RGC Ref.: M-CityU107/13
(please insert ref. above)

The Research Grants Council of Hong Kong
SRFDP & RGC ERG Joint Research Scheme
Completion Report

(Please attach a copy of the completion report submitted to the Ministry of Education by the Mainland researcher)

Part A: The Project and Investigator(s)

1. Project Title

Inference of Large Epidemic-like Information Spreading: Theories and Network Forensics

2. Investigator(s) and Academic Department/Units Involved

	Hong Kong Team	Mainland Team
Name of Principal Investigator *(with title)*	Dr Chee Wei TAN	Prof Wenyi ZHANG
Post	Associate Professor	Professor
Unit / Department / Institution	Department of Computer Science, City University of Hong Kong	Department of Electronic Engineering and Information Science, University of Science and Technology of China
Contact Information		
Co-investigator(s) *(with title and Institution)*		
PhD student(s) (with period of involvement)	Name: Liang Zheng, Chen Jiaxing, Lou Xin Institution: City University of Hong Kong Period from 1 Jan 2014 to 31 Dec 2016	Name: Dong Wenxiang, Zhaoxu Wang Institution: USTC, China Period from 1 Jan 2014 to 31 Dec 2016

Note: The Hong Kong project team must involve at least one research postgraduate student pursuing a Doctor of Philosophy degree at the UGC-funded university (PhD student) at any time throughout the project period.

3. Project Duration

	Original	Revised	Date of RGC/

			Institution Approval *(must be quoted*)
Project Start date	1-Jan-2014		1-Jan 2014
Project Completion date	31-Dec-2016		
Duration *(in month)*	36		
Deadline for Submission of Completion Report	30-Sep-2017		

Part B: The Completion Report

5. Project Objectives

5.1 Objectives as per original application

Develop mathematical theories for inference of epidemic-like information spreading in large networks that exploit network connectivity and suspect characteristics.

2. Develop scalable and reliable algorithms to identify a single and multiple epidemic-like information sources in probabilistic spreading models.

3. Develop distributed source inference algorithms for very large networks with general heterogeneous spreading models.

4. Understand similarities and differences between epidemic spreading in online social-technological networks such as Twitter over mobile phone networks and classical probabilistic spreading models in epidemiology.

5. Develop distributed network forensics tools and protocols for online social-technological networks using Twitter and Weibo dataset to detect epidemic-like information spreading.

5.2 Revised Objectives

Date of approval from the RGC: _____

Reasons for the change: _____

_____

1.
2.
3. ....

6. Research Outcome

Major findings and research outcome
(maximum 1 page; please make reference to Part C where necessary)

Our major findings were to answer the question: “How to identify the source of the spreading given an observation of the infected nodes and the underlying network structure”, which remains a challenging open problem. This has important applications in cyber-security, for example, law enforcement agencies may be interested in identifying the perpetrators of false information that are virally spread in online social networks and used to manipulate the market prices of certain stocks.

Our research outcome has led to the rigorous development of mathematical theories and distributed algorithms to identify infection sources in a network, and quantify the detection performance in terms of the network structure by leveraging on probability theory, and, particularly, to utilize a novel link between Polya's urn model in probability theory and statistical inference with multiple observation data graphs to quantify the

exact detection performance in large networks. This research outcome has led to publication in 2014 ACM Sigmetrics conference and 2015 IEEE Journal on Selected Topics in Signal Processing (both conference and journal are top tiered with high impact factors). We have also developed network forensics algorithms to make insightful predictions in large-scale online social networks such as Twitter and Facebook. This research outcome has led to publication in 2016 IEEE Transactions on Knowledge and Data Engineering – a top-tiered journal with high impact factor. Other research outcome has led to joint collaborative work using Facebook data work in a 2016 CISS conference with its journal publication and also an authoritative monograph under preparation. Our research is expected to play important role in many security applications to combat cyber-crimes.

Potential for further development of the research and the proposed course of action *(maximum half a page)*

A recent 2017 article in the reputable Scientific American publication has analyzed the spread of memes and fake news in online social networks and highlighted that this is indeed an important research area that requires much more mathematical analysis and study (available from https://www.scientificamerican.com/article/how-fake-news-goes-viral-mdash-heres-the-math/). Our research does indeed lay down important mathematical foundations that go towards building more robust and reliable online social networks for transmitting information. Given that nowadays many people do indeed receive their news from sorely online social networks such as Twitter and Facebook, combatting against fake news, especially malicious ones, in online social networks will remain a long standing problem to come. This is even more important when we take geo-political constraints (e.g. Mainland China has different online social network platforms that differ from some other countries) into account, the SRFDP project has built important collaborative relationships between the two PI's in fundamental theoretical research and system work in an important area of cyber-security. Further proposed course of action includes continuing work to generate more joint research publications and to prepare for the application of further joint research project schemes, e.g., NSFC/RGC research scheme.

7. The Layman's Summary

(describe in layman's language the nature, significance and value of the research project, in no more than 200 words)

Detection and identification of malicious information sources in a network, be it in the case of a computer virus spreading in Internet or a misinformation or rumor propagating in an online social network, allows timely quarantine of the epidemic-like spreading to limit the damage caused. For example, law enforcement agencies may be interested in identifying the perpetrators of false information used to manipulate the market prices of certain stocks. How to identify the source of the spreading given an observation of the infected nodes and the underlying network structure remains a rather unexplored and challenging problem.

