Collaboration Pattern and Topic Analysis on Intelligence and Security Informatics Research

Wenli Liu, *National University of Defense Technology, China*

Xiaolong Zheng, *Chinese Academy of Sciences*

Tao Wang and Hui Wang, *National University of Defense Technology, China*

Here, the researcher collaboration patterns and research topics on Intelligence and Security Informatics (ISI) are investigated using social network analysis approaches. The collaboration networks exhibit a scale-free property and small-world effects.

Terrorism, cybersecurity threats, and criminal attacks have caused great threats to our societies and commercial infrastructures, especially with the rapid and large-scale migration of security information from traditional media to social media platforms. This leads to the need of developing more advanced technologies to protect our national/homeland security. Starting in 2003, scholars led by Hsinchun Chen, Fei-Yue Wang, and Daniel Zeng introduced a new research field—Intelligence and Security Informatics (ISI),[1–3] which mainly concerns the development of information technologies (IT) and AI approaches to address security-related challenges.[2–4]

In the past decade, with the significant work conducted by pioneering researchers, ISI has grown quickly into a highly interdisciplinary field,[5–10] with a large number of related papers getting published in many influential journals. Since 2005, IEEE has sponsored the flagship ISI annual international conference series, and since 2006 and 2008 in Pacific Asia and Europe, respectively, ISI workshops have been held annually.

As the body of ISI literature continues to grow, newcomers to this field or researchers in other fields might find it difficult to grasp the state of the field, along with the direction of current research avenues. They might have the following main questions: Who are the most influential authors in this field? How often (and in what capacity) do researchers collaborate with each other? What are the driving interests of researchers in this field?

This issue isn't unique to this discipline—other fields, in fact, may face similar problems. To contend with this prevailing issue, Mark Newman has constructed a social network graph, in which the nodes represent scientists and links represent the instances where they've co-authored a paper.[11] By analyzing this network, Newman commendably answered some

1541-1672/14/$31.00 © 2014 IEEE
Published by the IEEE Computer Society

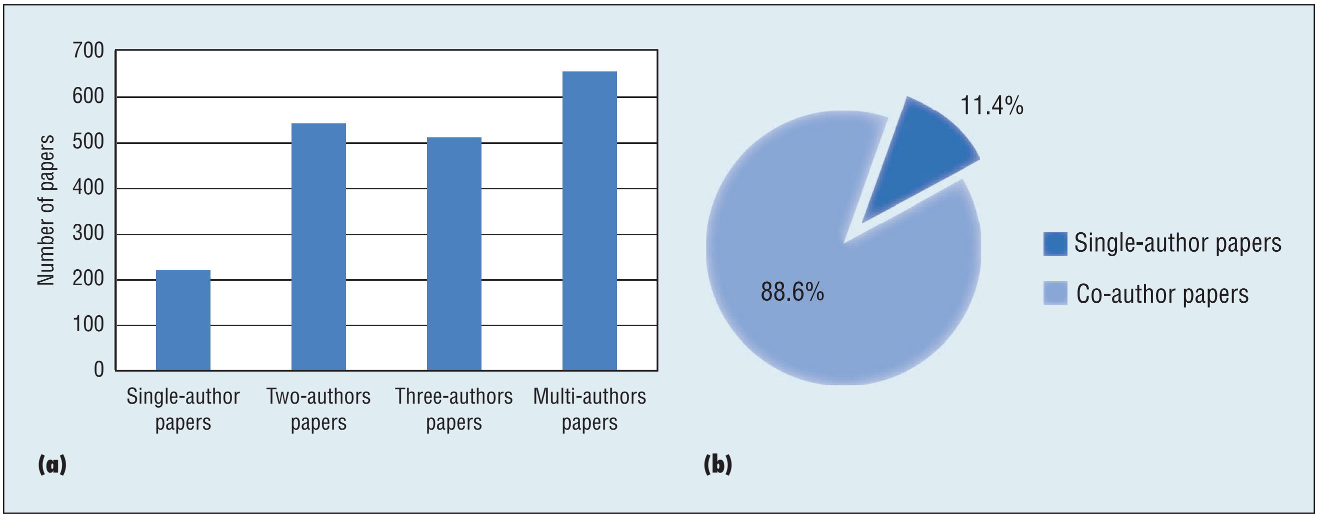

Figure 1. Breaking down the number of authors per paper. (a) Bar chart showing the number of authors per paper within the dataset. (b) Pie chart showing that single-authored papers are in the minority.

