

Social Cognition Construction of the Avian Flu based on Social Media Big Data

Yuejiao Wang

State Key Laboratory of Management and Control for Complex Systems

Institute of Automation, Chinese Academy of Sciences

Beijing, China

wangyuejiao2018@ia.ac.cn

Zhidong Cao

State Key Laboratory of Management and Control for Complex Systems

Institute of Automation, Chinese Academy of Sciences

Beijing, China

zhidong.cao@ia.ac.cn

Abstract—During the high incidence of avian flu, the mainstream media and social media report a lot on the epidemic, mobilizing the people to prevent and control avian flu. This paper collects reports on avian flu from News, Forums, Apps, WeChat and Microblog and forms five data sets. We extract agenda-settings from the News dataset and build agenda-setting networks of the five datasets. Then we use the QAP test to verify the relevance of these agenda-setting networks. We also project the agenda-setting dissimilarity matrices into a two-dimensional space using the MDS method to form cognitive maps, analyzing the cognitive drift of media platforms relative to News. Results show that the agenda-setting networks of Apps and News have the highest correlation coefficient of 0.9193, while Microblog and News have the lowest correlation coefficient of 0.5611. The cognitive maps of Apps, Forum and WeChat have a slight translation and rotation relative to the cognitive map of News. But their relative positional relationship among agenda-settings are similar with News, expect Microblog.

Keywords—avian flu, agenda-setting theory, QAP, MDS, cognitive construction

I. INTRODUCTION

Pestilence, especially the zoonosis, will cause great damage to the poultry farming and the transportation industry. More seriously, it threatens human life and spreads panic. The avian influenza virus, circulating worldwide, cause hundreds of millions of birds to be hunted and the economic losses can't be estimated. The mainstream news media use news reports to enhance the public's understanding of avian flu. Thereby, we can control the spread of the avian flu and allay social panic. So, using Internet media big data to build social cognition is an important informational means of biosafety.

Social cognition reflects the focus of the masses on events. The official reports and personal opinion on social media are mixed together to form social cognitive networks. Analyzing the links and differences between mainstream media reports and social cognition can help the government to control the direction of public opinion and deliver the prevention information more efficiently. And such biosafety information analysis methods are equally applicable to other public health events.

The cognitive structure of the human brain is reticulated, and our information sources are diverse. The different agenda-settings of an event are related to each other and they merge to form a cognitive network of the event. Lei Guo and McCombs proposed network agenda-setting (NAS) theory [1]. They used

the co-occurrence matrices of agenda-settings to characterize the agenda-setting networks. The agenda-setting effects are explained by the concept of dual psychological paths and need for orientation [2]. NAS theory explains the important influence of news media on public participation in social affairs [3]. Vu et al. expanded the NSA model's scope using five years (2007-2011) of data from national news media and polls [4]. Chris J. et al. found support for agenda-setting melding and validated the NAS model with large datasets from Twitter during the 2012 U.S. presidential election [5]. So, the NAS model can guide us to build an agenda-setting network for news, as well as social cognitive networks.

To analyze the correlation and cognitive drifts between agenda-setting networks, we use the quadratic assignment procedure (QAP) to calculate the network correlation coefficients of different media platforms. The multidimensional scaling (MDS) analysis is a geometric representative method that can project the dissimilarity measure of agenda-settings into a multidimensional space. It is widely used in the perception and preference of research subjects. By performing MDS analysis on the agenda-setting dissimilarity matrices of different media, we can analyze the cognitive drift of the avian flu.

The paper is structured as follows. Section II describes the collection rules and contents of the data sets. Section III explains the data analysis method. And in Section IV, we show the final analysis results on correlation and differences of agenda-setting networks. Section V summarizes the research and gives future directions of the project.

II. DATA

Through our institute's public opinion monitoring system, we have collected reports of avian flu on different media platforms. These data form into five data sets: "News", "Forum", "Apps", "WeChat", and "Microblog". The time span of these data are from January 20th, 2016 to October 20th, 2018. The data sets are all in Chinese.

