Finding Characteristics of Influencer in Social Network using Association Rule Mining

Watcharapan Ponchai[1], Bunthit Watanapa[2], Kaneungjit Suriyathumrongkul[3]
School of Information Technology, King Mongkut's University of Technology Thonburi, Thailand[1,2]
Cititel Co., Ltd., Bangkok, Thailand[3]
kookengarm@gmail.com[1], bunthit@sit.kmutt.ac.th[2], kaneungjit@hotmail.com[3]

***Abstract*— Relying on the Association Rule principle, this study aims to find significant characteristics of a powerful sharer or influencer who can influence others to actively join the activities (play game or answer quiz and share result) on Dek-D.com, which is a website affiliated with Facebook. Certain parts of users profiles and activities logs are treated as sources of data for the mining. Examples of discovered knowledge are (1) overall, highly influential sharers are those who share in various categories of games or quizzes and are active in both business and prime times, and (2) for specific dimensions of interest, e.g. by gender group as women tend to be influenced by people who regularly share love topic. By age, only teenagers have an influence on teenagers. By age and gender interplay, males, especially adult males, can convince males. These and other findings in this study could be beneficial to marketers for selectively approaching candidates to make effective social marketing campaigns.**

Keywords—Influencer; Social Network Analysis; Association rule mining; Big data analytic;

I. INTRODUCTION

Social Network Marketing has become an important tool for marketers to gain access to target groups of customers more effectively. It also provides lower cost when compared with other traditional approaches since social media become a norm in daily life of people. Nowadays, the number of social network users is still skyrocketing [11].

Social networks create an active communication channel between marketers and customers. It can cause a viral marketing. The routine process of this marketing is started by selecting the first group of people to receive the message about products and services. After a moment, each will swiftly spread the message to other people by word-of-mouth.

Finding the group of people who can spread the message effectively is an important step in social network marketing. Marketers attempt to find the characteristics of an influencer who can get the attention of followers.

Previous research has usually used the Social Network Analysis model (SNA). This model represents a networked structure of individual actors (each node) and perceived relationships (edges) that interconnect them. There are 3 popular centrality measures: Degree Centrality calculated by number of degree in each node, Closeness Centrality representing the relationship in terms of shortest path of network and Between Centrality representing the number of connected network at each node [1].

Data mining has been applied in social network analysis to identify influencers for a decade or so [2]. Among popular ones, Association Rule Mining is a method for analyzing and interpreting data in a perspective of knowledge discovery. It can find some hidden relationships and create new knowledge based on the voluminously recorded transactions or logged activities.

This research proposes a data mining approach to find characteristics of great influencers on real data of user profiles and activity logs collected from the web application named FunnyQuiz on Dek-D.com website.

The rest of this paper is presented in the following structure. Firstly, the review of previously related work is presented. Section 3 provides information about the methodology of this research. Next, experiment and result insights are discussed before making conclusion in the final section.

II. RELATED WORK

The survey of related literatures discloses that a variety of research approaches have been performed to find the characteristics of an influencer. On one direction, Aral et al. (2012) found that the sensitiveness to convincing is up to gender and age. Older people are more difficult to influence. When comparing men and women, men are more convincing. Interestingly, women are good influence on men, but not on those of the same gender [3]. A year later, by leveraging the power of social media, Babutsideze et al. (2013) used a questionnaire on Facebook that is equipped with a feature called "Invite Mechanism" that enables the users to invite their friends to answer the questionnaire. The data analyzed by statistical methods shows that younger women are often at important positions (or strategic positions) and have a rather high influential degree in social network [4].

On the other directions, many researches attempt to determine influence measured by network location. Lim et al. (2011) proposed guidelines to search the content creators who influence other users called "Content Power Users (CPUs)" by calculating the centrality in 3 forms such as Degree Centrality, Closeness Centrality and Between Centrality. They believed that users at the center position are likely to be influencer. The concern with this approach is the complexity of computation [5]. To address the issue, Han et al. (2012) applied a relaxed model called "Random Walk" [6]. Heidemann et al. (2010) proposed an algorithm called "PageRank" using scoring method to rank the affiliating value of a web page based on the probability that users will click on the links on the corresponding page. They also proposed a model applying this concept to identify the key users on social networks [7].

