Connectors, Mavens, Salesmen and More: An Actor-Based Online Social Network Analysis Method Using Tensed Predicate Logic

Joshua S. White, PhD
Department of Computer Science
State University of New York Polytechnic Institute

Jeanna N. Matthews, PhD
Department of Computer Science
Clarkson University

ASE SocialInformatics2014
December 16, 2014

Outline

Initial Motivation

Partially inspired by Gladwell's book, *The Tipping Point* [1], in which he discusses how life can be thought of as an epidemic. Some criticism exists as to Gladwell's rigor, however for our use it is about inspiration and motivation not accuracy.

The Books Key Points *"for our purposes"*

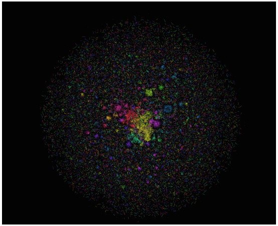
- Actors (Connectors, Mavens, Salesmen).
- Information spreads like disease.
- Ideas reach a tipping point (critical mass).

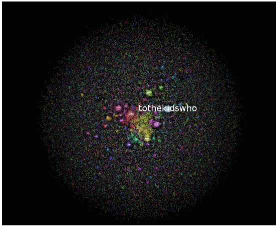
Let's Face It - Social Networks Are Fun

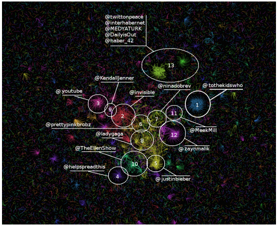
- We are a social species, that enjoy communicating and self adulation.

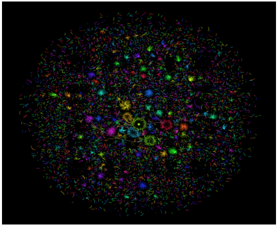
Problem Questions

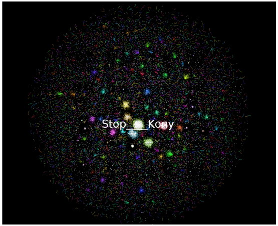
- Are there information security applications for social network data-mining?
  - ✓ Can we detect malicious social network use?
  - ✓ Can we analyze the spread of a major malware campaign?
  - ☆ Can we detect phishing in near-real-time
- Can we determine how information spreads on these networks?
  - ☆ Can we determine if a user is unique?
  - ★ Is there a way of classifying users based on actor types?
  - ☆ Can we determine who the opinion leaders or influencers are?

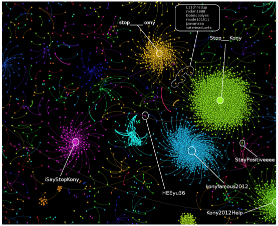



tothekidswho


@twittonpeace
@interhabernet
@MEDYATURK
@DailyisOut
@haber_42
13
@KendallJenner
@youtube
@invisible
@ninadobrev
@tothekidswho
@prettypinkbrobz
@ladygaga
@MeekMill
@TheEllenShow
@zaynmalik
@helpspreadthis
@justinbieber




Stop Kony


stop___kony
Stop_Kony
StayPositiveeee
iSayStopKony
HEEyu36
konyfamous2012
Kony2012Help

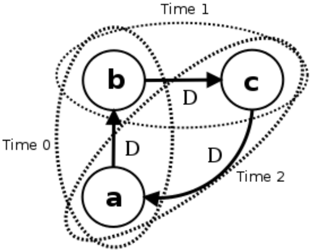
Actor Descriptions

- Isolate (Developmental Psychology) [27]
- Connector (Tipping Point) [1]
  - Star (Small World Problem) [26]
  - Bridge (The Hidden Organizational Chart) [2]
  - Liason (The Hidden Organizational Chart) [2]
- Maven (Tipping Point) [1]
- Salesmen (Tipping Point) [1]