"News" data set contains news reports on avian flu from the mainstream official media. For example, the mode of transmission of avian flu, the characteristics of the disease and preventive measures. These reports are relatively accurate, not mixed with rumors and interests. Because there are strict regulations on the reports of disease in China, the information in this data set can represent the views of government authorities. The data in "Forum" data set come from the major forums and post bars on the Internet. Some of these information are

The study was funded by National key research and development program (Nos. 2016YFC1200702, 2016QY02D0305) and National Natural Science Foundation of China (Nos. 91546112, 71621002).

reprinted from other channels, while there are also many users' personal opinions on avian flu. The panic, loss or other sentiments of the masses can be found in "Forum" data set. Therefore, it can reflect the public's social cognition of avian flu. "Microblog" contains data from Sina microblog. Similar to the "Forum", its contents are also news reprints and user's opinion. However, the length of a microblog is very short. Avian flu news in "Apps" come from different personalized news recommendation apps, e.g. Headlines Today. Many of them are news reprints, but there are also some unofficial organizations which write news for other interests, such as draw attention or disguised sales. "WeChat" data set contains information of different WeChat public accounts. These public accounts may belong to individuals, enterprises or government organizations. In particular, WeChat public accounts of individuals and enterprises often attract users' attention to achieve their own business purposes. Therefore, for their commercial intentions, the information they convey is not always objective and true.

Through the above introduction, we think the "News" data set well represents the government's attitude towards avian flu and the other four data sets can reflect social cognition to different extent.

III. METHODOLOGY

In essence, an event can set multiple agenda-settings. For example, the agenda-settings of the avian flu include patients, virus, measures and so on. Different agenda-settings are intertwined in our mind and constitute our overall understanding on avian flu. Our general understanding of agenda-settings can includes: which two agenda-settings are more closely related? And which agenda-setting is more important? Based on NAS theory, we extracted the agenda-setting network of mainstream news media from the "News" data set. In the same way, we can represent social cognitions using the agenda-setting networks of social media: Apps, Forum, WeChat and Microblog. The agenda-setting networks formed by the masses and social organizations on social media are the embodiment of social cognition.

Later, we can use the QAP test to analyze the relevance of agenda-setting networks of "News" and other social media. Further, using network visualization technology and multidimensional scaling analysis, we can visually observe the drift of the agenda-settings of different media. The specific data processing flow is shown in the Fig. 1.

A. Data Preprocessing

We obtained data by simply setting the keyword "avian flu" as a screening criterion. However, after a random review of these news data in the five data sets, we found that many data have important fields missing: the "content" field may be blank for some data were not properly obtained. Many data just contain the keyword "avian flu" in their content field, but from the title, we could see that the theme of the news is not about avian flu. We even found a same piece of data has been repeatedly retrieved many times. Therefore, these missing data, irrelevant data and duplicate data all need to be eliminated, which will make subsequent analysis more accurate. Although processing of big data places more emphasis on relevance than on precise causality, the preprocessing of data is still meaningful.

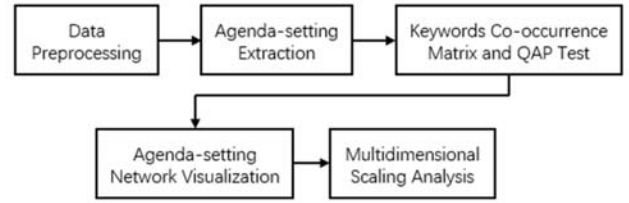


Fig. 1. Data processing flow

1) *Data deduplication*: We believe that if the URL and published time of the two news items are same, they are duplicate data. Using Excel's internal deduplication function, we can easily de-duplicate these five data sets separately.

2) *Remove missing data and irrelevant data*: We customize the Excel conditional function: a piece of news is retained when both its "content" and "title" fields both have keywords associated with avian flu. Under such screening conditions, missing data and irrelevant data can be removed.