Some researches rely on the user behaviors. Trusav et al. (2010) found the times of use (Time spent) such as number of log-ins per day as an influential factor [8]. In addition, Bakshy et al. (2011) analyzed the influence of sharing URLs. He believes the logic that if Mr. A share URL to Mr. B and Mr. B shares that URL to others, then Mr. A has an influence on Mr. B [9]. While Cha et al. (2010), measured the influence of 3 attributes: Followers, Retweet and Mention. They found that the number of followers is not the cause of increment in the number of shares or referring. They concluded that the influential power is gained from continuous sharing in the same topics [10].

In 2013, Probst et al. (2013) categorized the characteristics of influencers in social networks into 4 groups. The first is the personification of each person called "who one is". The second is the ability or knowledge called "what one knows". The third is the strategic position in social networks called "whom one knows". The last is the relationship between the positions of a person in the network and the activity called "how active one is" [1].

Using Probst's concept, we can summarize the inclusion of factors determining influencer accomplished by existing literature and our proposed research as shown in TABLE I.

TABLE I. SUMMARY OF THE FACTOR INCLUSIONS OF EACH RESEARCH

Researcher	**Method**	**Who one is**	**What one know**	**Whom one knows**	**How active one is**
S. Aral et al. [3]	Gender and age	✓			
Z. Babutsideze et al. [4]	Centrality measure with gender and age	✓		✓	
S. Lim et al. [5]	Centrality measure			✓	
B. Han et al. [6]	Random Walk algorithm			✓	
J. Heidemann et al. [7]	PageRank			✓	
M. Trusav et al. [8]	Time spent activity				✓
E. Bakshy et al. [9]	URL cascade				✓
M. Cha et al. [10]	Follower Retweet and Mention		✓		✓
This Research	**Association Rule in characteristics**	✓	✓		✓

III. METHODOLOGY

By applying the Association Rule Mining, this study aims to find the relationship between prospective characteristics (listed in TABLE V) and the degree of influence. The obtained relationships will be interpreted by the mined rules which are in the forms of antecedence (Left Hand Side of the rule, or LHS) and the consequence (Right Hand Side of the rule, or RHS). The process of the mining in this particular study can be broken down into steps of selecting data, preparing data, transforming data, and inferring interesting rules steps.

A. Data Selection

For this study, we selected the user data from FunnyQuiz of the Dek-D.com having Facebook as a affiliating gateway for social networking. As shown in Fig. 1, data collection starts when a user called "sharer" answers a quiz. Number of shares per dimension of interest, namely, parts of the day, gender, age range, and overall are incrementally computed and stored as defined in TABLE II. Then the quiz result is shared on sharer's News Feed (Fig. 2). Sharer ID is attached to the URL for reference. Sharer's friends (called "viewer") who view the posts can click URL to answer quiz. Doing so will automatically turn the "viewer" to "responder" and increase degree of influence for this sharer. Finally, the responder becomes a sharer and shares quiz result to other responders. This process will continuously occur as cycles as shown in Fig. 1.

Fig. 1. The process of data collection in Dek-D.com system

Fig. 2. Sample post of FunnyQuiz result in Facebook news feed

TABLE II. CHARACTERISTICS OF EACH SHARER

Characteristic	**Description**
Facebook ID	User ID in Facebook
Gender	Gender of sharer
Age	Age of sharer (calculated from birthday)
Number of shares per category	Number of shares in each category of quiz in 5 category. (individuality, love, entertainment, knowledge and hobby)
Number of shares per time range	Number of shares in 3 periods of time (morning time, business time and prime time)
Total number of shares	Number of shares in overall or share frequency
Responses	Number of responders for each of the 4 dimensions of interest: overall, age ranges, gender and share time

B. Data Preprocessing

After having collected data for 4 months (October 2014 – January 2015), we get 390,141 user profiles in total. Although it comprises a large dataset, most data are with low degree of influence and, hence, not much meaningful for mining purpose. The skewed distribution caused us to adopt only data with values above mean (15 responses in 4 months). We also filtered incomplete data out and make the candidate dataset remains only 38,152 rows, or 10% of all data.

