Identification of Opinion Leaders and Followers in Social Media

Chun-Che Huang[1], Li-Ching Lien[1], Po-An Chen[1], Tzu-Liang Tseng[2] and Shian-Hua Lin[3]

[1]*Department of Information Management, National Chi Nan University, Puli, Taiwan, R.O.C.*
[2]*Department of Industrial, Manufacturing and Systems Engineering, The University of Texas at El Paso,*
El Paso, TX, U.S.A.
[3]*Department of Computer Science and Information Engineering, National Chi Nan University, Puli, Taiwan, R.O.C.*

Keywords: Opinion Leaders and Followers Identification, Matrix Analysis, Social Media, Web Analytics.

Abstract: In recent years, with the development of Web2.0, opinion leaders on the Web go up onto the stage and lead the will of the people. Many time, government, private companies and even traditional news media need to understand the opinion leaders' ideas on the Web. Identifying opinion leaders and followers becomes a very important study. To study the characteristics of opinion leaders and the impact of opinion leaders on followers, our research evaluates whether every speaker in social media satisfies characteristics of opinion leader. The characteristics of opinion leader and relationship between opinion leader and follower are studied. By observing relational matrix, the interacting relations between users in social media are analysed and opinion leaders and followers are identified.

1 INTRODUCTION

Nowadays the opinion leaders have significant impact on the Internet. To recognize the opinion leaders' ideas is crucial from government, private companies to traditional news media. Spanish scholar Ramonet (2003) proposed that there should be the fifth right, a public power to balance the increasingly chaotic fourth right. In 'Towards a Civil Society' seminar in 2008, when discuss the role of blog in improving social development and discussion promotion, Blog is defined as the fifth right.

New network media such as Facebook, Twitter and Taiwan local PTT as the fifth right have huge users in discussing different issues. In these forums, there are many opinion leaders with professional knowledge to provide opinions to discuss. In this way, there are many followers who could be influenced by opinion leaders. However, the studies related to opinion leader identification emphasize on business sales rather than opinion leaders of social policies. Previous studies utilize quantitative methods to analyse and without considering from the qualitative perspectives. It is difficult to accurately identify the potential opinion leaders and followers without exploring the relationship between opinion leaders and followers.

This study surveys literature in Section II. Section III proposes a methodology to identify the opinion leaders and followers. Based on opinion leaders' social network characteristics and whether their speeches could get resonation, the matrix approach is proposed to model the relationship between users and identify opinion leaders set Section IV concludes this study. In addition, this study contributes offering three patterns to help the enterprises/government to implement their corresponding strategy. For example of general pattern, business take advantages of opinion leaders to change follower behaviour; therefore, their products/services can be promoted effectively. According to Discussion pattern, the opinion leaders and followers is not obvious such that business can ignore them to decrease investment. For example of blindly pattern, business can focus on negotiation professional to improve recruiting the opinion leaders and followers.

2 LITERATURE REVIEW

The word 'opinion leader' comes from study of president election in 1940 by Lazarsfeld, Berelson, and Gaudet (1948). They observed the behaviors of voters in election and found that the election advertisements were not as effective as opinion

leaders. However, opinion leaders are defined broadly. Opinion leaders are very important in social network because they are capable to informally influence others' attitudes and behaviors (Cho, Hwang & Lee, 2012, Li & Du, 2011; Li, Xing, Wang, & Ren, 2012,; Venkatraman, 1989). Opinion leaders usually have access to far more information on certain topic and have profession in this topic. These people are usually core of the social network and easy to get along with. The roles of opinion leaders vary by different themes, fields, cultures, social environments and eras (Weimann & Brosius, 1994). Those who receive the information and change their behaviors and attitudes are called followers (Burt, 1999). Compared with opinion leaders, the followers are not as important. In other words, the followers are those who will be influenced by opinion leaders and change their attitudes and behaviors. (Li et al., 2012)

According to the study listed above, we can conclude the properties of opinion leaders in the following:

1. The life experience is rich and the understanding of the knowledge is though. Many of them are highly educated.
2. Strong social skills, strong connection with broad masses, good reputation due to professional and knowledgeable. They have great influence and appealing power.
3. Sensitivity to the information, willingness to accept new things, with innovative spirit.

In this study, we define the *opinion leaders* are those who have a lot of followers. They are celebrities or important leaders in some areas. They have good reputation and higher education level. So their speeches and behaviors could influence others. And the followers are defined as those who are usually influenced by opinion leaders' speeches and behaviors. As a result, the followers will change their attitudes on the issues.