After data preprocessing, the size of the five data sets have been reduced. The data size of "News", "Forum", "Microblog", "Apps" and "WeChat" are 10954, 2134, 7881, 4280 and 11122 respectively. We can't guarantee that all the rest of the data are related to avian flu. There is also no guarantee that these five data sets can cover all the valid news on the media. But in the correlation analysis, these data are enough for us to show the agenda-setting network of news and the social cognitive structure of the avian flu.

B. Agenda-setting Extraction

"Agenda-setting" is a terminology for journalism, which means a social event that is widely concerned by the public. This concept has been described in detail in McComb's research of agenda-setting [6]. Before the Internet era, scholars used the statistical methods such as questionnaires to study agenda-setting. After the emergence of big data and advanced text data mining techniques, researchers use text keyword extraction methods to obtain agenda-settings in the NAS theory.

Because we want to study the impact of mainstream news reports on social cognition, the key point is the correlation and difference analysis of the news agenda-setting network and social cognitive network. Therefore, in the process of building networks of five data sets, the agenda-settings should be extracted from the "News" data set. These agenda-settings were set by the mainstream media in the reports of avian flu, but they changed as they spread through the social media platforms. The level of attention on the agenda-settings will fluctuate and the degree of association between different agenda-settings will change. In this way, we can monitor the agenda-settings drift in the social cognition of avian flu, whether they are generally consistent with the mainstream media, and whether there are large-scale rumors on social media. We can even track the concerns of the masses and conduct news reports to meet their needs.

There are many mature algorithms for keyword extraction. We use the most basic methods based on word frequency and part-of-speech. The specific process of keyword extraction can be divided into the following steps.

1) *Word segmentation and filtering*: Using “jieba”, the open source natural language processing library for Chinese in Python, we divide Chinese sentences into collections of words. And we filter out stop words, which are commonly meaningless words or punctuation. At the same time, the words are automatically marked with their part-of-speech, such as nouns, verbs, adjectives and so on. Then, we filter out some words that are not related to the agenda-settings according to their part-of-speech, because some words can’t represent the name or attribute of an event, such as conjunctions.

2) *Count words frequency and rank*: We extract the leaving words with the frequency in the top 200.

3) *Manual screening*: Even after the above operation, there still are many words that are not related to avian flu in the two hundred keywords. For example, the word “work” is a common vocabulary in news reports and does not reflect the characteristics specially associated with avian flu. So, it can’t be included in the agenda-settings. So, based on human experience, we conduct further screening and retain one hundred keywords as agenda-settings.

C. Agenda-setting Co-occurrence Matrix and QAP Test

We need to build agenda-setting networks for the five data sets and their equivalent representation are the agenda-setting co-occurrence matrices. In NAS theory, researchers introduced the concept of “co-occurrence”, which means that two words appear in the same paragraph at the same time. Because humans’ brain cognition generally shows such a pattern: when two words appear simultaneously in a text, people will think the two words are related. For example, “avian flu” often appears in the news report along with “fever”, so in the cognitive, people know that fever is one of the symptoms of avian flu. The link between the agenda-settings may be inherent, or it can be deliberately constructed by the news publisher for their commercial interests or other purposes. In summary, the more co-occurrence of these two agenda-settings, the higher the correlation between them.

The agenda-setting co-occurrence matrix of this paper is a matrix with 100 rows and 100 columns. The row and column labels are the 100 agenda-settings and the elements of the matrix are the number of co-occurrences of the corresponding agenda-settings. The acquisition of the co-occurrence matrix is quite simple, and we only need to use the “jieba” library for word segmentation and counting. Therefore, we can get five agenda-setting co-occurrence matrices from our five data sets respectively. The co-occurrence matrix corresponds to a weighted undirected network. The nodes of the network are the agenda-settings, and the weight of the edge is the times of co-occurrence of the two agenda-settings. In essence, this is exactly the agenda-setting network.