Actor Identification Example: Liaison

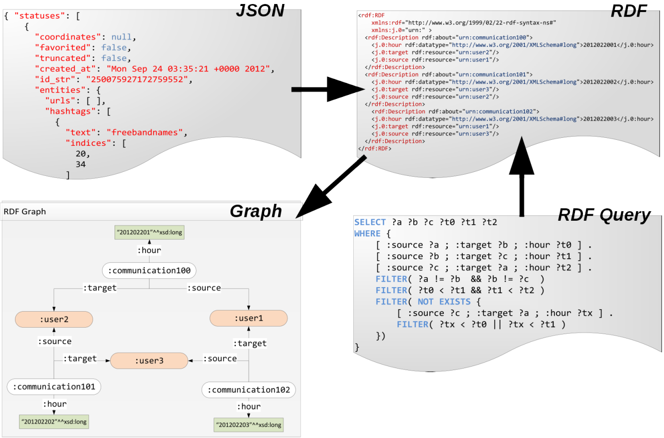
- Liaison: (Noun not Verb)
  - A person (b) who connects party 1 (a) and party 2 (c) through a requested introduction.
  - Like requesting for a first level contact on Linkedin to introduce you to someone in their network
- Not all social networks have a special features like Linkedin, we need to derive this relationship... Time is important!
- Previous methods did not take event sequence into account

Actor (b): Liaison - Logical



For the graph (a,b,c), It will at some time be the case that edge (a,b) exists and It will at some time be the case that edge (b,c) exists and It will at some time be the case that edge (c,a) exists and It has always been the case that edge (c,a) did not exist.

Actor Identification Example: Liaison



Actor Identification Continued

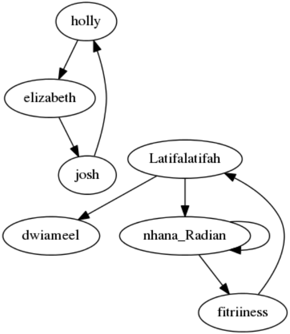
```
There were 2 liaisons found
##########################################################################

 The Following Liaison (b) relationship was found:
-------------------------------------------------------------------------
| (a) is: holly | (b) is: elizabeth | (c) is: josh  |
-------------------------------------------------------------------------

  Slot        Communication          Edge         Hour
--------------------------------------------------------------
|  t0  |  First Communication  | (a) --> (b) | 2012022001  |
|  t1  |  Second Communication | (b) --> (c) | 2012022002  |
|  t2  |  Third Communication  | (c) --> (a) | 2012022003  |
--------------------------------------------------------------


 The Following Liaison (b) relationship was found:
-------------------------------------------------------------------------
| (a) is: Latifalatifah | (b) is: nhana_Radian | (c) is: fitriiness  |
-------------------------------------------------------------------------

  Slot        Communication          Edge         Hour
--------------------------------------------------------------
|  t0  |  First Communication  | (a) --> (b) | 2012021904  |
|  t1  |  Second Communication | (b) --> (c) | 2012021912  |
|  t2  |  Third Communication  | (c) --> (a) | 2012021922  |
--------------------------------------------------------------
```