questions, such as the number of papers authors have written and the number of coauthors they had. Similarly, some other researchers have studied science collaborations in intelligent transportation, social computing, biology, mathematics, astrophysics, and nanoscience.[12–16] However, these researchers didn't consider how patterns of collaboration vary between institutions. Edna Reid and Hsinchun Chen created a systematic research process to identify the key researchers by using an integrated knowledge-mapping framework.[17] These significant studies mainly focused on the collaboration patterns at the researcher level, the institutional level, as well as the national level. To further understand the patterns of scientific collaboration, Ketan Mane and Kany Borner analyzed a certain field of topics and topic bursts.[18] However, as we investigated this process, we noticed that although the research of ISI is becoming increasingly important, its scientific collaboration patterns and topic evolution still haven't been uncovered. Yet, uncovering this information is important and applicable, because it can provide significant insights for researchers to better understand and evaluate key impact factors and quickly grasp emerging directions in ISI.

Thus, here we analyze collaboration patterns and conduct empirical analysis of topic evolution in ISI. We first construct an ISI co-authorship network and obtain some interesting results by analyzing its topological characteristics. To identify the key researchers in ISI, we adopt four measures to evaluate them. These measures encompass influence, productivity, collaboration ability, and the average number of times cited. Topic analysis is further conducted to detect the evolution process of research interests in this field.

Dataset

For this work, our dataset was retrieved from the Institute for Science Information Web of Science (WoS), an online academic citation index provided by Thomson Reuters. Based on the influence of WoS, we mainly focus on discussing the publications retrieved from WoS. The dataset contains the metadata and derived data. The metadata contains 505 papers that were published in the *Proceedings of the IEEE International Conference on ISI* from 2003 to 2010. And the derived data contains 1,431 papers which have directly and indirectly cited the metadata. It's well known that when one paper was cited by another paper, this meant the two papers were closely related to some extent.

All in all, our dataset contains 1,936 papers. From Figure 1, we can see that most of these papers, approximately 88.6 percent, have more than one author. This demonstrates that most ISI researchers are likely to collaborate, and the dataset may provide us with more significant information to understand collaboration patterns among these researchers.

ISI Collaboration Networks

After extracting authors' names in the dataset, we constructed a collaboration network, in which each node corresponds to an ISI author and each link between two authors represents a shared co-authorship relationship. The ISI collaboration network includes 4,014 nodes and 7,884 edges, and contains 731 components. Figure 2 shows the degree distribution of the ISI collaboration network, which is virtually scale-free. In this kind of network, only a few of researchers have a large number of collaborators, and most researchers have a few collaborators. Compared to other collaboration networks, such as the terrorism domain (1990–2003) and social network domain (1996–2002),[16] the ISI collaboration network has a longer tail. In ISI collaboration, about 1.5 percent of the researchers have more than

15 collaborators, and about 73.9 percent of the researchers have less than 5 collaborators.

Meanwhile, we also compute the average degree, average distance, largest distance, and clustering coefficient of the ISI collaboration network, and compare these parameters to three kinds of collaboration networks (see Table 1), including physics domain collaboration networks,[11] terrorism collaboration networks, and social network collaboration networks.[16]

Table 1 shows a summary of the basic statistics of four collaboration networks. In the ISI collaboration network, the average degree is 3.92 and the total number of nodes is 4,014. Theoretically, if the network is random graph, its clustering coefficient is 3.92/4,014 = 0.00098. However, the real-world network has a clustering coefficient 0.37, which is significantly higher than expected by random chance. In addition, the average distance of this network is only 1.58. Therefore, we can see that the ISI collaboration network exhibits the small-world effect.

As we know, some researchers began their study on their own and some collaborated with only a few researchers—for this reason, some components in the whole ISI co-authorship network only contain two nodes or are even isolated. However, the largest component contains 713 nodes, which means that the largest collaboration center includes 713 researchers. Figure 3 shows the largest component of the co-authorship network.

In the co-authorship network, Hsinchun Chen has the highest degree (degree = 90), which means he has the largest number of collaborators. The authors with the second and third highest degrees are Daniel Zeng (degree = 57) and Fei-Yue Wang (degree = 50), respectively.