We use the Pearson correlation coefficient to measure the similarity between the two co-occurrence matrices. However, since the co-occurrence matrix itself is a kind of relational data, it has strong collinearity. Therefore, we need to use other methods to test the significance of the correlation coefficient. The QAP test is a widely used method of significance test [7]. The specific steps of QAP are as follows. The original two co-occurrence matrices are A and B, and the diagonal elements of A and B should be set to zero before QAP test. Fix matrix A and

permute the rows and columns of matrix B. But be careful not to break the correspondence between two labels of B. So, we get a new matrix E. Calculate the Pearson correlation coefficient between A and E. Repeat the random permutation for hundreds or even thousands of times, and we can get the probability density distribution of the correlation coefficients. If the correlation coefficients of A and B are within the confidence interval of the probability density distribution, we can think that the correlation between them is significant, not random.

D. Agenda-setting Network Visualization

The co-occurrence matrix of the agenda-settings portrays the agenda-setting network. However, data in the form of chart are not conducive to people to discover the patterns implicit in them. In order to observe the agenda-setting network more intuitively, we use Gephi for data visualization. Gephi is an open source data visualization software. Using the co-occurrence matrix as input, we can get an initialized weighted undirected graph. Its built-in module partitioning function can automatically divide different agenda-settings nodes into different modules based on their associations. The size of the node is proportional to its degree. In order to make the layout of the network more reasonable, we apply Yifan Hu layout, so that all nodes of the agenda-setting network are arranged into a circular outline. After the above processing, we can easily analyze the similarities and differences between the five agenda-setting networks.

E. Multidimensional Scaling Analysis

Multidimensional scaling (MDS) analysis is an exploratory approach for data analysis [8]. It makes data more meaningful through dimensionality reduction and visualization. This method can identify implicit structural within the data.

Enter a dissimilarity matrix that represents the distances between the agenda-settings. The dissimilarity matrix is calculated based on the co-occurrence matrix. The calculation of dissimilarity matrix is shown as in (1):

$$D_{ij} = 1 - \frac{C_{ij}}{\sqrt{C_{ii} \cdot C_{jj}}}, D_{ij} \in [0,1] \quad (1)$$

D_{ij} is the element in dissimilarity matrix D and C_{ij} is the element in co-occurrence matrix C . If the D_{ij} is close to one, it indicates that the relation between the corresponding two agenda-settings is small. Through the eigenvectors, we project the agenda-setting co-occurrence matrix into a two-dimensional map. Not only can we measure the similarity between the two agenda-settings with the Euclidean distance, but we can also cluster them through locations in the two-dimensional map. The basic algorithm principle of MDS is shown as in (2):

$$\|z_i - z_j\| = dist_{ij}, dist_{ij} = D_{ij} \quad (2)$$

$\|z_i - z_j\|$ is the Euclidean distance after two agenda-settings are dimension reduced, and $dist_{ij}$ is the dissimilarity between the two agenda-settings in the original space. The dissimilarity matrix D is a differential description of the agenda-settings. Importing the built-in function “cmdscale” in MATLAB, we project one hundred agenda-settings into a two-dimensional cognitive map. The five data sets correspond to five cognitive maps. Using the cognitive map of “News” as a reference, we can analyze the offset of cognitive maps on social

media. Obviously, in addition to the agenda-setting network, this is also a way of constructing social cognition. And the two-dimensional spatial map is a representation that is more in line with our cognitive habits.

IV. RESULTS

A. Agenda-setting Extration

Through a series of important processes such as word segmentation and manual screening, we got 100 agenda-settings for the avian flu epidemic in the mainstream media. Because the data sets we grabbed are from China, the agenda-settings are all Chinese vocabularies. For the sake of understanding, we translated the agenda-settings into English. Although there are some nuances in semantics, it does not prevent us from conveying the main ideas and methods of this paper. Due to space limitations, we only listed the top 10 agenda-settings in Table I, with their corresponding English translations and frequencies.