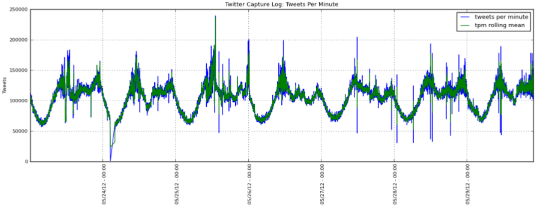


Actor Identification Sample Logics

Actor Type	Logic	
Isolate	$\forall a\,[Isolate(a) \leftrightarrow \mathbf{G}\,[\forall b \neg edge(a,b)]]$	(1)
Connector: Star	$\forall a\,\left(Star(a) \leftrightarrow \neg\exists b\left(cent(b) > cent(a)\right)\right)$	(2)
Connector: Bridge	$\forall b\left(Bridge(b) \leftrightarrow \exists c,e\begin{pmatrix} c \neq e \wedge edge'(b,c) \wedge edge'(b,e) \wedge \\ \forall x\,(edge'(b,x) \rightarrow (x = c \vee x = e)) \wedge \\ cent(b) > cent(c) \wedge cent(b) > cent(e)\end{pmatrix}\right)$	(3)
Connector: Liaison (Prospective)	$\forall a,b,c\,(Liaison(a,b,c) \leftrightarrow \mathbf{F}\,(edge(a,b) \wedge \mathbf{F}\,(edge(b,c) \wedge \mathbf{F}\,(edge(c,a) \wedge \mathbf{H}\neg edge(c,a)))))$	(4)
Connector: Liaison (Retrospective)	$\forall a,b,c\,(Liaison(a,b,c) \leftrightarrow \mathbf{P}\,(edge(c,a) \wedge \mathbf{H}\neg edge(c,a) \wedge \mathbf{P}\,(edge(b,c) \wedge \mathbf{P} edge(a,b))))$	(5)
Maven	$\forall m\,(Maven(m) \leftrightarrow \exists i,g\,\mathbf{F}\,(edge(i,m,msg) \wedge \mathbf{F}\,(edge(g,m) \wedge \mathbf{F}\,(edge(m,g,msg)))))$	(6)
Salesman	$\forall s\,(Salesman(s) \leftrightarrow \exists i,g\,\mathbf{F}\,(edge(i,s,msg) \wedge \mathbf{F}\,(edge(s,g,msg) \wedge \mathbf{H}\neg edge(g,s))))$	(7)

Established Dataset

- In 2012 we collected 165 TB of Twitter Data (Uncompressed)
  - 175 Days Collected, 147 Full Days
    - Estimated 45 Billion Tweets
  - Estimates place total Twitter traffic at 175 million tweets/day-2012
  - Daily collection rates between 50% and 80% of total traffic



Actor Identification Example: Results

- Remember those pretty plots from earilier?
- We take our entire dataset and filter it for 31 days between February 20th and March 20th, and for only #KONY2012 related Tweets

Query	Number of Records
Edges	1,070,910
Isolates	48,060
Liaisons	37,530
Mavens	1,790
Salesmen	391

Approach	Time
Conversion of CSV to RDF using Python	18 sec
RDF file procd. w/Jena (8 thr.)	6.285 min
RDF file procd. w/RDFLib (1 thr.)	13.151 hr
RDF file procd. w/RDFLib (8 thr.)	35.854 min
Serialized CSV–RDF procd. w/RDFLib (1 thr.)	13.159 hr
Serialized CSV–RDF procd. w/RDFLib (8 thr.)	36.762 min

Conclusions

- We aimed to answer the following subset of questions when we started this portion of our work:
  - Can we come up with a way of classifying users based on actor types?
  - Can we determine who the opinion leaders or influencers are?
  - Can we determine how information spreads on these networks?

Future Work

- We have established a more perminant test facility and dataset location in the COSI (Clarkson Open Source Institute)
- We are pursuing the semantic side of social network analysis
  - Currently only one true SNA semantic ontology exists that is openly available and it's only on paper.
  - We are planning on rolling both the actor and event analysis into one approach which will be part of a new ontology
- We have grown our team to include a number of individuals affliated with multiple institutions.
- We recently finished a project using machine learning to process URLs and web-pages on-mass to detect Phishing
- We recently finished a project that analyzed Twitter accounts for duplication, or single ownership

References

[1] Gladwell, M. (2000). "The tipping point". Boston: Little, Brown and Company.

[2] Allen, H. T. (1976). "Communication networks - The hidden organizational chart". The Personnel Administrator, 21(6), 31-35.

[3] Arun Phadke, James Thorp. (1978). "Contracts and Influence". Social Netowrks, 1:1-48

[4] Davis, A., et. al. (1941). "Deep South: A social Anthropological Study of Caste and Class". University of Chicago Press. Chicago, Ill.

[5] Freeman, L. (2004) "The Development of Social Network Analysis: A Study in the Sociology of Science". BookSurge, LLC. North Charleston, SC.