As Figure 3b shows, the subnetwork is obtained by removing the nodes with a degree of less than 16. In this subnetwork, three nodes, called H.C. Chen, D. Zeng, and F.Y. Wang, respectively, are much larger than others. Furthermore, the edges among them are also much bolder. This indicates that these three researchers have collaborated frequently with others and played a significant role in the development of the ISI field. Especially, to our investigation, a famous article[2] written by these three researchers in *IEEE Transactions on Intelligent Transportation Systems* in 2004, with the title of "Intelligence and Security Informatics for Homeland Security: Information, Communication, and Transportation," is regarded as the initial research work. This article has played a critical role in promoting the development of ISI as a field. As Table 2 shows, there are several other examples of significant researchers collaborating within this field.

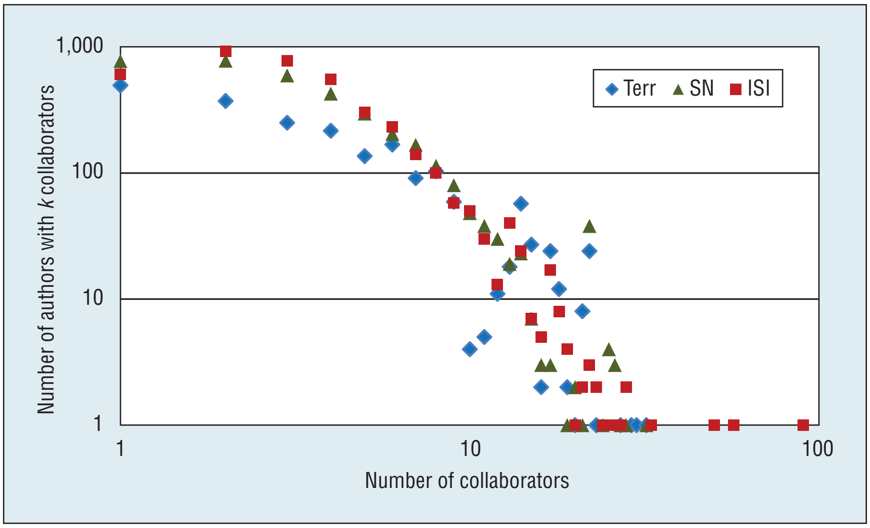

Figure 2. Degree distribution of the Intelligence and Security Informatics (ISI) collaboration network. Terr represents the terrorism domain collaboration network and SN represents the social network domain collaboration network.

Table 1. Summary statistics for four collaboration networks.

Criteria	Physics[11]	Terrorism[16]	Social network[16]	ISI
Total authors	52,909	3,038	4,072	4,014
Average degree	9.7	4.66	3.98	3.92
Average distance	5.9	1.26	1.29	1.58
Largest distance	20	9	13	14
Clustering coefficient	0.43	0.45	0.42	0.37

Institutional-Level Collaborations

We've examined the frequency of researchers collaborating, but to better understand the relationships involved, how often is this occurring between institutions? To ascertain this information, we constructed an institutional-level collaboration network of ISI. Collaborations between two institutions are considered when there's a co-occurrence in one paper in the dataset. The institutional-level collaboration network is constructed in just the same way as the researcher level. It includes 1,047 nodes and 529 edges, and contains 14 components. It's also an unconnected graph. The largest component contains 187 nodes. Figure 4