TABLE I. AGENDA-SETTINGS LIST OF TOP 10

Order	Agenda-settings (Chinese)	Agenda-settings (English)	Word Frequency
1	禽流感	avian flu	34860
2	传染病	pestilence	34174
3	疫情	epidemic	29436
4	病毒	virus	27852
5	疫苗	vaccine	23192
6	动物	animal	21872
7	报告	report	21079
8	工作	job	18845
9	H7N9	H7N9	16821
10	流感	flu	14642

Through observation, the avian flu agenda-settings contained in these 100 words are diverse. These agenda-settings can be classified into the following aspects.

1) *The symptoms of avian flu:* There are many agenda-settings that describe the pathogenic properties of avian flu. For example, the word “pathogenicity” emphasizes the infectious characteristic of avian flu. And “respiratory tract” indicates the route of transmission of avian flu.

2) *Market factors affected by avian flu:* Some of the agenda-settings reflect the fluctuations in China's domestic poultry market after the outbreak of avian flu. The words “price” and “sales” are the agenda-setting vocabularies of market price volatility.

3) *China's emergency prevention and control measures against avian flu:* The “vaccine” is one of the government's preventive measures against avian flu.

There is a phenomenon that needs attention. Chinese vocabularies are varied and two different words will express the same meaning. For example, “kids” and “children” are both the top 100 agenda-settings. The usage of “children” is more formal

while “kids” often appears in spoken language. We don't merge these two words because they reflect the difference in expression of various groups. Experts prefer to use “children” in formal settings while non-professionals use “kids” more often. In short, these 100 words are enough to represent the main agenda-settings of avian flu.

B. Results of QAP Test

We obtained co-occurrence matrices of five data sets using one hundred agenda-setting keywords and did QAP test on these five co-occurrence matrices. The Pearson correlation coefficients are shown in Table II. It can be find that the context of News has the highest similarity with Apps (correlation coefficient=0.9193, $p<0.01$), the second similarity with Forum (correlation coefficient=0.8244, $p<0.01$), the third similarity with WeChat (correlation coefficient=0.7169, $p<0.01$), and the lowest similarity with Microblog (correlation coefficient=0.5611, $p<0.01$). The QAP test shows that these correlation coefficients are all statistically significant. These five agenda-setting networks are significantly related, especially between News, Apps, and Forum. The relevance of Microblog and several other media platforms are relatively weak.

TABLE II. CORRELATION COEFFICIENT OF CO-OCCURRENCE MATRIX

	Forum	Microblog	Apps	WeChat
News	0.8244*	0.5611*	0.9193*	0.7169*
Forum	--	0.4857*	0.7589*	0.5861*
Microblog	--	--	0.5002*	0.2911*
Apps	--	--	--	0.9048*

* means a result with $p<0.01$

It is certain that the agenda-setting of the mainstream news media is effective and dominates the direction of public opinion. Factors that affect the correlation coefficients of the co-occurrence matrices are considered as follows. News, Apps and Forum have higher correlation coefficients, indicating that their news contents are closer and the agenda-setting structures are highly similar. But there are some drifts between the agenda-setting networks of News and WeChat public accounts. This shows that the information dissemination of the public accounts is biased. The WeChat public accounts may have potential purpose of attracting users' attention. Or the publishers of the public accounts are more concerned about the avian flu agenda-settings that are relevant to their own interests. The context of News has the lowest similarity with Microblog. Because the information on Microblog is relatively short, the co-occurrence probability of agenda-setting words are much smaller, resulting in a decrease in the correlation coefficient.

TABLE III. CORRELATION COEFFICIENT OF DISSIMILARITY MATRIX

	Forum	Microblog	Apps	WeChat
News	0.7768*	0.5936*	0.9403*	0.8254*
Forum	--	0.4594*	0.7413*	0.6661*
Microblog	--	--	0.5821*	0.4138*
Apps	--	--	--	0.9277*

* means a result with $p<0.01$

Based on the co-occurrence matrices, we get the corresponding dissimilarity matrices. The QAP test is still used

for dissimilarity matrix correlation detection. The Pearson correlation coefficients we obtained are shown in Table III.