[6] Stanley Wasserman, Katherine Faust. (1994). "Social Network Analysis: Methods and Applications", Structural Analysis in the Social Sciences, 25 November 1994

[7] Donald Triner. (2010). "Publicaly Available Social Media Monitoring and Situational Awareness Initiative," Office of Operations Coordination and Planning: Departmetn of Homeland Security, June 22 2010.

[8] Juris, Jeffrey. (2012). "reflections on #Occupy Everywhere: Social media, public space, and emerging logics of aggregation". American Ethnologist. Vol 39, No. 2, pp. 259-279.

[9] Sheedy, Caroline. (2011). "Social Media for Social Change: A Case Study of Social Media Use in the 2011 Egyptian Revolution". Capstone Project.

[10] Stark, Rodney. (1987). "Deviant Places: A Theory of the Ecology of Crime". Criminology, 25: 893âĂŞ910.

[11] Brett Stone-Gross, et al. (2011). "The underground economy of spam: a botmaster's perspective of coordinating large-scale spam campaigns." In Proceedings of the 4th USENIX conference on Large-scale exploits and emergent threats (LEET'11). USENIX Association, Berkeley, CA, USA, 4-4.

[12] Taylor Dewey, et al. (2012). "The Impact of Social Media on Social Unrest in the Arab Spring". Stanford University - Defense Intelligence Agency Final Report.

[13] Christian Sturm and Hossam Amer. (2013). "The effects of (social) media on revolutions: perspectives from egypt and the arab spring". In Proceedings of the 15th international conference on Human-Computer Interaction: users and contexts of use - Volume Part III (HCI'13), Masaaki Kurosu (Ed.), Vol. Part III. Springer-Verlag, Berlin, Heidelberg, 352-358.

[14] Woods, Richard. (2010). "Privacy is Dead?: Facebook's Mark Zuckerberg says privacy is dead. So why does he want to keeps this picture hidden?". Times Newspapers Ltd.

[15] Statistics Brain. (2013). "'Facebook Statistics". Statistic Brain Research Institute, publishing as Statistic Brain. 6/23/2013. http://www.statisticbrain.com/facebook-statistics/

[16] CBS News. (2012). "Twitter's censorship plan rouses global furor". Associated Press. January 27, 2012

[17] Statistics Brain. (2013). "Twitter Statistics". Statistic Brain Research Institute, publishing as Statistic Brain. 5/7/2013. http://www.statisticbrain.com/twitter-statistics/

[18] Bagley, Nick. (2012). "The Decline of Myspace: Future of Social Media". Dreamgrow Digital. 8/13/2012. http://www.dreamgrow.com/the-decline-of-myspace-future-of-social-media/

[19] alton, Antony, "Temporal Logic", The Stanford Encyclopedia of Philosophy (Fall 2008 Edition), Edward N. Zalta (ed.)

[20] Shea Bennett. "Just How Big Is twitter In 2012 [INFOGRAPHIC]". All Twitter - The Unofficial Twitter Resource, February 2013

[21] Mallon, Shanna. (2012). "50 Facts about Social Media for Business". Straight North, LLC publishing as The Straight North Blog. Downers Grove, IL.

[22] D. Karaiskos, et. al. (2010) "Social network addiction : a new clinical disorder?". European psychiatry : the journal of the Association of European Psychiatrists. volume 25, Page 855. DOI: 10.1016/S0924-9338(10)70846-4)

[23] Helms, R, Ignacio, et al.(2010) "Limitations of Network Analysis for Studying Efficiency and Effectiveness of Knowledge Sharing" Electronic Journal of Knowledge Management Volume 8 Issue 1 (pp53 - 68)

[24] Dhar, Vasant. (2013) "Data Science and Prediction". Communications of the ACM. Vol. 56 No 12, Pages 64-73. 10.1145/2500499

[25] Sullivan, Danny. (2011). "Why Second Chance Tweets MAtter: After 3 Hours, Few Care About Socially Shared Links". Thrid Door Media Inc. Publishing as Search Engine Land.

