REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
31-08-2016	Final Report	9-Jan-2012 - 8-Jan-2015

4. TITLE AND SUBTITLE
Situational Awareness for Social Media: Theories, Models and Algorithms

5a. CONTRACT NUMBER
W911NF-12-1-0034

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHORS
Aram Galstyan, Kristina Lerman, Eduard Hovy, Yan Liu, Ram Nevatia

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
University of Southern California
Department of Contracts and Grants
3720 South Flower Street
Los Angeles, CA 90089 -0701

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)
61769-NS-DRP.78

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for Public Release; Distribution 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 goal of this project is to develop technological capabilities
for monitoring social media platforms to detect propaganda campaigns and influence operations and predict their effect. The intended outcome of this research will be set of automated and semi–automated tools to monitor emergence of topics and memes, predict their spread, detect attempts to propagate rumors and misinformation, recognize participant activities and intent, and detect covert attempts to manipulate public opinion.

15. SUBJECT TERMS
social media, social networks

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Aram Galstyan
UU	UU	UU	UU		19b. TELEPHONE NUMBER 310-448-9183

Report Title

Situational Awareness for Social Media: Theories, Models and Algorithms

ABSTRACT

The goal of this project is to develop technological capabilities
for monitoring social media platforms to detect propaganda campaigns and influence operations and predict their effect. The intended outcome of this research will be set of automated and semi–automated tools to monitor emergence of topics and memes, predict their spread, detect attempts to propagate rumors and misinformation, recognize participant activities and intent, and detect covert attempts to manipulate public opinion.

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
08/24/2015	53.00	Shuyang Gao, Greg Ver Steeg, Aram Galstyan. Understanding Confounding Effects in Linguistic Coordination: An Information-Theoretic Approach, PLoS ONE, (06 2015): 0. doi: 10.1371/journal.pone.0130167
08/25/2015	61.00	Rumi Ghosh, Kristina Lerman. The Impact of Network Flows on Community Formation in Models of Opinion Dynamics, The Journal of Mathematical Sociology, (03 2015): 0. doi: 10.1080/0022250X.2014.905776
08/27/2012	2.00	Kristina Lerman, Tad Hogg. Social dynamics of Digg, EPJ Data Science , (06 2012): 1. doi: 10.1140/epjds5
08/27/2012	1.00	Kristina Lerman, Rumi Ghosh. Network structure, topology, and dynamics in generalized models of synchronization, Physical Review E, (08 2012): 0. doi: 10.1103/PhysRevE.86.026108
08/28/2014	41.00	P. Jeffrey Brantingham, Aram Galstyan, Yoon-Sik Cho, George Tita. Latent self-exciting point process model for spatial-temporal networks, Discrete and Continuous Dynamical Systems - Series B, (04 2014): 0. doi: 10.3934/dcdsb.2014.19.1335
08/28/2014	42.00	Vasanthan Raghavan, Greg Ver Steeg, Aram Galstyan, Alexander G. Tartakovsky. Modeling Temporal Activity Patterns in Dynamic Social Networks, , (03 2014): 0. doi: 10.1109/TCSS.2014.2307453
08/28/2014	43.00	Aram Galstyan, Pablo Branas-Garza, Armen E. Allahverdyan. Opinion Dynamics with Confirmation Bias, PLoS ONE, (07 2014): 0. doi: 10.1371/journal.pone.0099557
08/28/2014	44.00	Kristina Lerman, Tad Hogg, Hussein Suleman. Leveraging Position Bias to Improve Peer Recommendation, PLoS ONE, (06 2014): 0. doi: 10.1371/journal.pone.0098914
08/28/2014	45.00	Rumi Ghosh, Kristina Lerman. Rethinking centrality: The role of dynamical processes in social network analysis, Discrete and Continuous Dynamical Systems - Series B, (04 2014): 0. doi: 10.3934/dcdsb.2014.19.1355
08/28/2014	46.00	Nathan O. Hodas, Kristina Lerman. The Simple Rules of Social Contagion, Scientific Reports, (03 2014): 0. doi: 10.1038/srep04343
08/28/2014	48.00	Cristina Garcia-Cardona, Allon G. Percus, Kristina Lerman, Rumi Ghosh, Laura M. Smith. Spectral clustering with epidemic diffusion, Physical Review E, (10 2013): 0. doi: 10.1103/PhysRevE.88.042813

TOTAL: 11

Number of Papers published in peer-reviewed journals:

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

Received	Paper
08/27/2013 38.00	Georgios Fellouris, Alexander G. Tartakovsky. ALMOST OPTIMAL SEQUENTIAL TESTS OF DISCRETE COMPOSITE HYPOTHESES, Statistica Sinica, (12 2013): 0. doi:
TOTAL:	**1**

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
TOTAL:	

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

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

Received	Paper
08/24/2015 54.00	Yoon Sik Cho, Greg Ver Steeg, Aram Galstyan. Where and Why Users “Check In", Twenty-Eighth AAAI Conference on Artificial Intelligence. 28-JUL-14, . : ,
08/24/2015 58.00	Greg Ver Steeg, Aram Galstyan. Maximally Informative Hierarchical Representations of High-Dimensional Data, The 18th International Conference on Artificial Intelligence and Statistics (AISTATS) . 11-MAY-15, . : ,
08/24/2015 56.00	Shuyang Gao, Greg Ver Steeg, Aram Galstyan. Estimating Mutual Information by Local Gaussian Approximation , The 31st Conference on Uncertainty in Artificial Intelligence . 13-JUL-15, . : ,
08/24/2015 55.00	Shuyang Gao, Greg Ver Steeg, Aram Galstyan. Efficient Estimation of Mutual Information for Strongly Dependent Variables, International Conference on Artificial Intelligence and Statistics (AISTATS) . 11-MAY-15, . : ,
08/25/2015 67.00	Philipp Geiger, Kun Zhang, Bernhard Schoelkopf, Mingming Gong, Dominik Janzing . Causal Inference by Identification of Vector Autoregressive Processes with Hidden Components, The 32nd International Conference on Machine Learning. 06-JUL-15, . : ,
08/25/2015 73.00	Xinran He, Theodoros Rekatsinas, James Foulds, Lise Getoor, Yan Liu . HawkesTopic: A Joint Model for Network Inference and Topic Modeling from Text-Based Cascades, The 32nd International Conference on Machine Learning. 06-JUL-15, . : ,
08/25/2015 72.00	Rose Yu, Dehua Cheng, Yan Liu . Accelerated Online Low Rank Tensor Learning for Multivariate Spatiotemporal Streams, The 32nd International Conference on Machine Learning. 06-JUL-15, . : ,
08/25/2015 70.00	Greg Ver Steeg, Aram Galstyan. Discovering Structure in High-Dimensional Data Through Correlation Explanation, Advances in Neural Information Processing Systems . 08-DEC-14, . : ,
08/25/2015 69.00	Mohammad T. Bahadori, Qi Rose Yu, Yan Liu. Fast Multivariate Spatio-temporal Analysis via Low Rank Tensor Learning, Advances in Neural Information Processing Systems. 08-DEC-14, . : ,
08/25/2015 66.00	Mingming Gong, Kun Zhang, Bernhard Schoelkopf, Dacheng Tao, Philipp Geiger . Discovering Temporal Causal Relations from Subsampled Data, The 32nd International Conference on Machine Learning. 06-JUL-15, . : ,
08/25/2015 65.00	Sidharth Gupta, Xiaoran Yan, Kristina Lerman. Structural Properties of Ego Networks, International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction . 31-MAR-15, . : ,
08/25/2015 64.00	Jeon-Hyung Kang, Kristina Lermam. VIP: Incorporating Human Cognitive Biases in a Probabilistic Model of Retweeting, International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction. 31-MAR-15, . : ,

08/25/2015 63.00 Farshad Kooti, Luca Maria Aiello, Mihajlo Grbovic, Kristina Lerman, Amin Mantrach. Evolution of Conversations in the Age of Email Overload,
24th International World Wide Web Conference. 18-MAY-15, . : ,

08/25/2015 62.00 Jeon-Hyung Kang, Kristina Lerman. User Effort and Network Structure Mediate Access to Information in Networks,
The Ninth International AAAI Conference on Web and Social Media. 26-MAY-15, . : ,

08/25/2015 59.00 Rumi Ghosh, Shang-hua Teng, Kristina Lerman, Xiaoran Yan. The interplay between dynamics and networks,
the 20th ACM SIGKDD international conference. 23-AUG-14, New York, New York, USA. : ,

08/25/2015 60.00 Suradej Intagorn, Kristina Lerman. Placing user-generated content on the map with confidence,
the 22nd ACM SIGSPATIAL International Conference. 03-NOV-14, Dallas, Texas. : ,

08/25/2015 68.00 Kun Zhang, Jiji Zhang, Bernhard Schölkopf. Distinguishing Cause from Effect Based on Exogeneity,
Fifteenth conference onTheoretical Aspects of Rationality and Knowledge. 04-JUN-15, . : ,

08/27/2012 3.00 Greg Ver Steeg, Aram Galstyan. Information transfer in social media,
the 21st international conference. 15-APR-12, Lyon, France. : ,

08/27/2012 11.00 Macskassy, S.. Characterizing Retweeting Behaviors in Twitter: On the use of Text vs. Concepts,
ECML/PKDD workshop on Collective Learning and Inference on Structured Data (CoLISD) . 28-SEP-12, . : ,

08/27/2012 10.00 Macskassy, S.. On the Study of Social Interactions in Twitter,
Proceedings of the Sixth International Conference on Weblogs and Social Media. 05-JUN-12, . : ,

08/27/2012 9.00 Tartakovsky, A.G.. Sequential Hypothesis Testing and Changepoint Detection: Past and Future,
Proceedings of the International Conference on Stochastic Optimization and Optimal Stopping. 24-SEP-12, . : ,

08/27/2012 8.00 Kang, J.-H., Lerman, K.. Using Lists to Measure Homophily on Twitter,
AAAI workshop on Intelligent Techniques for Web Personalization and Recommendation. 22-JUL-12, . : ,

08/27/2012 7.00 Hodas, N., Lerman, K.. How Limited Visibility and Divided Attention Constrain Social Contagion,
ASE/IEEE International Conference on Social Computing. 03-SEP-12, . : ,

08/27/2012 6.00 Liu, Y., Bahadori, M.T.,, Li, H.. Sparse-GEV: Sparse Latent Space Model for Multivariate Extreme Value Time Series Modeling,
29th International Conference on Machine Learning. 26-JUN-12, . : ,

08/27/2012 5.00 Wenjun Zhou, Hongxia Jin, Yan Liu. Community discovery and profiling with social messages,
the 18th ACM SIGKDD international conference. 11-AUG-12, Beijing, China. : ,

08/27/2013 23.00 Yoon Sik Cho, Greg Ver Steeg, Aram Galstyan. Socially Relevant Venue Clustering from Check-in Data,
KDD Workshop on Mining and Learning with Graphs. 11-AUG-13, . : ,

08/27/2013 40.00 Vasanthan Raghavan, Greg ver Steeg, Aram Galstyan, Alexander G. Tartakovsky. COUPLED HIDDEN MARKOV MODELS FOR USER ACTIVITY IN SOCIAL NETWORKS,
ICME International Workshop on Social Multimedia Research 2013. 15-AUG-13, . : ,

08/27/2013 37.00 G. Fellouris, A.G. Tartakovsky. Unstructured Sequential Testing in Sensor Networks,
52nd IEEE Conference on Decision and Control, Florence. 12-DEC-13, . : ,

08/27/2013 36.00 Kristina Lerman, Prachi Jain, Rumi Ghosh, Jeon-Hyung Kang, Ponnurangam Kumaraguru. Limited Attention and Centrality in Social Networks,
International Conference on Social Intelligence and Technology (SOCIETY2013). 08-MAY-13, . : ,

08/27/2013 35.00 Laura M. Smith, Linhong Zhu, Kristina Lerman, Zornitsa Kozareva. The Role of Social Media in the Discussion of Controversial Topics,
SocialCom'13. 08-SEP-13, . : ,

08/27/2013 34.00 Jeon-Hyung Kang, Kristina Lerman. LA-CTR: A Limited Attention Collaborative Topic Regression for Social Media,
The Twenty-Seventh AAAI Conference on Artificial Intelligence. 14-JUL-13, . : ,

08/27/2013 33.00 Jeon-Hyung Kang, Kristina Lerman, Lise Getoor. LA-LDA: A Limited Attention Topic Model for Social Recommendation,
SBP'13 Proceedings of the 6th international conference on Social Computing, Behavioral-Cultural Modeling and Prediction . 03-APR-13, . : ,

08/27/2013 32.00 Kristina Lerman, Jeon-Hyung Kang. Structural and cognitive bottlenecks to information access in social networks,
the 24th ACM Conference. 30-APR-13, Paris, France. : ,

08/27/2013 31.00 Tad Hogg, Kristina Lerman, Laura M. Smith. Stochastic Models Predict User Behavior in Social Media,
ASE/IEEE Social Computing Conference. 08-SEP-13, . : ,

08/27/2013 30.00 Nathan O. Hodas, Farshad Kooti, Kristina Lerman. Friendship Paradox Redux: Your Friends Are More Interesting Than You,
ICWSM. 08-JUL-13, . : ,

08/27/2013 29.00 Siddharth Jain, Eduard Hovy. DETERMINING LEADERSHIP IN CONTENTIOUS DISCUSSIONS,
ICME International Workshop on Social Multimedia Research 2013. 15-JUL-13, . : ,

08/27/2013 28.00 Yi Chang, Xuanhui Wang, Qiaozhu Mei, Yan Liu. Towards Twitter context summarization with user influence models,
the sixth ACM international conference. 03-FEB-13, Rome, Italy. : ,

08/27/2013 27.00 Mohammad Taha Bahadori, Yan Liu. An Examination of Practical Granger Causality Inference,
SIAM Conference on Data Mining (SDM'13). 02-MAY-13, . : ,

08/27/2013 26.00 Mohammad Taha Bahadori, Yan Liu, Eric P. Xing. Fast structure learning in generalized stochastic processes with latent factors,
the 19th ACM SIGKDD international conference. 10-AUG-13, Chicago, Illinois, USA. : ,

08/27/2013 21.00 Jihie Kim, Jaebong Yoo, Ho Lim, Huida Qiu, Zornitsa Kozareva, Aram Galstyan. Sentiment Prediction using Collaborative Filtering,
ICWSM-2013. 08-JUL-13, . : ,

08/27/2013 20.00 Greg Ver Steeg, Aram Galstyan. Information-theoretic measures of influence based on content dynamics,
the sixth ACM international conference on Web Search and Data Mining. 03-FEB-13, Rome, Italy. : ,

08/28/2014 50.00 Dehua Cheng, Yan Liu. Parallel gibbs sampling for hierarchical dirichlet processes via gamma processes equivalence,
the 20th ACM SIGKDD international conference. 23-AUG-14, New York, New York, USA. : ,

08/28/2014 47.00 Linhong Zhu, Aram Galstyan, James Cheng, Kristina Lerman. Tripartite graph clustering for dynamic sentiment analysis on social media,
the 2014 ACM SIGMOD international conference. 21-JUN-14, Snowbird, Utah, USA. : ,

08/28/2014 49.00 Laura M. Smith, Linhong Zhu, Kristina Lerman, Zornitsa Kozareva. The Role of Social Media in the Discussion of Controversial Topics,
2013 International Conference on Social Computing (SocialCom). 07-SEP-13, Alexandria, VA, USA. : ,

08/28/2014 51.00 Dehua Cheng, Mohammad Taha Bahadori, Yan Liu. FBLG: a simple and effective approach for temporal dependence discovery from time series data,
the 20th ACM SIGKDD international conference. 23-AUG-14, New York, New York, USA. : ,

08/28/2014	52.00	J. He, Y. Liu, Q. Yang. Linking Heterogeneous Input Spaces with Pivots for Multi-Task Learning. , SIAM Conference on Data Mining (SDM'14). , . Philadelphia, PA: Society for Industrial and Applied Mathematics, Society for Industrial and Applied Mathematics