It can be found that the dissimilarity matrix of News has the highest similarity with Apps (correlation coefficient=0.9403, $p<0.01$), and the lowest similarity with Microblog (correlation coefficient=0.5936, $p<0.01$). Fig. 2 shows the probability density distribution of the correlation coefficients generated by the QAP test in its permutation process. The true Pearson correlation coefficients are within the confidence interval. We only show the distribution of correlation coefficients between the co-occurrence matrices of News and Apps. The probability density distributions of other correlation coefficients are similar to Fig. 2.

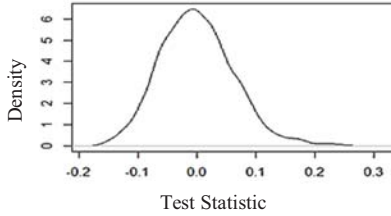


Fig. 2. The probability density distribution of the correlation coefficients

C. Agenda-setting Networks

Another equivalent representation of the co-occurrence matrix is a weighted undirected graph, which is called the agenda-setting network in NAS theory. The agenda-setting network visualization helps us to observe and analyze the similarity of agenda-setting of different media from the structural level. In this way we can have a more intuitive understanding of the agenda-setting network.

Due to space limitation, we only show the agenda-setting networks of three data sets in Fig. 3: News, Apps and Microblog. Because using the News dataset as a reference, as shown in Table II, the Apps dataset has the highest correlation with News, while the Microblog dataset has the lowest correlation. And after modular partitioning, the nodes of different modules are of different colors. In order to make the nodes in the center of the network more obvious, the greater the degree of the nodes, the larger the size of the nodes.

In Fig. 3, the three agenda-setting networks are divided into two modules by the community discovery algorithm [9]. In Fig. 3a and Fig. 3b, the ratio of the number of red nodes and green nodes is very close, and the size of the nodes is also very uniform. However, there are significant differences between Fig. 3a and Fig. 3c. In the Microblog's agenda-setting network, the number of green nodes is reduced, and the degree of green nodes is also declining. This shows from another perspective that the agenda-setting networks for the two media platforms, News and Apps, are very similar, with a correlation coefficient of 0.9193. But the Pearson correlation coefficient between networks of Microblog and News is significantly lower.

The data table given by Gephi confirms our observations and conjectures. The red-green nodes ratios in Fig. 3a and Fig. 3b are 66:34 and 65:35, respectively. And there are many overlaps in the agenda-setting words in the same module. However, the

red-green node ratios in Fig. 3c is 74:26, and the agenda-setting overlaps between Fig. 3a and Fig. 3c is less.

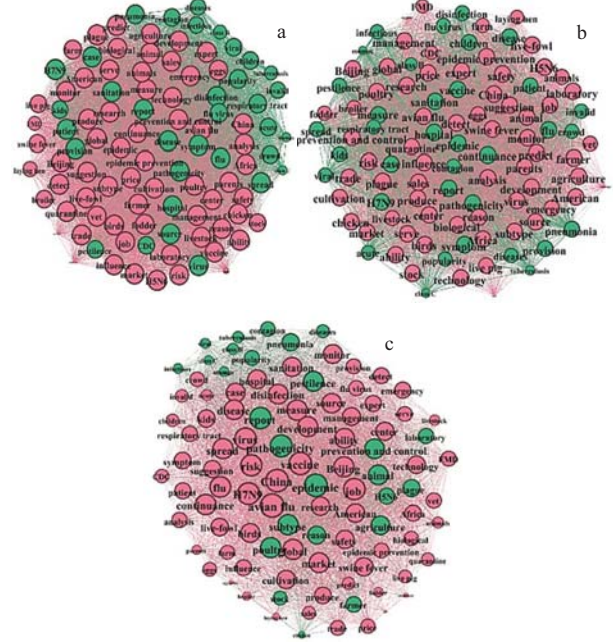


Fig. 3. Agenda-setting networks of News (a), Apps (b) and Microblog (c)

D. The Results of Multidimensional Scaling Analysis

The differences in the agenda-setting of different media platforms are not specific enough in the meshed agenda-setting networks. Multidimensional scaling (MDS) analysis can visually depict the differences between avian flu agenda-settings in the physical space using Euclidean distance.