[26] Travers J., Milgram S. (1969) "An Experimental Study of the Small World Problem," Sociometry, Vol. 32, No. 4. pp. 425-443, doi:10.2307/2786545

[27] Harrist, A. W., Zaia, A. F., Bates, J. E., Dodge, K. A. and Pettit, G. S. (1997). "Subtypes of Social Withdrawal in Early Childhood: Sociometric Status and Social-Cognitive Differences across Four Years". Child Development, 68: 278âĂŞ294. doi: 10.1111/j.1467-8624.1997.tb01940.x

[28] Taylor, J. (2013). "Personal communication". August 12, 2013.

[29] Galton, Antony. (2008). "Temporal Logic". The Stanford Encyclopedia of Philosophy. Edward N. Zalta (ed.). URL = http://plato.stanford.edu/archives/fall2008/entries/logic-temporal/.

[30] Minker, Jack. (1982). "On indefinite databases and the closed world assumption". Lecture Notes in Computer Science. 6th Conference on Automated Deduction. Springer Berlind Heidelberg. pp. 292-308 doi=10.1007.BFb0000066

[31] Jeremy J. Carroll, Ian Dickinson, Chris Dollin, Dave Reynolds, Andy Seaborne, and Kevin Wilkinson. (2004). "Jena: implementing the semantic web recommendations," In Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters (WWW Alt.'04). ACM, New York, NY, USA, 74-83. DOI=10.1145/1013367.1013381

[32] Claudio Gutierrez, et al. (2005) "Temporal RDF". In Proceedings of the Second European conference on The Semantic Web: research and Applications (ESWC'05), Asuncion Gomez-Perez and Jerome Euzenat (Eds.). Springer-Verlag, Berlin, Heidelberg, 93-107.

[33] Andrew Page.(2012). "Know Your Meme: Kony 2012". http://www.knowyourmeme.com/memes/events/kony-2012

[34] Goutam Kumar Saha. 2007. "Web ontology language (OWL) and semantic web." Ubiquity 2007, September, Article 1 (September 2007), 1 pages. DOI=10.1145/1295289.1295290 http://doi.acm.org/10.1145/1295289.1295290

[35] John Guare, "Six Degrees of Seperation," A Play, May 1990

[36] Lada Adamic, et al. (2003). "A social network caught in the Web," First monday, 8(6)

[37] David Liben-Nowel, et al. (2005). "Geographic Routing in Social Networks," Proceedings of the National Academy of Sciences (PNAS), 102:11623-1162, 2005

[38] Ravi Kumar, et al. (2006). "Structure and Evolution of Online Social Networks," In the Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mininig (KDD;06), Philadelphia, PA.

[39] Michelle Girvan, Mark Newman. (2002). "Community structure in social and biological networks," Proceedings of the National Academy of Sciences (PNAS), 99(12):7821-7826.

[40] Ceren Budak, et al. (2010). "Where the blogs tip: connectors, mavens, salesmen and translators of the blogosphere". In Proceedings of the First Workshop on Social Media Analytics (SOMA '10). ACM, New York, NY, USA, 106-114. DOI=10.1145/1964858.1964873

[41] Steven Levitt, Stephen J. Dubner. (2005) "Freakonomics: A Rogue Economist Explores the Hidden Side of Everything," New York: Morrow-Harper.

[42] George Kelling, Catherine Coles. (1998). "Fixing Broken Windows: Restoring Order and Reducing Crime in Our Communities," January 20, 1998

[43] Roe v. Wade, 410 U.S. 113 (1973)

[44] Jonah Beger. (2013). "Contagious: Why Things Catch On," Simon and Schuster Publishing, March 5, 2013

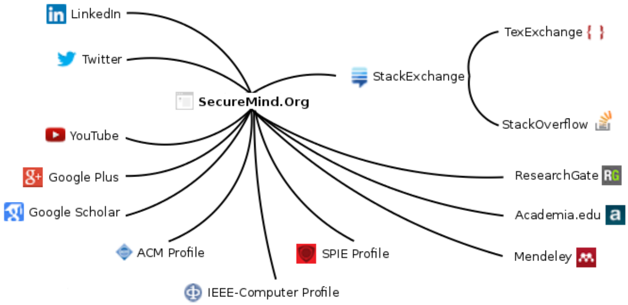
[45] R. S. Renfro. (2001). "Modeling and Analysis of Social Networks',' PhD thesis, Air Force Institute of Technology.