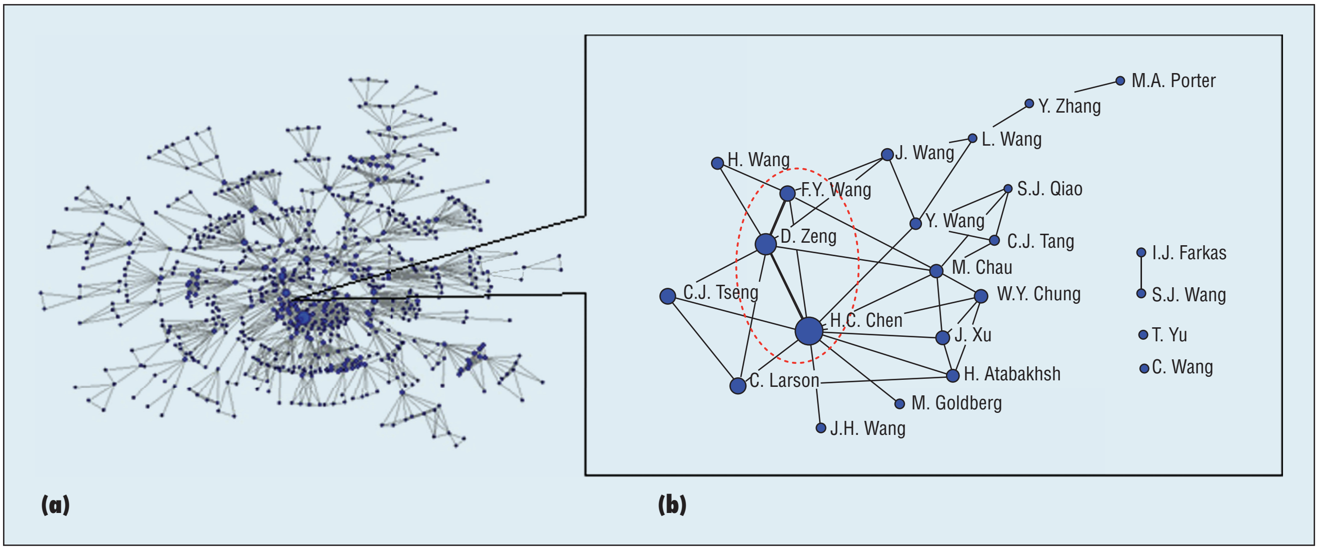

Figure 3. Researcher-level collaboration network. (a) The largest component. (b) Rich collaboration scientists. Node size is proportional to degree. Link strength is proportional to the frequency of collaboration. All the scientists in (b) have more than 15 collaborators.

Table 2. Frequent collaboration authors.

Two author collaborations	No. of collaboration times	Three author collaborations	No. of collaboration times
Hsinchun Chen and Daniel Zeng	13	Hsinchun Chen, Daniel Zeng, and C.J. Tseng	7
Fei-Yue Wang and Daniel Zeng	11	Fei-Yue Wang, Daniel Zeng, and Wenji Mao	7
Mark Goldberg and Malik Magdon-Ismail	10	Daniel Zeng, Hsinchun Chen, and Fei-Yue Wang	5
William Zhu and Fei-Yue Wang	9	Hsinchun Chen, Daniel Zeng, and C. Larson	5
Judee K. Burgoon and Jay F. Nunamaker	9	Judee K. Burgoon, Jay F. Nunamaker, and Mark Adkins	3

shows the largest component of the institutional collaboration network.

In the institutional collaboration network, betweenness represents the knowledge diffusion ability. We can clearly see how ISI knowledge spreads between institutions.

In this network, there are two clusters which were obtained using the Girvan-Newman algorithm. The University of Arizona node has the highest degree (degree = 20), which means it has directly collaborated with 20 institutions. It's a collaboration center in the blue cluster. In fact, the ISI conference was held in Arizona in 2003 and 2004. The University of Arizona collaborated with many institutions at that time, such as Michigan State University, the University of Utah, and the University of Pennsylvania. At the same time, the node with the second highest degree is the Chinese Academy of Science. It's obviously a collaboration center in the red cluster. In the year 2005, the "Intelligence and Security Informatics innovation group of the international collaboration partner plan" was carried out by the Chinese Academy of Science and the University of Arizona. Going forward, many international institutions joined in ISI research, including Zhejiang University, the University of Copenhagen, and Indiana University.

On the other hand, it's obvious that "Univ_Arizona" and "Chinese_Acad_Sci" are the two key nodes that connect two major components. So, the University of Arizona and the Chinese Academy of Science are the key institutes that should be considered. In our investigation, we saw that great efforts have been carried out by research scientists in University of Arizona (such as Hsinchun Chen, Daniel Zeng, Edna Reid, Judee K. Burgoon, and Jay F. Nunamaker) and research scientists in the Chinese Academy of Science (such as Fei-Yue Wang, Jue Wang, and Wenji Mao). There are 21 collaboration papers that contained journey articles, conference papers, and reviews between these two science teams. These papers mainly discuss homeland security,[2] wisdom of crowd,[10] and information sharing.[19]

Key Researchers

To identify the key researchers, we adopted four measures, including productivity, influence, collaboration ability, and average number of times cited. Productivity is measured by the number of papers published by one researcher. Influence represents the cited counts of one researcher's total papers. Collaboration ability, which is measured by degree, represents the total number of collaborators for one researcher. The