TOTAL: **46**

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

(d) Manuscripts

Received	Paper	
08/27/2012	14.00	Lerman, K., Ghosh, R., Surachawala, T. . Social Contagion: An Empirical Study of Information Spread on Digg and Twitter Follower Graphs, PLoS ONE (06 2012)
08/27/2012	18.00	Fellouris, G., Tartakovsky, A.G.. Almost Minimax Sequential Tests of Composite Hypotheses, Statistica Sinica (03 2012)

TOTAL: **2**

Number of Manuscripts:

Books

Received	Book

TOTAL:

Received	Book Chapter

TOTAL:

Patents Submitted

Patents Awarded

Awards

Graduate Students

NAME	PERCENT_SUPPORTED	Discipline
Gao, Shuyang	0.31	
Agharwal, Amav	0.01	
FTE Equivalent:	**0.32**	
Total Number:	**2**	

Names of Post Doctorates

NAME	PERCENT_SUPPORTED
Zhu, Linhong	0.02
Yan, Xiaoran	0.06
FTE Equivalent:	**0.08**
Total Number:	**2**

Names of Faculty Supported

NAME	PERCENT_SUPPORTED	National Academy Member
Galstyan, Aram	0.16	No
Ferrara, Emilio	0.40	
Ver Steeg, Greg	0.22	No
Lerman, Kristina	0.13	
Nevatia, Ram	0.80	No
Liu, Yan	0.10	
Allahverdyan, Armen	0.18	
Hovy, Eduard	0.10	No
FTE Equivalent:	**2.09**	
Total Number:	**8**	

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

Total Number:

Names of other research staff

NAME PERCENT_SUPPORTED

FTE Equivalent:

Total Number:

Sub Contractors (DD882)

1 a. Carnegie Mellon University

1 b. 406 Mellon University
5000 Forbes Avenue
Pittsburgh PA 15213

Sub Contractor Numbers (c): 167593
Patent Clause Number (d-1):
Patent Date (d-2):
Work Description (e): CMU was responsible for developing content reach models and algorithms for analyzing
Sub Contract Award Date (f-1): 1/1/13 12:00AM
Sub Contract Est Completion Date(f-2): 5/31/16 12:00AM

---

1 a. University of Connecticut - Storrs

1 b. Sponsored Program Services
438 Whitney Road Ext., Unit 1133
Storrs CT 062691133

Sub Contractor Numbers (c): 46485418
Patent Clause Number (d-1):
Patent Date (d-2):
Work Description (e): Developed statistical models
Sub Contract Award Date (f-1): 1/1/14 12:00AM
Sub Contract Est Completion Date(f-2): 12/31/14 12:00AM

---

1 a. University of Connecticut - Storrs

1 b. U-133, Room 117
438 Whitney Road Extension
Storrs CT 062691133

Sub Contractor Numbers (c): 46485418
Patent Clause Number (d-1):
Patent Date (d-2):
Work Description (e): Developed statistical models
Sub Contract Award Date (f-1): 1/1/14 12:00AM
Sub Contract Est Completion Date(f-2): 12/31/14 12:00AM

---

1 a. Universidad Carlos III de Madrid

1 b. Madrid Street, 126

GETAFE (MADRID) SPAIN 28903

Sub Contractor Numbers (c): 56489683
Patent Clause Number (d-1):
Patent Date (d-2):
Work Description (e): Investigated the role that network structure plays in individual's access to information in c
Sub Contract Award Date (f-1): 8/1/14 12:00AM
Sub Contract Est Completion Date(f-2): 8/31/14 12:00AM

Inventions (DD882)

Scientific Progress

See attachment

Technology Transfer

FINAL REPORT
Award # DARPA-W911NF-12-1-0034
Situational Awareness for Social Media: Theories, Models and Algorithms

August 31, 2016

AUTHORS

Aram Galstyan, Kristina Lerman, Eduard Hovy, Yan Liu, Ram Nevatia

LEADING PERFORMING ORGANIZATION

University of Southern California

Technical Point of Contact:
Aram Galstyan
Information Sciences Institute
University of Southern California
4676 Admiralty Way, Suite 1001
Marina del Rey, CA 90292
Tel: 310-448-9183
Fax: 310-822-0751
Email: galstyan@isi.edu

Adminstrative Point of Contact:
Brigidann Cooper
USC Department of Contracts & Grants - Marina Office
University of Southern California
4676 Admiralty Way, Suite 1001
Marina del Rey, CA 90292
Tel: 310-448-9161
Fax: 310-823-6714
Email: brigidannc@usc.edu

Contents

List of Figures

List of Tables

1. Statement of the Problem

Online social media has emerged as a critical platform for real-time information dissemination, search, marketing, and strategic communication. Like any other technology, social media can be used for both constructive and malicious purposes, albeit with unprecedented effectiveness. While some malicious use of social media can be counteracted through timely detection of rumors and orchestrated propaganda campaigns, manually monitoring the vast volume of social media communication is highly impractical. The goal of this project is to develop technological capabilities for monitoring social media platforms to detect propaganda campaigns and influence operations and predict their effect. Our work will lead to automated and semi–automated tools to monitor emergence of topics and memes, predict their spread, detect attempts to propagate rumors and misinformation, recognize participant activities and intent, and detect covert attempts to manipulate public opinion. Our proposed research is organized into three concurrent and inter–related research thrusts:

Research Thrust 1: Causal Models of Participants and their Interactions. The objective of this research thrust is to develop a framework for modeling and classifying participant behavior and inferring intent, uncovering and quantifying possible persuasion campaigns, and understanding causal-behavioral relationships between participants in social media. One of the central problems in social science is to understand how information diffuses through social channels and how different participants influence this process. Our objective is to go well beyond network-based metrics and develop novel methods for uncovering causal relationships and assessing influence based on time–resolved, dynamical analysis of participant activities. In the course of this project, we will develop information-theoretic methods for revealing causal relationships between participants based on observations of their activity and construct predictive models that link individual activity to collective network behavior. The intended outcome of the project will consists of models and algorithms for classifying participant behavior based on temporal signatures of their activities, and data-driven methods for learning such behaviors over a range of relevant time scales.

Research Thrust 2: Information Diffusion and Change Prediction. The objective of this research thrust is to develop models for understanding and quantifying information propagation in networks. This understanding will be crucial for developing accurate algorithms to detect buzz, monitor spread of ideas and memes, and predict how people will respond to them. In the couse of the project we will develop scalable algorithms for detecting small but sustained changes in activity patterns that will automatically flag important changes as quickly as possible within an acceptable threshold of false alarms. Another important goal of this thrust is to identify possible precursors of changes based on fine-grained analysis of collective user dynamics, to estimate transmissibility of topics and memes across social groups and use these estimates to predict how participants will respond to them.

Research Thrust 3: Content-Rich Models of Participants and Social Media. The last research thrust is concerned with leveraging communication content for better understanding of processes unfolding in social media. In particular, the objective is to produce fine-grained, content–based models of social media participants and their interactions, a deeper understanding of topics they discuss, and tools for monitoring meme diffusion throughout the network. In the course of the project, we will develop methods for inferring different types of participants and links, both at

individual and group level, by combining a wide array of textual analysis techniques with methods for analyzing patterns of their interpersonal interactions. We are also examining how link heterogeneity impacts information diffusion in the network, and aggregate these features to support models of information diffusion.

2. Summary of Addressed Tasks and Accomplishments

In the course of the SMISC proigram, we have made significant advances in each of three research areas outlined above. Below we briefly highlight the main accomplishments.

For ***Research Thrust 1***, we have developed a family of algorithms based on information-theoretic principles that allow to find predictive relationships between social media participants. For example, we have introduced *content transfer* as a novel information-theoretic measure with a predictive interpretation that directly quantifies the strength of the effect of one user's content on another's in a model-free way. We have demonstrated on Twitter data collected for thousands of users that content transfer is able to capture non-trivial, predictive relationships even for pairs of users not linked in the follower or mention graph. We have also developed a a number of algorithms that use Granger causality to infer directed influence links based on traces of users activities. We have also developed computational model of user behavior based on temporal signatures of their activities. In particular, we have developed a Coupled Hidden Markov Model where each user's activity is coupled to the aggregated activity profile of all the neighbors. We have shown that those models can be used for clustering the users according to their activity patterns. (Section **??**)

One of our most important contribution has been the development of unsupervised information-theoretic learning method called CorEx, for Correlation Explanation. Social media data is high dimensional, heterogeneous, and sparse, which makes it challenging to extract useful knowledge from such data. To address those challenges, CorEx learns regularities from high dimensional multivariate data by trying to explain correlations observed in the data. In contrast to most other learning methods, CorEx does not make any assumptions about how the data is generated. Our experiments with various datasets, including human-behavioral data, suggest that CorEx is indeed able to extract hidden patterns in the data that can be used for different predictive purposes.

Within ***Research Thrust 2***, we have studied how different factors affect information diffusion, or more general social contagion, in social media. Social contagion is the mechanism by which information spreads through the follower graph on these sites. We have examined how users on different social media platforms respond to an incoming stimuli, e.g., a tweet (message) from a friend on Twitter or a recommendation for a news story on Digg, using both empirical analysis and mathematical modeling. Our results indicate that psychosocial factors play a crucial role in the process of social contagion. For instance, visual salience, or visibility, of the stimulus plays an important role in determining whether it will acted on, e.g., retweeted. Visibility of the stimulus decays in time as a user's friends add new messages to his queue, increasing the cognitive effort required for discovering and acting upon the stimulus.

It is generally believed that information spreads between individuals like a pathogen, with each exposure by an informed friend potentially resulting in a naive individual becoming infected. However, empirical studies of social media suggest that individual response to repeated exposure to information is significantly more complex than the prediction of the pathogen model. As a

proxy for intervention experiments, we compared user responses to multiple exposures on two different social media sites. We showed that the position of the exposing messages on the user-interface strongly affects social contagion. Accounting for this visibility significantly simplifies the dynamics of social contagion. The likelihood an individual will spread information increases monotonically with exposure, while explicit feedback about how many friends have previously spread it increases the likelihood of a response. We showed that our model can accurately forecast user behavior on two social media sites.

We have also looked that the problem of buzz modeling and predictions, which are fundamental problems in computational social science. Those problems are extremely challenging since the distributions of these time series observations have heavy tails and most existing models fail miserably. To address those issues, we have developed a temporal point process model with generalized extreme value (GEV) distributions. Instead of modeling the temporal dependence directly, which is nonlinear and has no sufficient data to establish meaningful dependence, we resort to the location parameters of GEV distributions because they capture the mode of extreme value variables and can be modeled reasonably well by linear Gaussian models. We have tested Sparse-GEV on two different tasks: uncovering the underlying dependency among time series, and predicting the future values of time series, and demonstrated that it performs better than several other baselines.

For ***Research Thrust 3***, we have done extensive work on developing efficient models that augment social networks with rich content data. For instance, in a very recent paper published at WWW, we proposed a shared latent space model that uses the same generative model to describe both user interactions, and their individual attributes. This type of joint multi-modal approach results in more accurate predictive models. Furthermore, it enables use to do across-modality predictions, when, for instance, we use a participant's network information to make predictions about his/her behavior, or vice versa. As another example, we have developed the HawkesTopic model (HTM) for analyzing text-based cascades, such as retweeting a post or publishing a follow-up blog post. HTM combines Hawkes processes and topic modeling to simultaneously reason about the information diffusion pathways and the topics characterizing the observed textual information. We have validated our approach on both synthetic and real-world data sets, including a news media dataset for modeling information diffusion, and an ArXiv publication dataset for modeling scientific influence. Our results show that HTM is significantly more accurate than several baselines for both tasks.

In another subproject, we have studied the problem of identifying roles of sentiment in information propagation and generating a predictive model of twitter users using sentiment information. We have focused on a corpus of political (UK) tweets, and examined the impact of emotion expressions on the generated response. We extract message sentiment using Linguistic Inquiry and Word Count (LIWC), and then use logistic regression to find out which of these affective variables explain the number of replies or retweets generated from the message. In particular, we have compared differences between user reply vs. retweet behavior with respect to sentiment variables. Prior work suggests that that degree of emotion expressions in twitter messages can affect the number of replies generated as well as retweet rates. However, there are no comparable studies for response. We have found that, due to the difference in the nature of endorsement (retweet) vs. responses (replies or conversation), some of the variables present opposite roles in explaining the degree of responses the message receives.

A large part of our effort in this research thrust has been focused on developing methods for automatically identifying important characteristics of social media participants and groups based on the text they produce and share, especially in contentious discussions. In these discussions, the outcome often depends critically on the discussion leader(s). Recent work on automated leadership analysis has focused on collaborations where all the participants have the same goal. We have analyzed discussions on the Wikipedia Articles for Deletion (AfD) forum, and defined several complimentary models to quantify the basic leadership qualities of participants and assign leadership points to them. We have also investigated the problem of identifying participants in the discussion whose contribution can be considered sensible. We have proposed a model of sensibility that uses features based on participants contribution patterns and content, as well as on the discussion domain, to investigate their effect on the analysis.

One of the main objectives of the SMISC program is to develop capabilities for automatically detecting orchestrated campaigns and influence operations on social media platforms. To test the technologies developed so far, DARPA held a four-week competition in February/March 2015, where multiple teams competed to identify a set of *influence bots* within Twitter. The USC team successfully participated in the challenge, and obtained the only 100% accuracy results, by detecting all 39 influence bots while not making any false positives. It is important to note that there was no single method that could detect all the bots. Instead, we used a combination of techniques and methods developed in each of the three aforementioned research thrust.

Below we provide more technical details on certain methods that were not covered in our previous yearly reports.

3. Causal Models of Participants and Their Interactions

3.1. Emergence of Leadership in Communication

Identifying leaders can be important for applications such as viral marketing, accelerating (or blocking) the adoption of innovations, *etc*. Communication research postulates that informational influence in groups often happens via a two-step process, where information first flows from news media to opinion leaders, and then is spread further to followers. This theory is believed to adequately account for consumer behavior and has been refined in several ways.

During the last reporting period, we have developed formal, mathematical model for the emergence and the type of leadership in a collective of interacting agents modeled via neurons. Our model mimics a discussion process, where opinion expression by one agent facilitates activation of other agents. Specifically, we postulate tractable rules for the agent's behavior. The model incorporates major factors that are relevant for the leadership, e.g. activity, attention, initial social capital (i.e. well-connectedness in the network), and score (credibility). The rules depend on parameters that characterize the agent's "conservatism" with respect to changing its state.

The leader is naturally defined as an agent that influences other agents strongly (i.e. stronger than those agents influence the leader), and that actively participate in the group activity, in the sense that blocking the leader will diminish (or at least essentially decrease) the activity of the group.

Let us consider N agents. At a given moment of discrete time t, each agent can be active—give an opinion, ask question *etc*—or passive. For each agent i ($i = 1, ..., N$) we introduce a variable $m_i(t)$ that can assume two values 0 (passive) and 1 (active).

Following a tradition in quantitative sociology, we model agents via thresholds elements, i.e. we postulate that each agent i has an information potential $w_i(t) \geq 0$, and i activates whenever $w_i(t)$ is larger than a threshold u_i:

$$m_i(t) = \vartheta[\, w_i(t) - u_i \,], \;\; t = 0, 1, 2, \dots , \tag{1}$$

$$w_i(t+1) = (1 - m_i(t)\,) \sum_{j=1}^{N} q_{ij}(t) m_j(t), \tag{2}$$

where $\vartheta(x)$ is the step function: $\vartheta(x < 0) = 0$, $\vartheta(x \geq 0) = 1$. The factor $(1 - m_i(t))$ in (2) nullifies the potential after activation; hence an agent cannot be permanently active. The influence $q_{ij}(t)m_j(t)$ of j on i is non-zero provided that j activates, $m_j(t) = 1$. We assume that $q_{ij} \geq 0$ and $q_{ii} = 0$, i.e. connections can only facilitate the potential generation. Given the freedom in choosing q_{ij}, we take $u_i = 1$.