[46] C. Clark. (2005). "Modeling and analysis of clandestine networks," Masters thesis, Air Force Institute of Technology.

[47] J. T. Hamill. (2006). "Analysis of Layered Social Networks," PhD thesis, Air Force Institute of Technology.

[48] G. Ereteo , F. Gandon, M. Buffa, O. Corby. (2009) "Semantic Social Network Analysis," Proceedings of the WebSciâĂŽ09. http://journal.webscience.org/141/

Contact



LinkedIn
Twitter
SecureMind.Org
StackExchange
TexExchange
StackOverflow
YouTube
Google Plus
Google Scholar
ACM Profile
IEEE-Computer Profile
SPIE Profile
ResearchGate
Academia.edu
Mendeley

Questions

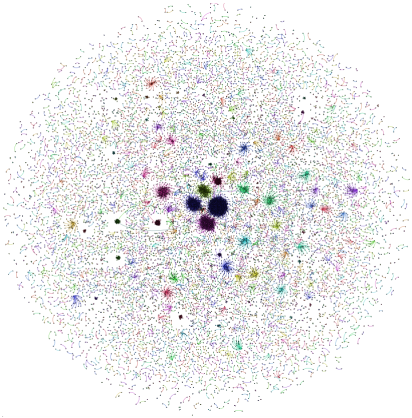
Questions?

Suplimental Material

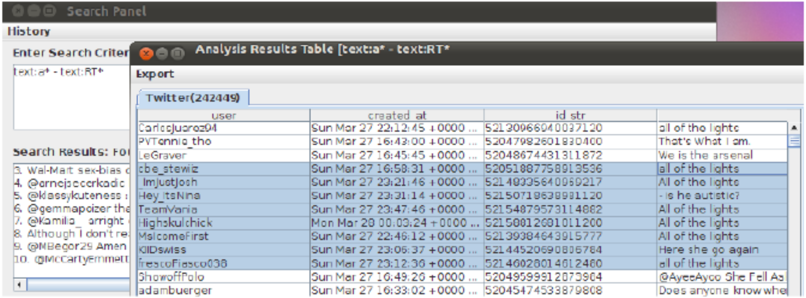
- **Twitter JSON Key Fields**

profile_link_color	Coordinates	verified
In_reply_to_screen_name	Geo	time_zone
In_reply_to_status_id	text	statuses_count
In_reply_to_status_id_str	entities	Contributors
In_reply_to_user_id	place	protected
profile_background_color	contributors_enabled	trunkated
profile_background_title	default_profile	retweeted
default_profile_image	description	id_translator
follow_request_sent	followers_count	location
friends_count	geo_endabled	favorites_count
profile_image_url_https	listed_count	following
profile_background_image_url	notifications	retweet_count
background_image_url_https	name	created_at
profile_image_url	lang	Favorited
sidebar_border_color	use_background_image	Id_str
sidebar_fill_color	screen_name	Created_at
profile_text_color	show_all_inline_media	Id
url	utc_offset	

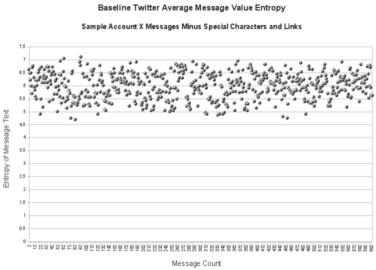
- **BEK Infectious Account Visualization**