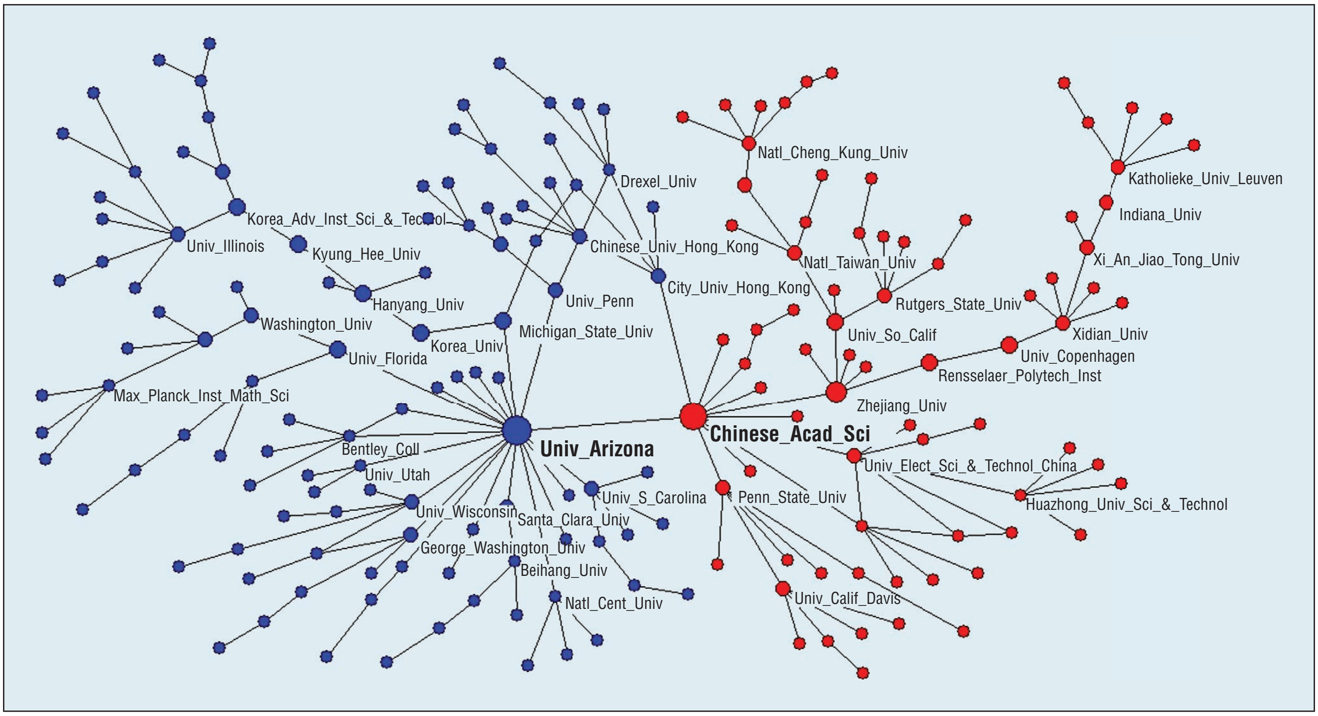

Figure 4. Institutional-level collaboration network. Node size is proportional to the betweenness.

average cited times represent the quality of one researcher's papers. In Table 3, key researchers are identified and ranked using these four aspects.

We assumed that all the authors of one paper made the same contribution to that paper. The more that an author appears in papers, the more productivity he is accredited. Table 3 shows the productivity and influence of the most prolific authors. Hsinchun Chen is the most productive author and has 75 papers contributing to the field of ISI. The productivity of his papers account for 3.88 percent of the total, following by Fei-Yue Wang at 2.17 percent (42), and Daniel Zeng at 1.86 percent (36). Their papers are cited 84, 46, and 32 times, respectively.

We determined this information using the four aspects we previously mentioned: collaboration ability (see Figure 3), productivity, influence, and the average number of times cited (see Table 3). Figure 5 shows the top 10 researchers identified in our dataset, selected and ranked by productivity (primarily considering the number of papers).

Table 3. Prolific authors in ISI.*

Author	Institute of the author	Productivity	Influence	Average no. of times cited
Hsinchun Chen	University of Arizona	75	84	1.12
Fei-Yue Wang	Chinese Academy of Sciences	42	46	1.10
Daniel Zeng	Chinese Academy of Sciences and University of Arizona	36	32	0.89
Jennifer Xu	University of Arizona	15	22	1.47
J. K. Burgoon	University of Arizona	14	20	1.43
Edna F. Reid	University of Clarion	12	18	1.50
Christopher C. Yang	The Chinese University of Hong Kong	12	9	0.75
Jia-Lun Qin	University of Massachusetts, Lowell	11	19	1.73
J. F. Nunamaker	University of Arizona	11	17	1.55
E.G. Im	Hanyang University, South Korea	9	6	0.67

* This information was gathered in December 2011.