In (2), $q_{ij}(t)$ quantifies the influence of j on i. We parametrize it as

$$q_{ij}(t) = q\, \tau_{ij}(t), \quad \tau_{ii}(t) = 0, \tag{3}$$

$$\sum_{j=1}^{N} \tau_{ij}(t) = 1, \tag{4}$$

where q is the maximal possible value of $q_{ij}(t)$. Now τ_{ij} is the weight of influence. Eq. (4) reflects the fact that agents have limited attention. This characteristics was also noted for neurons, and it is achieved via introducing non-normalized weights $\widetilde{\tau}_{ij}(t)$:

$$\tau_{ij}(t) = \widetilde{\tau}_{ij}(t) \Big/ \sum_{j=1}^{N} \widetilde{\tau}_{ij}(t) \; . \tag{5}$$

Importantly, we do not pre-determine the network structure. Hence the sum $q\sum_{j=1}^{N} \tau_{ij}(t) m_j(t)$ in (2) is taken over all the agents, where

$$\widetilde{\tau}_{ij}(t+1) = \tau_{ij}(t) + f\, \tau_{ij}^{\alpha}(t)\, m_j(t+1), \qquad \alpha > 0, \tag{6}$$

where a non-active j ($m_j(t+1) = 0$) does not change τ_{ij}. In one version of the model f =const. Thus $\widetilde{\tau}_{ij}(t)$ changes such that more active and more credible agents get more attention from neighbors. In (6), $\tau_{ij}^{\alpha}(t)$ controls the extent to which i re-considers those links that did not attract its attention previously (conservatism): for $\alpha \not\approx 0$ the weight with $\tau_{ij}(t) \approx 0$ is not reconsidered in the next step. A similar structure was employed for modeling confirmation bias [1]. It also appears in a preferential selection model for network evolution.

Our analysis suggest that even for f =const, the above model yields non-trivial leadership scenario. Introduce additional variables results in even richer behaviors.

Scores. For each agent i we now introduce its *credibility score* [1]) $\sigma_i(t) \geq 0$, which is a definite feature of an agent at a given moment of time. Scores interact with m_i and τ_{ij} by modulating the function f in (6):

$$f = f[\sigma_j(t) - \beta\,\sigma_i(t)], \qquad \beta = 0, 1, \tag{7}$$

where we assume

$$f[x] = x \quad \text{for} \quad x > 0; \qquad f[x] = 0 \quad \text{for} \quad x \leq 0. \tag{8}$$

Thus for $\beta = 1$, the agent i reacts only on those with score higher than σ_i, whereas for $\beta = 0$ every agent j can influence i proportionally to its score σ_j. For convenience, we restricted $\beta = 0, 1$ to two values. (Note that if we define $f[x]$ to be a positive constant for $x \leq 0$, then the situation without scores can be described via $\beta \to \infty$.)

Dynamics of σ_i is determined by the number of agents that follow i and by the amount of attention those followers pay to the messages of i:

$$\sigma_i(t+1) = (1 - \xi_1)\sigma_i(t) + \xi_2 \sum_{k=1}^{N} m_i(t)\tau_{ki}(t), \tag{9}$$

where ξ_1 and ξ_2 are constants: $1 \geq \xi_1 > 0$ quantifies the score loss (forgetting), while the term with $\xi_2 > 0$ means that every time the agent i activates, its score increases proportional to the weight $\tau_{ki}(t)$ of its influence on k. If i is not active, $m_i(t) = 0$, its score decays.

Development of complex network theory motivated many models, where the links and nodes are coupled; see [9] for an extensive review. Neurophysiological motivation for studying such models comes from the synaptic plasticity of neuronal connections that can change on various time scales.

The ingredients in the evolution of the score (9) do resemble the notion of fitness, as introduced for models of competing animals. There the fitness determines the probability of winning a competition. Similar ideas were employed for modeling social diversity.

Initial conditions . We assume that initially all agents are equivalent:

$$\sigma_i(0) = 0. \tag{10}$$

The initial network structure is random:

$$\tau_{ij}(0) = n_{ij} \Big/ \sum_{k=1}^{N} n_{ik}\ , \quad n_{ij} \in [0, b], \quad n_{ii} = 0, \tag{11}$$

where n_{ij} are independent random variables homogeneously distributed over the interval $[0, b]$. Now

$$\phi_i \equiv \sum_{k=1}^{N} \tau_{ki}(0), \tag{12}$$

[1]Credibility refers to the judgments made by a message recipient concerning the believability of a communicator. A more general definition of credibility [not employed here] should account for its subjective aspect; a message source may be thought highly credible by one perceiver and not at all credible by another.

measures the initial cumulative influence of i (i.e. its initial social capital) and estimates the initial rate of score generation; cf. (9).

For $m_i(0)$ we impose initial conditions, where some agents are activated initially (by a news or discussion subject), i.e. $m_i(0)$ are independent random variables:

$$\Pr[m(0)=1]=\gamma, \quad \Pr[m(0)=0]=1-\gamma, \quad \gamma \le 1/2. \tag{13}$$

Thresholds of collective activity. The only way the initial activity can be sustained is if the agents stimulate each other (as it happens in a real discussion process). Our numerical results show that there are two thresholds $\mathcal{Q}^+$ and $\mathcal{Q}^-$, so that for $q \ge \mathcal{Q}^+$ the initial activity is sustained indefinitely,

$$\sum_{i=1}^{N} m_i(t) > 0, \quad \text{if} \quad q \ge \mathcal{Q}^+. \tag{14}$$

If $q \le \mathcal{Q}^-$ the initial activity decays in a finite time t_0 (normally few time-steps)

$$\sum_{i=1}^{N} m_i(t \ge t_0) = 0, \quad \text{if} \quad q \le \mathcal{Q}^-. \tag{15}$$

For $\mathcal{Q}^+ > q > \mathcal{Q}^-$ the activity sustaining depends on the realization of random initial conditions $m_i(0)$ and $\tau_{ij}(0)$: either it is sustained indefinitely or it decays after few steps. Qualitatively, the activity is sustained if sufficiently well-connected agents are among the initially activated ones; see below. $\mathcal{Q}^+$ and $\mathcal{Q}^-$ depend on all the involved parameters, their numerical estimates are given below in (17). Note that $\mathcal{Q}^- > 1$, since for $q < 1$ no activity spreading is possible: even the maximal weight $\tau_{ik} = 1$ cannot activate i; see (2, 3).

Our experiments suggest that the emergent network structures depend mainly on parameters α and β. Other parameters can be important for supporting activity, i.e. $\mathcal{Q}^+$ and $\mathcal{Q}^-$ depend on them, but are not crucial for determining the type of the emerging network. So we fix for convenience [cf. (13, 11, 9)]:

$$N=100, \;\; \gamma=0.5, \;\; b=10, \;\; \xi_1=0.1, \;\; \xi_2=1. \tag{16}$$

For these parameters and (for example) $\alpha=\beta=1$, we got numerically

$$\mathcal{Q}^- = 2.19, \qquad \mathcal{Q}^+ = 2.65. \tag{17}$$

Behavioral noise. We note that the deterministic firing rule (1) can be modified to account for agents with behavioral noise.The noise is implemented by assuming that the threshold $u_i + v_i(t)$ has (besides the deterministic component u_i) a random component $v_i(t)$. These quantities are independently distributed over t and i. Now $v_i(t)$ is a trichotomic random variable, which takes values $v_i(t) = \pm V$ with probabilities $\frac{\eta}{2}$ each, and $v_i(t) = 0$ (no noise) with probability $1-\eta$. Hence η describes the magnitude of the noise. We assume that V is a large number, so that with probability η, the agent ignores w_i and activates (or does not activate) randomly. Thus, instead of (1), we now have

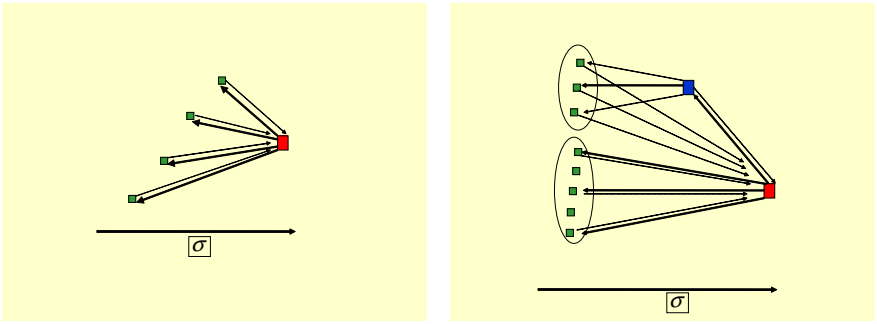
$$m_i(t) = \vartheta[\,\phi_i(t)(\,w_i(t) - u_i\,)\,], \tag{18}$$

$$\Pr[\phi_i(t)=1]=1-\eta, \qquad \Pr[\phi_i(t)=-1]=\eta, \tag{19}$$

where $\phi_i(t)$ are independent (over i and over t) random variables that equal ± 1 with probabilities $\Pr[\phi_i = \pm 1]$.

Qualitatively the same predictions are obtained under a more traditional model of noise, where the step function in (1) is replaced by a sigmoid function.

Extensive simulations with the above model demonstrate that the three basic leadership scenarios (laissez-faire, participative and autocratic) emerge under different behavioral rules. In laissez-faire leadership is characterized by a relatively weak guidance of autonomous followers; in participative (democratic) systems leaders do influence their followers strongly, but encourage and accept feedback from them; and finally, no feedback is accepted by autocratic leaders. For instance, Fig. 1 depicts and example of a structure describing participative leadership.



(a) (b)

Figure 1: Participative leadership (a) *Single Participative leader*: $\alpha = \beta = 1$ (b) *Hierarchic leadership*: $\alpha = 2$, $\beta = 1$. The followers are divided into two groups, strongly driven by respectively leader (red) and sub-leader (blue). The feedback is collected by the leader only.

3.2. Granger causality analysis of participation interactions

A major challenge in social influence analysis is the incompleteness of the data, i.e., certain data are missing in the real-world datasets we can collect. For example, in social media analysis, the external events influence large clusters of users, while the news propagates through the local connections in the network. In order to identify the true influence patterns among the users, we need to take into consideration the impact of external unobserved events. Therefore understanding and quantifying the impact of unobserved processes is one of the major challenges in social influence analysis.

In this project, we have made progresses in three aspects to address the problem: (1) theoretical analysis of Granger causality in presence of latent variables; (2) advanced stochastic models for uncover social influence networks in presence of latent variables; (3) further improvement of Granger causality analysis. Next, we discuss our research outcomes in details:

Theoretical Analysis of Granger Causality in Presence of Latent Variables [SDM 2013]. There are two major challenges in discovering temporal causal relationship in large-scale data: (i) not all the influential confounders are observed in the datasets and (ii) enormous number of high dimensional time series need to be analyzed. The first challenge stems from the fact that in most datasets

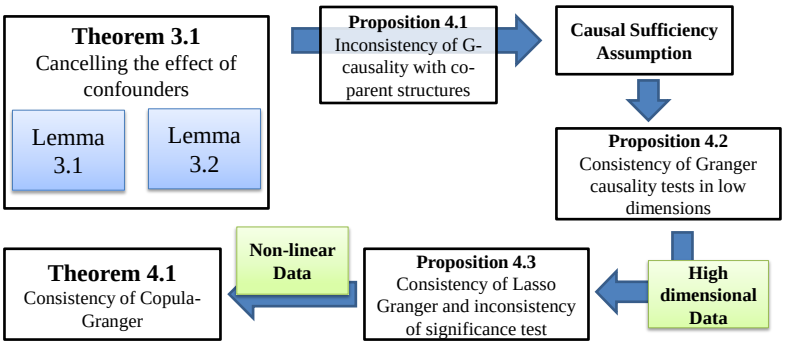


Figure 2: The sequence of the theoretical results: Theorem 3.1 utilizes path delays to find a subset of time series that via conditioning on them we are able to cancel out the spurious confounder effects. Proposition 4.1 shows that when unobserved time series are parents of multiple observed time series, there is no consistent Granger causality test. The *Causal Sufficiency* assumption excludes these structures and with this assumption both significant test and Lasso-Granger become consistent in low dimensions (Proposition 4.2). Proposition 4.3 shows that in high dimensions significant test is inconsistent but Lasso-Granger is consistent. When the data deviates from the linear model assumed in Lasso-Granger, Theorem 4.1 shows that while *Copula-Granger* is consistent in high dimensions, it can efficiently capture non-linearity in the data.

not all confounders are measured. Some confounders cannot even be measured easily which makes the spurious effects of unobserved confounders inevitable. The question in these situations is how can we utilize the prior knowledge about the unmeasured confounders to take into account their impact. The second challenge requires us to design scalable discovery algorithms that are able to uncover the temporal dependency among millions of time series with short observations.

In this project, we addressed both challenges and the key theoretical results are shown in Figure 2 (see [2] for details). By analysis of the effects of unobserved confounders in simple structures, we derive a new set of criteria which utilizes the aggregate delay in the confounding paths. The new criteria requires smaller subset of time series to be observed; hence it is more likely to achieve theoretical guarantee for Granger causality results. We show that under *causal sufficiency* assumption which excludes one particular structures, the two main Granger causality inference techniques, *Significance Test* and *Lasso-Granger*, are consistent. However, we observe that in higher dimensions only Lasso-Granger is consistent. Utilizing the high dimensional advantages of L_1 regularization, we design a semi-parametric Granger causality inference algorithm called *Copula-Granger* and show that while it is consistent in high dimensions, it can efficiently capture non-linearity in the data.

Stochastic Models for Uncover Social Influence Networks in Presence of Latent Variables [KDD 2013]. In order to effectively uncover the social influence networks in presence of latent variables, we developed a flexible stochastic process model, the *generalized linear auto-regressive process* (GLARP) and identify the conditions under which the impact of hidden variables appears as an additive term to the evolution matrix estimated with the maximum likelihood. In addition, we also developed a fast greedy algorithm based on the selection of *composite atoms* in each iteration and provide a performance guarantee for it (see [3] for details).

Consider the following model for generalized linear autoregressive processes with hidden factors:

$$g\left(\mathbb{E}_{\mathcal{H}(t)}\begin{bmatrix}\mathbf{x}(t)\\ \mathbf{z}(t)\end{bmatrix}\right) = \sum_{\ell=1}^{K}\begin{bmatrix}A^{(\ell)} & B^{(\ell)}\\ C^{(\ell)} & D^{(\ell)}\end{bmatrix}\begin{bmatrix}\mathbf{x}(t-\ell)\\ \mathbf{z}(t-\ell)\end{bmatrix} + \mathbf{b} \tag{20}$$

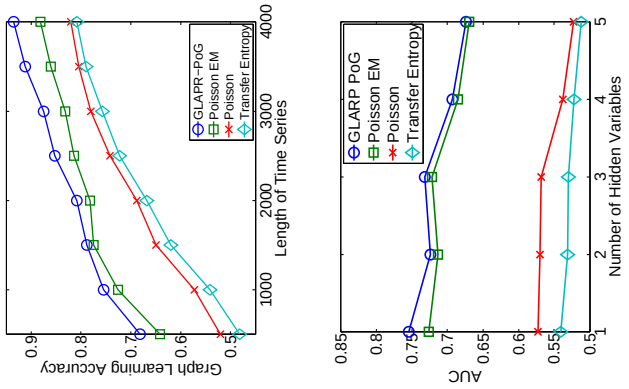
for $t = K+1, \ldots, T$, where $\mathbf{x}(t)$, a $p \times 1$ vector, represents the observed variables, $\mathbf{z}(t)$, a $r \times 1$ vector, denotes the unobserved variables and the function g is the link function. The density function of the observations at time t is denoted by $f(\mathbf{x}(t), \boldsymbol{\theta}(t))$ where $\boldsymbol{\theta}(t)$ denotes the set of parameters of the distribution that are functions of the evolution matrices $A^{(\ell)}, B^{(\ell)}, C^{(\ell)}$ and $D^{(\ell)}$ and the past values of time series $\mathbf{x}(t)$ and $\mathbf{z}(t)$. We show that the maximum likelihood estimate of their evolution matrices can be decomposed into a sparse and a low-rank matrix with the latter capturing the impact of unobserved processes as follows:

$$\begin{aligned}&\min_{A^{(\ell)},L^{(\ell)},\mathbf{b}} \mathcal{L}\left(\mathbf{x}(t), A^{(\ell)}, L^{(\ell)}\right)_{t=1:T}^{\ell=1:K} \\ \text{Subject to:}\quad &\sum_{\ell=1}^{K}\left\|A^{(\ell)}\right\|_0 \le \eta_S, \quad \sum_{\ell=1}^{K}\text{rank}\left(L^{(\ell)}\right) \le \eta_L,\end{aligned} \tag{21}$$

where the L_0 norm of the matrices is equal to the number of non-zeros elements of the matrices and $\mathcal{L}$ denotes the likelihood of the stochastic process defined in Eq(20). As an example, we analyze the Poisson and Conway-Maxwell Poisson auto-regressive processes for count data, and heavy-tailed auto-regressive time series model based on Gumbel distribution for extreme value time data.