Taking the dissimilarity matrix of the News dataset as a reference, we project the agenda-setting dissimilarity matrices of the other four media platforms into a two-dimensional cognitive map using the MDS method. In cognitive maps, if the Euclidean distances of the two agenda-settings are short, it means that they are more closely related in the cognitive structure. More importantly, with the news data as a reference, we can measure the drifts on agenda-setting of the other four media platforms. Fig. 4 shows the cognitive maps of avian influenza constructed by the MDS method, comparing with the cognitive map of News. Limited by space, we only put the cognitive maps of Apps and Microblog in Fig. 4, because the dissimilarity matrix of Apps has the highest similarity with News and the matrix of Microblog has the lowest similarity with News in Table III.

The cognitive maps of the four media platforms, Apps, WeChat, Microblog and Forum, are the same as the cognitive maps of News from the outline. We have marked the names of three nodes in the Fig. 4 to facilitate the observation of agenda-setting drift. Using the cognitive map of News as a reference, the cognitive maps of Apps, WeChat and Forum all show a slight shift and rotation overall, but the relative positional relationship of their agenda-settings is not much different. However, as shown in Fig. 4a, the microblog's cognitive maps have changed significantly in the distance and relative position

of the agenda-settings. The differences reflected in Fig. 4 are consistent with our previous similarity calculations in Table II. And according to the location of the agenda-settings in the two-dimensional map, they can be roughly clustered into two types. This is consistent with the module partitioning results in Fig. 3. So, MDS is not only an effective way to build social cognition, but also intuitive method for agenda-settings classification and association.

and Apps have the strongest relevance, while News and Microblog have the weakest relevance. The consistencies of the conclusions indicate that the combination of NAS theory and MDS is an effective means for social cognition construction. The same method is also suitable for analyzing other public health safety events. The following aspects need further research: cognitive maps drift analysis in time and geospatial dimensions, which can reveal the impact of news reports on social cognition in a more detailed dimension.