C. Data Transformation

Some attributes, e.g. age and time, are continuous data that are unusable in Association Rule Mining. We converted them to discretely nominal values. Other scale values (e.g. number of shares) are dichotomized (High or Low degree) using Median as a threshold. (refer to TABLE III and TABLE V)

TABLE III. DATA TRANSFORMATION

Continuous Data	**Period**	**Discrete Data**
Age	1 - 5 years old	Early childhood
	6 - 12 years old	Child
	13 - 18 years old	Teen
	19 - 30 years old	Adult
	31 - 60 years old	Middle age
Share time	01.00 – 08.00	Morning time
	09.00 – 16.00	Business time
	17.00 – 00.00	Prime time
Scale values	>= Median	High
	< Median	Low

Noted that the transaction data, including the derived number of responders and number of shares had been derived and archived by a special software without date stamper. This limitation causes bias on all frequency-oriented data, esp. unequal length of time each user had been in the system. The adjustment schemes as shown in Equation (1) and (2) are introduced for mitigating the effect. By this scheme, all values of number of shares and responders in any group of interests are transformed into a closed range of [0, 1] as shown in TABLE V.

$$\text{Adjusted number of shares} = \frac{Number\ of\ shares}{Total\ number\ of\ shares} \qquad (1)$$

$$\text{Adjusted number of responders} = \frac{Number\ of\ responders}{Total\ number\ of\ shares} \qquad (2)$$

D. Finding Interesting Rules

By principle of Association rule analytic, a set of three threshold values are determined to ensure the statistical significance or the mined rules of the required form of LHS implies RHS (LHS = Left-Hand-Side and RHS = Right-Hand-Side). (1) Support, or sup(X=>Y), is defined as number of instances which contain LHS (or Antecedent) and RHS (or Consequence). It is usually represented in percentage of the total number of instances in the system. (2) Confidence, or Conf(X=>Y) is defined as the probability of occurring RHS given the LHS is occurred. (3) Lift, or lift(X=>Y), measures whether the occurrence of LHS and that of RHS are independent of each other or not. Lift =1 means the two events of LHS and RHS are independent of each other and hence no rule can be inferred. On the contrary, Lift > 1 implies that there are dependency between the two events. The higher the Lift is, the more meaningful the interpretation of the relationship of LHS and RHS will be [25]. The computational formula for the three measures are shown in TABLE IV.

$$\text{LHS} \Rightarrow \text{RHS} \qquad (3)$$

TABLE IV. COMPUTING FORMULA FOR THE THREE THRESHOLD VALUES

Thresholds	**Formula**	**Values**
Support	P(LHS and RHS)	0 - 100 %
Confidence	$\frac{\text{P(LHS and RHS)}}{\text{P(LHS)}}$	[0, 1]
Lift	$\frac{\text{P(LHS and RHS)}}{\text{P(LHS) P(RHS)}}$	[0, +∞)

Where P(X) = probability of X

To comprehend the mining results into insightful rules for seeking characteristics of influencer, the Apriori algorithm was adapted to 6 steps procedure as listed below:

1) Load the preprocessed data for each dimension of interest to the system, namely, Age (only Teen up to Middle Age), Share time, Gender, and Overall.

2) Perform association rules mining using WEKA[1] package, with control parameters of minimum support = 10%, minimum confidence = 70% and lift > 1.1.

3) Select only the rules which RHS is the number of responders for the corresponding group of interest.

4) Divide rules into 2 groups by confidence value. The first group is the top-notch with confidence of 80% or higher.

5) Sort rules by the number of factors on LHS to systematically gain insights on the relationship between characteristics and responses.

6) Infer meaningful rules (for each group of interests in TABLE V) which indicate characteristics of users as antecedence (LHS) and the level of responses as consequence (RHS).

[1] WEKA is a collection of open source machine learning algorithms of the University of Waikato, New Zealand [21].

TABLE V. AVAILABLE CHARACTERISTICS AND VALUES

Characteristic Group	**Characteristic**	**Adjusted Values**	**Median**
Demographic	Gender	Male, Female	-
	Age	Child, Teen, Adult, Middle Age	-
Number of share category (Interests)	Individuality	[0, 1]	0.396
	Love	[0, 1]	0.258
	Entertainment	[0, 1]	0.1
	Knowledge	[0, 1]	0.053
	Hobby	[0, 1]	0.357
	Diversity	[0.002, 1]	0.294
Number of share time per interval	Morning	[0, 1]	0.417
	Business	[0, 1]	0.317
	Prime	[0, 1]	0.581
	Diversity	[0.002, 1]	0.386
Number of responders in each group of interest	Overall	[0.021, 109]	1.625
	Female	[0, 106]	0.92
	Male	[0, 51]	0.5
	Teen	[0, 49]	0.313
	Adult	[0, 87]	0.632
	Middle Age	[0, 38]	0
	Morning time	[0, 33]	0.09
	Business time	[0, 57]	0.462
	Prime time	[0, 86]	0.846

IV. EXPERIMENT RESULT AND DISCUSSION

The experiment had been conducted according to the guideline given in Section III and yielded interesting results as follows.