We tested our proposed model on both synthetic datasets and real application datasets. We generate a synthetic dataset according to the Poisson autoregressive point process model in Eq. (20). We fix the number of observed variables at 60 and vary the number of hidden variables from $r = 1$ to 5. We also varied the length of observed time series to study the asymptotic behavior of the algorithms. The results are shown in Fig. 3. As we expect, the performance of the algorithms uniformly increases with the length of the time series. The algorithms that capture the impact of hidden variables outperform the other algorithms by a large margin and our proposed algorithms achieve the best performance.

We used a *complete* Twitter dataset to analyze the tweets about "Haiti earthquake" by applying different temporal dependency analysis methods to identify the potential top influencer on this topic (i.e. those Twitter accounts with the highest number of effect to the others). We divided the 17 days after the Haiti Earthquake on Jan. 12, 2010 into 1000 intervals and generated a multivariate time series dataset by counting the number of tweets on this topic for the top 1000 users who tweeted most about it. As shown in Figure 4, the performance of all algorithms increase as we increase the number of retweets requirement n for the ground truth influence graph. As we expected the GLARP-COMG algorithm outperforms the EM counterpart by avoiding the local minima. The prediction performance in Table 1 confirms this trend as well.



(a) (b)

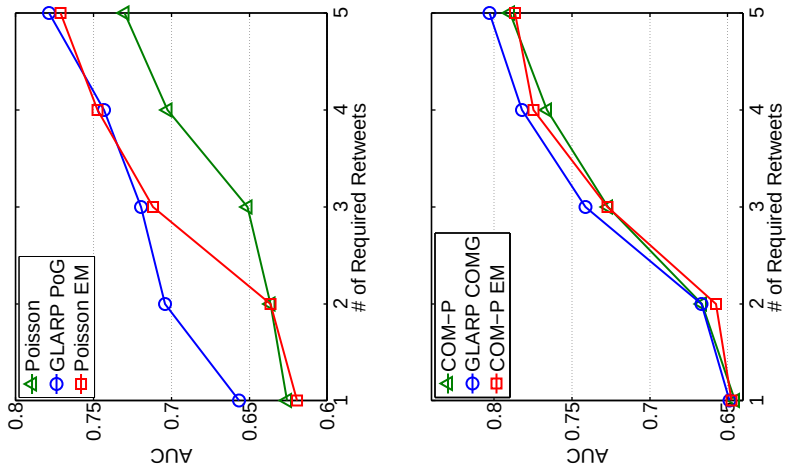
Figure 3: Synthetic dataset results on the point process dataset (a) Graph learning accuracy as the length of the time series increases. (b) Graph learning accuracy as the number of hidden variables increases.

Table 1: The RMS prediction error of the algorithms in the Twitter dataset. Results have been normalized by the the mean.

Method	RMSE	Norm-RMSE
GLARP-COMG	0.0059	0.3014
COM-P EM	0.0113	0.5739
COM-P	0.0096	0.4876
GLARP-PoG	0.0017	0.0887
Poisson EM	0.0062	0.3148
Poisson	0.0017	0.0847
Transfer Entropy	0.0030	0.1519

Forward Backward Granger Causality for High-dimensional Time Series [KDD 2014]. Penalized regression techniques (e.g. lasso or lasso-type regressions) have achieved major improvement for discovering *sparse* temporal dependence structures from multivariate time series data. However, the overall performance of existing Granger causality techniques still leaves room for improvement. In this project, we explore a new direction by considering the procedure of reversing the time in time series data. The inspiration for our work comes from classical mechanics where it is well-known that the basic equations of the classical physics remains valid when we look in reversed order of time, i.e., replacing time stamp t with $-t$. In [6], we propose a novel but simple approach, namely forward backward (FB) Granger causality, to infer the temporal dependence structures from the original time series and the time-reversed time series by simple averaging. We provide both theoretical analysis and empirical studies on the effectiveness of the proposed approach.

We define the forward time series as the original multivariate time series $\{\mathbf{y}^{(t)}\}, t = \ldots, 0, 1, \ldots,$ and the backward time series $\{\mathbf{z}^{(t)}\}$ as $\mathbf{z}^{(t)} := \mathbf{y}^{(-t)}$. $\{\mathbf{y}^{(t)}\}$ and $\{\mathbf{z}^{(t)}\}$ both contain N time series; univariate time series are denoted by $\{y_i^{(t)}\}$ and $\{z_i^{(t)}\}$ for $i = 1, \ldots, N$. Suppose that the assumptions for correctness of Granger causality are satisfied such that the coefficients estimated by Granger causality indicate the existence of temporal dependence relationships, a simple way



(a) (b)

Figure 4: The graph learning accuracy when the number of retweets requirement n for the ground truth influence graph $G_{RT}(n)$ is varied. The performance of (a) Poisson and (b) COM-Poisson autoregressive processes confirms that they make better predictions for the stronger influence edges.

Algorithm 1 Naive Forward Backward Lasso Granger Causality

Input: Time series $\{\mathbf{y}^{(t)}\}$, lag P, penalty parameter λ.
Output: Coefficients $\mathbf{A}_\ell^{FB}$, $\ell = 1, 2, \ldots, P$.
Define the backward time series $\{\mathbf{z}^{(t)}\}$ by $\mathbf{z}^{(t)} = \mathbf{y}^{(-t)}$.
Get forward coefficients $\mathbf{A}_\ell$ via Lasso-Granger with $\{\mathbf{y}^{(t)}\}$, P, and λ.
Get backward coefficients $\mathbf{B}_\ell$ via Lasso-Granger with $\{\mathbf{z}^{(t)}\}$, P, and λ.
Return $\mathbf{A}_\ell^{FB} = \frac{1}{2}(\mathbf{A}_\ell + \mathbf{B}_\ell^\top)$, $\ell = 1, 2, \ldots, P$.

to take advantage of the reverse time series is to add the coefficients for (i, j, k) in the forward time series and (j, i, k) in the backward time series. The proposed *Naive Forward Backward Lasso Granger Causality Algorithm* is shown in Algorithm 1.

We also developed the copula extension for nonlinear time series. Given a time series $\{\mathbf{x}^{(t)}\}$, we can map the data using the empirical marginal distribution of time series to the Gaussian distribution by

$$y_i^{(t)} = s_i \Phi^{-1}(\hat{F}(x_i^{(t)})), \quad \text{for } i = 1, \ldots, N, \tag{22}$$

where $\hat{F}$ is the empirical *cumulative distribution function* (CDF) of the ith time series, Φ is the CDF for standard Gaussian distribution and s_i is the standard derivative of the ith time series, which helps to retain original information. $\{\mathbf{y}^{(t)}\}$ will be treated as a linear representation for the original time series, to which we can apply Granger causality based algorithms. The Naive Forward Backward Copula Lasso Granger Causality as described in algorithm 2.

In addition, we investigated several well-known time series models and establish the connection between forward and backward time series in terms of temporal dependence structures. We have proved that for VAR, the strength $\{\mathbf{A}_k\}_{ij}$ of triplet (i, j, k) in forward time series is approximately the same as that for $\{\mathbf{B}_k\}_{ji}$ of triplet (j, i, k) in backward time series. For binary time series

Algorithm 2 Naive Forward Backward Copula Lasso Granger Causality

Input: Time series $\{\mathbf{y}^{(t)}\}$, lag P, penalty parameter λ.
Output: Coefficients $\mathbf{A}_\ell^{FB}$, $\ell = 1, 2, \ldots, P$.
for each $i = 1, 2, \ldots, N$ **do**
  Transform $y_i^{(t)} \rightarrow w_i^{(t)}$ by equation 22.
end for
Get coefficients $\mathbf{A}_\ell^{FB}$ by calling algorithm 1 with $\{\mathbf{w}^{(t)}\}$, P, and λ.
Return $\mathbf{A}_\ell^{FB}$, $\ell = 1, 2, \ldots, P$.

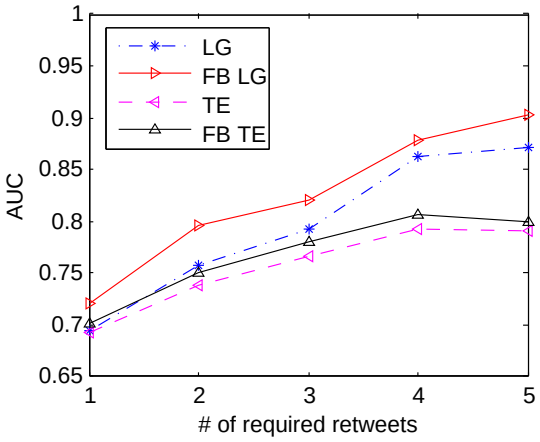


Figure 5: The performance of temporal dependence recovery on Haiti dataset. The performance of FB LG and FB TE outperform the LG and TE, respectively.

models that satisfy certain mild conditions, the strength of triplets (i, j, k) and (j, i, k) for forward and backward time series, respectively, share the same order of magnitude. These connections justify our approach to combine the results from the forward and the backward time series for better temporal dependence inference.

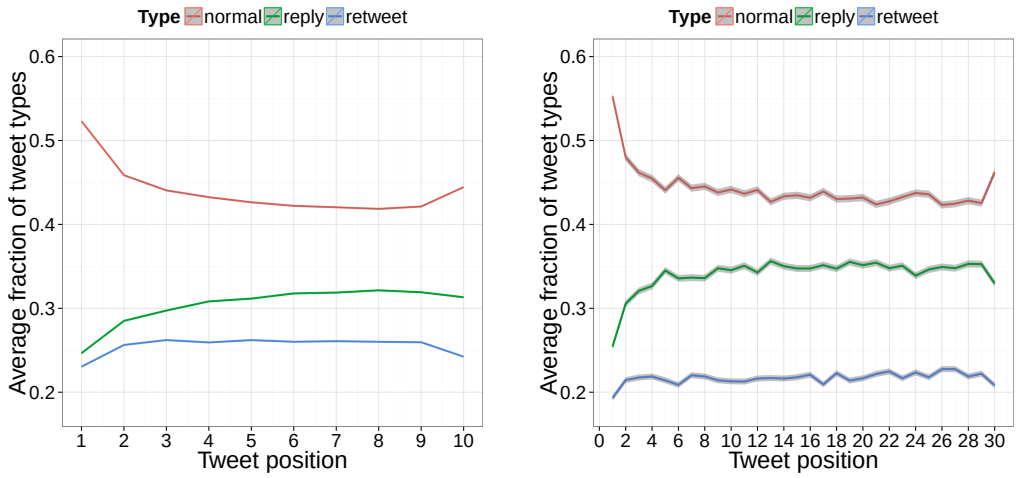
We conduct experiments on both the synthetic datasets and two real application datasets. On the synthetic datasets, the forward and backward approach can improve the performance of Lasso-Granger (LG), especially for the high-dimensional case (detailed results see full paper). We test LG, FB LG, transfer entropy (TE), and FB TE on the Twitter Haiti dataset (see Figure 5 for results). The AUC is calculated against the retweet graph $\mathcal{G}_{RT}$, and we vary the required number of retweets, so that only if the retweets from j by i passes the required number n, we establish an edge from i to j. Intuitively, n screens the weak influence between users. As we can see, all algorithms perform better as we increase n. In addition, the forward backward approach improves the performance for both baseline algorithms.

4. Information Diffusion and Change Prediction

4.1. Cognitive depletion

Human behavior shows strong daily, weekly, and monthly patterns. In this project we have also demonstrated changes in online behavior that occur on a much smaller time scale: minutes, rather

than days or weeks, which we attribute to cognitive effects, such as mental fatigue or boredom. Specifically, we studied how people distribute their effort over different tasks during periods of activity on the Twitter and Reddit social platforms. We demonstrate that as an online session progresses, people prefer to perform simpler tasks, such as replying and retweeting others' posts, rather than composing original messages on Twitter, or posting shorter comments on Reddit.



Sessions with 10 tweets

Sessions with 30 tweets

Figure 6: Change in the fraction of tweets of each type over the course of sessions in which users posted 10 or 30 tweets.

We studied a large dataset of tweets posted by over 100 thousand users on Twitter. First, constructed activity sessions from the time series of user's tweets. To do this, we examined the time interval between successive tweets and considered a break between sessions to be a time interval greater than some threshold a 10 minute threshold. Tweets posted by a user within 10 minutes of his or her previous tweet are considered to belong to the same session. The exact value of the threshold is not important and does not qualitatively affect our findings.

We studied the types of tweets users post at different times during a session. Since user behavior during longer sessions could be systematically different from their behavior during shorter sessions, we aggregate sessions by their length, i.e., the number of tweets posted. Then for each tweet position within a session, we calculate the fraction of tweets that belong to each of our three tweet types: *normal* tweets the user composes, *replies* to the messages of others, and *retweets* of the messages of others. Figure 6 shows that tweets are more likely to be normal tweets early in a session, and later in a session, users prefer cognitively easier (i.e., retweet) or socially more rewarding (i.e., reply) interactions.

We measure the strength of this effect empirically and find that the first post of a session is up to 25% more likely to be a composed message, and 10–20% less likely to be a reply or retweet. Qualitatively, our results hold for different populations of Twitter users segmented by how active and well-connected they are. Although our work does not resolve the mechanisms responsible for these behavioral changes, our results offer insights for improving user experience and engagement on online social platforms.

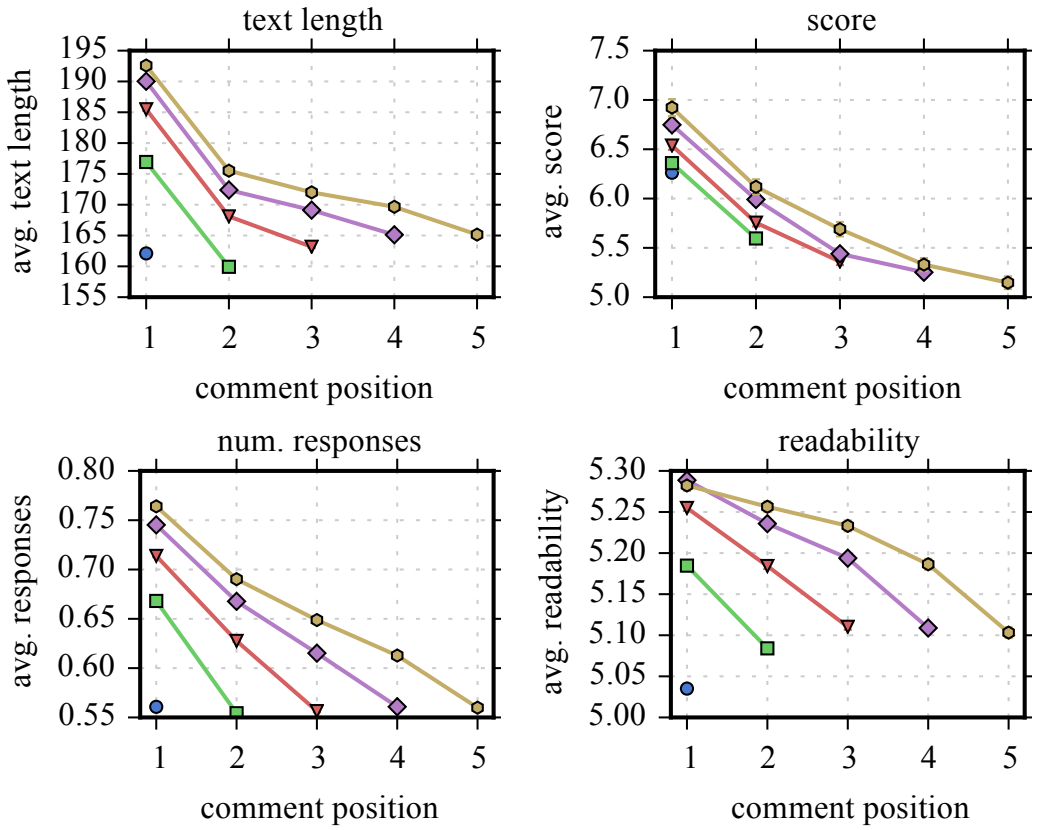


Figure 7: Average of all four proxies of the quality of comments posted on Reddit at different positions in a session. The colors (different markers) indicate different session lengths (number of comments written in a session, 1 up to a length of 5). The x-axis depicts a comment's index within the session, and the y-axis gives the average feature value (with error bars). The results indicate that earlier comments in a session tend to be of higher quality than later ones. Additionally, there appears to be a relation between the session length and the performance of the first comment in a session (stacking of lines). Overall, these empirical insights suggest performance deterioration over the course of sessions.

