REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
20-06-2016	Final Report	1-Sep-2011 - 31-Aug-2014

4. TITLE AND SUBTITLE
Final Report: Multi-modal Social Networks A MRF Learning Approach

5a. CONTRACT NUMBER
W911NF-11-1-0265

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
611102

6. AUTHORS
Sanjay Shakkottai, Sujay Sanghavi

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
University of Texas at Austin
101 East 27th Street
Suite 5.300
Austin, TX 78712 -1532

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
59441-NS.14

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT
The work primarily focused on two lines of research.

1. We propose new greedy algorithms for learning the structure of a graphical model of a probability distribution, given samples drawn from the distribution. Our research modifies greedy algorithms through appropriate node pruning, to result in fast algorithms that provide analytical guarantees on correctness.

15. SUBJECT TERMS
social networks, learning and inference

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Sanjay Shakkottai
UU	UU	UU	UU		19b. TELEPHONE NUMBER 512-471-5376

Report Title

Final Report: Multi-modal Social Networks A MRF Learning Approach

ABSTRACT

The work primarily focused on two lines of research.

1. We propose new greedy algorithms for learning the structure of a graphical model of a probability distribution, given samples drawn from the distribution. Our research modifies greedy algorithms through appropriate node pruning, to result in fast algorithms that provide analytical guarantees on correctness.

2. The objective of this line of work is to use noisy measurements from cascades – stochastic processes for spread on graphs – to learn the spread of information / opinion / malware. Our approach for this learning problem has been to view this as hypothesis testing on graphs – given noisy and partial information on both node states and the network graph, we formulate the problem as distinguishing between a benign hypothesis (no spreading process) and a malicious hypothesis (spreading process such as malware). This approach has been used in a sequence of studies, starting from distinguishing with partial information, to that with nodes with are adversarial (nodes could lie about their state), to dealing with noisy network knowledge. We have also been able to use this approach to learn the identity of communities with shared interests.

Enter List of papers submitted or published that acknowledge ARO support from the start of the project to the date of this printing. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

Received Paper

TOTAL:

Number of Papers published in peer-reviewed journals:

(b) Papers published in non-peer-reviewed journals (N/A for none)

Received Paper

TOTAL:

Number of Papers published in non peer-reviewed journals:

(c) Presentations

Number of Presentations:

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received		Paper
06/09/2016	8.00	. Localized Epidemic Detection in Networks with Overwhelming Noise, the 2015 ACM SIGMETRICS International Conference. 15-JUN-15, Portland, Oregon, USA. : ,
06/09/2016	11.00	. Epidemic thresholds with external agents, IEEE INFOCOM 2014 - IEEE Conference on Computer Communications. 27-APR-14, Toronto, ON, Canada. : ,
06/09/2016	12.00	. Topic modeling from network spread, The 2014 ACM international conference. 16-JUN-14, Austin, Texas, USA. : ,
06/20/2016	13.00	. Network forensics: random infection vs spreading epidemic, Proceedings of ACM Sigmetrics. 11-JUN-12, London, UK. : ,
TOTAL:	**4**	

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received		Paper
06/09/2016	4.00	Constantine Caramanis, Shie Mannor, Sanjay Shakkottai, Chris Milling. Detecting epidemics using highly noisy data, the fourteenth ACM international symposium. 28-JUL-13, Bangalore, India. : ,
TOTAL:	**1**	

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts):

(d) Manuscripts

Received	Paper
TOTAL:	

Number of Manuscripts:

Books

Received Book

TOTAL:

Received Book Chapter

TOTAL:

Patents Submitted

Patents Awarded

Awards

Graduate Students

NAME	PERCENT_SUPPORTED	Discipline
Avik Ray	0.25	
Chris Milling	0.20	
Siddhartha Banerjee	0.20	
Praneeth Netrapalli	0.10	
Sharayu Moharir	0.05	
FTE Equivalent:	**0.80**	
Total Number:	**5**	

Names of Post Doctorates

NAME	PERCENT_SUPPORTED
Javad Ghaderi	0.05
FTE Equivalent:	**0.05**
Total Number:	**1**

Names of Faculty Supported

NAME	PERCENT_SUPPORTED	National Academy Member
Sanjay Shakkottai	0.08	
Sujay Sanghavi	0.08	
FTE Equivalent:	**0.16**	
Total Number:	**2**	

Names of Under Graduate students supported

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period: 0.00

The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:..... 0.00

The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:..... 0.00

Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):..... 0.00

Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:..... 0.00

The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense 0.00

The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields:..... 0.00

Names of Personnel receiving masters degrees

NAME
Total Number:

Names of personnel receiving PHDs

NAME
Siddhartha Banerjee
Praneeth Netrapalli
Chris Milling
Total Number: **3**

Names of other research staff

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Sub Contractors (DD882)

Inventions (DD882)

Scientific Progress

Please see attached document for details of progress and accomplishments.