From Figure 5, we can see that researchers who have less productivity may have a higher collaboration ability, while researchers who have more productivity may have fewer instances of being cited on average. Authors such as Jennifer Xu and Christopher C. Yang have more productivity, but are cited fewer times. Meanwhile, Hsinchun Chen plays a key role in the first three aspects, while Jia-Lun Qin has the most times cited per paper.

ISI Researchers' Interests

Since its first proposal in 2003, ISI has attracted many researchers' attention

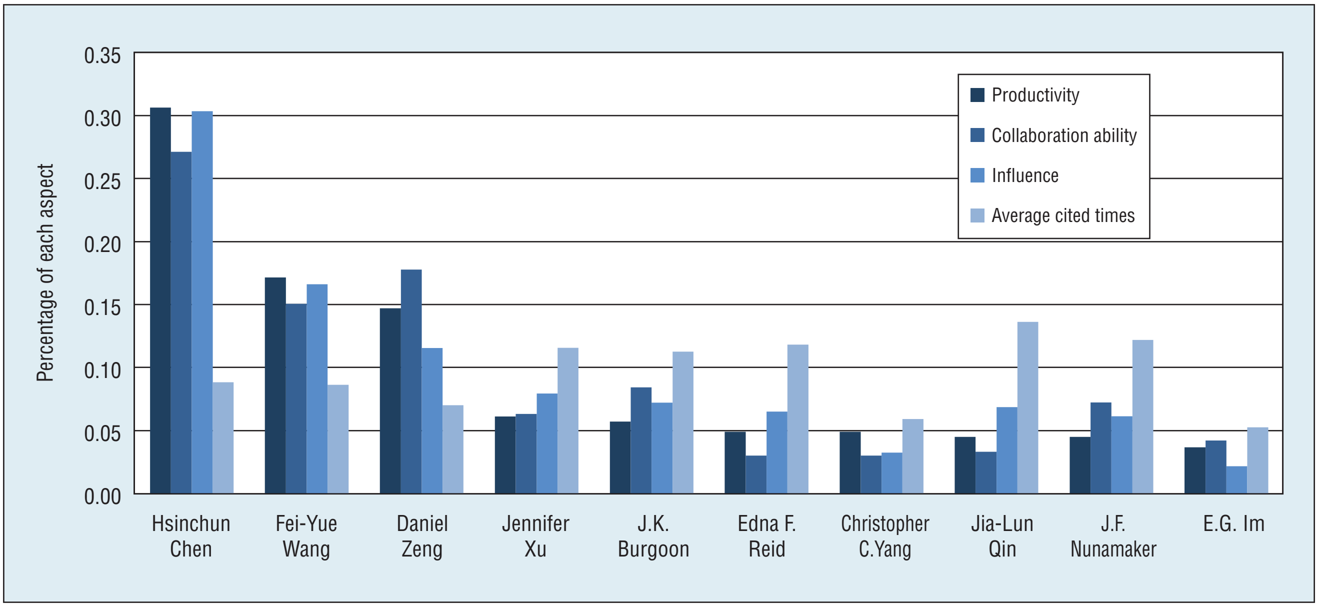

Figure 5. Top 10 researchers in the ISI field ranked by their productivity. For clarity, we normalized these results.