We studied a dataset around 40 million comments posted by people on the social networking site Reddit in April 2015. We sessionized user behavior by periods of commenting activity without breaks longer than 60 minutes

For measuring performance, we looked at four proxies of comment quality: text length, readability, the score a comment receives from others, and the number of responses it triggers. Figure 7 shows the changes in performance over the course of sessions. Different colors and markers distinguish sessions of distinct length (i.e., number of comments written during the session) of up to a length of 5. The x-axis shows the session index of a comment, the y-axis shows the (population-wide) average of respective feature (with error bars). For example, in the first plot of Figure 7, the red triangle at $x = 2$ refers to the average text length of all comments written in second position of all sessions of length 3.

We measure the strength of the declines in comment quality using mixed-effects models, which

control for individual variability. Even after accounting for variations among users, each comment posted during a session is 5%–11% higher quality than the next comment, depending on the measure of quality. Although our work did not resolve the mechanisms responsible for these performance changes, our results offer insights for creating algorithms to predict user behavior and improve user experience and engagement on online social platforms.

4.2. Cycle Modeling of Social Media Information Diffusion and Predictions

One important challenge in pursuit of social media information diffusion and predictions is how to find distinct temporal patterns within a set of buzz time-series. For example, as seen in Figure 8(a), the *Euro 2012* event attracted more and more frequent mentions when it came close to its final match day, a sudden increase of frequency happened when the final match started, and the large volume was maintained for a couple of hours during the final match, after which the mention volume dropped dramatically in just a few hours on social media. In contrast, for the sudden breaking news of a gun shooting at *Aurora Colorado* in 2012, as seen in Figure 8(b), we observe that a huge amount of frequent mentions appeared all of a sudden, and the public's attention dropped slowly in the next 24 hours. Intuitively, these two events have quite distinct temporal patterns. Therefore capturing the temporal patterns and modeling the inherent structures of these temporal patterns can provide us useful insights into information propagation and prediction in social media. Within this project, we have developed a hierarchical graphical model approach based on life cycle modeling as effective and scalable solutions.

Product life cycle (PLC) models [11] was originally introduced by economists to model the life span of a product: introduction to the market (initial sales), growth (sudden increase of sales), equilibrium (maturity phase defined by approximately constant sales) and decline (when the sales decrease dramatically). Usually we can observe 4 different stages, which are segmented with dotted lines. In general, for different types of buzz events, the growth and equilibrium stages might be relatively short, while the decline stage varies. Moreover, a buzz sequence often has multiple peaks, and the number of peaks is unknown in advance.

To handle a time-series sequence with multiple peaks, we first propose a parametric approach, namely K-Mixture of Product Life Cycle (*K-MPLC*), to automatically group buzz time-series based on the PLC mixture parameters [4]. An efficient $L1$-regularized approach to achieve a sparse solution for the PLC mixture models. Second, we propose a non-parametric approach to automatically infer the number of mixtures and peaks in K-MPLC models and demonstrate the effectiveness of the proposed approach via timeline summarization.

K-Mixture of Product Life Cycle for Diffusion Modeling and Prediction [ICDM 2014]. Given a topic, we count its mentions on social media during a pre-determined time interval (e.g., an hour), and generate a time-series sequence of this topic over a number of intervals during an observation window. Since a buzz sequence may consist of several obvious peaks, it could be modeled with multiple PLC models. As the number of peaks is not known in advance, we model a buzz sequence

with a mixture of PLCs as:

$$f(t;\boldsymbol{w},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\mu}) = \sum_{\ell=1}^{L} w_\ell f(t;\alpha_\ell,\beta_\ell,\mu_\ell), \tag{23}$$

$$f(t;\alpha,\beta,\mu) = \begin{cases} {Z_\ell}^{-1}(t-\mu)^{\alpha-1}e^{-\beta(t-\mu)} & (t \geq \mu) \\ 0 & (\text{Others}) \end{cases},$$

where L is the number of PLC models, $\boldsymbol{w} = [w_1, \ldots, w_L]^\top$ denotes the weight vector, $^\top$ denotes the matrix transpose, $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_L]^\top$ and $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_L]^\top$ are the vectors of Gamma distribution parameters, $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_L]^\top$ refers to locations of PLC models, and Z_ℓ is the normalization factor.

To discover the underlying similar patterns of buzz time-series, we propose a probabilistic graphical model, K-Mixture of Product Life Cycle (K-MPLC), to cluster N time-series into K groups. Throughout this paper, we assume K is known. Suppose that we are given N buzz time-series and their corresponding parameters $\boldsymbol{W} = [\boldsymbol{w}_1, \ldots, \boldsymbol{w}_N] \in \mathbb{R}^{L\times N}$, $\boldsymbol{\tau} \in \mathbb{R}^N$, and $\boldsymbol{\sigma} \in \mathbb{R}^N$. Given distinct characteristics over different parameter vectors, we use a Dirichlet distribution to model the weight vector $\boldsymbol{w}$ as its sum needs to be 1. Since τ and σ take positive values, we use the Gamma distribution to model them separately. Specifically, the probability for each instance is:

$$p(\boldsymbol{w},\tau,\sigma|\boldsymbol{\pi},\boldsymbol{\Theta},\boldsymbol{a},\boldsymbol{b},\boldsymbol{a'},\boldsymbol{b'}) = \sum_{k=1}^{K} \pi_k p(\boldsymbol{w}|\boldsymbol{\theta}_k) p(\tau|a_k,b_k) p(\sigma|a'_k,b'_k).$$

Here, $\boldsymbol{\pi} = [\pi_1, \ldots, \pi_K]^\top$ are the mixture weights, and

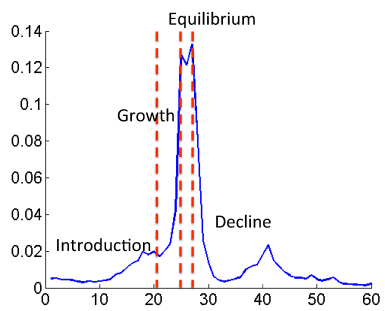
$$p(\boldsymbol{w}|\boldsymbol{\theta}_k) = C(\boldsymbol{\theta}_k) \prod_{i=1}^{L} w_i^{\theta_{ki}-1},$$

$$p(\tau|a_k,b_k) = \frac{b_k^{a_k}}{\Gamma(a_k)} \tau^{a_k-1} e^{-b_k\tau},$$

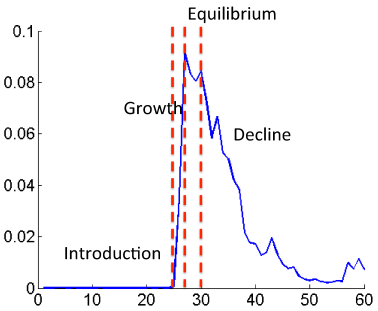
$$p(\sigma|a'_k,b'_k) = \frac{{b'_k}^{a'_k}}{\Gamma(a'_k)} \sigma^{a'_k-1} e^{-b'_k\sigma}$$

are Dirichlet and Gamma distributions, $C(\boldsymbol{\theta}) = \frac{\Gamma(\sum_{i=1}^{L}\theta_i)}{\Gamma(\theta_1)\Gamma(\theta_2)\ldots\Gamma(\theta_L)}$, and $\Gamma(a) = \int_0^\infty t^{a-1}e^{-t}\mathrm{d}t$.

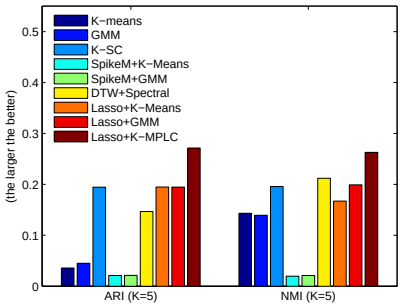
To evaluate the effectiveness of the K-MPLC algorithm, we build two benchmark datasets. In this experiment, we select thousands of high frequency search queries as candidate buzz topics and collect mentions from social media sites from June 22nd to August 8th, 2012. Considering the number of mentions of a topic per hour, we generated a time-series sequence for the topic within a time window. If the number of mentions at time t in a topic is 10 times higher than the average mention numbers in the past 48 hours, we regard the topic at that time as a buzz topic. Figure 9 represents the clustering results of different algorithms on the general buzz dataset. In this experiment, each time-series sequence is modeled by a mixture of 3 PLC models. It clearly shows that *Lasso+K-MPLC* performs the best among all algorithms: its ARI and NMI scores are 0.2714 and 0.2628, which is much better than others. K-SC performs the second best among all algorithms.



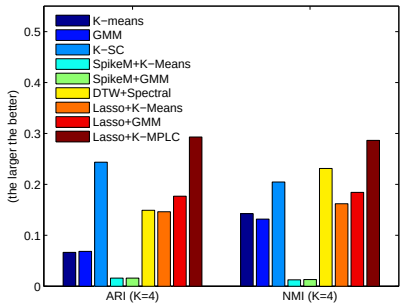
(a) Time-series of *Euro 2012*.



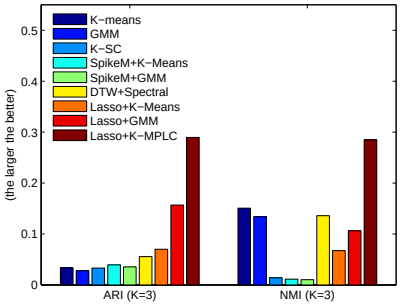
(b) Time-series of *Aurora Colorado*.



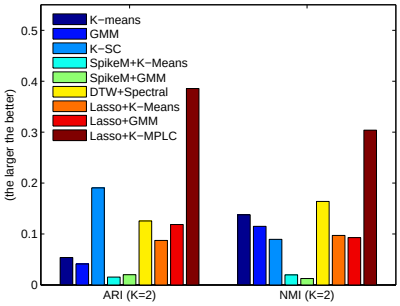
(a) Clustering results on 5-label ground truth.



(b) Clustering results on 4-label ground truth.



(c) Clustering results on 3-label ground truth.



(d) Clustering results on 2-label ground truth.

Figure 9: Clustering results on General Buzz Dataset with different Labels.

Nonparametric Models for Diffusion Modeling and Prediction [IJCAI 2016]. One of the major challenges in applying K-MPLC to practical applications is lack of guidance in setting the number of peaks in the PLC model. Therefore, in [5], we propose a nonparametric generative model where social media posts with temporal information are observations and the number of peaks are latent variables to be detected.

Similar as K-MPLC, we use T to denote temporal information for the entity with D posts and adopt a Gamma distribution with parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ to model temporal information, which can capture life cycles with sudden-spike-and-heavy-tail patterns as:

$$p(\boldsymbol{t}|\alpha_k, \beta_k) = \frac{\beta_k^{\alpha_k}}{\Gamma(\alpha_k)} t^{\alpha_k - 1} e^{-\beta_k t}, \;\; \Gamma(a) = \int_0^\infty t^{a-1} e^{-t} \mathrm{d}t.$$

In order to automatically determining the number of peaks in PLC Z, we adopt a Bayesian nonparametric method, where we assume that Z follows Chinese Restaurant Process with the parameter τ. To make the model fully Bayesian, we place the conjugate priors on the model parameters $\boldsymbol{\theta}$, $\boldsymbol{\theta}'$ and $\boldsymbol{\alpha}, \boldsymbol{\beta}$ respectively as: 1.) multinomial distribution $\boldsymbol{\theta}$ has the Dirichlet prior $\mathrm{Dir}(\boldsymbol{\eta})$; 2.) multinomial distribution $\boldsymbol{\theta}'$ has the Dirichlet prior $\mathrm{Dir}(\boldsymbol{\eta}')$ and 3.) $p(\alpha, \beta|\hat{p}, \hat{q}, \hat{r}, \hat{s})$ is the conjugate prior of Gamma distribution: $p(\alpha, \beta|\hat{p}, \hat{q}, \hat{r}, \hat{s}) \propto \frac{\hat{p}^{\alpha-1}e^{-\beta\hat{q}}}{\Gamma(\alpha)^{\hat{r}}\beta^{-\alpha\hat{s}}}$. The generative process is described as follows.

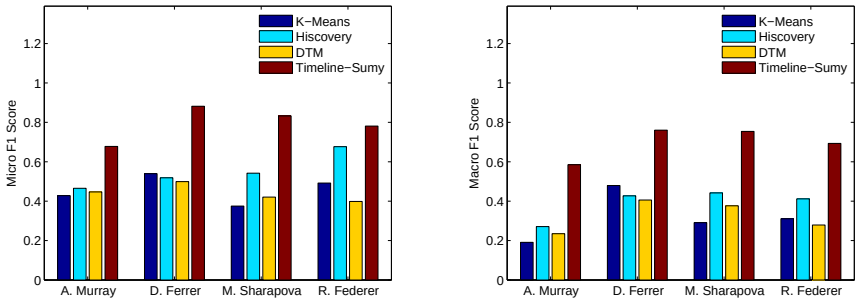
for each post j, **do**
  Draw $z_j \sim \mathrm{CRP}(\tau)$
  if z_j is a new episode **then**
    draw $\theta_{z_j} \sim \mathrm{Dir}(\boldsymbol{\eta})$ and $\theta'_{z_j} \sim \mathrm{Dir}(\boldsymbol{\eta}')$
    draw $\alpha_{z_j}, \beta_{z_j} \sim p(\alpha, \beta|\hat{p}, \hat{q}, \hat{r}, \hat{s})$
  end if
  Draw $C_j \sim \mathrm{Multinomial}(\theta_{z_j})$
  Draw $L_j \sim \mathrm{Multinomial}(\theta'_{z_j})$
  Draw $t_j \sim \mathrm{Gamma}(\alpha_{z_j}, \beta_{z_j})$
end for

To demonstrate the effectiveness of the proposed framework, we apply our proposed model to timeline summarization and compare it with several state-of-the-art timeline summarization algorithms on social media datasets with ground truth.

	Andy Murray dataset		*David Ferrer* dataset		*Maria Sharapova* dataset		*Roger Federer* dataset	
	ROUGE-2	ROUGE-L	ROUGE-2	ROUGE-L	ROUGE-2	ROUGE-L	ROUGE-2	ROUGE-L
LexRank†¶	0.07453	0.19753	0.08421	0.20942	0.15541	0.35570	0.11865	0.26967
ETS†¶	0.09765	0.30739	0.08738	0.26087	0.03580	0.15776	0.10909	0.29779
TMP†¶	**0.24242**	0.44295	0.10614	0.26111	0.12618	0.27586	0.13953	0.29730
LTR‡¶	0.17578	0.36576	0.30612	0.43655	0.23245	0.40964	0.12620	0.34286
K-Means + LTR‡¶	0.12062	0.34496	0.32376	0.44156	0.29146	0.43000	0.28906	0.43580
Hiscovery + LTR‡¶	0.13044	0.40157	0.34626	0.48329	0.44501	0.54453	0.25433	0.44529
DTM + LTR‡¶	0.18255	0.42829	0.37811	0.50990	0.32161	0.50000	0.15663	0.35200
Timeline-Sumy‡§	0.22594	**0.45000**	**0.45078**	**0.59278**	**0.49869**	**0.62663**	**0.35590**	**0.53180**

Table 2: Timeline Summarization Comparison. († indicates an unsupervised learning approach, while ‡ means a supervised learning approach; § indicates the algorithm could learn the number of timeline episode K automatically, while ¶ means the algorithm needs to predefine K.)