Technology Transfer

Greedy Learning of Graphical Models

We propose new greedy algorithms for learning the structure of a graphical model of a probability distribution, given samples drawn from the distribution. While structure learning of graphical models is a widely studied problem with several existing methods, greedy approaches remain attractive due to their low computational cost.

The most natural greedy algorithm would be one which, essentially, adds neighbors to a node in sequence until stopping; it would do this for each node. While it is fast, simple and parallel, this naive greedy algorithm has the tendency to add non-neighbors that show high correlations with the given node. Our new algorithms overcome this problem in three different ways. The *recursive greedy algorithm* iteratively recovers the neighbors by running the greedy algorithm in an inner loop, but each time only adding the last added node to the neighborhood set. The second *forward-backward greedy* algorithm includes a node deletion step in each iteration that allows non-neighbors to be removed from the neighborhood set which may have been added in previous steps. Finally, the *greedy algorithm with pruning* runs the greedy algorithm until completion and then removes all the incorrect neighbors. We provide both analytical guarantees and empirical performance for our algorithms. We show that in graphical models with strong non-neighbor interactions, our greedy algorithms can correctly recover the graph, whereas the previous greedy and convex optimization based algorithms do not succeed.

A. Ray, S. Sanghavi and S. Shakkottai, "Improved Greedy Algorithms for Learning Graphical Models". *IEEE Transactions on Information Theory*, Volume 61, Number 6, pp. 3457 – 3468, June 2015.

Epidemic Spread and Detection

Objective: The objective of this project is to use noisy measurements from cascades – stochastic processes for spread on graphs – to infer the spread of information / opinion / malware.

Approach: Our approach has been to treat the problem as hypothesis testing on graphs – given noisy and partial information on both node states and the network graph, we formulate the problem as distinguishing between a benign hypothesis (no spreading process) and a malicious hypothesis (spreading process such as malware).

Scientific Barriers: The high dimensionality of the problem, along with noisy, partial and potentially adversarial node information renders the classical maximum likelihood approach to be intractable. We have instead developed novel methods based on concentration of measures on graphs to develop simple algorithms (low degree polynomial with respect to graph size) to distinguish between the hypotheses, and with theoretical guarantees on performance (in the regime where the network size is large).

Significance: This binary hypothesis view has been used in a sequence of works, starting from distinguishing with partial information, to that with nodes with are adversarial (nodes could lie about their state), to dealing with noisy network knowledge. We have also been able to use this approach to learn the identity of communities with shared interests (topic modeling with network

state). Finally, we have used these theoretical ideas to develop a practical platform to detect the presence of malware on smartphone platforms.

As part of this broader study, we have also developed new understanding of epidemic spread with a variety of assumptions (e.g. what if the spread has bounded susceptibility, and there are adversarial nodes that assist the spread).

Accomplishments: A number of papers in premier conferences and journals in the field (listed below). Further, a student working on this project (Dr. Siddhartha Banerjee) graduated with his Ph.D. and has joined the Operations Research and Information Engineering (ORIE) Department at Cornell University as a tenure-track Assistant Professor.

"Localized epidemic detection in networks with overwhelming noise", E. Meirom, C. Caramanis, S. Mannor, S. Shakkottai, A. Orda, *Proceedings of ACM Sigmetrics* (poster paper), Portland, OR, June 2015.

"Distinguishing Infections on Different Graph Topologies", C. Milling, C. Caramanis, S. Mannor and S. Shakkottai. *IEEE Transactions on Information Theory*, Vol. 61, No. 6, June 2015.

"Local Detection of Infections in Heterogeneous Networks", C. Milling, C. Caramanis, S. Mannor and S. Shakkottai, *Proc. of IEEE Infocom*, Hong Kong, 2015.

