect relationship. The four large nodes in the figure are {severity of consequences = negligible}, {number of fatalities = 0 fatalities}, {shipwreck = yes}, {Causation factors = human factors}. This further indicates that in the studied 2012-2016 years, the Jiangsu section of the Yangtze River is mainly dominated by negligible maritime accidents, and the human factors should be drawn much attention to enhancing maritime safety.

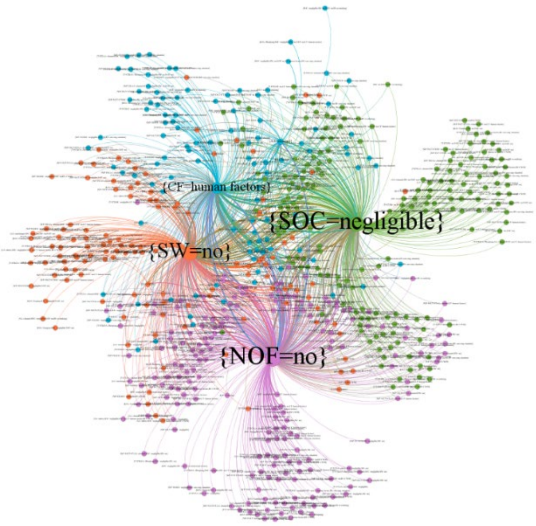

Figure 7. Network visualization of all rules

4 CONCLUSIONS

For exploring the characteristics of ship collision accidents, the association rules method is intuitive to represent the cause-effect relationships. By establishing historical database and using association rules, it is easier to discover the causation patterns of the collision accident. For example, it is found that anchorage and berthing area are more likely to collide with stationary ships; the Zhenjiang channel has a large probability of accident occurrence, and the flooding period may easy to cause collision accidents. These findings of causation factors make the maritime safety administration, the organization in charge of maritime safety in Yangtze River, take the countermeasures to prevent the occurrence of the collision accident.

Although association rules have some advantages in the field of analyzing collision accident, there are still some shortcomings. Specifically, the evaluation standard of historical database has great influence on data results, and when dealing with different problems, the determination of the threshold of the association rules needs further analyzed.

ACKNOWLEDGEMENTS

The research presented in this paper was sponsored by a grant from National Key Technologies Research & Development Program (2017YFC0804900, 2017YFC0804904), a grant from National Science Foundation of China (Grant No. 51809206), Hubei Natural Science Foundation (Grant No. 2017CBF202) and the Hong Kong Scholar Program (NO.2017XJ064).

REFERENCES:

Agrawal, R., & Srikant, R. 1994. Fast Algorithms for Mining Association Rules. *Proceedings of the 20th Very Large Data Bases (VLDB' 94)Conference*: 487-499.

Akten, N. 1999. Analysis of Shipping Casualties in the Bos phorus. *Journal of Navigation* 57(3): 345-56.

Borah, A., & Nath, B. 2018. Identifying Risk Factors for Adverse Dis eases Using Dynamic Rare Association Rule Mining. *Expert Systems with Applications*113: 233-63.

Barati, M., Bai, Q., & Liu, Q. 2017. Mining semantic association rules from RDF data. *Knowledge-Based Systems*133: 183-196.

Chai, T., Weng, J., & Deqi, X. 2017. Development of a Quantitative Risk Assessment Model for Ship Collisions in Fairways. *Safety Science* 91: 71-83.

Fan, S., Yan, X., Zhang, J., & Zhang, D. 2017. Review of Human Factors in Water Traffic Accidents. Journal of Transport Information and Safety (02): 01-08. (In Chinese)

Geurts, K., Thomas, I., & Wets, G. 2005. Understanding Spatial Concen trations of Road Accidents Using Frequent Item Sets. *Accident Analysis & Prevention* 37(4): 787-99.

Grabot, B. 2018. Rule mining in maintenance: Analysing large knowledge bases. *Computers & Industrial Engineering*, in press.

Hu, S., Fang, Q., Xia, H., & Xi, Y. 2007. Formal Safety Assessment Based On Relative Risks Model in Ship Navigation. *Reliability Engineering and System Safety* 92(3): 369-77.

Jiang, X. 2010. Statistics and Analysis of Ship Collision Accidents in Jiangsu Section of Yangtze River in Recent Five Years. China Water Transport (the second half of the month 09): 46-48. (in Chinese)

Kujala, P., Hänninen, M., Arola, T., & Ylitalo, J. 2009. Analysis of the marine traffic safety in the Gulf of Finland. *Reliability Engineering and System Safety* 94(8): 1349-1357.

Ochin, Kumar, S., & Joshi, N. 2016. Rule Power Factor: A New Interest Measure in Associative Classification. *Procedia Computer Science* 93: 12-18.

Sahoo, J., Das, A., & Goswami, A. 2015. An Efficient Approach for Min ing Association Rules From High Utility Itemsets. *Expert Systems with Applications42*(13): 5754-78.

Sedova, N., Sedov, V., & Bazhenov, R. 2018. Neural Networks and Fuzzy Sets Theory for Computer Modeling of Ship Collision Avoidance in Heavy Traffic Zone, *IEEE 2018 International Russian Automation Conference (RusAutoCon)*: 01-05.

Sii, H., Ruxton, T., & Wang, J. 2001. A Fuzzy-Logic-Based Approach to Qualitative Safety Modelling for Marine Systems. *Reliability Engineering and System Safety*73(1): 19-34.

Wang, L., & Yang, Z. 2018. Bayesian Network Modelling and Analysis of Accident Severity in Waterborne Transportation: A Case Study in China. *Reliability Engineering & System Safety*180: 277-89.

Weng, J., Zhu, J., Yan, X., & Liu Z. 2016. Investigation of Work Zone Crash Casualty Patterns Using Association Rules. *Accident Analysis & Prevention* 92: 43-52.

Weng, J., & Li, G. 2017. Exploring Shipping Accident Contributory Factors Using Association Rules. *Journal of Transportation Safety & Security*: 01-22.

Xu, C., Bao, J., Wang, C., & Liu, P. 2018. Association Rule Analysis of Factors Contributing to Extraordinarily Severe Traffic Crashes in China. *Journal of Safety Research* 67: 65-75.

Zhang, J., Teixeira, Â., Soares, CG, & Yan, X. 2018. Quantitative Assessment of Collision Risk Influence Factors in the Tianjin Port. *Safety Science*110: 363-71.