Since there are no benchmark datasets for the studied task, we manually label 4 social media datasets for evaluation. We collect 684.9k social media posts about *Andy Murray*, 20.8k posts about *David Ferrer*, 72.9k posts about *Maria Sharapova*, and 336.9k posts about *Roger Federer* from June 22 to August 7, 2012, which overlaps with two premium tennis tournaments: Wimbledon Open Tournament and London Olympics Tournament. We manually label each timeline episode according to each tennis star's major sports activities which are reflected on the corresponding time-series, and finally generate 13, 10, 11, 13 timeline episodes for these four datasets respectively. In each dataset, all social media posts not belonging to any timeline episode are labeled as the background episode. For each timeline episode, we manually label 1 representative social media post as the episode summary. We perform standard preprocessing steps on the datasets such as removing all stop-words and filtering low frequent terms and hashtag labels. Furthermore,



(a) Micro F1 Comparison (b) Macro F1 Comparison

Figure 10: Micro F1 and Macro F1 Comparison for Timeline Episode Detection.

the granularity of time-series is per hour, and our data from June 22 to August 7 can be represented as 1128-dimension time-series. We use ROUGE-2 and ROUGE-L [12] as the metrics to assess the quality of timeline summarization.

To evaluate timeline summarization, we compare our proposed framework with the following state-of-the-art summarization algorithms: 1.) *LexRank* is a widely used traditionaltext-based summarization algorithm, and it builds a sentence to sentence graph and uses the centrality to select sentences [8]; 2.) *ETS* is a timeline summarization for news corpus, and it is a graph based approach with optimized global and local biased summarization [14]; 3.) *TPM* is a timeline summarization algorithm based on Twitter data, and it infers dynamic probabilistic distributions over interests and topics for tweets summarization [13]. Note that these three baseline approaches are not supervised learning approaches.

We demonstrate timeline summarization results in Table 2, which suggests that our method outperform state-of-art methods significantly. To deeply understand the reason why our approach outperforms *K-Means + LTR*, *Hiscovery + LTR* and *DTM + LTR*, we further compare the quality of detected timeline episodes by their corresponding episode detection algorithms. We use micro F1 score and macro F1 score as the metrics to evaluate the performance of timeline episode detection. Since baseline episode detection algorithms cannot determine the number of episodes automatically, we assume that the number of timeline episodes, K, is known in advance for a fair comparison. We illustrate timeline episode detection results in Figure 10. We observe that the proposed nonparametric generative model often obtains the best performance since it uses the life cycle models to capture unique temporal properties of timeline episodes and it captures content and temporal information simultaneously.

5. Content-Rich Models of Participants and Social Media

5.1. Point Process Models for Diffusion Network Inference from Text Cascades

In many social media platforms, we are often confronted with time series data with text observations. For example, tweets with time stamps recording when the tweets are published, blog posts with time stamps, and so on. There has been an increasing interest in understanding the processes and dynamics of *information diffusion* through networks. Most prior work focuses on modeling

the diffusion of information by solely exploiting the observed time stamps when different users can social media post (i.e., publicize) information. These techniques do not leverage textual information and focus mainly on *context-agnostic* tasks. It has been shown that models that are based on observed time stamps become more effective at discovering topic-dependent transmission rates or diffusion processes when combined with the textual information associated with the information propagation. Nevertheless, previous work along this line assumes that either the topics associated with the diffusion process are specified in advance or that the influence paths are fully observed.

In our recent work [10], we develop a novel framework that combines topic models and the Hawkes process under a unified model referred to as the *HawkesTopic* model (HTM). HTM uses the Marked Multivariate Hawkes Process to model the diffusion of information and simultaneously discover the hidden topics in the textual information. Superficially, it captures the posting of information from different nodes of the hidden network as *events* of a Hawkes process. The *mutually exciting nature* of a Hawkes process, i.e., the fact that an event can trigger future events, is a natural fit for modeling the propagation of information in domains such as the ones mentioned above. To address the limitation that the thematic content of the available textual information is unknown, HTM builds upon the Correlated Topic Model (CTM) and unifies it with a Hawkes process to discover the underlying influence across information postings, and thus, the hidden influence network. We derive a joint variational inference algorithm based on the mean-field approximation to discover both the diffusion path of information and its thematic content. Experiments on both synthetic and real-world data sets, including a EventRegistry dataset for modeling information diffusion, and an ArXiv publication dataset for modeling scientific influence, demonstrate that HTM is significantly more accurate than several baselines for both influence network inference and topic modeling.

Theoretical Analysis of Latent Models. Latent models (e.g., hidden Markov model) have achieved significant successes in modeling and analyzing large-scale time series data. In practical applications, we are always confronted with the challenge of model selection, i.e., how to appropriately set the number of latent variables. Following recent advances in latent models via tensor decomposition, we make a first attempt in our recent work in [7] to provide theoretical analysis on model selection in latent models. Using *latent Dirichlet allocation* and *Gaussian mixture model* as examples, we derive the upper bound and lower bound on the number of latent variables K for a data collection of finite size. We show that under mild conditions, K can be effectively bounded by the data statistics, such as the number of examples, the number of features, and the document length (if available). Experimental results demonstrate that our bounds are correct and tight.

Twitter Buzz Summarization via Granger Causality. In this work, we leverage Granger causality to infer user influence models and improve the task of Twitter buzz summarization. For ease of description, we first define a *twitter context tree*, a tree structure of tweets that are connected by the reply-to relationship, and the root of a context tree is the original tweet that starts the buzz. The generated summary would be useful to present the buzz in a succinct form along with the original tweet, which helps the user to quickly understand the whole contextual information. For each tweet tree, we compute the pairwise user influence score between the Twitter user generating the root tweet and any responding Twitter users using Granger causality, and use these influence scores as input features in a ranking model for tweet summarization. The motivation is that if a user is strongly influenced by the user generating the root tweet, chances are his replies are

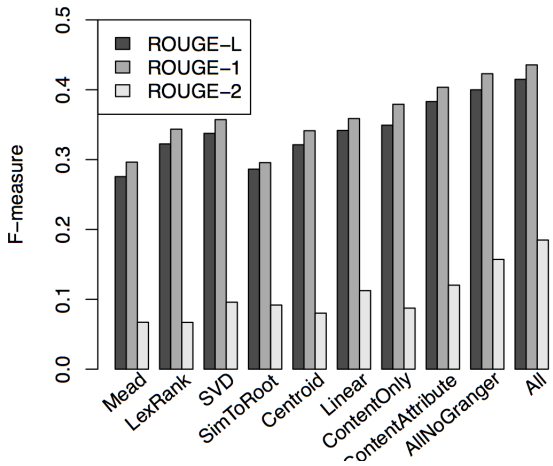


Figure 11: Overall comparison of different methods for Tweet buzz summarization

strongly coherent with the root tweet, and such a reply is more likely to be an appropriate summary candidate than other replies.

We manually select 10 representative Twitter Buzz topics from March 7th to March 20th, 2011. Among these 10 topics, 4 out of 10 topics are about Japan Tohoku earthquake and tsunami, another 4 topics are related to music shows, while the remaining 2 are gossip. The largest buzz collection contains 11,394 tweets, and the smallest context tree includes 1,106 tweets. On average, there are 4,265 tweets among these 10 topics. We use the ROUGE package[2] to evaluate the generated summaries.

In Figure 11, we compare all the methods using the F-measure of all three ROUGE metrics. The first 7 methods are text-based baselines and the last 3 methods incorporate user influence information. From this figure, we can see that the last 3 methods are much better than all the text-based methods. For example, using ROUGE-L as the metric, we can improve over the Linear method by 21% by including the Granger causality influence scores (ALL).

5.2. Richer Models of Participants and Links from Text

Goal of this Work. In this portion of the project we developed methods to automatically identify influential people in contentious and other discussions on social media, based on the characteristics of their communications. Our approach is to automatically identify the salient characteristics of types of individuals and groups from the text they produce and share, and to construct models of how these characteristics reflect leadership and other social qualities expressed by participants. We offered these characteristics for inclusion into the general information diffusion models developed by others in the project, which would provide a richer basis over which their research can discover statistical regularities that signify persuasion campaigns, emergent communities, etc. Our work focused mainly on identifying individuals who lead others in forming opinions, and who shape others opinions about controversial topics.

[2]http://berouge.com

Approach: We worked in social media domains centering on controversial topics (such as politics, religion, joint projects, etc.), in which participants argue about specific attitudes or courses of action. We defined several roles typically played by participants, including

- **Leader**: someone who guides the discussion and whom others follow
- **Follower**: someone who doesnt add much, but simply follows another
- **Rebel**: someone who disagrees with the prevailing opinion, but does so in a way that the others take seriously by responding
- **Idiot**: like a Rebel, someone who disagrees with the prevailing opinion, but is irrelevant, irritating, or incoherent, so that no-one takes him or her seriously
- **Voice in the Wilderness**: someone who someone who disagrees with the prevailing opinion but is ignored, despite having valid arguments

For each role, we defined one or more models, and, through text annotation and analysis, identified the characteristics of the text of each role. We then built rules or applied machine learning techniques to build systems that instantiate the models. We evaluated their performance against held-out data.

These models vary in complexity. Some of them, like parts of Leadership, require simple counting of the number of turns/participants who adopt the terminology they introduce, etc. Others, like Idiocy, require determining whether the contributions made by a participant are sensible. This requires analyzing various aspects of the text, including its argumentation structure (to determine logical flow and *non sequiturs*), he emotional content (to determine *ad hominem* and other remarks), and adherence to rules of argumentation (such as those defined in Wikipedia). We therefore developed methods to decompose the structure of contentious discussions/arguments, in order to help determine which participants support which arguments, and how logical their reasoning is when they do so.

In this section, we first describe early work on determining the Speech Act distributions of pairwise discussions within blogs, then move to the social leadership roles adopted by people in contentious discussions.

Data and Annotation. During the project we used several datasets, including some we created and made it available for public use.

Set 1: Initially we worked exclusively with one dataset: the discussion corpus from the **Wikipedia Article for Deletion (AfD) forum**[3]. Wikipedia, being a very large peer production system, has its own decentralized governance system to maintain the quality of articles created by users. Any Wikipedia user can nominate any article on Wikipedia for consideration for deletion. Any interested participants are then welcome to state their stance (usually in their first contribution to the discussion), after which an open discussion ensues. Participants may change their stance during the discussion. The discussion continues for at least 7 days. When it has died down, a

[3]http://en.wikipedia.org/wiki/Wikipedia:Articles_for_deletion

Wikipedia administrator officially assigned to the discussion declares the final consensus, which may or not reflect the numerical majority of the participants opinions.

We downloaded 92066 discussions with 781909 distinct comments and 47066 distinct users spanning the time Jan 1, 2009 to June 30, 2012. Each discussion focuses on one Wikipedia page, which someone initially suggests should be changed, deleted, moved elsewhere, etc. A discussion between supporters and detractors ensues. When discussion has died down, a Wikipedia expert employee makes the final decision.

Each discussion includes the title of the article in question, the nominator of the discussion, all the comments in the discussion, and the final outcome alone with the admin who imposed it. Each comment in the discussion includes the user who posted it, the stance of the user if specified, the time of the comment, and the level of the comment. The level refers to information about whether the comment is a new thread or a reply to some previous comment/reply. The data extracted was verified by cross checking 25 random discussions with the original discussions on the website. The corpus consists of a total of 457668 distinct instances of stance. Since of the 6 possible stances, merge, redirect, and transwiki combine to a total of only 7.16%, and since they rarely occur together in the same discussion, we simplified the data by combining them to the single stance compromise. Similarly, the outcome of the discussion was also converted to compromise if appropriate. Also, as an outcome, withdraw implicitly means keeping the article. Therefore, it was converted to keep. We create a timeline for each discussion based on the chronological order of each comment. Whenever users state their stance for the first time in the discussion, it is propagated to their subsequent comments unless they explicitly state a change in stance.

Set 2: The second set contains 10 discussions from the online **4forums.com forum**[4]. These discussions are political debates on controversial topics. Unlike the AfD corpus, participants do not have to state their stances explicitly, there is no time limit for how long the discussion can continue, and there is no official moderator or facilitator. Participants on this forum express their views and argue with others without any visible goals to achieve.

We selected these rather different corpora in order to ensure that our model of social roles and their signals holds up in general. Although both sets consist of contentious discussions, the nature of the discussion is very different. Where the AfD discussions have a measured and polite tone, the 4forums discussions can become quite heated and ad hominem. Participants on the Wiki forum are goal-oriented —they want their stance to be the final consensus of the discussion— while participants on the 4forums.com forum are opinion-oriented — they are primarily focused on presenting their own viewpoint. This is a classic example of argumentation by reason vs. argumentation by insistence. Statistics for both Wikipedia:AfD and 4forums.com corpus are given in Table 3.

Set 3: In Year 1, we also worked with the **W3C blog corpus**. We obtained approximately one dozen blogs in the W3C community, involving about 160 participants. Each blog contains a multi-turn discussion about a specific W3C topic. The blogs are: agenda@ietf.org, discuss@apps.ietf.org, public-esw@w3.org, public-usability-workshop@w3.org, public-ws-chor@w3.org, w3c-ietf-xmldsig@w3.org, w3c-rdfcore-wg@w3.org, w3c-wai-au@w3.org, w3c-wai-er-ig@w3.org, w3c-wai-ig@w3.org, w3c-wai-gl@w3.org, w3c-wai-eo@w3.org, w3c-wai-gl@w3.org,

[4]http://www.4forums.com/political/

Table 3: Corpus statistics

	Wikipedia	4forums
no of Discussions	80	10
no of Participants	500	107
no of Turns	1487	624
no of Words	96138	51695

Table 4: Example Speech Act predominance and associated social role

Person	Predominant Speech Act	Inferred Social Role
A	Request, Order	Boss
B	Inform	Worker
A	Request, Acknowledge	Client
B	Commit	Provider
A	Ask	Pupil
B	Inform, Explain	Teacher

www-font@w3.org, www-amaya@w3.org. We analyzed these for utility but found them not controversial enough to support our research. Nonetheless, they did provide some useful results showing the distribution of Speech Acts.

Speech Act Model. Some sentences just convey facts. Others actually constitute acts by themselves. For example, "I promise to…" is both a fact (that the speaker promises) and a promise (an actual undertaking by the speaker). **Speech Acts** are such actions that are conveyed by utterances. Four basic types have been defined in the discourse linguistics literature: *Request/Order (Directive), Provide-Information (Constative), Commit-to (Commissive)*, and *Acknowledgement*.

Our initial project focused on determining Speech Acts in one class of social media, namely email. The premise of this work was: if you can determine asymmetries in Speech Act behavior between pairs of people, you can predict their social roles relative to one another; examples are shown in Table 4.

We hand-annotated 1305 examples (each example one sentence in a W3C blog discussion) with one of four Speech Act labels Acknowledgment (138 sentences), Constative (709), Directive (60), and Commissive (237). In addition, 234 sentences carried no Speech Act and received no label. Using 0%, 3%, and 10% of this as a test set, we trained two classifiers on the respective remainders: the J48 Decision Tree and SVM classifiers. Results are shown in Table 5.

Our hypothesis was that we would find clusters of individuals: those who provide far more Acknowledgement than Directives (the followers) and those who issue more Directives than Acknowledgements (the leaders). Analyzing the results, we searched for pairs of individuals with asymmetrical Speech Act behavior. For example, Brian McBride issues Directives to several people, in particular Jeremy Carroll, Dan Connolly, and Eric Miller, but they provide only Constatives (information) in return, but few if any Directives. We hypothesized that McBride is socially and/or organizationally recognized as a leader in the relevant W3C subcommunity.

To determine all such asymmetries we plotted the ratios of Directive to Constative, Acknowledge to Directive, and other pairs of Speech Acts for all pairs of individuals.