In previous studies, Ma and Liu (2014) used SuperedgeRank algorithm to analyze 3 seed network property. This was applied on identifying opinion leaders in Fukushima nuclear issue. Cho et al. (2012) analyze the role of opinion leaders in marketing strategy with innovatively spreading social network research. Leal, Hor-Meyll, and de Paula Pessoa (2014) use interviews to study how opinion leaders influence users' purchase decision in virtual communities. Li and Du (2014) propose a framework by building key words base and apply this framework to identify opinion leaders in Twitter users. William, McMurray, Kurz and Lambert (2015) use key words labels and social network analysis to analyze the behaviors and attitudes of Twitter users towards climate change. In summary, most previous studies use methods such as social measure and self-report which are based on questionnaire or interview to identify opinion leaders. Few of them are based on articles published by opinion leaders or social pattern of opinion leaders and followers. Our study proposes a method to analyze the articles published by the users, users' characteristics and their interactive relationship. Combined with relational matrix method, we identify opinion leaders and followers as well as social pattern when they interact with each other.

3 METHODOLOGY

In this study, the users' published contents is classified according to expert's judgement (Table 1). Positive speech means that the spokesman's statement is the same with responder's attitude, the opinions are the same. Meaningless speech means the statement and spokesman are unrelated or undetermined. Negative speech means spokesman's statement are different from responder's attitude, the opinions are different.

Table 1: Speech Contents Classification.

Contents Classification	
Speech Contents	Score
Uniform Speech	+1
Meaningless Speech	+0
Not uniform Speech	-1

Next, the relationship between users is represented by a matix formation. Matrix M is a relational matrix between users. The row of matrix is defined as i, the column of the matrix is defined as j. The count of users are set as C={1 … n}. The elements in matrix T are count of responses and being responded. The elements in matrix I are influence factor between users and social community support level.

$$M=[T|I]$$

Matrix T represents count of responses and being responded between users. T_{n*n} represents count of responses between each pair of users.

n: count of users.

T_{ij} : count of user's speech.

i: index of the user who responds to other people's opinion.

j: index of the user whose opinion is responded.

Matrix I represents users' power to influence and being influenced and social community support level.

I_{ij}: social community support level of the user.

i: index of the user who influences to other people.
j: index of the user who is influenced.

In matrix T, $i \neq j$, then T_{ij} means counts of responses of i to j. When i=j, relationship T_{ij} means counts of speech of i. In matrix I, when characteristic and property $i \neq j$, I_{ij} means influence of i on j. In this study, the influence of i and j can be classified into 3 different layers according to job title, professional knowledge and social community support level. Each criteria have three levels, no influence (NI), general influence (GI), high influence (HI).

Table 2: Influence Power Comparison Table.

Knowledge status	Social Status			Social Community Support
	low	General	high	
Wrong	NI	NI	NI	Low Social Community Support
Popular	NI	NI	GI	
Professional	NI	GI	GI	

Knowledge status	Social Status			Social Community Support
	low	General	high	
Wrong	NI	NI	GI	Neutral Social Community Support
Popular	NI	GI	GI	
Professional	GI	GI	HI	

Knowledge status	Social Status			Social Community Support
	low	General	high	
Wrong	NI	GI	HI	High Social Community Support
Popular	GI	HI	HI	
Professional	HI	HI	HI	

In this study, we compare and analyze two users based on these three criteria. In table 2 influence comparison table, we compare influence relationship between users. For the higher job title, professional knowledge or social community support level, the influence power is high influence (HI). For the general job title, professional knowledge and social community support level, the influence power is general influence (GI). For those with no job, inaccurate professional knowledge or fair social community support level, the influence power is no influence (NI). Note that when compare these users, we could only analyze the users' speech contents and professional knowledge. So in our study, we need not to complete data for each user. We can list the part we could not get or judge as missing. The missing part is listed as blank.

In the Influence Power Comparison Table, the Social Community Support refers to the power of the forum and how close between the user and the media. For example, if users can easily get and pass the message to the media, we define the user has high social community. Besides, if a particular user posts the article that are read by large of users, we also defined it is high social community.

In matrix I, when characteristic and property are equal i=j, then the relationship I(ij) means the social community support level of this user. This study classifies social community support into four levels. For the well-known, well followed social media, it is classified as social community support high (H) (Table 3). For the less-known, less followed social media, it is classified as low (L). However, due to some users' privacy setting or anonymous discussion, our study could not know this user's social community support level. Our study defines these users as social community support level missing (O).