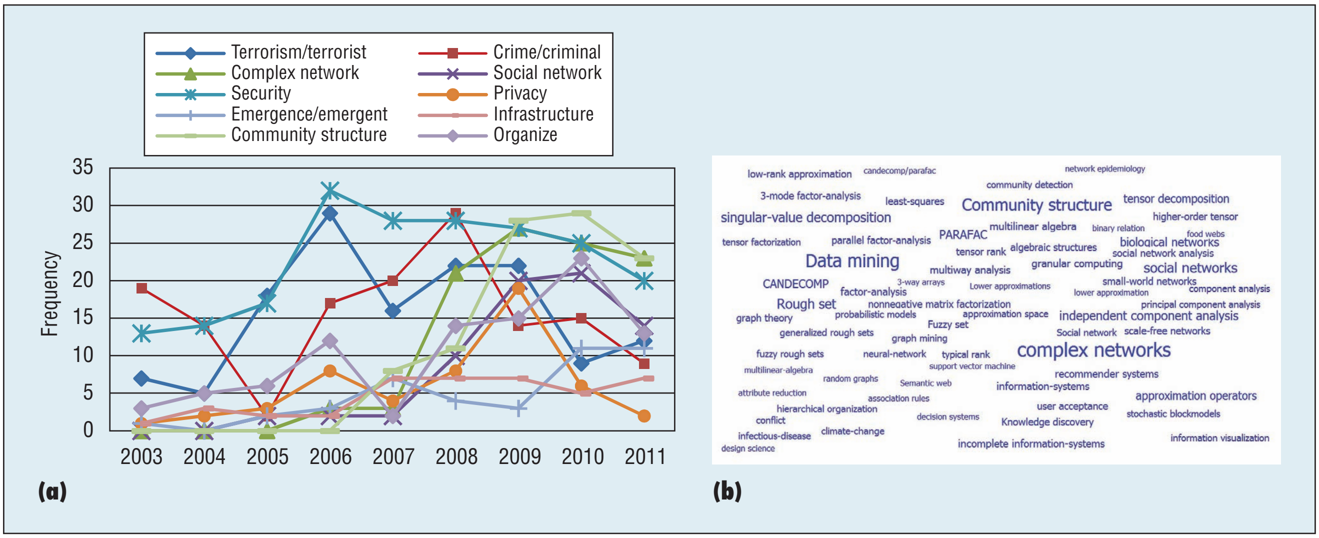

Figure 6. Topic analysis in ISI. (a) Evolution of 10 ISI keywords from 2003 to 2011, and (b) ISI keyword tag cloud of 2011.

from all over the world. However, what does this field care about and what topics do the researchers study?

To answer this question, we examined every paper's keywords. In general, the keywords in papers were extracted from the title, abstract, or the content itself. Thus, we assumed that the keywords of a paper represented the topics which authors cared about. If papers have the same keywords, we assumed that they discussed the same topic. A keyword with higher frequency means that researchers cared more about that particular topic. Here, different topics are described by using different sets of related keywords.

Figure 6a shows the trends of high-frequency keywords. We can clearly see that research on *security* and *terrorism/terrorist* both had a peak in 2006. After 2008, they both experienced a decline. Studies on *organize*, *privacy*, *community structure*, *complex network* and *social network* stayed at the same level from 2003 to 2007. In this stage, research interested in these words remained stable. After 2007, there was a rapid increase in research related to these words.

To show the topics that ISI researchers most cared about in recent years, we selected parts of keywords, with a frequency higher than 7, to make an ISI keyword tag cloud (see Figure 6b). In this keyword tag cloud, the size of the keywords are not only related to their frequency, but also related to the degree, which is calculated

THE AUTHORS

Wenli Liu is a PhD candidate in the Center of Military Computational Experiments and Parallel Systems Technology, National University of Defense Technology, China. His research interests include data mining, business intelligence, and parallel management theory. Liu has an MSc in engineering from the National University of Defense and Technology. Contact him at lwl_david@163.com.

Xiaolong Zheng is an associate professor in the State Key Laboratory of Management and Control for Complex Systems at the Chinese Academy of Sciences' Institute of Automation. His research interests include social computing and data mining. Zheng has a PhD in Control Theory and Control Engineering. He is the corresponding author of this article; contact him at xiaolong.zheng@ia.ac.cn.

Tao Wang is a PhD candidate in the Center of Military Computational Experiments and Parallel Systems Technology, National University of Defense Technology, China. His research interests include social computing, link prediction, link mining, and parallel management theory. Wang has an MSc in engineering from the National University of Defense and Technology. Contact him at peter_king171@126.com.

Hui Wang is a professor in the Center of Military Computational Experiments and Parallel Systems Technology, National University of Defense Technology, China. His research interests include social computing and data mining. Wang has a PhD in Management Science and Engineering. Contact him at huiwang@nudt.edu.cn.

in a co-keywords network. So, the size of a keyword indicates the number of papers that used this keyword.

As Figure 6b shows, the keywords *complex networks*, *community structure*, and *data mining* are clearly keywords related to topics that are of high importance. We search back into the dataset and find that there are 285 papers that contain these keywords. The types of papers include journey articles, conference papers, and book chapters. And their related researches include *terrorism informatics*, *information system security*, *infrastructure protection*, and *data/text mining*.

We investigated the researcher collaboration patterns and research topics using social network analysis methods from different levels and perspectives. By analyzing topological characteristics, we obtained that the degree distribution of ISI coauthorship networks are scale-free and exhibit small-world effects. Several researchers—such as Hsinchun Chen, Fei-Yue Wang, and Daniel Zeng—have a large number of collaborators. These researchers have conducted some useful studies and played a significant role in promoting the development of ISI as a field. Two institutions, including the University of Arizona and the Chinese Academy of Sciences, have been found to be the bridges linking other components. We also discerned that researchers in ISI have increasingly paid attention to three topics: complex networks, data mining, and community structure, respectively. We believe that these results can provide significant insights for researchers to better understand the development of ISI and quickly grasp emerging directions in this field. ■

Acknowledgments

The authors would like to thank professors Fei-Yue Wang and Xin Zhang for their fruitful comments. This work was supported in part by the National Natural Science Foundation of China under grants 71103180, 91124001, and 61105124, and by the Research Fund for the Doctoral Program of Higher Education of China under grant 20114307120023.