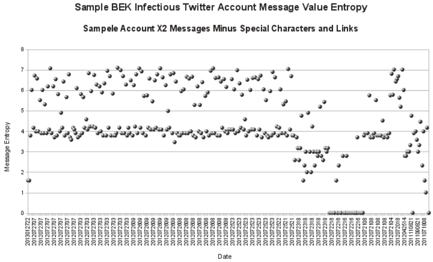


- **Coalmine User Interface**

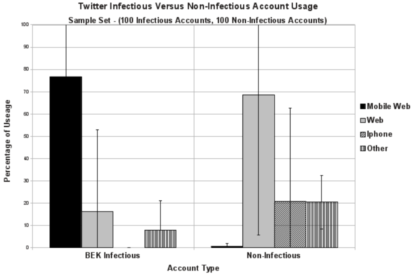


- **Malware Infection Vector Detection Continued**





- **Malware Infection Vector Detection Continued**

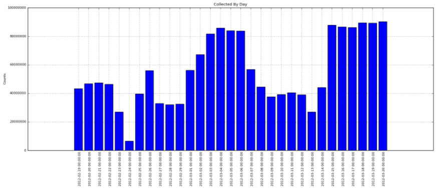


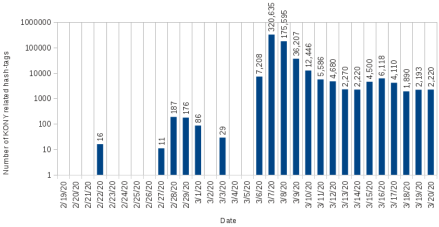
Total Tweets Processed	6,531,319,202
Total Number of Unique Accounts	265,163,290
Number of Suspicious Accounts	729,609
Total Number of Suspicious Tweets	8,286,480
Calculated Percentage of Infectious Accounts	0.000275
Calculated Percentage of Infectious Tweets	0.127
Dataset Processing Time with Regex	22H 48M
Dataset Processing Time w/Fig. 5.10 filter	23H 21M

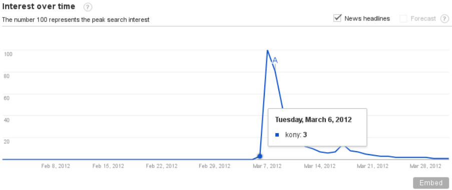
Event Identification

- Still in the initial stages of this part of our work
- Given a general topic, “search term, hashtag,” we can identify most of the related content from the dataset
- We have a means for alerting on all new posts regarding that term
- We can dig historically through the data and trace the path that an itea took
- We can identify the influential individuals, “accounts,” that played a part in the information spread
- Our test case was the KONY2012 Event

Event Identification Continued







Event Identification Continued

- Top 10 Twitter Accounts, sending and receiving KONY2012 related Tweets

Directed @ Account Names	In-Degree	Origin Account Names	Out-Degree
tothekidswho	625	twittonpeace	47
Invisible	125	interhabernet	44
youtube	118	DailyisOut	44
helpspreadthis	95	MEDYA_TURK	42
justinbieber	83	haber_42	35
prettypinkprobz	48	gundem_haber	30
ninadobrev	48	twittofpeace	22
MeekMill	47	korkmazhaber	19
ladygaga	43	tarafsiz_haber	14
KendallJenner	39	Son_DakikaHaber	13

Event Identification Continued

- Top 10 Twitter Accounts, retweeting and being retweeted regarding KONY2012

Retweeting Accounts	In-Degree	Message Source	Out-Degree
MedyaKonya	8	Stop_____Kony	2642
twittonpeace	8	tothekidswho	753
haber_42	7	konyfamous2012	716
gundem_haber	7	Kony2012Help	615
korkmazhaber	7	stop________kony	353
DailyisOut	7	WESTOPKONY	225
interhabernet	6	zaynmalik	221
KONYA_ZAMAN	6	iSayStopKony	127
konya_time	6	Stop_2012_Kony	80
konyagazetesi	5	Kony_Awareness	72