S. Banerjee, A. Gopalan, A. Das and S. Shakkottai, "Epidemic Spreading with External Agents". *IEEE Transactions on Information Theory,* Volume 60, Issue 7, pp. 4125 – 4138, July 2014.

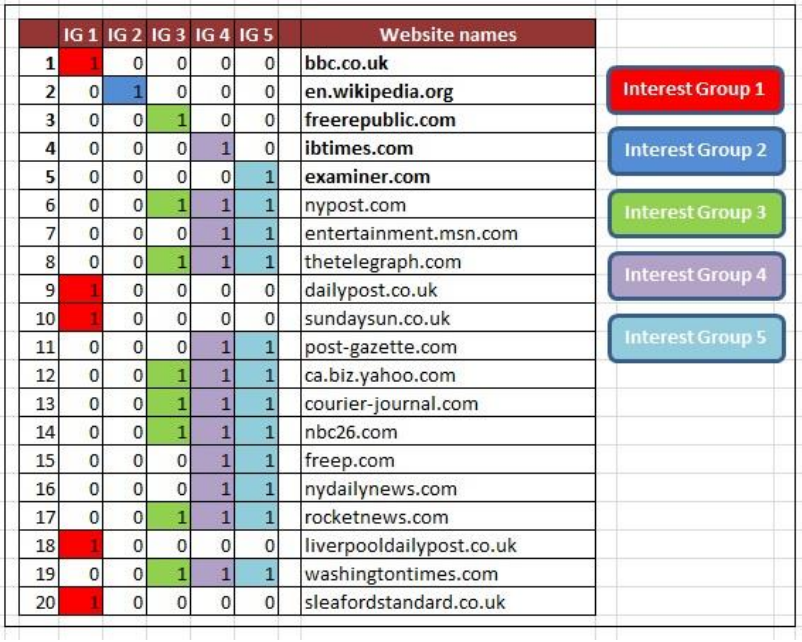
S. Banerjee, A. Chatterjee and S. Shakkottai, "Epidemic Thresholds with External Agents". *Proceedings of IEEE Infocom*, Toronto, Canada, April 2014. (19% acceptance)

S. Krishnasamy, S. Banerjee and S. Shakkottai, "The Behavior of Epidemics under Bounded Suceptability". *Proceedings of ACM Sigmetrics,* Austin, TX June 2014. (17% acceptance)

"Topic Modeling from Network Spread," A. Ray, S. Sanghavi and S. Shakkottai. *Proceedings of ACM Sigmetrics* (poster paper), Austin, TX June 2014.

Conclusions: Our approach based on hypothesis testing on graphs provides a new tool for detecting malware (or more generally, spreading processes). This has a variety of applications, ranging from detecting communities with shared interests in networks, to detecting new malware. We have explored various applications using real datasets, and shown the benefits of our approach.

Image:



	IG 1	IG 2	IG 3	IG 4	IG 5	Website names
1	1	0	0	0	0	**bbc.co.uk**
2	0	1	0	0	0	**en.wikipedia.org**
3	0	0	1	0	0	**freerepublic.com**
4	0	0	0	1	0	**ibtimes.com**
5	0	0	0	0	1	**examiner.com**
6	0	0	1	1	1	nypost.com
7	0	0	0	1	1	entertainment.msn.com
8	0	0	1	1	1	thetelegraph.com
9	1	0	0	0	0	dailypost.co.uk
10	1	0	0	0	0	sundaysun.co.uk
11	0	0	0	1	1	post-gazette.com
12	0	0	1	1	1	ca.biz.yahoo.com
13	0	0	1	1	1	courier-journal.com
14	0	0	1	1	1	nbc26.com
15	0	0	0	1	1	freep.com
16	0	0	0	1	1	nydailynews.com
17	0	0	1	1	1	rocketnews.com
18	1	0	0	0	0	liverpooldailypost.co.uk
19	0	0	1	1	1	washingtontimes.com
20	1	0	0	0	0	sleafordstandard.co.uk

Topic modeling from network spread – new algorithms for determining topics for each website using network spread models. The dataset for this study is from the Stanford Network Analysis Project (SNAP). The algorithm ideas are described in the paper: “Topic Modeling from Network Spread,” A. Ray, S. Sanghavi and S. Shakkottai. *Proceedings of ACM Sigmetrics* (poster paper), Austin, TX June 2014.