After defining influence power, compared with expert justified speech counts in table 1 classification of speech contents, our study lists the user with negative speech count as non –follower. At the same time, if the contents being responded is negative, it is listed as Non-opinion leader.

Table 3: Social Community Support Level.

Social Community Support Level	
Category	Code
Low Social Community Support	L
Neutral Social Community Support	N
High Social Community Support	H
Missing	O

In this study, there are three ways to show the social relationship between opinion leaders and followers in relational matrix, as shown in below social community pattern.

Social community Pattern:
  General Pattern:
    In this pattern, the opinion leaders usually have high social community support and their posts are very professional. So they can broadly influence many followers. This is most common pattern in general social community.
  Discussion Pattern:
    The appearance of discussion pattern comes from discussion space provided for social community platform users. Users may propose different specific advice when discuss with each other and thus influence other users.

Blindly Follow:

This pattern comes into exist because the followers follow the opinion leaders by their personal charm. Generally, followers in this pattern do not care about the correctness of the contents published by opinion leaders.

This study defines the opinion leader and follower identification problem as searching of two mutual influenced relational matrices, where the speech counts and influence power in two matrices are separated following the limitations:

(1) The users with negative count of responses are not followers.

(2) Consecutive speech is viewed as one speech.

(3) Users' speech contents should be first justified by experts.

(4) The influence power of users' speech is decided by speech contents and social community support level.

(5) The high social community support between users is judged as "GI" or "HI"

(6) The low social community support is judged as "NI"

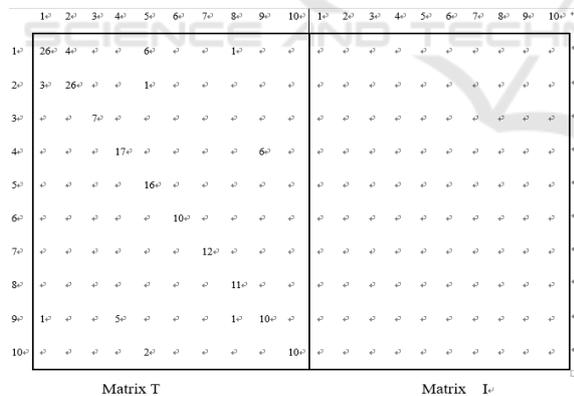Next, the solution approach to resolved the problem is proposed:

Figure 1: Matrix modeling.

Step 1 Crawling data: According to forum (fan pages, posts) in social media according to a particular topic, data are crawled including text and responses information.

Step 2 Convert data to the matrix formulation: Crawled data are cleaned and converted to matrix T from social media according to counts of interaction between users (Figure 1).

Step 3 Remove meaningless users: Remove meaningless users from matrix. The users with no speech and never being responded are removed (Figure 2).

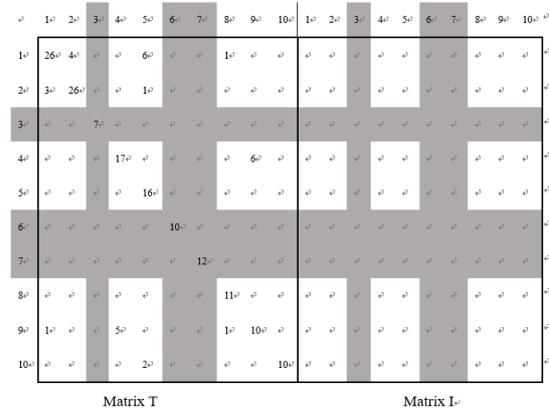Figure 2: Remove meaningless users.

Step 4 Label users' class: Label users' social community support level based on collected materials and put into matrix T according to table 3 social community support level (Figure 3).
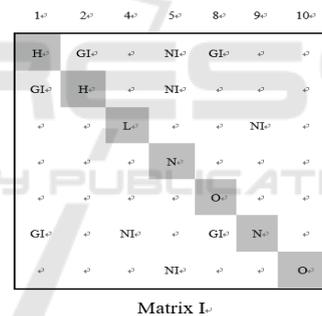

Figure 3: Label users' class.

Step 5 Influence power analysis: According to influence power level in table 2 influence power comparison, analysis the users' influence power based on experts' judgement and input the influence power into matrix I (Figure 4).

Step 6 Sorting: Based on counts of speeches, responses, being responded and influence power, move forward the related users. The users with greater counts of responses and being responded are listed in the more front.