A. High confidence group (Conf. >= 0.8)

The group has the rules with the most confidence (>= 0.8) obtained in this experiment. There are only rules with 2 characteristics on LHS: topic of interest and share time.

As shown in TABLE VI, the implication of the findings could be interpreted as follows. People who are interested in various topic and share only a few in the topics of hobby get a high response (conf.=0.82, lift=1.64, sup.=22%). In women group, people who share topic of love have more influence (conf.=0.82, lift=1.64, sup.=12%). The adult group is less interested in the topic of knowledge than people in the other age ranges (conf.=0.81, lift=1.62, sup.=11%).

Diversity of share time is another major factor that impacts the degree of influence. The appropriate sharing times are business time and prime time (TABLE VII).

TABLE VI. RULES OF INTEREST WITH HIGH CONFIDENCE

LHS						**RHS**	**Conf.**	**Lift**	**Sup.**
Interests					*Diversity*	*Response*			
Ind.	*Lov.*	*Ent.*	*Kno.*	*Hob.*					
				Low	**High**	**overall = High**	0.82	1.64	22%
			Low	Low	**High**	**overall = High**	0.88	1.75	11%
			Low	**High**	Low	overall = Low	0.81	1.62	12%
				Low	**High**	**female = High**	0.8	1.61	22%
	High			Low	**High**	**female = High**	0.82	1.64	12%
	Low			**High**	Low	female = Low	0.81	1.63	13%
			Low	Low	**High**	**adult = High**	0.81	1.62	11%

TABLE VII. RULES OF SHARE TIME WITH HIGH CONFIDENCE

LHS				**RHS**	**Conf.**	**Lift**	**Sup.**
Share time			*Diversity*	*Response*			
Morning	*Business*	*Prime*					
Low			**High**	**overall = High**	0.81	1.61	18%
Low	**High**		**High**	**overall = High**	0.8	1.6	10%
Low		**High**	**High**	**overall = High**	0.8	1.6	11%

B. Medium Confidence Group (0.7 <= Conf. < 0.8)

In this group, each rule has medium confidence. Some rules overlap in the high confidence group. There are many rules that comprise 3 characteristics: demographic, topic of interests and share time. These rules effectively identify focal points of the potential influencers.

In demographic, males have an influence on males, especially those who are in adult age (conf.=0.77, lift=1.49, sup.=21%). On the other hand, Female teens get low responses from male group (conf.=0.72, lift=1.5, sup.=11%). Teenagers are always attractive in teen group (conf.=0.72, lift=1.43, sup.=18%). In addition, Female teens have low influence in adult group (TABLE VIII).

TABLE VIII. RULES OF DEMOGRAPHICS WITH MEDIUM CONFIDENCE

LHS		**RHS**	**Conf.**	**Lift**	**Sup.**
Demographic		*Response*			
Gender	*Age*				
male		**male = High**	0.75	1.46	32%
male	adult	**male = High**	0.77	1.49	21%
female	teen	male = Low	0.72	1.5	11%
	teen	**teen = High**	0.72	1.43	18%
	teen	adult = Low	0.75	1.51	19%
female	teen	adult = Low	0.79	1.59	12%

From TABLE IX, diversity of share category proportionately impacts degree of influence. This is in line with the result in the first group. Males are rarely interested in people who share the topic of love (conf.=0.72, lift=1.4, sup.=28%). People prefer the sharing in topic of individuality in business time. (conf.=0.7, lift=1.4, sup.=15%). Lastly, people prefer sharing the topic of love in prime time (conf.=0.71, lift=1.42, sup.=22%).