References

1. H.C. Chen et al., "Intelligence and Security Informatics," LNCS 3495, Springer. 2005.
2. H.C. Chen, F.Y. Wang, and D. Zeng, "Intelligence and Security Informatics for Homeland Security: Information, Communication, and Transportation," *IEEE Trans. Intelligent Transportation Systems*, vol. 5, no. 4, 2004, pp. 329–341.
3. F.Y. Wang and J. Wang, "Intelligence and Security Informatics: The State of the Art and Outlook," *China Basic Science*, no. 2, 2005, pp. 24–29.
4. H.C. Chen and F.Y. Wang, "Artificial Intelligence for Homeland Security," *IEEE Intelligent Systems*, vol. 20, no. 5, 2005, pp. 12–16.
5. W. Zhu and F.Y. Wang, "Reduction and Axiomization of Covering Generalized Rough Sets," *Information Sciences*, vol. 152, 2003, pp. 217–230.
6. H.C. Chen, "Intelligence and Security Informatics: Information Systems Perspective," *Decision Support Systems*, vol. 41, no. 3, 2006, pp. 555–559.
7. F.Y. Wang, "Toward a Paradigm Shift in Social Computing: The ACP Approach," *IEEE Intelligent Systems*, vol. 22, no. 5, 2007, pp. 65–67.
8. F.Y. Wang et al., "Social Computing: From Social Informatics to Social Intelligence," *IEEE Intelligent Systems*, vol. 22, no. 2, 2007, pp. 79–83.
9. H. Mo and F.Y. Wang, "Linguistic Dynamic Systems Based on Computing with Words and Their Stabilities," *Science in China Series F: Information Sciences*, vol. 52, no. 5, 2009, pp. 780–796.
10. F.Y. Wang et al., "A Study of the Human Flesh Search Engine: Crowd-Powered Expansion of Online Knowledge," *Computer*, vol. 43, no. 8, 2010, pp. 45–53.
11. M.E.J. Newman, "Coauthorship Networks and Patterns of Scientific Collaboration," *Proc. Nat'l Academy of Sciences of the United States of America*, 2004; doi:10.1073/pnas.0307545100.
12. L.J. Li et al., "Research Collaboration and ITS Topic Evolution: 10 Years at T-ITS," *IEEE Trans. Intelligent Transportation Systems*, vol. 11, no. 3, 2010, pp. 517–523.
13. T. Wang et al., "On Social Computing Research Collaboration Patterns: A Social Network Perspective," *Frontiers of Computer Science*, vol. 6, no. 1, 2012, pp. 122–130.

14. D. Jansen, R. von Gortz, and R. Heidler, "Knowledge Production and the Structure of Collaboration Networks in Two Scientific Fields," *Scientometrics*, vol. 83, no. 1, 2010, pp. 219–241.
15. R. Guimera et al., "Team Assembly Mechanisms Determine Collaboration Network Structure and Team Performance," *Science*, vol. 308, no. 5722, 2005, pp. 697–702.
16. C.M. Chen, "CiteSpace II: Detecting and Visualizing Emerging Trends and Transient Patterns in Scientific Literature," *J. Am. Soc. for Information Science and Technology*, vol. 57, no. 3, 2006, pp. 359–377.
17. E.F. Reid and H.C. Chen, "Mapping the Contemporary Terrorism Research Domain," *Int'l J. Human-Computer Studies*, vol. 65, no. 1, 2007, pp. 42–56.
18. K.K. Mane and K. Borner, "Mapping Topics and Topic Bursts in PNAS," *Proc. Nat'l Academy of Sciences of the United States of America*, 2004; doi:10.1073/pnas.0307626100.
19. P.J.H. Hu et al., "System for Infectious Disease Information Sharing and Analysis: Design and Evaluation," *IEEE Trans. Information Technology in Biomedicine*, vol. 11, no. 4, 2007, pp. 483–492.

cn *Selected CS articles and columns are also available for free at http://ComputingNow.computer.org.*