Table 5: Results of W3C blog discussion sentence classification of 4 Speech Acts

J48 Decision Tree		TP Rate	FP Rate	Precision	Recall	F-Measure
Acknowledgement	Cross Validation	0.975	0.150	0.975	0.975	0.975
	Test	0.978	0.097	0.977	0.978	0.977
Constative	Cross Validation	0.809	0.198	0.808	0.809	0.808
	Test	0.806	0.192	0.811	0.806	0.805
Commissive	Cross Validation	0.951	0.936	0.920	0.951	0.932
	Test	0.948	0.956	0.912	0.948	0.930
Directive	Cross Validation	0.911	0.326	0.908	0.911	0.905
	Test	0.903	0.403	0.913	0.903	0.890
None	Cross Validation	0.904	0.379	0.901	0.904	0.895
	Test	0.851	0.512	0.848	0.851	0.826
SVM, Radial Basis		TP Rate	FP Rate	Precision	Recall	F-Measure
Acknowledgement	Cross Validation	0.959	0.359	0.961	0.959	0.954
	Test	0.978	0.144	0.978	0.978	0.977
Constative	Cross Validation	0.854	0.164	0.860	0.854	0.852
	Test	0.821	0.177	0.832	0.821	0.820
Commissive	Cross Validation	0.954	0.954	0.910	0.954	0.931
	Test	0.955	0.955	0.912	0.955	0.933
Directive	Cross Validation	0.909	0.412	0.917	0.909	0.896
	Test	0.888	0.465	0.902	0.888	0.869
None	Cross Validation	0.965	0.042	0.968	0.965	0.966
	Test	0.948	0.119	0.947	0.948	0.947

Table 6: Inter-annotator agreements

	Kappa coefficient
Annotator1–Annotator2	0.74
Annotator2–Annotator3	0.71
Annotator3–Annotator1	0.69

Unfortunately, we did not find any strong patterns. We concluded that our starting hypothesis was not provable in the W3C blog domain. We therefore abandoned this line of research.

Social Role Model. We next turned to identifying and analyzing the social roles that people play in contentious online discussions. Our conclusions do not apply only for contentious discussions, but having a topic of contention makes it easier to identify the distinct role types.

As before, we started by annotating data, and during the annotation process refining our understanding and crafting our models. When we had obtained a clear enough picture of the problem, and therefore could annotate reliably, we also had enough data to support machine learning and automated classification.

Data annotation: In order to ensure the quality of the research, we hired two annotators who, for a low hourly rate, assign judgments to each turn in our corpora. They are invaluable in helping us refine the models and formulate precise annotation instructions, which in turn help define the conceptualization of the various participant roles. We made available this work not only the annotated corpus but also the annotation manuals, enabling others to take our work further.

Initially we mainly focused on the Leader role, but later broadened our work to the other roles. To determine signals of leadership, we constructed two complementary models: *Content Leader* and *SilentOut Leader*. The models quantify the basic leadership qualities of participants and assign leadership points to them. To evaluate their effectiveness we correlated the leaders ranks produced by the two models using the Spearman Coefficient. We also proposed a method to verify the quality of the leaders identified by each model.

The annotators began with a training annotation set consisting of 10 AfD articles. Three annotators were asked to identify two basic social roles performed by participants (Leaders and Rebels, where Rebels were described as the participants who have enough contribution but are unable to exercise any kind of influence). Also, they were asked to assign any other role that would identify a participant with characteristics different from the two given roles. After the completion of the initial task, the annotators agreed upon a set of social roles for the initial coding manual. The annotators came up with a set of characteristics that define each role and also the criteria to assign values to each characteristics for each participant. Using the initial coding manual, the annotators were asked to annotate 8 more sets, each consisting of 10 AfD articles. After annotating each set, the annotators discussed the annotations and refined and/or added any roles, characteristics, or criteria that were agreed to be helpful. After completing all 8 sets, the annotators re-annotated all the discussions again using the final coding manual. The annotations for the 10 discussions from the 4forums.com forum started using the same coding manual. The final coding manual included 4 characteristics and 8 social roles, described below. Inter-annotator agreement for annotated social roles using Cohen's Kappa coefficient are shown in Table 6.

As annotation progressed, we increased the number of roles and refined their definitions and characteristics they are comprised of. Eventually, we produced the following principal roles: **Leader, Follower, Rebel, Idiot, Voice in Wilderness, Nothing Sensible, Nothing**, and **Other**.

In order to distinguish these roles we identified three aspects of participants' overall contribution, further subdivided into four identifiable characteristics, which in various combinations reflect the behaviors of these social roles: **Participation Type** (Stubbornness, Sensibleness), **Attendedness Value** (Ignored-ness), and **Influence Value** (Influence).

- **Stubbornness** captures the intransigence of a participant in the discussion. This characteristic has two components: the amount of participation and the degree of unwillingness to change opinion or stance. This characteristic differentiates between participants who form the heart of the discussion from participants who may comment only once or twice, or in minor ways only. A combination of the following criteria determine participation: the number of comments by the participant, the arguments/claims presented by the participant, and the level of engagement with other participants. We compare the number of comments by each participant against the average number of comments for each discussion, if the number of comments is higher than the average, the participant is considered stubborn. However, note that while calculating the average, we do not consider the outliers. For example, if most of the participants comment between 1–5 times, but there is a participant who comments 15 times, then while calculating the average, the participant with 15 comments is considered an outlier.

- **Sensibleness** measures the degree of 'sense' of the arguments/claims presented by a participant, which is very important for measuring his or her impact on the discussion. Sensibleness analysis is highly dependent on the domain or nature of the discussion. Therefore we defined somewhat different criteria for assigning sensibleness values in the two corpora. As mentioned, the AfD discussions are goal-oriented: each participant tries to sway the decision of the discussion in favor of their own stance. Also, since Wikipedia pages should meet the requirements stated in Wikipedia policies, as one would expect discussions on this forum sometimes revolve around such policies. Therefore the main criterion for someone to be sensible in such discussions is that they appeal to authority in support of their arguments/claims. Examples of such authority can be Wikipedia policies, links to external sources of recognized expertise, etc. In contrast, the discussions on 4forums.com are opinion-oriented, where participants primarily focus on presenting their own opinions and reasoning, and do not seriously consider that of others except to dispute it.

- **Ignored-ness** captures the attitude of participants towards each other. This attitude can be an indicator of the importance or relevance of a comment. Generally, for example, participants ignore spammers. The main criterion for ignored-ness analysis is to see whether the participant receives any replies to his or her comments. Therefore, when someone mentions a participant in his/her comment then that participant is considered not ignored. Wikipedia:AfD discussions are stored in a structured way (replies to a comment are indented), making determination of replies easy. In contrast, in 4forums discussions we have to rely on a participant quoting arguments presented by another participant, or naming them explicitly, to decide that the latter is not ignored.

Table 7: The relation between the values of characteristics and social roles corresponding to them. Stubbornness, sensibility, and influence can take values +1, 0, or -1. Ignored-ness can take values +1 or -1. An 'x' in the table indicates that the corresponding characteristic for the role can take any possible value. As one can notice, participants can thus have multiple roles. A Rebel can become Rebel-Leader if the participant has influence value +1. Similarly, a Nothing can become a Nothing-Follower if the influence value is -1.

Stubb	Sense	Ignore	Infl	Role
x	x	x	+1	Leader
x	x	x	-1	Follower
+1	+1	-1	x	Rebel
+1	+1	+1	0/-1	Voice in Wilderness
+1	-1	x	x	Idiot
-1	+1	x	x	Nothing Sensible
-1	0/-1	x	x	Nothing

- **Influence Value** helps in mainly to identify leaders and followers among the other roles. This has two aspects: (i) influence on others: was a participant able to influence another participant through their contribution?, and (ii) endorsement: did a participant acknowledge another participant's influence explicitly? This aspect helps in mainly to identify leaders among the other roles. The primary characteristic of leaders in contentious discussions is that they are able to influence others by their actions or arguments/claims. Therefore influence analysis becomes a key part of identifying leaders and followers in such discussions. The most observable indication of influence occurs when another participant changes his/her stance during the discussion and acknowledges the influential participant for influencing him/her for the change. In such cases, the participant who was influential in engendering the change is considered a leader and the participant who changes his/her stance or endorses other participants is considered a follower. Another example of indication of influence is when other participants acknowledge the influential participant through expressions such as "according to ..., or "as per ...", etc. An early implementation of a simplified model of Leadership appears in (Jain and Hovy, 2013).

Social Roles Overview. The final coding manual comprised 8 principal roles distinguished by 4 characteristics. Table 7 shows the relationship between the characteristics values and the corresponding social roles. Note that the roles Leader and Follower have a special status in that all the characteristics' values except influence are unspecified. This means that any participant, irrespective of their sensibility, stubbornness etc., may be seen as a Leader/Follower as long as they have the appropriate influence value. Hence any other roles could additionally acquire the qualities of being a Leader or Follower. Any combination of characteristics not specified in the table is annotated as the role Other.

- We define a **Leader** as a participant who manages to influence another participant to change his/her stance or influence them to follow him by endorsing through his/her actions or arguments/claims. We believe that the defining characteristic of a Leader is the amount of

influence he/she is able to induce regardless the amount or type of contribution. The characteristic combination for a Leader is (x, x, x, +1) for stubbornness, sensibleness, ignored-ness, and influence respectively.

- We define a **Follower** as a participant who is influenced by other participants and changes his/her stance or endorses other participants for their actions or arguments/claims. A Follower is also a participant who doesn't have his/her own arguments/claims, but instead restates arguments/ claims made by other participants. These participants provide support to leaders by endorsing them or by re-stating the same arguments/claims. Therefore, the contribution amount or type doesnt matter for such participants to define them as a Follower, making the characteristic combination for a Follower (x, x, x, -1) for stubbornness, sensibleness, ignored-ness, and influence respectively.

- A **Rebel** is a participant who forms the heart of the discussion and drives it in some direction. One of the main characteristics of a rebel is his/her devotion to the discussion, based on the amount of contribution and their level of engagement with other participants. The arguments/claims presented by a rebel are sensible and he/she is not ignored by other participants, which provides justification for the importance of his/her presence in the discussion. The characteristic combination for a rebel is (+1, +1, -1, x) for stubbornness, sensibleness, ignored-ness, and influence respectively.

- A **Voice in the Wilderness** is very similar to a Rebel in the amount and type of contribution which forms the heart of the discussion. The arguments/claims presented by a voice in wilderness are sensible as well. The only difference between a Voice in the Wilderness and a Rebel is that the former is ignored by other participants for some reason. Therefore the sensibility value for a Voice in the Wilderness is important to distinguish them from potential spammers, since spammers are never regarded as sensible contributors. The characteristic combination for a Voice in the Wilderness is (+1, +1, +1, x) for stubbornness, sensibleness, ignored-ness, and influence respectively.

- An **Idiot** is a participant whose contribution to the discussion is not towards serving the purpose of it. He/she does participate a lot in the discussion but the content may be either emotional, illogical, or completely off-topic. The ignorance characteristic has no significance in defining the role of an idiot since the main criterion for labeling a participant an idiot is the non-sensible contribution by himself/herself. The characteristic combination for an idiot is (+1, +1, +1, x) for stubbornness, sensibleness, ignored-ness, and influence respectively.

- **Nothing** and **Nothing-Sensible** are participants who make minimal contribution to the discussion and hence cannot be considered stubborn enough to stick to their arguments/claims. As a result, they may not have a major influence on the course or the outcome of the discussion. We distinguish between Nothing and Nothing-Sensible based on the number of sensible arguments/claims, in order to distinguish potential spammers from those who may have minimal but legitimate contribution. The latter may or may not be ignored based on the type of contribution. The characteristic combination for a nothing is (-1, -1, x, x) for stubbornness, sensibleness, ignored-ness, and influence respectively, and the characteristic combination for a nothing sensible is (-1, +1, x, x) for stubbornness, sensibleness, ignored-ness and influence respectively.

Detail of Social Role Leadership. As stated above, we modeled Leadership in two ways: the Content Leader and the SilentOut Leader models.

Models: We defined the **Content Leader model** to quantify the language use of a participant to (i) encourage others to follow his arguments (*Attract Followers, AT*) and (ii) counter the arguments from the opposing groups (*Counterattack, CA*). The AF component we model by matching the n-grams (word sequences) of a user against the n-grams used by another user having the same stance. The user whose language ngrams others copy in subsequent statements acquires leadership points. In the CA component we model the quality to stand up against opponents and try to nullify the arguments presented in oppose to the users stance, by matching the n-grams of a user against the n-grams of users holding a different stance. The user who counterattacks acquires leadership points. In addition, the user who is getting counterattacked also gets some leadership points. This is because of his/her ability to attract the attention of opposing groups, which implies a significant contribution in regards to the discussion. Both these attributes implicitly quantify another ability of a leader, which is having command over the course of the discussion. The fact that other users having the same or different stances are using the same argumentation words implies that the leader is guiding the discussion.

We define the equation for leadership points for Content Leader model for any user in any discussion by:

$$\alpha.AF + \beta.CA^{oppose} + \gamma.CA^{opposed}$$

where

AF = weight for n-grams originated by the user that are also used by other users having same stance

CA^{oppose} = weight for n-grams originated by user found also in n-grams of users with different stance

$CA^{opposed}$ = weight for n-grams originated by users having different stance, found also in n-grams of user

α, β, γ = weights of corresponding attributes (see below)

We defined the **SilentOut Leader model** to quantify the ability of a leader to silent opposing users out with his arguments: (i) giving arguments that cannot be countered (*Factual Arguments, FA*) and (ii) winning the small battles in the discussion (*Small Wins, SW*). The FA component considers whether a user presents an argument that none of the users from the opposing stances attack, which contributes to the leadership qualities of that user. We model this attribute by giving a constant amount of leadership points to users for each comment that elicits no reply from any opposing stance user. The SW component refers to the ability of a leader to silence out other users by countering their individual arguments and thus winning a small battle over other users. This counterargument not only nullifies the original argument from the opposing user but also strengthens the arguments for the leader's own stance. To model this attribute, we divide the discussion into small argument sections. Each argument section contains an original argument and at least one reply from a user from opposing stance. For each such argument section, the user who has the last say (i.e., whose reply gets no counterargument) acquires a constant amount of leadership points.

The equation for the leadership points for SilentOut Leader model for any user in any discussion is given by:

Table 8: Comparison of accuracy for different coefficient values for Content Leader model

α	β	γ	% accuracy
1	1	1	53.59
1	1	2	52.88
1	2	1	59.60
1	2	2	55.33
2	1	1	54.68
2	1	2	53.48
2	2	1	60.90

$$\alpha.FA + \beta.SW$$

where
FA = no of comments from user which didn't have any reply from any user from opposing stance
SW = no of small battles user won
α, β = weights of corresponding attributes (see below)

Experiments: To determine the weight of any n-gram, we use the inverse document frequency (idf) of that n-gram across all the discussions in the whole corpus. The idf value of an n-gram denoted by t is given by

$$idf(t) = log(N/|n \in N : t \in n|)$$

where
N = no of discussions in the corpus
$|n \in N : t \in n|$ = no of discussions where the n-gram t appears
We use unigrams, bigrams, and trigrams as part of our n-grams and calculate the weights across the whole corpus, not only the contentious discussions.

We processed discussions comment by comment, extracting n-grams from the text and storing them in the bag of n-grams for the user. We also kept track of the stance of each user as the discussion progresses. Given a new comment by user B, we extracted the n-grams in this comment. For each user A, who has already participated in the discussion, we matched the n-grams used by user B to the n-grams in the bag of user A.

As experimentation showed, the absolute values of α, β, and γ are not significant. Tables 8 and 9 show an analysis of relative values of model coefficients on the outcome prediction accuracy for Content Leader and SilentOut Leader models respectively (for details on the evaluation see below). We choose the coefficient values related to the best accuracy for each model to create the leadership rankings, which are used to calculate the Spearman correlation between the two models. But their relative values are important because that reflects their relative importance to the corresponding attributes of the model. Because of the lack of gold standards for leaders, we set the values for coefficients manually.

Evaluation: To evaluate the effectiveness of the models, we conducted two evaluations.

First, we calculated the correlation between Content Leaders and SilentOut Leaders. Using the two leadership models, we calculated leadership points for each user in each of the 8292 contentious Wikipedia AfD discussions. We then aggregated the leadership points for user across all the discussions and calculate average leadership points per discussion. Using the average score,

Table 9: Comparison of accuracy for different coefficient values for SilentOut Leader model

α	β	% accuracy
1	1	64.05
1	2	65.01
2	1	62.25

Table 10: Spearman correlation between models

Min participation	No of users	ρ
1	9489	0.23
5	2166	0.30
10	1144	0.46
20	533	0.48
50	183	0.64

we created a ranked list of leaders for both the models. We compared the ranked list using the Spearman's rank correlation coefficient. For any two ranked lists, the coefficient is calculated as

$$\rho = 1 - 6 \sum d^2 / N(N^2 - 1)$$

where
d = difference in statistical rank of corresponding user
N = total number of users under consideration

Table 10 shows the Spearman rank correlation between the Content Leader and SilentOut Leader models. The first column limits the users by minimum number of different discussions they must have participated in to be considered for correlation calculation. The second column shows the number of users satisfying the criteria for minimum number of discussions and the last column shows the correlation coefficient for those users. Positive correlation coefficient implies that the models complement each other for identifying same users as leaders.