TABLE IX. RULES OF INTEREST WITH MEDIUM CONFIDENCE

LHS						**RHS**	**Conf.**	**Lift**	**Sup.**
Interests					*Diversity*	*Response*			
Ind.	*Lov.*	*Ent.*	*Kno.*	*Hob.*					
	Low				**High**	**male = High**	0.72	1.4	28%
	High		Low		Low	male = Low	0.71	1.47	16%
				Low	**High**	**adult = High**	0.75	1.51	27%
High					**High**	**adult = High**	0.71	1.42	25%
		Low			**High**	**adult = High**	0.7	1.39	24%
				High	Low	adult = Low	0.75	1.5	27%
		High			Low	adult = Low	0.7	1.4	24%
Low				**High**	Low	adult = Low	0.76	1.51	14%
				Low	**High**	**time_business = High**	0.7	1.41	27%
		Low		Low	**High**	**time business = High**	0.72	1.43	14%
High				Low	**High**	**time_business = High**	0.7	1.4	15%
				High	Low	time_business = Low	0.7	1.41	27%
		High		**High**	Low	time_business = Low	0.7	1.39	15%
				Low	**High**	**time_prime = High**	0.75	1.5	27%
	High				**High**	**time_prime = High**	0.71	1.42	22%
				High	Low	time_prime = Low	0.73	1.47	27%
	Low			**High**	Low	time_prime = Low	0.77	1.53	16%

In the dimension of share time, as shown in TABLE X, most rules conform with the findings in the first group except that people who are active in the morning seem to be a niche group. They prefer the people who play and share in the morning part of the day (conf.=0.72, lift=1.42, sup.=13%).

TABLE X. RULES OF SHARE TIME WITH MEDIUM CONFIDENCE

LHS				**RHS**	**Conf.**	**Lift**	**Sup.**
Share time			*Diversity*	*Response*			
Morning	*Business*	*Prime*					
Low			**High**	**female = High**	0.75	1.49	17%
			High	**male = High**	0.7	1.36	35%
Low			**High**	**male = High**	0.72	1.4	16%
			High	**adult = High**	0.72	1.44	36%
			Low	adult = Low	0.72	1.44	36%
High			**High**	**time_morning = High**	0.71	1.41	20%
High	Low		**High**	**time_morning = High**	0.75	1.49	11%
High		Low	**High**	**time_morning = High**	0.72	1.42	13%
Low			Low	time_morning = Low	0.72	1.45	20%
Low	**High**		Low	time_morning = Low	0.75	1.51	10%
	High		**High**	**time_business = High**	0.76	1.53	20%
		Low	**High**	**time_business = High**	0.76	1.52	21%
	Low		Low	time_business = Low	0.74	1.48	20%
	Low	**High**	Low	time_business = Low	0.76	1.51	17%

V. CONCLUSION AND FUTURE WORK

This study proposes an analytic process for mining the association or relationship between characteristics of users and their influenceability on social media based on the Association Rule principle. A 4-month activity logs and user profiles on the selected web application, Dek-D.com connected with Facebook as a gateway to social network, were archived for the experiment. The obtained results show many valuable relationships which may be directly beneficial to the marketing planners of Dek-D.com in terms of market positioning and activities promotion strategies. A sample of interesting findings is that, overall with about 80% confidence, influencer are people who are active at various times of the day and involve assorted categories of games and quizzes. A better focus for positioning can be achieved when combining another two rules in the same confidence level: (1) influential persons share very little in hobby and knowledge categories and (2) they rarely share in the morning. On a particular group, females are influenced by the people who regularly share the love category. If accepting confidence of 70% or lower, interesting relationships are found in the dimension of age and gender. Only teenagers can attract teenagers and female teenagers cannot influence older males or females. Males are influenced by males, especially those who are adult age.

We are in the direction to extend the proposed model by integrating the centralities (products of social network analysis) as a new factor of “whom one knows” characteristics on the LHS. This shall enable our research to cover all possible types of characteristics, according to Probst’s definition.

REFERENCE

[1] F. Probst, L. Grosswiele, and R. Pleger, “Who will lead and who will follow : Identifying Influential Users in Online Social Networks”, Business & Information Systems Engineering, Vol.3, pp. 179-193, 2013.

[2] M. Adedoyin-Olowe, M. Medhat Gaber, and F. Stahl, 2013, “A Survey of Data Mining Techniques for Social Media Analysis”, Social and Information Networks, Vol.2.