Step 7 Grouping: Base on Clustering Identification (CI) Algorithm, the Group are formed. Group users in matrix T and I by the users' relationship. When influence power between users is high or there are responses between users, we classify these users in

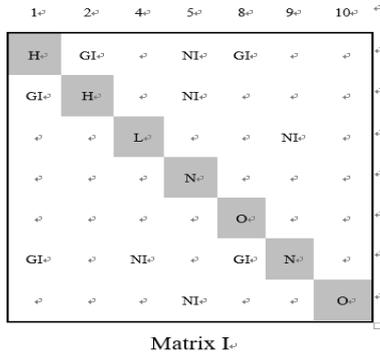the same group. If there is no relationship, we do not group them.

Figure 4: Influence power analysis.

Step 8 Identification: In each group, based on opinion leaders' definition in this study, define interaction between users, opinion leaders and followers.

At final, the ABC network platform is used as an example to explain application of our study method in practical. The platform is open since 02/10/2015. This example is based on national energy conference held by bureau of energy, ministry of economic affair. After four months' discussion, this national energy conference produces a great of common opinions and other opinions. This study utilizes relational matrix to analyze green energy related issues in ABC network platform. The users in ABC network platform could use anonymous style to discuss or use Facebook, Google or Yahoo and so on social media username and password to log in. This study collects, classifies and analyzes the discussion data in 'Stable Energy' issue on ABC network platform.

After remove meaningless users, M=[T|I] (partial data in Figure 5&6) is formulated. Note that the related users is moved forward based on count of users' speech, responses, being responded and influence power. The users with greater count of responses and being responded are listed more front.

In Step 7, Group A {1, 2, 3, 4, 5, 6, 18, 20, 21, 24}, Group B {2, 16, 24, 26, 27, 34, 36}, Group C {17, 18, 19}, Group D {19, 20, 21, 36}, Group E {2, 27, 28, 29, 30, 36}, Group F {7, 20, 26}, Group G {2, 9, 27, 31, 34}, Group H {7, 27}, Group I {7, 8, 15}, Group J {21, 22, 23}, Group K {25, 26}, Group G {34, 35} are formed. In Step 8, for example of Group A, both response and responded contents of user 1 have quite influence power.} So user 1 is an opinion leader. User 18, 24, 21, 3, 5 and 6 are positive towards user 1's opinion and they are followers. There are discussions in this group, so this is a discussion pattern (Figure 6).
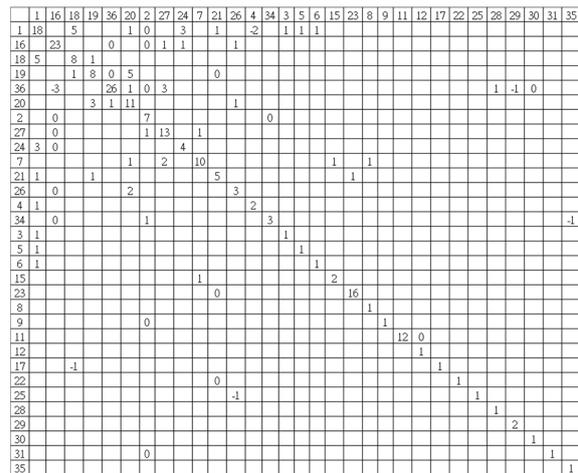

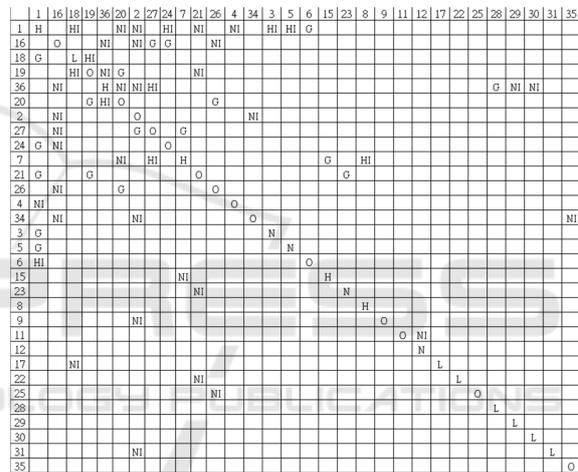Figure 5: Matrix elements conversion.
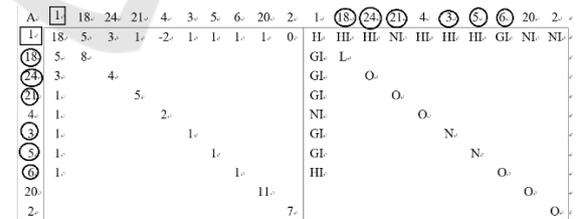

Figure 6: Matrix elements conversion.