As a second evaluation, we predicted the outcome of the discussion. Our work presents a possible method to test the quality of leaders identified by each model. The criterion to qualify as a quality leader is to be able to turn the outcome of the discussion into ones favor. We compared our models with some other nave models: *Majority Stance* (outcome of the discussion is predicted as the stance with the majority of votes); *Majority Content Stance* (outcome of the discussion is predicted as the stance which has largest number of words in support); *Talkative Leader* (outcome is predicted as the stance of the user who has contributed the largest number of words in the discussion); *Content Leader* (outcome is the stance of the user with the highest leadership points per the Content Leader model); *SilentOut Leader* (outcome is predicted as the stance of the user with the highest leadership points per the SilentOut Leader model); *Content and SilentOut Leader* (outcome is predicted as the stance of the user with the highest linear combination of Content and SilentOut Leadership points. The combination coefficients are set manually because of the lack of gold standard data. After experimenting with various coefficient values we set them to 1 and 2 respectively).

Table 11 shows the comparison of accuracies for predicting the outcome of Wikipedia AfD contentious discussions based on the different models described above.

Table 11: Comparison of outcome prediction accuracy

Model	% accuracy
Majority Stance	45.59
Majority Content Stance	42.56
Talkative Leader	40.28
Content Leader	60.90
SilentOut Leader	65.01
Content and SilentOut Leader	**68.34**

After a long period of refinement of the model, our annotation of contentious discussions stabilized to the degree where we do not expect many significant changes if we were to address a new contentious corpus. While we do not expect to find many additional roles, we continue actively to discuss the nature and organization of their underlying characteristics.

Sensibleness Heuristics. While most of the above-mentioned factors can be computed relatively easily, measuring the degree of sense a speaker is making is much more difficult.

We therefore analyzed the text of contributors in our two corpora and compared the nature of those we considered *sensible* to those we considered *nonsense*. After considerable discussion and annotation (see below), we identified the following factors that contribute toward making a person be deemed *sensible*. Note that this is never something that a single turn can establish; it always requires multiple turns in an ongoing interaction, the longer the better.

We identified three principal relevant areas, each with specific heuristics (most of which could be codified using rules):

- Speaker's adherence to the standard rules of argumentation. Generally, the better adherence, the more sensible the speaker looks. Specifically:
  - explicitly mentions Wiki argumentation rules/policies
  - refers to external sources (most of the time sensible)
  - provides conceptual starting points
  - reasons logically from premise to conclusion
  - doesn't start tangential discussion (relevancy)
  - expands other people's comments (refer to previous comment)
- Speaker's expertise/knowledge of the domain. Generally, the higher expertise, the more likely the speaker is sensible.
  - as far as possible, judge whether the comment is logical and not crazy
  - doesn't start tangential discussion (relevancy)
  - sounds knowledgeable; use expert terminology
- Speaker's level of emotion. This is a somewhat secondary factor where the argumentation is strong.

- the general tone of the contribution — how many emotion words are included in the turns
- when emotion is shown, how it appears relative to what came before (in response to others, or just strengthening one's point)
- in the emotion-bearing statement, the theme is an action and not the other person (therefore feels 'softer')
- no *ad hominem* attacks

However, to properly characterize and quantify each turn, we had to determine its role with respect to the overall argument. We therefore had to determine argument structure.

The Structure of Online Arguments. Since antiquity, the discipline of rhetoric has studied the nature of argumentation and persuasion (Aristotle, 1954). Modern social media discussions offer a rich area for study. For example, both sentences "This page violates Wikipedia policies" and "The page violates Wikipedia policies because it has no sources" express an opinion about some page that violates the Wikipedia policies, but the second is more persuasive because it provides the reason for the opinion.

Automatically creating the rhetorical argumentation structure, in this case recognizing that the second sentence contains a claim that justifies the opinion, is required for various aspects of discourse analysis, including identifying claims and their authors, finding which claims function as justifications, and for which others, which claims contradict others, etc. In our work we explored the problem of automatically creating the rhetorical argumentation structure of controversial online discussions: how people state and connect in an attempt to persuade others. We used the word **claim** to refer to any assertion made in a discussion, which the author intends the reader to believe to be true, and that can be disputed. Our contribution included procedures that:

- detect sentences that express claims,
- identify specific claims within the sentences,
- discover links between claims if any,
- create the rhetorical argumentation structure.

Annotation: The coding manual for annotating claims was created by 3 annotators. The annotators were initially told that a claim is any assertion by a user that can be disputed. Rather than annotating whether a sentence is a claim or not, the annotators were asked to identify specific claims within the sentences, using xml tags. This enables treatment of the many sentences containing more than one claim, and helps in identifying specific claims automatically. The first annotation round consisted of 10 AfD discussions (313 sentences). After the first round, the annotators discussed results, and two annotators created the coding manual. Using this coding manual, the annotators then annotated another 10 AfD discussions (422 sentences). To calculate agreement scores, we say that two annotators agree when a part of a sentence is marked as a claim only if the starting word of the claim by annotator1 is within 2 words of the starting word of the claim by annotator2, with the same criteria holding for the end word of the claim. The agreement scores using

Table 12: Inter-annotator agreements for claim annotation

	Kappa coefficient
Annotator1-Annotator2	0.72
Annotator2-Annotator3	0.68
Annotator3-Annotator1	0.67

Cohen's Kappa coefficient appear in Table 12. Using this final coding manual, the third annotator annotated 70 more AfD discussions, which were used for our experiments.

The coding manual for annotating links between claims was created by two annotators. Once claims have been identified with sentences, determining which claims are connect becomes much easier. The annotators were asked to annotate for links between claims that may establish relations such as conclusion, justification, contradiction, etc. The annotators were also asked to annotate the direction of the relation between (i.e., whether claim1 is dependent on claim2 or vice versa). The annotators were given 5 AfD discussions consisting of 201 sentences. The Cohens Kappa agreement score for these annotations was 0.84.

Automated Claim Identification and Link Detection: We approached the problem of automated claim identification as a tagging problem. We pre-processed the annotated sentences to assign B_C, I_C and O tags to each word, where B_C indicates word at the beginning of the claim, I_C indicates word inside any claim and O indicates word outside any claim. We used Conditional Random Fields (CRF) to determine tags for each word. For each word w_i, we extract the two words before it and two words after it. Using these words we create five types of features (question mark, POS tags, etc.), described in (Jain and Hovy, submitted).

We also treated link detection as a classification problem. We formed claim pairs and determine whether they are linked. As noted earlier, most of the claims linked together are either in the same sentence or in the adjacent sentences. Therefore for a particular claim c_i in sentence s_i, we create pairs (c_i, c_j) for all the claims $c_j(ij)$ in sentence s_i and pairs (c_i, c_j) for all the claims c_j in sentence $s_i + 1$. We did not create a separate entry for the reversed pair (c_j, c_i). Each pair (c_i, c_j) was classified as either having no link between them, ci_i is dependent on c_j or c_j is dependent on c_i. For each claim pair (c_i, c_j) we used features such as word and POS ngrams, claim distance between c_i and c_j, etc.

We used the machine learning package Weka for the Claim Detection and Claim-Link Detection classification experiments. In Weka, we selected SVM with radial basis function and J48 Tree classifiers. We used CRFsuite for the Claim Detection and Claim Identification experiments. For each experiment, we trained the model on 80% of the data and tested it on the remaining 20% of the data.

Results for claim detection and claim linkage are reported in Tables 13 and 14 respectively.

Sensibleness Features:

We calculated several features for each turn, and summed them for each participant, to build a vector characterizing the participant's degree of sensibleness.

1. Tangentiality: Participants who tend to deflect from the main subject of the discussion are considered to be non-sensible as they contribute nothing constructive towards the discussion. For

Table 13: Claim detection results

Experiment	F-score
Majority	0.70
Question	0.72
SVM	0.83
J48	0.82
CRF	0.85

Table 14: Claim-Link detection results

Experiment	F-score
Adjacent	0.70
Same Sentence	0.77
SVM	0.86
J48	0.89

each participant we categorize each of his/her comments as tangential to the discussion or not. To quantify this we used *itf-ipf*, a slightly modified version of *tf-idf*, to approximate tangentiality of any comment. We used the fact that any tangential comment would tend to be ignored by other participants, i.e., the words used in the tangential comments would be used relatively less than other words overall and would be used by relatively fewer participants. We calculated *tf* (term frequency) and *pf* (participant frequency: total number of participants who used the word in the discussion) and compute the *itf-ipf* value for each word *w* in a comment as:

$$w_{itf-ipf} = 1/w_{tf} * log(N/w_{pf})$$

where N = total number of participants. Using the *itf-ipf* value for each word, we calculate the tangential quotient (TQ) for a comment (C) as:

$$TQ_C = (\sum_{w \in C} w_{itf-ipf})/N_w$$

where N_w = total number of words in the comment.

We divided the total *itf-ipf* value by the total number of words to nullify the effect of the length of the comment on the tangential quotient. To determine whether a comment is tangential, we compared the tangential quotient value of that comment to the average tangential quotient for that discussion. If the value is more than 1.5 standard deviations of the average (μ+1.5σ), then we said that the comment is tangential.

We used the percentage of tangential comments from a participant as one of the features for the model.

2. Reference: Reference by others in the discussion (of which peer reviews are an example in published literature) provide an external opinion on the sensibility of a participants contribution to the discussion. They can therefore play a significant part in determining the sensibility of a participant, as a system without domain knowledge of the discussion cannot verify the validity of claims.

We first identified all sentences that contain references to other participants. We used the NLTK toolkits NER (Name Entity Recognition) module to identify reference to participants. We also

identified second person pronouns in replies to other participants as reference. Next we analyzed the sentence that contains the reference using NLTKs sentiment analysis module. If the sentence was polar, then we checked the polarity of the sentence; if it was positive, then we considered it a positive peer review towards the participant referenced in the sentence, and vice versa. We used the number of positive reviews and the number of negative reviews as features for the sensibleness analysis.

3. Questions: The percentage of sentences that were questions: It can be a good strategy to ask questions related to the discussion. But asking too many questions can be considered as non-sensible.

4. Comments: The percentage of domain-irrelevant comments like personal attacks: We used this feature to identify participants who constantly attack others rather than presenting their own arguments. We use a similar method to that for peer reviews to identify comments that are targeted towards other participants and have negative polarity.

Classification of Sensibleness:

We used the Weka classifier to determine the sensibility value for each participant. We trained an SVM with radial basis function classifier using the Wikipedia discussion over the features explained above. We report the F1-score from Weka using 10-fold cross validation. We compared our model with the following:

- Everyone: The most naive baseline would be to classify everyone to be sensible. This model gives us an idea of sensibility trend for that particular domain.
- Argumentation model: We create a model using features from argumentation structure only and classify participants for sensibleness.
- Peer review model: We classify any participant as sensible if he/she has equal or more positive reviews than negative reviews.
- Tangential model: We classify any participant as sensible if he/she has fewer than 25% tangential comments.
- Policy guy: We classify any participant as sensible if he/she mentions Wikipedia policy in any of his/her comments.

To verify whether the insights regarding the domain of the discussions help in our sensibility analysis, we create another model with an additional feature "percentage of comments that mentions Wikipedia policy. We report the results of our models in Table 15.

Since there are no discussion policies for 4forums.com, we cannot create the correspond-ing models for it. We use the model trained on Wikipedia to classify sensibleness on 4forums.com in order to compute cross domain accuracy; see Table 16.

Application of Ideas. During the project we applied our methods also to the SMISC Social Action Storm exercise corpus that was available for some period.

Table 15: Sensibility analysis result for Wikipedia

Model	F1-score
Everyone	0.76
Argumentation model	0.82
Peer review model	0.78
Tangential model	0.79
Policy guy	0.74
Sensibility Model	0.87
Sensibility Model + Policy	0.89

Table 16: Sensibility analysis result for 4forums.com

Model	F1-score
Everyone	0.74
Argumentation model	0.75
Peer review model	0.72
Tangential model	0.71
Sensibility Model	0.77

Overall Conclusion. We were quite pleased with the quality of the results. Earlier work in automated discourse structure analysis suggested that it is a very difficult problem. But the Articles for Deletion corpus, which contains argumentative / contentious discussions with relatively clear-cut topics and only one topic per discussion, and is well-structured with (mostly) grammatical language, makes a big difference. Also, breaking down the procedure into a series of smaller steps, first focusing just on identifying claims, makes the task a lot easier. A fruitful line of future work would be to identify the principal kinds of links that hold between claims (justifications, elaborations, contradictions, etc.), using idea from Rhetorical Structure Theory (Mann and Thompson, 1988) and similar.

In this work we only scratched the surface of the problem of identifying whether someone makes sensible comments. Still, the success of the approach of counting some surface features to determine sensibleness is encouraging. Sensibleness analysis can be highly subjective and domain dependent — what is sensible in a gaming forum may not be sensible in health support forum. It should be possible to create a model that can be used across domains that allows tweaking the model based on the target domain, especially when using features that are harder to model, such as differentiating between global level and local level comments in political discussions.

References

[1] Armen Allahverdyan and Aram Galstyan. Opinion dynamics with confirmation bias. *PLoS ONE*, 9(7):e99557, 07 2014. doi: 10.1371/journal.pone.0099557. URL http://dx.doi.org/10.1371%2Fjournal.pone.0099557.

[2] T. Bahadori and Y. Liu. An examination of practical granger causality inference. In *SIAM Conference on Data Mining (SDM'13)*, 2013.

[3] Taha Bahadori, Yan Liu, and Eric P. Xing. Glarp: Fast structure learning in generalized stochastic with latent factors. In *KDD'2013*, 2013.

[4] Yi Chang, Antonio Otega, and Yan Liu. Ups and downs in buzzes: Life cycle modeling for temporal pattern discovery. In *Submitted to ICDM*, 2014.

[5] Yi Chang, Jiliang Tang, Dawei Yin, Makoto Yamada, and Yan Liu. Timeline summarization from social media with life cycle models. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016*, 2016.

[6] Dehua Cheng, Mohammad Taha Bahadori, and Yan Liu. Forward backward granger causality: a simple and effective approach in mining time series. In *KDD*, 2014.

[7] Dehua Cheng, Xinran He, and Yan Liu. Model Selection for Topic Models via Spectral Decomposition. *AISTATS*, 2015.

[8] Gunes Erkan and Dragomir R. Radev. Lexrank: Graph-based lexical centrality as salience in text summarization. *Journal of Artificial Intelligence Research (JAIR)*, 22:457–479, 2004.

[9] Thilo Gross and Bernd Blasius. Adaptive coevolutionary networks: a review. *Journal of The Royal Society Interface*, 5(20):259–271, 2008. ISSN 1742-5689. doi: 10.1098/rsif.2007.1229. URL http://rsif.royalsocietypublishing.org/content/5/20/259.

[10] X. He, T. Rekatsinas, J. Foulds, L. Getoor, and Y. Liu. HawkesTopic: A Joint Model for Network Inference and Topic Modeling from Text-Based Cascades. *International Conference on Machine Learning (ICML)*, 2015.

[11] Theodore Levitt. Exploit the product life cycle. *Harvard Business Review*, 1, November 1965.

[12] Chin-Yew Lin. Rouge: a package for automatic evaluation of summaries. In *Proceedings of the Workshop on Text Summarization Branches Out (WAS)*, 2004.

[13] Zhaochun Ren, Shangsong Liang, Edgar Meij, and Maarten de Rijke. Personalized time-aware tweets summarization. In *Preceedings of SIGIR*, 2013.

[14] Rui Yan, Xiaojun Wan, Jahna Otterbacher, Liang Kong, Xiaoming Li, and Yan Zhang. Evolutionary timeline summarization: a balanced optimization framework via iterative substitution. In *Preceedings of SIGIR*, 2011.