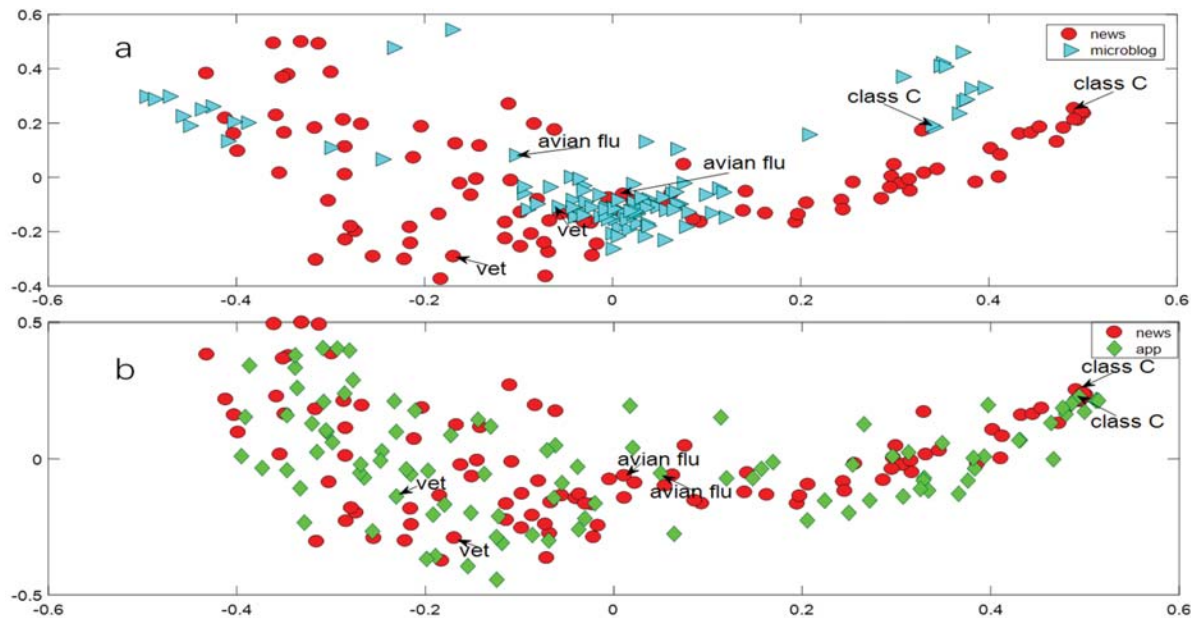


Fig. 4. Cognitive maps contradistinction between News and Microblog (a), News and App (b)

V. CONCLUSION

The focuses of the paper are the social cognitive construction of avian flu and the correlation analysis of agenda-setting of different media for avian flu. We have acquired five avian flu-related data sets from News, Forum, Apps, WeChat and Microblog, representing different media platforms. Drawing on NAS theory, we extracted the agenda-settings words of avian flu in the News data sets and built networks of agenda-setting. The correlation coefficients between the five agenda-setting networks were calculated by the QAP test. We found that the context of News has the highest similarity with Apps (correlation coefficient=0.9193, $p<0.01$), the second similarity with Forum (correlation coefficient=0.8244, $p<0.01$), the third similarity with WeChat (correlation coefficient=0.7169, $p<0.01$), and the lowest similarity with Microblog (correlation coefficient=0.5611, $p<0.01$). Then, with Gephi for network visualization, we analyzed the relevance and difference of agenda-setting from the graphical structure level. Finally, using the MDS method, the agenda-setting networks were projected onto two-dimensional maps to measure the difference in agenda-settings between News and other media.

Whether it is NAS theory or MDS method, our conclusions are consistent: News and other social media have a strong correlation in the avian flu agenda-setting. Among them, News

REFERENCES

- [1] G. Lei, H. T. Vu, and M. McCombs, "An expanded perspective on agenda-setting effects: Exploring the third level of agenda-setting setting," *Revista de comunicaci3n*, vol. 11, pp. 51-68, 2012.
- [2] M. McCombs, and N. J. Stroud, "Psychology of agenda-setting effects: Mapping the paths of information processing," *Review of Communication Research*, vol. 2, pp. 68-93, 2014.
- [3] M. McCombs, "A look at agenda-setting: Past, present and future," *Journalism studies*, vol. 6, no. 4, pp. 543-557, 2005.
- [4] H. T. Vu, L. Guo, and M. E. McCombs, "Exploring "the world outside and the pictures in our heads" A network agenda-setting study," *Journalism & Mass Communication Quarterly*, vol. 91, no. 4, pp. 669-686, 2014.
- [5] C. J. Vargo, L. Guo, M. McCombs and D. L. Shaw, "Network issue agenda-settings on Twitter during the 2012 US presidential election," *Journal of Communication*, vol. 64, no. 2, pp. 296-316, 2014.
- [6] M. E. McCombs, and D. L. Shaw, "The agenda-setting function of mass media," *Public opinion quarterly*, vol. 36, no. 2, pp. 176-187, 1972.
- [7] D. Krackardt, "QAP partialling as a test of spuriousness," *Social networks*, vol. 9, no. 2, pp. 171-186, 1987.
- [8] A. Buja, D. F. Swayne, M. L. Littman, N. Dean, H. Hofmann, and L. Chen, "Data visualization with multidimensional scaling," *Journal of Computational and Graphical Statistics*, vol. 17, no. 2, pp. 444-472, 2008.
- [9] V. D. Blondel, J. L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of statistical mechanics: theory and experiment*, no.10, P10008, 2008.