[3] S Aral. and D. Walker, 2012, “Identify Influential and Susceptible Members of Social Networks”, SCIENCE, Vol.337, pp.337-341

[4] Z. Babutsideze, T. Lomitashvili and K. Turmanidze, 2013, “Influential Individuals on Social Networks : An Example of Facebook in Georgia”, European Scientific Journal, Vol.4, pp. 320-330.

[5] S. Lim, S. Kim, S. Park and J.H. Lee, 2011, “Determining Content Power Users in a Blog Network: An Approach and Its Applications”, IEEE Transactions on Systems, Man, and Cybernetics – part-A: System and Human, Vol41, NO.5, pp. 853-862.

[6] B. Han and A. Srinivasan, 2012, “Your Friends have More Friends Than You do-Identifying Influential Mobile Users Through Random Walk”, ACM

[7] J. Heidemann, M. Klier and F. Probst, 2010, “Identifying Key Users in Online Social Networks: A PageRank Based Approach”, Proceedings of the 31th International Conference on InformationSystems, Saint Louis, Missouri, pp. 1-22.

[8] M. Trusav, A.V. Bodapati and R.E. Bucklin, 2010, “Determining Influential Users in Internet Social Networks”, Amercian market Association, pp. 643-658.

[9] E. Bakshy, J.M. Hofman, W.A. Mason and D.J. Watts, 2011, “Everyone's an influencer: quantifying influence on twitter”, WSDM '11 Proceedings of the fourth ACM international conference on Web search and data mining, pp.65-74.

[10] M. Cha, H. Haddadi, F. Benevenuto and K.P. Gummadi, 2010,“Measuring User Influence in Twitter: The Million Follower Fallacy”, ICWSM ’10: Proceedings of international AAAI Conference on Weblogs and Social

[11] Statista, 2014, Social networks: global sites ranked by users 2014 | Statistic, Available : http://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/ [2014, October 25].

[12] Internetlivestats.com, 2014, Number of Internet Users (2014) - Internet Live Stats , Available : http://www.internetlivestats.com/internet-users/ [2014, October 25].

[13] C. Kiss and M. Bichler, 2008, “Identification of Influencers – Measuring Influence in Customer Networks”, Decision Support Systems, Vol.46, pp. 233-253.

[14] P. Tan, M. Steinbach and V. Kumar, 2006, Introduction to Data Mining, Addison-Wesley Companion Book , Ch.6, Available: http://www-users.cs.umn.edu/~kumar/dmbook/index.php

[15] G. Shmueli, N.R. Patel, P.C. Bruce, 2010, Data Mining for Business Intelligence, Wiley & Sons, Inc., pp.13 / pp.263.

[16] H. Wu, I. Ting and K. Wang, 2009, “Combing Social Network Analysis and Web Mining Techniques to Discover Interest Groups in the Blogspace”, Innovative Computing Information and Control ,pp. 1180-1183.

[17] Orgnet.com, 2013, Social Network Analysis, A Brief Introduction, Available : http://www.orgnet.com/sna.html [2014, October 25].

[18] Marketing-schools.org, Influencer Marketing, Available : http://www.marketing-schools.org/types-of-marketing/influencer-marketing.html [2014, October 25].

[19] Zentut.com, What is Data Mining, Available : http://www.zentut.com/data-mining/what-is-data-mining/ [2014, October 25].

[20] Gephi, Gephi - The Open Graph Viz Platform, Available : http://gephi.github.io/ [2014, October 23].

[21] University of Waikato, Weka 3: Data Mining Software in Java , Available : http://www.cs.waikato.ac.nz/ml/weka/ [2014, October 23].

[22] B.V. Carolan, 2014, Social Network Analysis and Education, SAGE Publications, Inc, California, pp.1-6, pp.154

[23] D.M. Scott,2010, The New Rules of Marketing & PR, John Wiley & Sons, Inc., Hoboken, New Jersey, pp. 173.

[24] Education-portal.com, Havighurst's Developmental Task Model, Available : http://education-portal.com/academy/lesson/havighursts-developmental-task-model.html#lesson [2014, December 27].

[25] L. M. Sheikh, B. Tanveer and S. M. A. Hamdani, 2004, "Interesting Measures for Mining Association Rules", INMIC IEEE, pp.641-644

[26] J. Nahar, T. Imam, K. Tickle and Y. Chen, 2013, "Association rule mining to detect factors which contribute to heart disease in male and females", Expert System with Applications, pp. 1086 - 1093