We have developed mathematical theories and distributed algorithms to identify infection sources in a network, and quantify the detection performance in terms of the network

structure by leveraging on probability theory, and, particularly, to utilize a novel link between Polya's urn model in probability theory and statistical inference with multiple observation data graphs to quantify the exact detection performance in large networks. We have also developed network forensics algorithms to make insightful predictions in large-scale online social networks such as Twitter and Facebook. Our research is expected to play important role in many security applications to combat cyber-crimes.

Part C: Research Output

8. Peer-reviewed journal publication(s) arising directly from this research project

(Please attach a copy of each publication and/or the letter of acceptance if not yet submitted in the previous progress report(s). All listed publications must acknowledge RGC's funding support by quoting the specific grant reference.))

The Latest Status of Publications				Author(s) ***(bold** the authors belonging to the project teams and denote the corresponding author with an asterisk*)*	Title and Journal/ Book *(with the volume, pages and other necessary publishing details specified)*	Submitted to RGC *(indicate the year ending of the relevant progress report)*	Attached to this report *(Yes or No)*	Acknowledged the support of this Joint Research Scheme *(Yes or No)*	Accessible from the institutional repository *(Yes or No)*
Year of publication	Year of Acceptance *(For paper accepted but not yet published)*	Under Review	Under Preparation *(optional)*						
2015				Zhaoxu Wang, Wenxiang Dong, Wenyi Zhang* and Chee Wei Tan	Rooting out Rumor Sources in Online Social Networks: The Value of Diversity from Multiple Observations, **IEEE Journal of Selected Topics in Signal Processing**, vol. 9, no. 4, pp. 663-677, 2015.	Yes	No	Yes Page 1	Yes
2016				Felix Ming Fai Wong, Chee Wei Tan*, Soumya Sen and Mung Chiang	Quantifying Political Leaning from Tweets, Retweets and Retweeters, **IEEE Transactions on Knowledge and Data Engineering**, Vol. 28, No. 8, pp. 2158-2172, 2016.	No	Yes	Yes Page 13	Yes
			2018	Pei-Duo Yu, Chee Wei Tan, Ching Nam Hang, Wenyi Zhang and Hung-Lin Fu	Optimal Detection of Influential Spreaders in Online Social Networks, **IEEE/ACM Transactions on Networking** (under preparation)	No	No	No	No

9. Recognized international conference(s) in which paper(s) related to this research project was/were delivered *(Please attach a copy of each delivered paper. All listed papers must acknowledge RGC's funding support by quoting the specific grant reference.)*

Month/Year/ Place	Title	Conference Name	Submitted to RGC *(indicate the year ending of the relevant progress report)*	Attached to this report *(Yes or No)*	Acknowledged the support of this Joint Research Scheme *(Yes or No)*	Accessible from the institutional repository *(Yes or No)*
June/2014/Texas	Rumor Source Detection with Multiple Observations: Fundamental Limits and Algorithms	2014 ACM Special Interest Group on Measurement and Evaluation (SIGMETRICS)	Yes	No	Page 12	June/2014/Texas
July/2015/Singapore	A Probabilistic Characterization of the Rumor Graph Boundary in Rumor Source Detection	2015 IEEE Digital Signal Processing Conference	Yes	No	Page 1	July/2015/Singapore
July/2015/Singapore	On Inferring Rumor Source for SIS Model under Multiple Observations	2015 IEEE Digital Signal Processing Conference	Yes	No	Page 1	July/2015/Singapore
March/2016/ Princeton	Optimal Detection of Influential Spreaders in Online Social Networks	2016 Annual Conference on Information Science and Systems	Yes	Yes	Page 1	March/2016/ Princeton

10. Student(s) trained *(Please attach a copy of the title page of the thesis.)*

Name	Degree registered for	Date of registration	Date of thesis submission/ graduation
Zheng Liang	PhD	29-05-2013	29-5-2015
Chen Jiaxing	PhD	1-10-2014	30-9-2018
Lou Xin	PhD	1-6-2012	30-11-2015

11. Other impact *(e.g. award of patents or prizes, collaboration with other research institutions, technology transfer, etc.)*

The project has led to the following prizes received by Prof. Tan, Chee Wei:

1) Recipient of a 2014 Young Scientist Award from the URSI International Union of Radio Science. Official announcement at: http://www.ursi.org/en/young_scientists.asp?year=2014

2) Selected Participant as Outstanding Scientist at the **2015 U.S. National Academy of Engineering China-America Frontiers of Engineering Symposium** at Irvine, California in June 2015. Selection was coordinated by the U.S. National Academy of Engineering.
Presented a Poster on "Cyber-security and web intelligence in complex engineered networks". Official announcement at:
http://www.naefrontiers.org/Symposia/CAFOE/25153/44847.aspx
http://www.naefrontiers.org/File.aspx?id=49727

3) Selected Participant as Outstanding Scientist at the 2015 1st Indian Academy of Engineering- China Academy of Engineering Youth Engineering Leader Conclave (ICCON-I) at Gandhinagar, India in October 2015. One out of seventeen Chinese Young Scientists representing China. Selection was coordinated by the China Academy of Engineering.
Presented a Poster on "Cyber-security and web intelligence in complex engineered networks".

We have collaborated with the Electrical Engineering Department at Princeton University on the viral spread of information in Twitter online social networks.