Figure 7: Matrix elements conversion.

Note that opinion leader is labeled with ☐. Follower is labeled with ◯. Mutual influenced discussion pattern is labeled with ⬭.

In group D, the influence power of user 19's responses to user 20 and 21 are high. But the responses influence is not high. User 19 and 20 have discussion relationship, so this is a discussion pattern (Figure 8).
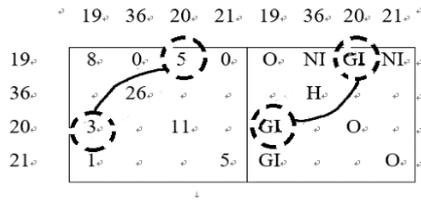
Figure 8: Group D social pattern analysis.

Note that refers Mutual influenced discussion pattern is labeled with ⟨∴⟩.

All other groups are analyzed. User 1 and 36 are summarized as opinion leaders. Group A, C, D, F and H are discussion pattern.

Table 4: Analysis Results.

Opinion Leader	Discussion Pattern Group
User 1, 36	A, C, D, F, H

The results are validated by domain experts and shows that this study can effectively identify opinion leaders and define social community pattern in the social communities, where users' support level could not be obtained because users disagree with each other, users are high controversial, less persons are involved in discussion or many users are anonymous. Through observation of social community pattern among users, we could know users will not support a user's opinions because the user has many speeches.

4 CONCLUSIONS

This study utilizes relational matrix to find the relationship between opinion leaders and followers. Create criteria of social community support level and influence power level between users. Combined with experts' judge to identify opinion leaders and followers. According to social community support level and influence power level in this study, utilize relational matrix to identify opinion leaders in green power issue in the social communities. Utilize our study method to identify social pattern between users. Then we can know, when identify opinion leaders, the social community support level and posted contents are very important to identify opinion leaders. Users could have positive and negative opinions toward the issue. Only considering connection between users' posts are not enough. Even the users' post a lot, if they can't get support from others, they could not be defined as opinion leaders.

Future study and suggestions:

Currently, this study is applied on green energy low carbon issue. In future, this could be applied in marketing filed. Opinion leaders and followers could get comments and reviews of products from social community.

This study proposes social community support level and influence power level. Also we apply relational matrix to analyze relationship between users and social community pattern. If this could be applied in the social media with many information and highly discussed. This study could be more complete.

This study analyzes with static information. In future, we could collect dynamics information to analyze and get instant identification.

REFERENCES

Aguado, E., & Burt, J. E., 2007. *Understanding weather and climate*, Pearson Prentice Hall Upper Saddle River, NJ, USA.

Burt, R. S. 1999., *The social capital of opinion leaders*. The Annals of the American Academy of Political and Social Science, 566(1), 37-54.

Cho, Y., Hwang, J., & Lee, D., 2012. Identification of effective opinion leaders in the diffusion of technological innovation: A social network approach. *Technological Forecasting and Social Change, 79*(1), 97-106

Katz, E., & Lazarsfeld, P. F. 1955., *Personal Influence: The part played by people in the flow of mass communications* ,Transaction Publishers

Lazarsfeld, P. F., Berelson, B., & Gaudet, H., 1948., *The people's choice*. New York: Columbia UniversityPress, 1948

Leal, G. P. A., Hor-Meyll, L. F., & de Paula Pessôa, L. A. G., 2014., *Influence of virtual communities in purchasing decisions: The participants' perspective.* Journal of Business Research, 67(5), 882-890.

Li, F., & Du, T. C., 2011, *Who is talking? An ontology-based opinion leader identification framework for word-of-mouth marketing in online social blogs.* Decision Support Systems

Li, F., & Du, T. C., 2014., *Listen to me—Evaluating the influence of micro-blogs.* Decision Support Systems.

Ma, N., & Liu, Y., 2014. , *SuperedgeRank algorithm and its application in identifying opinion leader of online public opinion supernetwork.* Springer

Weimann, G., & Brosius, H.-B., 1994. *Is there a two-step flow of agenda-setting?* International Journal of Public Opinion Research

Williams, H. T., McMurray, J. R., Kurz, T., & Lambert, F. H., 2015. *Network analysis reveals open forums and echo chambers in social media discussions of climate change*. Global Environmental Change